

Active contours and prior knowledge for change analysis: application to digital urban building map updating from optical high resolution remote sensing images

Dissertation Submitted to

Institute of Automation, Chinese Academy of Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Engineering

by

Timothée BAILLOEUL

Pattern Recognition and Intelligent Systems

Dissertation Supervisors: Baogang Hu, Philippe Marthon

Acknowledgments

Firstly, I would like to thank Baogang Hu and Marc Jaeger for having welcomed me at LIAMA and for all the care they have taken to make my stay at the laboratory as pleasant as possible.

I am grateful to Véronique Prinet and Jean-Guy Planès for having proposed this thesis subject to me and for the confidence they have placed in me. Thank you to Véronique for having supervised my research works, for the freedom she left me and for always having been present at the delicate moments along this PhD journey.

My thanks also go out to Bruno Serra and Philippe Marthon who have also followed up my works. Thank you to Philippe for his remarks, constructive criticism and for having received me in his research unit. I extend all my warmth to Bruno for having been an active, determined and effective correspondent at Alcatel Alenia Space. I would like to thank him for his support and kindness.

I very warmly thank the members of the jury for the time that they are going to spend reading this document.

I thank Josiane Zérubia and all her team for having received me for several months within the Ariana project. I would particularly like to thank Ian Jermyn and Marie Rochery for their assistance and advice on quadratic models. Thank you to Ian for his patience, our numerous discussions and the interest he has shown in my works.

I am also grateful to Pinxiang Chen and Haitao Zhang and Sun Lebing from the Beijing Institute of Surveying and Mapping for having lent us the maps and aerial images used in this study. I would like to very sincerely thank Haitao Zhang for his availability and diligent help.

Thank you to Véronique Prinet, Philippe Marthon, Martine Lacoste, Baogang Hu, Bizhen Hong and Serge Lefeuvre for having helped me overcome the difficulties putting in place the first co-tutelage between the Academy of Sciences of China and the INPT.

My thanks also go out to Jean-François Barczi for his good humor, and for having ceaselessly answered my questions in the area of Computer Sciences.

I warmly thank all the students and interns at LIAMA who contribute to the labo-

ratory's friendly atmosphere. I am thinking in particular of Jinghui Duan and Fengfeng Tang who have always been prompt to solve my problems with the Chinese language. I extend all my friendship to Wanlin Zhu for our weekly discussions on active contours and our frenzied games of badminton. Thanks to all of you for your efficient revision of the thesis abstract in Chinese.

Lastly, a big thank you to my French and Chinese family for their tireless support. Thank you to my parents who have always encouraged me from afar (in the cartographic sense of the term): I thank my mother for the many times she has proofread the manuscript and my father for his inexhaustible curiosity in my scientific activities. Thank you to Jing, my wife, for her optimism and encouragements that have been an essential comfort during the years spent on this doctoral study.

Abstract

With the commercial advent of optical satellite imagery with sub-meter resolution (Ikonos, Quickbird), a higher frequency of map revision and monitoring is affordable and realistic at a large scale. Automatic object recognition and change detection from remote sensing data has been studied for more than thirty years by the Image Processing and Computer Vision communities, yielding remarkable results over rural and suburban areas. Dense urban areas are however more difficult to analyze at a high resolution: the presence of shadows, occlusions and details profusion make their interpretation difficult.

In the present work, we tackle the issue of urban building map updating from a single very high resolution (VHR) Quickbird image merged with a Digital Surface Model (DSM) encoding the altitude of the scene. In particular, we address the problem of removing exogenous variabilities between the map and imagery data (map-to-image registration inaccuracies, simplified or mistaken objects in the map) to achieve a final reliable change detection. To achieve such a goal, we make use of active contours segmenting techniques to accurately match symbolized cartographic objects to their counterpart representations in the image. We investigate how well these deformable models can deal with urban artifacts while embedding prior shape knowledge derived from the map as well as collateral sources of information (DSM). Besides a literature review and a problem statement analysis exposed in the first part of the thesis, the major achievements and contributions described in the next chapters can be summarized as follows:

1. **Camera calibration parameters optimization.**

We propose an optimization scheme to accurately retrieve interior/exterior camera orientation parameters of aerial images. This method estimates the parameters a posteriori and uses the redundancy of n stereo-pair images. A non-linear cost function built from the camera model is minimized by a Simplex scheme using ground control points and tie points gathered from all images. Experimental results show the effectiveness of the approach to refine the absolute (geocoding) and relative (images consistency) precision of the parameters. Pairwise computed DSMs were subsequently geocoded in the same cartographic referential system as other data, orthorectified and mosaicked.

2. **Digital map refinement from shape constrained active contours.**

- (a) A novel method based on active contours is proposed to match each building

symbolized in the map to its counterpart representation in the satellite image. Geometric and extrinsic shape prior knowledge is derived from the digital map to regularize the active contour: its shape is constrained to be akin to the one of the cartographic building which is supposed to be matched in the image. We experimentally show the effectiveness of this method to overcome urban artifacts. Comparative studies are carried out with region-based (Mumford-Shah, Bayesian models) and edge-based (Gradient Vector Flow) active contours.

- (b) The proposed shape penalty is invariant from a global transformation. Invariance is traditionally achieved by an optimization sub-procedure based on the gradient descent scheme. We propose an alternative based on the Simplex algorithm which is showed to be more robust and to outperform the gradient descent.
- (c) As active contours schemes are sensitive to local minima of the energy they minimize, we propose to add in the functional a new data term related to the DSM in complement to the satellite image. As a building is more discriminated from the background in the DSM than in the image, it drives the curve to the target more robustly. Results show the lower sensitivity to remote initialization and to local minima of the segmenting curve in this case.
- (d) Finally, we propose a new soften shape prior incorporation to address the issue of local minima without the support of the exogenous DSM. Based on a spatio-temporal weighting technique, this approach allows relaxation and reinforcement of the prior shape penalty during the active contour convergence process. The comparison with the traditional shape prior incorporation confirms the potential of the proposed scheme to alleviate the sensitivity to local minima.

3. Digital map correction using a new shape energy.

As a map is an abstracted representation of the reality, it may embody simplification and generalization effects which introduce exogenous variabilities between the map and the image. Besides, a map is usually achieved manually and may contain locally mistaken objects. Since the prior shape derived from the corrupted map constrains the active contour in a hard fashion, it is impossible to overcome these variabilities which then bias the matching process. We address this issue with a new shape prior energy allowing parallel discrepancies of the active contour from the prior shape. This model includes a new formulation of quadratic active contour models (Rochery *et al.*) to enhance corners and straight lines of the building to be segmented in the satellite image.

4. Map-to-image change detection from data fusion and Hough voting.

Simple change indicators computed from the DSM and a multispectral image are merged to detect obvious changes between the map and the more recent remote sensing data. In case no obvious change is detected, map-to-image exogenous variabilities are alleviated by the active contour matching process. An edge-based Hough voting approach is then proposed to confirm the “changed” or “unchanged”

status of each matched building: correspondences between segments of the refined cartographic building and segments primitives extracted from the image are evaluated and accumulated in a Hough voting space. The Hough voting score as well as the geometric variation due to the active contours matching are integrated in a stochastic framework to evaluate a “no-change” probability for each cartographic object.

Experimental results are showed with a Quickbird satellite image (0.6 m/pixel) and 1:10,000 scale cartographic data provided by the Beijing Institute of Surveying and Mapping.

KEY WORDS: Remote sensing images, GIS, high resolution, urban, satellite optical sensors, active contours, prior shape constraint, optimization, change detection, map updating, DSM, fusion.

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 1 |
| 1.1 | Application context and the need for updating urban maps | 1 |
| 1.2 | Scientific goal and approach adopted | 2 |
| 1.3 | Innovations | 3 |
| 1.4 | Thesis framework | 4 |
| 1.5 | Context of the PhD study | 5 |
| 2 | Analysis and formulation of the problem | 6 |
| 2.1 | Introduction | 6 |
| 2.2 | State of the art relative to scene analysis and map updating from high-resolution remote sensing images | 6 |
| 2.2.1 | Specific aspects and strategies for map updating | 6 |
| 2.2.1.1 | Comparing images | 6 |
| 2.2.1.2 | Map/image comparison | 7 |
| 2.2.2 | Some works on remote sensing data fusion for scene analysis at high-resolution | 9 |
| 2.2.2.1 | Fusion of optical images | 9 |
| 2.2.2.2 | Optical and altimetric data fusion | 10 |
| 2.2.3 | Utilization of constraints derived from prior knowledge | 10 |
| 2.2.3.1 | Generic knowledge | 10 |
| 2.2.3.2 | Specific knowledge | 12 |
| 2.2.4 | Summary | 15 |
| 2.3 | Position of the problem of map updating | 15 |
| 2.3.1 | General conceptual framework | 16 |
| 2.3.1.1 | Definition of the remote sensing related universes | 16 |
| 2.3.1.2 | Definition of $U_{interpretation}$ | 17 |
| 2.3.2 | Problem addressed in the thesis | 18 |
| 2.3.2.1 | map-to-image change analysis considered in the thesis | 18 |
| 2.3.2.2 | Difficulties with urban environments | 20 |
| 2.4 | Proposed methodology | 21 |
| 2.5 | Conclusion | 23 |
| 3 | Generating an orthoscopic digital surface model | 24 |
| 3.1 | Introduction | 24 |
| 3.2 | Photogrammetric stereoscopy | 24 |
| 3.2.1 | Stereoscopy - reminder and general principle | 25 |

| | | |
|----------|--|-----------|
| 3.2.2 | Automated matching of images using adaptive correlation masks | 28 |
| 3.2.2.1 | Resemblance criterion by crossed correlation | 28 |
| 3.2.2.2 | Correlation by adaptive masks | 31 |
| 3.2.3 | Principle of the reconstruction chain used | 34 |
| 3.3 | Optimizing the image acquisition parameters | 35 |
| 3.3.1 | Image acquisition parameters | 35 |
| 3.3.1.1 | Calculation of the affinity parameters: internal orientation | 37 |
| 3.3.1.2 | Assessment of the quality of the image acquisition parameters | 38 |
| 3.3.2 | Overall optimization | 40 |
| 3.3.2.1 | Cost function | 40 |
| 3.3.2.2 | Geo-centered Cartesian reference frame | 41 |
| 3.3.2.3 | Results | 43 |
| 3.3.3 | Assessment of the geocoding quality of the DSMs | 46 |
| 3.4 | Orthorectification of an epipolar DSM | 48 |
| 3.4.1 | Orthorectification by the direct method | 48 |
| 3.4.1.1 | Triangular interpolation | 49 |
| 3.4.1.2 | Algorithm | 50 |
| 3.4.1.3 | Results | 51 |
| 3.4.2 | Pre-processing on the disparity image | 53 |
| 3.4.2.1 | Median filtering | 53 |
| 3.4.2.2 | Rejection filter | 54 |
| 3.4.2.3 | Summary | 55 |
| 3.5 | Merging DSMs | 56 |
| 3.6 | Conclusion | 59 |
| 4 | Presentation and choice of shape constrained active contours for map-to-image fine matching | 61 |
| 4.1 | Introduction | 61 |
| 4.2 | State of the art - Prior shape constraints in active contours | 62 |
| 4.2.1 | Active contour framework | 62 |
| 4.2.1.1 | Representation of an active contour | 62 |
| 4.2.1.2 | Geometric and variational approach | 67 |
| 4.2.1.3 | Edge-based active contours | 68 |
| 4.2.1.4 | Region-based active contours | 72 |
| 4.2.2 | Segmentation with specific shape constraint | 77 |
| 4.2.2.1 | Non statistical approach | 78 |
| 4.2.2.2 | Statistical approach | 80 |
| 4.2.2.3 | Shape constraint invariance from geometrical transformation | 81 |
| 4.2.3 | Segmentation with a generic shape constraint: quadratic models | 82 |
| 4.3 | Active contours for fine map-to-image matching: choice and implementation | 83 |
| 4.3.1 | Choice of the active contours representation | 83 |
| 4.3.2 | Choice of a term for the dependence on data | 84 |
| 4.3.2.1 | Region-based models | 84 |

| | | |
|----------|---|------------|
| 4.3.2.2 | Edge-based model | 85 |
| 4.3.3 | Choice of the shape constraint | 86 |
| 4.3.4 | Algorithm and optimization | 87 |
| 4.3.4.1 | Reducing calculation time: use of the Fast Narrow Band method | 87 |
| 4.3.4.2 | Evolution diagram of the shape constrained active contour | 90 |
| 4.4 | Conclusion | 91 |
| 5 | Robust shape constrained active contours | 93 |
| 5.1 | Introduction | 93 |
| 5.2 | Robust estimation of the T_{sim} parameters by the simplex algorithm . . | 93 |
| 5.3 | Fusion of exogenous terms in the attachment to the data energy functional | 95 |
| 5.4 | More flexible incorporation of the prior shape: spatio-temporal flexibility of λ | 99 |
| 5.5 | Application and analysis of the experimental results | 102 |
| 5.5.1 | Experimental protocol | 102 |
| 5.5.2 | Simplex vs gradient descent comparison for estimating T_{sim} pa- rameters | 102 |
| 5.5.3 | Contribution of the prior shape knowledge for segmentation . . . | 104 |
| 5.5.4 | Sensitivity to the initialization of the shape constrained active contours | 107 |
| 5.5.4.1 | Comparison of the region-based models | 107 |
| 5.5.4.2 | Edge-based model | 116 |
| 5.5.5 | Solutions for a lower sensitivity to the initialization and to the local minima | 119 |
| 5.5.5.1 | Results with fusion of the exogenous data into the at- tachment to data term | 119 |
| 5.5.5.2 | Results with a spatio-temporal shape constraint | 123 |
| 5.6 | Limits of the model with shape constraint | 129 |
| 5.7 | Conclusion | 131 |
| 6 | Shape constraint authorizing parallel variations | 133 |
| 6.1 | Introduction | 133 |
| 6.2 | Shape constraint energy authorizing parallel variations | 133 |
| 6.2.1 | Model | 133 |
| 6.3 | Insertion of generic prior knowledge by quadratic models | 134 |
| 6.3.1 | Corrective quadratic energy | 135 |
| 6.3.1.1 | Model | 135 |
| 6.3.1.2 | Energy gradient | 136 |
| 6.3.2 | Algorithm | 137 |
| 6.3.2.1 | Calculation of the quadratic force | 138 |
| 6.3.2.2 | Extension of the force | 138 |
| 6.3.3 | Implementation difficulties | 139 |
| 6.4 | Experiments and analysis of the results | 141 |
| 6.4.1 | Results without quadratic correction | 141 |
| 6.4.2 | Results with quadratic correction | 144 |

| | | |
|----------|---|------------|
| 6.5 | Conclusion | 147 |
| 7 | Map-to-image change analysis | 148 |
| 7.1 | Introduction | 148 |
| 7.2 | Methodology | 148 |
| 7.3 | Calculation and fusion of simple change indices | 150 |
| 7.3.1 | Index calculated from the altimetric data (DSM) | 151 |
| 7.3.1.1 | Limitations of the map-to-DSM comparison | 151 |
| 7.3.1.2 | Approach | 151 |
| 7.3.1.3 | Experimental results | 153 |
| 7.3.2 | Index calculated from a multispectral image | 157 |
| 7.3.3 | Change indicators fusion | 160 |
| 7.4 | Map-to-image change analysis | 161 |
| 7.4.1 | Matching map and image segments by Hough voting | 161 |
| 7.4.1.1 | Principle | 161 |
| 7.4.1.2 | Algorithm | 162 |
| 7.4.1.3 | Energy associated with the Hough voting | 163 |
| 7.4.2 | Measurement of geometric variation due to active contours matching | 164 |
| 7.4.3 | Probability of non-change | 165 |
| 7.5 | Experimental results | 166 |
| 7.5.1 | Experimental protocol | 166 |
| 7.5.2 | Study of some cases of change/non-change | 167 |
| 7.5.2.1 | Change detection | 167 |
| 7.5.2.2 | Detection of non-change | 169 |
| 7.5.2.3 | Ambiguity and false alarm with a map marred by local errors | 176 |
| 7.5.2.4 | Summary of results | 177 |
| 7.5.3 | Statistical analysis | 178 |
| 7.5.3.1 | Results with a good quality map | 179 |
| 7.5.3.2 | Results with a medium quality map | 180 |
| 7.6 | Conclusion | 181 |
| 8 | Conclusion | 183 |
| 8.1 | Summary | 183 |
| 8.2 | Discussion of the limitations | 185 |
| 8.2.1 | Limits of the shape constraint model | 185 |
| 8.2.2 | Limits of the change analysis | 186 |
| 8.2.3 | Limits of the active contours | 186 |
| 8.3 | Perspectives | 186 |
| A | Data presentation and pre-processing | 188 |
| A.1 | Data presentation | 188 |
| A.1.1 | Remote sensing data | 188 |
| A.1.1.1 | Aerial images | 188 |
| A.1.1.2 | Satellite images | 190 |
| A.1.2 | Cartographic data | 190 |
| A.2 | Pre-processing of the remote sensing data | 191 |

| | | |
|----------|--|------------|
| A.2.1 | Subsampling of the aerial images | 191 |
| A.2.2 | Rectification of the satellite images | 193 |
| B | Simplex algorithm | 194 |
| C | Direct and inverse localization modeling | 197 |
| C.1 | Inverse modeling | 197 |
| C.2 | Direct modeling | 198 |
| D | Fast calculation of the mean and of the variance inside and outside an active contour represented by a levels set | 201 |
| E | Calculation of the shape constraint and quadratic energy variations | 203 |
| E.1 | Evolution equation derived from the shape constraint functional by difference of normals | 203 |
| E.2 | Derivative of the quadratic correction energy | 205 |
| E.2.0.1 | Energy $J_1(C)$ | 205 |
| E.2.0.2 | Energy $J_2(C)$ | 206 |
| E.2.1 | Total variation of the energy $J_{correction}(C)$ | 207 |
| F | 3D representation of an urban scene | 208 |
| F.1 | Objective and methodology | 208 |
| F.2 | Modeling principle | 208 |
| F.3 | Results over Beijing city | 208 |
| | Bibliography | 210 |

Notations and conventions

Typography

- * A scalar variable is noted in italic light face: x
- * A column vector is noted in bold straight lower case: \mathbf{x}
- * A row vector is a transposed column vector: \mathbf{x}^T
- * A matrix is noted in bold straight upper case: \mathbf{M}

Notations

| | |
|--|---|
| $\langle \mathbf{x}, \mathbf{y} \rangle$ | inner product of two vectors \mathbf{x} and \mathbf{y} |
| $ x $ or $ \mathbf{x} $ | modulus of a scalar value or Euclidean norm of a vector |
| \overline{AB} | length of the segment $[AB]$ where A and B are two points of the Euclidean space |
| ∇f | gradient of a scalar function |
| $\operatorname{div}(\mathbf{x})$ | divergence of a vector |
| Δf | Laplacian of a scalar function |
| $\min(a, b)$ | minimum between two real numbers |
| $\max(a, b)$ | maximum between two real numbers |
| $\operatorname{sign}(a)$ | Characteristic function of the real argument a . It is equal to +1 if a is strictly positive; -1 if a is strictly negative; 0 if a is null |
| $H(\cdot)$ | Heaviside step function |
| $\delta(\cdot)$ | Dirac distribution |
| $H_a(\cdot), \delta_a(\cdot)$ | Derivable and regular approximations of $H(\cdot)$ and $\delta(\cdot)$ |
| $g(x)$ | function with real values asymptotically tending towards zero. $g: \mathbb{R}^+ \rightarrow \mathbb{R}$ $g(x) = \frac{1}{1+x^p}$, with $p \in \mathbb{N}$, typically $p = 1, 2$ |
| $I(c, r)$ | value of the image I in column c and row r |

Acronyms

- * MAP: Maximum *A Posteriori*
- * PCA: Principal Component Analysis
- * MDL: Minimum Description Length

- * **GIS**: Geographic Information System
- * **BISM**: Beijing Institute of Surveying and Mapping
- * **DSM**: Digital Surface Model
- * **LIDAR**: LIght Detection And Ranging
- * **UTM**: Universal Transverse Mercator
- * **GPS**: Global Positioning System
- * **RGB**: Red Green Blue
- * **HSI**: Hue Saturation Intensity
- * **RANSAC**: RANdom SAmple Consensus
- * **IGN**: Institut G  ographique National (France)
- * **BDTopo**: Base de Donn  es Topographiques, IGN product
- * **PPS**: Principal Point of Symmetry
- * **BNS**: Beijing National System
- * **SAR**: Synthetic Aperture Radar
- * **GVF**: Gradient Vector Flow
- * **NDVI** : Normalized Difference Vegetation Index

Chapter 1

Introduction

1.1 Application context and the need for updating urban maps

For some years, remote sensing optical images with a sub-metric resolution in the civil sector have no longer been the prerogative of aerial images. Satellite imaging, as shown by the Quickbird and Ikonos programs, has now reached similar resolutions with the advantages of lower costs and a higher acquisition frequency. It is an unprecedented advantage for the regular maintenance of cartographic data on a large scale. It is essential to have up-to-date maps for many applications. For example, it is essential to have frequent updates on the urban fabric for planning the sustainable development of a city in the long term. A reliable and recent representation of the road network makes it possible to find solutions to the problems of road traffic; knowing the location of residential areas with precision makes it possible to optimize the transport services in poorly served areas and to know where to develop new housing projects. Urban pollution, whether atmospheric or sound-related, can be better understood and better controlled thanks to a map enhanced with altitude information relative to the buildings. In a very different context, that of natural (earthquakes, flooding) or less natural (military conflicts) catastrophes, the quick production of maps of the damage - in terms of infrastructure and buildings destroyed - is a key element for providing the population with rapid relief and allow efficient reconstruction. We can then understand that above all a frequently updated map represents a decision-support tool whether it has to be used in an emergency or for long-term planning purposes.

Whereas the need for reliable and up-to-date maps is growing and the flow of remote sensing images is increasing, the bottleneck in the production of cartographic data lies in the manual processing applied to the data. Map maintenance is traditionally carried out either by surveys in the field, or by operators detecting changes between an old database and more recent remote sensing data. In both cases, this process is slow and therefore costly, and consequently does not make it possible to meet the demand. An urgent need to automate the photo-interpretation process, calling on computerized image understanding techniques, has made itself felt. Over the last twenty years a large number of works have been carried out aiming to facilitate map updating or to attempt

to automate it. But the fact is that the utilization of imagery as the sole source of information for maintaining cartographic data has failed.

1.2 Scientific goal and approach adopted

In this thesis we address the problem of updating cartographic data concerning urban constructions using high-resolution optical satellite images (0.6 m/pixel) and auxiliary geospatial data, in particular altimetric data of the DSM type (Digital Surface Model). The goal of this research work consists of studying and proposing a methodology for analyzing changes between a symbolic representation of buildings in an urban scene (digital map) and more recent remote sensing data. In this study, the analysis of changes concerns the verification of the presence or disappearance in an image of a building symbolized on an older digital map.

The comparison between a map and a high-resolution optical image presents two major difficulties. The first one concerns the automated or semi-automated interpretation of the image of an urban landscape: the profusion of detail, the presence of shadows, the low contrast of objects and other artifacts, are all obstacles that cause the traditional image processing and pattern recognition approaches to fail. The second one is due to the errors and simplifications contained in the map which introduce variabilities - which we will call *exogenous* - between the semantic data and the image. These exogenous variabilities are specific to each object and are liable to be interpreted, wrongly, as an effective change.

With a view to overcoming these problems we propose a methodology based on data fusion and the utilization of prior knowledge. The merging of geospatial data makes it possible to take advantage of the complementarity of their information with the goal of ensuring greater robustness for the processing. As for the prior knowledge of the position and shape - provided by the digital map - it makes it possible to restrict the search space to the objects to be found in the image for a comparison with the map.

Consequently, we propose an approach based on the following principles:

- * Detect the obvious changes between the map and the remote sensing data using a simple method. A detection of obvious changes of buildings in the map will make the most of the fusion of an altimetric data (DSM) and a remote sensing multispectral image. Since these changes are free of ambiguities, no complex processing of the images is required. We will therefore use tried and tested types of processing acknowledged by the remote sensing and image processing communities.
- * Develop a more sophisticated method for the more complex cases of change/non-change. By complex situations we mean the cases where the information in the map is partially erroneous or when the image is altered by the artifacts of the urban environment. Under these circumstances, the detection of change cannot be carried out by means of simple processing methods. To achieve this, we have adopted a two-step approach:

- ↪ Making a correction to the exogenous variabilities between the map and a panchromatic satellite image using the shape constrained active contours technique. Active contours offer a theoretical framework that is particularly well-suited to the integration of prior shape knowledge and to the merging of data. Furthermore, they are perfectly suited to the integration of vector (digital map) and raster (image) data. However, the constrained active contour models proposed in the literature do not fully answer the problems posed by our application. In particular, the profusion of detail in very high-resolution images in an urban region tends to multiply the occurrence of local minima of the energy functional minimized by the active contours. The presence of local errors in the prior shape derived from the map also limits the technique's matching capability. One of the goals of this thesis will be to overcome the above-mentioned problems by proposing new solutions for a more robust and faster convergence of the contour with, at the same time, a greater flexibility with respect to the shape known beforehand.
- ↪ Analyzing the changes and non-changes by comparing the cartographic object refined by the active contours with the panchromatic image. This analysis aims to produce a change indicator whose reliability is improved by the fine matching carried out by the active contours.

1.3 Innovations

The main innovations of this work concern the active contour models. The first originality concerns the adaptation of the constrained active contours to the problem of map-to-image matching and the analysis of changes between a map and high-resolution remote sensing optical data.

Our other contributions are relative to the improved robustness of these models with respect to the local minima. We propose two solutions to compensate for this problem: one concerns the merging of a DSM into the attachment-to-data term in addition to the analyzed image. The better discrimination of the constructions in the DSM makes the active contour less sensitive to the initialization and shortens the convergence time. The second solution consists of spatially and dynamically relaxing the shape constraint during the convergence of the active contour. This then reaches an approximate solution of the object to be matched before the shape constraint is uniformly restored in order to regularize the contour.

The most innovative contribution of this thesis is the creation of new shape constraint energies enabling local and parallel discrepancies of the active contour with respect to the prior shape. The presence of the cartographic objects' local shape errors has led us to create these types of model.

Lastly, we have taken advantage of the robustness of the simplex optimization algorithm in circumstances where its use has not until now been explored. In particular, its integration in the shape constraint on the active contours will be found to have better performances than the gradient descent technique conventionally used.

1.4 Thesis framework

To begin with, chapter 2 provides a brief state of the art concerning the updating of cartographic data and the extraction of objects from high-resolution remote sensing data. We will see that the utilization of specific prior knowledge extracted from the map and data fusion are the current trends used to facilitate the analysis of scenes. We then precisely describe the problem addressed in this thesis, its difficulties, and the approach that we are proposing to successfully carry out an analysis of map/image changes. Our methodology relies on the utilization of prior knowledge and on the merging of auxiliary data. We will use altimetric data as a source of additional information. The following chapter describes how this data was generated from stereoscopic aerial images.

Chapter 3 describes how a Digital Surface Model (DSM) is generated from a pair of aerial stereoscopic images. The DSM is used as a source of information in addition to the satellite data to help analyze the urban constructions. The problem of camera calibration has been considered with using a new optimization algorithm designed to estimate precisely the aerial image acquisition parameters. This approach makes it possible to estimate the parameters by using the information redundancy relative to the control points extracted from n stereopair images. A non-linear cost function constructed from the model of the camera is minimized in order to estimate the calibration parameters. The DSMs generated are then geocoded, orthorectified and merged. The information contained in the DSM will be used later at the time of unequivocal change detection (chapter 7) and with the active contour models (chapter 5). The following chapter gives a reminder of the theoretical foundations of these mathematical tools.

Chapter 4 starts by presenting a brief state of the art concerning active contours. We classify the various models according to their representation mode, their attachment-to-data and regularization terms, with special attention being paid to the insertion of exogenous prior shape constraints. Lastly, we will substantiate our choice of representing active contours by level sets and will describe the attachment-to-data and shape constraint models which will be used in this study.

Chapter 5 shows how to improve the robustness of the shape constrained active contours and designed to achieve a match between the cartographic buildings and the image. The first contribution consists of proposing a different optimization algorithm for estimating the pose parameters making the shape constraint invariant from a global transformation. The second one concerns the insertion of the DSM in the active contours' attachment-to-data term. This type of representation is complementary to the satellite image and enables a better discrimination of the constructions from the rest of the scene. The third innovation concerns the incorporation of a shape constraint that varies according to space and time. This is designed to give the active contours' greater flexibility so as to make them less sensitive to the minimized energy local minima. We will also see that the shape constraint model is not, despite everything, suited to overcoming the local errors contained in the map. The following chapter addresses this problem with the formulation of new energies.

Chapter 6 describes the formulation of a new shape constraint energy criterion, combining linear and quadratic energies. This is designed to overcome the errors present in the map in order to improve the matching with the image. We propose a solution that consists of authorizing certain deviations of the active contour with respect to the reference shape. The class of authorized deviations is that of movements constrained by parallelism and corresponds to the errors the most often found in the map. This model is completed by a quadratic energy imposing generic rectilinearity and orthogonality constraints on the active contour. The works presented in chapters 5 and 6 are designed to improve the map-to-image matching, i.e. the consistency of the two representations. This contributes to improving the reliability of the change analysis described in the following chapter.

Fine matching by active contours has enabled us to reduce the variabilities between the cartographic object and its representation in the image. These variabilities are signs of relatively minor differences between the map and the image, and may or may not be due to real changes between the two types of data. In chapter 7, we describe a methodology for analyzing changes between a map of buildings and multi-source remote sensing data. The approach that we are proposing is based on the merging of altimetric, multispectral and panchromatic data processing and is split into two steps. The first one merges change indices with a view to only validating the unequivocal disappearance of a building. These indices are calculated by means of simple methods using the multispectral image and the altimetric data. In most cases where it has not been possible to identify a change clearly, a second more complex phase is carried out in order to characterize the degree of resemblance between the cartographic object and its representation in the panchromatic satellite image. This processing takes advantage of the fine matching achieved by active contours and makes it possible to formulate a non-change probability for each element of the map.

Chapters 3, 5, 6 and 7 are each illustrated with experimental results which are analyzed and compared with the traditional approaches.

1.5 Context of the PhD study

This thesis, organized in co-tutelage, is the fruit of a Sino-French cooperation between the LIAMA (Chinese Academy of Sciences), Alcatel Alenia Space and LIMA (ENSEEIH/INPT). The PhD study was co-financed by Alcatel Alenia Space and LIAMA/NLPR in the framework of projects 01.10 (*Digital map updating on urban areas using high resolution images*) and MOST 863 (*Multi-source data fusion and urban development planning decision support*).

Chapter 2

Analysis and formulation of the problem

2.1 Introduction

This chapter gives an overview of the recent works carried out on the interpretation of urban scenes using high-resolution remote sensing data. This assessment of the state of the art makes no claim to being exhaustive and will be oriented towards the description of methods using the fusion of remote sensing and cartographic data in order to facilitate the difficult analysis of urban landscapes. We will then describe the problem of updating maps in a general conceptual framework. This will lead us to more specifically formulate the problem that we are attempting to resolve through this study, that is to say the updating of digital maps of buildings using high-resolution optical images of an urban environment. Lastly, we will describe a methodology designed to resolve this problem and which is based on the attenuation of the exogenous variability between the map and imagery data.

2.2 State of the art relative to scene analysis and map updating from high-resolution remote sensing images

2.2.1 Specific aspects and strategies for map updating

The automatic or semi-automatic updating of maps using remote sensing data is a problem at the crossroads between several disciplines of image processing and pattern recognition. It refers to change detection between two multi-temporal representations of a given scene [14, 97] and to pattern recognition for the extraction of objects (roads, buildings, etc.) [63, 45, 8, 71, 73].

2.2.1.1 Comparing images

The approaches used for detecting changes in remote sensing are traditionally based on a comparison of images that is made either at the level of the pixel, of the primitives extracted from the images - typically invariants - or of the object. A complete review of inter-image change detection techniques can be found in [97]. Whatever the algorithm

used, a preliminary data alignment phase is essential to compare them at a later stage. This matching is both geometric and radiometric.

In the case of images obtained with only slightly different acquisition views, an affine transformation is sufficient for the geometric registration of the data; otherwise, non-global transformations are then estimated such as the optical flow [10]. The choice of primitives derived from the image to be matched is delicate. Indeed, the precision of the data fusion and therefore the reliability of the change detection method will depend on the quality of the primitives extraction. In [14], P. Blanc proposes an alignment method based on the pixels intensity (correlation) with a sub-pixel precision. In the case of image acquisitions with significantly different resolutions, Dufournaud *et al.* matched point primitives made invariant from rotation and translation [37]. Noting that a low-resolution image is the result of a convolution of a high-resolution one by a Gaussian kernel of variance $s\sigma$, the authors discretize the scale space s and estimate the best direct plane similarity matching the points extracted from two images whose difference of resolution can reach a factor as high as six.

The radiometric correction of the data consists of attenuating the differences of illumination and of lighting represented in the images in order to make their later comparison efficient. The most commonly used technique consists of normalizing the images so that they have the same mean and radiometric variance [67, 118]. Other more sophisticated techniques based on linear transformations of the lighting [34, 57] or the modeling of the illumination have been proposed [75, 127].

2.2.1.2 Map/image comparison

The problem of comparing a map with an image is, however, different from that of comparing images with each other, because the cartographic data is a synthetic, high semantic level and simplified representation of the reality. Like the comparison of images for change detection, map/image comparison includes a preliminary data fusion phase. However, no radiometric correction is applied since the map is not a remote sensing image. The problem of superimposing a map with a SPOT-4 satellite image has been studied by M. Roux in [105]. A map in paper format was digitized and then pre-processed to extract the road network. This was then registered by dynamic programming on linear and “junction” type primitives extracted from the image. Present-day maps are generally taken from Geographic Information Systems (GIS) and are in fact digital. The availability of such maps presents a certain advantage: the cartographic data is vector-based and does not require any pre-processing as is required for digitized paper maps; the map is digital and geocoded, its interfacing with remote sensing images is thus facilitated. In particular, if the image is also geocoded, the superimposition of the data prior to change detection is straight-forward.

To date, much less work has been carried out on the comparison of cartographic and remote sensing data to achieve map updating than on detecting changes between two or more images. Whereas the research carried out on map updating using high-resolution optical imagery remained marginal in the nineties [54, 109], it is tending to multiply now

as shown by the European agencies cartography programs: ATOMI project in Switzerland [38, 6, 133, 134], revision of the ATKIS map in Germany [124, 122, 123, 42] and TOP10DK in the Netherlands [62, 82, 61]. Among the works already carried out on the subject of cartographic data updating, a distinction can be made between those that take advantage of the specific prior knowledge derived from the map in order to facilitate the processing, and those that neglect this high-level information. In this latter case, bottom-up approaches are often used: prior extraction of the pertinent primitives from the image, then higher level reasoning by making assumptions on whether the object extracted from the image belongs to the map. As a general rule, the problem with such an approach is the difficulty of exhaustively and correctly extracting from the image all the primitives being sought. So, those that are missing could be incorrectly interpreted as a change while being compared with the map. The extraction of primitives corresponding to an object from the image whose nature is coherent with those symbolized on the map is also uncertain. S. Servigne is confronted with these problems in the framework of updating a cadastral map of the city of Padua using aerial images [109]. The result of a low level segmentation of the images is used to attempt to match it with the cartographic data. To begin with, a relaxation technique is used to attempt to associate an element in cadastral database with its segmented representation in the image. The results are not very satisfactory because the segmentation produces a large number of contours, often fragmented, for a given object. In order to improve matching, another region-based method is applied by injecting the prior knowledge relative to the texture of the objects. A textures database is learnt manually from the image listing nine different types: land, crops, tarmac, water, vegetation, roof, façade, shadow, pedestrian. The comparison of the cartographic and image objects is finally carried out with respect to shape, texture and surface resemblance criteria. In [54], O. Jamet justifies the choice of not using the information derived from the map because it is not recent and is sometimes not reliable and would therefore bias the change detection process. He then proposes a method of pairing off segment primitives from a stereoscopic pair of aerial images which are compared with the vectors of a digital map of the urban constructions (IGN, BDTopo). The comparison, formalized in the framework of the Dempster-Shafer belief theory [111] enables to model the uncertainty associated with the map.

The prior knowledge derived from the map is however an undeniable advantage for improving objects extraction from urban scenes or for updating cartographic data [35, 36, 120, 122, 124, 48, 3, 123, 4, 9, 62, 12, 61, 82, 42]. Urban scenes imaged at high resolution contain a great amount of detail and other artifacts which make their automatic or semi-automatic analysis difficult (shadows, distortion due to the image acquisition perspective, occlusions). The specific information derived from the map informs us about the object location to be found in the image, which limits both the search space and the false alarms rate. Besides the specific knowledge provided by the map, the fusion of auxiliary remote sensing data is also the tendency for the analysis of urban scenes at high-resolution. These additional data are found useful for eliminating ambiguities relative to the interpretation of the scene (multispectral, infrared, altimetric, stereoscopic images) [124, 122, 48, 123, 47, 46, 62, 83, 82, 61]. In the rest of this chapter, we propose to describe the recent works on object extraction and map updating from high-resolution images in an urban environment according to these two approaches, that

is to say the fusion of remote sensing data and the utilization of prior knowledge, in particular the specific knowledge provided by existing maps.

2.2.2 Some works on remote sensing data fusion for scene analysis at high-resolution

2.2.2.1 Fusion of optical images

Data fusion is a formal framework in which are expressed means and tools for the alliance of data originating from different sources. It aims at obtaining information of greater quality; the exact definition of “greater quality” will depend upon the application [121]. In the framework of high-resolution optical images fusion, we can mention the methods that aim to improve the spatial resolution of multispectral images from better resolution panchromatic images. A complete review of these fusion techniques is given in [125]. The gain in quality is then a more complete description of the scene, combining a high geometric resolution with numerous spectral signatures, thus facilitating the analysis and extraction of objects. This gain is nevertheless tainted by merged image reconstruction artifacts which introduce biases in the spatial and spectral domain. A study comparing the performances of some of these techniques has been carried out by F. Laporterie in [64]. In [42], M. Gerke uses an IKONOS satellite image resulting from the fusion of a panchromatic and multispectral image (*Pan-sharpening* or PXS) for updating the ATKIS road network maps. The multispectral information makes it possible to characterize the road efficiently (thanks to the near infrared band) while meeting the map’s topographic precision constraints.

It should be noted that the merging of optical data does not only concern images with different resolutions. Indeed, panchromatic and stereoscopic images with identical resolutions are also used to facilitate scene analysis, particularly at the level of built-up areas [107, 80, 77, 53]. In [77], Nevatia *et al.* propose a bottom-up approach for detecting buildings from aerial images: linear primitives are firstly extracted from an image, and are then associated to formulate roof hypotheses in the form of parallelograms. The hypotheses are verified by the detection of the building’s shadows and walls. The best hypotheses are selected and merged with those taken from the analysis of other images. A complete review of the extraction and 3D reconstruction of buildings is available in [63, 45, 71, 8]

Lastly, it should be noted that images provided by optical sensors with the same resolution and different spectral signatures are also used. The multispectral measurement is particularly interesting for classifying the objects in the scene. In the works of V. Walter [122], aerial multispectral images (RGB, 2 m/pixel) are used for a supervised classification by maximum likelihood of agricultural or urban zones, roads and water. The multi-band information coupled with the images’ texture information enable to obtain encouraging results on rural and peri-urban images. In the next section we will see that for purely urban environments, an additional altimetric data is required for an efficient discrimination of the constructions.

2.2.2.2 Optical and altimetric data fusion

The availability of a Digital Surface Model (DSM) representing the altitude of the objects in the scene is particularly interesting for analyzing urban environments. The DSM is generally the result of a 3D reconstruction from at least two images, or from a direct measurement of the reflection of a laser signal (LIDAR). A DSM enables to distinguish the aerial structures from the ground, thus facilitating the extraction of the buildings. This latter operation is nevertheless made delicate in the presence of vegetation which becomes confused with the buildings in the altimetric image.

In [48], N. Halla and V. Walter filter the vegetation from a Lidar DSM by means of a multi-echo analysis: the laser signal passing through the canopy generates two echoes. The faster echo corresponds to the vegetation's leaves whereas the slower one corresponds to the ground. If we only take the second echo into consideration, the generated Lidar DSM contains very little aerial vegetation and drastically limits the ambiguities between the canopy and the buildings. In a second phase, the authors pursue the works of [122] by applying a classification by maximum likelihood of the scene: the DSM is then considered like the fourth band of a multispectral image (RG and near infrared). The joint utilization of the DSM and the near infrared band makes it possible to achieve an efficient classification of the buildings and vegetation. The DSM data also enables to estimate the shadow areas in the scene and thus avoid their incorrect classification.

In [82], color images (RGB, 0.5 m/pixel) and a normalized Lidar DSM are the input data to an algorithm for the supervised classification of constructions. A first step consists of splitting the acquired class of buildings into sub-groups with homogeneous spectral characteristics (ISODATA algorithm [5]). A second step deals with a classification based on the Mahalanobis distance carried out on the remote sensing data. The result of the classification is then compared with a digital building map to detect any changes.

In [47] snake type active contours are used to segment the buildings represented in an IKONOS image (1 m/pixel) and a Lidar DSM. The active contours are initialized thanks to a coarse detection of the buildings from a DSM, they then move towards the high gradients of the satellite image while being constrained by inflation forces derived from the DSM [27].

2.2.3 Utilization of constraints derived from prior knowledge

In the following sections we distinguish the works on object extraction or map updating depending on whether they are based on the generic or specific (map) knowledge of a studied landscape. A complete review of the different types of prior knowledge used in remote sensing is available in [9, 35].

2.2.3.1 Generic knowledge

The generic knowledge provides information on general properties relative to the radiometry and/or geometry of an object to be extracted from the image. These properties

are the result of human reasoning on the basis of acquired knowledge and most often enable to define object models (typically the roads and buildings). The following are examples of generic knowledge:

- * Roads have parallel edges, a constant width and slight curvature except at intersections.
- * Roads have a homogeneous radiometry that is clearer than their environment.
- * Buildings have straight edges and corners.
- * Buildings have an altitude that is higher than the ground and project shadows onto it.

This knowledge is derived from observations and is therefore considered to be empirical. In [94, 96, 95] Péteri and Ranchin use snake type active contours ([58], see section 4.2.1.3) to extract the road network from an Ikonos satellite image representing a dense urban environment with a resolution of one meter. The interest of the active contour is twofold: in the end it allows a non-fragmented extraction of the road network, it also easily incorporates high level information such as geometric constraints. Notably in [95], a parallelism constraint is used to segment the edges of the road. This geometric constraint taken from generic knowledge enables a more robust segmentation. A multi-scale analysis increases the robustness of the method with respect to the image's level of detail.

M. Rochery *et al.* also used parallelism in [100, 99, 101] for roads extraction using active contours. The constraint is formulated thanks to the introduction of a quadratic energy which enables a long range interaction between points of the active contour. Unlike the works of Péteri [94, 96, 95], this approach is not sensitive to initialization and naturally incorporates the notion of network. However, the results are obtained from medium resolution peri-urban images.

In [83], H. Oriot uses statistical snakes to segment the buildings in a pair of high-resolution stereoscopic images. A cost function is defined so as to favor the integration of regions with great disparities inside the contour. The extraction of the building is refined by the images' gradient information and by a generic constraint favoring angles of 90° and 180° . This approach is semi-automated insofar as the initialization of the active contour is carried out manually. The movement of the contour minimizing a cost function (non-variational approach) is based on the insertion, updating and elimination of the snake's nodes.

In [119], Vinson *et al.* use rectangular deformable models in order to segment the buildings in an ortho-image. A pre-detection of the buildings by segmentation of a digital surface model is used to initialize the model which will converge on the high gradient areas of the image, i.e. the edges of buildings. On convergence, the parameters of the rectangular model (barycenter, length, width, orientation) are estimated. The results of rectangular extractions can then be merged. The choice of a rectangular model reflects the empirical knowledge relative to the shape of the buildings that are the most often

encountered.

Whereas the generic prior knowledge is largely used to help the extraction of objects, it is dependent on societal and cultural aspects and can thus vary at a national or worldwide scale (for example the road network in the old quarters of Beijing is in no way comparable with the road network in Manhattan). This information therefore cannot be generalized with respect to the geographical location represented by remote sensing data. It must be completed by a more dedicated and specific knowledge that is less constrained by human empirical knowledge.

2.2.3.2 Specific knowledge

Typically, specific knowledge is derived from cartographic data. The map indicates the location, shape and nature of the objects represented by the remote sensing data. This is a strong piece of prior knowledge which makes the extraction or map updating task more robust. The map is essentially used in three ways. The first one concerns the reduction of the search space for an object being looked for in the remote sensing data. This enables to reduce the calculation time and the false alarms rate. The second one concerns the training areas for the supervised classification of scenes. Whereas these zones are traditionally defined manually, the cartographic data enables to automate their input. Lastly, the map can be used to enrich a knowledge base with specific information from the analyzed image.

Determination of training areas

The works of B.P. Olsen [82], N. Haala [48] and V. Walter [122] are all based on the supervised classification of multi-sensor images (see section 2.2.2). These classification methods require training areas defined manually by experts. These authors use an existing map to define these zones in order to make the system automatic. However, the utilization of this type of knowledge derived from the map makes the assumption that the number of cartographic objects that have changed is low with respect to the total number of objects. Indeed, too high a number of changes would deteriorate the subsequent classification. To avoid learning errors on the boundaries of cartographic objects, A. Busch reduces the learning space by means of morphological filterings [16]. Unlike the works of [82, 48, 122], those of A. Busch concern medium resolution optical images. In the case of very high-resolution images such as those used by B.P. Olsen, a distortion due to the perspective can be seen at the level of the buildings. Since the map is in orthoscopic geometry (seen at the nadir), the data are not compatible for superimposition prior to the learning phase. Nevertheless, the TOP10DK map used is three-dimensional: the knowledge of the remote sensing image acquisition parameters then makes it possible to efficiently project each cartographic building on to its homologous representation in the image.

Reduction of the search space

In [106], M. Roux et H. Maître use a digitized paper map at a scale of 1:25000 of

the city of Paris to improve the 3D reconstruction of buildings. The map, in which the blocks of buildings are isolated, enables to restrict the search space for the homologous points obtained by correlation for the calculation of disparities from stereoscopic aerial images. Since the search space is reduced, the pairing errors are less common and a smaller correlation window is used to obtain finer results.

In [114, 35], the authors use a road map as the initialization of a change detection process with a more recent high-resolution aerial image. In [35], M. de Gunst makes a comparison between the cartographic data and the image by correlation between the map and a radiometric profile taken from the image. When the correlation is satisfactory, there is no change. When it is low, a second phase in the process based on a contextual reasoning is triggered: depending on the environment in which a change has been detected, a list of objects likely to be detected in the image is drawn up by order of priority. Each type of candidate object triggers a particular type of segmentation the result of which is confronted by hypothesis tests with models defined beforehand. If all of the tests are unfruitful, the search window is enlarged, and the parameters of the extraction methods are adapted. In [114], C. Steger proposes a radiometric profile analysis of the road guided by the map in order to validate the non-change of the parts that it is made up of. When the map is not very precise, generic road parallelism and rectilinearity constraints are used to overcome these inaccuracies. The elements of the map can be rejected, verified or ambiguous. A second algorithm detects the new roads by inspecting the vicinity of the parts of the map validated during the previous phase.

In the framework of the ATOMI project [38], the authors use redundant sources of information (digital map at 1:25000, digital surface model (DSM), multispectral images) for the extraction of roads [133] and the reconstruction of buildings [78]. The map is used to restrict the search space. In [133], stereoscopic RGB aerial images and a digital terrain model (DTM) are used for the extraction of roads. In [78], vector data of the residential zones are coupled with a prediction of buildings provided by the DSM. These data and the multispectral data are used simultaneously for an unsupervised classification of the buildings in the image. A second phase consists of matching the cartographic objects on the results of the classification. Since the scales of the map and of the image are different, the symbolized buildings are often far from their representation in the ortho-image. An affine transformation estimated by empirical scores enables the map-to-image matching and a subsequent 3D reconstruction.

Agouris *et al.* [4] have developed a method enabling to detect the geometric variations of a road symbolized in a map and represented in a more recent remote sensing image using snakes. The originality of their approach lies in the introduction of a new external constraint energy term which depends on the degree of uncertainty associated with the map's planimetric precision. This uncertainty is calculated from a remote sensing image whose acquisition date corresponds to the date on which the map was made: the active contour is initialized thanks to the cartographic objects and moves without any constraint in the image. Depending on the total energy variations, and according to its final value, a degree of uncertainty is assigned to each of the snake's nodes. This degree is allocated according to fuzzy logic rules inspired by linguistics. So, a great

energy variation and a high final energy for a given node will lead to a great uncertainty being assigned to the point of the cartographic object on which that node was initialized. This uncertainty allows the authors to express a recall force similar to that of a spring fixed to the node of the active contour in its initial state. Finally, the external constraint is imposed on the active contours which move on the most recent image. In qualitative terms, the method used makes the external constraint energy proportional to the lengthening of a spring in a certain vicinity, then it becomes constant outside of that vicinity. The extent of the vicinity and the stiffness of the spring depend on the previously determined uncertainty. The lower the uncertainty, the shorter the range of the constraint energy with a great stiffness associated with the spring: since the map's degree of veracity is high, if there is really a change, the gradients of the new road's contours should be sufficiently intense to overcome the constraint. On the contrary, if the uncertainty is high (long range, slight stiffness), it does not seem to be justified to attract the active contour towards its initial position defined by the map. This method appears effective on simple cases of medium-resolution optical images: the road is deviated manually on portions of the road, thus simulating a slight change. The method does not take into account the appearance of new roads whether they intersect old one or not.

An approach similar to [4] is introduced into [12] by Bentabet *et al.* for the extraction of roads from SAR images. Snakes, initialized by means of a map, move to segment the representation of the road more faithfully in the image. The active contours used are attracted by the image's high gradient zones. They are consequently sensitive to the noise of the remote sensing data. The authors propose a filtering of the speckle which preserves the contour information. Considering the problem of the parameterization of the snakes, the authors estimate the stiffness coefficient of the active contour from the mean curvature of the road calculated from the map.

In [3, 2], Agouris *et al.* propose a change detection method for the buildings symbolized in a map using two diachronic aerial images. One of the images was acquired at the same date than the map was made, whereas the other one is more recent. Each cartographic object is projected onto the image that is time-consistent with the map. The subsampling of the cartographic polygon considered provides a list of points whose presence will be verified in the more recent image. The change detection is carried out by assessment of a correlation score originally developed by Grün in [44]. The goal of this method is to find iteratively the six parameters of an affine transformation making it possible to minimize, in the least squares sense, the radiometric differences of the pixels in two windows belonging to each of the images. This approach enables to obtain a very good matching accuracy, however it often encounters parameter estimation non-convergence problems. The authors of [3] resolve these problems by giving more weight to the contours of the buildings than to the rest of the image. Analysis of the radiometric profile of the contours in the reference image, guided by the map, makes it possible to classify the contours in three categories: single (transition between two radiometric levels), double and triple contour. These profiles are then mathematically modeled by a Gaussian combination whose maxima are located at the level of the contours; they are finally injected into the estimation model of the affine transformation's parameters and significantly improve the technique. The advantage of this method is

that it is able to detect partial changes made to a given building. The difficulty is that an image acquired on the same date as the map was made is required; the method also relies on a perfect superposing between the map and the image which is not often the case because of the errors due to the cartographic inaccuracies.

In [65], Leitloff *et al.* propose to extract queues of vehicles from a Quickbird panchromatic image. This method is based on the extraction of linear primitives derived from the image's gradient. The knowledge of a digital map enables the authors to filter numerous primitives that have no link with the cars: by assuming that they are aligned with the road, they apply a morphological filtering that only keeps the primitives that are parallel with the road network. This pre-processing reduces the complexity of the calculation and the subsequent queue recognition errors based on the analysis of the radiometric profiles.

Enriching a knowledge base

In [133], the prior knowledge is made up of rules concerning the arrangement of the roads, of characteristic primitives such as the ground marking and the existing maps. This latter specific information is inserted in a knowledge base system. This base will be enriched with the image's radiometric characteristics at the places where the coherence with the map is verified. An unsupervised classification based on the acquired knowledge extracts the road network.

2.2.4 Summary

In view of this bibliographic analysis, we can see that the merging of remote sensing data significantly improves the interpretation of the landscape and, in particular, the merging of urban scenes which are difficult to analyze. The injection of specific and prior knowledge makes it possible to improve the object recognition and map updating performances, thus avoiding the stumbling blocks of the traditional bottom-up methods. This specific information is most often used to restrict the search space, define learning zones or specifically enrich knowledge bases. However, the taking into account of the geometric information provided by the map is more marginally used. The difficulty, on the other hand, is naturally due to the fact that the map and the image do not contain exactly the same information (due either to changes or to errors on the map). A map-to-image comparison must consequently be capable of moving away from this prior knowledge in order to find the object such as it is in the image. In the next section we clarify these aspects in a general way by formulating the problem of map updating that we are attempting to solve.

2.3 Position of the problem of map updating

In order to pose and formulate the problem of map updating from the remote sensing data, we propose to use and extend the conceptual framework described in [55] and used in [14] to analyze change detection between multi-date and multi-source data.

This framework is of interest insofar as it is well suited to the problem that we are addressing, but it needs to be adapted.

2.3.1 General conceptual framework

2.3.1.1 Definition of the remote sensing related universes

This formalism consists of considering the objective universe $U_{objective}$ containing the exhaustive knowledge set that we may have relative to a given scene. This objective knowledge is inaccessible to us in its entirety. It manifests itself in the form of physical phenomena called *observables*. The knowledge of all of the observables does not make it possible to exhaustively deduce the $U_{objective}$ because the function for the physical transcription of this universe to $U_{observable}$ is not reversible (ill-posed problem). So, the only way of partially knowing $U_{objective}$ is equivalent to measuring one or more observables by means of a measuring apparatus. It should be noted however that for a given scene, the observables vary according to certain factors. In the case of optical imaging these factors are for example:

- * The lighting, which varies according to the position of the Sun with respect to Earth. Depending on the illumination, the shadows cast due to the presence of buildings will vary.
- * The observation geometry (zenith viewing angle). In the specific case of a dense urban environment, this geometry explains why occlusions due to buildings are or are not present due to the image acquisition angle.
- * The weather conditions.
- * The stray phenomena of multiple reflections (by the atmosphere, the ground or the objects making up the scene) will disturb and bias the observable that we really want to measure.

The choice of the means of observation of course depends on the observable we want to measure. In the case of optical imaging, we choose to observe the electromagnetic rays radiated by the Sun and then reflected by the scene on to a spectral window in the visible and near infrared domain (passive sensor). The measurement made of this observable, called observation, is also tainted by errors specific to the physical measuring instrument. For example, the noise of the measuring sensor, or a poor calibration of the latter will alter the final information which will be available to us and from which we hope to obtain the knowledge of $U_{objective}$.

We then understand that $U_{observation}$ is only a partial and biased representation of $U_{objective}$. These alterations are not necessarily stable over time. Thus, when comparing the diachronic images for the change detection, these artifacts introduce undesirable exogenous variabilities because they are likely to be wrongly interpreted as a change. Overcoming exogenous variabilities for detecting effective changes that have been made

in $U_{objective}$ is a difficult task which is at the heart of the whole problem of detecting changes between two images.

2.3.1.2 Definition of $U_{interpretation}$

The works carried out by P. Blanc in [14] attempt to address this question in a general way for a geometric and radiometric matching of the data. However, in [14] the data compared for detecting change all belong to $U_{observation}$. Our case is different in that we compare an observation with a map. A map is definitively not an observation, it is the result of an interpretation, schematization, simplification of an observation. The map, which belongs to $U_{interpretation}$, is also likely to contain errors inherent to the cartographic transcription of $U_{observation}$ towards $U_{interpretation}$. These errors are due to interpretation errors made by the operator who made the map (or of an automated process, if one was used). There are two types of photo-interpretation error. One concerns an interpretation error relative to the nature of the manually extracted object; for example, a lawn or a parking space could be interpreted as the roof of a building observed at the nadir given the great similarity of their characteristics (homogeneous radiometric zone whose contours are characterized by steep gradients). These errors are very rare because the ambiguous cases of photo-interpretation are usually verified on the terrain. The most common errors concern the delineation of the objects' contours. In [79], N. Nideröst illustrates these delineation errors which vary according to the interpretation made by cartographers (figure 2.1).



Figure 2.1: Contour delineation errors: different segmentations of the same object by several photo-interpreters. The diagram is taken from [79].

In addition to this, there is the contour simplification phase relative to the objects inserted in the map which, even if this is not an interpretation error, represents a deviation from the reality of the objects observed in the image. In the case of buildings, some of them have a complex shape which it is pointless to transcribe with precision on to the map for the scale concerned. The shape is therefore simplified by the operator in a more or less arbitrary way. It is also difficult to outline when each of numerous and of small sized buildings are agglomerated together. One solution adopted consists of including a group of objects in a single cartographic entity. Figure 2.4 illustrates this generalization effect. Lastly, a small-scale map superimposed over a high-resolution image will exhibit significant generalization and location imprecision effects, which will, in the end, make the detection of change arduous or even incoherent (figure 2.2).

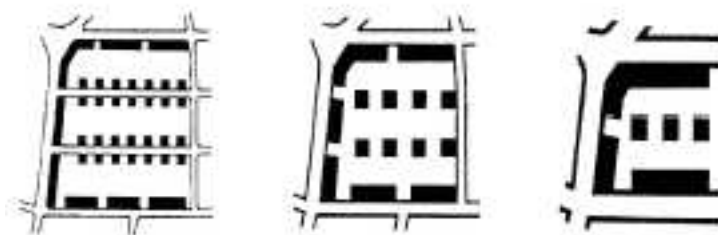


Figure 2.2: Generalization effect due to scale. The same scene is represented at different scales: 1:25000 / 1:50000 / 1:100000 (from left to right). The diagram was taken from [79].

A reliable method of map-to-image change detection must therefore compensate for these exogenous variabilities present in the $U_{observable}$, $U_{observation}$ and $U_{interpretation}$ universes so that they are not interpreted as an effective change (table 2.1).

| Universe | | Artifacts that are sources of exogenous variabilities |
|----------------------|----------------------|--|
| $U_{observable}$ | Lighting geometry | Shadows due to the objects in the scene according to the height of the Sun |
| | Observation geometry | Different occlusions according to the acquisition angle |
| $U_{observation}$ | | Sensor noise, quantification error (digitization), calibration error |
| $U_{interpretation}$ | | Interpretation error relative to the nature of the cartographic object |
| | | Object extraction inexactitudes |
| | | Simplification, generalization |

Table 2.1: Classification of urban artifacts according to the different universes.

2.3.2 Problem addressed in the thesis

2.3.2.1 map-to-image change analysis considered in the thesis

In this thesis work, we examine the updating of digital maps of two buildings in two dimensions from a more recent high-resolution panchromatic satellite image. There are essentially four aspects to the maintenance of cartographic data:

1. Validation/deletion of the cartographic objects that are detected as being present/-absent in the more recent satellite image.
2. Insertion in the map of new objects detected in the more recent satellite image.
3. Improvement of the planimetric precision of the cartographic objects that exist in the more recent satellite image.

4. The enrichment of the map by inserting new attributes. The addition of the 3D information to a 2D map is an example.

In this study, we will explore points 1, 3 and 4 on this list. We will therefore not detect the new buildings to add them to the map. We will limit ourselves to a *map-to-image* detection that consists of individually taking each building symbolized on the map and checking whether it still exists in the more recent remote sensing data. If the building on the map still exists, it may be possible to refine its position and its shape thanks to the remote sensing data (we will consider that the information provided by the image is more reliable than that of the cartographic data). Table 2.2 summarizes the different types of change likely to be found.

| Nature of change | New situation |
|-----------------------------------|--|
| a. Building completely demolished | Bare ground |
| b. Building completely demolished | Any building |
| c. Building completely demolished | Smaller buildings intersecting the former footprint. |
| d. Building completely demolished | Building with the same footprint |
| e. Building completely demolished | Objects \neq construction (roads, trees, urban furniture, ...) |
| f. Building partially demolished | Building partially demolished |
| g. Building with new extensions | Building partially extended |
| h. Building currently being built | Building with the same footprint but with a higher altitude |

Table 2.2: Possible changes of a building in $U_{objective}$

The change of type *d* is marginal and will not be examined, change type *h* is difficult to detect because the map that will be used only contains two-dimensional information. It will therefore be ignored.

It should be noted that insofar as we consider the potential presence of local shape errors on the map, the cases of *f* and *g* represent poorly posed problems of change detection. Indeed, it is impossible to know whether a partial map-to-image change of low amplitude is due to a delineation error on the map or to an effective change in the real world (extension, partial demolition of the building). So, we will not examine these partial changes. Nevertheless, we will develop a methodology that will attempt to correct the map-to-image incoherencies, whether they are due to the artifacts on the map or to real changes.

We will assume that the scales of the map and of the remote sensing data used are similar, in this way we will not be concerned by generalization effects due to the scale. However, the delineation errors and the generalization effects on the map inherent to $U_{interpretation}$ will have to be taken into account (figures 2.3-2.4).

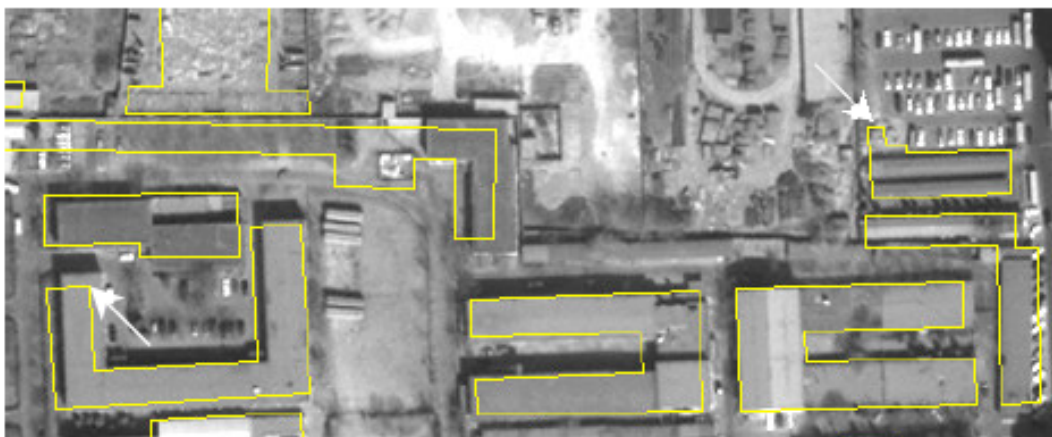


Figure 2.3: Superposing of a map of the constructions (yellow polygons) with a satellite image. The white arrows point to the maps local shape inexactitudes.



Figure 2.4: Superposing of a map of the constructions (yellow polygons) with a satellite image illustrating the generalization effect: a single cartographic polygon encompasses numerous adjacent buildings.

2.3.2.2 Difficulties with urban environments

We will focus our study on dense urban environments. The difficulties intrinsic to $U_{observable}$ and $U_{observation}$ are numerous for this type of landscape for which the utilization of very high-resolution remote sensing images reveals artifacts that do not exist at medium and small scales. These artifacts may be shadows due to the buildings or

flyovers, a serious geometric distortion due to perspective effects at the level of the constructions, or to the occlusion effects. The geometric distortions of the image inherent to the effects of the perspective and of the image acquisition angle make the planimetric location of the objects seen in the image somewhat imprecise while modifying their shape. One way of correcting this artifact is to orthorectify the image but it was not possible to carry out this processing with our data. The shadows are effectively a problem because they mask information and create undesirable artificial contours for any computerized photo-interpretation processing. In the works of D. Boldo and H. Le Men [15], a method is proposed for attenuating the shadow and sky occlusion effects. In an urban environment, the occlusions are due to the high-altitude structures (buildings), however a distinction must be made between an occlusion observed on the image and the effect of an occlusion which is not observed. An occlusion observed on the image generates the occultation of a piece of information (one building hides another one). The effect of an occlusion that is not physically observed degrades the information by partially masking it: a street may be shady because of a sky occlusion (the zone of shadow is not directly lit by the Sun, but by the surrounding sky which results in an attenuation of the object's reflection in the scene) while being visible on the image. Another typically urban effect is the back-lighting of facades which light the objects on the ground, thus introducing a bias in the measurement of their luminance [15]. Lastly, other difficulties are linked to the human activity that it is possible to detect on very high-resolution images. Thus, a road with a traffic jam - where the density of vehicles is very high - will have a radiometry that is very different from that of a road with fluid traffic. It must be noted however that if these artifacts represent additional problems, that may also contribute a non-negligible quantity of information. Thus, the detection of a shadow means that a building exists in its vicinity, the detection of vehicles confirms the existence of a road (generic prior knowledge).

2.4 Proposed methodology

Our methodology assumes that we have two panchromatic high-resolution and multispectral satellite images (RGB and near infrared) of the same scene and an older two-dimensional digital map of the buildings. We will assume that the map and the satellite images are initially superposable (either by the control point data or by a registration algorithm or the geocoding information). This overall matching is not sufficient to resolve the exogenous variabilities specific to each building on the map.

We will use an auxiliary Digital Surface Model (DSM) aiming to help discriminate the constructions from the rest of the scene. In our study, the DSM will be generated from stereoscopic aerial images that are more recent than the map to be updated.

The method that we are proposing for carrying out this change analysis with the map is as follows:

1. Optimization of aerial image acquisition parameters and generation of a DSM by correlation of the stereo-pair aerial images (chapter 3).
2. For each building symbolized on the map:

- (a) Detection of unequivocal changes by merging multispectral and altimetric data. This change detection is limited to case *a* in table 2.2. It cannot in any event confirm a non-change. It is designed to detect the unambiguous cases of change between the map and the remote sensing data (chapter 7).
- (b) Attenuation of the exogenous variabilities occurring in $U_{interpretation}$ thanks to the panchromatic satellite image. We only study the buildings that have not been detected as having changes in step (a). We use the active contours designed to finely match each building symbolized on the map with its homologous representation in the image (chapters 4,5,6). In the case of a cartographic building that is absent from the image (change), the match fails and the change will be detected later in step (c). In order to perform the fine map-to-image matching by active contours, we will take advantage of the prior and specific knowledge derived from the map in order to overcome the difficulties encountered with dense urban environments. This knowledge is twofold:
 - i. **Geographic.** Since the cartographic and remote sensing data are initially superimposed, we know where to look in the image to carry out the fine matching. So, the active contour will be initialized close to the roof of the building to be matched thanks to the geographical information contained in the map. It is important to note that when the image is not orthorectified, the superposing of the map-to-image is only effective for buildings of low to medium height (the distortion due to the perspective visible at the level of tall buildings moves the roof away from the orthoscopic cartographic object). Since the active contour technique is sensitive to the initialization, our method cannot take into account very tall buildings. When the remote sensing data are orthorectified, this limitation disappears.
 - ii. **Geometric.** We know in advance the shape of the building to be found in the image if no change has occurred. This information will be inserted as a shape constraint in the active contours model and will guarantee a greater success for the matching with a lower sensitivity to urban artifacts. The merging of a DSM in the active contours model and a variation in the shape constraint will show a greater fine matching capability with respect to the geographic uncertainty of the buildings on the map. Whereas the exogenous variabilities are attenuated, the precision of the cartographic is likely to be improved.
- (c) Panchromatic map-to-image change detection. A coherence score is calculated between the segments of the resectioned cartographic object and the segments extracted from the panchromatic image. This score, combined with the geometric variation recorded by the matching of the active contours will make it possible to calculate a probability of non-change indicating the presumption of a change/non-change for the building on the map considered (chapter 7).

The result of the method is a change/non-change probability for each building on the map. Optionally, the altitude derived from the DSM may be used to enrich the map

which was originally in two dimensions (phase 2.b.ii, see Appendix F). If the satellite image is orthorectified then the planimetric localization of the buildings will be improved by fine matching (phase 2.b).

The satellite images and digital maps that we will use to experimentally verify the efficiency of the method are presented in Appendix A. The remote sensing images come from a sensor on board the Quickbird satellite. The resolutions of the panchromatic and multispectral images are respectively 0.6 m/pixel and 2.4 m/pixel. The image acquisition angle with respect to the nadir varies between 5° and 9° . The distortion due to the perspective is therefore relatively low but nevertheless corresponds to the sub-optimal configuration of non-orthorectified data mentioned in 2.b.1. These satellite images and the aerial data (0.2 m/pixel) represent all the city of Beijing which is currently undergoing a rapid expansion and restructuring of its urban fabric.

2.5 Conclusion

In this chapter we have outlined the state of the art carried out recently on the extraction of objects (roads, buildings) and map updating using high-resolution remote sensing data. It can be seen that the merging of auxiliary data represents the current trend for resolving the ambiguities in urban scenes. More recently, the utilization of maps in the fusion process has shown its efficiency. The prior knowledge provided by the map is specific and enables either to restrict the search space or to define learning zones, or enrich a knowledge base with characteristics specific to the image. After having described the problem of updating maps from the general angle of universes that are specific to the observables, observations and interpretations, we have more specifically formulated the problem addressed by this thesis. Lastly, we have proposed our methodology for analyzing changes between a digital map of buildings and more recent remote sensing images. This is based on the fusion of remote sensing data and proposes to resolve the problem of exogenous variabilities between the map and the images using the active contours technique in order to complete a reliable detection of a map-to-image change. The proposed approach is limited to the analysis of changes a, b, c, e of table 2.2 and is restricted to buildings of *low* or *medium* height when the remote sensing data are not orthorectified.

Chapter 3

Generating an orthoscopic digital surface model

3.1 Introduction

This chapter proposes to describe the process for generating an orthoscopic digital surface model (DSM) from pairs of stereoscopic aerial images. It must be remembered that in the general methodology, this DSM is a source of additional and discriminating information at the level of buildings which will aim to facilitate the updating of maps on the basis of a single satellite image. To begin with, a brief reminder will be given concerning the fundamental principles of stereovision, and of the 3D reconstruction algorithm that was used to generate a non-orthoscopic DSM. This algorithm is based on a stereoscopic image correlation technique which preserves the contour of the objects in the landscape [85, 86]. This software requires precise knowledge of the acquisition parameters for the images used to generate the DSM. The second section of this chapter proposes a method making it possible to refine the precision of such parameters known beforehand. This approach, which represents the main contribution in this chapter, requires the optimization of non-linear functionals by the simplex algorithm. Quantitative results will demonstrate the efficiency of the proposed method. Lastly, we will end this chapter with the geocoding, orthorectification and merging of the DSMs generated by the 3D reconstruction chain.

3.2 Photogrammetric stereoscopy

Stereoscopy concerns the methods enabling to obtain a measurement of a scene's relief from two photographs taken from two distinct viewing angles. Inspired by the human visual system, stereoscopy is commonly used in photogrammetry when creating relief maps, and also in optical and electronic microscopy.

In this chapter, we will only examine the automated restitution of the relief of a scene observed using remote sensing means (figure 3.1). The result of a reconstruction such as this is the localization in 3D space of each of point of the scene seen in at least two images. In the particular case of remote sensing, the reconstructed points of the

landscape are often referenced with respect to the terrain, i.e. with respect to a cartographic baseline. A landscape estimated elevation difference is called the Digital Terrain Model (DTM) or the Digital Surface Model (DSM) depending on whether we are representing the relief of the ground alone or of the ground and its superstructures (buildings, etc.). Unlike a DSM obtained by LIDAR imaging, a DSM made by stereophotogrammetry is the result of a complex reconstruction processing, which renders it sub-optimal. Indeed, this reconstruction is subject to the errors present in the knowledge of the image acquisition (calibration) and to the performance of the mathematical models used. The authors of [7] have, however, shown that their 3D reconstruction from stereoscopic satellite images taken by Ikonos is better than the LIDAR data over a small extent of the generated DSM.

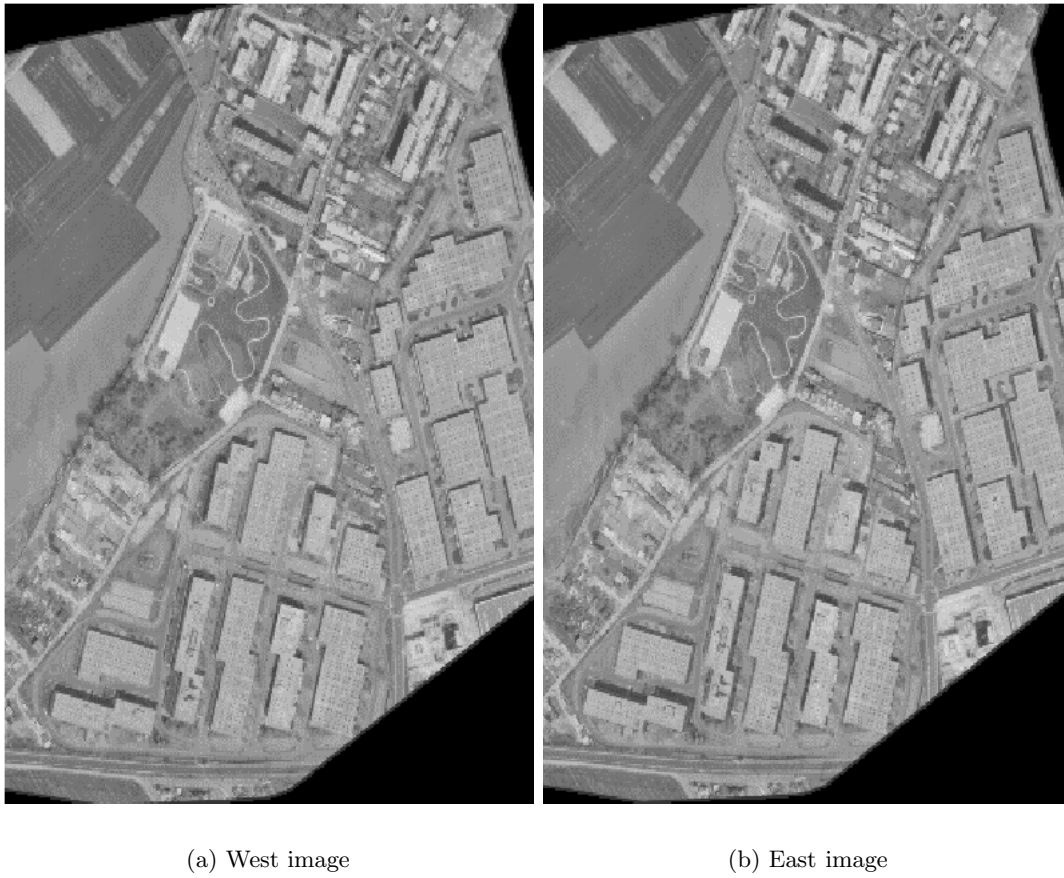


Figure 3.1: Example of a pair of stereoscopic images. The images are satellite data simulated from aerial images.

3.2.1 Stereoscopy - reminder and general principle

Let P be a point in space belonging to a landscape represented by two different remote sensing images I_A and I_B . These images were taken from different camera positions. C_A

and C_B denote the positions of the optical centers for each of the images. The purpose of stereoscopy is to determine the location of point P in the scene by calculating its three-dimensional coordinates from its projections P_A and P_B in the images (figure 3.2).

The relative difference between P_A and P_B is called the parallax and it is inversely proportional to the distance between point P and the focal planes of the images. The parallax representing the distance in meters in the measurement plane between the two representations of any given point, can be expressed in pixels in the image, this is then called the disparity. By definition, the disparity d is the ratio between the parallax and the resolution ρ of the image in m/pixel:

$$d = \frac{\text{parallax}}{\rho}$$

The intersection of the plane (P, C_A, C_B) with the focal planes forms the conjugated epipolar straight lines. The distance $\overline{C_A C_B}$ is called the base which we will note B . The optical centers are assumed to be at an altitude H from the ground.

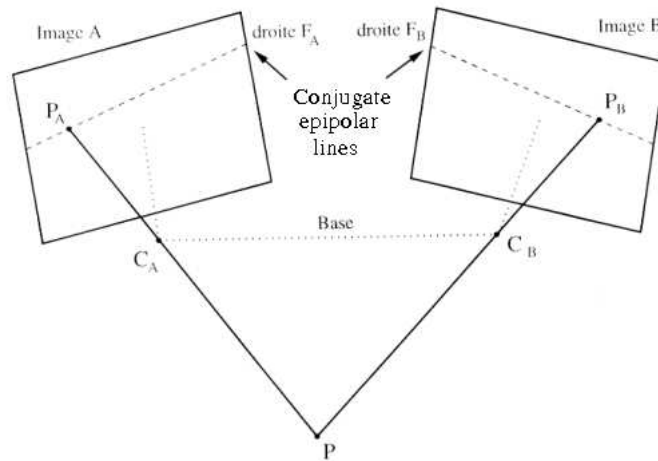


Figure 3.2: For the stereoscopic restitution, two images are acquired from two different viewpoints. One point P in the scene is projected onto each image at the intersection of the image plane and of the straight line joining P to the optical center C_A or C_B .

The knowledge of the image acquisition conditions (orientation and intrinsic parameters of the camera) enables to determine the viewing directions of pixels P_A and P_B , i.e. the straight lines $(P_A C_A)$ and $(P_B C_B)$. We can then calculate the three-dimensional position of P by considering the intersection of the viewing directions of the two homologous points P_A and P_B (this phase is called *aero-triangulation*). The 3D reconstruction by stereoscopy is therefore equivalent to solving two problems:

1. **Calibration.** Precise knowledge of the image acquisition parameters is required for a reliable calculation of the viewing directions. This information can be estimated from the images and the ground control points or from the prior knowledge.

In section 3.3, we will propose a method making it possible to refine the precision of these parameters, for which only an imprecise estimate is known beforehand.

2. **Matching.** For any given pixel P_A , we must find its homologous point P_B in the conjugated image in order to calculate the intersection of the viewing axes. This shape recognition problem is the most delicate. It should be noted that the exploration space of P_B for a given P_A can be drastically reduced if the conjugated epipolar straight lines are known. In section 3.2.2 we will describe the surface matching method implanted in the reconstruction chain that was used to generate the DSM.

The important parameters in stereoscopic image acquisition are the base B and the height H , or rather their ratio B/H . In the case of a satellite, H is fixed by the latter's orbit and B is the only parameter on which it is possible to exercise any influence. In aerial imaging, the setting of these parameters is more flexible.

- * A high B/H ratio means two image acquisitions with large angles with respect to the nadir of the observed scene. The advantage of this type of configuration is that it gives a good altimetric location of the objects. Indeed, the closer the angle of the inter-image viewing directions is to 90° , the less the uncertainty on the viewing directions will induce large errors on the altimetry. However a high B/H ratio significantly increases the risk of occlusion (point seen in one image but hidden in the other one due to the scene's relief) along with the failure rate relative to the recognition of the homologous points. A high B/H ratio will therefore not be suitable for an urban relief since it varies greatly and quickly thus inducing numerous occlusions.
- * On the contrary, a not very high B/H ratio implies two very similar image acquisitions with a low angle with respect to the scene's nadir. The number of occlusions is low, and the points to be matched are easier to recognize. However, the altimetric precision is degraded since it is determined from virtually parallel viewing axes generating large errors if there is any imprecision relative to the calibration of the remote sensing sensor. It is therefore necessary to choose a B/H ratio that is suited to the nature of the relief in the observed scene. Besides these two limit cases that have just been described, B/H varies typically between 0.1 and 1.2.

Finally, it should be noted that we have implicitly assumed until now that the viewing directions of each pair of homologous points intersect with each other. This is obviously untrue because of the errors and imprecision of the measuring instruments used to calculate the viewing directions. In practice, the intersection point P is generally defined as the center of the smallest segment joining the two viewing directions (figure 3.3).

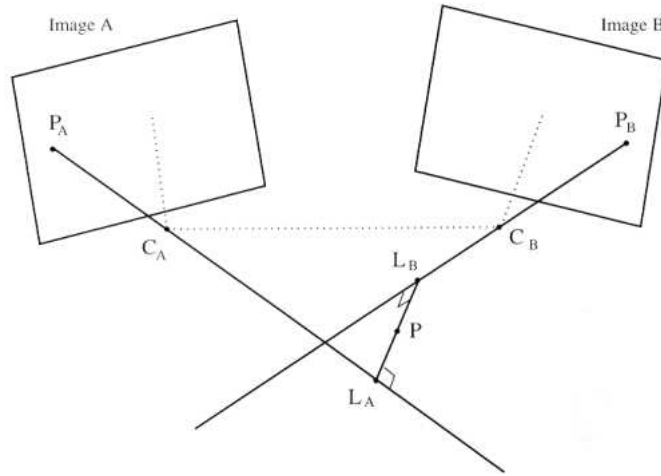


Figure 3.3: For two conjugated points P_A and P_B , the relative orientation enables to calculate the position of point P in the coordinates system linked to the pair of images. P is defined as the center of the smallest segment joining the straight lines (P_A, C_A) and (P_B, C_B) .

3.2.2 Automated matching of images using adaptive correlation masks

The purpose of matching is to find for each pixel in the reference image I_A its homologous point in image I_B . The exact geometric relationship making it possible to pass from point (i_A, j_A) in I_A to (i_B, j_B) in I_B is as follows:

$$(i_B, j_B) = (i_A + d_c, j_A + d_l)$$

where: (d_c, d_l) is the true disparity at point (i_A, j_A) according to the columns and lines respectively.

The automated matching of images requires the maximization of a resemblance function between two points or two regions belonging to each of the images. The surface approach consists of looking for homologous points from their resemblance with the respective *blocks*. Let us take a pixel in the reference image, we consider a window centered on that point, then we look for its homologous point by sweeping a window of the same size in the other image. We just have to formulate a resemblance criterion in order to quantify the similarity of the windows.

3.2.2.1 Resemblance criterion by crossed correlation

Techniques based on the correlation of the intensities of the images have been used in numerous commercial stereo-photogrammetry applications, they also represent one of the oldest methods developed in the area of computer vision. The principle of crossed correlation for measuring the similarity between two points is to calculate the inter-correlation coefficient from two vignettes of size $K \times L$. The higher the correlation score, the greater the presumption of similarity will be. The normalized centered correlation

coefficient at pixel (i_A, j_A) in the reference image I_A and at pixel $\mathbf{s} = (s_c, s_l)$ in image I_B is written:

$$c_{i_A, j_A, \mathbf{s}} = \frac{\gamma_{i_A, j_A, \mathbf{s}}(I_A, I_B)}{\sigma_{i_A, j_A}(I_A) \sigma_{i_A, j_A, \mathbf{s}}(I_B)}$$

with the inter-correlation function:

$$\gamma_{i_A, j_A, \mathbf{s}}(I_A, I_B) = \sum_{m=i_A-K}^{i_A+K} \sum_{n=j_A-L}^{j_A+L} (I_A(m, n) - \bar{I}_A) (I_B(m + s_c, n + s_l) - \bar{I}_B)$$

and the variances:

$$\sigma_{i_A, j_A}^2(I_A) = \sum_{m=i_A-K}^{i_A+K} \sum_{n=j_A-L}^{j_A+L} (I_A(m, n) - \bar{I}_A)^2$$

and

$$\sigma_{i_A, j_A, \mathbf{s}}^2(I_B) = \sum_{m=i_A-K}^{i_A+K} \sum_{n=j_A-L}^{j_A+L} (I_B(m + s_c, n + s_l) - \bar{I}_B)^2$$

where:

- * \bar{I}_X represents the average of the intensity of the pixels in image X inside the window of size $K \times L$ considered.
- * The correlation window in the conjugated image is centered on pixel $\mathbf{s} = (s_c, s_l)$. s_c and s_l belong to the interval $[-W, W]$ which defines the search area in the conjugated image ($W > K, W > L$).

The normalization of the correlation coefficient by the standard deviations of the vignettes, and the fact that it is centered, makes it possible to overcome a linear radiometric bias present between the two images. The correlation process is reiterated for each pixel in the conjugated image within the search area, which enables to construct a correlation surface comprised between 0 and 1 for each pixel in the reference image (figure 3.4). The correlation surface illustrated in figure 3.4 shows several modes. Here it is a complex surface. The ideal correlation surface corresponds to the case where the homologous point is identified in a unique and unambiguous way, it then only includes a single, high-level peak. The position of the maximum corresponds to that of the processed pixel's homologous point. This maximum is extracted and enables to calculate the disparity associated with that pixel. An interpolation of this correlation surface in the vicinity of the discrete maximum enables to obtain a sub-pixel estimation of the real position of the maximum, and therefore of the disparity. The search for the sub-pixel disparity is reiterated for each pixel in the reference image in order to make a disparity map.

The assessment of the quality of matching is then carried out according to the following criteria:

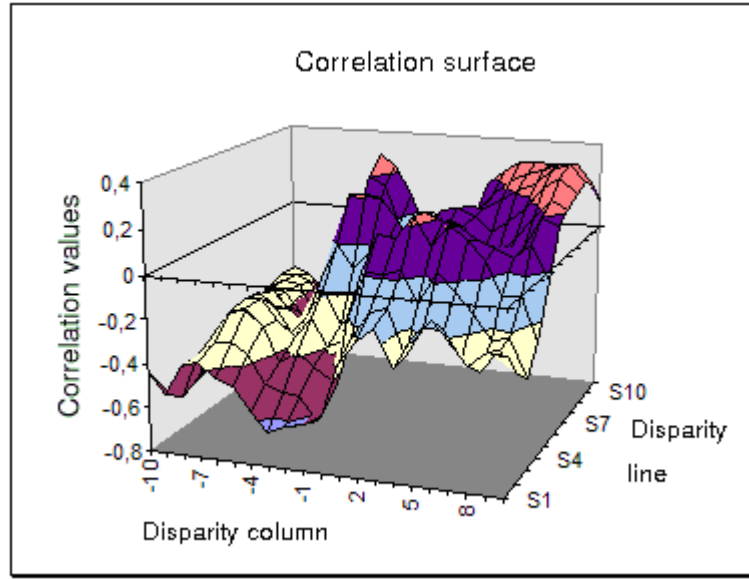


Figure 3.4: Example of correlation surface.

- * height of the correlation peak: the correlation value must be greater than a threshold ε_{corr} .
- * width of the correlation peak: the spread of the peak at the summit must be restricted.
- * unicity: limited presence of local maxima (ambiguities) in the vicinity of the correlation peak.
- * divergence of the interpolation function: the sub-pixel maximum must not be detected too far away from the discrete maximum.

These quality criteria make it possible to filter out the pairing errors and ambiguities in order to obtain a disparity image that is as reliable as possible. If, for a pixel in the reference image, none of the pixels in the conjugated image satisfy the above-mentioned conditions, the disparity is not filled. This case will occur when the resemblance between the two pixels is not sufficiently frank, or when a point in the scene is only visible in a single scene (occlusion).

Other correlation constraints may be taken into account to improve the reliability of the matching and speed it up:

- * **Reciprocity constraint:** if X is the conjugated homologous point of Y , Y is the homologous point of X , hence the interest of carrying out a crossed correlation (with one image and then the other as reference) and then a filtering to eliminate the incoherent disparity points.

* **External constraints:** they limit the complexity when looking for the homologous point in order to increase the processing speed. So, rectifying the two images in epipolar geometry will make it possible to look for the homologous point of a pixel in a reference image on a straight line in the other image, rather than a zone in the plane. Furthermore, if we know beforehand the scene's maximum and minimum altitudes, we limit the search to one part only of the corresponding epipolar straight line. This enables to limit the calculation times since the problem of the search for homologous pixels is then reduced to one dimension. Furthermore, the restriction of the search space limits the occurrence of correlation errors. The rectification in epipolar geometry nevertheless requires a precise knowledge of the image acquisition parameters. We do not detail this image re-sampling process which is illustrated in figures 3.5-3.6 and detailed in [129].

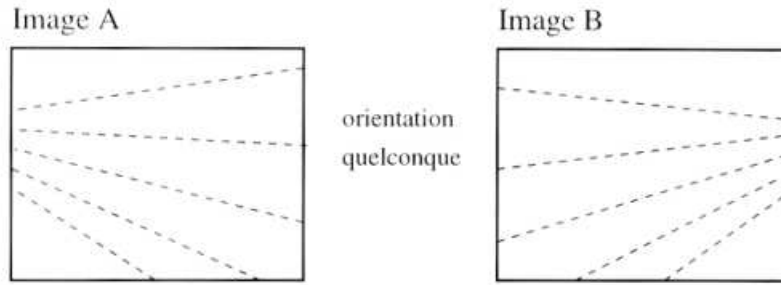


Figure 3.5: Pair of images in any geometry.

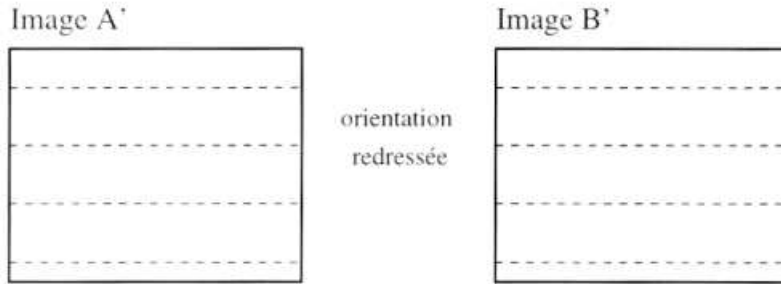


Figure 3.6: Pair of images in epipolar geometry.

3.2.2.2 Correlation by adaptive masks

The correlation method using vignettes with fixed dimensions has the advantage of providing maps with dense disparities. This technique is well-suited to the restitution of DTMs, however it does have a limitation for the generation of urban DSMs. Indeed, it tends to smooth out the altimetric transitions because of the non-homogeneous disparities contained in the vignettes. This is particularly critical for the reconstruction of the

relief of the constructions.

The correlation method using adaptive masks was designed to compensate for this problem and restore urban DSMs that have great altimetric discontinuities. This technique was developed by N. Paparoditis in [85, 86]. This approach enables to adapt the shape of the window to the contours of the objects present in the image so as to reduce the disparity variations. The principle of correlation by adaptive masks is illustrated in figure 3.7. It should be noted that this approach is based on the hypothesis that the altimetric discontinuities and, more generally, the scene's slope ruptures are characterized by high intensity gradients in the image space. Figures 3.8 and 3.9 show the

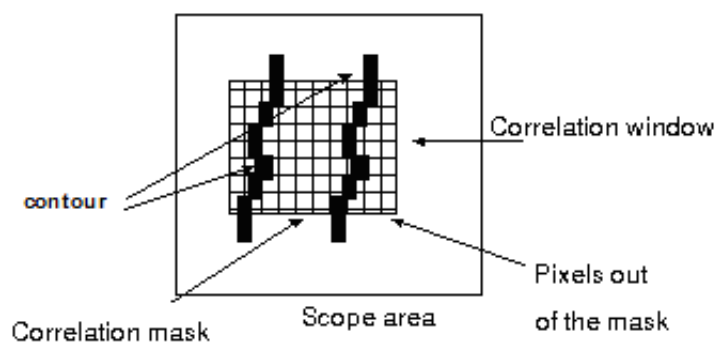


Figure 3.7: Adaptive correlation mask.

contribution of correlation by adaptive masks: by taking into account the contours it is possible to avoid the pairing errors in the vicinity of the constructions.

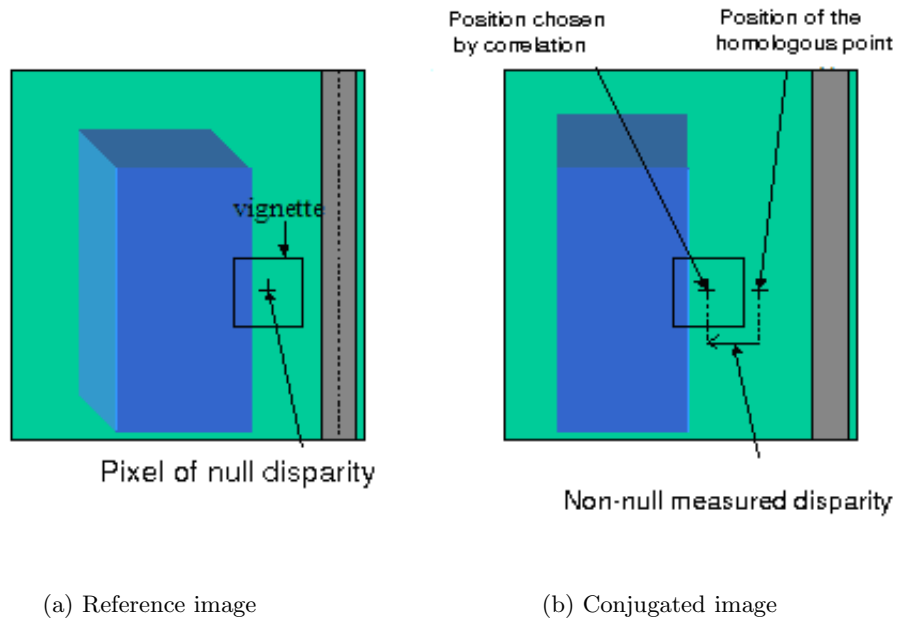


Figure 3.8: Error in the case of occulted zones with a fixed correlation window. (a) a null disparity pixel is considered. (b) its homologous point in the conjugated image is not found because part of the blue building is taken into account in the correlation window.

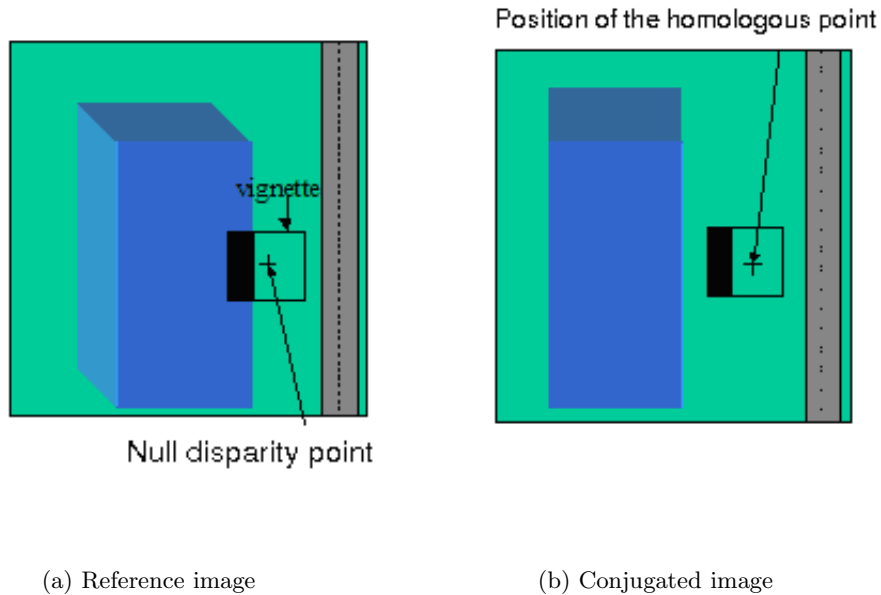


Figure 3.9: Application of the mask on the context window. The part of the blue building is withdrawn from the correlation window thanks to the mask adaptive geometry. The homologous pixel is found.

The size of the adaptive correlation mask is adjusted locally and automatically. This local adaptation is carried out considering a minimum size for which the signal-to-noise ratio is sufficiently great to achieve a high-quality correlation. The quality is assessed by measuring the so-called Cramer-Rao statistical bound which is a sub-determinant of the standard deviation of the disparity's estimation error. Figure 3.10 illustrates the best performances of adaptive masks with respect to fixed vignettes.

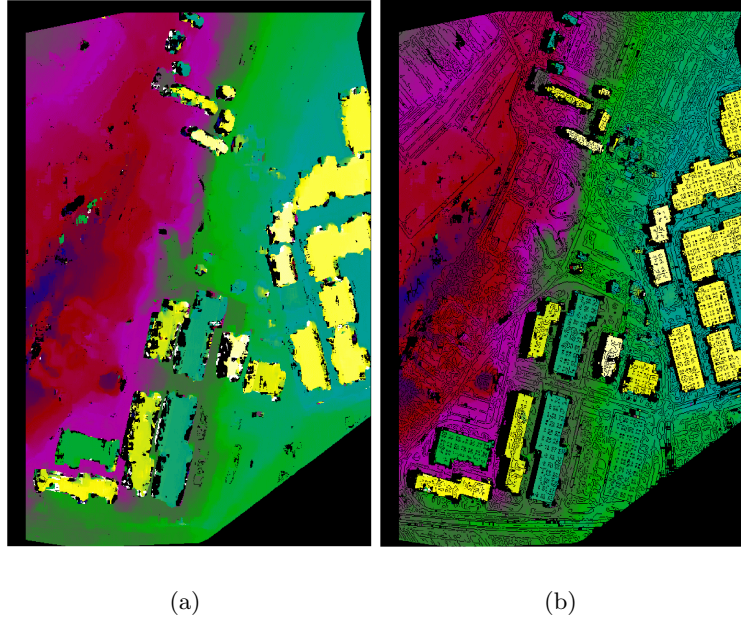


Figure 3.10: Disparity image generated from the pair in figure 3.1: (a) Conventional correlation (b) Correlation by adaptive masks.

3.2.3 Principle of the reconstruction chain used

The 3D reconstruction chain used enables to process pairs of aerial and of satellite images. In order to reduce the calculation times and the correlation errors, the images are first of all rectified in epipolar geometry thanks to the knowledge of the image acquisition parameters. The images are then matched using the adaptive masks technique described in section 3.2.2.2. The result of the matching is a disparity image for each pair of correlated images. The roles of the reference image and of the conjugated image are then exchanged to generate a second, equivalent disparity image. The two disparity images are then merged and the aerial triangulation is finally implemented to produce a DSM geocoded in epipolar geometry. Each pixel of this DSM contains an item of spatial localization information coded by geodetic coordinates (longitude, latitude and elevation with respect to a reference ellipsoid). The detail of the reconstruction chain process is illustrated in figure 3.11.

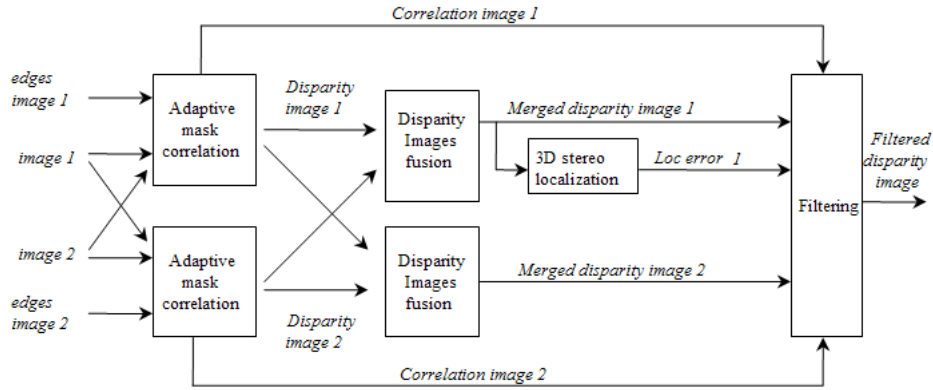


Figure 3.11: Summary of the 3D reconstruction process.

3.3 Optimizing the image acquisition parameters

In the rest of this document we will consider the data of three aerial images representing the city of Beijing. These images are the result of the digitization of analog films (see appendix A). The accuracy of the image acquisition parameters delivered with the three images is crucial for transformation into epipolar geometry preceding the correlation phase making it possible to make the DSM. It is also a determining factor concerning the accuracy of the spatial information which will be associated with each DSM by aero-triangulation.

3.3.1 Image acquisition parameters

The image acquisition parameters in aerial optical photography are as follows:

1. Position of the camera's optical center expressed as three-dimensional coordinates with respect to a reference frame \mathfrak{R} .
2. The camera's attitude angles with respect to \mathfrak{R} .
3. The camera's focal length f .
4. The focal plane's Principle Point of Symmetry (PPS).
5. Distortion of the optical system (negligible for the images of Beijing).
6. Affinity parameters: six parameters corresponding to the rotation, translation and scale factor modeling the image's digitization. It is necessary to introduce a rotation and a translation to take into account the non-alignment of the photograph with the scanner and the offset between the first pixel digitized and the one corresponding to the first pixel imaging the photograph. The scale factor corresponds to the scanner's enlargement (zoom).

The parameters provided by the Beijing Institute of Surveying and Mapping (BISM) are referenced with respect to the Beijing National System 1954 (BNS) which is a cartographic projection reference frame of the Gauss-Krüger type (UTM type with a central meridian scale factor equal to 1) associated with Krasovsky's ellipsoid. The camera's attitude angles are the parameters $(\varphi, \omega, \kappa)$ commonly used in photogrammetry (and not the pitch, roll and yaw (heading) angles (ϕ, θ, ψ) used in air navigation [11]). The angles (ϕ, θ, ψ) illustrated in figure 3.12 are referenced with respect to a plane tangential to the terrestrial ellipsoid (horizon) and to the North and East directions, whereas $(\varphi, \omega, \kappa)$ express the rotation between the camera reference marks and the terrestrial reference associated with a *cartographic projection* (figure 3.13). The relationship of the transition between these two triplets of angles is therefore not trivial and involves non-linear equations taken from the cartographic projection. The affinity parameters are not supplied directly but must be calculated using eight fiducial marks present on the borders of the image (figures A.1-3.14). These fiducial marks are expressed in millimeters in the optical system's focal plane, and can be identified manually or automatically in the image in pixel coordinates. The affinity parameters can then easily be estimated by resolving an over-determined linear system by the least squares' technique.

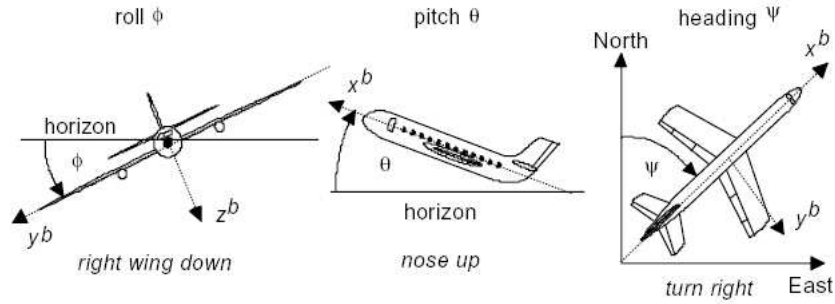


Figure 3.12: Definition of the attitude angles (roll, pitch, heading) of a camera on-board an aerial platform (assuming that the aircraft and the camera rotation axes are identical). The diagram is taken from [11].

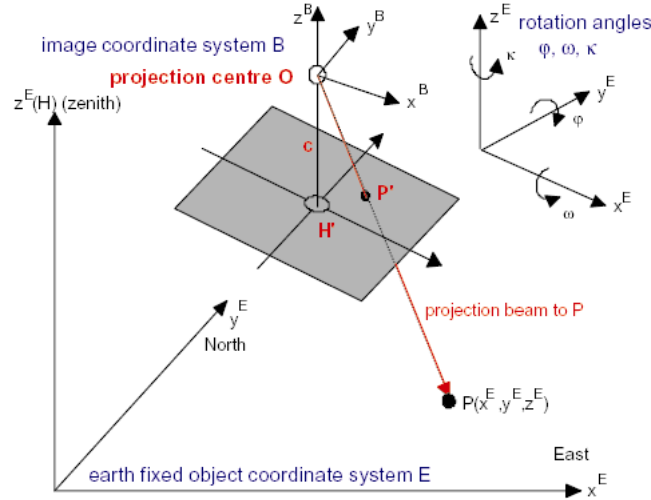


Figure 3.13: Illustration of the angles $(\varphi, \omega, \kappa)$ used in photogrammetry. These angles express the rotation between the camera reference (x^B, y^B, z^B) and the terrain reference (x^E, y^E, z^E) associated with a cartographic projection. The diagram is taken from [11].

3.3.1.1 Calculation of the affinity parameters: internal orientation

Eight fiducial marks are expressed in the camera focal plane coordinates.

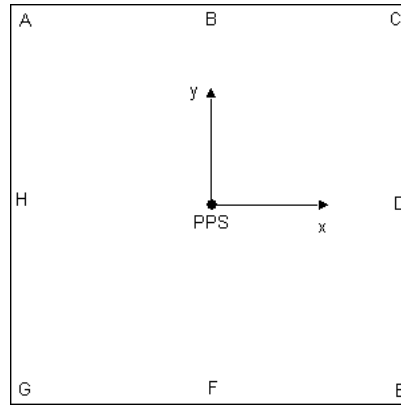


Figure 3.14: Eight fiducial marks represented in the camera's focal plane. The focal plane's reference mark is centered on the optical system's principal point of symmetry (PPS).

The coordinates in pixels of these same marks can also be acquired manually or automatically on the digitized images. It is then possible to solve the following system if you have at least three marks:

$$\begin{cases} lig = T_{lig} + a_{00}X_{camera} + a_{01}Y_{camera} \\ col = T_{col} + a_{10}X_{camera} + a_{11}Y_{camera} \end{cases}$$

where:

- * (col, lig) are the coordinates in pixels of a fiducial mark considered Φ in the image. These are measured on the image.
- * (X_{camera}, Y_{camera}) are the coordinates in millimeters in the focal plane of the mark Φ . These coordinates are provided by the camera manufacturer's specifications.
- * (T_{lig}, T_{col}) are the translation affinity parameters to be estimated.
- * $(a_{ij})_{i,j \in \{0,1\}}$ are the rotation/scale factor affinity parameters in to be estimated.

The resolution of such a linear system is carried out using the least squares technique. We have chosen to use the eight marks simultaneously to ensure more precise estimations. The residues on the eight marks vary between 0.33 and 0.36 pixels on the three images. These parameters are estimated on the non-subsampled aerial images in order to preserve the precision of the fiducial marks.

3.3.1.2 Assessment of the quality of the image acquisition parameters

As we underlined in the introduction to this section, the quality of the image acquisition parameters is a determining factor for the utilization of images with the 3D reconstruction chain. This quality must be examined according to two aspects. One concerns the precision of the geo-referencing information of the image's pixels in the 3D scene (absolute precision), the other is relative to the precision of the localization of a point of the scene present in both images of a stereographic pair (relative precision) with a view to putting it into epipolar geometry at a later time. The assessment of these two quality criteria relative to the image acquisition parameters requires given additional information. Indeed, the Ground Control Points (GCPs) expressed in the reference frame \mathfrak{R} and in pixel coordinates in the images are essential for assessing the absolute precision. Conjugated points identified in both images of a stereographic pair (Tie Points - TPs) make it possible to assess the relative precision. For the same reasons as in paragraph 3.3.1.1, the assessment of the quality is carried out on the non-subsampled images.

▷ Absolute precision

Formalization

The assessment of an image's absolute precision consists of projecting all its GCPs (expressed in the reference frame \mathfrak{R}) in the image by inverse modeling (appendix C.1). Inverse modeling enables, thanks to the image acquisition parameters, to calculate the position of one of the scene's 3D points in the image. It is then possible to calculate the mean distance between the points projected in the image and the corresponding imaged GCPs. Approximately ten GCPs per image have been provided by the BISM.

Let us note:

- * $GCP_{k,\mathfrak{R}}^l$ the k^{th} GCP of image l expressed in three dimensional coordinates with respect to the terrain reference frame \mathfrak{R} .
- * $GCP_{k,Im}^l$ the k^{th} GCP of image l expressed in pixel coordinates.
- * $GCP_{k,proj}^l$ the k^{th} GCP of image l projected into the image by inverse modeling from $GCP_{k,\mathfrak{R}}^l$.

An image l contains N_l GCPs, in our case $l = \{1, 2, 3\}$ and $N_l \simeq 10$. The absolute precision assessment criterion conferred by the image acquisition parameters for image l becomes:

$$d_{abs,l} = \frac{\sum_{k=1}^{N_l} \overline{GCP_{k,proj}^l GCP_{k,Im}^l}}{N_l}$$

Results

The absolute precision results with the image acquisition parameters provided are of the order of two or three pixels for images 1 and 3 (table 3.1). The parameters are not so good on the second image with nearly eight pixels' difference on average and a large standard deviation. An absolute optimization seems necessary in view of the above results.

| l | $d_{abs,l}$ in pixels | $\max \left(\overline{GCP_{k,proj}^l GCP_{k,Im}^l} \right)$ | $\sigma \left(\overline{GCP_{k,proj}^l GCP_{k,Im}^l} \right)$ |
|---|-----------------------|--|--|
| 1 | 2.3 | 4.8 | 1 |
| 2 | 7.5 | 14.1 | 4.7 |
| 3 | 2.7 | 6.4 | 1.5 |

Table 3.1: Results in pixels of the absolute precision quantification of the image acquisition parameters for the three aerial images.

▷ Relative precision

Formalization

The assessment of the relative precision consists of projecting all of an image's TPs in the form of epipolar straight lines into the conjugated stereographic image. The relative precision criterion is the average of the distances between the calculated epipolar straight lines and the TPs in the second image. To determine an epipolar straight line in the second image of a stereographic couple, a TP of the first image is projected into the terrain reference frame \mathfrak{R} by direct modeling at two different altitudes: alt_{min} and alt_{max} (appendix C.2). These two 3D points are then projected into the second image by inverse modeling. The projection into the reference frame \mathfrak{R} requires the image acquisition parameters of the first image whereas those of the second image are used for the inverse modeling.

Let us note:

- * $TP_{k,1}^m$ the k^{th} TP of the first image in the m^{th} stereoscopic pair. It is expressed in pixel coordinates.
- * $TP_{k,2}^m$ the k^{th} TP of the second image in the m^{th} stereoscopic pair. It is expressed in pixel coordinates.
- * $EPI_{k,proj}^m$ is the k^{th} epipolar straight line projected into the second image of the m^{th} stereoscopic pair. It is calculated from $TP_{k,1}^m$.

The two images of the m^{th} stereoscopic pair each contain N_m TPs, in our case $m = \{1, 2\}$ and $N_m = 10$. The relative precision assessment criterion for the pair m becomes:

$$d_{rel,m} = \frac{\sum_{k=1}^{N_m} \text{dist} \left(EPI_{k,proj}^m, TP_{k,2}^m \right)}{N_m} \quad (3.1)$$

The distance function $\text{dist}(\cdot)$ expressed in equation 3.1 is the distance from a point to a straight line, i.e. the distance between the point and the intersection between the straight line and the normal to the latter passing through the point. The TPs were extracted manually and then refined by sub-pixel correlation [14]. Twenty TPs are therefore selected and the ten best ones are kept in the end.

Results

The relative precision results obtained are not good enough to be used by the 3D reconstruction chain (table 3.2). Indeed, a relative precision lower than three pixels is required for the prior transformation into epipolar geometry phase. The precision of the image acquisition parameters provided by the BISM is not sufficient to ensure a high-quality geo-referencing and the transformation into epipolar geometry of the images. These parameters will therefore have to be optimized.

| m | $d_{rel,m}$ in pixels | $\max \left(\text{dist} \left(EPI_{k,proj}^m, TP_{k,\in}^m \right) \right)$ | $\sigma \left(\text{dist} \left(EPI_{k,proj}^m, TP_{k,\in}^m \right) \right)$ |
|---|-----------------------|---|---|
| 1 | 1.7 | 3 | 0.6 |
| 2 | 3.9 | 7.7 | 1.8 |

Table 3.2: Results in pixels of the quantification of the relative precision of the image acquisition parameters for each of the pairs of aerial images.

3.3.2 Overall optimization

3.3.2.1 Cost function

The optimization of the image acquisition parameters for the three images must take into account both the absolute and the relative precision. The optimized parameters will thus make it possible to transform the images into epipolar geometry (to ensure the success of the 1D correlation of the reconstruction software) while providing high quality geographical localization information (each pixel of the resulting DSM has longitude,

latitude and elevation coordinates calculated by aerial triangulation). The optimization is carried out on the non-subsampled images. It will be sufficient to divide the affinity parameters by the subsampling factor to make them compatible with the reduced-size images at input to the reconstruction software. The fact that we have more than two images is an advantage for the quality of the optimization. Indeed, the information redundancy will make it possible to better minimize each distance $d_{abs,l}$ and $d_{rel,m}$ involved in the optimization. The cost function J_{param} to be minimized is then:

$$J_{param} = \sum_{l=1}^n d_{abs,l} + \sum_{m=1}^{n-1} d_{rel,m} \quad (3.2)$$

Remark: if one zone in the scene is imaged by more than two photographs, we will only consider the pair that has the greatest overlap. So, for n images, we will assume that there are $n - 1$ stereoscopic pairs.

The cost function given in equation (3.2) is not a linear function of the image acquisition parameters. The approach commonly used to minimize J_{param} consists of linearizing the cost function in order to achieve the optimization. We have chosen to keep J_{param} in its original non-linear form and use the simplex algorithm to estimate the optimum image acquisition parameters. The simplex is a robust optimization algorithm suited to non-linear functions; the detail of this method is given in appendix B. Our reasons for choosing the simplex algorithm are as follows:

- * since the simplex optimizes the non-linear functions directly, it is not necessary to linearize the cost function J_{param} .
- * since J_{param} is not linearized, we are likely to obtain a more precise estimation of the image acquisition parameters.
- * the simplex does not have any setting parameters, which makes it easier to use.

The degrees of freedom for optimization are all the image acquisition parameters, however the focal length and the PPS are assumed to be identical whichever image we consider. The total number of degrees of freedom for an image is 15. For n images there will be $12n + 3$ (that is to say 39 in this study).

3.3.2.2 Geo-centered Cartesian reference frame

The image acquisition parameters at input to the reconstruction software must be referenced with respect to a geo-centered direct Cartesian reference frame (this concerns the expression of the camera's position and attitude). The origin of such a reference frame is the Earth's center of gravity. The (OZ) axis is the Earth's rotation axis and it points northwards. The (OX) axis is such that the Greenwich meridian is included in the (OXZ) plane. (OY) is such that the $(OXYZ)$ reference is direct. Now, \mathfrak{R} designates this reference frame (see [39] for other commonly used reference frames).

The position of the camera's optical center for each image acquisition must now be expressed in this reference \mathfrak{R} . Since we have the position in the BNS projection reference, the latter had to be converted using the equations of [39, pp. 100-101]. As for the optical system's attitude angles, it is delicate to attempt to convert them from the BNS system to \mathfrak{R} . The main difficulty is the different nature of the reference frames used: one is in three dimensions (\mathfrak{R}), and the other is the result of a transverse cylindrical projection (BNS). It was simpler to estimate these angles by absolute optimization on each image by using the GCPs converted beforehand into the system \mathfrak{R} . It is legitimate to ask whether these angles, initialized at 0, could not have been estimated at the time of the overall optimization. Experience has shown that the calculation of these angles fails in this case, hence the need to do this beforehand by absolute optimization on each image. The diagram in figure 3.15 summarizes the overall functioning of the image acquisition parameters optimization program.

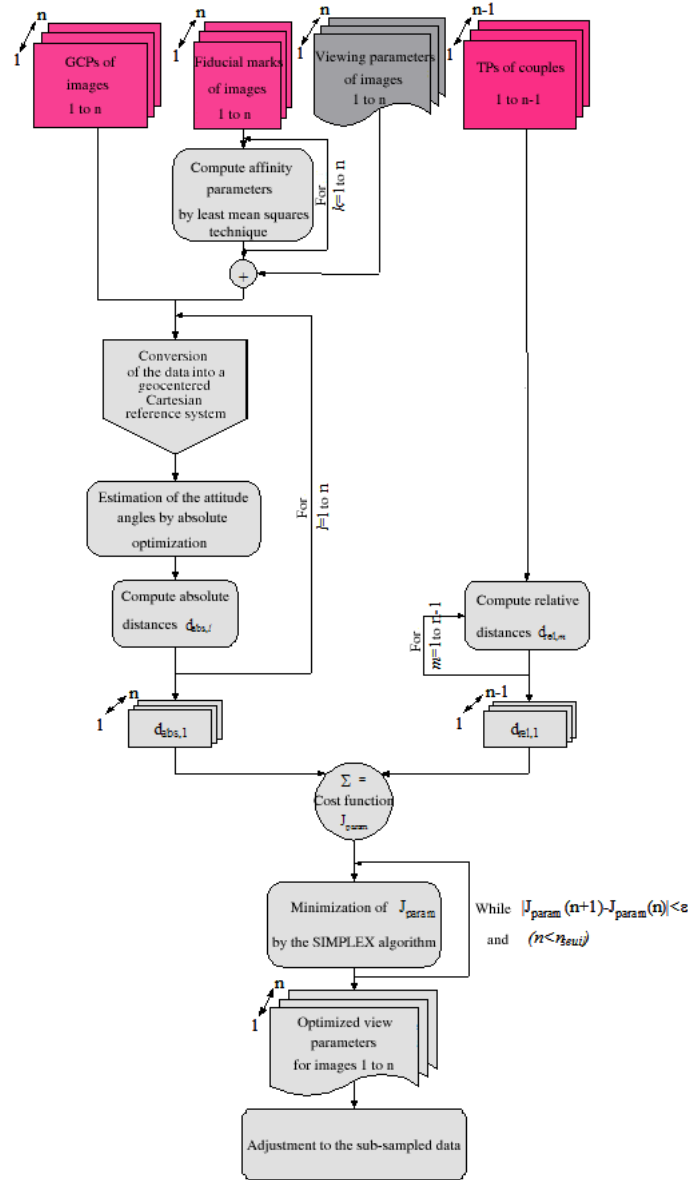


Figure 3.15: Algorithmic summary of the optimization of the image acquisition parameters of n aerial images.

3.3.2.3 Results

The results obtained with the image acquisition parameters of the three images of Beijing are given in table 3.3.

| l | $d_{abs,l}$ in pixels | $\max \left(dist \left(GCP_{k,proj}^l, GCP_{k,Im}^l \right) \right)$ | $\sigma \left(dist \left(GCP_{k,proj}^l, GCP_{k,Im}^l \right) \right)$ |
|-----|-----------------------|--|--|
| 1 | 1.5 | 4.5 | 1.5 |
| 2 | 1.8 | 4.1 | 1.2 |
| 3 | 1.8 | 5.1 | 1.5 |
| m | $d_{rel,m}$ in pixels | $\max \left(dist \left(EPI_{k,proj}^m, TP_{k,\in}^m \right) \right)$ | $\sigma \left(dist \left(EPI_{k,proj}^m, TP_{k,\in}^m \right) \right)$ |
| 1 | 0.3 | 1 | 0.3 |
| 2 | 0.3 | 0.65 | 0.25 |

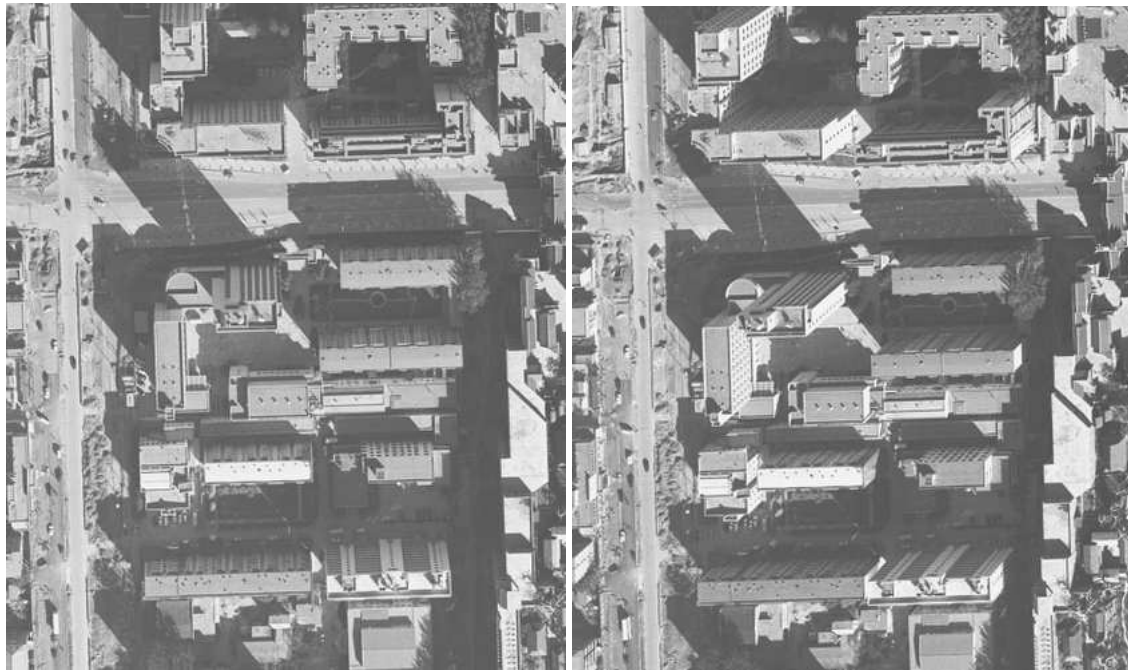
Table 3.3: Results of the optimization of the image acquisition parameters on the images with a resolution of 0.21 m/pixel.

If we carry these results over to the subsampled images used by the 3D reconstruction chain, we have:

| l | $d_{abs,l}$ in pixels |
|-----|-----------------------|
| 1 | 0.5 |
| 2 | 0.6 |
| 3 | 0.6 |
| m | $d_{rel,m}$ in pixels |
| 1 | 0.1 |
| 2 | 0.1 |

Table 3.4: Results of the optimization of the image acquisition parameters on the images subsampled by a factor of three.

It can be seen that the optimization has made it possible to improve the precision of the image acquisition parameters, and significantly so for the relative precision. It is nevertheless important to mention that these results are sensitive to the distribution and to the number of control points in the images. Approximately ten points uniformly distributed over each image are required to obtain satisfactory results. Figure 3.17 illustrates the DSM generated by the 3D reconstruction chain with the optimized parameters and the stereoscopic pair shown in figure 3.16.



(a) West image

(b) East image

Figure 3.16: Pair of stereoscopic images rectified in epipolar geometry.

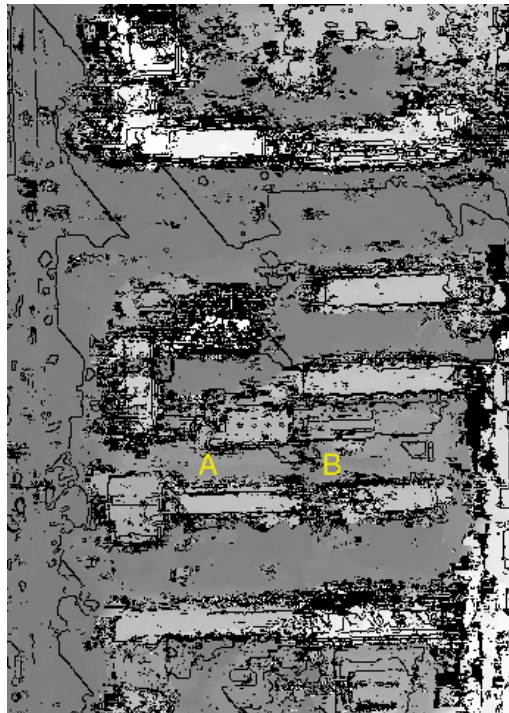


Figure 3.17: Generated DSM in epipolar geometry. Unfilled pixels are in black; intensity is proportional to the altitude.

The DSM generated has numerous pixels unfilled with altitude information (in black on the image in figure 3.17) at the level of the constructions. These are in fact zones where there are occlusions and where the facades of the buildings are difficult to correlate. The artificial contours due to the shadows and the objects present on one of the images (cars, etc.) result in small unfilled zones on the road. However, the contours of the buildings are well preserved, and most of them have a relatively homogeneous altitude. The shadows cast on the ground are also managed well (no detection of false buildings). Even if this image is visually imperfect, it does nevertheless contain a pertinent information enabling to locate the buildings and distinguish them from the ground in most cases. The limits of the correlation are illustrated at the places annotated A and B on the image in figure 3.17: these zones seem to belong to the ground (region of low intensity, consequently of low elevation) when in fact they correspond to two buildings (figure 3.16). Buildings A and B represent an extreme case of the dense urban environment. They are in fact enclaves in a zone of shadow that makes it difficult to observe their disparity and contours.

3.3.3 Assessment of the geocoding quality of the DSMs

The assessment of the quality of the DSM geocoding has been carried out using the given Ground Control Points (GCPs). Each control point is expressed in pixel coordinates in the aerial images and in 3D coordinates in the Beijing cartographic system: East (E), North (N) and elevation (Z). The cartographic coordinates are expressed in meters, the altitude Z is referenced with respect to the Krasovsky ellipsoid. It will be possible to quantify the quality of the DSM when the coordinates of these GCPs are measured in the DSM and then compared with their ground truth. The first step consists of identifying the control points on the subsampled epipolar geometry reference images that were used to generate the DSM. Since there is an exact correspondence between the reference images and the epipolar DSMs generated from them, it is possible to read the values (E, N, Z) in the DSMs for the GCPs considered. We have generated two epipolar DSMs from 3 stereo-pair images. For each DSM, we have about ten GCPs some of which were not used in the optimization of the image acquisition parameters. The results are given in tables 3.5 and 3.6.

| <i>GCP</i> | <i>Absolute Error</i> | | | |
|---------------------------|-----------------------|------------|------------|----------------|
| | ΔN | ΔE | ΔZ | $\Delta Plane$ |
| 1 | 1.0 | 0.1 | 0.8 | 1.0 |
| 2 | -0.5 | -0.1 | 0.4 | 0.5 |
| 3 | 1.5 | 0.6 | 2.9 | 1.6 |
| 4 | -0.5 | 0.8 | 1.5 | 0.9 |
| 5 | 0.0 | 0.6 | 2.4 | 0.6 |
| 6 | 0.0 | -1.5 | 0.2 | 1.5 |
| 7 | 0.0 | 0.0 | -0.5 | 0.0 |
| 8 | -0.5 | -0.5 | 0.5 | 0.7 |
| 9 | 0.0 | -0.2 | 0.8 | 0.2 |
| 10 | -1.0 | -0.1 | 0.2 | 1.0 |
| 11 | -0.5 | 0.5 | -0.3 | 0.7 |
| 12 | 0.0 | -0.3 | 0.2 | 0.3 |
| Mean | 0.0 | 0.0 | 0.8 | 0.7 |
| Standard deviation | 0.7 | 0.6 | 1.0 | 0.5 |

Table 3.5: Absolute error between the GCPs and the 3D reconstructions of the DSM_1 . ΔN : error in the northern direction, ΔE : error in the eastern direction, ΔZ altimetric error, error in the plane: $\Delta Plane = \sqrt{\Delta N^2 + \Delta E^2}$.

| <i>GCP</i> | <i>Absolute Error</i> | | | |
|---------------------------|-----------------------|------------|------------|----------------|
| | ΔN | ΔE | ΔZ | $\Delta Plane$ |
| 1 | 0.5 | -0.9 | 1.0 | 1.0 |
| 2 | 0.0 | -0.2 | -0.2 | 0.2 |
| 3 | -0.5 | 0.6 | 2.7 | 0.8 |
| 4 | -0.5 | 0.1 | 1.6 | 0.5 |
| 5 | 0.5 | 0.0 | 1.4 | 0.5 |
| 6 | 2.5 | -3.4 | -0.8 | 4.2 |
| 7 | -1.5 | 0.5 | 0.4 | 1.5 |
| 8 | 0.0 | 0.2 | 1.1 | 0.2 |
| 9 | 0.0 | -0.2 | 0.5 | 0.2 |
| 10 | -0.5 | 0.2 | 2.2 | 0.5 |
| 11 | -1.5 | 0.4 | -0.2 | 1.5 |
| 12 | 0.0 | 0.8 | -1.3 | 0.8 |
| 13 | 0.0 | -0.2 | 0.1 | 0.2 |
| 14 | 0.0 | -0.5 | -0.3 | 0.5 |
| Mean | -0.1 | -0.2 | 0.6 | 0.9 |
| Standard deviation | 1.0 | 1.0 | 1.1 | 1.1 |

Table 3.6: Absolute error between the GCPs and the 3D reconstructions of the DSM_2 . ΔN : error in the northern direction, ΔE : error in the eastern direction, ΔZ altimetric error, error in the plane: $\Delta Plane = \sqrt{\Delta N^2 + \Delta E^2}$.

The pertinent quantities in these tables are the standard deviations which represent the precision of the geocoding in the North, East, Z direction or in the plane. It can be seen that the planimetric precision varies between 0.5 m and 1 m, and the altimetric precision is 1 m. These results seem satisfactory and validate the preliminary phase of optimization of the camera's image acquisition parameters, and the aerial triangulation calculation of the 3D chain.

3.4 Orthorectification of an epipolar DSM

Orthorectification consists of correcting the distortion due to the perspective that is present in the image. Each pixel of an image in orthoscopic geometry is such that it will be seen at the nadir. The need to orthorectify the images of urban scenes only arises for very high resolution images representing a high relief or tall buildings. Once the image has been orthorectified, it can be superimposed over the cartographic data which are also in this geometry. The DSM provided by the 3D reconstruction chain is in epipolar geometry. The geo-referencing information associated with each pixel of the DSM will make it possible to orthorectify the latter.

3.4.1 Orthorectification by the direct method

Direct method orthorectification consists of considering each pixel of the DSM in epipolar geometry and projecting it into a geocoded terrain grid [81]. Several pixels of the epipolar DSM can be projected into any given planimetric localization in the orthoscopic grid. Only the highest altitude pixel will be kept (*z-buffer* technique). The orthoscopic grid can be coded in longitude/latitude or according to a cartographic projection. In our case, we want to superpose the DSM with the cartographic data. Consequently, the longitude/latitude information present in the epipolar DSM generated has been converted into (E, N, Z) coordinates of the BNS cartographic projection [39, pp. 99]. Each pixel is then projected into the orthoscopic grid which is itself associated with the BNS reference frame. The relationship between the pixels in the orthoscopic grid and the BNS terrain coordinates is of the type:

$$\begin{cases} E = E_{min} + iR \\ N = N_{max} - jR \end{cases} \quad (3.3)$$

where:

- * R is the grid's pitch in meters.
- * (E_{min}, N_{max}) are the BNS terrain coordinates of the origin contained in the epipolar DSM.
- * (i, j) are the pixel coordinates of the orthoscopic grid, respectively column and line.
- * (E, N) are the BNS terrain coordinates associated with the pixel (i, j) .

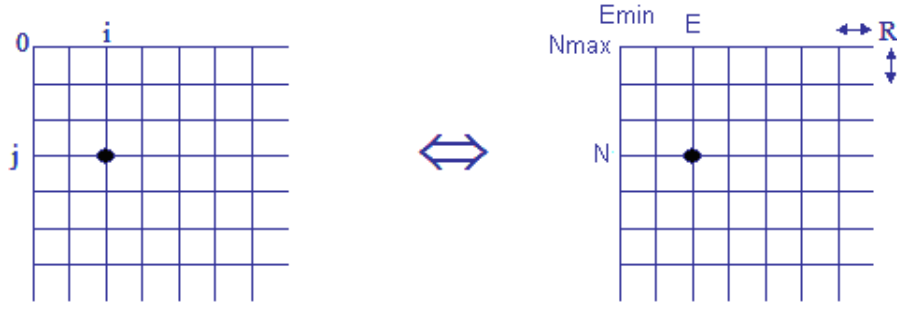


Figure 3.18: Orthoscopic grid associated with the cartographic projection.

3.4.1.1 Triangular interpolation

It is not possible to obtain a high-quality ortho-DSM by individually projecting each pixel of the epipolar DSM into the orthoscopic grid. Indeed, the pixel projected into the orthoscopic grid will have non-integer coordinates. The contribution of the projected pixel towards its block will have to be subject to interpolation. The main question is to decide to which block the pixel will contribute. If the block is too small, certain points of the orthoscopic grid will not be filled and the result will not be very dense. If the block of influence is too high, it is possible that points in the grid will have an incorrect altitude value assigned to them. This problem will be all the more sensitive at the level of the roofs of the buildings.

One way of settling this question is to adopt a triangular interpolation. Instead of considering each pixel in the epipolar DSM individually, let us consider an isosceles triangle whose equal sides are one pixel long. This triangle (ABC) of the epipolar DSM is shown in figure 3.19 (on the left). This triangle is then projected into the orthoscopic grid thanks to the geocoding information associated with each pixel A , B and C and thanks to the equations system (3.3).

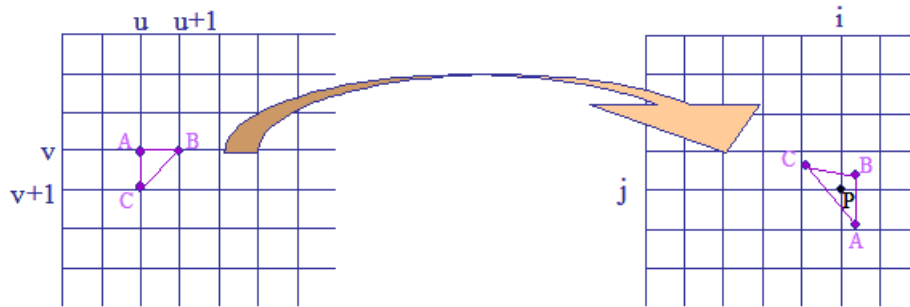


Figure 3.19: Projection of a triangle of the epipolar DSM to the orthoscopic grid.

The points of the orthoscopic grid included in the projected triangle will have an altitude and a radiometry assigned to them according to the following interpolation

formulas:

$$z_{ij} = \frac{z_A \overline{PB} \overline{PC} + z_B \overline{PA} \overline{PC} + z_C \overline{PA} \overline{PB}}{\overline{PA} \overline{PB} + \overline{PB} \overline{PC} + \overline{PC} \overline{PA}} \quad (3.4)$$

where:

- * z_{ij} is the altitude of point P of the grid and coordinates (i, j) . P is included in the projected triangle (A, B, C) .
- * z_X is the altitude of the point X ($X = A, B$ or C).

If the distortion due to the perspective is large, a grid pixel will only rarely be included in a single projected triangle. To illustrate this, let us take the example of an edge of a parallelepiped building subject to a perspective distortion in the image. All of this edge pixels have the same planimetric position (E_0, N_0) with different altitudes. In the orthoscopic grid, this point will therefore be included in several triangles (figure 3.20).

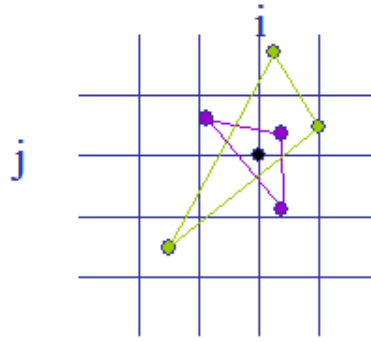


Figure 3.20: Multiple triangles projected into the orthoscopic grid and intersecting with the same pixel.

Our goal is to orthorectify the epipolar DSM, i.e. to obtain a vertical view for each pixel in the orthoscopic grid. So, for a given pixel in the grid, we will assign the maximum altitude amongst those calculated by the various triangles that include this point.

3.4.1.2 Algorithm

The direct orthorectification algorithm with triangular interpolation is as follows:

1. Initialization of the image of the orthoscopic DSM “ DSM_{ortho} ” to -1 (-1 is the value symbolizing indetermination: no altitude value provided); fix the value of the orthoscopic grid R ’s pitch in meters.
2. Convert the geodetic coordinates of the epipolar DSM “ DSM_{epi} ” into BNS cartographic coordinates per [39, pp. 99]. Determine the values E_{min} and N_{max} of the converted DSM_{epi} .

3. For each pixel (u, v) of DSM_{epi} , we note $B = (u + 1, v)$ and $C = (u, v + 1)$:
 - (a) For $A = (u, v)$ and $A = (u + 1, v + 1)$
 - i. If the pixels A, B and C are all filled:
 - * Projection of the triangle (ABC) onto the orthoscopic grid per equation 3.3 which becomes $(A'B'C')$.
 - * For each pixel (i, j) of the orthoscopic grid included in $(A'B'C')$:
 - \leadsto Calculate the altitude z_{ij} per equation (3.4).
 - \leadsto $DSM_{ortho}(i, j) = \max(DSM_{ortho}(i, j), z_{ij})$

3.4.1.3 Results

Figure 3.21 illustrates a DSM orthorectified and projected into the BNS system with a pitch $R = 0.65$ m. The comments relative to the large number of unfilled pixels in figure 3.21 and to their localization at the level of the constructions are identical to those made for the DSM in epipolar geometry (figure 3.17). However, we note the appearance of light scratch type artifacts. These artifacts are inherent to the triangular interpolation technique developed in section 3.4.1.1 and to the correlation errors contained in the epipolar DSM. Indeed, certain pixels of the epipolar DSM are the result of an erroneous correlation generating the calculation of an abnormally high altitude with respect to the pixel's environment. The planimetric coordinates of this point are also false. A triangle involving one or more of these corrupted pixels will include an abnormally high number of pixels at the time of its projection into the orthoscopic grid. Furthermore, these pixels will receive an altitude that is too high which explains the light-colored appearance of these artifacts.



Figure 3.21: DSM orthorectified with a pitch of $R = 0.65$ m and $\varepsilon_{corr} = 0.4$.

There are several ways of attenuating this problem. The first one consists of modifying the rejection threshold for the inter-correlation score ε_{corr} involved when pairing the stereoscopic images. The result given in figure 3.21 was generated with a threshold $\varepsilon_{corr} = 0.4$. The danger of increasing ε_{corr} is that it may make the decision to fill a pixel more selective, and the indeterminate zones will then grow larger.

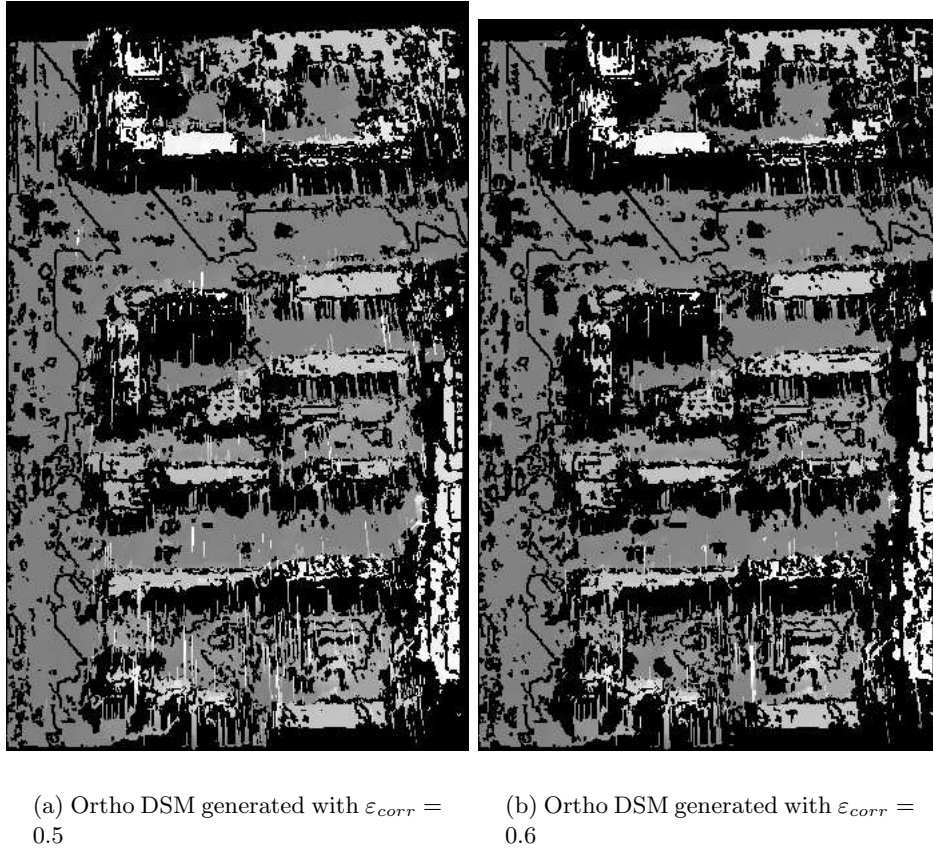


Figure 3.22: Orthorectified DSM with different correlation thresholds.

The comparison of figures 3.21, 3.22.a and 3.22.b shows the propagation of the indeterminate zones according to the correlation threshold ε_{corr} . The artifacts are attenuated when ε_{corr} increases.

3.4.2 Pre-processing on the disparity image

Another way of reducing these artifacts consists of filtering the disparity image which is used to generate the epipolar DSM. Two filters have been tried to attempt to eliminate the erroneous pixels. A median filtering of size $N \times N$ ($N = 3$ or 7) and a rejection filter of size $N \times N$ ($N = 7$). The rejection filter considers the current pixel to be indeterminate if there are more than x% of unfilled pixels in the window. This filter was designed subsequent to the observation of erroneous pixels isolated by unfilled pixels.

3.4.2.1 Median filtering

The median filtering whose results are shown in figure 3.23 was only carried out on the filled pixels in the disparity image. It can be seen that this filtering enables to attenuate the artifacts without eliminating them completely. Applying a large window does not

reduce the artifacts better. The artifacts are also less well attenuated than by increasing the correlation threshold.

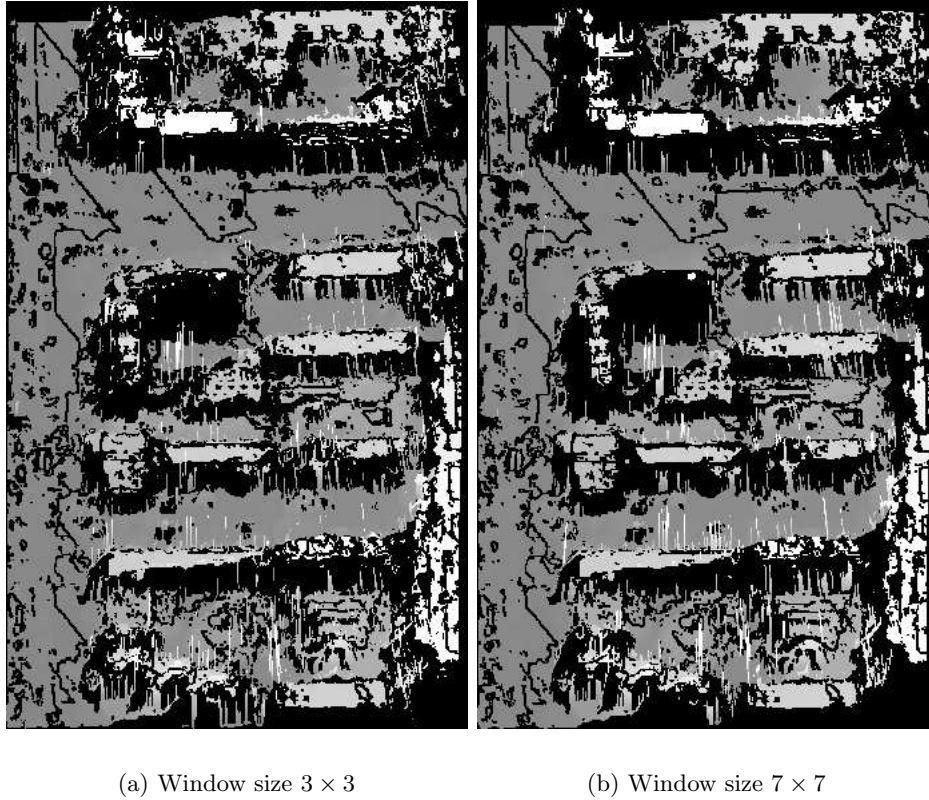
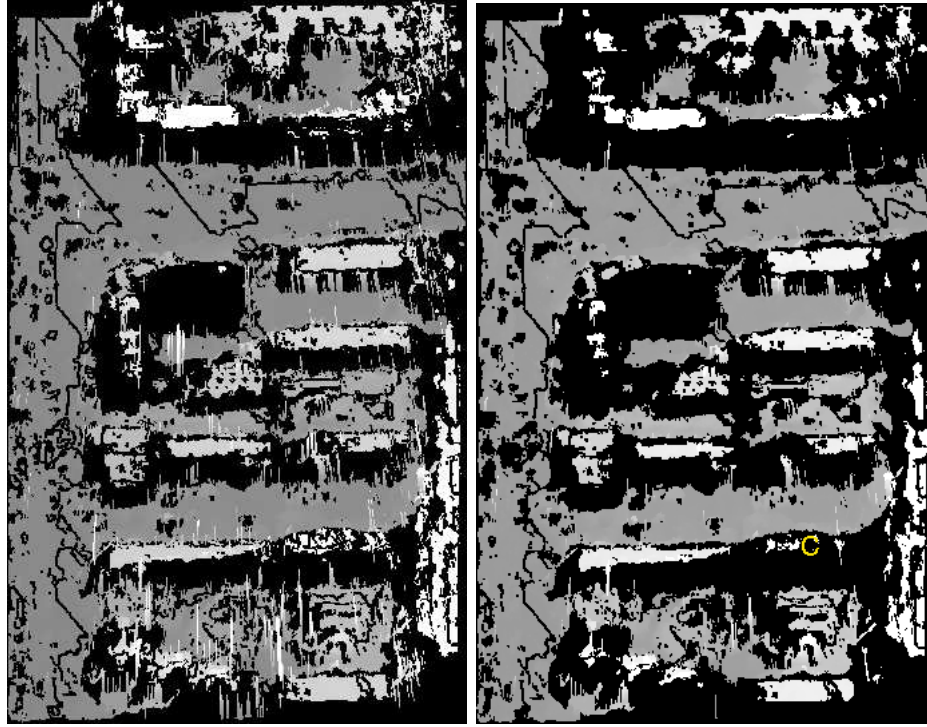


Figure 3.23: Orthoscopic DSM made with a disparity image filtered with a median filter.

3.4.2.2 Rejection filter

The comparison of figures 3.21, 3.24.a and 3.24.b clearly shows the influence of the rejection threshold for a window of fixed size. When the threshold is lowered, the artifacts are more or less eliminated but the unfilled zones spread to such an extent that they mask some of the pertinent information (e.g. building eaten away and annotated with the letter C in figure 3.24.b).



(a) Rejection at 50 per-cent

(b) Rejection at 30 per-cent

Figure 3.24: Ortho-DSM made with a disparity image filtered with a rejection filter of size 7×7 .

The edges of the roofs of buildings are also affected by the filter's rejection filter. Whereas the correlation algorithm inserted in the 3D reconstruction chain preserves the edges of the buildings, this filter tends to eat away at them and to break up the roof altitude information.

3.4.2.3 Summary

Several methods have been tested for reducing the artifacts due to the disparity image's erroneous pixels. None of them is perfect in absolute terms since each one significantly deteriorates the information contained in the DSM. Experimentally, we have observed that the best results were obtained by successively using all the methods described and with parameters that do not deteriorate the processed image too much. We can thus slightly increase the correlation threshold when generating the epipolar DSM ($\varepsilon_{corr} = 0.5$) in order to avoid generating too many indeterminate pixels while eliminating some erroneous pixels. The disparity image can then be filtered by a median filter (3×3), then by the 7×7 rejection filter with a threshold of 60% unfilled pixels. Figure 3.26.b shows the result of orthorectification with this type of pre-processing of the disparity image on another zone of the DSM.

3.5 Merging DSMs

The precise geocoding image contained in each orthorectified DSM will allow us to merge the two geocoded DSMs. The algorithm developed for this purpose is broadly based on the *correlar* function used by the 3D reconstruction chain. Contrary to the *correlar* function which merges the disparity images, we merge orthorectified and geocoded DSMs. This algorithm attempts to merge these DSMs to the best taking into account the inter-DSM coherence (validation of the information if the values delivered by the other DSMs are close) and inter-DSM coherence (validation of the merger information with respect to its coherence relative to the neighborhood). In our case the fusion is more like a “concatenation” or a “mosaicking” since the DSMs’ overlap area only represents a quarter of their surface. The aim of the fusion algorithm is to be as general possible while considering the merging of N and not just two orthorectified DSMs. This algorithm is expressed as follows:

I) Determining the encompassing grid: Each DSM has a perfectly known extent in the cartographic system in which they are projected. The first step consists of building the geocoded grid of the final DSM, resulting from the merging. This grid is the smallest grid containing the N DSMs to be merged.

II) Loop on each pixel in the encompassing grid:

For each pixel pix in the grid:

1. Test of belonging to the N DSMs
 - (a) For $i \in \{1, \dots, N\}$ test whether $pix \in MNS_i$;
 - (b) If pix does not belong to any DSM: **rejection**
 - (c) If pix belongs to a single DSM: **assign the DSM value to pix** (whether the value is filled or not)
 - (d) If pix belongs to m DSMs of which m_r are filled ($m \in \{2, \dots, N\}$)
 - i. If $m_r = 0$, **rejection**
 - ii. if $m_r = 1$, **assign the value of the single DSM filled at that point to pix**
 - iii. If $m_r \in \{2, \dots, m\}$
 - A. If the m_r points of the m_r DSMs are coherent with each other:
 - * **assign to pix the mean value of the points that are coherent with respect to their block**¹
 - * If none of the points are coherent with the block: **rejection**
 - B. If the $m_r - k$ points of the m_r DSMs are coherent with respect to each other ($k \in \{1, \dots, m_r - 2\}$), we choose the two points that are the most coherent with each other:
 - * If these two points are coherent with respect to the block: **we assign their mean value to pix .**

¹We perform the same operation if it is impossible to calculate the coherence with respect to the block of m_r points.

- * If only one point is coherent with the block: **we assign the value of that point to pix .**
- * Otherwise: **rejection**
- C. If the $m_r - k$ points of the m_r DSMs are not coherent with each other:
 - * If the coherence with respect to the block can be calculated: **we assign to pix the value of the point that is most coherent with respect to the block.**
 - * Otherwise: **rejection**

III) End.

In this algorithm, the notions of coherence are defined as follows:

- * Coherence between two points p_i and p_j taken from two DSM_i, DSM_j : p_i and p_j are coherent with each other $\iff |altitude(p_i) - altitude(p_j)| < \varepsilon_{inter-coherence}$.
- * Coherence with respect to the block of a point p_i taken from DSM_i : p_i is coherent with respect to the block $\iff |altitude(p_i) - altitude(voisinage)| < \varepsilon_{intra-coherence}$.

The altitude of the block is the mean of the points of the merged DSM (result) localized in a causal window (the pixel of interest is in the bottom right-hand part of the window). Typically, the calculations have been carried out with a 5×5 pixel window. Since the DSMs have unfilled pixels that can hinder the calculation of the mean of the coherence with respect to the block (if there are too few filled points in the window this calculation will be meaningless), we consider that this coherence is "incalculable" if more than 50% of the pixels in the window are unfilled.

The result of merging the DSMs illustrated in figures 3.25-3.26 was achieved with $\varepsilon_{inter-coherence} = \varepsilon_{intra-coherence} = 5m$. Figure 3.25 illustrates the merger of two DSMs. Figure 3.26 shows more precisely the zone framed in red.

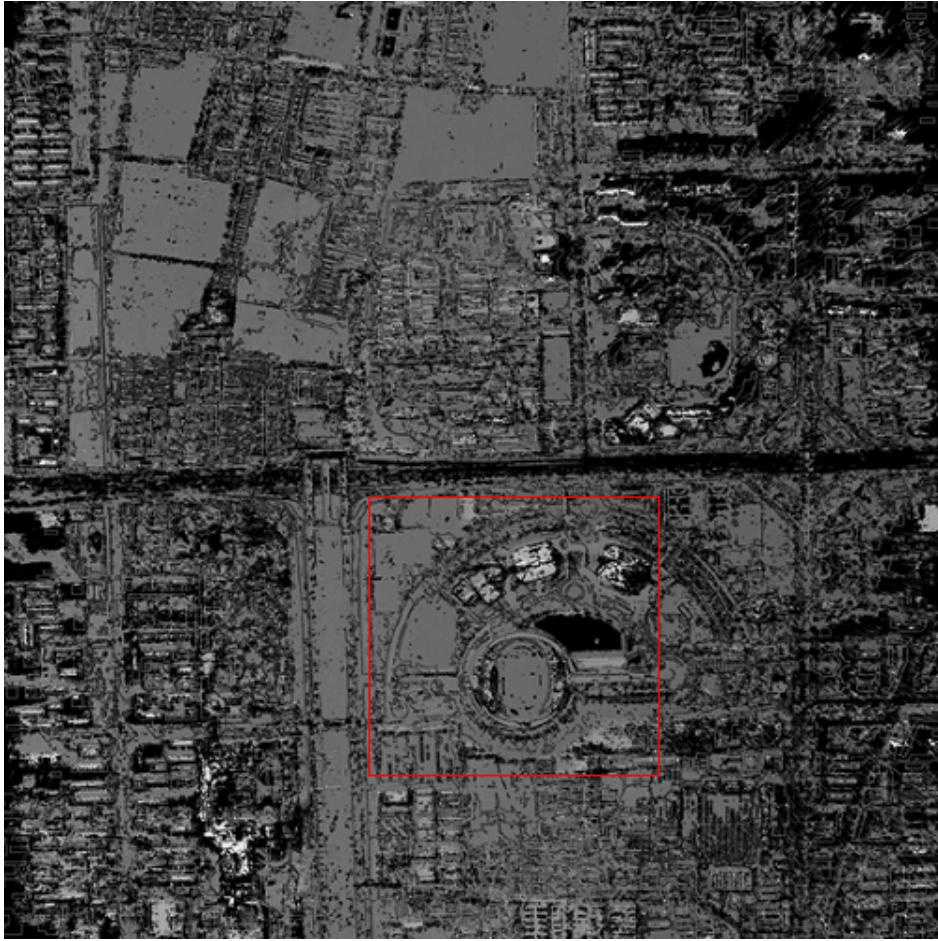
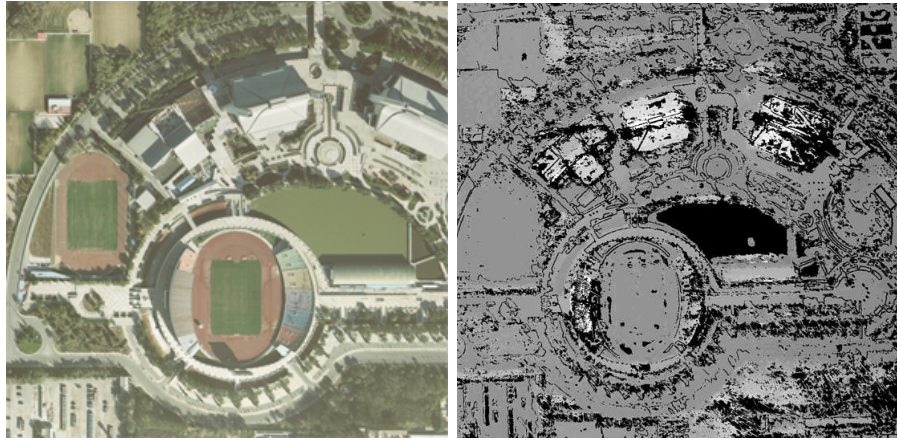


Figure 3.25: Merged DSM: $DSM_1 + DSM_2$.



(a) Extract from the aerial image

(b) Merged DSM corresponding to image (a)

Figure 3.26: Enlargement of the red rectangular zone in figure 3.25.

3.6 Conclusion

This chapter presented the work carried out on generating orthoscopic DSMs. The main contribution of this study is the optimization of the image acquisition parameters of aerial images. These parameters known beforehand with insufficient precision have been optimized by the simplex algorithm thanks to the given GCPs and TP. Particular attention has been paid to the quantitative assessment of the accuracy of the geocoding of the DSMs generated by the reconstruction chain with the optimized parameters. The reconstructed DSMs reach a planimetric and altimetric precision of the order of one meter, which validates the method of optimization of the image acquisition parameters. It is nevertheless important to note that the quality of the optimization depends on the number and on the uniform distribution of the GCPs used. We have also detailed the orthorectification and merging of several geocoded ortho-DSMs. The final merged and orthoscopic DSM can then be superimposed with the cartographic data and the satellite image thanks to the geocoding information. This superposing will be all the more coherent with cartographic data that are also in orthoscopic geometry.

Lastly, we have been able to note, thanks to the results shown here, that the 3D reconstruction produces DSMs with many unfilled pixels. In the DSM in figure 3.25, they represent 33% of the image. These indeterminate zones are mainly localized around buildings because of the occlusions. Indeed, two images are not sufficient for providing the altitude of the whole scene. The current trend is to use triplets or quadruples of images to reconstruct a landscape relief [136]. The DSMs generated in this study therefore have a coarse appearance and would seem to be difficult to use. The correlation errors also contribute to this. These are inherent to the difficulty of urban areas which have numerous details and deteriorate the correlation. It should also be noted that the

signal-to-noise conditions were not optimal: since the images were analog, they were scanned, and digitization inevitably introduces noise. We noted that the blue spectral band was particularly noisy which is due to a defect either in the camera or in the scanner. Nevertheless, we will see in the next chapter that the imperfect appearance of the ortho-DSM produced will not be an obstacle to its utilization as an additional source of information completing the satellite image.

Chapter 4

Presentation and choice of shape constrained active contours for map-to-image fine matching

4.1 Introduction

This chapter presents the deformable models called “active contours” and which will be used to match the buildings symbolized in the map with their homologous representation in a high-resolution panchromatic satellite image. We use the flexibility of active contours and their potential for incorporating high-level information to achieve this fine matching, the goal of which is twofold: i) the cartographic objects will have a finer spatial localization if the image has greater precision than what is available in the cartographic data, which corresponds to one of the components of *map updating*. ii) this matching enables to reduce the map-to-image exogenous variabilities which are source of errors whenever detecting later changes. The urban scenes represented in the Quickbird satellite images used have a high level of detail and certain *urban artifacts* that can only be detected at very high resolution (occlusions, shadows, low contrast of the objects, etc.). In order to overcome these difficulties, we propose to take advantage of the prior knowledge contained in the map. As a general rule, there are three aspects to the information provided by the map. It indicates the nature of the object to be matched in the image (building, road, etc.), its location (the cartographic and remote sensing data are assumed to be globally superimposed), and its shape. The location information will make it possible to initialize the active contour close to the building represented in the image, and its shape will be constrained by the silhouette of the building symbolized in the map in order to overcome the difficulties associated with urban images. In the first part of this chapter we will describe the state of the art in the area of active contours. We will classify the various models according to their mode of representation, attachment to the data and regularization, with special attention being paid to the insertion of exogenous shape constraints known beforehand. Lastly, we will justify our choice for a representation of the active contours by level set functions and will describe the attachment to data and shape constraint models that will be used in this study.

4.2 State of the art - Prior shape constraints in active contours

4.2.1 Active contour framework

Active contours represent segmentation techniques that enable to extract an object of interest from an image. The segmentation is not immediate, it requires a dynamic phase of the contour (hence the designation "active") which will move iteratively during artificial time t , from its initial position towards the boundaries of the object to be extracted. Such an evolution over time can be formalized mathematically in the form of an *evolution equation* explicitly or implicitly expressing the speed of the active contour at each of its points. There are several ways of obtaining an evolution equation. One consists of deriving this equation from the minimization of an energy functional, in which case we speak of a variational approach. An alternative consists of constructing an evolution equation by analogy with other scientific disciplines such as Physics, this is the geometrical approach. We will focus on describing, and then using the variational approach in this thesis. With an approach such as this, the energy functional can be outlined as the sum of two classes of energy terms. The first class concerns the internal energy of the contour aiming to control its intrinsic constraints, such as its regularity for instance. In section 4.2.2 we will see that it is possible to insert more specific geometrical constraints derived from the prior knowledge that we have of the object to be segmented in the image. The second class concerns the *attachment to data term* which will make the active contour interact with the characteristics extracted from the image. In the works of Foulonneau [40] and Jehan-Besson [56], this external energy term is called *criterion* and is constructed from *descriptors*. A descriptor is a measurement taken on the image enabling to characterize a boundary or a region. A boundary descriptor could, for example, be the image's gradient map, a descriptor of a region R could be the mean of the image's pixels contained in R . Depending on the descriptor chosen in the energy functional, we will derive different types of active contours that will be more or less effective according to the nature of the image to be analyzed.

We can therefore see different types of active contours taking shape. They are differentiated by the way with which the evolution equation is deduced, by the active contour's representation mode and, lastly, the attachment to data term which can be based either on the boundaries, the regions or on both. The goal of this section is to detail and explain these different ways of classifying the active contours. Hereinafter, we will restrict ourselves to two-dimensional active contours moving in the plane. We recommend that the reader should consult the works of [13, 110, 87, 108, 56, 102, 40] for an exhaustive view of the state of the art in the area of active contours.

4.2.1.1 Representation of an active contour

Explicit representation

Historically, it was the explicit representation of the active contour that emerged first in the Computer Vision community thanks to the pioneering works of Kass, Witkin and Terzopoulos in 1987 [58]. This type of representation is achieved by parameterizing an

oriented active contour¹ \mathbf{C} with respect to a parameter p and to a time factor t : $\mathbf{C}(p, t)$.

$$\mathbf{C} = \{ \mathbf{C}(p, t) \in \mathbb{R}^2, C^2 \mid p \in [a, b] \subset \mathbb{R}, t \in \mathbb{R}^+ \} \quad (4.1)$$

It is important to note the existence of other ways of representing a contour in an explicit manner (B-Splines, Fourier base, ...). For simplicity, we shall however explain only the method presented in equation (4.1). \mathbf{C} 's evolution equation can then be formalized in a general way, by breaking down the speed of $\mathbf{C}(p, t)$ according to its normal component \mathbf{N} and tangential component \mathbf{T} in (p, t) in a Frénet Frame (\mathbf{T}, \mathbf{N}) :

$$\frac{\partial \mathbf{C}(p, t)}{\partial t} = v_N(p, t) \mathbf{N}(p, t) + v_T(p, t) \mathbf{T}(p, t) \quad (4.2)$$

where $\mathbf{N}(p, t)$ is the normal unitary vector of the contour \mathbf{C} in $\mathbf{C}(p, t)$; $\mathbf{T}(p, t)$ is the unitary tangent vector of the contour \mathbf{C} in $\mathbf{C}(p, t)$. In [108], Sapiro shows that if the normal speed does not depend on the parameterization, then the tangential component of the speed does not influence the deformation of the active contour, but only the parameterization. We can, therefore, without any loss of generality, consider that in the rest of the document, the active contour moves according to its normal:

$$\frac{\partial \mathbf{C}(p, t)}{\partial t} = v_N(p, t) \mathbf{N}(p, t) \quad (4.3)$$

The definition presented in equation (4.1) corresponds to a continuous curve \mathbf{C} in the plane. Since the image is physically and discretely represented by a regular pixel grid, the contour \mathbf{C} must also be made discrete. In practice, making the contour discrete is an under-sampling with regard to the variable p (figure 4.1).

The active contour is therefore made up of a list of nodes whose positions are updated in accordance with the equation (4.3) as \mathbf{C} moves. The discretization of equation (4.3) then becomes:

$$\mathbf{C}(m_i, t + 1) = \mathbf{C}(m_i, t) + \Delta t v_N(m_i, t) \mathbf{N}(m_i, t) \quad (4.4)$$

where m_i is the i^{th} node among N . If we consider that the number of N nodes is fixed over time, we foresee the following problems: if the length of \mathbf{C} increases over time, the polygonal approximation of the object to be segmented will be very coarse: if the length decreases, there is a risk that the nodes will be very close to each other and this will create numerical instability if the finite differences scheme is used to estimate derivatives along \mathbf{C} in order to compute the value of v_N (since the denominator tends to zero). Decreasing the time step Δt in the equation (4.4) allows us to avoid these instabilities but causes an increase of the computational time so that this quickly becomes impractical. There are three ways of solving the problem: i) re-sampling the \mathbf{C} nodes so as to keep the distance constant between the nodes, ii) adding diffusion terms in the expression of speed and thus try to soften the curve of the active contour (the distance between the nodes is thus decreased but does not approach zero), or iii) filtering the oscillations caused by the instability. These non intrinsic solutions are less than optimal and create other problems not associated with the segmentation issue *per se* and thus influence to a large degree the resulting object extraction. Two other problems persist

¹In this document we consider that \mathbf{C} is counter-clockwise oriented.

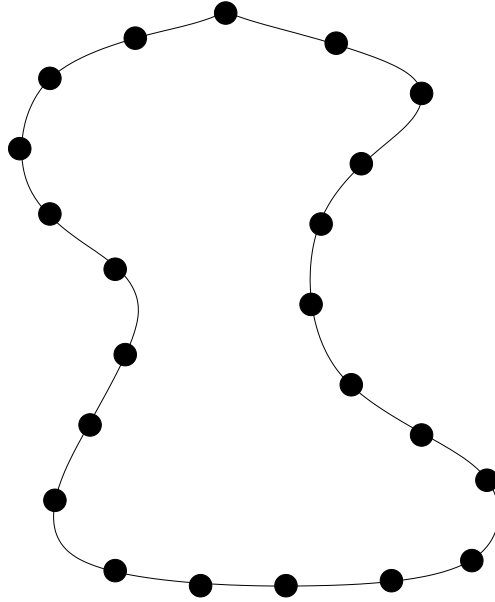


Figure 4.1: Under-sampling of an explicitly represented contour that is parameterized by its arc length.

that are directly related to the explicit representation: the contour \mathbf{C} can not undergo topological changes, and, in addition, cross-artifacts of the contour can appear. The first problem is solved by re-parameterizing the active contour [72]. The second difficulty can be handled by use of "disconnection" "de-looping" techniques [51].

Implicit representation

In 1988, the work of Osher and Sethian [84] presented a less natural way of representing a contour. This is an *implicit* representation taken from the Physics of interfaces, called "level sets". The principle is as follows: a closed contour $\mathbf{C}(t)$ is represented as the zero of a function with $n + 1$ dimensions (in the plane, $n = 2$, see figure 4.2). Originally, it was a function of the signed Euclidean distance that was proposed to represent the $n + 1^{st}$ dimension of the level set. However, other functions can be chosen, for example, Haker *et al.* represent a 3D surface as a function u , solution of Laplace's harmonic equation $\Delta u = 0$ [49].

Within the framework illustrated in [84], ϕ is the level set function verifying the following properties:

- * ϕ is a Lipschitz function with real number values such that $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$
- * whatever the time t , the zero of $\phi(\mathbf{x}, t)$ is the closed contour $\mathbf{C}(t)$:

$$\forall \mathbf{x} \in \mathbb{R}^2, \mathbf{C}(t) = \{ \mathbf{x} \in \mathbb{R}^2 | \phi(\mathbf{x}, t) = 0 \} \quad (4.5)$$

- * $\phi(\mathbf{x})$ is the signed distance between the point \mathbf{x} and the contour $\mathbf{C}(t)$. The sign of $\phi(\mathbf{x})$ depends on \mathbf{x} belonging to the internal or external part of the contour

that we note respectively as $\Omega_{in}(t)$ and $\Omega_{out}(t)$. By convention, we choose $\phi(\mathbf{x})$ positive as the inside of $\mathbf{C}(t)$ and negative as the outside:

$$\begin{cases} \phi(\mathbf{x}, t) = 0 & \text{if } \mathbf{x} \in \mathbf{C}(t) \\ \phi(\mathbf{x}, t) = d(\mathbf{x}, \mathbf{C}(t)) & \text{if } \mathbf{x} \in \Omega_{in}(t) \\ \phi(\mathbf{x}, t) = -d(\mathbf{x}, \mathbf{C}(t)) & \text{if } \mathbf{x} \in \Omega_{out}(t) \end{cases} \quad (4.6)$$

with $d(\mathbf{x}, \mathbf{C}(t))$ as the smallest Euclidean distance from point \mathbf{x} to the contour $\mathbf{C}(t)$:

$$d(\mathbf{x}, \mathbf{C}(t)) = \min_{\mathbf{x}_c \in \mathbf{C}(t)} |\mathbf{x} - \mathbf{x}_c| \quad (4.7)$$

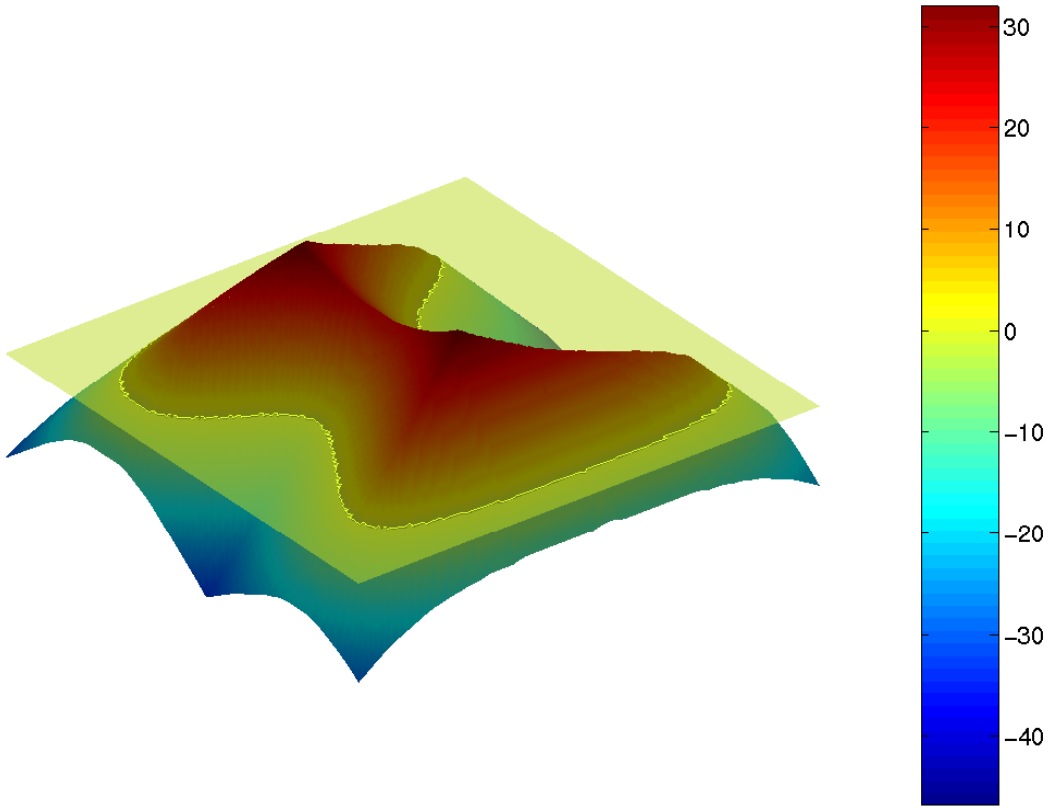


Figure 4.2: Representation in three dimensions of the intersection of a level set with the image plane (zero level). The contour implicitly represented is that of figure 4.1. The caption indicates the signed Euclidean distance of the contour.

Such a representation has certain advantages over explicit representations. First, a contour represented by a level set can change its topology with time, so it is not necessary to have *a priori* knowledge of it. The evolution equation is numerically more stable and avoids having to remap the explicit representation nodes. The problem of the contour loops is resolved naturally by the flexible topology of $\mathbf{C}(t)$. And finally, using level sets, it is possible to obtain several interesting geometric properties that are

intrinsically part of the contour or of any other level line. In this way, the normal can be set at ϕ or at any other point of the image and therefore, all the more so, at any other point of the contour if only one knows the point set such as $\phi(\mathbf{x}, t) = 0$. The inside normal contour is given by the following:

$$\mathbf{N}(\mathbf{x}) = \frac{\nabla \phi(\mathbf{x})}{|\nabla \phi(\mathbf{x})|} \quad (4.8)$$

The curvature κ is also easily derived by:

$$\kappa = \text{div}(\mathbf{N}) = \nabla \cdot \left(\frac{\nabla \phi(\mathbf{x})}{|\nabla \phi(\mathbf{x})|} \right) \quad (4.9)$$

$$= \frac{\phi_{xx}(\mathbf{x}) \phi_y^2(\mathbf{x}) - 2\phi_x(\mathbf{x}) \phi_y(\mathbf{x}) \phi_{xy}(\mathbf{x}) + \phi_{yy}(\mathbf{x}) \phi_x^2(\mathbf{x})}{(\phi_x^2(\mathbf{x}) + \phi_y^2(\mathbf{x}))^{3/2}} \quad (4.10)$$

The level sets also allow logical operations on the sets such as the calculation of the union or of the intersection.

The evolution equation does not address specifically the contour $\mathbf{C}(t)$ as in (4.2), but the level set function. We derive the following evolution equation from $\phi(\mathbf{x}, t)$ by differentiating the equation (4.5) relative to the time t :

$$\forall \mathbf{x}(t) \in \mathbf{C}(t), \quad \frac{\partial \phi(\mathbf{x}(t), t)}{\partial t} = 0 \quad (4.11)$$

It then becomes:

$$\frac{\partial \phi(\mathbf{x}, t)}{\partial t} + \left\langle \nabla \phi(\mathbf{x}, t), \frac{\partial (\mathbf{x}(t))}{\partial t} \right\rangle = 0 \quad (4.12)$$

$$\phi_t(\mathbf{x}, t) + \langle \nabla \phi(\mathbf{x}, t), \mathbf{x}_t(t) \rangle = 0 \quad (4.13)$$

Then, after decomposing the expression $\mathbf{x}_t(t)$ in a similar manner as (4.2) we obtain:

$$\phi_t(\mathbf{x}, t) = -\langle \nabla \phi(\mathbf{x}, t), v_N(\mathbf{x}, t) \mathbf{N}(\mathbf{x}, t) \rangle \quad (4.14)$$

It is important to note that the implicit formulation frees itself from the parameterization p in the active contour $\mathbf{C}(t)$, and this parameterization caused a numerical instability and topological problem in the explicit representation. Making the $\phi(\mathbf{x}, t)$ function discrete can be done through a regular footprint grid Δx and Δy (Eulerian approach), as opposed to the Lagrangian explicit according to p which is more delicate (see figure 4.1). For a given contour $\mathbf{C}(t)$, the associated level set is the calculation of the signed distance of this contour (equation 4.6). For a given level set, if one want to retrieve the contour $\mathbf{C}(t)$, the zeros of $\phi(\mathbf{x}, t)$ have to be calculated (equation (4.5)). The extracted contour is therefore a linear approximation of the true contour represented by $\phi(\mathbf{x}, t)$ (figure 4.5).

Comments:

- * It is important to note that a level set is not merely a calculation artifice that allows us to use different evolution equations for an active contour. A level set is an intrinsic representation different from explicit ones and, as a consequence, results in an alternative numerical solution.

- * The implicit representation presents a loss of spatial information with respect to the representation by nodes (*snakes*). It is possible to know if one is placed on the contour (level zero); on the other hand, it is impossible to decide *where* one is situated. The sequencing relation between the points is lost. The problem of nodes is dual: it allows for localization on the contour since the nodes are organized as an ordered sequence. However the calculation of the distance between any pixel of the image to the contour is not direct.
- * Note that with an explicit representation by nodes, it is possible to model open or closed contours. It is more difficult to process open contours using level sets.
- * And finally, the computational complexity with level sets is greater than with explicit methods. This is inherent to the encoding in the form of a signed distance function of dimension $n + 1$ (n being the dimension of the image), whereas the explicit procedure governs a much smaller number of nodes.

4.2.1.2 Geometric and variational approach

Geometric approach

After having chosen an implicit or explicit presentation model, the normal speed needs to be determined $v_N(\mathbf{x}, t)$ which will allow the evolution of the active contour. The ability of the active contour to segment a particular object represented in the image depends on the calculation of this velocity vector field.

The geometric approach, by analogy with Physics, or by purely mathematical considerations establishes directly the evolution equation of the active contour. In [17] and [68, 69], the authors describe an active contour subject to an Euclidean heat flow or to the average curvature κ (by analogy with the equation for the propagation of heat), $v_N = \kappa$. This flow has diffusing and smoothing properties and prevents the active contour from presenting singularities [110]. This is only an internal regularity constraint of the active contour. To make the contour develop, the authors introduce a constant normal speed, $v_N = \kappa + c$. This constant is analogous to the effect of a swelling or retracting force introduced by Cohen *et al* [27]. The data fit term is a function g weighting the normal speed. In their experiments, the authors use an edge-based term: when the contour reaches the high gradient zones of the image, the weighting function tends to zero. The expression of the normal speed is, finally,

$$v_N = g(|\nabla I|)(\kappa + c) \quad (4.15)$$

The main problem in this approach is the sensitivity of the initialization of the active contour. This is inherent in the data fit term based on the limits as well as on the unequivocal swelling/retracting forces used during contour evolution. The initialization needs a strong a priori knowledge of the localization of the object being extracted. This technique has been shown to be effective in a semi-automatic approach on medical images.

Variational approach

It is possible to formulate a functional J which minimization through the calculation of variations gives an evolution equation for the active contour $\mathbf{C}(t)$; this is the variational approach. The functional is composed of energy terms or *criteria* which are intrinsic (internal constraints on $\mathbf{C}(t)$) or extrinsic (attachment to the data). While the intrinsic criteria are often based on the limit defined by $\mathbf{C}(t)$, the extrinsic criteria are either based on the edges or on the regions. We will use this dichotomy working on the extrinsic criteria to differentiate the edge-based from the region-based active contours. The gradient descent technique is traditionally used to deduce the evolution equation of the contour \mathbf{C} using the calculation of the variations,

$$\frac{\partial \mathbf{C}}{\partial t} = -\frac{\partial J}{\partial \mathbf{C}} \quad (4.16)$$

An active contour evolving in accordance with the equation (4.16) tends to minimize the energy J whose minimum corresponds to the segmentation of the desired objects in the image.

4.2.1.3 Edge-based active contours

An edge-based active contour takes account only of the information present in the vicinity of $\mathbf{C}(t)$. Thus, only the information present at the level of the edges of the object being segmented will be used. In order to formulate the functional, in the first place a *descriptor* of the boundary must be chosen, called $k_b(\mathbf{x})$. This descriptor can be a map of the image gradient module or the field of the gradient vectors in each pixel of the image I . The criterion built from the descriptor is then an integral along the contour $\mathbf{C}(t)$:

$$J_b(\mathbf{C}(t)) = \int_{\mathbf{C}(t)} k_b(\mathbf{x}) d\mathbf{a}(\mathbf{x}) \quad (4.17)$$

where $d\mathbf{a}(\mathbf{x})$ is a length element. We choose to present two types of edge-based active contours, the *snakes* and the *geodesic active contours*.

- Snakes

In [58], Kass, Witkin and Terzopoulos in 1987 represent the active contour in an explicit manner in formulating the functional as the sum of three edge-based terms,

$$J_b(\mathbf{C}(p, t)) = \alpha \int_a^b \left| \frac{\partial \mathbf{C}(p, t)}{\partial p} \right|^2 dp + \beta \int_a^b \left| \frac{\partial^2 \mathbf{C}(p, t)}{\partial p^2} \right|^2 dp - \gamma \int_a^b |\nabla I(\mathbf{C}(p, t))|^2 dp \quad (4.18)$$

with $\{\alpha, \beta, \gamma\} \in \mathbb{R}^+$. The first two terms are intrinsic criteria, a particular case of Tikhonov's operator, developed to the second order and with the weights w_k constant,

$$J_{b,Tikhonov}(\mathbf{C}(p, t)) = \sum_{k=0}^n \int_{\mathbf{C}(t)} w_k(p) \left| \frac{\partial^k \mathbf{C}(p, t)}{\partial p^k} \right|^2$$

The first of the intrinsic criteria can be assimilated with an elastic constraint on the nodes of the discretization of $\mathbf{C}(p, t)$, which causes neighborhood nodes to get closer together. The second of the intrinsic criteria is a rigidity constraint on the active contour, regularizing the curvature (fine sheet model). The last term of the equation (4.18) is the extrinsic criterion of attachment to the data based on the gradient image amplitude. We note that the more $\mathbf{C}(p, t)$ is placed on an elevated image gradient, the weaker is the global functional $J_b(\mathbf{C}(p, t))$. In a variational approach, the second step is to derive the evolution equation of the active contour of which the solution is a minimum of the functional. In [58], the authors use the Euler-Lagrange equations so as to minimize $J_b(\mathbf{C}(p, t))$. In Lagrange's formalism, S is the "action" and is defined by,

$$S = \int \mathcal{L}(t, q_i, \dot{q}_i) dt \quad (4.19)$$

where \mathcal{L} is the Lagrangian, difference between kinetic and potential energy in a mechanical system of n particles. \mathcal{L} depends on the positions and the speeds of the particles as well as the time. Minimizing the "action" (least action principle), is to annul an infinitesimal variation \tilde{S} of S created by the infinitesimal Lagrangian variation $\tilde{\mathcal{L}}$. The minimization of S is successful in resolving the following n differential equations,

$$\frac{\partial \mathcal{L}}{\partial q_i} = \frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i} \right), i \in \{1, \dots, n\} \quad (4.20)$$

Given a closed active contour (cyclic conditions at the limits), the equation (4.20) is successful in resolving the following partial derivatives equation,

$$\alpha \frac{\partial^2 \mathbf{C}(p, t)}{\partial p^2} + \beta \frac{\partial^4 \mathbf{C}(p, t)}{\partial p^4} + \gamma \frac{\partial \int_{\mathbf{C}(t)} |\nabla I(\mathbf{C}(p, t))|^2 dp}{\partial \mathbf{C}(p, t)} = 0 \quad (4.21)$$

In discretizing the active contour in n nodes m_i , the authors of [58] solve the equation (4.21) using the finite differences scheme and obtain finally the following matrix equation,

$$\mathbf{A}\mathbf{V} + \mathbf{f}_v = 0$$

with \mathbf{A} a pentadiagonal Toeplitz matrix,

$$\mathbf{V} = (\mathbf{C}(m_0), \dots, \mathbf{C}(m_n))$$

$$\text{and } \mathbf{f}_v = \left(\frac{\partial \int_{\mathbf{C}(t)} |\nabla I(\mathbf{C}(p, t))|^2 dp}{\partial \mathbf{C}(m_i)} \right)_{i \in \{1, \dots, n\}}.$$

The desired solution is \mathbf{V} , however, \mathbf{A} is not invertible (0 is an eigenvalue of \mathbf{A}). The solution must therefore be sought out in an iterative manner. The method proposed by [58] produces numerical instabilities, in particular because of the calculation of the fourth order derivative in equation (4.21). In addition, the active contour tends to contract in the part of the contour which has the strongest gradient. In [126], the authors propose a rapid algorithm which avoids the difficulties of [58]. The main contribution of this method is to propose an alternative to the estimations by finite differences of the first and second derivatives in the functional, increasing thus the calculation speed and

the precision in the results. The weight coefficients (α, β) , of these derivatives are also variable during the iterations: in accordance with a criterion relative to the curvature (threshold) and to the value of the greatest gradient in the vicinity of the current node, their values can diminish in order to model geometric singularities (angular points). This category of active contours is however limited,

- * The functional expressed in equation (4.18) depends on the parameterization p of the active contour, and this makes it non intrinsic.
- * The result of the segmentation is sensitive to the parameters α , β , and γ , that need to be fixed and adjusted according to the analyzed image. There is no theoretical framework for determining these values, but only empirical choices.
- * Inherently within the manner of representation, the topology of the contour cannot evolve. In addition, it can only segment convex forms.
- * As we noted above, this type of contour is numerically unstable because of its explicit representation.
- * Finally, the initialization of the contour must be very near the object to be segmented in the image which, in general, is of little interest.

- *Geodesic active contours*

In [18], Caselles *et al.* show that minimizing the equation functional (4.18) with $\beta = 0$, is the same as finding a geodesic curve in a Riemann space whose metric depends on the analyzed image. The proposed functional is then,

$$J_b(\mathbf{C}(p, t)) = \alpha \int_a^b \left| \frac{\partial \mathbf{C}(p, t)}{\partial p} \right|^2 dp + \gamma \int_a^b g(|\nabla I(\mathbf{C}(p, t))|) dp \quad (4.22)$$

The functional expressed in equation (4.22) has the same difficulties as that proposed in (4.18): it is non intrinsic and depends on arbitrary parameters α and γ . In accordance with the Maupertuis principle [18], Caselles *et al.* show that minimizing the equation functional (4.22) is the same as minimizing the active contour length L_R (hence the name *geodesic*) in accordance with a metric derived from the image,

$$L_R = \int_0^{L(C(t))} g(|\nabla I(\mathbf{C}(s, t))|) ds \quad (4.23)$$

Compared to the Euclidean length of $\mathbf{C}(t)$ which is by definition $L_E = \oint ds$, we observe that the new definition of the length includes a weighting in $g(|\nabla I(\mathbf{C}(s, t))|)$ which is a term depending on the data. Thus, the more the active contour crosses high image gradient areas, the more the length is decreased. The segmentation problem becomes therefore to find a geodesic contour in which the metrics depend on the image. In minimizing the functional which from here on is intrinsic (4.23) using the calculation of variations, the authors establish the evolution equation of the contour by gradient descent. Representing the contour by a level set, they obtain as follows,

$$\phi_t(\mathbf{x}, t) = g(|\nabla I(\mathbf{C}(p, t))|) \kappa |\nabla \phi(\mathbf{x}, t)| - \langle \nabla g(|\nabla I(\mathbf{C}(p, t))|), \nabla \phi(\mathbf{x}, t) \rangle \quad (4.24)$$

The first term of the equation (4.24) is related to the curvature of the active contour and has a regulatory effect. This term appears naturally and justifies the suppression of the term controlling the rigidity of the contour given in the equation (4.18). This term guarantees the regularity of the weighted curve by the function g , thus stopping the progression of the contour in the zones of high gradient. The second term is the truly new element of the geodesic curve. Indeed the terms based on the curvature have already been introduced with geometric approaches. However, the second term of the equation (4.24) allows us to stabilize the active contour on the edges characterized by a true gradient value. Usually, a contour progression in which the speed is weighted by g is workable if the gradient value is sufficiently high to cause g to tend toward zero. In practice, the value of the gradient is not necessarily very high, thus bringing about a “leak” of the active contour crossing those zones where $|\nabla I|$ is not significant. This type of phenomenon appears also when the value of the gradient is not constant along the edge of an object to be segmented. In this latter case, the second term of the equation allows us to partially alleviate the problem in applying a rigid reverse force, thus stabilizing the contour and avoiding the problems of the purely geometric approach.

The geodesic active contours are an improvement over the “snakes” on several levels. First, the functional of the geodesic contours is intrinsic. These latter take account of the case of a variable gradient value along the edge of the object to be segmented. Secondly, they allow us to segment non convex objects and can easily be transposed to a level set representation. And finally, we can note and underline the contribution of the geodesic contours, which unify the variational and geometric approaches.

The drawbacks of edge-based active contours

Even though they are largely used for image segmentation, the edge-based active contours are subject to a number of problems limiting their application to relatively simple images. The main difficulty in this approach is the purely *local* character of the information used to move the active contour. This makes it necessary to initialize the active contour very near to the object to be segmented. The descriptor that is currently chosen for these types of active contours is the image gradient module. Thus, if all or a part of the active contour is found in the zones of homogeneous luminance, not having a high gradient, the contour will not be moved by any external force. Finally, the other problem caused by the local character of the edge-based contours is the sensitivity to image noise. Since the edge descriptors are calculated locally, the neighborhoods considered are relatively small compared to the size of the image and are therefore influenced by the noise to a large degree. Pre-filtering of the image is a solution that is always available to diminish the problem. However, the filtering has the effect of delocalizing the edge of the object thus making the segmentation imprecise.

To partially alleviate the problem caused by the sensitivity to the initialization of the edge-based active contours, Cohen [27] proposes to add an artificial swelling or retracting force aimed at bringing the active contour to the edges of the object to be segmented. This is the same as adding a constant c into the evolution equation of the

geometric or geodesic contours in the following manner,

$$\phi_t(\mathbf{x}, t) = [\kappa + c] g(|\nabla I(\mathbf{C}(p, t))|) |\nabla \phi(\mathbf{x}, t)| - \langle \nabla g(|\nabla I(\mathbf{C}(p, t))|), \nabla \phi(\mathbf{x}, t) \rangle \quad (4.25)$$

As a function of the sign of c , the active contour extends outward or contracts according to its normal. The choice of the sign is therefore critical since it requires a priori knowledge regarding the localization of the object to be extracted in relation to the initial state of the active contour $\mathbf{C}(t = t_0)$. Cases where the active contour is partially inside and partially outside the object to be segmented is equally difficult.

- *Diffusion of the gradient information using the Gradient Vector Flow (GVF)*

In order to make the image gradient information not local, Xu and Prince [128] propose to diffuse the image gradient by using a vector field of components (\mathbf{u}, \mathbf{v}) . Given f , the normalized gradient module of the image I , the diffusion of the gradient information is carried out by the minimization of the functional:

$$J_{GVF}(u, v) = \int_{\Omega} \mu_{GVF} (\mathbf{u}_x^2 + \mathbf{u}_y^2 + \mathbf{v}_x^2 + \mathbf{v}_y^2) + |\nabla f|^2 |(\mathbf{u}, \mathbf{v}) - \nabla f|^2 d\mathbf{x} \quad (4.26)$$

where μ_{GVF} is a positive constant.

The first term of the equation (4.26) depends on the appearance of the field (\mathbf{u}, \mathbf{v}) . For the weak gradient areas, this term is predominant and regularizes (\mathbf{u}, \mathbf{v}) . On the contrary, for strong gradient areas, the second term connected to the data causes (\mathbf{u}, \mathbf{v}) to resemble the image gradient ∇f . In [128], the authors propose to deduce the GVF (\mathbf{u}, \mathbf{v}) from the gradient descent minimizing J_{GVF} :

$$\begin{cases} \mathbf{u}_t = \mu_{GVF} \Delta \mathbf{u} - (\mathbf{u} - f_x) |\nabla f|^2 \\ \mathbf{v}_t = \mu_{GVF} \Delta \mathbf{v} - (\mathbf{v} - f_y) |\nabla f|^2 \end{cases} \quad (4.27)$$

The GVF thus obtained can also be incorporated with the active contours represented explicitly [128] or implicitly [91]. In [91], N. Paragios uses this diffused vector field with active contours represented by level sets. Here is one of the models presented in [91]:

$$\phi_t(\mathbf{x}, t) = g(|\nabla I(\mathbf{C}(p, t))|) (\kappa |\nabla \phi(\mathbf{x}, t)| - \langle (\mathbf{u}, \mathbf{v}), \nabla \phi(\mathbf{x}, t) \rangle) \quad (4.28)$$

The first term of the equation (4.28) regularizes the curve and the second lines up the normal of the active contour with the GVF and tends to direct \mathbf{C} towards the areas of high gradient. This *gradient vector flow* technique allows us to make the edge information less local and at the same time allows the active contour to segment non convex objects.

4.2.1.4 Region-based active contours

An effective alternative to edge-based active contours is the use of region-based active contours which use intrinsically the *global* information of the image. The descriptors of a region R are often statistical quantities such as the average, the variance, the texture or the histogram of the region concerned. The energy functional constructed from such

descriptors $k(\mathbf{x}, R)$ is then the double integral of the region $R(t)$, defined by $\mathbf{C}(t)$. Optionally, it is possible to add an edge-based descriptor $k_b(\mathbf{x})$ either for introducing a regularizing term or for inserting a data fit term that is complementary to terms derived from region-based information,

$$J_r(R(t)) = \int_{R(t)} k(\mathbf{x}, R(t)) d\mathbf{x} + \int_{\mathbf{C}(t)} k_b(\mathbf{x}) d\mathbf{a} \quad (4.29)$$

Region-based active contours are part of a very active research area ever since the 1990s in Computer Vision. In the beginning, the work of Zhu *et al.* [137] presented a method of region competition within the mixed framework which was at the same time Bayesian and tended to minimize the *Minimal Length Description* criterion (MLD). The authors consider the image pixels to be like the realization of a random variable in which the density of Gaussian probability has parameters which differ depending on its belonging or not belonging to a region defined by the active contour. Based on this work and on that of [18], N. Paragios (using an approach completely Bayesian) introduces active *geodesic regions* which integrate the edge-based term of the geodesic contours for the supervised segmentation of scalar images, and moving objects tracking from the video sequences [87, 90]. The work of M. Rousson [103, 102] continues that of N. Paragios, extending them to color images and to the segmentation of textured zones which no longer respond to Gaussian criteria.

At the same time, the work of D. Mumford and J. Shah, carried out in 1985, brought about the opening up of another more geometric branch in the area of region-based segmentation [74]. This work has recently become popular in the Computer Vision community thanks to T. Chan and L. Vese [19]. Rousson shows in [102] that the minimal partition limit in the work of Mumford and Shah is a particular case of the Bayesian approach. There is a plethora of recent works on region-based segmentation [19, 20, 21, 25, 40, 74, 88, 87, 92, 90, 103, 102, 117, 137]. They are related to the supervised and unsupervised segmentation or classification of N image partitions, to the use of texture information, and to the analysis of sequences of images. In the framework of the present work, we limit ourselves to the bimodal segmentation of a scalar image. Even if the number of partitions is known a priori ($N = 2$), their statistical properties in the image are not known (the opposite of supervised approaches). We present the two main region-based segmentation models that will be used for our application: the Mumford-Shah model (and its derived forms) as well as the model based on the Bayesian approach.

The Mumford-Shah functional

In 1985, D. Mumford and J. Shah proposed, in [74], a segmentation method applicable for any type of signal. In the framework of image processing, the idea is as follows: for a given image I , an ideal image is sought out, namely U which is piecewise homogeneous and which is an approximation of I . The image U is a collection of homogeneous regions in which the intersections form a set B made up of regular edges. The authors formalized this idea in the form of the following functional:

$$J_{MS} = \mu^2 \int \int_R (U - I)^2 + \mu \int_{R-B} |\nabla U|^2 + \nu L_B \quad (4.30)$$

where $(\mu, \nu) \in \mathbb{R}^+$ and L_B is the length of the boundaries making up the set B . The first term of the functional causes U to resemble the image I . The second term forces U to be homogeneous within each region (the inter-regional discontinuities are preserved). And finally, the last term causes the boundaries partitioning the image U to have a minimal length L_B . The minimization of this functional is therefore the same as finding the set of the boundaries of B segmenting the image I into homogeneous parts. This segmentation technique also allows us to eliminate image noise at the same time.

More recently, Chan and Vese applied this segmentation technique to the limiting case of two regions which is the minimal partition limit [19]. The authors reduce U to the average of I inside and outside region R . This simplification, called *cartoon limit*, or more poetically, *Mondrian Universe* [40], constrains U to be piecewise constant (and not homogeneous). The expression of the functional in the form of level sets is as follows:

$$J_{CV}(\phi, t) = \int_{\Omega} (I(\mathbf{x}) - c_{in}(\phi(\mathbf{x}, t)))^2 H_a(\phi(\mathbf{x}, t)) d\mathbf{x} \\ + \int_{\Omega} (I(\mathbf{x}) - c_{out}(\phi(\mathbf{x}, t)))^2 (1 - H_a(\phi(\mathbf{x}, t))) d\mathbf{x} + \nu \int_{\Omega} |\nabla H_a(\phi(\mathbf{x}, t))| d\mathbf{x} \quad (4.31)$$

The quantities c_{in} and c_{out} being by definition the average of the image pixels I inside and outside the active contour in time t :

$$c_{in}(\phi(\mathbf{x}, t)) = \frac{\int_{\Omega} I(\mathbf{x}) H_a(\phi(\mathbf{x}, t)) d\mathbf{x}}{\int_{\Omega} H_a(\phi(\mathbf{x}, t)) d\mathbf{x}} \quad (4.32)$$

$$c_{out}(\phi(\mathbf{x}, t)) = \frac{\int_{\Omega} I(\mathbf{x}) (1 - H_a(\phi(\mathbf{x}, t))) d\mathbf{x}}{\int_{\Omega} (1 - H_a(\phi(\mathbf{x}, t))) d\mathbf{x}} \quad (4.33)$$

where H_a is a regular approximation of the Heaviside function. The minimization of the functional of the equation (4.31) by the method of gradient descent produces the following evolution equation,

$$\phi(\mathbf{x}, t)_t = - \left\{ (I(\mathbf{x}) - c_{in}(\phi(\mathbf{x}, t)))^2 - (I(\mathbf{x}) - c_{out}(\phi(\mathbf{x}, t)))^2 + \kappa\nu \right\} \delta_a(\phi(\mathbf{x}, t)) \quad (4.34)$$

where δ_a is a regular approximation of Dirac's distribution. This approach allows for a segmentation less sensitive to the initialization of the active contour compared to edge-based methods. It is also an easily used method that is relatively inexpensive in terms of calculations. However, this model restricts segmentation to homogeneous objects. It is, nonetheless, possible to move away from the hypothesis of strict homogeneity if the object in question is sufficiently set off from the image background. That is the case for certain results of [19] illustrating the segmentation of galaxies (star constellations) that are not homogeneous but which stand out singularly from the interstellar space. Chan and Vese then introduced in [20] a less drastic simplification of the image U in allowing it to be piecewise homogeneous. The values of U depend then on the pixel position \mathbf{x} and on its positioning on the inside or the outside of the region set apart by the active contour:

$$U(\mathbf{x}) = \begin{cases} U_{in}(\mathbf{x}), & \text{if } \mathbf{x} \in \Omega_{in}(t) \\ U_{out}(\mathbf{x}), & \text{if } \mathbf{x} \in \Omega_{out}(t) \end{cases}$$

The Mumford-Shah functional then becomes,

$$\begin{aligned}
 J_{CV}(\phi, t) = & \int_{\Omega} (I(\mathbf{x}) - U_{in}(\mathbf{x}))^2 H_a(\phi(\mathbf{x}, t)) d\mathbf{x} \\
 & + \int_{\Omega} (I(\mathbf{x}) - U_{out}(\mathbf{x}))^2 (1 - H_a(\phi(\mathbf{x}, t))) d\mathbf{x} \\
 & + \mu \int_{\Omega} |\nabla U_{in}(\mathbf{x})| H_a(\phi(\mathbf{x}, t)) d\mathbf{x} \\
 & + \mu \int_{\Omega} |\nabla U_{out}(\mathbf{x})| (1 - H_a(\phi(\mathbf{x}, t))) d\mathbf{x} \\
 & + \nu \int_{\Omega} |\nabla H_a(\phi(\mathbf{x}, t))| d\mathbf{x}
 \end{aligned} \tag{4.35}$$

Deriving the equation (4.35) respectively relative to $U_{in}(\mathbf{x})$, $U_{out}(\mathbf{x})$ and $\phi(\mathbf{x})$, the authors deduce the evolution equation of the ideal image U and of the active contour implicitly represented by ϕ . Using a different approach, Tsai *et al* estimate $U_{in}(\mathbf{x})$ and $U_{out}(\mathbf{x})$ using a stochastic theory of optimization [117].

Bayesian approach

We recall the formulation of the *maximum a posteriori* (MAP) method given in [87, 90, 102] and its link with previous works based on a Bayesian approach.

Given $p(\mathcal{P}(\Omega)|I)$ the probability of getting a partition $\mathcal{P}(\Omega)$ of a given image I , an optimal segmentation of the image is obtained when this probability is maximized. In accordance with Bayes formula, this probability is expressed in the form,

$$p(\mathcal{P}(\Omega)|I) = \frac{p(I|\mathcal{P}(\Omega))}{p(I)} p(\mathcal{P}(\Omega)) \tag{4.36}$$

with $p(I)$ being the probability of observing a realization I of the image. $p(\mathcal{P}(\Omega))$ is the probability of obtaining an image partition $\mathcal{P}(\Omega)$ among all the possible partitions. $p(I|\mathcal{P}(\Omega))$ is the probability of obtaining an image I for a partition $\mathcal{P}(\Omega)$ which is known *a priori*. The hypothesis of equiprobability is commonly retained for the observation probability of a realization I of the image. Thus, the equation (4.36) becomes,

$$p(\mathcal{P}(\Omega)|I) \sim p(I|\mathcal{P}(\Omega)) p(\mathcal{P}(\Omega)) \tag{4.37}$$

In [104, 102], Rousson injects into $p(\mathcal{P}(\Omega))$ an a priori knowledge in the form of optimal partitions that can be obtained. In another way, this term allows the active contour to be regularized if it depends on the length $L(\mathbf{C})$ of the contour,

$$p(\mathcal{P}(\Omega)) = \frac{\nu}{2} \exp^{-\nu L(\mathbf{C})} \tag{4.38}$$

with ν being a constant belonging to \mathbb{R}^+ . Maximizing the probability above is the same as finding a partition of the minimal length. This term is the same as the one that was introduced into the Mumford-Shah functional in equation (4.30). In [87, 90], N.

Paragios considers $p(\mathcal{P}(\Omega))$ constant (equiprobability). He adds to the region-based functional derived from MAP the edge-based term from the geodesic contours which has a diffusing and regularizing term.

The calculation of $p(I|\mathcal{P}(\Omega))$ requires other hypotheses in order to be successfully carried out:

* *The regions of the optimal partition are not correlated:*

$$p(I|\mathcal{P}(\Omega)) = p(I|\mathcal{P}(\Omega_{in}, \Omega_{out})) = p(I|\mathcal{P}(\Omega_{in}))p(I|\mathcal{P}(\Omega_{out})) \quad (4.39)$$

This hypothesis is reasonable since the aim of the segmentation is to separate out the regions of the image where the properties are different.

* On the other hand, the following hypothesis is stronger and more restrictive: *the pixels of a single region have the same probability to be observed*. This hypothesis is shown to be in error in textured regions or in those with repeated patterns where there is a local interaction between the pixels. The equation (4.37) then becomes,

$$p(I|\mathcal{P}(\Omega)) = \prod_{\mathbf{x} \in \Omega_{in}} p(I(\mathbf{x})|\mathcal{P}(\Omega_{in})) \prod_{\mathbf{x} \in \Omega_{out}} p(I(\mathbf{x})|\mathcal{P}(\Omega_{out})) \quad (4.40)$$

The maximization can be reformulated as an expression of energy minimization as follows,

$$J(\Omega_{in}, \Omega_{out}) = -\ln\{p(\mathcal{P}(\Omega)|I)\} \quad (4.41)$$

In replacing the expressions of equations (4.38) and (4.40) in the formulation of the equation (4.41), we arrive at the following energy functional,

$$J(\Omega_{in}, \Omega_{out}) \sim - \int_{\Omega_{in}} \ln(p(I(\mathbf{x})|\mathcal{P}(\Omega_{in}))) - \int_{\Omega_{out}} \ln(p(I(\mathbf{x})|\mathcal{P}(\Omega_{out}))) + \nu L(\mathbf{C}) \quad (4.42)$$

In [87, 90], [102] and [137], the authors presuppose that the probability densities follow normal distributions in which the only parameters are the average c and the variance σ calculated on Ω :

$$p(I(\mathbf{x})|c, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{-\frac{(I(\mathbf{x})-c)^2}{2\sigma^2}} \quad (4.43)$$

In [87, 90], the Gaussian distribution parameters are known in advance through a supervised process. In [102] and [137], these parameters are dynamically calculated during the evolution of the active contour. In [137], the expression of the functional does not come from MAP but from the method referred to as MDL (Minimum Description Length).

With the hypothesis of a Gaussian distribution, and in representing the active contour by a level set, the energy functional becomes,

$$J_{Bayes}(\phi, t) = \int_{\Omega} \left(\frac{(I(\mathbf{x}) - c_{in}(\phi(\mathbf{x}, t)))^2}{2\sigma_{in}^2} + \ln\left(\sqrt{2\pi\sigma_{in}^2}\right) \right) H_a(\phi(\mathbf{x}, t)) d\mathbf{x}$$

$$+ \int_{\Omega} \left(\frac{(I(\mathbf{x}) - c_{out}(\phi(\mathbf{x}, t)))^2}{2\sigma_{out}^2} + \ln \left(\sqrt{2\pi\sigma_{out}^2} \right) \right) (1 - H_a(\phi(\mathbf{x}, t))) d\mathbf{x} + \nu L(\mathbf{C}) \quad (4.44)$$

in which the variances on the inside and the outside of the active contour are, by definition,

$$\sigma_{in}^2(\phi(\mathbf{x}, t)) = \frac{\int_{\Omega} (I(\mathbf{x}) - c_{in}(\phi(\mathbf{x}, t)))^2 H_a(\phi(\mathbf{x}, t)) d\mathbf{x}}{\int_{\Omega} H_a(\phi(\mathbf{x}, t)) d\mathbf{x}} \quad (4.45)$$

$$\sigma_{out}^2(\phi(\mathbf{x}, t)) = \frac{\int_{\Omega} (I(\mathbf{x}) - c_{out}(\phi(\mathbf{x}, t)))^2 (1 - H_a(\phi(\mathbf{x}, t))) d\mathbf{x}}{\int_{\Omega} (1 - H_a(\phi(\mathbf{x}, t))) d\mathbf{x}} \quad (4.46)$$

The evolution equation of the active contour derived from the gradient descent of the functional (4.42) is then,

$$\begin{aligned} \phi(\mathbf{x}, t)_t = & \left\{ -\frac{(I(\mathbf{x}) - c_{in}(\phi(\mathbf{x}, t)))^2}{2\sigma_{in}^2(t)} + \frac{(I(\mathbf{x}) - c_{out}(\phi(\mathbf{x}, t)))^2}{2\sigma_{out}^2(t)} \right. \\ & \left. + \ln \left(\frac{\sigma_{out}^2(t)}{\sigma_{in}^2(t)} \right) + \kappa\nu \right\} \delta_a(\phi(\mathbf{x}, t)) \end{aligned} \quad (4.47)$$

The Gaussian distribution parameters are evaluated at each iteration by the maximum likelihood method, which is the simple estimation of the average and variance in this case. In [137], Zhu *et al* arrive at a functional that is similar to the equation (4.47) without the regularizing term. The procedure based on the MDL dissimulates however the clearly announced hypotheses of the Bayesian framework in order to arrive at this result. We note that for $\sigma_{in}^2(t) = \sigma_{out}^2(t) = cst$, we return to the Mumford-Shah model expressed as equation (4.34) by Chan and Vese.

4.2.2 Segmentation with specific shape constraint

The models presented in the preceding section, whether they are edge-based or region-based models, are functions of the pixel intensity in the image. Thus, if the object to be segmented is altered by noise, occlusions or by a weak contrast at the boundaries, the result will be strongly influenced by those artifacts that degrade the quality of the extraction. This is particularly critical in the case of medical images (ultrasound images for example), which have a weak signal to noise ratio and present organs in which certain parts of the contour are extremely weak or even hidden. Because of this, the Computer Vision community in Medical Imagery was the first and the most active to resolve this problem by inserting prior shape knowledge into the segmentation process.

As regards an object to be segmented, the a priori knowledge is of many types. It can be related to the radiometry of the object (color, luminance), its textural properties or its shape. In this chapter, we focus solely on the a priori knowledge of the shape. That allows us to constrain the active contour so it resembles a shape (or family of shapes) of the reference and to regularize specifically on those parts of the active contour in the image where the artifacts are present.

Since the beginning of the 90s, many papers have shown how the input of a specific shape prior could improve the segmentation results [33, 29, 31, 32, 30, 41, 40, 66, 21, 24, 23, 25, 22, 92, 104, 102, 113, 131, 130]. We shall examine briefly the most important works about the insertion of prior shape knowledge and differentiate between statistical approaches (which use a set of reference shapes) and those that are deterministic (single shape prior). In general, the shape constraint is introduced by a distance that allow us to compare the active contour in time t with the prior shape. Within the framework of variational approaches, this distance is used for the formulation of a shape constraint energy, J_{shape} which is then added to the term respecting the connection with the data,

$$J = J_{image} + \lambda J_{shape} \quad (4.48)$$

where J is the global energy functional, J_{image} is the energy of connection to the data and $\lambda \in \mathbb{R}^+$. We propose to describe these distances as a function of the way in which the active contour is represented (implicitly or explicitly). This dichotomy is not perfect since certain approaches can be formulated with one or the other of these representation systems. We will identify these approaches when necessary.

4.2.2.1 Non statistical approach

Explicit representation

In [41] A. Foulonneau proposes to use geometric moments to characterize both the active contour and the reference shape. A distance based on these moments is then created to measure the distance between the active contour and the a priori shape. In order to guarantee a more compact and less redundant representation, the geometric moments are projected on an orthogonal base of Legendre polynomials. The parameters of this explicit representation have the advantage of being intrinsically invariant from translation and scaling. A detailed analysis of the invariance extension to rotation, or more generally to affine transform, is available in [40]. As is the case in all representations using a base (Legendre or Fourier), the order truncation of the base is a delicate decision. In order to represent shapes having singularities (corners), a high order must be used in order to obtain an approximation of quality. This is a problem that increases the calculation complexity and restricts the representation possibilities of such explicit methods.

Implicit representation

Y. Chen was the first to introduce a shape constraint in the form of a level set [24, 23]. However, we note that this formulation can also be presented explicitly. The author proposes an energy constraint term invariant from similarity transform, T_{sim} which, within the variational approach framework, takes the following shape,

$$J_{shape}(\mathbf{C}, T_{sim}) = \int_0^1 d^2(T_{sim} \mathbf{C}(p)) |\mathbf{C}_p(p)| dp \quad (4.49)$$

in which $d(\mathbf{x}) = d(\mathbf{C}^*, \mathbf{x})$ is the distance between the coordinate point \mathbf{x} and the reference shape \mathbf{C}^* . The term is weighted by a constant λ and added to the data fit term of

geodesic active contours. The results obtained from this approach show a marked improvement for medical image segmentation. However, this scheme is still very sensitive to the initialization because of the edge-based term connected to the data. Invariance by similarity requires an optimization of the parameters of T_{sim} during the active contour evolution. The author estimates their values by gradient descent. Adjusting the weight λ is not a trivial matter. If the weighting is too weak, a poor segmentation of the object risks being obtained and a too strong weight will prevent the active contour from respecting the information derived from the image.

In [32], D. Cremers proposes the following term for the energy constraint,

$$J_{shape}(\phi, \psi_0) = \int_{\Omega} (\phi(\mathbf{x}) - \psi_0(\mathbf{x}))^2 d\mathbf{x} \quad (4.50)$$

This quadratic term is the difference between the level set ϕ representing the active contour and the one encoding the reference shape ψ_0 . This formulation has the advantage of preserving the intrinsic characteristics of the flexible topology created by the level sets. On the other hand, this energy depends on the size of the integration domain and is not invariant from any transformation. D. Cremers notes that such a distance only allows us to extract the object known a priori in the image, operating like a filtering system in the shape space of the active contour. He then proposes a new energy capable of locally applying the shape constraint on the image,

$$\begin{aligned} J_{shape}(\phi, \psi_0, L) = & \int_{\Omega} (\phi(\mathbf{x}) - \psi_0(\mathbf{x}))^2 (L(\mathbf{x}) + 1)^2 d\mathbf{x} + \int_{\Omega} \lambda^2 (L(\mathbf{x}) - 1)^2 d\mathbf{x} \\ & + \gamma \int_{\Omega} |\nabla H(L(\mathbf{x}))| d\mathbf{x} \end{aligned} \quad (4.51)$$

The function L called “dynamic labelling”, only takes the values of $+1$ or -1 . The function L reinforces the shape constraint as it tends towards 1 and annuls it as it converges towards -1 . This last behavior is favored when the active contour is more distant from the prior shape (static in the present case). The last term of the equation (4.51) assures the regularity of the function L . This energy allows us to segment a corrupted object thanks to the shape prior as well as the surrounding objects having the same radiometry without knowing in advance the place where the constraint has to be applied.

In [104] the authors propose the following energy constraint,

$$J_{shape}(\phi, \psi_0) = \int_{\Omega} (\phi(\mathbf{x}) - \psi_0(T_{sim}\mathbf{x}))^2 H(\phi(\mathbf{x})) d\mathbf{x} \quad (4.52)$$

This formulation has the advantage of no longer depending on the integration space Ω since the integral is now applied to the inside of the active contour. It is also invariant from direct plane similarity as in the equation (4.49). But as D. Cremers shows in [31], this quadratic distance is not symmetric and is not capable of taking account of objects with several components. He therefore proposes a pseudo-distance verifying the positiveness and symmetrical properties without, however, satisfying the triangular inequality,

$$J_{shape}(\phi, \psi_0) = \int_{\Omega} (\phi(\mathbf{x}) - \psi_0(\mathbf{x}))^2 \frac{h(\phi(\mathbf{x})) + h(\psi_0(T_{sim}\mathbf{x}))}{2} d\mathbf{x} \quad (4.53)$$

in which $h(\phi(\mathbf{x})) = \frac{H(\phi(\mathbf{x}))}{\int_{\Omega} H(\phi(\mathbf{x}))d\mathbf{x}}$. This energy allows us also, in this way, to successfully segment a multi-component object partially erased or noised over.

In [21], is formulated another symmetrical constraint energy that is independent from the integration domain,

$$J_{shape}(\phi, \psi_0) = \int_{\Omega} (H(\phi(\mathbf{x})) - H(\psi_0(T_{sim}\mathbf{x})))^2 d\mathbf{x} \quad (4.54)$$

This term compares the inside areas of the active contour to the reference shape with respect to the standard norm L_2 .

4.2.2.2 Statistical approach

The object to be segmented in the image may present a certain variability when compared to the reference shape. This is the case, for example, for medical images in which the organ shape varies from one individual to another, depending on the means of acquisition, the orientation of the camera or the time in the case of video sequences (cardiac echography). Since these variations compared to the reference cannot be modeled by an affine linear transformation, many works have focused their efforts on the learning of diverse reference shapes so as to confer a maximum degree of flexibility to the shape constraint. Thus the active contour is permitted to be deformed in the sub-space of the shapes defined by the learning. Since the learned forms are redundant, they are often projected into an orthogonal sub-space by a principal component analysis (PCA) in order to establish the main variability modes. A problem inherent in the processing of the shapes for the learning process, is their optimal alignment: they have to be placed in the same reference system so they can be compared without bias. The approach frequently adopted is to estimate the greatest direct plane similarity permitting optimal alignment (the Procrustes method). When certain shapes are too different, a clustering method is used to classify them in sub-families [24].

Explicit representation

Pioneering work on the insertion of statistical shape constraint is attributed to Staib and Duncan [113]. The authors propose representing explicitly the active contour using a Fourier base. The learning samples allow an evaluation of the density of probability for each parameter of the explicit representation (which are the coefficients of each element of the base). Assuming that the prior shape probability is Gaussian, they formulate their segmentation criterion by MAP. The optimal segmentation is thus obtained when the active contour satisfies the dependence on the data with a maximum degree of likelihood with the learned samples.

A similar approach is used by Cootes *et al* in [28]. However they use a different explicit way to represent the contour in using snakes. These pick up “control points” at chosen points of the learning shapes. The process is manual or assisted and represents a strong constraint, i) the choice of control points can affect the final result, ii) it is a slow stage which restricts the flow of data to be processed. We will see that the problem was

resolved in [33, 104]. After the learning shapes are lined up, they are projected onto an orthogonal sub-base by PCA. There is no energetic criterion allowing a comparison of the active contour with the reference shapes: at each iteration, the spatial variation of each of the nodes attracted by the high gradient zones is projected on the shape subspace and this constrains the displacement, and in the end, the active contour shape. This work is translated into a Bayesian framework and generalized to invariance from affine transformation in [131].

In [33], D. Cremers introduces the *Diffusion Snakes* based on a representation by B-splines. The advantage of the B-splines base is a greater compactness of the representation of the learning shapes as well as their local characteristics (unlike a Fourier base, a control point variation modifies the curve locally). In addition, this approach is more general as it can model singular shapes with corners. A statistical analysis by PCA allows to determine the main variability modes. The shape constraint is formulated as the Mahalanobis distance between the active contour and the average reference shape. This term has the advantage to be intrinsically invariant from direct plane similarity.

Implicit representation

Leventon *et al* propose in [66] to represent the learning shapes in the form of level sets. As in the methods presented, the PCA is used to derive the principal deformation modes. The shape constraint is also derived from a MAP type approach. However, it appears in the evolution equation as a correcting term equal to the difference between the level set of the active contour and that of the most probable shape (linear combination of the eigenmodes). This rather unnatural way of introducing the shape constraint is formalized more generally in the energy functional of a variational approach in [104, 102]. In these works, the learning shapes are once again represented by their level sets. These are aligned in the first instance by an estimation of the optimal direct plane similarity using the equation (4.50). The average shape is represented as the average of the level sets, $\phi_{\mathcal{M}}(\mathbf{x}, t)$, the variance $\sigma_{\mathcal{M}}(\mathbf{x})$ is calculated for each pixel from the learning shapes. The shape constraint energy, which is incorporated in a MAP process, is then:

$$J_{Shape}(\phi, T) = \int_{\Omega} \left(\frac{(s\phi(\mathbf{x}, t) - \phi_{\mathcal{M}}(T\mathbf{x}, t))^2}{s\sigma_{\mathcal{M}}(T\mathbf{x})} + \ln(\sigma_{\mathcal{M}}(T\mathbf{x})) \right) \delta_a(\phi(\mathbf{x}, t)) d\mathbf{x} \quad (4.55)$$

The first term of the energy represents the shape constraint that is normalized using the average variance for measuring the confidence of the model. Then the strongly repetitive zones among the learning samples (with weak variance) will have a shape constraint that is particularly reinforced.

4.2.2.3 Shape constraint invariance from geometrical transformation

As we have seen, invariance of the shape constraint with respect to certain geometrical transformations is crucial. Indeed, without this invariance, the segmentation power would be drastically reduced to objects from the image corresponding exactly to the reference. The invariance from affine transformation is a recurring problem in Computer Vision. Some representations give a natural invariance to a sub-class of affine

transforms, for example direct plane similarities [33, 113]. Others require the addition of extrinsic parameters that allow the realization of invariance [104, 21, 23, 24].

In the case of an energy criterion for the shape constraint represented by a level set (equation (4.54)), it is possible to calculate analytically the level set ϕ_2 of a contour represented by ϕ_1 and transformed by direct plane similarity T_{sim} :

$$\phi_2(\mathbf{x}) = s\phi_1(T_{sim}^{-1}\mathbf{x}) \quad (4.56)$$

where s is the scale factor of T_{sim} . The equation (4.56) can be expressed in the form,

$$\phi_2(x, y) = s\phi_1\left(\frac{(x - \mu_x)\cos(\theta) + (y - \mu_y)\sin(\theta)}{s}, \frac{-(x - \mu_x)\sin(\theta) + (y - \mu_y)\cos(\theta)}{s}\right) \quad (4.57)$$

where θ is the angle of rotation, and (μ_x, μ_y) the translation of T_{sim} . The extension of the analytic relation between level sets for a transformation with an anisotropic scale factor or with an affine transformation is more delicate and is still an open question. In [30], D. Cremers proposes a formulation of the intrinsic constraint with no parallel estimate of the pose parameters. However, it is limited to translation and scaling invariance. And finally, the interesting work of Riklin-Raviv *et al.* [98] allows us to model a perspective type projection in intersecting the signed distance function with an inclined plane (and not a horizontal plan of the set 0).

4.2.3 Segmentation with a generic shape constraint: quadratic models

A new class of higher order active contours was proposed by M. Rochery and I. Jermyn in [101]. The finality of these recent models is to incorporate global geometric constraints within active contours. If this constraint is more specific than the generic constraints for contour regularization, it is less specific than for the incorporation of prior shapes described in the preceding paragraphs.

The proposed active contours are based on a variational approach and minimize the quadratic energy defined below in the space of one-chains,

$$J_{quad}(C) = \oint \oint \left\langle \vec{\mathbf{t}}(p), F(\mathbf{C}(p), \mathbf{C}(p')) \vec{\mathbf{t}}(p') \right\rangle dpdp' \quad (4.58)$$

where $\vec{\mathbf{t}}(p), \vec{\mathbf{t}}(p')$ are the tangent vectors of the active contour \mathbf{C} at points p and p' . F is the covariant tensor weighting the interaction between the two points of the contour $\mathbf{C}(p)$ and $\mathbf{C}(p')$. The derivation of J_{quad} with respect to the curve \mathbf{C} allows us to deduce a non local shape constraint force for each point $\mathbf{C}(p)$ which is expressed as a simple integral on the contour.

In [99], the authors use quadratic models to extract roads from remote sensing images. They are interested in the particular case in which the tensor F is the identity matrix and formulate an energy designed to improve the segmentation of linear structures with parallel edges (roads). The quadratic energy thus proposed is,

$$J_{quad}(C) = - \oint \oint \left\langle \vec{\mathbf{t}}(p), \vec{\mathbf{t}}(p') \right\rangle \Psi(|\mathbf{C}(p) - \mathbf{C}(p')|) dpdp' \quad (4.59)$$

in which Ψ is a decreasing function representing the interaction potential between the contour points on the abscissa p and p' :

$$\Psi(x) = \begin{cases} 1 & \text{if } x < d_{min} - \varepsilon \\ 0 & \text{if } x > d_{min} + \varepsilon \\ \frac{1}{2} \left(1 - \frac{x - d_{min}}{\varepsilon} - \frac{1}{\pi} \sin \left(\pi \frac{x - d_{min}}{\varepsilon} \right) \right) & \text{otherwise} \end{cases} \quad (4.60)$$

Note that the energy formulation in (4.59) is minimal if two neighboring points of the contour ($|\mathbf{C}(p) - \mathbf{C}(p')| < d_{min}$) have co-linear tangents. This aids in the extraction of rectilinear structures. On the other hand, this energy is high for two neighboring points with anti-parallel tangents. Thus, the two parallel edges of an extracted road, will have a minimal width d_{min} above which the repulsion caused by the energy tends to zero at the same time as the profile Ψ . Other variants of quadratic energy implying dependence on the data fit term have been proposed. The results of such active contours represented by their level sets are promising. The insertion of such strong geometric constraints which are non local has been effective, and the implicit representation allows the natural inclusion of the concept of network.

4.3 Active contours for fine map-to-image matching: choice and implementation

4.3.1 Choice of the active contours representation

Within the framework of fine map-to-image matching that we propose to study in this thesis, we have opted for an *implicit representation of the active contour by level sets* for the following reasons,

1. Implicit representation allows for a natural modeling of the corners. This point has particular significance for our application since we consider objects (buildings) which present these geometric singularities.
2. Contrary to explicit representations, representation by level sets is parameters free. The use of this method is therefore more general and easier.
3. Implicit representation allows natural modeling of the objects with a complex topology. It is therefore possible to model buildings with interior courtyards.
4. The topology of an active contour represented by a level set can change during the convergence. This property does not seem very pertinent for our application since we assume that the shape constraint determines the topology of the active contour. However, we will see that it is possible to use the flexible topology produced by level sets in order to solve the problem of local minima (see section 5.4).
5. The level sets are numerically more stable than the explicitly represented contours, in particular for the modeling of geometric singularities.
6. Comparison between two forms represented by their level sets is direct and effective. For this reason, the incorporation of prior shape constraints is easier.

On the other hand, the drawbacks of the implicit representation are as follows,

1. Greater computational complexity. That is inherent to the $n + 1$ dimensions order of the level set representing the active contour. We will see in section 4.3.4.1 how the level set restriction on a narrow band results in a reduction of the calculation time.
2. The distance function for the level sets is not preserved in the Hamilton-Jacobi evolution equations. This implies a periodic re-initialization of the level sets during the evolution of the active contour. The calculation cost is therefore increased.

4.3.2 Choice of a term for the dependence on data

In this study, we use two *region-based* models for the dependence on the data: that of Chan and Vese [19] and the Bayesian model [90]. Chan and Vese's model is more effective if the image to be analyzed is piecewise constant. This restricts the field of application of our method to buildings for which the radiometry of the roof is uniform. We will see how the Bayesian model frees us from this hypothesis and may make it possible to process homogeneous buildings. Our motivation in choosing region-based active contours is based on their decreased sensitivity to initialization and to noise compared with edge-based active contours. However, we shall also explore *edge-based* contours because of their ability to process inhomogeneous buildings. In order to make the contour less sensitive to initialization, we will use the Gradient Vector Flow (GVF) method which allows to diffuse the very local information of the image gradients. The level set evolution equations with respect to the active contours that we will use are as follows:

4.3.2.1 Region-based models

* *Chan and Vese model*

$$\phi_t(\mathbf{x}, t) = -\delta_a(\phi(\mathbf{x}, t)) \left(-[I(\mathbf{x}) - c_{out}(\phi(\mathbf{x}, t))]^2 + [I(\mathbf{x}) - c_{in}(\phi(\mathbf{x}, t))]^2 \right) \quad (4.61)$$

in which ϕ is the level set representing the active contour, δ_a a regularized approximation of Dirac's distribution, c_{in} and c_{out} are averages of the image I on the inside and outside respectively of the active contour.

* *Bayesian model*

$$\phi_t(\mathbf{x}, t) = \left\{ -\frac{(I(\mathbf{x}) - c_{in}(\phi(\mathbf{x}, t)))^2}{\sigma_{in}^2(t)} + \frac{(I(\mathbf{x}) - c_{out}(\phi(\mathbf{x}, t)))^2}{\sigma_{out}^2(t)} + \ln \left(\frac{\sigma_{out}^2(t)}{\sigma_{in}^2(t)} \right) \right\} \delta_a(\phi(\mathbf{x}, t)) \quad (4.62)$$

in which σ_{in}^2 and σ_{out}^2 are the variances of the image I on the inside and the outside respectively of the active contour. In the equations (4.61-4.62), the regularization term has been omitted for two reasons. It tends to round the corners which could

present the active contour and we try to avoid this effect since we want to segment buildings whose boundaries often present singularities. And, in addition, the shape constraint term has a regularizing effect.

*** Range of application**

- ↪ Building with small elevation (negligible perspective distortion). This constraint comes from the incompatibility of the geometries in which the cartographic and remote sensing data are projected. Since the map is orthoscopic, the satellite image is not rectified. Therefore, in the case of buildings with very high elevation, the initialization of the active contour coming from the map will be placed at the level of the building footprint in the image. The active contour will therefore be distanced from the roof to be segmented, and this will make the result of the fine matching hazardous.
- ↪ Building without generalization effect in the map. Generalization consists in clustering several buildings within a single map component, which may make the considered cartographic object inhomogeneous: this is not compatible with the proposed region-based active contours model.
- ↪ The cartographic representation of the building is not mistaken. We will discuss at the end of this chapter the presence of errors in the map and we shall illustrate the limits of the proposed shape constraint model (see section 5.6).

4.3.2.2 Edge-based model

*** Geodesic models with Gradient Vector Flow**

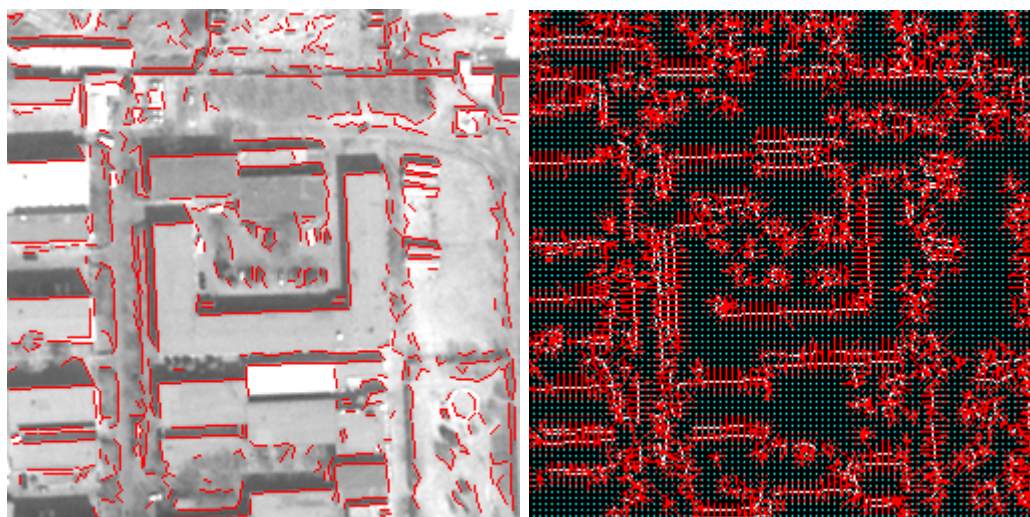
We adapt the model presented in equation (4.28) taken from [91] in the following manner:

$$\phi(\mathbf{x}, t)_t = -\langle(\mathbf{u}, \mathbf{v}), \nabla\phi(\mathbf{x}, t)\rangle \quad (4.63)$$

where (\mathbf{u}, \mathbf{v}) is the gradient vector flow (GVF) resulting from the diffusion of image gradient information. With respect to the equation (4.28) the regularization term has been omitted as well as the weighting term that is inversely proportional to the image gradient modulus. We justify the first omission by the insertion of a shape constraint which will have a regularization effect. The second is justified by the fact that the field (\mathbf{u}, \mathbf{v}) is null on the areas of high gradient which is redundant with the function $g(|\nabla I|)$ of the equation (4.28).

And finally, the last adaptation of the model we use is the information of the gradient which will be diffused by the calculation of the field (\mathbf{u}, \mathbf{v}) . We diffuse some segments primitives that are extracted from the areas of high image gradient. These primitives come not only from a detection of the contours, but also from a subsequent chaining which allows us to find the image segments. The segments primitives are of a higher level and are more likely to characterize buildings whose footprints or roofs are made

of rectilinear boundaries. The use of the segments is similar to a filtering of the image gradient, avoiding to take into account peripheral objects such as trees or urban furniture which do not respond to any rectilinear criterion. Figure 4.3 illustrates the extraction of these primitives.



(a) Primitive segments extracted from the chaining of the high gradient points (red)

(b) Gradient vector of the segments diffused by the Gradient Vector Flow method.

Figure 4.3: Example of the extraction of segments (a) and of the associated gradient vector flow (b).

* *Range of application*

- ~> Building with low elevation.
- ~> Building with or without generalization effect in the map, with a roof of any shape. The inhomogeneity of clustered buildings in a single cartographic object can be handled by the model that is based only on the image gradients.
- ~> The cartographic representation of the building is not mistaken.

4.3.3 Choice of the shape constraint

There are three aspects to the prior knowledge derived from the map. First, it is knowledge on the nature of the object that we intend to match into the image (we process only the buildings). Second, it is knowledge on the localization which allows us to look for the building at the appropriate point in the image, and we use this knowledge to initialize the active contour. And finally, the map give us the shape of the building that we are likely to find in the image. We constrain the active contour to be akin to the shape derived from the map in using the energy described in (4.54). The derivation

of this energy with respect to the level set ϕ gives the following equation,

$$\phi(\mathbf{x}, t)_t = -\frac{\partial J_{image}}{\partial \phi} + 2\lambda (H(\phi(\mathbf{x})) - H(\psi(\mathbf{x}))) d\mathbf{x} \quad (4.64)$$

with

$$\psi = \psi_0 \circ T_{sim}$$

where J_{image} is any energy criterion of the dependence on the data, ϕ is the level set representing the active contour; ψ_0 is the level set representing the prior shape derived from the map, $\lambda \in \mathbb{R}^+$.

The criterion of the equation (4.54) has the advantage of being independent of the integration domain Ω , and gives a discriminating measure of two shapes represented implicitly by ϕ and ψ_0 [102]. The shape constraint is invariant from the transformation T_{sim} which allows to introduce degrees of freedom between the active contour represented by ϕ and the static prior shape ψ_0 . T_{sim} is a direct plane similarity which allows the invariance of the shape constraint from rotation of angle θ , of the translation μ and scale factor s which is identical in image directions (Ox) and (Oy) ,

$$T_{sim}\mathbf{x} = s \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \mathbf{x} + \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} \quad (4.65)$$

According to the expression in the equation (4.54), it is not possible to deduce analytically the parameters $\xi = (s, \theta, \mu_x, \mu_y)$ of T_{sim} from the system:

$$\left(\frac{\partial J_{shape}}{\partial \xi_i} \right)_{i=1, \dots, 4} = 0 \quad (4.66)$$

Numerical optimization methods are then used to estimate the parameters $\hat{\xi}$ during the evolution of the active contour. The optimization by gradient descent is commonly used to resolve this problem. We will discuss an alternative to it in section 5.2.

4.3.4 Algorithm and optimization

4.3.4.1 Reducing calculation time: use of the Fast Narrow Band method

If we implement in a strict sense the evolution equation (4.64), we would have to review all the pixels \mathbf{x} of the image in order to completely fill the function ϕ of the instant $n+1$. That would be expensive in calculation time and would be useless. As a matter of fact, only the values of ϕ near to zero are of interest because the zero crossings represent the active contour. In practice, we carry out the updating of the active contour in a narrow band around the zero level of the set ϕ . When the active contour is on the verge of leaving the narrow band, this last is reconstructed from the last position of the contour.

The distance function of the level set is not preserved in the evolution equations of the Hamilton-Jacobi type which are usually derived from variational approaches. Therefore, the condition $|\nabla \phi| = 1$ is not verified through time. Two options are possible to solve the problem. The most satisfying is to derive evolution equations which are not

of the Hamilton-Jacobi type, as José Gomes proposed in his thesis, and in [43]. The calculation complexity is increased but the problem is solved in an elegant manner. The second way that we have chosen is to reinitialize the level set every N iterations when the contour is on the verge of leaving the narrow band. The concept of re-initialization in a narrow band was originally proposed in a paper by D.L. Chopp dealing with the physics of interfaces [26]. In [115] Sussman proposed a method for narrow band reconstruction by using the differential equation:

$$\phi_\tau(\mathbf{x}, t) = \text{sign}(\phi(\mathbf{x}, \tau)) (1 - |\nabla \phi(\mathbf{x}, \tau)|) \quad (4.67)$$

This approach proposes an approximation of the distance function, which is as fine as the number of iterations is large. The problem with such a method is the necessity of calculating the gradient of the level set ϕ , and this can be delicate in the neighborhood of a geometric singularity. A poor estimation of the gradient in these regions will bring about and propagate errors in the estimation of the distance function. In [60], R. Keriven discusses older algorithms of re-initialization. R. Keck discusses in considerable detail the influence of the re-initialization on the precision of the calculations within the framework of fluid mechanics [59]. In [89], N. Paragios and R. Deriche propose a faster local method – called *Hermes* – than [115], and in [110], Sethian describes a technique based on *Fast Marching*. We have chosen a method proposed by Yui in 2002, called *Fast Narrow Band* and which is faster than the aforementioned approaches [132]. This method is made up of two phases:

1. For each pixel in the coordinates \mathbf{x} belonging to zero level of the level set, an approximation of the distance to the neighborhood pixels is assigned according to the circular neighborhoods illustrated in figure 4.4: the process is repeated for the circular neighborhoods progressively larger until the size of the desired narrow band is reached. When a pixel is assigned several times a different value, the minimal value is kept. At the end of this phase, a narrow band containing coarse approximations of the distance function is constructed.
2. For each pixel in the narrow band, we look for pixels belonging to the contour (0 level of the level set) in the circular neighborhood whose radius value has been assigned in advance. Among all the pixels of the contour that are found, their Euclidean distances to the pixel of reference is calculated; the minimal value is finally attributed to the pixel being considered in the narrow band.

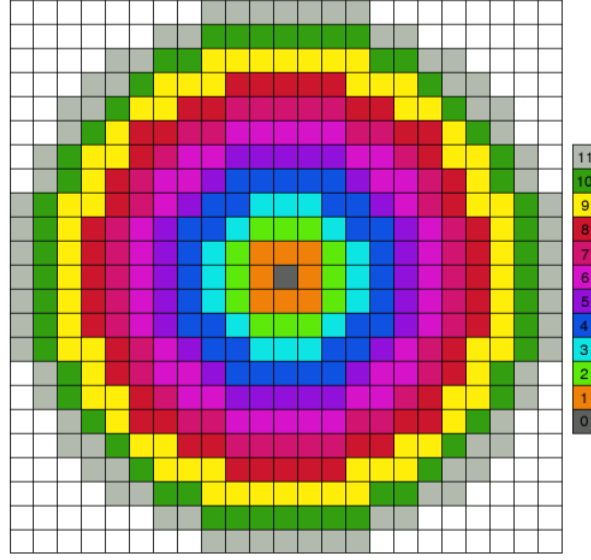


Figure 4.4: Circular neighborhoods used for the construction of the narrow rapid band. The caption gives an approximation of the pixel distance to the center.

This astute procedure allows for manipulation of circular neighborhoods thus limiting considerably the research space for the pixels of the contour. Nevertheless, we can be critical about the necessity of the contour to have a pixel precision and not sub-pixel. This loss of precision creates a penalty for future calculations of the curvature. The authors propose to correct this problem by diffusion at the end of the second algorithm phase. We propose to adapt the algorithm of the fast narrow band in order to calculate the exact distance function, and, while keeping a precise sub-pixel contour,

1. For each point with the sub-pixel coordinates $\mathbf{x}_{contour,sp}$ belonging to the contour, we round off the coordinates on the pixel grid while remembering the link $\mathbf{x}_{contour} \rightarrow \mathbf{x}_{contour,sp}$, then we assign an approximation of the distance function to the neighboring pixels in accordance with the circular neighborhood as shown in figure 4.4: we reiterate the process for the circular neighborhoods progressively larger until we reach the desired size of the narrow band. When a pixel is assigned several times a different value, the minimal value is kept. At the end of this phase, a narrow band containing coarse approximations of the distance function is constructed.
2. For each pixel in the narrow band, we look for pixels belonging to the contour (0 level of the level set) in the circular neighborhood whose radius value has been assigned in advance. For each pixel found in the contour, $\mathbf{x}_{contour,k}$, the Euclidean distance is evaluated to the sub-pixels $\mathbf{x}_{contour,sp,k,i}$ linked to $\mathbf{x}_{contour,k}$. The minimum value of all the $\mathbf{x}_{contour,sp,k,i}$ is finally attributed to the pixel that is being considered in the narrow band.

The building up of the narrow band occurs for three kinds of events: the construction of $\phi(\mathbf{x}, t = 0)$, the reconstruction of the narrow band when the active contour is on the

verge of leaving it, and when the level set needs to be reinitialize because it is so longer a distance function. Practically, we preferred to work with a very narrow band of two or three pixels and reinitialize at each iteration of the active contour.

The construction of a narrow band is indispensable and its dual operation is also necessary. Finding the contour from the level set by the detection of the zeros of the latter intervenes for two reasons: the display of the active contour, and the re-initialization of the fast narrow band. The extracted contour of the level set is a linear approximation of the real contour (figure 4.5). It is detected by the sign changes of the function $\phi(\mathbf{x}, t)$, and the sub-pixel coordinates are calculated by linear interpolation.

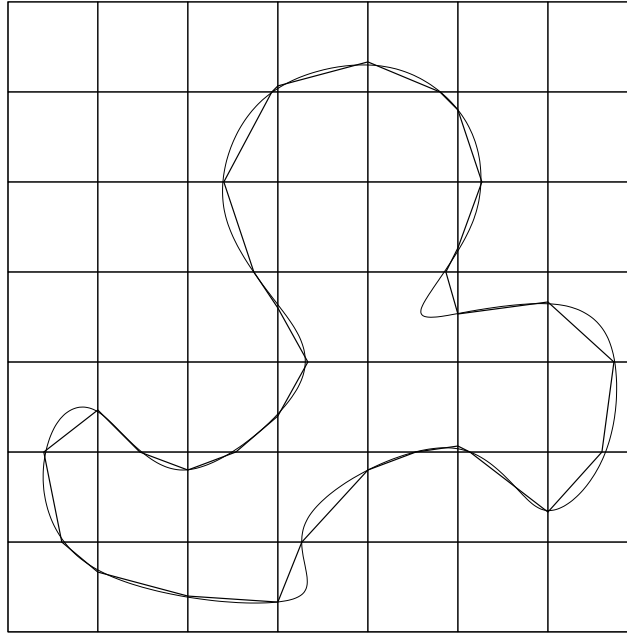


Figure 4.5: Linear approximation of a contour

4.3.4.2 Evolution diagram of the shape constrained active contour

The evolution diagram shown in figure 4.6 describes the updating algorithm of the active contour. In this figure, the function \mathcal{F} is defined by the evolution equation. In all the experiments carried out on real images, the polygonal prior shape derived from the map and the initial active contour are identical.

- * In practice, the image space considered in order to cause the active contour to evolve is a sub-set of Ω . Typically, a sub-image of 256 by 256 pixels is extracted from the satellite image Quickbird (which uses 30,000 by 30,000!) at the spot of the cartographic polygonal that we want to match and revise. If the size of the box encompassing the polygonal is greater than 256 by 256 pixels, we extract a larger sub-image having a margin of 100 pixels with respect to the encompassing box.

- * In a case where region-based models are used, the characteristic of homogeneity of the buildings represented in the image can be forced by pre-processing it with anisotropic diffusion (preserving the contour).
- * Generation of the level set ψ_0 representing the prior shape is carried out during initialization and covers the totality of the sub-image studied (unlike the set ϕ calculated only in a narrow band). In order to reduce the calculation time, we use Maurers algorithm [70] which allows for a calculation of the Euclidean distance on binary images in $O(N)$ where N is the number of pixels in the image. The cartographic polygon is thus rasterized before the application of this algorithm.
- * We decided to estimate the parameters of T_{sim} after each updating of the active contour.

4.4 Conclusion

This chapter retraced the state of art of active contours with a particular attention paid to the incorporation of the prior shape constraint. We exposed our choice of active contours represented by level sets in order to carry out the matching of the buildings of the map onto the image. Different terms expressing the relationship with the data were selected in order to bring the active contour close to the building represented in the image. Two region-based models were chosen because of their simplicity of implementation and their decreased sensitivity to the initialization. They are, nonetheless, limited to the segmentation of homogeneous buildings in the image. An edge-based model, intrinsically more sensitive to the initialization will be used to match inhomogeneous buildings. Finally, details related to the implementation and the reduction of the calculation complexity of the level sets were described. The following chapter will present our contributions for improving the robustness of the chosen models with shape constraint.

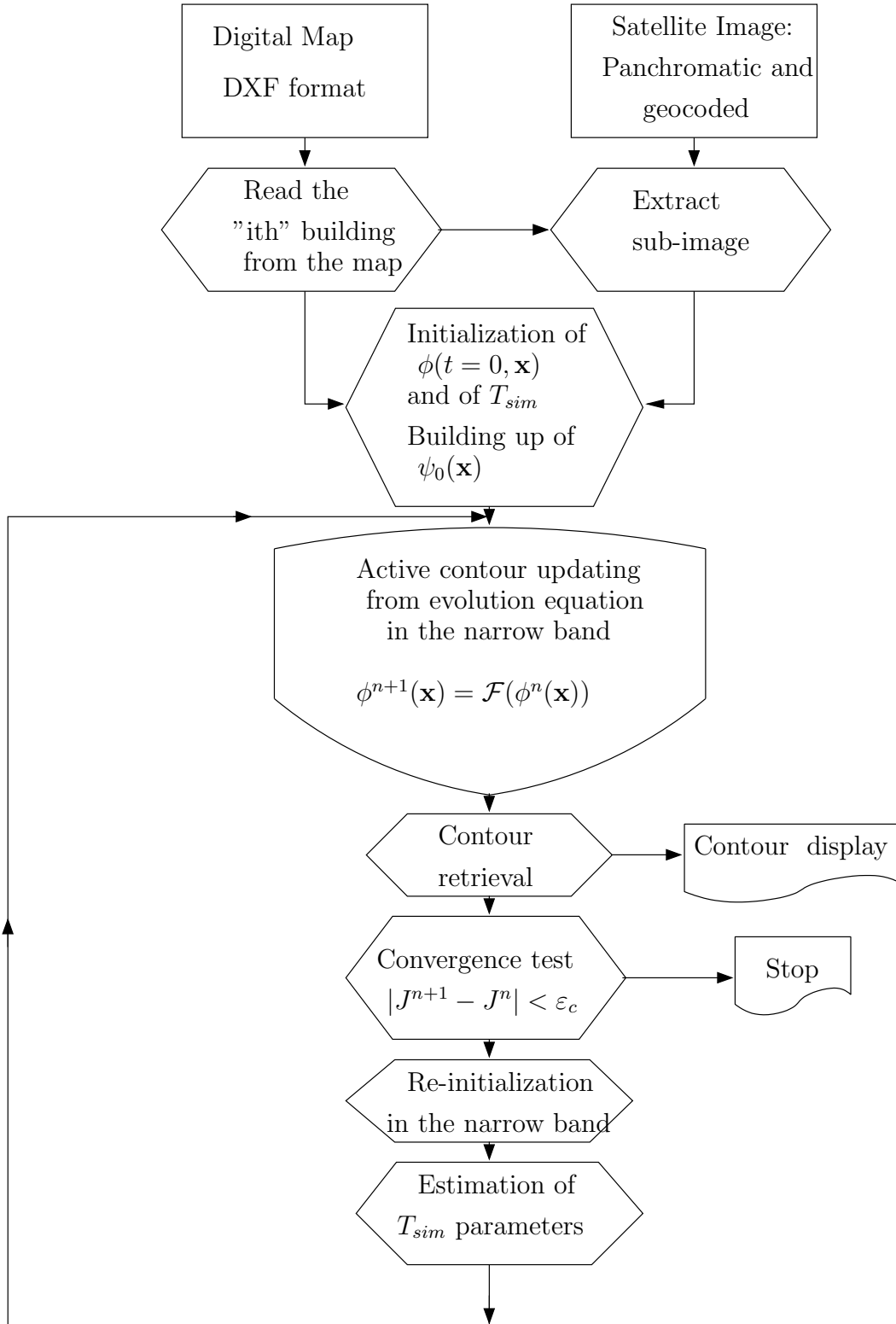


Figure 4.6: Evolution diagram of the shape constrained active contour

Chapter 5

Robust shape constrained active contours

5.1 Introduction

This chapter examines the improvement of the robustness of shape constrained active contours which are intended to achieve a matching between the cartographic buildings and the image. The goal of such a matching is to increase the map/image consistency to enable a later “no-change” detection that is reliable. We propose to improve the active contour models proposed in the previous chapter in order to increase their matching power. The first contribution consists of proposing a different optimization algorithm for estimating the pose parameters making the shape constraint invariant from a global transformation. The second one concerns the insertion of the Digital Surface Model (DSM) in the active contours’ attachment to data term. This type of representation is complementary to the satellite image and enables a better discrimination of the constructions from the rest of the scene. The third innovation concerns the incorporation of shape constraint varying according to space and time. This is intended to confer greater flexibility on the active contours to make them less sensitive to the local minima of the minimized energy. The experimental results proposed at the end of the chapter enable to assess the contribution of these three solutions to increasing the robustness and efficiency, in terms of calculation time, of the map-to-image matching.

5.2 Robust estimation of the T_{sim} parameters by the simplex algorithm

As we mentioned in section 4.3.3, it is possible that the expression of the shape constraint energy described in general terms in equation (5.1) may not have an analytical solution of the parameters $(\xi)_{i \in (1, \dots, N)}$ of the overall transformation T :

$$J_{shape}(\phi, \psi) = \int_{\Omega} \mathcal{D}(\phi(\mathbf{x}), \psi_0(T\mathbf{x})) d\mathbf{x} \quad (5.1)$$

where D represents a distance measuring the similarity between the shape of the active contour ϕ and the reference shape ψ_0 . This is the case of the energy formulated in equation (4.54) where $\mathcal{D}(\phi(\mathbf{x}), \psi_0(T\mathbf{x})) = (H(\phi(\mathbf{x})) - H(\psi_0(T\mathbf{x})))^2$ and $T = T_{sim}$ ($N = 4$). In the absence of a formal expression of the parameters that would have made the approach intrinsic, it is necessary to estimate them by numerical optimization as the active contour evolves. In [21, 23, 104, 24], the authors propose to calculate numerically the parameters minimizing the energy of equation (4.54) by gradient descent. We then have four equations relative to each of the T_{sim} parameters, that is to say the scaling s , rotation θ and translation μ :

$$\begin{cases} s^{n+1} = s^n + \Delta t_s F(\phi, \psi_0, s^n, \theta^n, \mu_x^n, \mu_y^n) \\ \theta^{n+1} = \theta^n + \Delta t_\theta G(\phi, \psi_0, s^n, \theta^n, \mu_x^n, \mu_y^n) \\ \mu_x^{n+1} = \mu_x^n + \Delta t_{\mu_x} M(\phi, \psi_0, s^n, \theta^n, \mu_x^n, \mu_y^n) \\ \mu_y^{n+1} = \mu_y^n + \Delta t_{\mu_y} N(\phi, \psi_0, s^n, \theta^n, \mu_x^n, \mu_y^n) \end{cases} \quad (5.2)$$

where n is the iteration rank, ϕ and ψ_0 are the level sets respectively representing the active contour and the prior shape. Δt_s , Δt_θ , Δt_{μ_x} and Δt_{μ_y} are the time steps associated with each of the gradient descents estimating the T_{sim} parameters. The detail of the expressions of the functions F , G , M and N is given in [21].

The estimation of the parameters using the four descents of system (5.2) may be found to be extremely perilous. The delicate point is the tuning of the time steps of the four descents relative to each of the parameters. Whereas each descent estimates a parameter with a different geometrical meaning (translation, rotation, scaling) it is highly likely that the speed of each descent will not be the same. For example, the descent estimating the rotation should be much slower than the translation descents. However, there is no theoretical framework making it possible to clarify this question which gives rise to empirical settings that are difficult, unstable and which influence the estimation of the parameters. Lastly, the inter-dependence of the four gradient descents means that a poor estimation of one parameter in a descent will affect all the others to such an extent that it will rapidly make the optimization diverge. Daniel Cremers mentions the issue of tuning the gradient descents in [30].

A more favorable alternative that we have chosen is to use the simplex algorithm (a reminder of the detail of this algorithm and some illustrations are given in appendix B). We then minimize the shape constraint energy detailed in equation 5.1) by a finding a set of N optimum parameters estimated simultaneously. The optimization is carried out by reflection, expansion and contraction of a polyhedron with $N + 1$ vertices in the parameters' space. The advantage of the simplex is the absence of setting parameters that may impact the result of the optimization. This approach, which is less sensitive to the energy local minima, is more robust than the gradient descent. It is a zero order method (it does not require the calculation of the cost function's gradient) that is better placed to find a minimum of a functional that is not strictly convex. Lastly, the simplex is able to minimize more complex shape constraint energies, with more parameters to be estimated, without requiring the theoretical calculation of the partial derivatives $\frac{\partial J_{shape}}{\partial \xi_i}$ inherent to the gradient descents.

However the simplex is more complex to calculate than the gradient descent. Indeed, the iterative construction of the vertices of the simplex requires the evaluation of the cost function J_{shape} which is therefore performed many times during the convergence process. J_{shape} is an integral over the whole image domain and involves a great calculation complexity, which will significantly increase the calculation times with respect to the gradient descent. It is nevertheless possible to limit the integral calculation of J_{shape} within a narrow band in order to lower the calculation workload. The expression of the energy of equation (4.54) then becomes:

$$J_{shape}(\phi, \psi) = \int_{\Omega} N(\phi(\mathbf{x}), \varepsilon_b) (H(\phi(\mathbf{x})) - H(\psi(\mathbf{x})))^2 d\mathbf{x} \quad (5.3)$$

where the function $N(\phi(\mathbf{x}), \varepsilon_b)$ is equal to 1 if $|\phi(\mathbf{x})| \leq \varepsilon_b$, otherwise it is equal to 0. The value of ε_b which is chosen arbitrarily is a trade-off between the calculation complexity and the desired precision of the parameter estimation. Indeed, a broad band would enable a good differentiation of shapes ϕ and ψ , which will guarantee a precise and robust assessment of the parameters, but for a high calculation cost. A band that is too narrow, on the contrary, will lead to a rapid convergence on an imprecise estimation, or even a local minimum of the energy functional. A half-band of three pixels $\varepsilon_b = 3$ has been found to be a satisfactory compromise. The only parameterization of the simplex algorithm is relative to the construction of the initial simplex. The first vertex of the simplex is calculated from a set of initial parameters ξ_{ini} , the remaining vertices are constructed by adding an individual variation $\delta_{\xi(i)}$ into each of the parameters of ξ_{ini} . In the case of T_{sim} , four parameters have to be estimated. The simplex will therefore have five vertices and its initial state will be:

- * Initial vertex 1: calculated from $(s_{ini}, \theta_{ini}, \mu_{x,ini}, \mu_{y,ini})$
- * Initial vertex 2: calculated from $(s_{ini} + \delta_s, \theta_{ini}, \mu_{x,ini}, \mu_{y,ini})$
- * Initial vertex 3: calculated from $(s_{ini}, \theta_{ini} + \delta_{\theta}, \mu_{x,ini}, \mu_{y,ini})$
- * Initial vertex 4: calculated from $(s_{ini}, \theta_{ini}, \mu_{x,ini} + \delta_{\mu_x}, \mu_{y,ini})$
- * Initial vertex 5: calculated from $(s_{ini}, \theta_{ini}, \mu_{x,ini}, \mu_{y,ini} + \delta_{\mu_y})$

The choice of values of $\delta_{\xi(i)}$ conditions the size of the initial simplex and therefore its capacity for estimating the parameters that are distant from ξ_{ini} . The safest choice is a relatively large initial simplex which will guarantee a better robustness of the optimization but at the price of a higher calculation cost. The comparison of the two algorithms on the basis of experimental results is given in section 5.5.2.

5.3 Fusion of exogenous terms in the attachment to the data energy functional

Examination of figure 5.1 shows that the buildings represented in the panchromatic satellite image are locally discriminated from the rest of the image, but not globally. So, the radiometry of a building may be different from its immediate vicinity without,

for all that, being discriminated from the rest of the image in the sense of the statistical quantities such as the mean. These quantities are however used in the region-based models that we have chosen. This difficulty, which is inherent to urban environments, will make the convergence of the region-based active contour sensitive to initialization. This will then have a greater risk of reaching a local minimum of the energy functional to be minimized.

However the Digital Surface Model (DSM) that we have previously generated (see chapter 3) seems to meet better the characteristics of the models chosen and based on the region information. Indeed the intensity of the DSM pixels is proportional to the altitude of the reconstructed objects, guaranteeing that the buildings will stand out from their immediate surroundings, or from the rest of the image if the construction density is not too high. It is possible to force the discrimination of a building represented in the DSM by modifying the latter's histogram. For our application, we have adopted the following way of doing this:

$$I_{MNS}^{clipped}(\mathbf{x}) = \frac{1}{2} \max_{\mathbf{x} \in I_{MNS}} (I_{MNS}(\mathbf{x})) \left[1 + \frac{2}{\pi} \arctan \left(\frac{I_{MNS}(\mathbf{x}) - m + s}{\zeta} \right) \right] \quad (5.4)$$

where m is the mean altitude of the DSM within the cartographic polygon superimposed with the DSM; $s = \zeta \tan \left(\frac{\pi(2ratio-1)}{2} \right)$ guarantees that the pixels of the DSM of value m will have an assigned value equal to $\max(I_{MNS}(\mathbf{x})) \times ratio$. The value of $ratio$ has been set at 0.95 in our experiments. The variable ξ enables to set the slope of the \arctan function, the lower the value of ξ with respect to 1, the steeper the slope will be when $I_{MNS}(\mathbf{x}) = m - s$, thus increasing the contrast between the building considered and the objects that are not so tall. On the contrary, the building will have an intensity that is only very slightly different from taller objects. We have chosen $\zeta = 0.5$. Figure 5.1 shows part of the original DSM and the result of the pre-processing formalized in equation (5.4). It can be seen that the pre-processed DSM is more like a two-phase image in the sense of the mean, and will have a better chance of working with the attachment to data term of the Chan and Vese model or of the Bayesian model. However, the 3D reconstruction of the DSM is very "noisy", there are a large number of unfilled pixels. So, it would seem difficult to obtain a precise segmentation of the building using the DSM alone.

We propose to merge the radiometric information from the satellite image and the altitude information from the DSM to accomplish the building segmentation. This merging of exogenous data is justified by the complementary representations of the scene proposed by the remote sensing image and the DSM:

- * In the satellite image, the boundaries of the building are most often easy to characterize and the image has a very high signal-to-noise ratio. However the presence of shadows, peripheral objects on the ground or partially masking the building make segmentation difficult. The heterogeneity of the radiometry of the building's roof or environment may cause the attachment to data models based on homogeneous regions to fail.

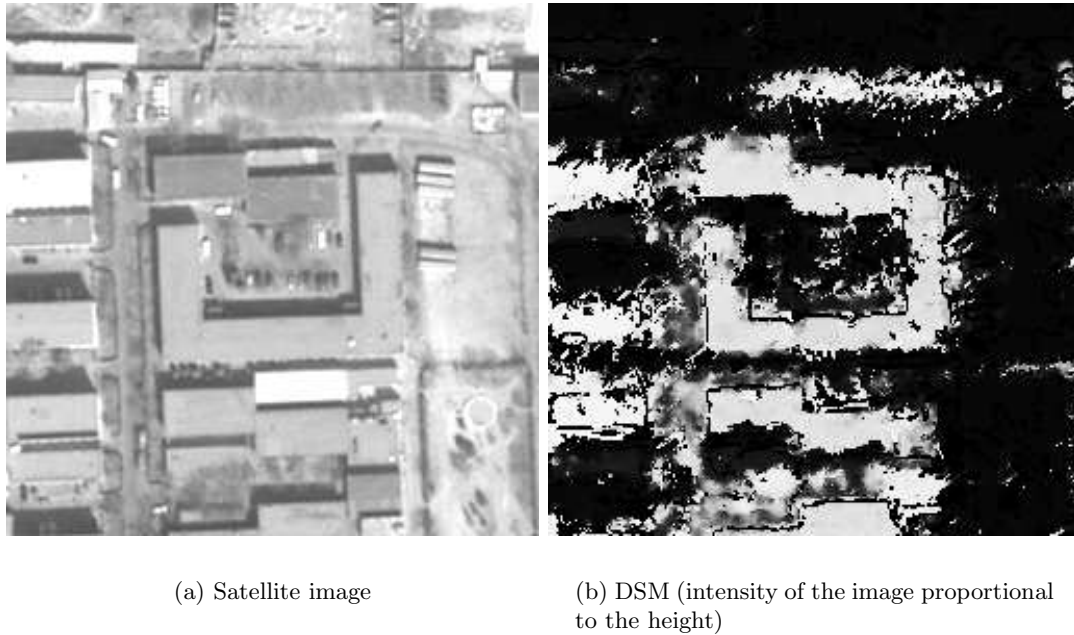


Figure 5.1: Different representations of the same scene: satellite image and 3D reconstruction

- * Inversely, the DSM represents the boundaries of the building's roof poorly because the three dimensional reconstruction is delicate in places of altitude transitions. However the problem of shadows and of peripheral objects located at the level of the ground is eliminated. The building represented in the DSM is better discriminated from the rest of the image.

These two representations complete each other in a satisfactory way. But, if we want to merge them, they must fulfill certain consistency conditions:

1. **The data must represent the same subject of interest:** this poses the problem of the date of data acquisition. Since the data are exogenous, they are likely to have been acquired at different times, times between which changes may have occurred. So, a building represented in the DSM generated from the aerial images taken in 1999 may have disappeared from the Quickbird image taken in 2002. This consistency condition means that we must test the merging of the exogenous data on areas where the object of interest in the scene is unchanged between the acquisition dates.
2. **Data must be superimposable:** each pixel of the data to be merged must represent the same part of an object in the scene. This clause raises two questions.

The first one concerns the precision of the data registration. In our case, the image and the DSM are geocoded in the same cartographic system. The image has a geocoding precision lower than the pixel whereas that of the DSM is equal to two pixels. This seems sufficient, given that at the end of the active contour convergence it is possible to reduce the influence of the DSM and thus give more weight to the image which is the more reliable of the two. The second issue concerns the geometry of the data to be merged. The DSM is in orthoscopic geometry whereas the Quickbird satellite image is not. This will certainly cause superimposition problems for the tall buildings which present a distortion due to the perspective in the image: the roof and the footprint are then not superimposed, contrary to the DSM. This will constrain us to only process the buildings that are not very tall, in the absence of an orthoscopic satellite image.

If it is assumed that the two consistency conditions are verified, we merge the satellite image and the DSM in the Chan and Vese's region-based attachment to data energy functional in the following way:

$$\begin{aligned} J_{\text{fusion,CV}}(\phi) &= J_{\text{CV}}(\phi) \\ &+ \lambda_{MNS} \left[\int_{\Omega} |I_{DSM}(x, y) - c_{DSM,in}|^2 H(\phi(x, y)) dx dy \right. \\ &\left. + \int_{\Omega} |I_{DSM}(x, y) - c_{DSM,out}|^2 (1 - H(\phi(x, y))) dx dy \right] \end{aligned} \quad (5.5)$$

where J_{CV} is defined in equation (4.31), I_{DSM} is the image of the DSM. The statistical quantities indexed by DSM are defined as in (4.32-4.33) but while referring to the DSM. The term that we have just added to the J_{CV} functional is the Chan and Vese attachment to data energy term relative to the DSM exogenous data term.

The extension of the Bayesian model of equation (4.44) is similar:

$$\begin{aligned} J_{\text{fusion,Bayes}}(\phi) &= J_{\text{Bayes}}(\phi) \\ &+ \lambda_{MNS} \left[\int_{\Omega} \left(\frac{|I_{MNS}(x, y) - c_{DSM,in}|^2}{2\sigma_{DSM,in}^2} + \ln(2\pi\sigma_{DSM,in}^2) \right) H(\phi(x, y)) dx dy \right. \\ &\left. + \int_{\Omega} \left(\frac{|I_{MNS}(x, y) - c_{DSM,out}|^2}{2\sigma_{DSM,out}^2} + \ln(2\pi\sigma_{DSM,out}^2) \right) (1 - H(\phi(x, y))) dx dy \right] \end{aligned} \quad (5.6)$$

It is also conceivable to merge the DSM data term with the attachment to data model based on the boundary information (GVF) (section 4.3.2.2). We choose to integrate the DSM in the form of a region-based term. Indeed, the DSM better meets the characteristic of a piecewise constant or homogeneous image. It would therefore seem to be better justified using a model based on the region information rather than on the boundaries. The integration in the GVF model of the DSM according to the Chan and Vese model is equivalent to the following equation:

$$\phi(\mathbf{x}, t)_t = -\langle (\mathbf{u}, \mathbf{v}), \nabla \phi(\mathbf{x}, t) \rangle$$

$$\begin{aligned}
& -\lambda_{MNS}\delta_a(\phi(\mathbf{x}, t)) \left\{ -[I_{MNS}(\mathbf{x}) - c_{MNS,out}(\phi(\mathbf{x}, t))]^2 \right. \\
& \quad \left. + [I_{MNS}(\mathbf{x}) - c_{MNS,in}(\phi(\mathbf{x}, t))]^2 \right\}
\end{aligned} \tag{5.7}$$

According to the models of equations (5.5) to (5.7), the active contour will move in parallel on two images: the DSM and the satellite image, and will be a trade-off between the information derived from these two representations and that of the prior shape constraint. It should be noted that the precision of the DSM geocoding ($\simeq 1\text{ m}$) is lower than that of the rectified Quickbird image which will be used as the most reliable source of information ($\simeq 0.2\text{ m}$). We therefore cancel the weight λ_{MNS} at the end of the convergence process so that the active contour is more faithful to the image. The preliminary phase consisting of giving more credit to the DSM is intended to bring us close to the final solution rapidly without the problems of local minima.

In [47], T. Guo moves a snake-type active contour on an orthoscopic remote sensing image and a DSM taken by LIDAR. A notable difference with respect to our approach is the utilization of an attachment to data term based on the boundaries in the altimetric and satellite image, which required the use of artificial inflation forces. Furthermore, no shape constraint was used and a DSM of higher quality, taken from a measurement rather from a reconstruction, was used.

The contribution of the DSM for a lower sensitivity to local minima is presented in section 5.5.5.1.

5.4 More flexible incorporation of the prior shape: spatio-temporal flexibility of λ

Still with a view to making the map-to-image matching more robust, we discuss in this section the influence of the weight λ which balances the image information and the shape constraint in equations (4.48) and (4.64). The setting of this constant, which weights the influence of the shape constraint, is not trivial for the following reasons:

1. Too weak a constraint would not allow the active contour to be regularized sufficiently and it would thus be sensitive to the alterations to the image. This would lead to an erroneous segmentation of the object which corresponds to a local minimum of the functional J of equation (4.48). In this case, the local minima, which we will call “minima of the first kind”, are numerous.
2. On the contrary, an excessive preponderance of the shape constraint limits the active contour’s possibilities for escaping from the local minima of the functional to be minimized J . In this type of situation, it will cost the active contour more to deform itself to reach the target in the image than to rest motionless and faithful to the shape imposed by the constraint. The chosen solution is therefore a local minimum (of the second kind) which corresponds also to a poor segmentation/matching of the object in the image.

Determining beforehand an optimal value λ_{opt} of the constraint weight appears to be a delicate task since it would depend both on the image and on the degree of alteration that it comprises; these characteristics are not, nevertheless, known in advance. In the context of medical imaging, Y. Chen proposes to try several values of the shape constraint weight and to adopt the value corresponding to the best solution assessed thanks to radiometric profiles [22]. This solution would seem to be sub-optimal since λ_{opt} is not estimated with certainty and, furthermore, the segmentation process must be repeated several times. Lastly, the delicate setting of λ is all the more difficult when the objects in the processed images are not well differentiated from the background. In this case, the information relative to the image is weakened compared with the shape constraint which becomes predominant, which increases the occurrence of local minima of the second kind. This effect is all the more significant when the active contour is initialized too far from the object to be found in the image.

To solve the problem of local minima inherent to an under or over-estimation of the shape constraint, we propose to convert the constant weight λ of equations (4.48) and (4.64) into a space function. The goal of this type of function is to relax the influence of the constraint in the vicinity of level 0 of the level set ψ , while keeping a strong and uniform constraint far from the prior shape. This provides a greater freedom and flexibility for the active contour which nevertheless remains bounded in space. Consequently, the active contour will be able to globally resemble the prior shape while allowing local variations. This flexibility is a determining factor for avoiding the local minima of the second kind. We propose to formalize the spatial profile of relaxation as a symmetrical function of the distance to the prior shape:

$$\lambda_{space}(\psi(\mathbf{x})) = \lambda \left(1 - e^{-\left(\frac{\psi(\mathbf{x})}{d}\right)^2} \right) \quad (5.8)$$

The spatial relaxation function of equation (5.8) is null on the loci of the shape constraint and asymptotically tends towards the constant λ far from the prior shape represented by ψ . The parameter d controls the size of the relaxation space. This formulation does not have any theoretical justification and other simpler expressions could satisfy the relaxation criteria. In section 5.5.5.2 we will verify experimentally that the derivability and the steady point of λ_{space} in $\psi(\mathbf{x}) = 0$ improve the insensitivity to the local minima.

The existence of the relaxation space parameterized by d is intended to prevent the active contour from being subject to the local minima of the second kind. However, the freedom granted to the active contour in the relaxation space will not make it possible to avoid the minima of the first kind. To resolve this problem, we propose to reduce the relaxation space at a later time (from t_1). When this space decreases, the efficiency of the shape constraint is restored uniformly over the image domain. The idea is therefore to achieve a coarse segmentation of the image's object with spatial relaxation ($t < t_1$), and then to re-establish the uniform prior shape subsequently to overcome the alterations to the image ($t \geq t_1$). The weight of the shape constraint thus becomes a function of space and of the iterative process's artificial time. The reduction of the relaxation space is achieved by replacing the constant d with a piecewise linear

function and decreasing as a function of time:

$$d(t) = \begin{cases} d_0 & \text{if } t < t_1 \\ (\varepsilon - d_0) \frac{t-t_1}{t_2-t_1} + d_0 & \text{if } t_1 \leq t < t_2 \\ \varepsilon & \text{if } t_2 \leq t \end{cases} \quad (5.9)$$

where $(d_0, \varepsilon, t_1, t_2) \in \mathbb{R}^{*+}$ with $d_0 > \varepsilon$ and $\varepsilon \ll 1$. d_0 is a constant that determines the initial size of the relaxation space and t_2 sets the speed of the decrease. In section 5.5.5.2 we will see that a simpler time decrease of the Heaviside step type is less efficient because of the inertia due to the estimation of the parameters of T_{sim} .

Finally, we strengthen the spatial relaxation and the restoration of the uniform constraint by replacing the factor λ of equation (5.8) by an increasing time function $\lambda_a(t)$ which will control the overall amplitude of the spatio-temporal function of the shape constraint weight. During the relaxation phase ($t < t_1$), the value is minimal to provide greater flexibility to the active contour. From t_1 , the amplitude increases to reach a maximum which will strengthen the re-establishment of the constraint in order to overcome the alterations to the image. The spatio-temporal function of the shape constraint's weight is finally:

$$\lambda_{flex}(\psi(\mathbf{x}), t) = \lambda_a(t) \left(1 - e^{-\left(\frac{\psi(\mathbf{x})}{d(t)}\right)^2} \right) \quad (5.10)$$

where $\lambda_a(t)$ varies between λ_{min} and λ_{max} with a profile opposed to that of equation (5.9). Figure 5.2 illustrates the spatio-temporal function $\lambda_{flex}(\psi(\mathbf{x}), t)$ in three dimensions.

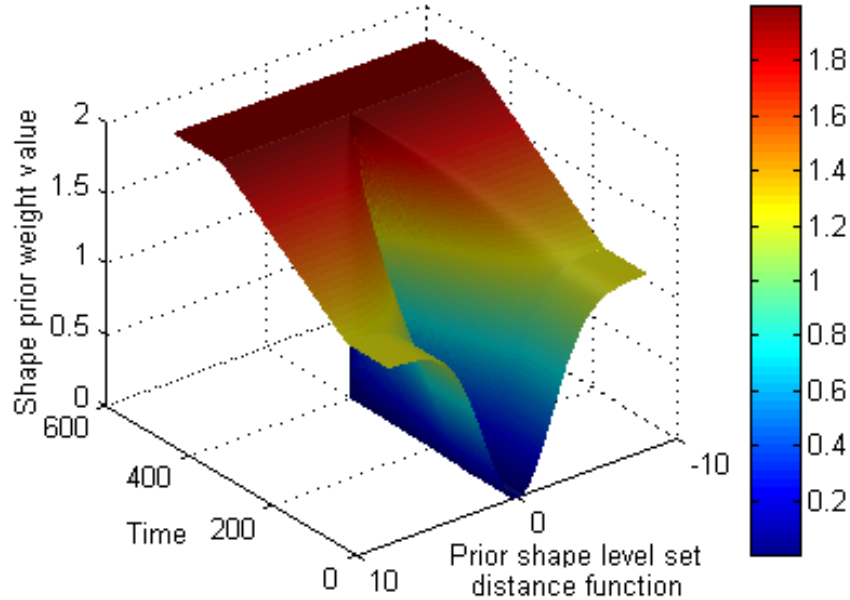


Figure 5.2: Spatio-temporal variation of the function weighting the shape constraint. $d_0 = 3$, $t_1 = 100$, $t_2 = 400$, $\lambda_{min} = 1.25$ et $\lambda_{max} = 2$.

5.5 Application and analysis of the experimental results

This section presents the result obtained with the three innovations proposed, that is to say the estimation of the global transformation's parameters by the simplex algorithm, fusion of exogenous data in the attachment to data term of the active contours and lastly the spatio-temporal variation of the shape constraint. We will compare these results with the most conventional models and will examine to what extent the solutions that we have proposed make it possible to overcome the problem of local minima. Lastly, we will not restrict ourselves to the remote sensing domain and will illustrate the results obtained with images used in Computer Vision and in medical imaging. These images share the same characteristics as the satellite images in an urban environment: a lack of overall discrimination of the object to be extracted with respect to the image's background.

5.5.1 Experimental protocol

In all the experiments described below, the active contour and the prior shape have the following characteristics:

1. The prior shape represented by the level set $\psi_0(\mathbf{x})$ corresponds *exactly* to the boundaries of the object to be segmented/matched in the image.
2. The initial active contour represented by the level set $\phi(\mathbf{x}, t = 0)$ is similar to the prior shape $\psi_0(\mathbf{x})$. It is then transformed by a similarity T_{sim} whose parameters are $\xi_0 = (s_0, \theta_0, \mu_{x,0}, \mu_{y,0})$. The goal of this type of transformation is to create a geometric deviation between the initial active contour and the image's target object in order to examine the method's matching capability.

In some cases, we will illustrate the active contour's evolution from the initial state to the final convergence in the form of a sequence either going from left to right or from top to bottom.

5.5.2 Simplex vs gradient descent comparison for estimating T_{sim} parameters

In order to compare the performances of the two algorithms, let us consider the following experiment: a prior shape ψ_0 is chosen from the two examples in figure 5.3. The active contour represented by ϕ and transformed by T_{sim} will not move over time (no external force due to the image is used, we only work in the shapes' space).

The goal is to determine whether the two methods manage to estimate ξ_0 by minimizing the energy given in equation (5.3).

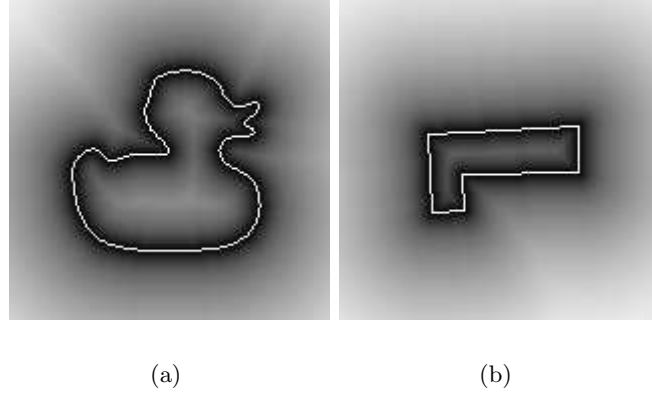


Figure 5.3: Prior shapes (in white, $\{\mathbf{x}/\psi_0(\mathbf{x}) = 0\}$) and their associated level sets (gray levels, $|\psi_0(\mathbf{x})|$). (a) toy (b) L-shaped building.

The number of iterations required to achieve convergence and the calculation time will be compared. We carry out the same experiment several times with different initializations $(s_0, \theta_0, \mu_{x,0}, \mu_{y,0})$:

1. Translation only: $\xi_{0,1} = (1, 0, 10, -5)$.
2. Rotation only: $\xi_{0,2} = (1, \pi/3, 0, 0)$.
3. Scaling only: $\xi_{0,3} = (1.2, 0, 0, 0)$.
4. Direct plane similarity: $\xi_{0,4} = (1.2, \pi/3, 10, -5)$.

In the experiments carried out, we placed the accent on the robustness of the optimizations by working with a narrow band of 11 pixels. The calculation times will be high, but the goal of these experiments is to compare the two algorithms rather than their absolute performances. The simplex and the gradient descent will have converged when they verify the criterion: $|J_{shape}^{n+1} - J_{shape}^n| < 10^{-4}$. The two algorithms were parameterized as follows:

1. Gradient descents: $\Delta t_s = 5 \cdot 10^{-7}$, $\Delta t_\theta = 2 \cdot 5 \cdot 10^{-7}$, $\Delta t_{\mu_x} = 2 \cdot 5 \cdot 10^{-3}$ and $\Delta t_{\mu_y} = 2 \cdot 5 \cdot 10^{-3}$
2. Simplex: $\delta_s = 0.25$, $\delta_\theta = \pi$, $\delta_{\mu_x} = 10$ and $\delta_{\mu_y} = 10$

Tables 5.1 and 5.2 respectively give the results obtained with the gradient descent and the simplex. The first thing we note is the robustness of the simplex which estimates ξ_{ini} correctly in every situation. The gradient is more sensitive to the local minima of the energy functional J_{shape} and therefore fails in three cases to provide a correct estimation of the parameters. The second remark concerns the calculation time. The simplex has shown itself to be much more efficient than the gradient descent. The number of iterations required to achieve convergence is lower with, in these experiments, a higher

speed varying between a ratio 5 to 68. The main reason for the slowness of the gradient descent is the numerical stability. A small time step must be used for each descent in order to meet the stability conditions, with the risk of the system becoming unstable or divergent. It is nevertheless important to note that a slightly different setting of the time steps could provide slightly faster or slightly longer convergences. The results in table 5.1 must therefore be qualified but they do give an order of magnitude concerning the convergence times.

| Shape | Experiment | Absolute error | | | Iterations | Time (s) |
|-------------------|-------------|-----------------------|------------|-----------------------|------------|----------|
| | | $\Delta\theta$ in rad | Δs | $\Delta\mu$ in pixels | | |
| Toy | $\xi_{0,1}$ | 0 | 0 | 0 | 440 | 15 |
| | $\xi_{0,2}$ | 0.8 | 1.05 | $(-14.3, -3.17)$ | 1680 | 58 |
| | $\xi_{0,3}$ | 0 | 0 | 0 | 550 | 21 |
| | $\xi_{0,4}$ | 0.8 | 1.05 | $(-14.3, -3.17)$ | 2780 | 98 |
| L-shaped building | $\xi_{0,1}$ | 0 | 0 | 0 | 420 | 11 |
| | $\xi_{0,2}$ | 0 | 0 | 0 | 2750 | 68 |
| | $\xi_{0,3}$ | 0 | 0 | 0 | 330 | 8 |
| | $\xi_{0,4}$ | 0.3 | 1.51 | $(16.47, 12.54)$ | 1170 | 30 |

Table 5.1: Absolute error between ξ_0 and its estimate $\hat{\xi}_0$ by the gradient descent algorithm and calculation time.

| Shape | Experiment | Absolute error | | | Iterations | Time (s) |
|-------------------|-------------|-----------------------|------------|-----------------------|------------|----------|
| | | $\Delta\theta$ in rad | Δs | $\Delta\mu$ in pixels | | |
| Toy | $\xi_{0,1}$ | 0 | 0 | 0 | 338 | 3 |
| | $\xi_{0,2}$ | 0 | 0 | 0 | 251 | 3 |
| | $\xi_{0,3}$ | 0 | 0 | 0 | 204 | 3 |
| | $\xi_{0,4}$ | 0 | 0 | 0 | 390 | 3 |
| L-shaped building | $\xi_{0,1}$ | 0 | 0 | 0 | 201 | <1 |
| | $\xi_{0,2}$ | 0 | 0 | 0 | 312 | 1 |
| | $\xi_{0,3}$ | 0 | 0 | 0 | 216 | 1 |
| | $\xi_{0,4}$ | 0 | 0 | 0 | 276 | 1 |

Table 5.2: Absolute error between ξ_0 and its estimate $\hat{\xi}_0$ by the simplex algorithm and calculation time.

5.5.3 Contribution of the prior shape knowledge for segmentation

In this section, we compare the performances of the active contours chosen in section 4.3.2 with and without incorporation of prior shape information. The goal is to illustrate the need for using the knowledge of the shape in order to correctly segment an object whose representation is altered in an image.

Figure 5.4 shows the results obtained with the Chan and Vese model and the

Bayesian model with a building meeting the homogeneity criterion. In this experiment, the initial active contour has not undergone any transformation ($\xi_0 = \mathbf{0}$) and is therefore exactly superimposed over the U-shaped building in the image. Despite an initialization of the active contour that matches the desired solution, it failed to segment the considered building. It can be seen in this case that the behavior of the active contour is as expected on the image's low-contrast areas: the contour "leaks" and then propagates itself throughout the image, encompassing the zones that have the same statistical properties. These results also illustrate a more selective segmentation of the Chan and Vese model due to its restriction to the partitioning of the image's constant zones. The Bayesian model, which is more permissive, segments a homogeneous and more extensive zone of the image.

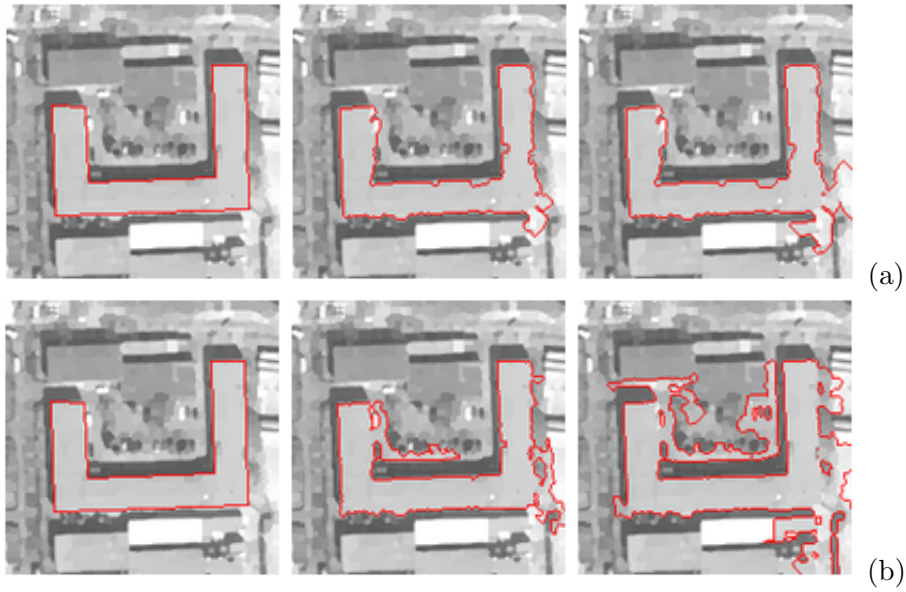


Figure 5.4: Segmentation without insertion of prior shape information and $\xi_0 = \mathbf{0}$. (a) Chan and Vese model (b) Bayesian model.

The experiments shown in figure 5.5 were conducted under the same initial conditions and the same models as in figure 5.4. However, the incorporation of prior shapes is taken into account. In this case we can see that the results are similar whatever the model used: a satisfactory segmentation is obtained and the artifacts due to the low-contrast zones are overcome. It was found necessary to opt for a different shape constraint weight λ according to the attachment to data model used. The empirical choice of λ is a trade-off between the information contained in the image and the prior shape constraint. If λ is too low, the active contour is free to be influenced by the attachment to data term which is likely to be corrupted by the noise, the occlusions or other urban artifacts. Inversely, too great a weight would be equivalent to minimizing the contribution of the image which leads the contour towards the building that we are attempting to segment. The active contour would therefore lose in flexibility and would be more inclined to reach the local minima of the functional to be minimized (we

will examine this point in section 5.5.4). If a weight $\lambda = 1$ is suitable for the Chan and Vese model, it is insufficient for the Bayesian model. Indeed, this latter model differs from that of Chan and Vese by the normalization of the image terms by the variance inside and outside the active contour. Whereas the outside of the active contour in the image is more “disordered” than the homogeneous building that we are seeking to segment, we have in a general way: $\sigma_{in}^2(t) < \sigma_{out}^2(t)$. When the contour converges on the homogeneous target we have $\sigma_{in}^2(t) \ll 1$: the division by this term then drastically increases the attachment to data term inside $\mathbf{C}(t)$. To counter this effect, we choose a higher value of λ ($\lambda = 10$).

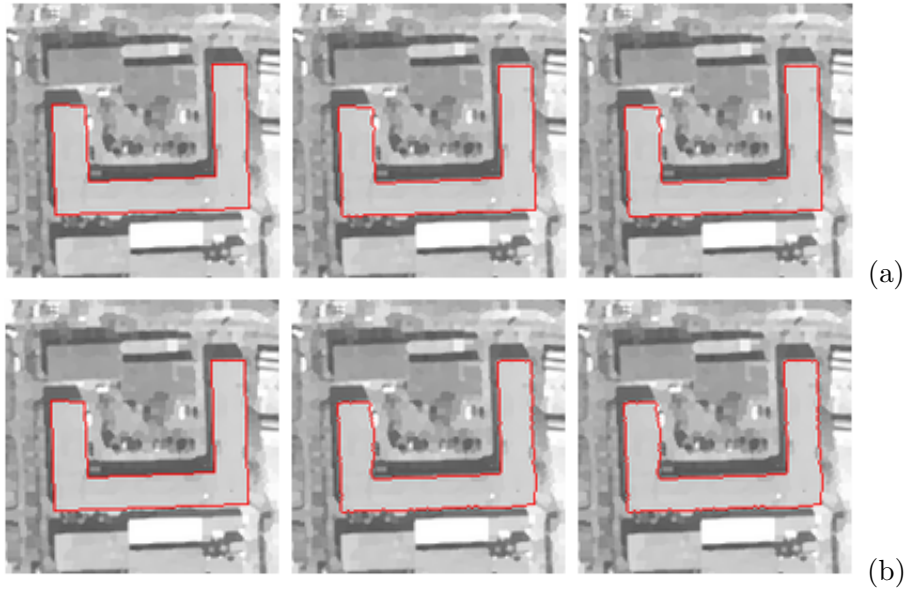


Figure 5.5: Segmentation with insertion of prior shape information: (a) Chan and Vese model and $\lambda = 1$ (b) Bayesian model and $\lambda = 10$.

The need to use the knowledge of the prior shape is once again illustrated in figure 5.6 with the active contours based on the diffused image gradient information (GVF). Unlike with the experiments in figures 5.4 and 5.5, the initial active contour is not exactly superimposed over the object to be matched in the image: the initial contour has undergone a parameter transformation $\xi_0 = (s_0 = 1.1, \theta_0 = 0.1, \mu_{x,0} = -4, \mu_{y,0} = 1)$. Figure 5.6.a shows that the segmentation of the building fails in the absence of a shape constraint. As we have already noted with the region-based active contours, the deformable model without constraints is too free and becomes sensitive to the artifacts of the urban images. In the case of figure 5.6.a, the great number of details produces numerous zones of high gradients which attract the active contour and distance it from the boundaries of the building. Figure 5.6.b shows the improvement brought by the shape constraint: the building in the image is correctly segmented despite an initialization that is far from the initial active contour. The invariance by direct plane similarity of the shape constraint makes it possible to overcome the initial superposing errors thanks to the dynamic estimate of the T_{sim} parameters.

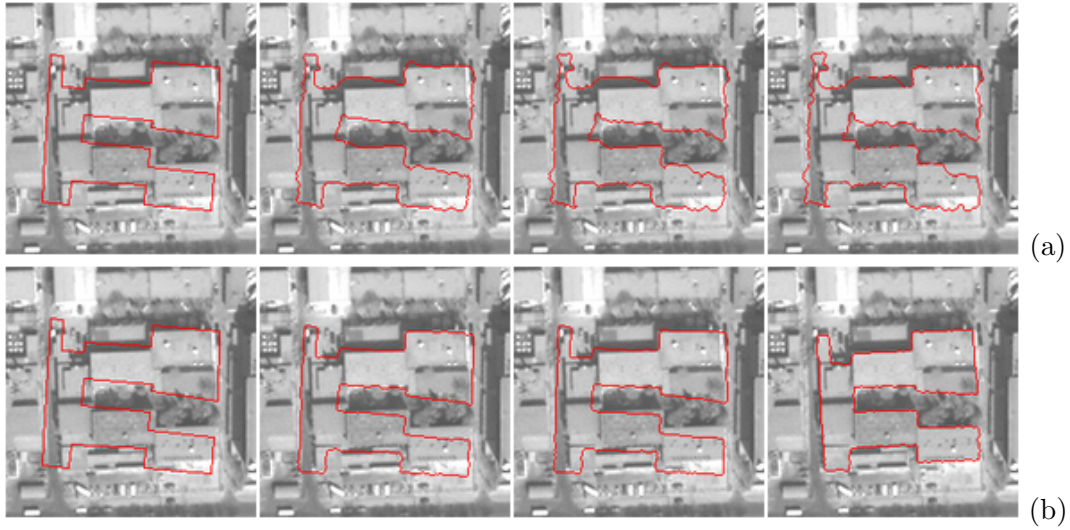


Figure 5.6: Segmentation with an edge-based active contour (GVF) and $\xi_0 = (s_0 = 1.1, \theta_0 = 0.1, \mu_{x,0} = -4, \mu_{y,0} = 1)$: (a) result without any prior shape constraint. (b) result with prior shape constraint ($\lambda = 5$).

5.5.4 Sensitivity to the initialization of the shape constrained active contours

The evolution of the active contours minimizes an energy functional by gradient descent. This optimization is global and not local. Thus, all the possible configurations of the contour in the image will not be tested so as to minimize the energy functional. On the contrary, the active contour must be initially close to the final solution so that it can converge on it. So, a non-null overlap between the initial contour and the target in the image is essential. In this section, we examine the sensitivity of the shape constrained active contours with respect to the initialization in order to assess the robustness of the algorithms and their capacity for solving the map-to-image matching problem. Nevertheless, we only examine a single class of initial matching error: the one that can be modeled by direct plane similarity.

5.5.4.1 Comparison of the region-based models

Figures 5.7 to 5.12 illustrate the sensitivity of the Chan and Vese and Bayesian models with respect to the initialization of the active contour. Figure 5.7 shows that the Chan and Vese model is not sensitive to a weak translation on the three examples illustrated. Figures 5.8.a and 5.8.c show that if the matching of the rectangular buildings does not seem to pose any problem with an initial rotation of 0.3 rad, we can see, however, that that of the U-shaped building has failed (figure 5.8.b). There are two reasons for this:

- * The first one is relative to the images that we use. In a dense urban environment, a building considered does not stand out singularly from its environment. Thus, if the initialization is too distant, a convergence on the desired object may be

highly random. This convergence is made even more hazardous for the Chan and Vese model where the only statistical quantity examined is the mean. It is very possible that during the iterative process, $c_{in}(\phi(\mathbf{x}, t)) \simeq c_{out}(\phi(\mathbf{x}, t))$ which would result in the active contour being immobilized, preventing the matching (local minimum).

- * The second reason is inherent to the shape constraint term. It is possible that the attachment to data term makes the active contour tend towards a state that does not belong to the space of shapes authorized by the shape constraint energy. So, even if the object represented in the image is the solution minimizing the energy functional, it is probable that it will cost more for the shape constraint term to come close to it than the gain made by the attachment to data energy. We then reach a *status quo* between these two antagonistic forces, making the active contour converge on a local minimum of the energy functional.

The Chan and Vese model seems to be very sensitive to a variation of the scaling factor. This is a convincing manifestation of the lack of radiometric discrimination of the building to be matched in the image in the sense of the mean (figures 5.9.a and 5.9.b).

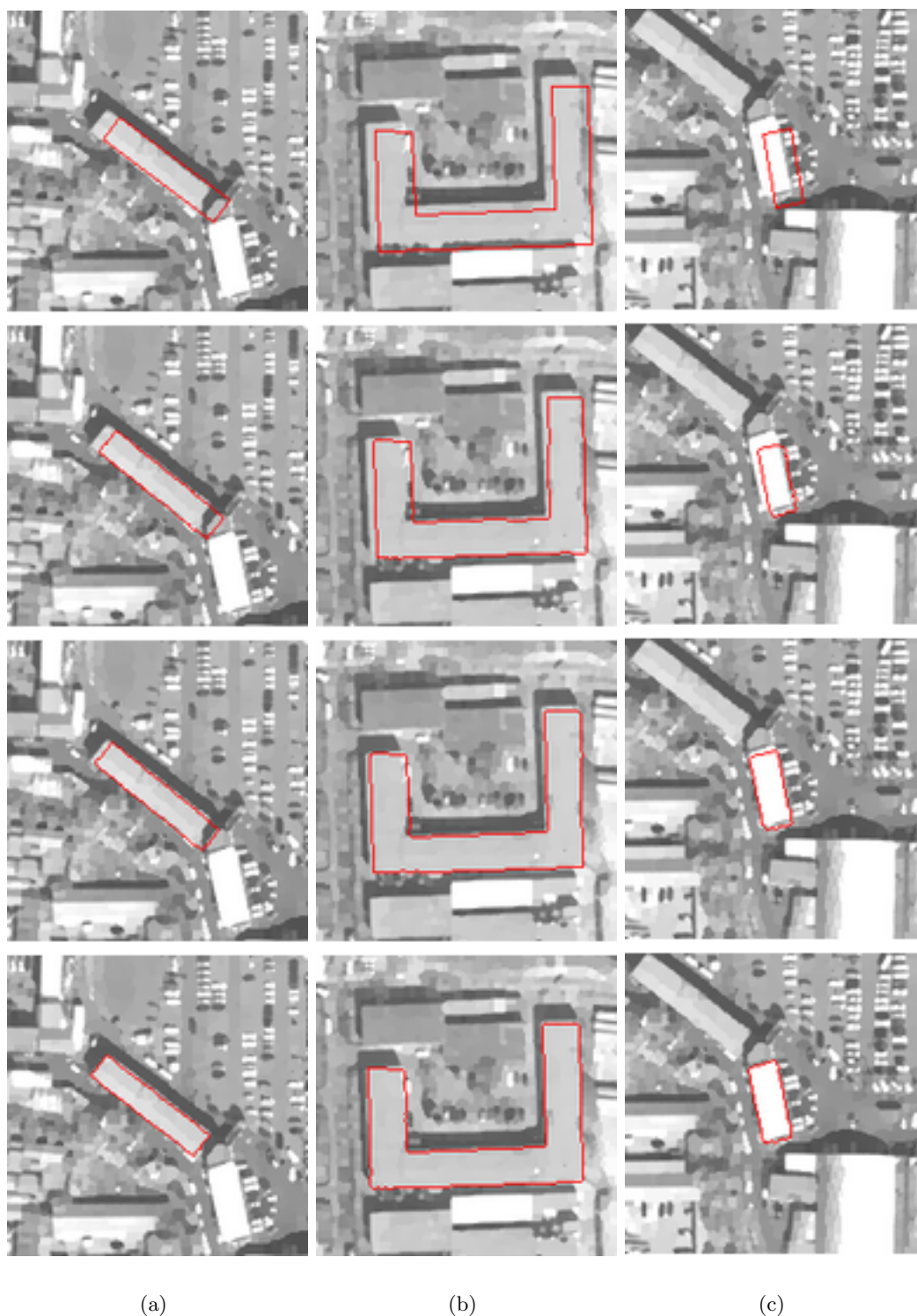


Figure 5.7: Sensitivity to translation: segmentation with the Chan and Vese model incorporating a shape constraint. The initial active contour of experiments (a), (b) and (c) has been transformed by a direct plane similarity of parameters $\xi_0 = (s_0 = 1.0, \theta_0 = 0.0, \mu_{x,0} = 5, \mu_{y,0} = 5)$.

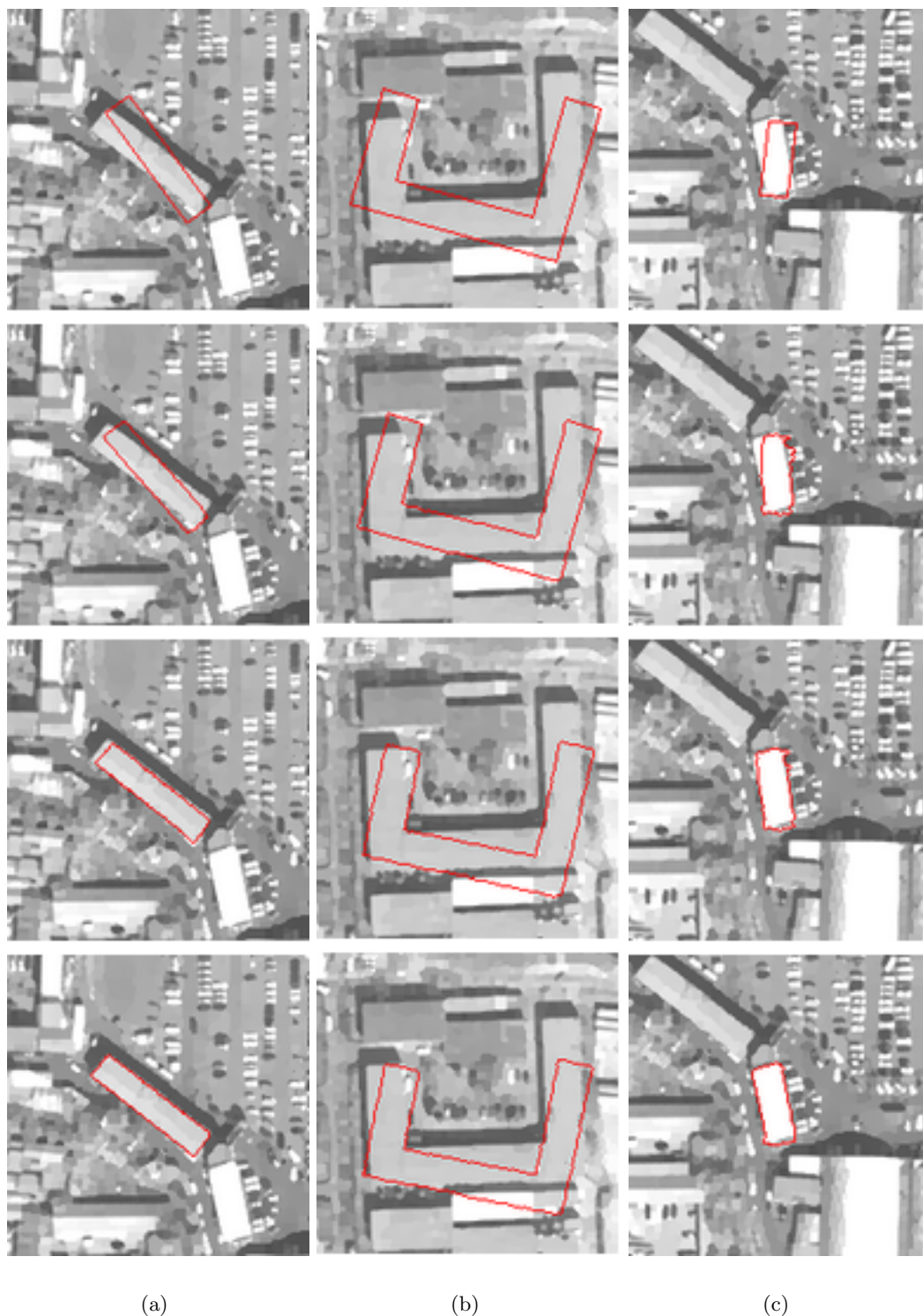


Figure 5.8: Sensitivity to rotation: segmentation with the Chan and Vese model incorporating a shape constraint. The initial active contour of experiments (a), (b) and (c) has been transformed by a direct plane similarity of parameters $\xi_0 = (s_0 = 1.0, \theta_0 = 0.3, \mu_{x,0} = 0, \mu_{y,0} = 0)$.

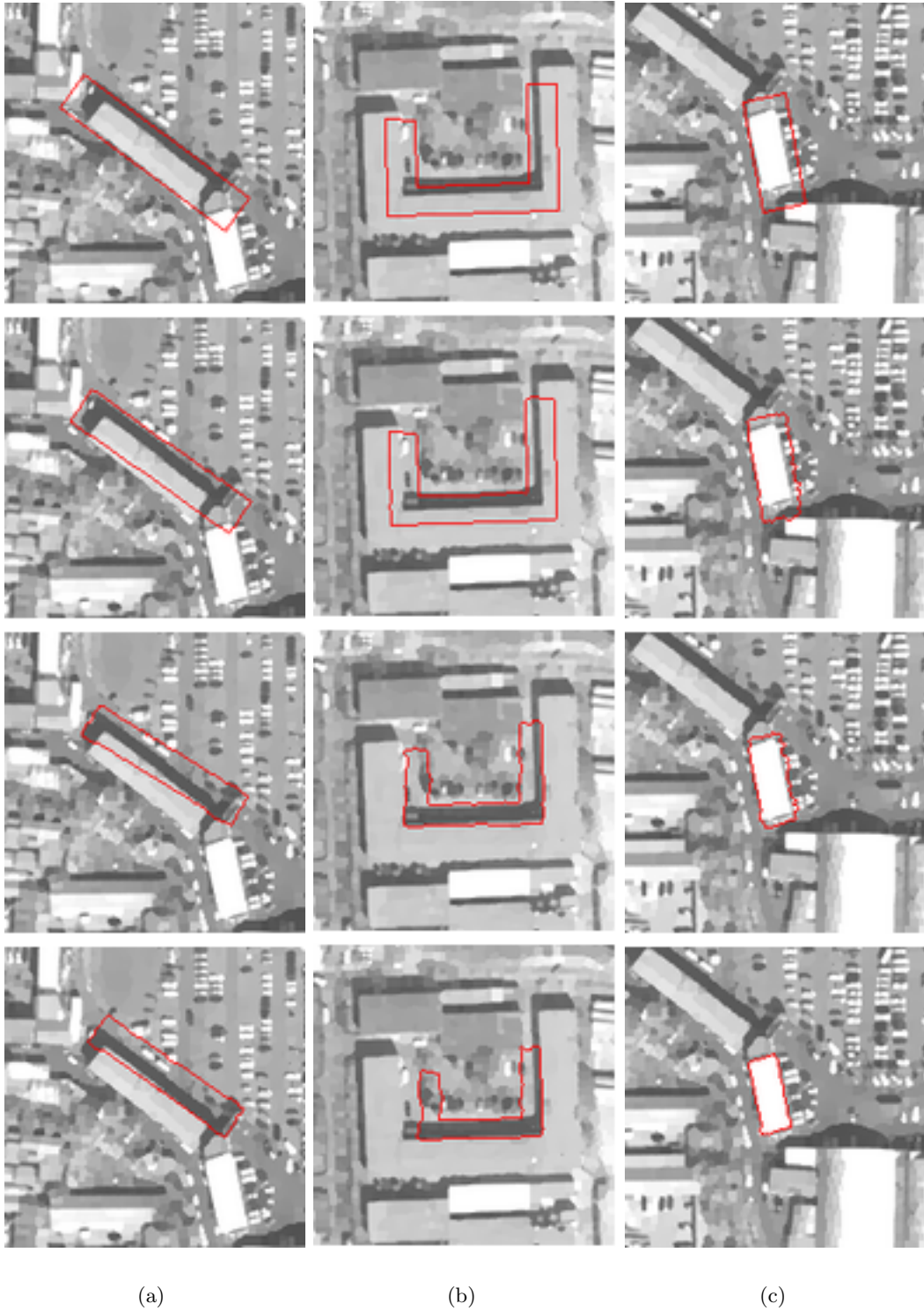


Figure 5.9: Sensitivity to the scaling factor: segmentation with the Chan and Vese model incorporating a shape constraint.

- (a) $\xi_0 = (s_0 = 1.5, \theta_0 = 0.0, \mu_{x,0} = 0, \mu_{y,0} = 0)$;
 (b) $\xi_0 = (s_0 = 0.8, \theta_0 = 0.0, \mu_{x,0} = 0, \mu_{y,0} = 0)$;
 (c) $\xi_0 = (s_0 = 1.5, \theta_0 = 0.0, \mu_{x,0} = 0, \mu_{y,0} = 0)$.

The results obtained with the Bayesian model in figures 5.10.b and 5.10.c are similar to those obtained with the Chan and Vese model concerning the sensitivity with respect to translation. However, we note that the segmentation of the first rectangular building fails (figure 5.10.a). This is inherent to the Bayesian model which is more permissive and authorizes the segmentation of objects of non-uniform intensity. This flexibility of the model associated with a shape constraint that is too great ($\lambda = 10$) makes it possible to explain the erroneous result obtained. We will see in section 5.4 how to solve this problem with a variable weight λ . As for the rotation (figure 5.11), it can be seen that the matching of the U-shaped building has succeeded (figure 5.11.b), just as that of the rectangular buildings has, whereas the Chan and Vese model has failed (figure 5.8.b). As we have mentioned earlier, the normalization by the variance of the image inside the active contour makes it possible to render the attachment to data term preponderant in order to overcome the shape constraint term which had tended to create local minima. Like with the Chan and Vese model, the sensitivity test with respect to the scaling factor fails for the experiments in the experiments in figures 5.12.a and 5.12.b. This is due to a poor initialization and a lack of discrimination of the building in the image. On the contrary, since the light-colored building in the third experiment is singularly different from the rest of the scene, the segmentation succeeds whatever the model used (figure 5.12.c).

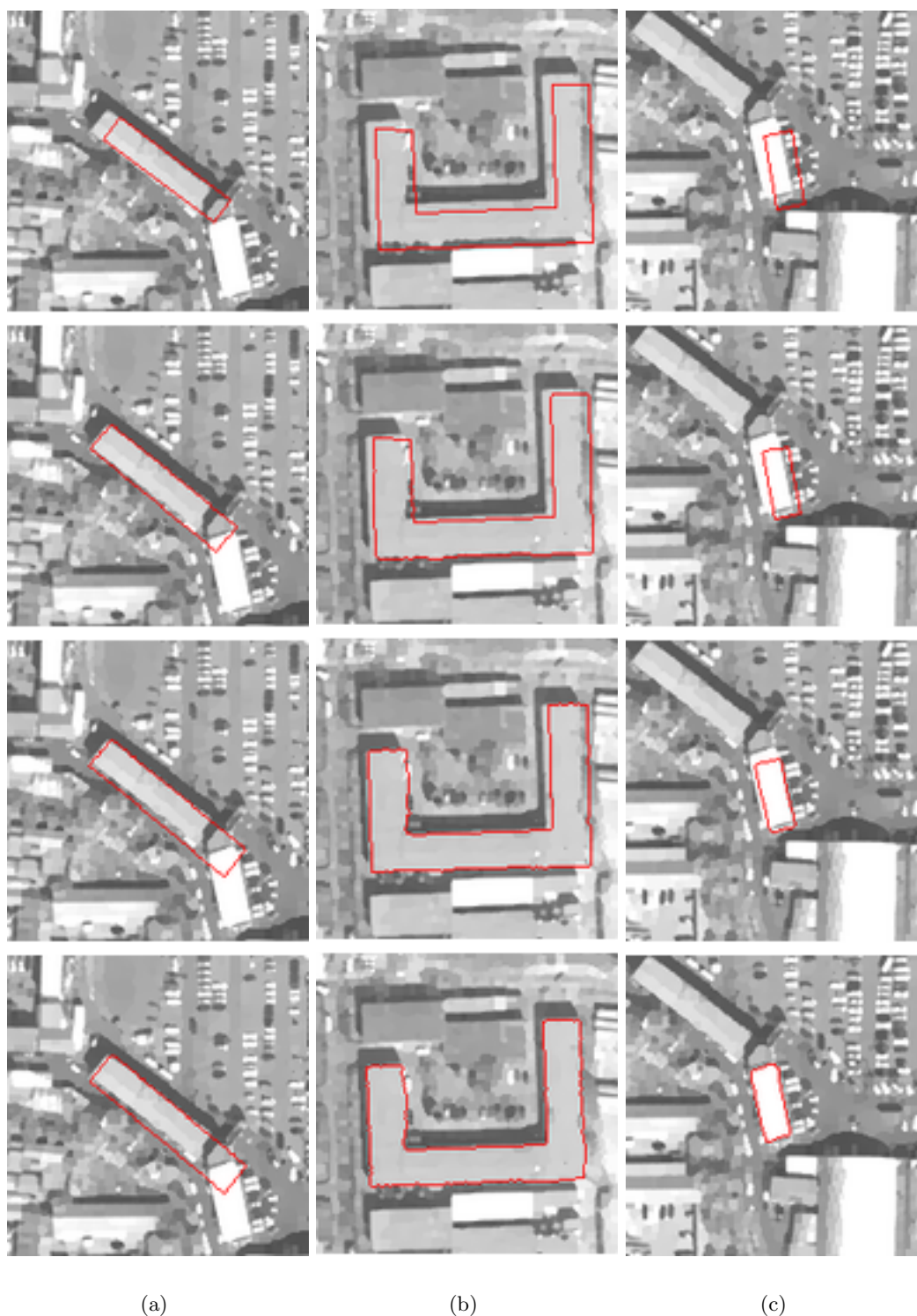


Figure 5.10: Sensitivity to translation: segmentation with the Bayesian model and shape constraint. The initial active contour of experiments (a), (b) and (c) has been transformed by a direct plane similarity of parameters $\xi_0 = (s_0 = 1.0, \theta_0 = 0.0, \mu_{x,0} = 5, \mu_{y,0} = 5)$.

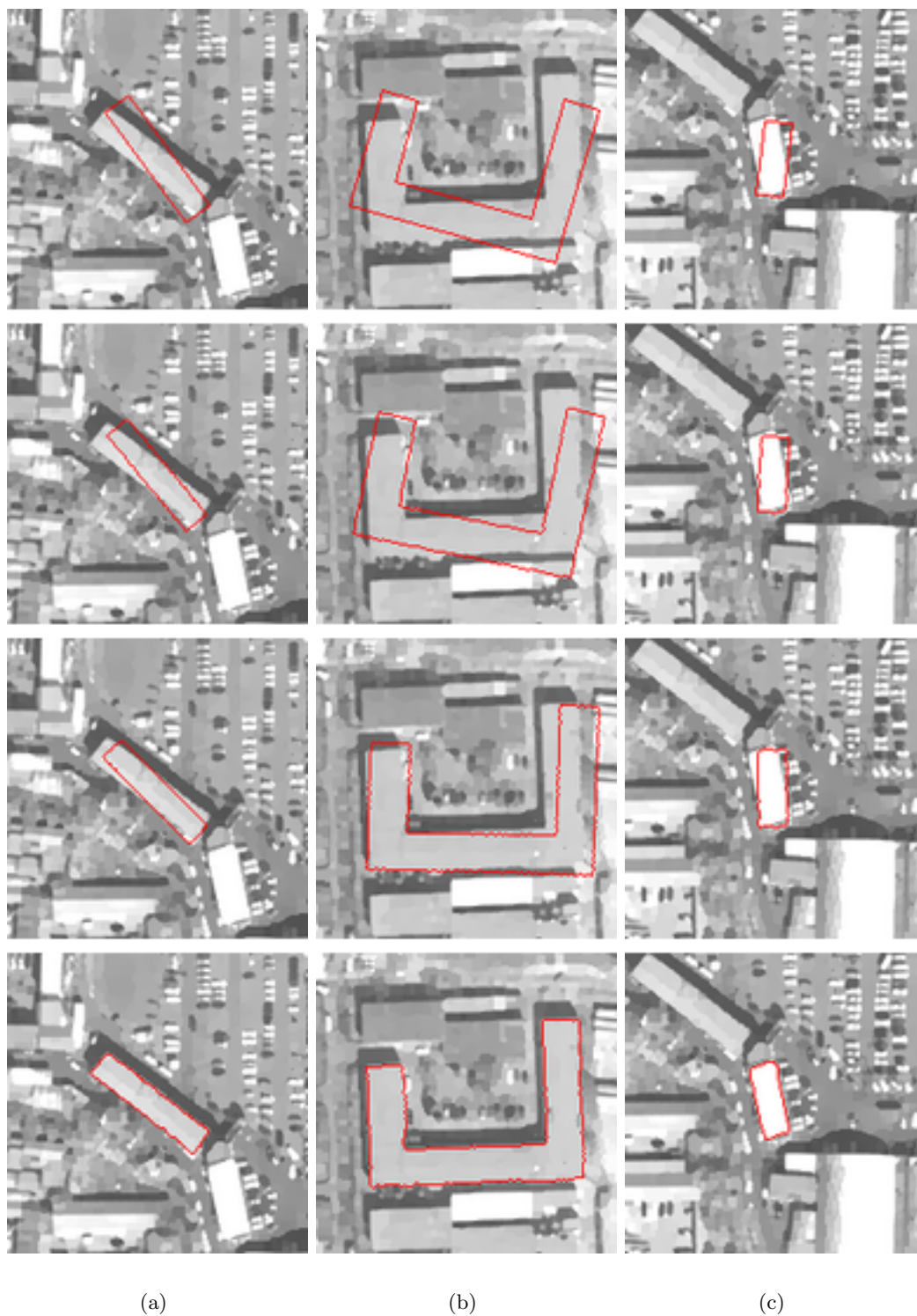


Figure 5.11: Sensitivity to rotation: segmentation with the Bayesian model and shape constraint. The initial active contour of experiments (a), (b) and (c) has been transformed by a direct plane similarity of parameters $\xi_0 = (s_0 = 1.0, \theta_0 = 0.3, \mu_{x,0} = 5, \mu_{y,0} = 5)$.

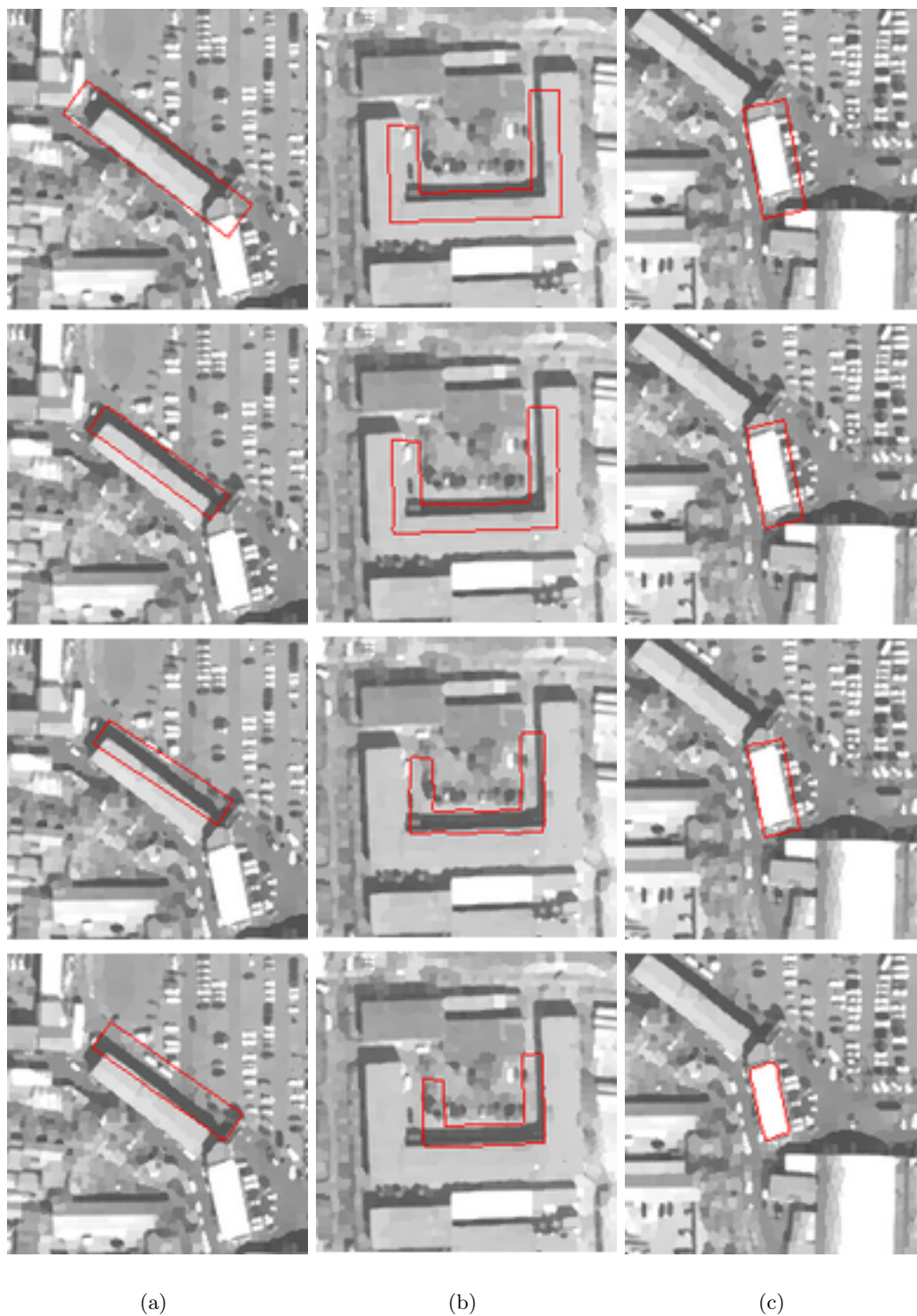


Figure 5.12: Sensitivity to the scaling factor: segmentation with the Bayesian model and the shape constraint.

- (a) $\xi_0 = (s_0 = 1.5, \theta_0 = 0.0, \mu_{x,0} = 0, \mu_{y,0} = 0)$
 (b) $\xi_0 = (s_0 = 0.8, \theta_0 = 0.0, \mu_{x,0} = 0, \mu_{y,0} = 0)$
 (c) $\xi_0 = (s_0 = 1.5, \theta_0 = 0.0, \mu_{x,0} = 0, \mu_{y,0} = 0)$.

5.5.4.2 Edge-based model

Figure 5.13 shows the matching result with the active contour based on the image's GVF. The illustrated matchings in this figure are satisfactory. It should however be noted that the initialization error is small. With a more distant initialization, the constrained segmentation fails (figure 5.14). These experiments demonstrate that even with the spatial distribution of the gradient information, the active contour remains sensitive to the initialization. The reason is the profusion of details surrounding the building which easily trap the active contour in a local minimum of the energy functional.



Figure 5.13: Sensitivity to initialization: segmentation with an active contour based on the contour information (GVF) and with the shape constraint.

- (a) $\xi_0 = (s_0 = 1.1, \theta_0 = 0.1, \mu_{x,0} = -4, \mu_{y,0} = 1)$
(b) $\xi_0 = (s_0 = 1.0, \theta_0 = -0.1, \mu_{x,0} = -1, \mu_{y,0} = -2)$
(c) $\xi_0 = (s_0 = 1.0, \theta_0 = 0.1, \mu_{x,0} = 0, \mu_{y,0} = 2).$

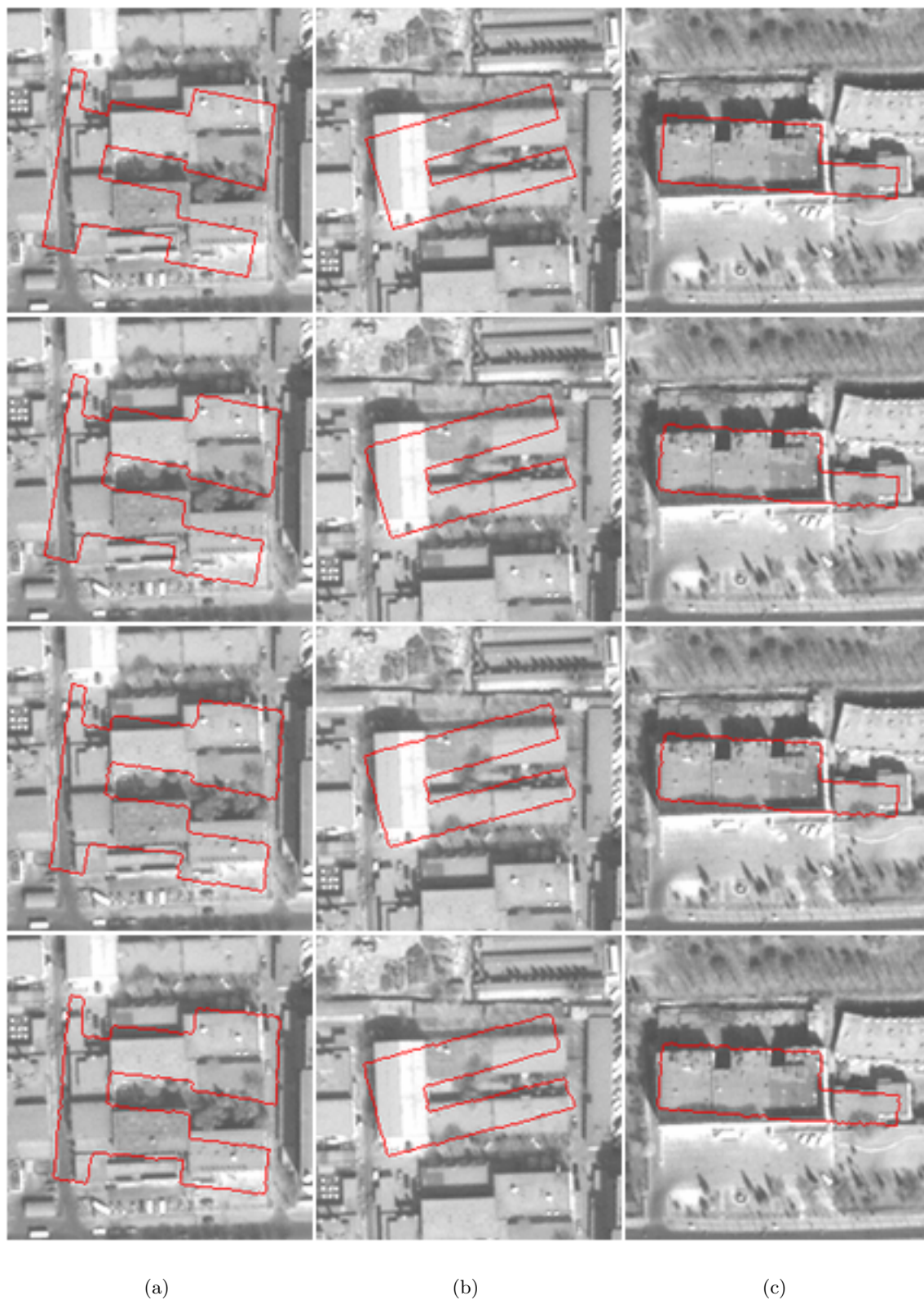


Figure 5.14: Sensitivity to a distant initialization: segmentation with an active contour based on the contour information (GVF) and with the shape constraint.

- (a) $\xi_0 = (s_0 = 1.1, \theta_0 = 0.25, \mu_{x,0} = -4, \mu_{y,0} = 1)$
 (b) $\xi_0 = (s_0 = 1.0, \theta_0 = -0.25, \mu_{x,0} = -2, \mu_{y,0} = -3)$
 (c) $\xi_0 = (s_0 = 1.0, \theta_0 = 0.1, \mu_{x,0} = -2, \mu_{y,0} = 2).$

5.5.5 Solutions for a lower sensitivity to the initialization and to the local minima

The experiments in the previous section have shown that the active contours used were sensitive to a distant initialization with respect to the object to be segmented in the image. We explain this sensitivity by the lack of discrimination of the object of interest with respect to the rest of the image and by the presence of the shape constraint which is inclined to make the contour converge on the local minima. In this section, we present the results concerning the fusion of the DSM with the satellite image and the spatio-temporal shape constraint. We will examine experimentally to what extent these two solutions resolve the problem of the local minima which prevent from an efficient matching.

5.5.5.1 Results with fusion of the exogenous data into the attachment to data term

Figure 5.15 illustrates the result obtained with the U-shaped building and the models proposed in equations (5.5) and (5.6). Despite the initialization intentionally made far from the building in the image, the matching succeeds. This shows that the good discrimination of the building in the DSM manages to lead the active contour to the target in the image. At the end of convergence, the contribution of the DSM is relaxed to give more weight to the image in which the boundaries of the building are better defined. Figure 5.16 illustrates the same experiment as figure 5.15 but without integration of the DSM. It can be seen that in this case the active contour does not manage to segment the building.

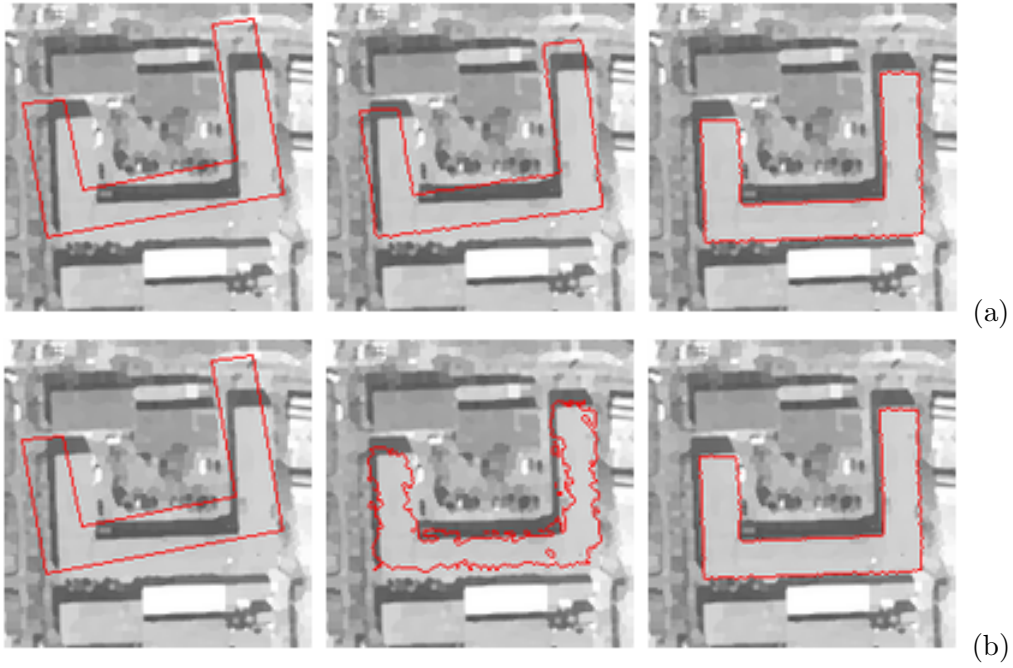


Figure 5.15: Successful matching with a distant initialization and insertion of the exogenous DSM in the energy functional. (a) Chan and Vese model, (b) Bayesian model.



Figure 5.16: Failed matching with a distant initialization and the Chan and Vese model (no incorporation of the DSM).

The comparison of figures 5.17 and 5.18 illustrates the efficiency of the proposed method. In the experiment in figure 5.18 where only the satellite image directs the active contour (Bayesian model), the matching fails. The distant initialization and the presence of shadow explain the result obtained. The shadow is an artifact that is absent from the DSM which then manages to direct the contour to the correct solution (figure 5.17).

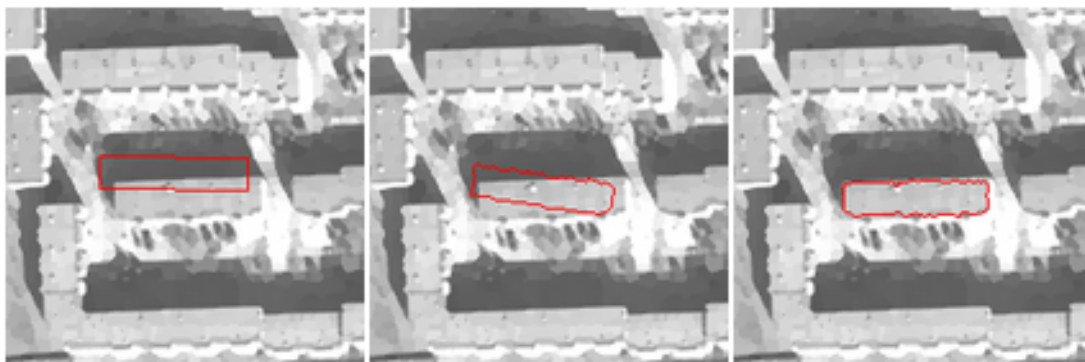
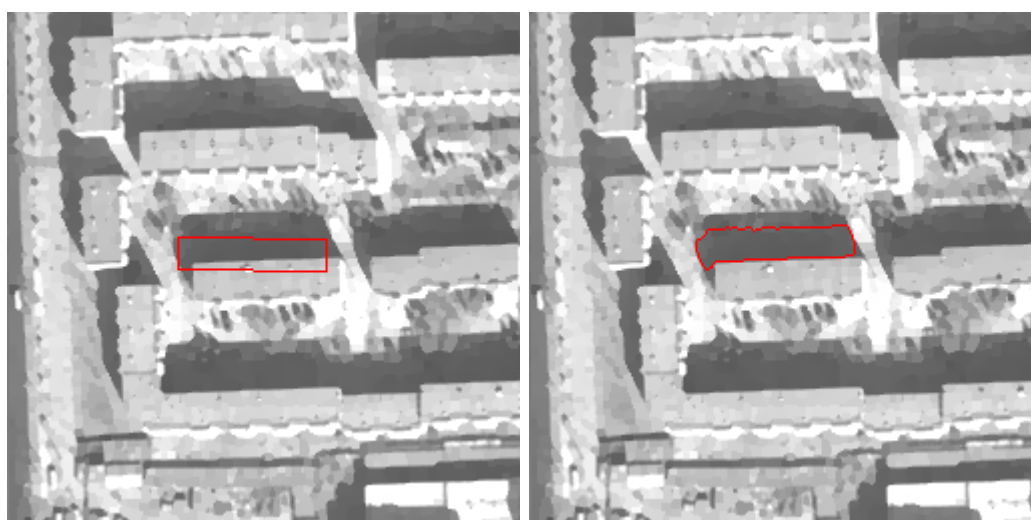


Figure 5.17: Successful matching with a distant initialization and insertion of the exogenous DSM in the energy functional (GVF model for the image, Chan and Vese model for the DSM).



(a) Etat initial

(b) Etat final: segmentation incorrecte

Figure 5.18: Failed matching with a distant initialization and the Chan and Vese model (no incorporation of the DSM).

Lastly, the experiment in figure 5.19 illustrates the complementarity of the region information from the DSM (Chan and Vese) and from the boundary extracted from the image's segments (GVF). The force derived from the DSM manages to bring the active contour close to the final solution by overcoming the artifacts likely to trap an active contour only moved by the image's gradients. The building is, in the end, successfully segmented (figure 5.19), unlike the case of figure 5.20 without incorporation of the DSM.



Figure 5.19: Successful matching with a distant initialization and insertion of the exogenous DSM in the energy functional (GVF model for the image, Chan and Vese model for the DSM).



(a) Initial state

(b) Successful matching

Figure 5.20: Failed matching with a distant initialization and the GVF model.

The last advantage of DSM fusion concerns the reduced convergence time. Table 5.3 compares the convergence times with and without merging of the DSM thanks to the calculation of their ratio. The experiments performed to establish these comparisons were carried out with buildings of figures 5.7.a and 5.7.b with different initializations. We see that the convergence times are shorter with the DSM fusion. This improvement is due to the great discrimination of the building in the DSM compared with the image. So, the joint utilization of the DSM makes it possible to surely and more efficiently lead the active contour to the target object in the image. The results in table 5.3 depend on the initialization of the active contour and of the image being analyzed. Nevertheless,

they show a real tendency to improve the efficiency.

| Experiment | ratio |
|------------------------------------|-------|
| U-shaped building, experiment 1 | 1.4 |
| Rectangular building, experiment 1 | 1.3 |
| U-shaped building, experiment 2 | 2.5 |
| Rectangular building, experiment 2 | 1.3 |

Table 5.3: Ratio between the convergence times without and with DSM merging, $ratio = \frac{t_{without\ fusion}}{t_{with\ fusion}}$.

5.5.5.2 Results with a spatio-temporal shape constraint

The experiment illustrated in figure 5.21 reproduces the experiment in figure 5.10.a with the spatio-temporal weight of the shape constraint expressed in equation (5.10). The amplitude of the constraint weight varies between $\lambda_{min} = 3$ and $\lambda_{max} = 30$ when the initial non-interaction distance is $d_0 = 7$ pixels. Whereas the matching of the rectangular building failed with a constant shape constraint weight and the Bayesian model (figure 5.10.a), it succeeds with its formulation which varies according to space and time (figure 5.21). The second image in the sequence illustrates the first phase of the algorithm: since the weight is low, the contour can undergo topological changes in order to segment the areas with similar radiometry. This segmentation is obviously limited in space. The third image in the sequence illustrates the second phase: the weight increases gradually and continuously which tends to penalize and therefore reduce the surface of the second region that appeared in the previous step. Finally, the contour converges on the correct solution.



Figure 5.21: Segmentation with a spatio-temporal variation of the shape constraint and the Bayesian model ($\lambda_{min} = 3$, $\lambda_{max} = 30$, $d_0 = 7$).

Figure 5.22 shows a successful matching with an initialization that is distant from the target in the image. The difficulty of the scene is the presence of a neighboring building (square) with radiometric properties that are similar to those of the building of interest. When a constant constraint weight is used (figure 5.23), we note that this peripheral building is also segmented, preventing a satisfactory matching. Inversely, the spatio-temporal constraint of the shape constraint provides the active contour with

sufficient flexibility to authorize a change of topology to achieve a coarse segmentation of the object (second image in the sequence in figure 5.22). When the uniform constraint is restored at a later time, the local errors in the segmentation situated in the relaxation space are eliminated and a successful matching is finally achieved.



Figure 5.22: Segmentation with a spatio-temporal variation of the shape constraint and the Bayesian model ($\lambda_{min} = 27$, $\lambda_{max} = 70$, $d_0 = 7$).



Figure 5.23: Segmentation with a constant shape constraint weight and the Bayesian model ($\lambda = 50$).

The experiments shown in figure 5.24 reproduce with a flexible constraint the experiments that failed in figure 5.14 with a constant shape constraint weight and the edge-based model (GVF). The improvement due to the constraint variation is more subtle with the GVF model. In the first case (figure 5.24.a), the adaptive weight does not make it possible to obtain a correct matching, whereas it is efficient with the other two examples. With an edge-based active contour based on the image's gradient, the spatio-temporal weight does not make it possible to achieve a successful matching with an initialization that is too distant. Indeed, these deformable models are affected by the image's local information, which is a problem when the shape constraint is relaxed. So, it is probable that the contour remains trapped during the period when the constraint weight is low, which prevents us from coming close to the building to be matched in the image.

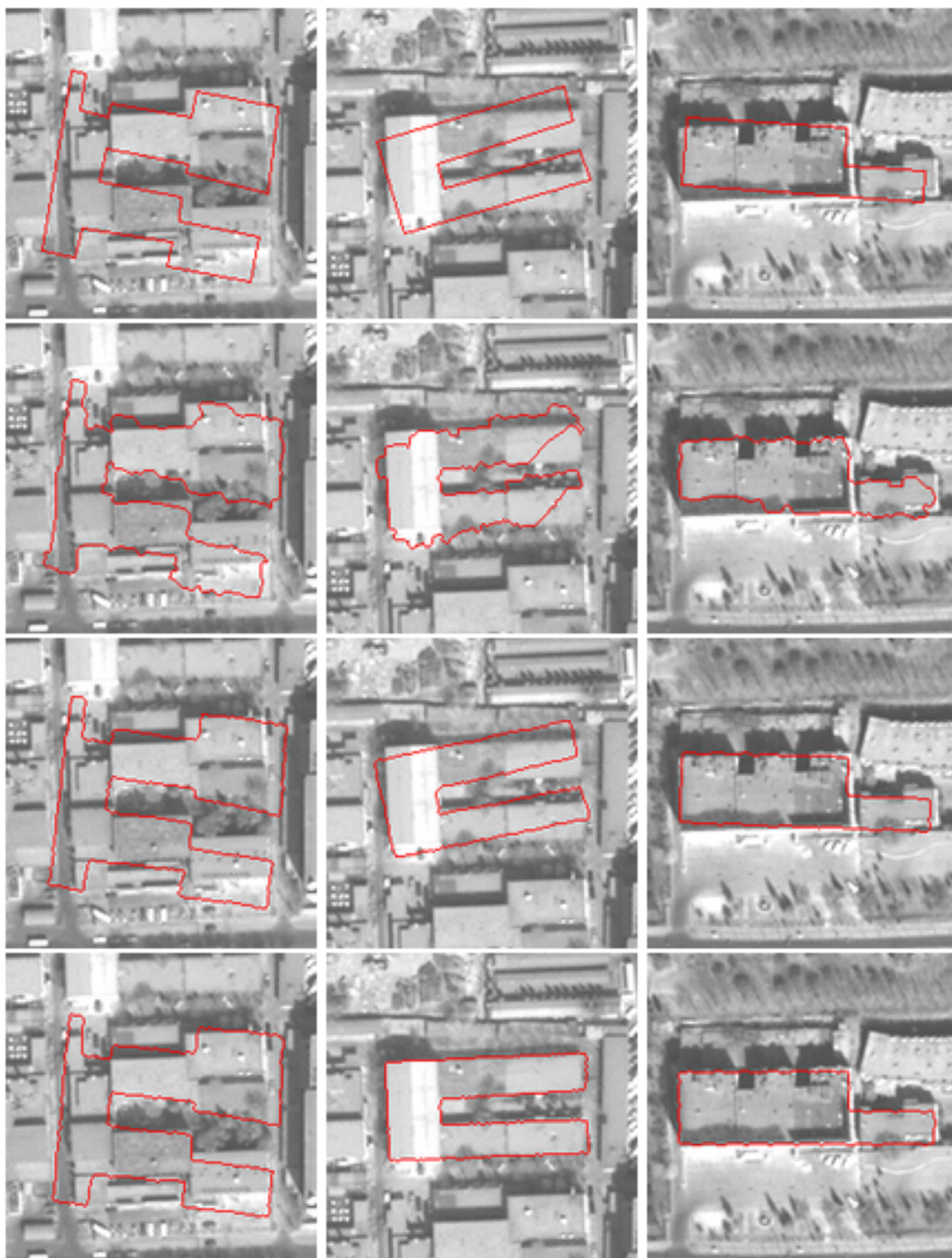


Figure 5.24: Spatio-temporal variation of lambda with the GVF model and $\lambda_{min} = 1$, $\lambda_{max} = 5$, $d_0 = 5$:

- (a) $\xi_0 = (s_0 = 1.1, \theta_0 = 0.25, \mu_{x,0} = -4, \mu_{y,0} = 1)$
- (b) $\xi_0 = (s_0 = 1.0, \theta_0 = -0.25, \mu_{x,0} = -2, \mu_{y,0} = -3)$
- (c) $\xi_0 = (s_0 = 1.0, \theta_0 = 0.1, \mu_{x,0} = -2, \mu_{y,0} = 2)$.

In figure 5.25 we propose to conduct experiments with images that do not belong to the remote sensing domain. These images present the same difficulties as the remote sensing data in an urban environment, that is to say a lack of discrimination of the object of interest with respect to its environment. The experiments with images 5.25.a and 5.25.b show the efficiency of the flexible shape constraint weight with the Bayesian model. The utilization of a constant weight prevents a successful matching: the contour's lack of flexibility quickly makes it founder in a local minimum of the energy functional (figure 5.26).

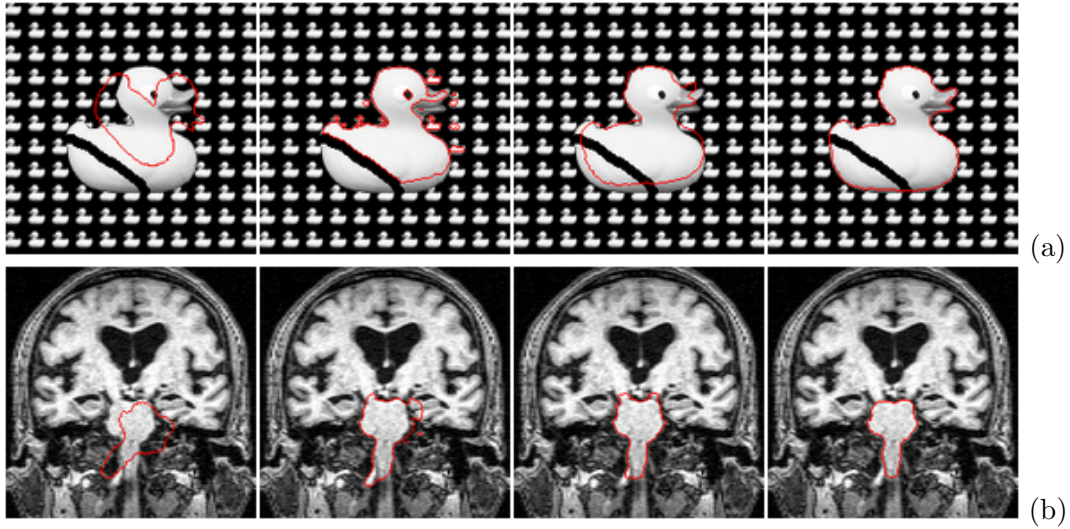


Figure 5.25: Segmentation with spatio-temporal variation of the shape constraint weight (Bayesian model): (a) $\lambda_{min} = 15$, $\lambda_{max} = 100$, $d_0 = 15$ (b) $\lambda_{min} = 50$, $\lambda_{max} = 100$, $d_0 = 10$.

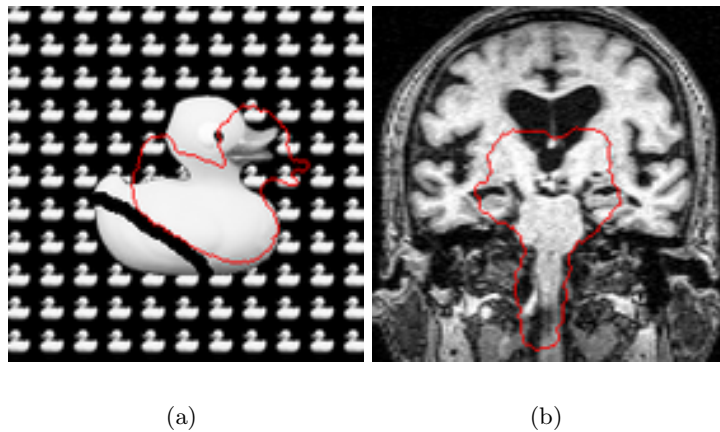


Figure 5.26: Failed matching with a constant shape constraint weight (Bayesian model): (a) $\lambda = 20$ (b) $\lambda = 70$.

Lastly, we propose experiments with different ways of making the shape constraint vary according to space and time. The experiment shown in figure 5.27.a reproduces the one in figure 5.22.a but without spatial relaxation of the constraint, only the temporal variation in amplitude of equation (5.10) is used. We note that contrary to the spatio-temporal case, the segmentation fails. The absence of relaxation in this case does not allow the active contour to be sufficiently flexible to coarsely segment the object of interest before the restoration of the uniform constraint. The utilization of a lower value λ_{min} to provide greater flexibility would fail insofar as the flexibility given would not be spatially bounded. The contour would then segment the peripheral objects that have the same statistical properties as the target. The experiment in figure 5.27.b reproduces the experiment shown in figure 5.25.b. A spatio-temporal weight is also used, but with a simpler formulation of the spatial profile:

$$\lambda_{flex}(\psi(\mathbf{x}), t) = \lambda_a(t) \begin{cases} \frac{|\psi(\mathbf{x})|}{2d(t)} & \text{if } |\psi(\mathbf{x})| < 2d(t) \\ 1 & \text{otherwise} \end{cases} \quad (5.11)$$

where ψ is the shape constraint level set, $d(t)$ and $\lambda_a(t)$ are defined in equation (5.10). The piecewise linear formulation of the spatial relaxation of equation (5.11) is certainly more intuitive than that of equation (5.10). Nevertheless, the experiment in figure 5.27.b using this type of formulation fails. The problem with this type of profile is an excessive strengthening of the shape constraint in the vicinity of the level zero of ψ . So, the constraint is not relaxed as much as with the presence of a steady point of the exponential formulation in equation (5.10). Increasing the size of the relaxation space d_0 would not be a viable solution since when the latter is too extensive, the active contour would be likely to be influenced by the surrounding artifacts and objects, preventing the coarse segmentation of the object of interest before the re-establishment of the uniform constraint. The advantage of the spatial profile of equation (5.10) is that it authorizes a great relaxation very close to the prior shape while keeping a high constraint at a greater distance. Lastly, the experiment in figure 5.27.c repeats the experiment in figure 5.22.a with a simpler formulation of the temporal profile. The temporal variations of $\lambda_a(t)$ and $d(t)$ are replaced by Heaviside step functions that are respectively increasing and decreasing. The restoration of the shape constraint is thus achieved abruptly and not gradually. In this case, figure 5.27.c shows that the segmentation fails. The reason for such a result concerns the dynamic estimation of the T_{sim} parameters enabling the invariance of the shape constraint from direct plane similarity. Whereas the relaxation space is effective, the active contour coarsely segments the object of interest in the image. The bounded spatial variations of the active contour with respect to the reference shape may slightly bias the estimation of the T_{sim} parameters. So, when the uniform constraint is restored suddenly, the active contour resembles the reference shape transformed by T_{sim} whose parameters are poorly estimated. Consequently, the active contour may be found far from the object to be segmented in the image and with a uniform and strong shape constraint, which makes it very sensitive to the local minima that we are seeking to avoid. Inversely, a gradual re-establishment of the shape constraint enables the estimator of the T_{sim} parameters (here the simplex) to make an estimation that is less and less biased whereas the active contour converges on the target (figure 5.27.c).

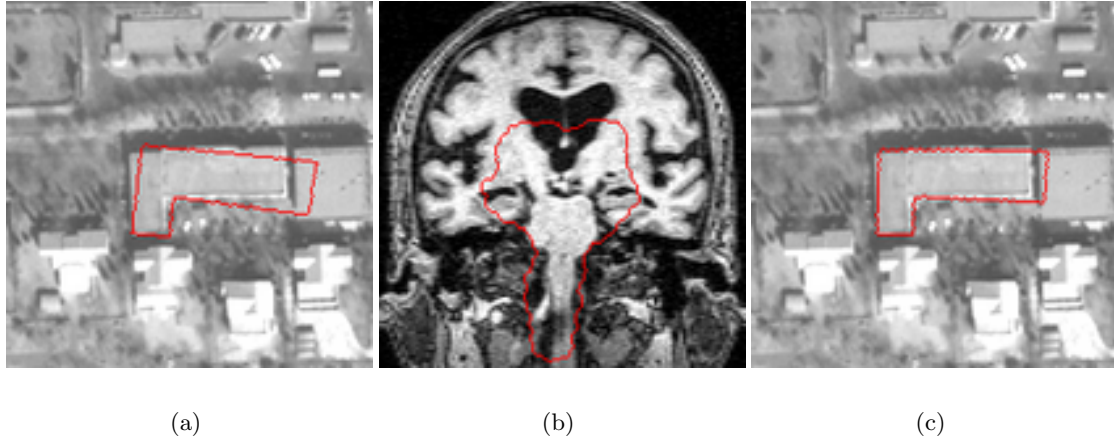


Figure 5.27: Failed matching (Bayesian model) with: (a) single temporal variation of the shape constraint with $\lambda_{min} = 30$, $\lambda_{max} = 70$ (no spatial relaxation); (b) spatio-temporal variation of the shape constraint with a spatial profile defined in equation (5.11) and $\lambda_{min} = 27$, $\lambda_{max} = 70$, $d_0 = 7$; (c) spatio-temporal variation of the shape constraint with Heaviside step functions for $\lambda_a(t)$ and $d(t)$.

Besides the possibility of ensuring a more robust convergence, the spatio-temporal flexibility of the shape constraint weight also makes it possible to significantly speed up the calculation time (tableau 5.4). Indeed, during the first phase, the weak shape constraint provides a sufficient degree of flexibility so that it efficiently comes close to the target in the image. The uniform and constant constraint which tends to slow down the contour's movement is only re-established at the end of the process to regularize the segmentation. Table 5.4 compares the convergence times without and with a spatio-temporal variation of the shape constraint using the calculation of their ratio. Like for the case of fusion of exogenous data, the experiments performed to establish these comparisons were carried out with the buildings in figure 5.7.a and 5.7.b with different initializations. Like with the results with DSM fusion, the tendency shown is an improvement in efficiency, which also depends on the initialization and on the scene being analyzed. .

| Experiment | ratio |
|------------------------------------|-------|
| U-shaped building, experiment 1 | 1.8 |
| Rectangular building, experiment 1 | 3.1 |
| U-shaped building, experiment 2 | 4.3 |
| Rectangular building, experiment 2 | 2.0 |
| U-shaped building, experiment 3 | 2 |
| Rectangular building, experiment 3 | 1.3 |

Table 5.4: Ratio between the convergence times without and with spatio-temporal shape constraint: $ratio = \frac{t_{uniform\ constraint}}{t_{spatio-temporal\ constraint}}$.

5.6 Limits of the model with shape constraint

Digital maps are often tainted with errors because they are the result of manual processing, which is in fact imperfect and subjective. If these errors can be modeled by composition of a translation, rotation and scaling, our method will be capable of overcoming this since the shape constraint energy term is invariant from direct plane similarity. Usually the errors observed on the map are local and cannot be modeled and corrected by a global transformation (see section 2.3 detailing the errors of $U_{interpretation}$). It is therefore possible that in places, the cartographic object and its homologous representation in the image may be incoherent without obeying any particular rules (figure 5.28).



Figure 5.28: Superimposition of the ground truth (grayed) and a locally erroneous cartographic polygon.

The cartographic errors will have an impact and will make themselves felt on the results of the segmentation as illustrated in figures 5.29 and 5.31. In figure 5.28 we can see that a correction involving a rotation would improve the match between the map and the ground truth made by hand. However, the local inconsistencies remain: the central bar of the U-shape is not wide enough on the map, the left-hand bar is not long enough either, whereas the size of the right-hand bar seems to be correct even if it appears to be too wide. The result of the matching with a locally erroneous map is then a trade-off between the image and the corrupted shape constraint. Figure 5.29 illustrates the example of the U-shaped building with the Bayesian model. Figures 5.30 and 5.31 illustrate the result with the GVF-based model without errors on the map (figure 5.30) and with local shape errors (figure 5.31). The presence of local errors in the map is a problem since we will use them to assist and constrain the matching process. In the next chapter we will propose to overcome this obstacle by making the incorporation of the shape constraint more flexible.



Figure 5.29: Final convergence of the active contour with an erroneous prior shape.



(a) Map-to-image initial superimposition

(b) After matching

Figure 5.30: Matching by GVF-based active contour and with a non-erroneous prior shape.



(a) Map-to-image initial superimposition

(b) After matching

Figure 5.31: Matching by GVF-based active contour and with an erroneous prior shape derived from the map.

5.7 Conclusion

The goal of this chapter was to show the contribution made by the insertion of the shape constraint for the fine map-to-image matching by active contours. The proposed matching technique makes it possible to match individually each building symbolized on the map on its homologous representation in a high-resolution panchromatic satellite image. The end-purpose is to reduce the exogenous map-to-image variabilities in order to improve the consistency between these two representations enabling a more reliable later detection of non-change.

Experimental results enabled to measure the contribution of the prior shape derived from the map in order to compensate for the difficulties inherent to urban images. Three types of active contours were used and they showed a certain sensitivity to initialization and to the local minima of the minimized energy functional. We have proposed two solutions to make the deformable models more robust. The first one concerns the injection of an orthoscopic DSM in the attachment to data energy and allows a better discrimination of the constructed layer with respect to the ground in the image. The force derived from the data in the DSM then directs the active contour more surely and efficiently towards the building in the image. The second contribution consists of spatially and temporally relaxing the shape constraint during the convergence process. The spatially bounded flexibility given to the active contour authorizes an approximate segmentation of the object in the image before being specifically regularized by the restoration of a uniform constraint. Experimental results have shown the efficiency of these two approaches with a reduction of the computational times. An alternative to

the gradient descent has also been proposed in order to ensure the invariance of the shape constraint from global transformations. Our alternative based on the simplex technique has been found to be more robust and efficient in terms of calculation times than the conventional descent method.

Lastly, we have assumed that the prior shape used as a constraint represented exactly the object to be matched in the image. We have not yet looked at the problem of possible disparities between the overall shape of the building in the image and that of the cartographic object. The presence of such variabilities is frequent however, and it is due either to local shape errors on the map or to partial changes that have been made to the constructions. In order to process the map-to-image matching with such artifacts, new shape constraint energies must be formulated.

Chapter 6

Shape constraint authorizing parallel variations

6.1 Introduction

It is possible that cartographic objects may contain *local* shape incoherencies with respect to the reality of the image. In the previous chapter we have illustrated the presence of these errors which prevent a totally satisfactory matching with the building represented in the satellite image. These inconsistencies are also due to local and partial changes that have appeared at the level of the building in the image. In both cases, the parts of the active contour, wrongly constrained by the prior shape, do not reach the boundaries of the object seen in the image because of the rigid way in which the shape constraint, which is only invariant from a *global* transformation, was incorporated. The problem of local deviations with respect to a prior shape has essentially been dealt with from a statistical viewpoint thanks to the contribution of learning samples [113, 28, 66, 104, 92, 102, 30]. We are proposing another, non-heuristic, solution which consists of authorizing certain discrepancies of the active contour with respect to the reference shape. The class of authorized deviations is the class of movements constrained by the parallelism with respect to the prior shape segments and corresponds to the local map-to-image inconsistencies most frequently found in cartographic data. To allow this type of deformation, we propose a new shape constraint energy. It is the sum of a linear energy comparing the normals of the active contour and of the prior shape, and a quadratic energy imposing generic rectilinearity and orthogonality constraints on the active contour. Preliminary but nevertheless encouraging experimental results are presented.

6.2 Shape constraint energy authorizing parallel variations

6.2.1 Model

The way of introducing the notion of parallelism into the shape constraint energy functional consists of imposing the alignment of the normals of the active contour and of

the reference shape. In terms of level sets, the normals are $\nabla\phi(\mathbf{x})$ and $\nabla\psi(\mathbf{x})$ ¹². The new energy term formulated on the basis of equation (4.54) is then:

$$J_{Shape}(\phi, \psi) = \lambda_{para} \int_{\Omega} |\nabla\phi(\mathbf{x}) - \nabla\psi(\mathbf{x})|^2 (H(\phi(\mathbf{x})) - H(\psi(\mathbf{x})))^2 d\mathbf{x} \quad (6.1)$$

The evolution equation associated with a functional $J = J_{image} + J_{Shape}$ requires the calculation of the gradient J_{Shape} with respect to ϕ . The detail of the calculations is given in appendix E, and results in:

$$\begin{aligned} \phi_t(\mathbf{x}, t) = & -\frac{\partial J_{image}}{\partial \phi} \\ & + 4\lambda_{para}\delta_{\alpha}(\psi(\mathbf{x}, t)) (\langle \nabla\phi(\mathbf{x}), \nabla\psi(\mathbf{x}) \rangle - 1) [H_a(\phi(\mathbf{x}, t)) - H_a(\psi(\mathbf{x}, t))] \\ & + 2\lambda_{para}(\Delta\phi(\mathbf{x}) - \Delta\psi(\mathbf{x})) [H_a(\phi(\mathbf{x}, t)) - H_a(\psi(\mathbf{x}, t))]^2 \end{aligned} \quad (6.2)$$

The last two terms of equation (6.2) are the result of the insertion of the new energy term. The first one tends towards zero when the normal to the active contour has the same orientation and direction as that of the prior shape. However, this term has a very local effect since it is weighted by the Dirac distribution which depends here on ψ . So, when the active contour will be far from the reference shape (beyond some pixels) this term will not have any influence. The second term is not local, and it imposes a constraint on the curvature of the active contour with respect to that of the prior shape. It is this term that will mainly contribute to imposing the parallelism constraint. The proposed parallelism constraint applies to the level lines of the prior shape ψ , however these are curved far from the geometric singularities (corners) of the level 0 of the prior shape. This will result in the active contour being rounded off when it moves away from the reference shape.

6.3 Insertion of generic prior knowledge by quadratic models

In order to resolve the unwelcome rounding effect at the level of the “corner” type geometric singularities, we will introduce a corrective energy that aims to favor the formation of corners at 90° and rectilinear segments. We thus introduce a generic shape constraint in addition to the specific one represented by the prior shape. The quadratic energies described in section 4.2.3 have been adopted to correct the rounding effect of the active contour.

¹We assume that the condition $|\nabla\phi| = 1$ is always verified, which is true in practice if we reinitialize the ϕ function at each iteration of the evolution equation. The condition $|\nabla\psi_0| = 1$ is always verified since the prior shape is not updated. So $|\nabla\psi_0| = |\nabla\psi| = 1$ since the level sets ψ_0 and ψ only differ by an isometric transformation which, by definition, does not affect the distance function.

²The local deformations forced by the parallelism are likely to be inhibited by the invariance by the direct plane similarity (scale) expressed in equation (4.54). Consequently, we exclude any invariance from global transformation in this chapter. If we use the notation of the previous chapters we then have $\psi = \psi_0 \circ T_{sim} = \psi_0$ where T_{sim} is the identity.

6.3.1 Corrective quadratic energy

6.3.1.1 Model

We propose to add the following corrective quadratic energy to (6.1):

$$J_{correction}(\mathbf{C}) = \oint \oint ds ds' (\sin 2\theta)^2 \Psi(|\mathbf{C}(s) - \mathbf{C}(s')|) \quad (6.3)$$

where θ is the angle between the tangents to the contour³ \mathbf{C} at the arc length points s and s' . Ψ is a decreasing function asymptotically tending towards 0 and defines the interaction potential between $\mathbf{C}(s)$ and $\mathbf{C}(s')$. It should be noted that the integration domain of $J_{correction}$ is this time the contour itself and not Ω . This type of energy will make it possible to favor the formation of rectilinear parts and corners within the active contour. The imposition of orthogonality constraints has already been proposed by H. Oriot in [83] in the framework of building segmentation by snakes. We took our inspiration from this idea to propose the formulation given in equation (6.3). However, contrary to the works of H. Oriot which do not use prior shapes, we propose to formulate a quadratic and not a linear energy criterion; furthermore we represent the active contour implicitly.

By noting that $\sin(2\theta) = 2 \sin(\theta) \cos(\theta)$, the correction energy can be written according to:

$$J_{correction}(\mathbf{C}) = 4 \oint \oint ds ds' \{\cos^2(\theta) - \cos^4(\theta)\} \Psi(|\mathbf{C}(s) - \mathbf{C}(s')|) \quad (6.4)$$

This energy is minimal for $\theta = 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}$:

- * If $\theta = 0$. This implies that the tangents are aligned in the interaction neighborhood defined by the profile Ψ . The rectilinearity of the objects extracted by the active contour will thus be strengthened. This property is desired since the buildings very often exhibit rectilinear roof edges (highly polygonal object).
- * If $\theta = \frac{\pi}{2} \text{ modulo } \pi$. This case represents the objective aimed for by the formulation of the corrective energy: the corners at 90° are favored to counteract the rounding effect.
- * If $\theta = \pi$. In this case the tangents are anti-parallel. The interaction neighborhood defined by Ψ has a spatial extent of the order of some pixels before tending towards 0. In the case of buildings, anti-parallel tangents occur when the distance $|\mathbf{C}(s) - \mathbf{C}(s')|$ is great and where the function Ψ is already quasi null. So, the effect of opposing tangents is negligible in our case. This case would occur if the object to be matched has fine or labyrinthine structures, which is the case of roads imaged on a small scale, but not that of buildings at very high resolution.

Before deducing the quadratic force from the derivation of $J_{correction}$ with respect to \mathbf{C} , equation (6.4) will have to be rewritten in a form that is independent from the

³In this document we will consider that \mathbf{C} is a counter-clockwise oriented contour.

parameterization of the contour. To do this, let us consider an arbitrary parameter p whose relationship with the arclength s is:

$$\frac{d\mathbf{C}}{ds} = \frac{d\mathbf{C}}{dp} \frac{dp}{ds} \quad (6.5)$$

which gives

$$\frac{ds}{dp} = \left| \frac{d\mathbf{C}}{dp} \right| \quad (6.6)$$

In the rest of this chapter, we adopt the following notations:

$$\begin{aligned} \mathbf{C} &\equiv \mathbf{C}(p) \\ \mathbf{C}' &\equiv \mathbf{C}(p') \\ R(p, p') &= |\mathbf{C} - \mathbf{C}'| \\ R(p, p') &\equiv R \\ \tilde{\mathbf{R}} &\equiv (\mathbf{C} - \mathbf{C}') R^{-1} \end{aligned}$$

The derivation with respect to the parameterization of the contour is symbolized by a point: $\dot{\mathbf{C}} = \frac{d\mathbf{C}}{dp}$. The tangent vectors at points p and p' then become: $\dot{\mathbf{C}}$ and $\dot{\mathbf{C}}'$. By using these notations, the quadratic energy to be minimized (6.4) becomes:

$$J_{correction}(\mathbf{C}) \sim \oint \oint dp dp' \left\{ \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 |\dot{\mathbf{C}}|^{-1} |\dot{\mathbf{C}}'|^{-1} - \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^4 |\dot{\mathbf{C}}|^{-3} |\dot{\mathbf{C}}'|^{-3} \right\} \Psi(R) \quad (6.7)$$

6.3.1.2 Energy gradient

With a view to finding the differential equation governing the evolution of the active contour with quadratic correction, the variation of $J_{correction}$ with respect to \mathbf{C} must be calculated:

$$\Delta J_{correction} = J_{correction}(\mathbf{C} + \delta\tilde{\mathbf{C}}) - J_{correction}(\mathbf{C}) \quad (6.8)$$

where $\Delta J_{correction}$ is the energy variation $J_{correction}$ generated by the infinitesimal deviation of the contour $\delta\tilde{\mathbf{C}}$. The calculation of $\Delta J_{correction}$ is long and fastidious and is summarized in appendix E. By assuming that the curve is parameterized again by the arclength, the energy variation is finally:

$$\begin{aligned} J_{correction}(\mathbf{C} + \delta\tilde{\mathbf{C}}) - J_{correction}(\mathbf{C}) &= \oint \oint ds ds' \Psi(R) \langle \tilde{\alpha}(s, s'), \delta\tilde{\mathbf{C}} \rangle \\ &\quad + \oint \oint ds ds' \Psi'(R) \langle \tilde{\beta}(s, s'), \delta\tilde{\mathbf{C}} \rangle \end{aligned} \quad (6.9)$$

where

$$\begin{aligned} \tilde{\alpha}(s, s') &= 2 \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 \left(1 - 3 \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 \right) \ddot{\mathbf{C}} \\ &\quad + 4 \left(1 - 6 \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 \right) \langle \dot{\mathbf{C}}', \ddot{\mathbf{C}} \rangle \left(\langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle \dot{\mathbf{C}} - \dot{\mathbf{C}}' \right) \end{aligned} \quad (6.10)$$

and

$$\begin{aligned}\tilde{\beta}(s, s') = & 2 \left(1 - \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 \right) \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 \tilde{\mathbf{R}} + 4 \langle \dot{\mathbf{C}}, \tilde{\mathbf{R}} \rangle \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle \left(2 \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 - 1 \right) \dot{\mathbf{C}}' \\ & + 2 \langle \dot{\mathbf{C}}, \tilde{\mathbf{R}} \rangle \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 \left(1 - 3 \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 \right) \dot{\mathbf{C}}\end{aligned}\quad (6.11)$$

By noting that $\ddot{\mathbf{C}} = \kappa \mathbf{n}$ where \mathbf{n} is the unit normal to \mathbf{C} , and κ is the contour's signed curvature, the quadratic force $F_{quad}(s)$ calculated at the arclength point s is:

$$F_{quad}(s) = \left\langle \frac{\partial J_{correction}}{\partial \mathbf{C}}, \mathbf{n} \right\rangle = \oint ds' \Psi(R) \langle \tilde{\alpha}(s, s'), \mathbf{n} \rangle + \oint ds' \Psi'(R) \langle \tilde{\beta}(s, s'), \mathbf{n} \rangle \quad (6.12)$$

where

$$\langle \tilde{\alpha}(s, s'), \mathbf{n} \rangle = 2\kappa \left\{ \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 \left(1 - 3 \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 \right) - 2 \langle \dot{\mathbf{C}}, \mathbf{n} \rangle^2 \left(1 - 6 \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 \right) \right\} \quad (6.13)$$

$$\begin{aligned}\langle \tilde{\beta}(s, s'), \mathbf{n} \rangle = & 2 \left(1 - \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 \right) \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 \langle \tilde{\mathbf{R}}, \mathbf{n} \rangle \\ & + 4 \langle \dot{\mathbf{C}}, \tilde{\mathbf{R}} \rangle \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle \left(2 \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 - 1 \right) \langle \dot{\mathbf{C}}, \mathbf{n} \rangle\end{aligned}\quad (6.14)$$

The shape constraint model authorizing parallel movements with respect to the prior shape with quadratic correction is finally governed by the equation:

$$\phi_t(\mathbf{x}, t) = - \frac{\partial J_{image}}{\partial \phi}$$

$$\begin{aligned}& + 4\lambda_{para} \delta_\alpha(\psi(\mathbf{x}, t)) (\langle \nabla \phi(\mathbf{x}), \nabla \psi(\mathbf{x}) \rangle - 1) [H_a(\phi(\mathbf{x}, t)) - H_a(\psi(\mathbf{x}, t))] \\ & + 2\lambda_{para} (\Delta \phi(\mathbf{x}) - \Delta \psi(\mathbf{x})) [H_a(\phi(\mathbf{x}, t)) - H_a(\psi(\mathbf{x}, t))]^2 - \lambda_{quad} F_{quad, ext}(\mathbf{x}) |\nabla \phi(\mathbf{x})|\end{aligned}\quad (6.15)$$

where $F_{quad, ext}(\mathbf{x})$ is the force extended over the image domain per equation (6.18).

6.3.2 Algorithm

The force $F_{quad}(s)$ is calculated for each point of contour $\mathbf{C}(s)$, it is then extended to every point in the domain defined by the narrow band surrounding the zero level of the levels set ϕ . The evolution algorithm of the active contour is then:

1. Construction of $\phi(\mathbf{x}, t = 0)$ in a narrow band; construction of $\psi(\mathbf{x}, t = 0)$ on the image domain Ω .
2. Extraction of the contour \mathbf{C} .
3. Calculation of the tangent and of the curvature for each point of the contour. When the point has non-integer coordinates, the geometric quantities are linearly interpolated between the two closest neighbors surrounding the point.

4. Calculation of the force $F_{quad}(s)$ for each point of the contour.
5. Extension of the force: $F_{quad}(s) \rightarrow F_{quad,ext}(\mathbf{x})$ according to equation (6.18).
6. Evolution of the active contour using (6.15).
7. Re-initialization of the distance function of level set ϕ .
8. Loop steps 2 to 7 until convergence is achieved.

6.3.2.1 Calculation of the quadratic force

The force $F_{quad}(s)$ is an integral on a closed contour which we approximate using the Newton-Cotes trapeze method. If we write $F_{quad}(s)$ in the form:

$$F_{quad}(s) = \oint ds' f(s, s')$$

The trapezoidal approximation will be:

$$\hat{F}_{quad}(s) = \sum_{s'=0}^{L-1} \frac{f(s, s') + f(s, s' + 1)}{2} |\mathbf{C}(s') - \mathbf{C}(s' + 1)| \quad (6.16)$$

which we can also write in the form:

$$\hat{F}_{quad}(s) = \sum_{s'=0}^{L-1} \frac{f(s, s')}{2} (|\mathbf{C}(s' - 1) - \mathbf{C}(s')| + |\mathbf{C}(s') - \mathbf{C}(s' + 1)|) \quad (6.17)$$

where: L is the length of the contour which has the cyclic conditions: $f(s, 0) = f(s, L)$; $\mathbf{C}(0) = \mathbf{C}(L)$. The numerical integration formula given in equation (6.16) implicitly imposes that the points $\mathbf{C}(s)$ of the contour should be ordered to make it possible to calculate the quantity $|\mathbf{C}(s') - \mathbf{C}(s' + 1)|$. When the active contour has been extracted by detection of the zeros of the level set ϕ , the contour tracing algorithm proposed by Pavlidis in [93] is used to order the points of \mathbf{C} .

6.3.2.2 Extension of the force

Extension of the force consists of filling $F_{quad,ext}(\mathbf{x})$ at each location \mathbf{x} of the narrow band with the value $\hat{F}_{quad}(s)$ of the contour's point closest to \mathbf{x} . The extended value of the force is thus constant along the normals to the level lines of ϕ . The force extension techniques are referenced in [110]. The most popular way consists of resolving the equation with partial derivatives proposed by the authors of [1]:

$$\frac{\partial F_{quad,ext}(\mathbf{x})}{\partial t} + \text{sign}(\phi) \left\langle \frac{\nabla \phi}{|\nabla \phi|}, \nabla F_{quad,ext}(\mathbf{x}) \right\rangle = 0 \quad (6.18)$$

The details relative to the extension of $F_{quad,ext}(\mathbf{x})$ are given in [101].

6.3.3 Implementation difficulties

The programming and implantation of the evolution equation (6.15) has raised many numerical instability problems. In particular, the precise evaluation of the geometric quantities such as the tangents or the curvature has been capital for attenuating the instabilities. The measures listed below in decreasing order of importance have been implemented to limit the instability problems:

1. Precise calculation of the geometric quantities

- (a) Calculation of the curvature using the formula proposed by the authors of [135] in the framework of fluid mechanics. The formulation proposed is robust with respect to the geometric singularities. The curvature is only calculated thanks to the level set ϕ .
- (b) Precise calculation of the tangents. The normals to the points of the contour are estimated, the associated tangents are deduced from this by a counter-clockwise rotation of $\pi/2$. We estimate the normal from ϕ as the mean of the normals in four directions [110, pp. 70]. Derivatives are estimated using the finite differences scheme. In order to increase the precision of the derivatives, we calculate them using an Essentially Non-Oscillatory (ENO) technique which consists of making a local approximation of ϕ by a polynomial of order n . The gradient of ϕ estimated in this way is more robust in the vicinity of the singularities [116].

- 2. **Re-initialization of the levels set.** The fast narrow band reconstruction technique described in section 4.3.4.1 has been found to be too coarse to be used with the quadratic models. Indeed, the so-called fast narrow band technique moves the contour (zero level of the level set) to a sub-pixel scale, which is a source of instabilities. We have opted for another re-initialization technique preserving the surface included in the contour, and which limits the movement of the front [116].
- 3. **Estimation of $\nabla\phi$.** The calculation of the gradient of the level set representing the active contour is used many times when calculating the quadratic energy: calculation of the tangents, re-initialization of ϕ , extension of the quadratic force. Given that we are seeking to create geometric singularities (corners) where the calculation of $\nabla\phi$ is delicate and sensitive to the calculation imprecision, we have used an ENO method with a polynomial of the order three to obtain a robust evaluation of this gradient.
- 4. **Choice of the interaction function Ψ .** We have conducted experiments with the interaction profile Ψ of the equation (4.60). This type of profile is not necessarily suited to our application since it is dedicated to the repulsion of anti-parallel tangents in a spatial vicinity smaller than d_{min} . Furthermore, the decrease of the function in equation (4.60) is likely to cause Gibbs disturbances in the spatial domain. The choice of a modified Bessel function of the second kind $\Psi(x) = K_0(x/a)$ ($a \in \mathbb{R}^{*+}$) whose shape is shown in figure (6.1) has given better results with attenuated numerical instabilities. $K_n(x)$ is the solution of the Bessel modified

differential equation, and its integral definition is:

$$K_n(x) = \frac{\sqrt{\pi}}{(n-1/2)!} \left(\frac{1}{2}x\right)^n \int_1^\infty e^{-zx} (z^2-1)^{n-1/2} dz \quad (6.19)$$

For $n = 0$, we have:

$$K_0(x) = \int_0^\infty \cos(x \sinh t) dt \quad (6.20)$$

The derivative of K_0 with respect to x is:

$$\frac{dK_0(x)}{dx} = -K_1(x)$$

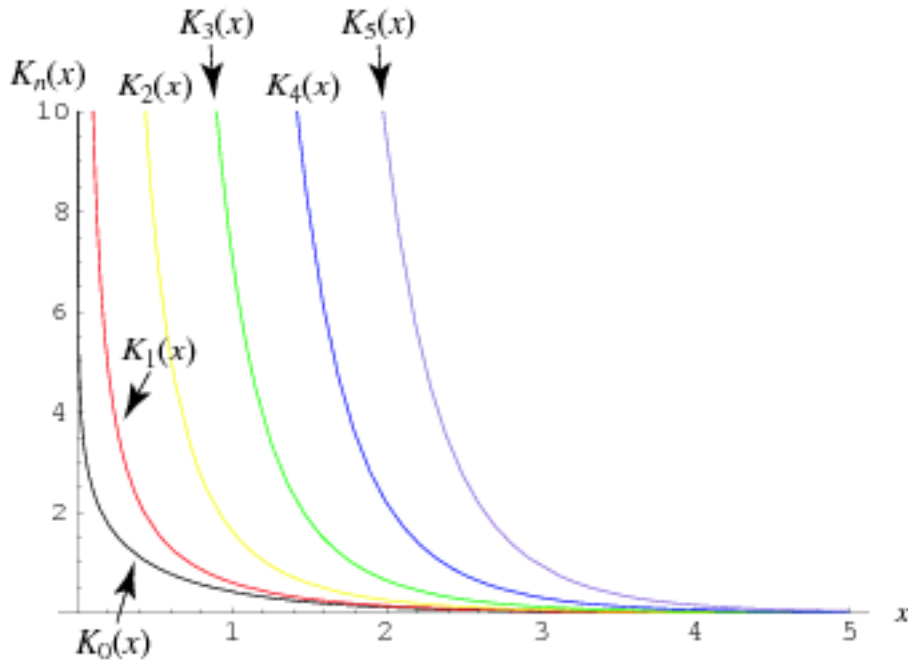


Figure 6.1: Modified Bessel functions of the second kind.

The function K_0 decreases asymptotically, but diverges when $x \rightarrow 0$. This singularity does not however prevent the integral of equation (6.12) from converging. However, this singularity must be taken into account for the numerical calculation of the integral expressed in equation (6.16). Indeed, the divergence of $\Psi(x)$ at 0 means that the term $f(s, s' = s)$ in (6.17) cannot be calculated. We make an approximation \hat{D} of the term:

$$D = \frac{f(s, s)}{2} (|\mathbf{C}(s-1) - \mathbf{C}(s)| + |\mathbf{C}(s) - \mathbf{C}(s+1)|)$$

by oversampling the segments $[\mathbf{C}(s-1) \mathbf{C}(s+1)]$ and $[\mathbf{C}(s) \mathbf{C}(s+1)]$ with a fixed pitch h (figure 6.2). The values of f at the oversampled places of the segments are calculated from the linear interpolations of the geometric quantities, \hat{D} is calculated using the trapezes method.

5. **Extraction of the contour.** The extraction of the contour by linear interpolation may be found to be inadequate because of a localization of the contour's points that is too coarse which, finally, biases the calculation of the geometrical quantities. We have opted for an ENO method making an approximation of the level set by a polynomial of degree two [112]. The roots of the polynomial designate the points of contour \mathbf{C} .
6. **Regularization due to the curvature.** A generic regularization term of the curvature has been used to lessen the remaining instabilities.

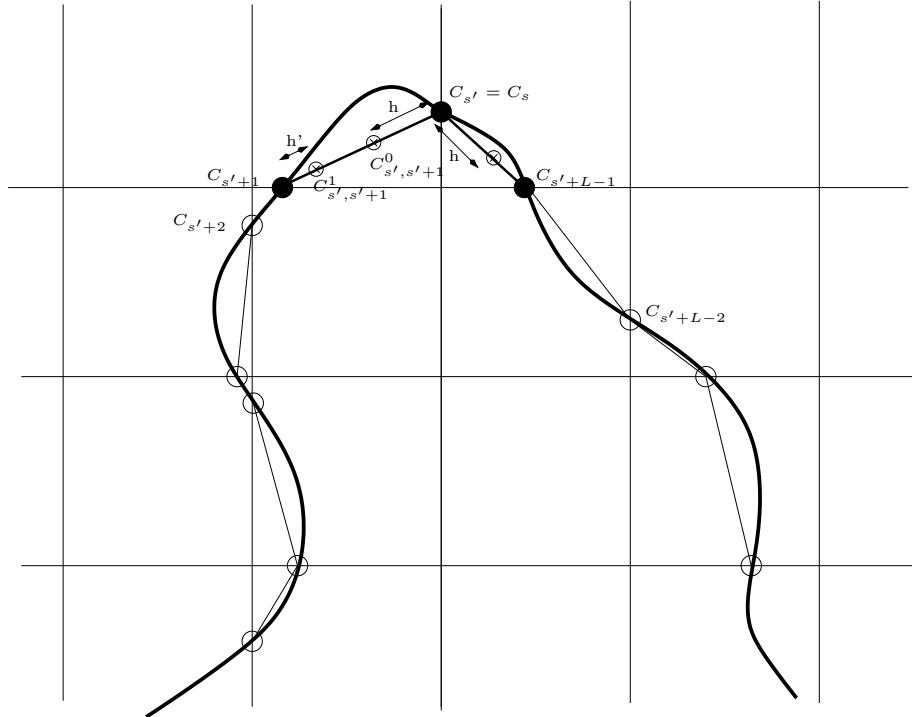


Figure 6.2: Diagram describing the sampling mode with a Bessel function for profile Ψ

6.4 Experiments and analysis of the results

6.4.1 Results without quadratic correction

We propose to conduct experiments with the model of equation (6.2) without quadratic correction with synthetic images. The purpose of these experiments is to evaluate the efficiency of the energy proposed and to measure the impact of the corner rounding effect. The synthetic image shown in figure 6.3 presents a U-shaped object locally altered by an erasure. For the experiments in this section, the initial contour is square and encompasses the object of interest (figure 6.3). The prior shape used is also U-shaped but with local differences with respect to the object of interest (figure 6.4). To

facilitate the comprehension of the results presented in figures 6.6-6.7, we display in blue the polygon that represents the prior shape on the image.

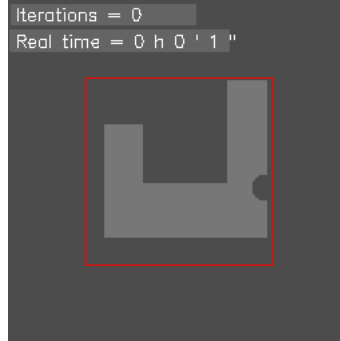


Figure 6.3: Synthetic image with the initial contour shown in red.

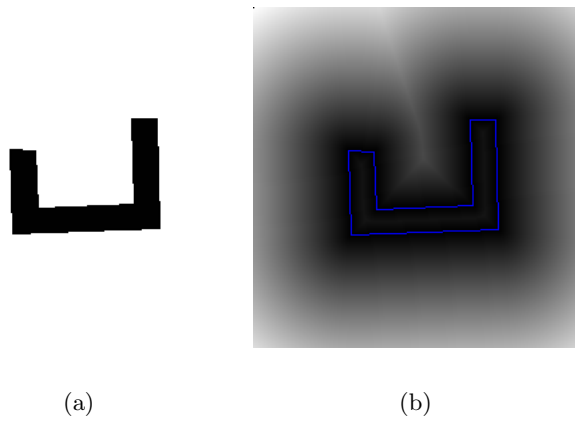


Figure 6.4: (a) Prior shape and its associated level set (b).

The experiment in figure 6.5 shows the result obtained with the conventional shape constraint expressed in equation (4.54) and the Chan and Vese attachment to data model.

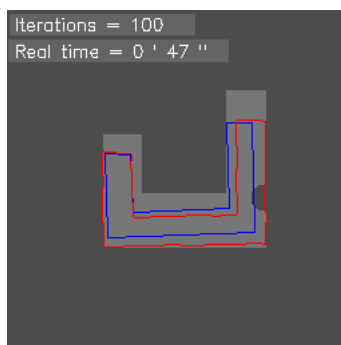


Figure 6.5: Result with the conventional shape constraint.

It can be seen that with a conventional shape constraint, the final contour is a trade-off between the image and the reference shape modulo a direct plane similarity. No local deviation with respect to the reference is permitted. The experiments in figures 6.6-6.7 illustrate the results with the new constraint energy and were carried out with different values of λ_{para} in order to assess its impact on the shape constraint. As shown in figures 6.6-6.7, the shape constraint with the new energy term makes it possible to move away from the prior shape, in order to segment the object in the image more precisely.

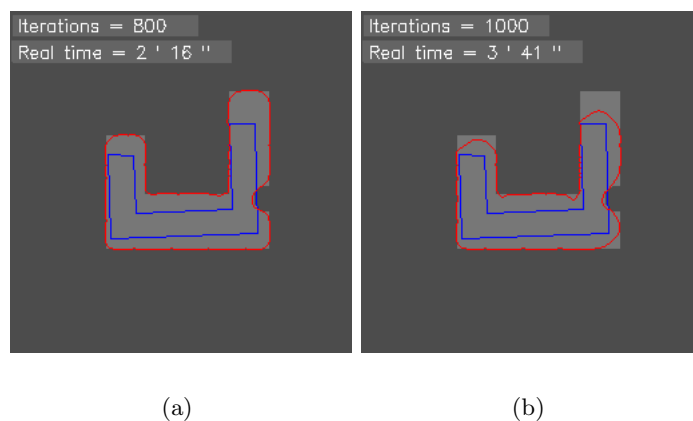


Figure 6.6: Segmentation with shape constraint authorizing parallel variations:
(a) $\lambda_{para} = 5$; (b) $\lambda_{para} = 10$

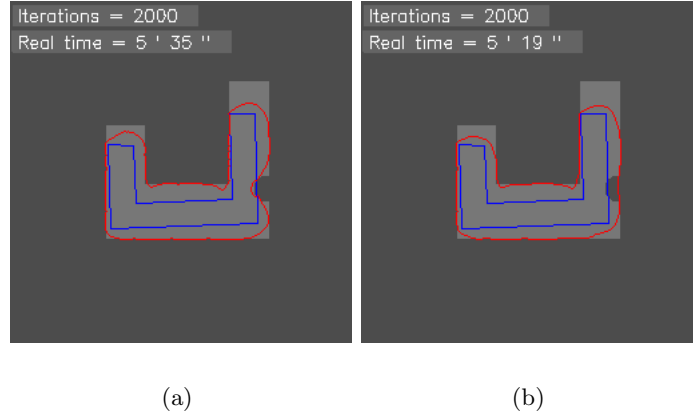


Figure 6.7: Segmentation with shape constraint authorizing parallel variations: (a) $\lambda_{para} = 15$; (b) $\lambda_{para} = 30$

The central bar of the U is now well extracted, even if the edges are far from the prior shape. It can be seen that naturally, the greater the weight of λ_{para} the more the model is capable of overcoming the alterations of the image (erasure of the object represented in the image in this case). However, this new energy introduces undesirable effects such as the rounding of the contour at the level of the outer corners; the dual distortion effect can also be seen on the inner corners of the reference shape. In the case of the outer corners, the level lines are curves outside the implicitly represented shape. Thus, the long-range shape constraint makes the contour respect this non-null curvature and induces the rounding of the active contour.

6.4.2 Results with quadratic correction

The experiments carried out with the quadratic force were conducted as follows:

1. The initial active contour is similar to the prior shape (represented in blue on the images) and moves according to the model in equation (6.2) without quadratic correction. The attachment to data model is that of Chan and Vese.
2. After n_0 iterations, the quadratic correction force is activated.

There are two reasons for such an experimental protocol: i) this makes it possible to reduce the calculation time. The quadratic force is a double integration to be carried out for each point of the contour, which drastically increases the complexity of the calculation. ii) The quadratic force seems to neutralize the contour's progression. This effect is not yet understood. One of the reasons could be the term proportional to the curvature of equation (6.13) which tends to reduce the length of the active contour.

Figure 6.8 compares the result obtained without and with quadratic correction with a perfect synthetic object. In figure 6.8.a we can note the effect of the rounding of the edges observed in the previous section. These roundings are attenuated with the

insertion of the regularization quadratic force (figure 6.10.b). It can be seen in particular that the active contour is rectilinear on the upper parts of the two branches of the U-shaped object, the inner parts are also correctly segmented.

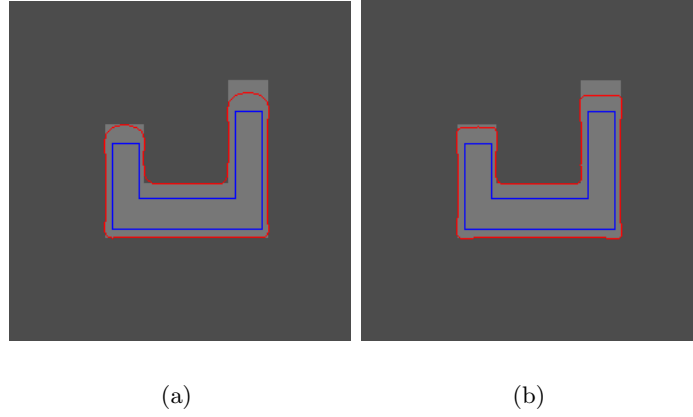


Figure 6.8: Segmentation with the Chan and Vese attachment to data model and the shape constraint authorizing parallel variations: (a) without quadratic correction and $\lambda_{para} = 20$ at instant n_0 ; (b) with quadratic correction, $\lambda_{para} = 20$, $\lambda_{quad} = 8$ and $a = 0.5$.

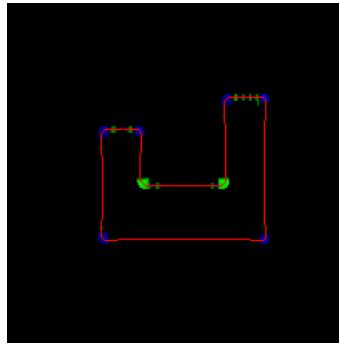


Figure 6.9: Representation of the quadratic force at instant n_0 . The blue zones correspond to a positive force (dilation effect), and the green zones represent a negative force (compression effect).

Figure 6.10 compares the result obtained without and with quadratic correction with an altered synthetic object. We note the same attenuation of the rounding effects. However, we note that the erasure is less well corrected. Figure 6.11 makes it possible to understand this result: the part of the active contour situated in the vicinity of the alteration receives a negative force which will tend to push back this part towards the inside. So, the attachment to data term and the quadratic force lead the contour to contract and overcome the force derived from the shape constraint. One solution for

resolving this problem would be to increase the weight of the shape constraint λ_{para} , however the contour would then not be able to move significantly away from the prior shape.

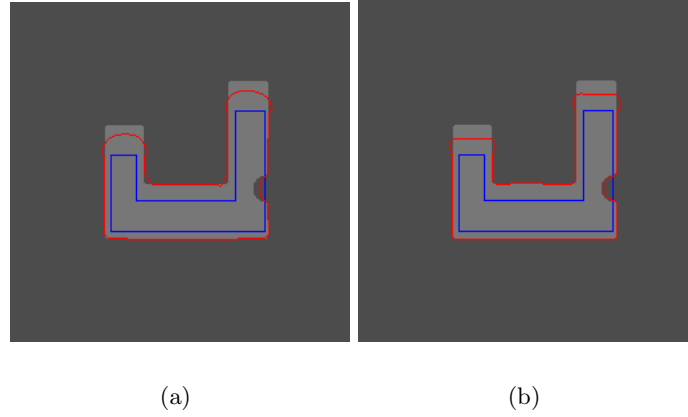


Figure 6.10: Segmentation with the Chan and Vese model and the shape constraint authorizing the parallel variations: (a) without quadratic correction and $\lambda_{para} = 50$ at instant n_0 ; (b) with quadratic correction, $\lambda_{para} = 50$, $\lambda_{quad} = 7$ and $a = 0.5$.

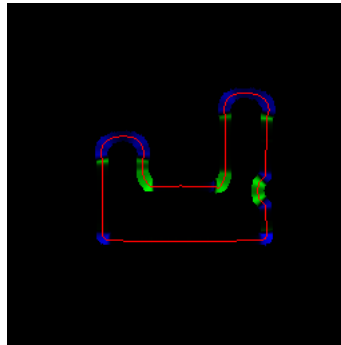


Figure 6.11: Representation of the quadratic force at instant n_0 . The blue zones correspond to a positive force (dilation effect), and the green zones represent a negative force (compression effect).

In general, it was difficult to set the respective weights of the constraint and of the quadratic term ($\lambda_{para}, \lambda_{quad}$) with respect to the attachment to data force. The example in figure 6.12 with a real image illustrates this difficulty. Since the building in the image is only slightly discriminated from the rest of the scene, we had to reduce the weight of the constraint ($\lambda_{para}, \lambda_{quad}$). In this experiment, the prior shape is the result of the matching of the cartographic polygon taken from the map with the conventional shape constraint used in the previous chapter. We note that the result with the real image is not satisfactory. Oscillations can be seen along the contour which seems to be sensitive

to the artifacts in the image. We had already noted this sensitivity with the synthetic image.



(a)

Figure 6.12: Experiment with a real image, the prior shape is the result of the matching with the conventional shape constraint. The segmentation is carried out with the Chan and Vese model, the shape constraint authorizing the parallel variations and with quadratic correction $\lambda_{para} = 4$, $\lambda_{quad} = 0.5$ and $a = 0.5$.

6.5 Conclusion

We have presented a new shape constraint energy model. It is more flexible and allows local variations of the active contour with respect to the prior shape. The proposed model includes a linear energy limiting the difference between the contour and the prior shape to belong to the class of parallel variations. A second quadratic energy imposes the rectilinearity of the contour and favors the formation of corners. These characteristics are suited to highly polygonal objects such as the buildings represented in remote sensing images. Experiments have given encouraging results with synthetic images. However, the application of the model to real images has remained unfruitful to date. Many aspects relative to the quadratic energy are still not understood, in particular the setting of the weighting of the quadratic force with respect to the shape constraint and to the image, the shape of the interaction profile as well as the size of the interaction neighborhood. All of these currently unanswered questions influence the numerical stability of the model.

Chapter 7

Map-to-image change analysis

7.1 Introduction

The active contours based fine matching enabled us to reduce variabilities between the cartographic object and its representation in the image. The variabilities we proposed to correct in the previous chapters are characterized either by a similarity transformation, or a normal displacement of the segments composing the cartographic object. These variabilities are signs of relatively minor differences between the map and the image, and may or may not derive from real changes between two data. In the present chapter, we propose a methodology to analyze changes between a map of buildings and multi-source remote sensing data. The approach we propose is based on merging processing of altimetric, multispectral and panchromatic data and is split into two stages. The first merges change indices intended solely to validate the unequivocal disappearance of a building. These indices are calculated from simple methods using the multispectral and altimetric image. In most cases where a change cannot be identified unequivocally, a second, more complex phase is carried out to characterize the degree of resemblance between the cartographic object and its representation in the panchromatic satellite image. This processing takes advantage of the fine matching by active contours and allows for the formulation of a probability of non-change for each element of the map. Experimental results illustrate the effectiveness of the proposed method, especially the contribution of fine matching in increasing the level of confidence conferred on the non-change decision. Finally, we shall examine the sensitivity of the proposed approach with respect to local shape errors contained in the map through statistical analysis of the results.

7.2 Methodology

The change detection method we propose implies a global registration of the map with remote sensing data and the Digital Surface Model generated (DSM). In our case, this registration is performed using the geocoding information of the data (figure 7.2). We then propose two levels of change detection:

1. The first works by merging low-level indices calculated from auxiliary data (DSM and multispectral image) likely to be acquired at different times but necessarily

more recent than production of the map (figure 7.1, phase A). The result of this first stage is to confirm only *unequivocal changes* of buildings but does not allow us to decide on non-changes. The buildings not detected as having changed are considered to be "indeterminate" and require the second phase more sophisticated processing.

2. If the first phase final result cannot confirm a change, we firstly aim at attenuating the low amplitude geometric exogenous variabilities between the cartographic object and its representation in the panchromatic satellite image. We do this using the active contours introduced in the previous chapters (figure 7.1, phase B). We then propose new change indices that are more elaborate than those in step 1. The first index is a measurement in a Hough space of the resemblance of segments taken from the image and of those from the refined cartographic object. The second evaluates the geometric variation between the initial cartographic object and its repositioning after active contours based matching. These indicators are merged in the form of an energy allowing a probability of non-change to be expressed in the Gibbs probabilistic framework (figure 7.1, phase C). This result is an indicator of change that can be used to facilitate and accelerate an operator-driven map updating procedure.

The methodology proposed is summarized in 7.1.

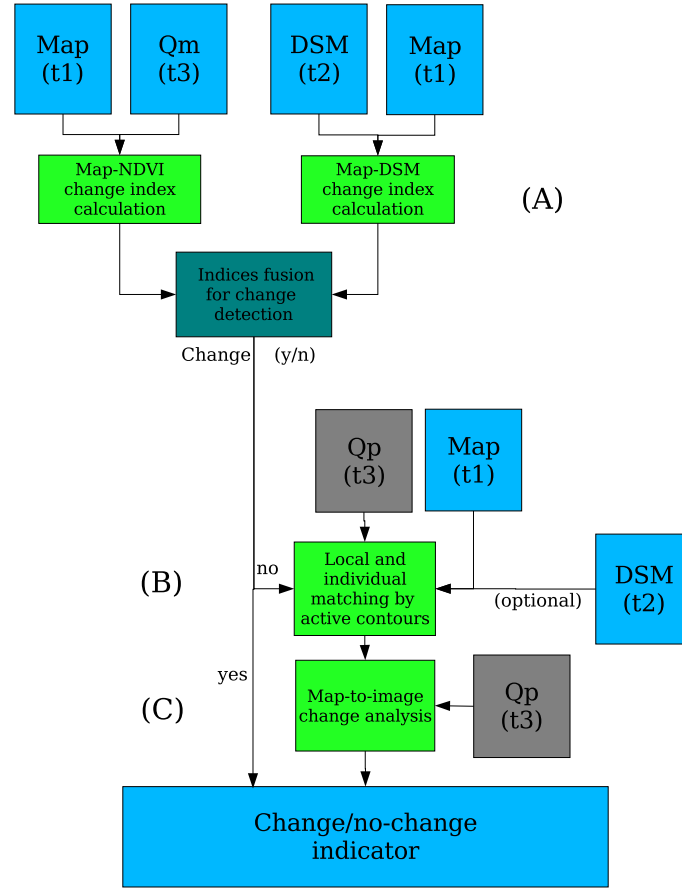


Figure 7.1: Proposed change detection methodology: (A) calculation and fusion of change indices. (B) attenuation of map/image exogenous variabilities (C) analysis of final changes. Legend: Q_m and Q_p are respectively the multispectral and panchromatic satellite images.

7.3 Calculation and fusion of simple change indices

This section introduces two simple methods to calculate indices confirming unequivocal changes of buildings from remote sensing auxiliary data. The first compares elements of the map with altimetric data (DSM) and generates a change index derived from statistical tests and altitude thresholding. The second uses the multispectral satellite image. A classification by maximum likelihood allows to detect the presence of vegetation and thus also the absence of buildings. Both indices produced are then merged to take advantage of the complementarity between the altimetric and multispectral representations.

7.3.1 Index calculated from the altimetric data (DSM)

7.3.1.1 Limitations of the map-to-DSM comparison

A method for change detection between a map and a DSM is intrinsically limited. Indeed, a DSM only provides altitude data, which is far from being exhaustive to fill in the reality of a scene. Thus, it will be difficult to distinguish buildings from the canopy with only a DSM generated from 3D reconstruction (conversely, LIDAR imagery analysis allows the difference with vegetation to be ascertained).

More specifically, the DSM we generated contains unfilled pixels inherent to the stereo-correlation method that was used to build it. These unfilled pixels represent 30% to 40% of the image and, due to occlusions, are localized in the neighborhood of buildings. Furthermore, these locally indeterminate zones may be enlarged using the pre-processing of the disparity images explained in section 3.4.2. This lack of altitude data is an additional limitation for change detection preventing decision-making. This is the reason why we decided to design a simple map-to-DSM change detection method. A more sophisticated approach would not necessarily be more effective due to the incomplete nature of the data provided by the DSM.

Finally, map-to-DSM registration errors must also be taken into account. It would appear to be difficult to apply local matching by active contours with the DSM as long as the latter is noisy. Fine matching using the panchromatic satellite image to eliminate map-to-DSM exogenous variabilities appears to be awkward as these data items are exogenous: their geometries will be different as are their dates of acquisition.

7.3.1.2 Approach

The algorithm we propose relies on calculating global statistical criteria such as the mean, median and variance from DSM altitude values contained within a cartographic object or in its vicinity. The method set up is limited to two cases of unequivocal changes that correspond to situations *a* and *b* of table 2.2.

1. Building completely demolished and replaced by a bare, flat, incline-free ground.
2. Building completely demolished and replaced by a bigger building with a flat, non-inclined roof.

We limit our attention to non-inclined, flat roofs or ground as these flatness characteristics are easy to verify using measurement of the variance in altitude of the DSM pixels inside a building on the map (polygon).

The change detection method can be described as follows: for each cartographic object superimposed on the DSM, we calculate the percentage of pixels filled in for the DSM and included in the polygon. If this percentage is not high enough, no change decision will be made and the case will be considered to be undecidable. Conversely, altimetric flatness is verified in the object and in its vicinity. If flatness is proven, a

final test will measure the variation in altitude (in the sense of the median) between the polygon and its neighborhood. If this variation is less than the normal size of a building, a case of change will be detected; otherwise the case will be considered to be ambiguous. The final result of this change detection is a binary score s_{MNS} equal to 1 when a change is validated. Undecidable and ambiguous cases are grouped together to be represented by a score $s_{MNS} = 0$.

While this method remains simple, it has to be reliable so as to make the least possible number of errors. We shall evaluate the method's success in comparison with its various parameters. These parameters are three in number:

1. $thres_{filled}$ percentage of DSM pixels filled for altitude below which change detection is not implemented.
2. $thres_{flat}$ altimetric variance threshold (in meters) below which a zone of the DSM is considered to be flat and non-inclined.
3. $thres_{build}$ altimetric threshold (in meters) above which a pixel of the DSM is considered to belong to the construction.

The algorithm for the method presented then becomes:

1. For each polygon of the GIS superimposed on the DSM:
 - (a) If the percentage of unfilled pixels inside the polygon or in its vicinity $< thres_{filled}$, then change detection will not be implemented. The result is *undecidable*.
 - (b) Otherwise if the polygon and its neighborhood have a standard deviation $< thres_{flat}$ AND the neighborhood has a rate of unfilled pixels $< thres_{filled}$ then:
 - i. If the difference in altitude between the polygon and its neighborhood $< thres_{build}$, then the building will be considered to have *changed*.
 - (c) Otherwise the case will be considered to be *ambiguous*.
2. End.

We perform calculations both on the vicinity of a polygon and on its inside to evaluate relative differences in altitude. An absolute measurement infers knowledge of altitude of the ground for the scene, which requires the a DTM that in turn allows the DSM to be normalized. Without the DTM, an absolute detection criterion would fail if the relief of the land at the scene was not strictly flat (mountainous or hilly regions).

As we have stated, the method needs to cater for possible imperfections in map-DSM registration. The proximity of an object in the DSM is consequently defined as being the crown resulting from subtraction of a dilatation of the cartographic polygon

by a 15×15 pixels structuring element with that of 3×3 pixels. The neighborhood is thus not too close to the edge of the polygon, thus preventing from considering pixels belonging to the building that could distort calculations of statistical criteria for change detection.

7.3.1.3 Experimental results

The change detection illustrated in figure 7.3 was conducted on 975 buildings with the following parameters: $thres_{filled} = 10\%$; $thres_{flat} = 1$ m; $thres_{build} = 3$ m.

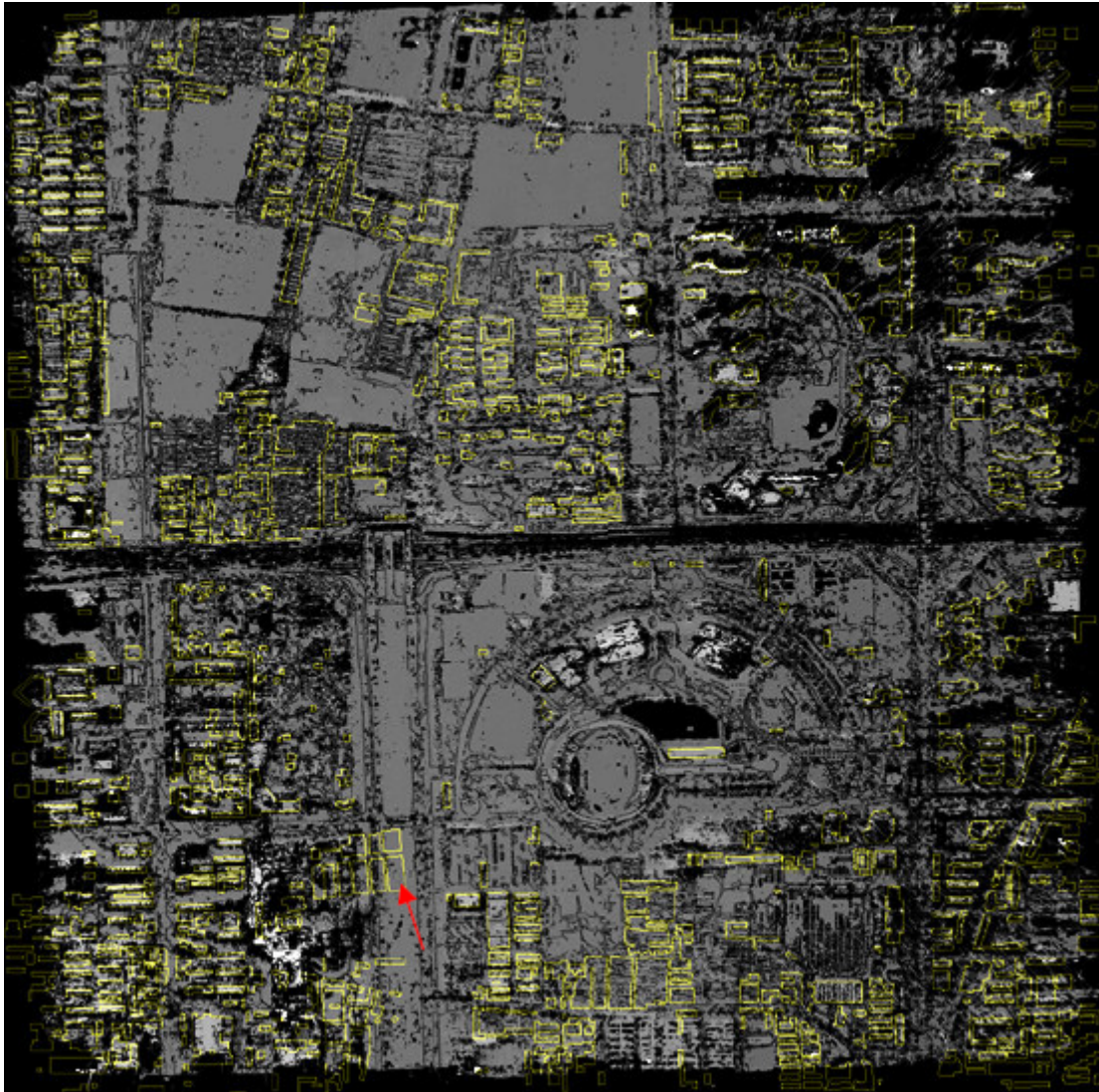


Figure 7.2: Registration of the building map (shown in yellow) with the DSM using geocoding information. The red arrow points to a change: building replaced by a terreplein.

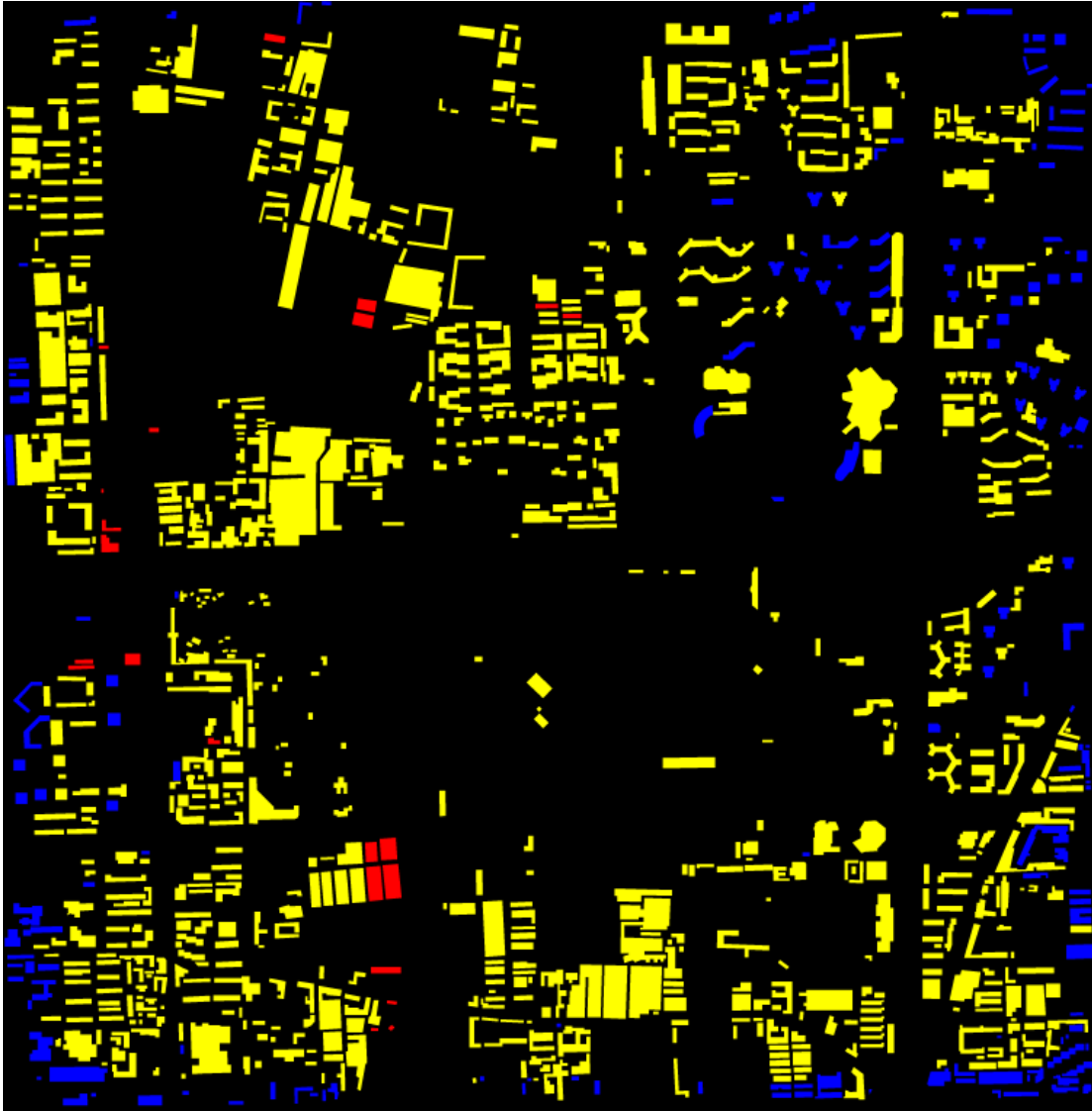


Figure 7.3: Result of change detection between the map and the DSM. Legend: red (change), yellow (ambiguous), blue (undecidable). Parameters: $thres_{filled} = 10\%$, $thres_{flat} = 1$ m, $thres_{build} = 3$ m.

As in any method, choosing the values for parameters remains a delicate matter. However, they have physical values that cannot be chosen at random. A flatness threshold $thres_{flat}$ higher than three meters is devoid of meaning. Setting too high a threshold for building detection will increase the risk of false alarms, as any building whose altitude is lower than the threshold will be wrongly classified as having changed. Finally, making decisions with too low a percentage of pixels being determined (filled with altitude information) will incur a greater risk of errors generated by the method.

* Variational analysis of parameters.

We conducted an analysis of the method's success in relation to the variation in its three parameters (this analysis was conducted with GIS cartographic data from 1996 as illustrated figure 7.2). Systematically, two parameters were fixed, with just one varying. We evaluated the method's success by visual examination of the map and an aerial view that had been used to produce the DSM. The results are summarized in figures 7.4 to 7.6.

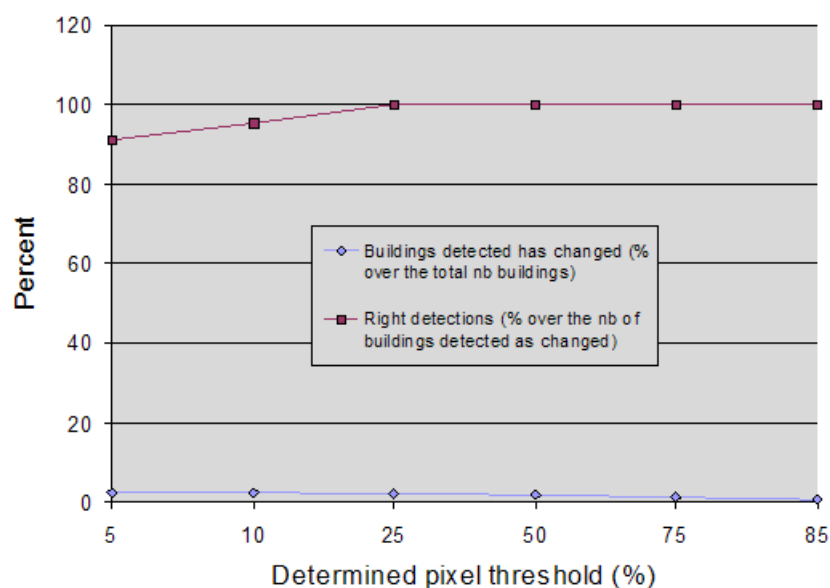


Figure 7.4: The method's success in relation to $thres_{filled}$. The rate of buildings correctly detected as having changed is shown in red. The blue curve represents the rate of buildings detected as having changed with respect to the total number of buildings in the map. $thres_{flat} = 1m$ and $thres_{build} = 3m$.

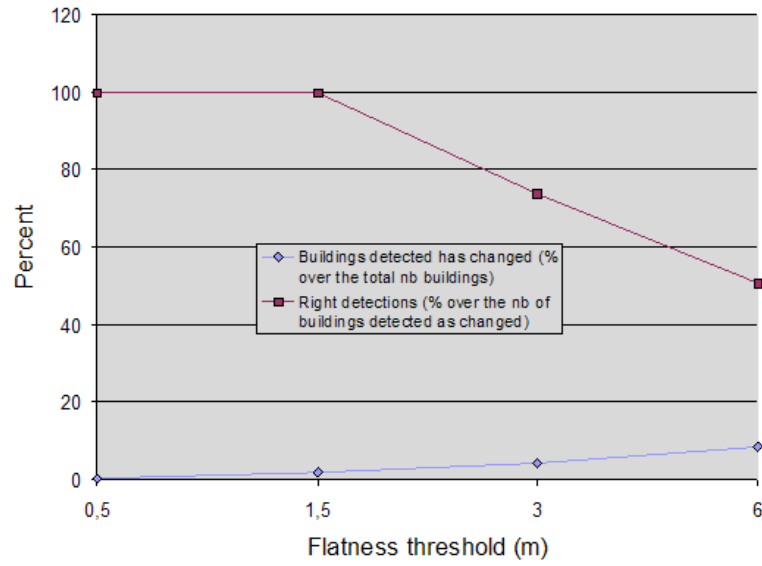


Figure 7.5: The method's success in relation to $thres_{flat}$. The rate of buildings correctly detected as having changed is shown in red. The blue curve represents the rate of buildings detected as having changed in relation to the total number of buildings on the map. $thres_{filled} = 50\%$ and $thres_{build} = 3m$.



Figure 7.6: The method's success in relation to $thres_{build}$. The rate of buildings correctly detected as having changed is shown in red. The blue curve represents the rate of buildings detected as having changed in relation to the total number of buildings on the map. $thres_{filled} = 65\%$ and $thres_{flat} = 1.5m$.

We can see that the more $thres_{filled}$ increases, the more the method is successful (figure 7.4). The more $thres_{flat}$ increases, the less successful the method becomes. It should be noted that the method seems to be more sensitive with respect to this parameter than with others. Beyond 1 m - 1.5 m for $thres_{flat}$, the method's success drops off critically (figure 7.5). Finally, as the $thres_{build}$ increases, so the rate of false alarms also rises, impairing the method's success (figure 7.6).

Examination of the values in the graphs shows the relatively few buildings for which a change decision is made. This is consistent with our initial assumption of restricting the method to the limited cases of change detection mentioned in table 2.2. One problem with the method involves having to systematically consider the vicinity of a polygon to be able to conduct the test for presence of the building comparing the difference in altitude of the neighborhood and of the polygon with $thres_{build}$. If the neighborhood cannot be calculated due to too high a rate of unfilled pixels (whereas the pixels inside the polygon are fairly well filled in), the case will not be processed. Normalizing the DSM with a DTM would provide a solution to dispense with the need to calculate the related altitude criterion and its intrinsic problem. We could then use an absolute criterion that tests only the median altitude of the pixels included in the polygon with respect to a threshold.

To conclude this section, we shall look into the factors influencing the proposed map-to-DSM change detection method and thus see whether it can be generalized to sites other than Beijing. One first major factor is the relief of the scene. If the relief is too rugged, ground flatness conditions will never be fulfilled and, as a result, there will be little decision-making. The method is therefore not suited to hilly or mountainous countryside unless we have first normalized the DSM using a DTM. As the Beijing terrain is extremely flat, we were able to dispense with this normalization stage. A second factor related to the architectural culture of the city concerns the shape of the buildings. The method can only detect changes of type *b* with flat-roofed buildings (table 2.2). A city with double slope roof and contiguous buildings will not satisfactorily lend itself to this change detection approach.

7.3.2 Index calculated from a multispectral image

The multispectral images we have in our possession have a band in the near infrared that allows us to characterize the vegetation. Assuming that there is no overlap between the vegetation and constructions, we shall then be able to confirm the disappearance of a building at locations on the map that are covered over with vegetation in the image. As for the previously explained method, we shall not be able to confirm a non-change since absence of vegetation does not imply the presence of a building (which may have been replaced by a road for example).

Calculation of the change index between the cartographic object and its representation in the multispectral image must answer two questions:

1. How to characterize the vegetation of the scene?

2. How to make a reliable change decision from detected vegetation?

Characterization of vegetation. We use an index that has conventionally been employed in remote sensing since 1969, the *NDVI* (*Normalized Difference Vegetation Index*). This involves to subtract the infrared band from the red one, then normalizing this term by summing both bands:

$$NDVI = \frac{IR - R}{IR + R}$$

This index varies between -1 and 1. Vegetation has high absorption in the red and strong reflectance in the infrared. Thus, a positive value for *NDVI* confirms the presence of vegetation. The closer the value is to 1, the stronger this assumption will be. A negative value confirms absence of vegetation. Figure 7.7 illustrates the values for this index in accordance with eleven classes. On figure 7.7 we can see that vegetation is pretty well characterized, with nevertheless relatively low *NDVI* values (on the illustrated examples, the *NDVI* does not exceed 0.4). There are several explanations for this: i) the seasonal effect has strong impact on the *NDVI* value that measures the vegetation's vitality. The Quickbird image studied was acquired in March, just before Spring. Thus, the aerial biomass was relatively underdeveloped. ii) the density of dust or pollution (which is high in Beijing) also impacts the amplitude of *NDVI*.

Decision-making with the NDVI. As with the *s_{MNS}*, we wish to formulate a binary criterion *s_{NDVI}* whose value 0 means indetermination while 1 corresponds to a high density of vegetation within the cartographic polygon considered, i.e. to a change indicator. We use a supervised classification by maximum likelihood to determine *s_{NDVI}*. This type of method has the advantage of avoiding the pitfalls of empirical approaches based on thresholding of the *NDVI* value to characterize vegetation. However, the method is supervised and requires a learning process. In our case, this learning process can be automated thanks to the “vegetation” layer of a GIS digital map (assuming that few changes have affected the vegetation between when the map was made and acquisition of the image).

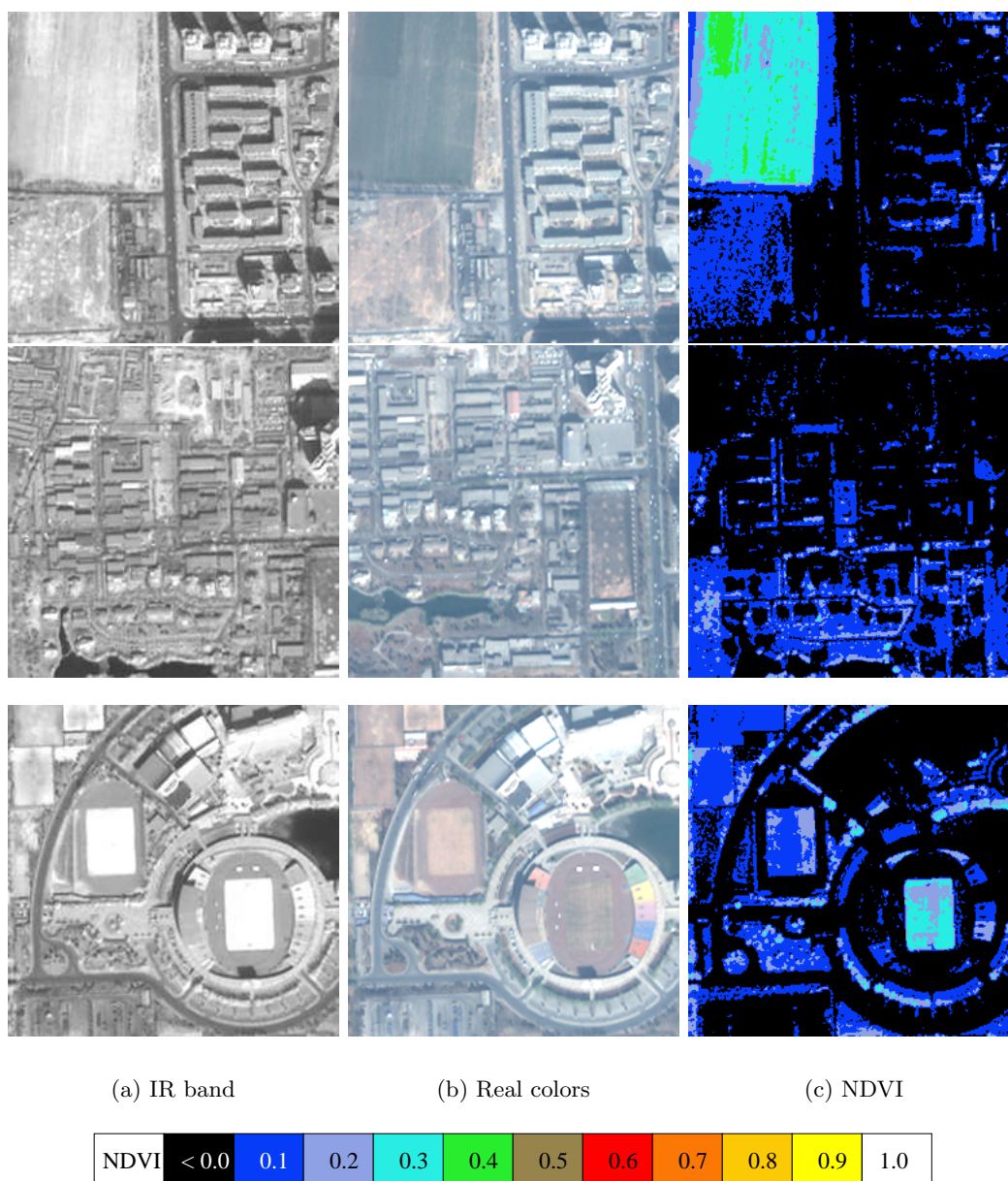
Let us consider n independent measurements of the *NDVI* inside the j^{th} cartographic polygon. We shall note these measurements y_i :

$$y_{i \in \{1, \dots, n\}} = NDVI(\mathbf{x}_{i \in \{1, \dots, n\}})$$

where \mathbf{x} is the pixel position of the i^{th} pixel of the *NDVI* image belonging to the j^{th} polygon of the map.

Let $p(y_i|\theta)$ be the conditional and normalized probability density for observing the measurement y_i for the given set of parameters θ . Estimation of parameters θ is achieved by maximization of the likelihood function:

$$\mathcal{L}(y_i|\theta) = \prod_{i=1}^n p(y_i|\theta) \quad (7.1)$$

Figure 7.7: Infra-red band, real colors (RGB), and $NDVI$.

where \mathcal{L} is the likelihood function measuring the probability of observing the special set of data y_i knowing θ . In our case, we shall consider the two classes $\{C_k\}_{k \in \{0,1\}}$ with C_0 and C_1 denoting non-vegetation and vegetation respectively. According to our initial hypothesis, there is no overlap between constructions and vegetation, thus class C_1 is associated with change in buildings, while C_0 corresponds to indetermination.

We also assume that $p(y_i|\theta)$ obeys a normal distribution with mean m and variance σ^2 . In this special case, each class is fully characterized by these two statistical quantities

that have to be measured by a learning process: $\theta_k = (m_k, \sigma_k^2)_{k=0,1}$. For the cases of experiments illustrated in this chapter, the training areas were determined manually. The problem of binary classification of a cartographic polygon thus comes down to choosing the class \mathcal{C}_k whose likelihood function is maximum from the measured data:

$$\mathcal{L}(y_i | \theta_k)_{k=0,1} = \frac{1}{(2\pi\sigma_k^2)^{n/2}} e^{-\frac{\sum_{i=1}^n (y_i - m_k)^2}{2\sigma_k^2}} \quad (7.2)$$

The change indicator related to the *NDVI* for the cartographic polygon j considered will then be:

$$s_{NDVI}(j) = \arg \max_{k,k=0,1} \mathcal{L}(y_i | \theta_k) \quad (7.3)$$

7.3.3 Change indicators fusion

The calculations of change indices based on the DSM and vegetation show many similarities: they are simple methods that can only disprove the presence of a building of the map in binary fashion. They are also methods that use non-redundant data and with results whose complementarity can be used for decision-making. We merge the map-to-DSM and map-to-NDVI change indices in accordance with the following strategy:

1. The DSM and the NDVI image are generated from data acquired at different dates. Nevertheless, it does not seem wise to use the acquisition time difference between the data to balance their influence in the final merged decision. Indeed, it would be erroneous to weaken or strengthen a confirmed change decision in relation to the date of acquisition as a proven change lasts over time. It is preferable to determine a method's contribution with respect to the degree of confidence assigned to it. For our application, we do not have prior knowledge allowing to state that one or other is less reliable, so we grant them equal levels of confidence when both methods confirm a change.
2. When one of the two methods refers to an indeterminate result while the other confirms a change, we choose not to penalize the one having detected the change in the fusion of results. The two methods based on the comparison with the DSM and NDVI image are simple and cover limited instances of change. Thus, an indetermination does not imply a strong assumption of non-change. It is often due to the inability of the method to make a decision.
3. When both methods refer to an indeterminate result, the fusion of results is also indeterminate.

To summarize, the fusion of change indices is simply formalized in the following form for the cartographic element j considered:

$$s_{fusion,MNS-NDVI}(j) = \max(s_{MNS}(j), s_{NDVI}(j))$$

7.4 Map-to-image change analysis

Whereas the previous section only covered unequivocal disappearances of buildings symbolized in the map, the method for change analysis between the map and the panchromatic satellite image is also intended to take into account cases of non-change. This method assumes a local and individual matching of each building with the image is first performed using the active contours technique (chapter 5). This pre-processing aimed at reducing map/image exogenous variabilities enhances the consistency of the cartographic object with its homologous representation in the image if no change has occurred. The map-to-image change analysis method is then applied on each matched/refined cartographic building. For each cartographic object considered, the proposed method is based on two kinds of information:

1. **Measurement of consistency between segment primitives extracted from the image and those taken from the map.** Consistency is measured by a voting in the Hough space that estimates the translation that best matches the two representations. We shall see that the maximum value of the Hough accumulator is an indicator of resemblance between the cartographic object and the one in the image. This will enable us to deduce whether a change or a non-change has occurred. The segments extracted from the image are the result of rectilinear chaining of high gradient pixels of the sub-image framing the considered cartographic building. These segments of the image are those used by the edge-based active contours (see section 4.2.1.3). The segments of the cartographic object refined by the active contours technique are the result of the vectorization of the final contour.
2. **Geometrical variation from map-to-image fine matching.** Too high a geometrical variation between the original cartographic object and its refined version by active contours is also a sign of change. This variation can be measured using the level sets ϕ_{final} and ψ_0 which represent respectively the final active contour (result of fine matching) and the initial cartographic object.

We shall formulate these both types information as data related terms to define a Gibbs energy. This will enable us *in fine* to calculate a probability $p_{NC}(j)$ of non-change for the cartographic object j .

7.4.1 Matching map and image segments by Hough voting

7.4.1.1 Principle

Let S and S' be two sets of segments. S is the set of segments of a considered cartographic building. This object was refined by active contours matching. S' is the set of segments taken from a sub-image surrounding the cartographic object.

We propose to use a Hough space to accumulate the evidence of resemblance between S and S' [52]. This type of bottom-up approach is robust and gets rid of the contribution from image segments that do not belong to the building we are interested in.

The method's principle is to consider each pair of map/image segments individually and estimate the translation allowing for them to be superimposed. This translation is projected in a Hough accumulator that thus allows to determine the most likely translation μ_{Hough} matching S and S' . Normalization of the maximum of the Hough accumulator by the perimeter of the cartographic object provides a resemblance score s_{Hough} between S and S' varying between 0 and 1. If $s_{\text{Hough}} \approx 1$ and $\mu_{\text{Hough}} \approx \mathbf{0}$ then the cartographic object's consistency with the image is proven. Indeed, the majority vote is significant and the estimation of the translation is consistent with the fact that the refined cartographic object is matched on its representation in the image. If $s_{\text{Hough}} \ll 1$ or $\mu_{\text{Hough}} \neq \mathbf{0}$ then a lack of resemblance will be detected, corresponding to a change. This method, based on the image gradient, will be sensitive to shadows of buildings. It is likely that the pairing will occur on the shadow cast by the building rather than its actual edges. We therefore first filter the segments of shadow in the image before the estimation of the translation between each pair of map/image segments.

7.4.1.2 Algorithm

The algorithm for evaluation of s_{Hough} and μ_{Hough} is as follows:

1. Let $\mathbf{u}_{\text{ombre}}$ be the unit vector known a priori giving the orientation and direction of the shadow in the image. This vector has the same orientation and the same direction as the Sun's rays. It can be calculated through the knowledge of the image acquisition parameters and the date of acquisition of the remote sensing image. In our case, we measure it manually. The measured vector is near-constant over the entire image.
2. Let S and S' be two sets of segments with sizes n and m respectively.

$$S = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n\} ; S' = \{\mathbf{s}'_1, \mathbf{s}'_2, \dots, \mathbf{s}'_m\} \quad (7.4)$$

3. For each segment $\mathbf{s}_i \in S$:
 - (a) For each segment $\mathbf{s}'_j \in S'$:
 - i. If the mean of radiometric intensity along the profile $\mathbf{s}'_j + \mathbf{u}_{\text{ombre}}$ is less than that of $\mathbf{s}'_j - \mathbf{u}_{\text{ombre}}$ plus a threshold, then:
 - A. If the segments \mathbf{s}_i and \mathbf{s}'_j are near collinear ($\frac{\langle \mathbf{s}_i, \mathbf{s}'_j \rangle}{|\mathbf{s}_i| |\mathbf{s}'_j|} \leq \cos \alpha_s$, with α_s a given threshold), we estimate the minimum $\mu_{\text{ij,min}}$ and maximum $\mu_{\text{ij,max}}$ translation allowing each of the ends of the two segments to be superimposed (figure 7.8).
 - B. In a 2-D accumulator (μ_x, μ_y) whose axes represent the space of possible translations in directions Ox and Oy , we add the segment $[\mu_{\text{ij,min}}, \mu_{\text{ij,max}}]$ with a weight equal to $\min(|\mathbf{s}_i|, |\mathbf{s}'_j|)$, where $|\mathbf{s}_i|$ represents the length of the segment \mathbf{s}_i . To make the method more robust with respect to local variations in the segments of the two representations (uncertainties due to the segment extraction method and errors present in one or other of the representations), the segment is enlarged by l pixels to form a band in the accumulator (figure 7.9).

(b) End

4. The μ_{Hough} translation is estimated at the location of the accumulator maximum value. This maximum is normalized by the perimeter P of the cartographic polygon ($P = \sum_{i=1}^n |s_i|$) to obtain s_{Hough} .

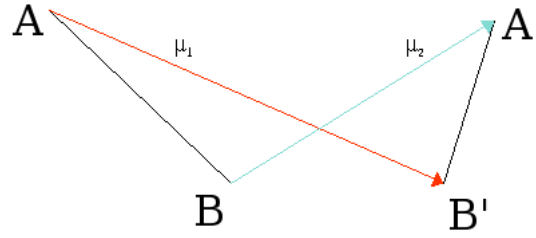


Figure 7.8: Estimation of the maximum μ_1 and minimum μ_2 translation between a map segment $[AB]$ and an image segment $[A'B']$.

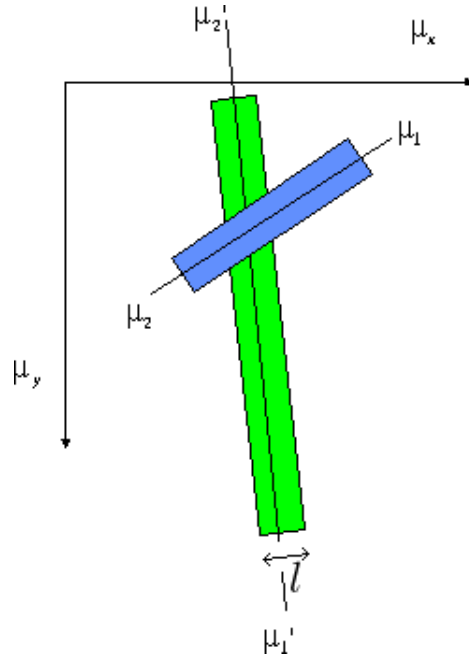


Figure 7.9: 2D Hough accumulator (μ_x, μ_y) . Accumulation between minimum and maximum translation is enlarged in a band of l pixels.

7.4.1.3 Energy associated with the Hough voting

With the aim of inserting the Hough voting decision within the Gibbs probabilistic framework, we formulate an energy E_{Hough} measuring consistency between a carto-

graphic object and the image. The proposed energy is as follows:

$$E_{Hough} = \frac{1}{P} \oint (1 - \alpha(\mathbf{C}_{\mu_{Hough}}(p))) dp \quad (7.5)$$

where P is the perimeter of the cartographic object \mathbf{C} parameterized by p . $\mathbf{C}_{\mu_{Hough}}$ is the contour \mathbf{C} translated by the μ_{Hough} estimate. The quantity $\alpha(\mathbf{x})$ is such that:

$$\alpha(\mathbf{x}) = \begin{cases} 1 & \text{if } I(\mathbf{x}) \in S'_{Hough} \\ 0 & \text{otherwise} \end{cases} \quad (7.6)$$

where $I(\mathbf{x})$ is the satellite image, and S'_{Hough} is the set of segments extracted from the image having contributed to the majority poll in the Hough space, i.e. having contributed to the estimation of μ_{Hough} . The energy of the equation (7.5) will be minimal when all the pixels of the image found under the plot of the translated cartographic object have contributed to the Hough vote. Otherwise, few pixels of the image will have contributed to the poll, which is an indication of change. The energy will then be maximal. Normalization by the perimeter P allows the energy of the equation (7.5) to be invariant with respect to the size of the considered cartographic object. The energy of the equation (7.5) can be expressed more simply:

$$E_{Hough} = 1 - s_{Hough} \quad (7.7)$$

where s_{Hough} is the normalized Hough score defined in the previous section.

Finally, let us recall that condition $E_{Hough} = 0$ is not enough to characterize a non-change. Indeed, this condition has to be associated with an estimate μ_{Hough} with low amplitude. To account for this property, we use the energy associated with a spring of stiffness k that will penalize the estimation of a too high translation:

$$E_{\mu_{Hough}} = k |\mu_{Hough}|^2 \quad (7.8)$$

7.4.2 Measurement of geometric variation due to active contours matching

When the initial map-to-image registration is of poor quality, the invariance from affine transformation by active contours allows this effect to be corrected by rotation, translation and scaling. This correction plays a role in reducing exogenous variabilities independent of an effective change and should not intervene in the change detection. Nevertheless, if geometrical variation is too significant, it may not be ascribable to the poor quality of the initial superimposition, but rather to a change (figure 7.11). We propose to formulate an energy criterion E_{geom} measuring the geometrical variation between the initial cartographic object and its transform by $T_{sim,final}$ which is the similarity estimated by the active contours fine matching:

$$E_{geom}(\phi, \psi_0) = \frac{1}{A_{\psi_{final}} + A_{\psi_0}} \int_{\Omega} (H(\psi_{final}(\mathbf{x})) - H(\psi_0(\mu_{persp}\mathbf{x})))^2 d\mathbf{x} \quad (7.9)$$

where ψ_0 is the level set representing the initial cartographic object; ψ_{final} is the level set representing the cartographic object transformed by $T_{sim,final}$: $\psi_{final} = \psi_0(T_{sim,final}\mathbf{x})$;

$A_{\psi_{final}} = \int_{\Omega} H(\psi_{final}(\mathbf{x})) d\mathbf{x}$ and $A_{\psi_0} = \int_{\Omega} H(\psi_0(\mathbf{x})) d\mathbf{x}$. The energy term of the equation (7.9) is inspired from that used for the prior shape constraint of the equation (4.54). However, two differences should be noted. Firstly, the energy E_{geom} is made independent of the size of the compared objects thanks to normalization by the sum of areas included in the cartographic object (A_{ψ_0}) and in the final active contour ($A_{\phi_{final}}$). Finally, this energy is made invariant by μ_{persp} translation only. This invariance is intended to avoid penalizing translations between the map and the active contour that are collinear with the perspective distortion direction \mathbf{V} . Remember that the satellite image is not rectified and there is thus a geometrical variation between the building's footprint (onto which the cartographic object is likely to be superimposed) and the roof segmented by the active contour. This geometrical distortion is intrinsic to the image and must not be interpreted as a change. In the Quickbird satellite image, distortion can be modeled by a translation whose direction is constant over the entire image. The orientation and direction of distortion is given by the unit vector \mathbf{V} we measured manually. The vector μ_{persp} is then defined as follows:

$$\mu_{persp} = w_{per} \mu_{final} \quad (7.10)$$

with μ_{final} the estimation of translation between the final active contour and the prior shape constraint. The prior shape is similar to the initial cartographic object and μ_{final} is estimated by the simplex algorithm (see section 5.2). The weight w_{per} favors invariance by μ_{final} only if the latter is collinear with \mathbf{V} :

$$w_{per} = \begin{cases} \left\langle \frac{\mu_{final}}{|\mu_{final}|}, \mathbf{V} \right\rangle^2 \mu_{final} & \text{if } \left\langle \frac{\mu_{final}}{|\mu_{final}|}, \mathbf{V} \right\rangle > 0 \\ 0 & \text{otherwise} \end{cases} \quad (7.11)$$

Local and parallel variations of the active contour (see chapter 6) with respect to the prior cartographic shape are also indicative of map/image inconsistencies. In this case, it is, however, difficult to discern whether these map/image local variations (at the scale of the cartographic object) are due to an effective change represented in the image or to an error present in the map. Because this problem is ill-posed, we decided not to compare the final active contour having allowed the fine matching with the cartographic object (whether transformed or not).

7.4.3 Probability of non-change

We propose to merge the Hough voting information with the geometrical variation of the active contours for each building j symbolized in the map. The total energy $E_{NC}(j)$ measuring the non-change is then:

$$E_{NC}(j) = \lambda_{Hough} E_{Hough}(j) + \lambda_{\mu_{Hough}} E_{\mu_{Hough}}(j) + \lambda_{geom} E_{geom}(j) \quad (7.12)$$

where $\lambda_{Hough}, \lambda_{geom}$ are positive real constants. Using the formalism of Gibbs distributions to define a non-change probability density $p_{NC}(j)$ associated with $E_{NC}(j)$, we can write:

$$p_{NC}(j) = \frac{1}{Z} e^{-E_{NC}(j)} \quad (7.13)$$

where Z is the partition constant. By defining $p_{NC}(j) + p_C(j) = 1$ (where $p_C(j)$ is the change probability), the constant Z will then take value 1. The probability $p_{NC}(j)$ is an indicator of non-change when it is greater than 0.5, otherwise a change is detected.

7.5 Experimental results

In section 7.5.2 we show some experimental results for various cases of changes and non-changes in order to measure the method's effectiveness and to illustrate difficulties that cannot be handled. In section 7.5.3, we propose to statistically analyze the results generated on a greater number of cases and with maps of varying quality.

7.5.1 Experimental protocol

All the experiments conducted in this section were performed in the following conditions:

- * **Data:** maps from 1996 and 2002 were used. When a map bears the wording “perfect”, this means that errors in the map were corrected manually. In this specific case, the cartographic object is consistent with its representation in the image; we then simulate a matching error that can be modeled by a similarity transform of parameters $\xi_0 = (s_0, \theta_0, \mu_{x,0}, \mu_{y,0})$. The (panchromatic and multispectral) Quick-bird satellite images dating from 2002 were used in these experiments
- * **Procedure:** When the preliminary test based on the DSM-NDVI data confirms a change, fine matching by active contours and the Hough voting are not carried out. When a change could not be validated by the method based on the DSM-NDVI, we illustrate the initial state of the cartographic object superimposed on the image as well as the final state after matching (illustrations only concern section 7.5.2).
- * **Initialization of active contours:** the prior shape as well as the initial active contour are similar to the cartographic object. The cartographic object is registered on the more recent satellite image using geocoding information.
- * **Hough voting:** the size of the accumulator used in experiments was:
 $(\mu_x = -30 \dots 30 \text{ pixels}, \mu_y = -30 \dots 30 \text{ pixels})$. The width l of the segments added to the accumulator was $l = 2 \text{ pixels}$.
- * **Vectorization:** in order to be able to take part in the Hough vote, the final active contour having permitted fine matching must be discretized in the form of segments. This active contour was originally a list of points taken from zeros detection of level set representing it. These points are firstly ordered in a list $\{\mathbf{C}(i)\}$, then chained to form a sequence of segments. The chaining algorithm involves forming segments $[\mathbf{C}(j), \mathbf{C}(j+k)]$, as long as the distance to the segment of points $\{\mathbf{C}(i)\}$ located between $\mathbf{C}(j)$ and $\mathbf{C}(j+k)$ is less than a threshold ε_{vect} . In our application, the fineness of vectorization is decisive for an effective matching between the refined cartographic object and the segments extracted from the image. We have thus chosen a restrictive threshold $\varepsilon_{vect} = 0.3$.

* **Non-change energy parameters:** change detection parameters used are: $\lambda_{Hough} = \lambda_{\mu_{Hough}} = 1$, $\lambda_{geom} = 2$ and $k = 0.01$. The value of λ_{geom} reflects the prior knowledge available on the quality of map/image initial registration. A high weight means a high confidence in quality, and thus a weak variation in E_{geom} will be penalized by a significant increase in total energy of non-change to finally reduce the probability p_{NC} . The stiffness of the spring k can be set in accordance with the uncertainty associated with estimation of μ_{Hough} . Remember that the segments of the Hough accumulator are broadened by l pixels, which deteriorates the quality of the translation estimation. We choose a low value for k so as not to penalize the final change decision with respect to these inaccuracies.

* **(No)-change decision:**

- We consider a change to be proven if $s_{fusion,MNS-NDVI} = 1$ or $p_{NC} < 0.4$.
- We consider detection of a non-change if $s_{fusion,MNS-NDVI} = 0$ and $p_{NC} > 0.6$.
- We consider decision-making impossible (indetermination) if $s_{fusion,MNS-NDVI} = 0$ and $p_{NC} > 0.4$ and $p_{NC} < 0.6$.

7.5.2 Study of some cases of change/non-change

7.5.2.1 Change detection

The figures below illustrate correct change detections conducted with the digital map of 1996. In figures 7.10.a and 7.10.b, the absence of buildings is validated by the change detection based on the use of the multispectral image. These are unequivocal cases of change where buildings have been completely demolished and replaced by ground covered with vegetation (case *a* of table 2.2). In both cases, classification by maximum likelihood identifies the areas of the image included in the cartographic polygon, as with vegetation, which negates the presence of buildings: $s_{fusion,MNS-NDVI} = 1$.

The two following cases of change were simulated (cartographic objects created manually). Decision-making is impossible through pre-processing ($s_{fusion,MNS-NDVI} = 0$) and consequently results from the method comparing only the map and the panchromatic satellite image. Figure 7.11 illustrates the contribution that taking E_{geom} into account in the decision-making model makes. In the illustrated case, the final Hough score is extremely high as the new building on the image has the same shape as the cartographic object, which tends to confirm (wrongly) a non-change. However, the spatial variation is too great to be ascribed to exogenous variabilities, E_{geom} is thus high and the building is detected as having changed ($p_{NC} = 0.3$).

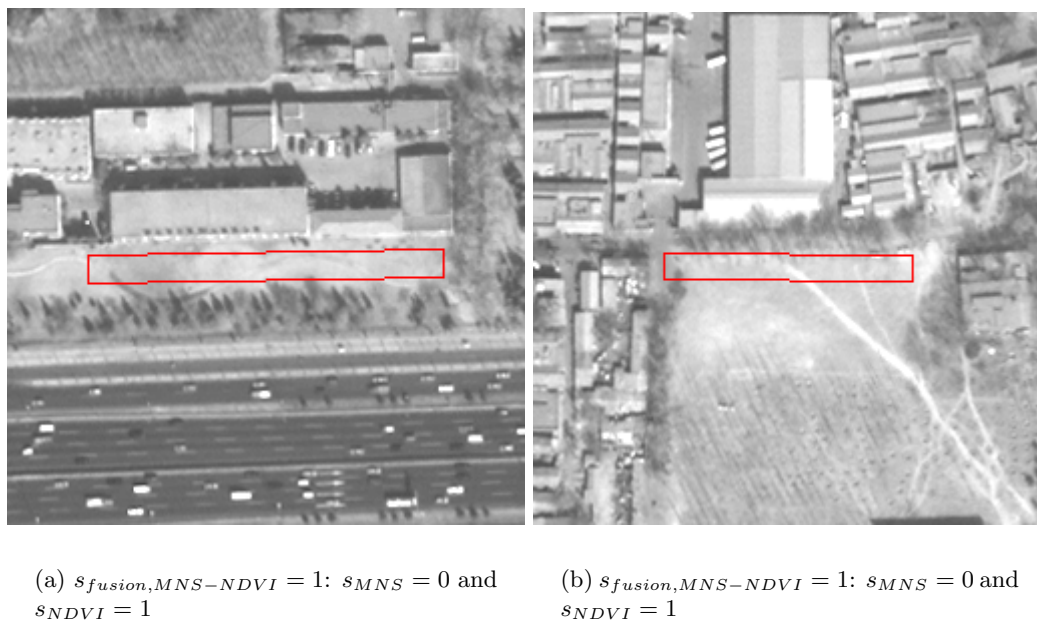


Figure 7.10: Detection of unequivocal changes. Changes were detected by comparison with the NDVI image. The figures illustrate cartographic objects superimposed on a more recent satellite image.

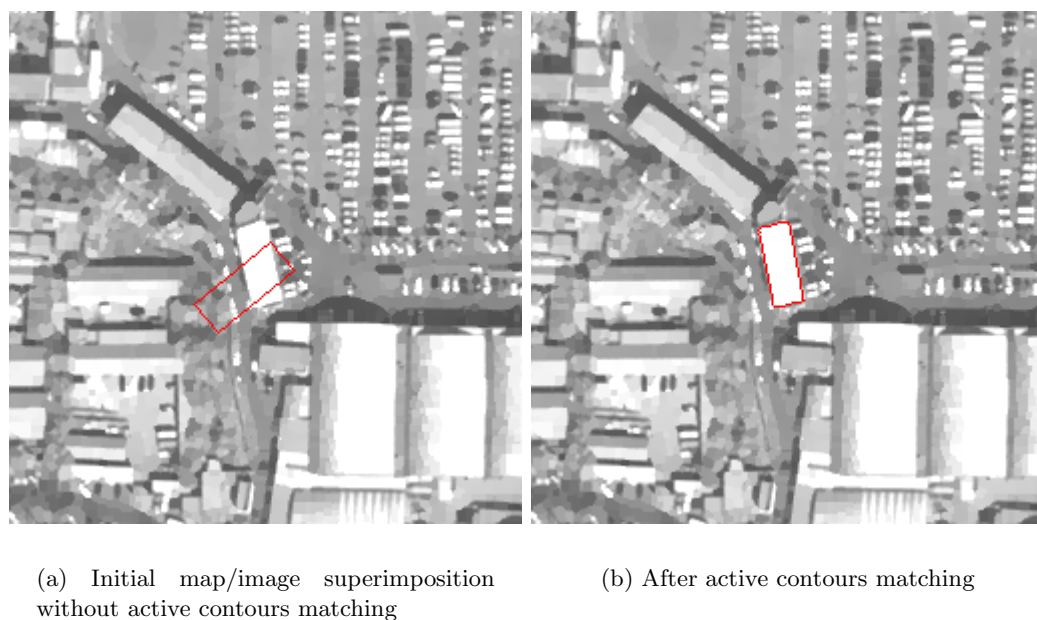
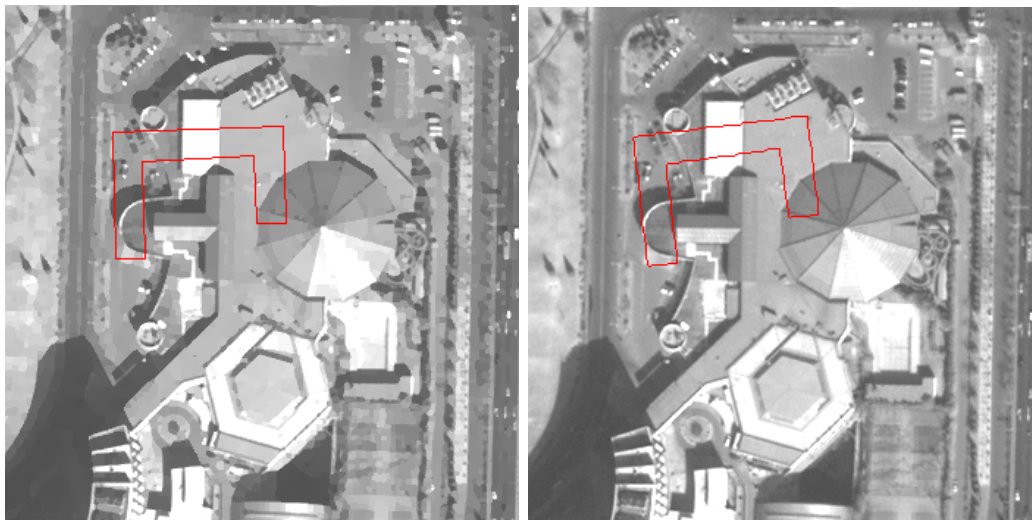


Figure 7.11: Change detection due to strong geometrical variation: $E_{geom} = 0.54$ and $p_{NC} = 0.30$. Change is simulated.

Finally, figure 7.12 illustrates another type of simulated and correctly detected change where one building is replaced by another. In this case, matching by active contours fails since the cartographic building no longer exists in the image. As the initial building has disappeared, few segment primitives on the image back up the hypothesis of the building being present under the plot of the cartographic object modified by the active contour. Thus, the Hough vote designates another location where the segments of the image are more consistent with the cartographic representation. While this remote location ($\mu_{\text{Hough}} = (-6, -26)$) is a better candidate, the voting score remains low $s_{\text{Hough}} = 0.27$ since no similar building is found in the image. The combination of a low Hough score with a translation reckoned to be significant makes p_{NC} tend drastically towards 0.



(a) Initial map-to-image superimposition without active contours matching

(b) Failed matching

Figure 7.12: Change detection due to a high μ_{Hough} translation. $s_{\text{Hough}} = 0.27$, $\mu_{\text{Hough}} = (-6, -26)$ and $p_{NC} = 0.00$. Change is simulated.

7.5.2.2 Detection of non-change

Perfect map

We propose to conduct non-change tests with a perfect map dating from 1996. In this map, the cartographic objects have been corrected to be similar to their representation on the image. A high probability of non-change is then guaranteed if the map-to-image matching is effective. Figures 7.13 to 7.16 illustrate an effective reduction of map/image variabilities that allows consistency between these two representations to be enhanced. The Hough final score is thus increased considerably compared with that for the initial state (table 2.2). This results in correct non-change detections of buildings. In figure 7.13, a rotation by 0.1 rad is applied to the cartographic object, simulating

a map-to-image superimposition error. Fine matching by active contours is effective and increases the final Hough vote score ($s_{Hough} = 0.76$). The probability $p_{NC} = 0.63$ confirms a non-change with nevertheless a moderately high value. This is due to a non-negligible geometrical variation $E_{geom} = 0.1$ that brings down the probability value. What figure 7.14 shows, illustrates the same phenomenon: a relatively high final Hough score ($s_{Hough} = 0.76$) with a low estimated translation ($\mu_{\mathbf{Hough}} = (1, 0)$) confirm a non-change. Nevertheless, the geometrical variation incurs the value for probability of non-change being brought down: $p_{NC} = 0.63$. With the experiment of figure 7.15, the

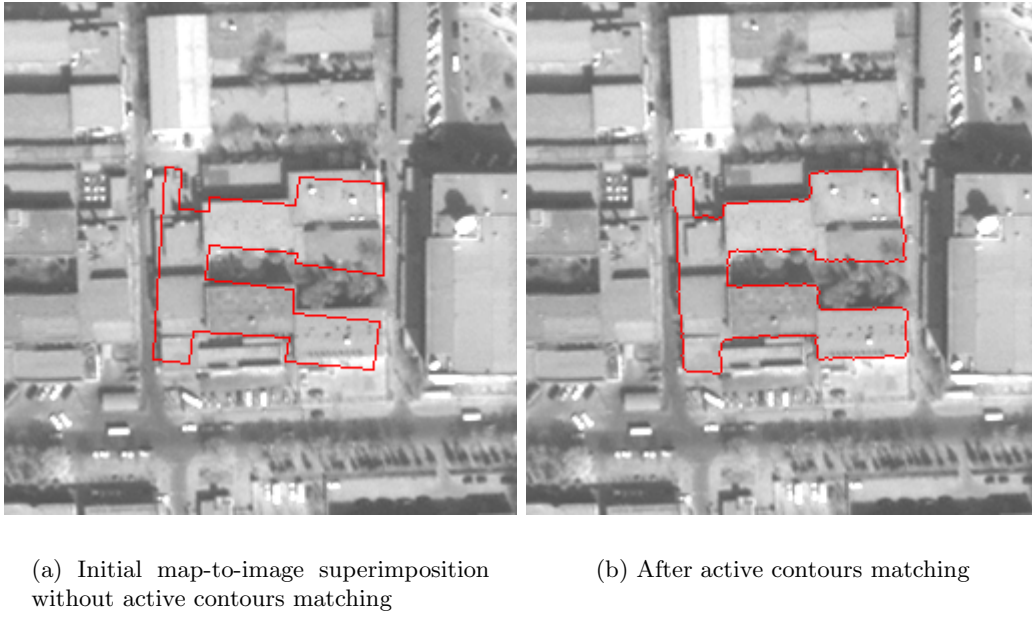
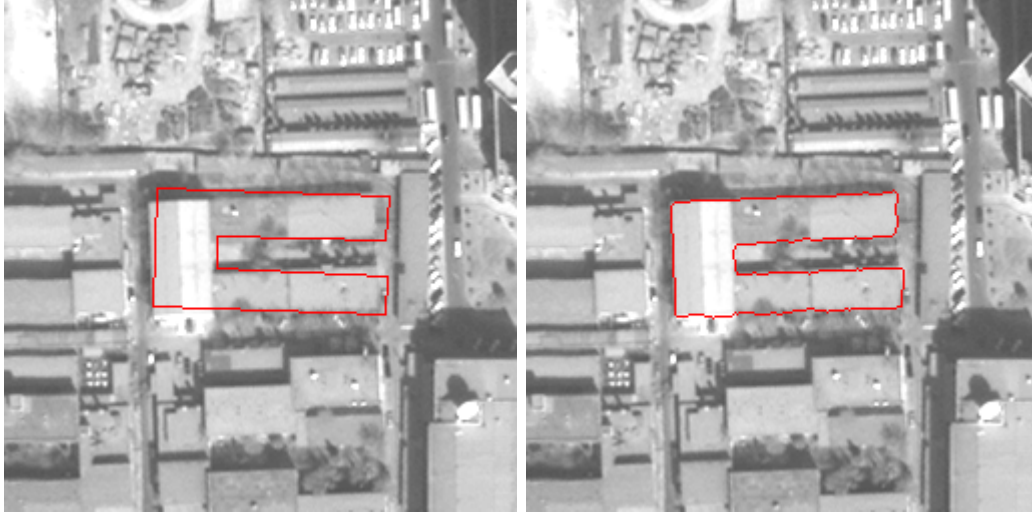


Figure 7.13: Detection of non-change with a perfect cartographic object transformed by similarity of parameters $\xi_0 = (s_0 = 1, \theta_0 = 0.1, \mu_{x,0} = 1, \mu_{y,0} = 1)$. Non-change is confirmed with $s_{Hough} = 0.76$, $\mu_{\mathbf{Hough}} = (1, 0)$, $E_{geom} = 0.1$ and $p_{NC} = 0.63$.

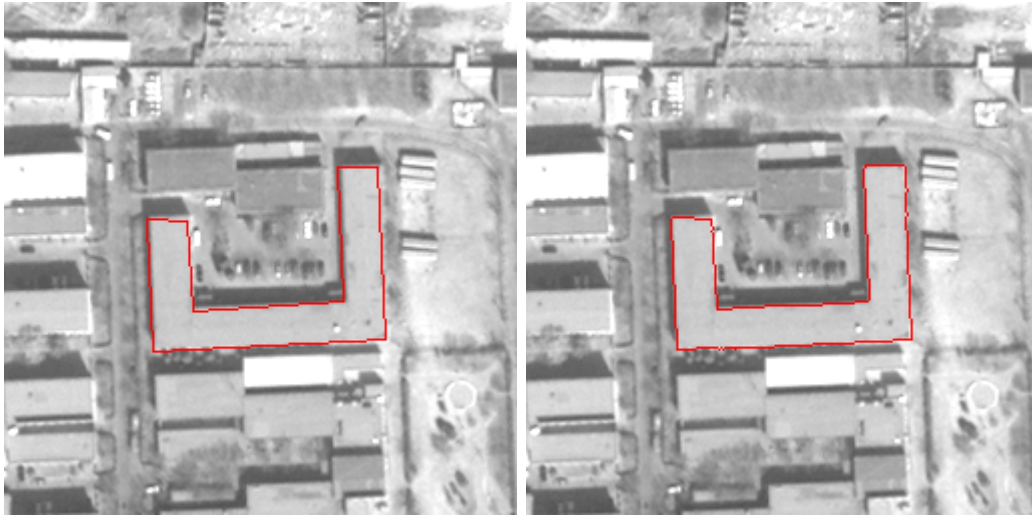
non-change presumption is higher ($p_{NC} = 0.76$) because of a more significant Hough score and a weaker geometrical variation.



(a) Initial map-to-image superimposition without active contours matching

(b) After active contours matching

Figure 7.14: Detection of non-change with a perfect cartographic object transformed by similarity of parameters $\xi_0 = (s_0 = 1.05, \theta_0 = 0.075, \mu_{x,0} = 4, \mu_{y,0} = -1)$. Non-change is confirmed with $s_{Hough} = 0.72$, $\mu_{Hough} = (1, 0)$, $E_{geom} = 0.08$ and $p_{NC} = 0.63$.

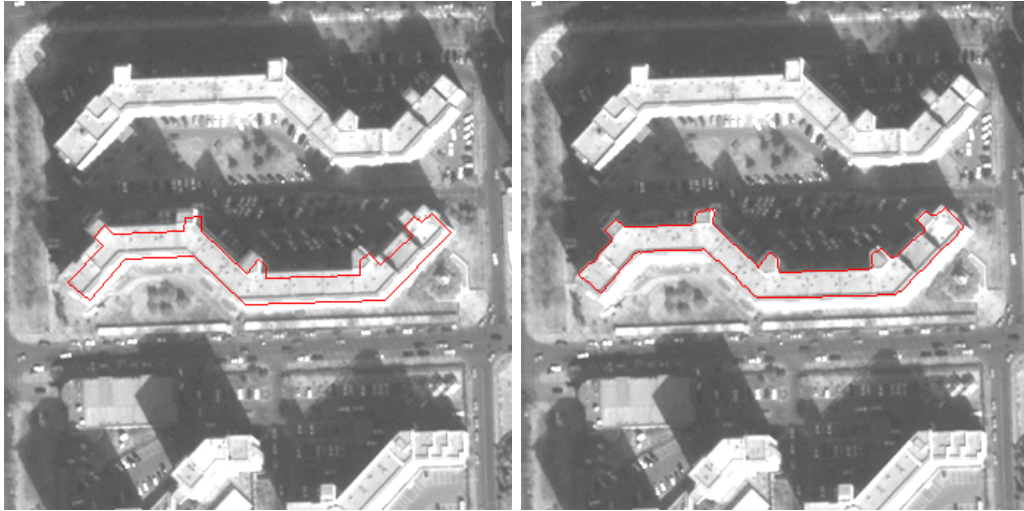


(a) Initial map-to-image superimposition without active contours matching

(b) After active contours matching

Figure 7.15: Detection of non-change with a perfect cartographic object transformed by similarity of parameters $\xi_0 = (s_0 = 1.0, \theta_0 = -0.02, \mu_{x,0} = -2, \mu_{y,0} = 1)$. Non-change is confirmed with $s_{Hough} = 0.81$, $\mu_{Hough} = (0, 0)$, $E_{geom} = 0.04$ and $p_{NC} = 0.76$.

Finally, the result of figure 7.16 shows the need not to penalize collinear translations with the direction \mathbf{V} of perspective distortion. Figure 7.16.a shows a cartographic object close to the footprint of a moderately high building in the image. Distortion is weak enough for the active contour to be capable of segmenting the roof. The translation vector between the initial state and final matching is near collinear with \mathbf{V} . Thus, the geometrical variation is not penalizing ($E_{geom} = 0.01$). The Hough score is high ($s_{Hough} = 0.84$) with low translation $\mu_{\mathbf{Hough}} = (-1, 0)$, which explains the high value of non-change probability: $p_{NC} = 0.83$ (which is also due to the perfect nature of the map).



(a) Initial map-to-image superimposition without active contours matching

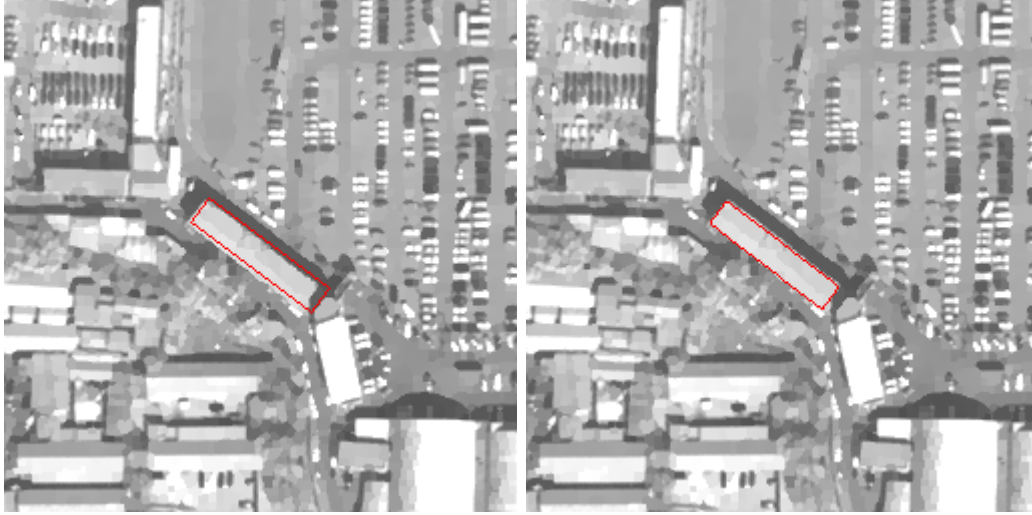
(b) After active contours matching

Figure 7.16: Detection of non-change with a perfect cartographic object transformed by similarity of parameters $\xi_0 = (s_0 = 1.0, \theta_0 = 0.0, \mu_{x,0} = 5, \mu_{y,0} = 5)$. Non-change is confirmed with $s_{Hough} = 0.84$, $\mu_{\mathbf{Hough}} = (-1, 0)$, $E_{geom} = 0.01$ and $p_{NC} = 0.83$.

Real map

We shall now a real map made in 1996, without manual correction of potential errors. The errors with respect to the image can either be modeled by a global (at the scale of the cartographic object) or local transformation. In the case of local shape errors of cartographic objects, the matching by active contours then provides a trade-off between the erroneous prior knowledge derived from the map and the reality of the image. Figures 7.17 and 7.18 show correct results of non-change detection in the case of matching errors that can be modeled by global transformations. Fine matching by active contours then allows for an almost perfect superimposition between the cartographic object and the building in the image. The consistency between the two representations is enhanced and, as a result, the Hough score is improved after the fine matching. The

probability p_{NC} is lower in the case of figure 7.17 due to a more significant geometric variation ($p_{NC} = 0.68$ in figure 7.17 and $p_{NC} = 0.83$ in figure 7.18). Figures 7.19 to 7.21

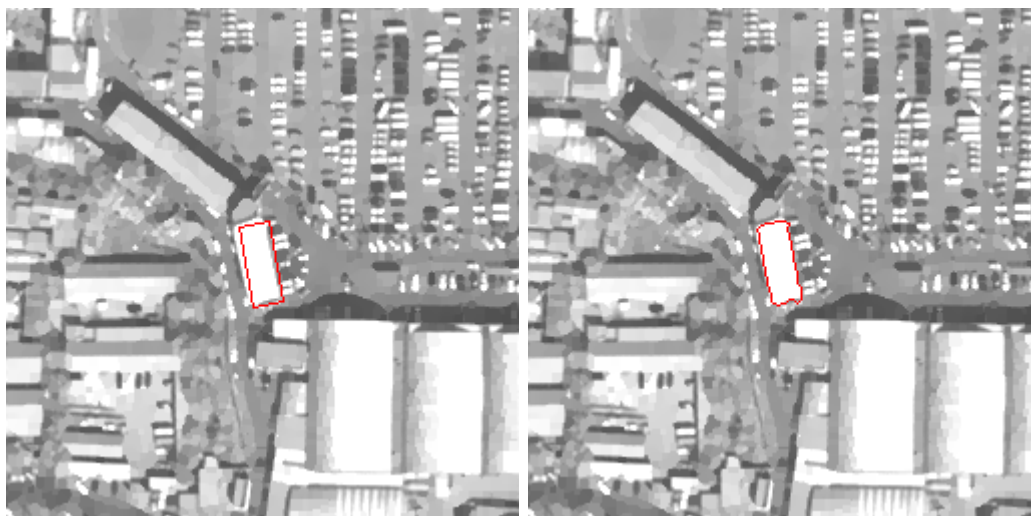


(a) Initial map-to-image superimposition without active contours matching

(b) After active contours matching

Figure 7.17: Detection of non-change with a matching error that can be modeled by a global transformation. Non-change is confirmed with $s_{Hough} = 0.86$, $\mu_{Hough} = (0, -1)$, $E_{geom} = 0.12$ and $p_{NC} = 0.68$.

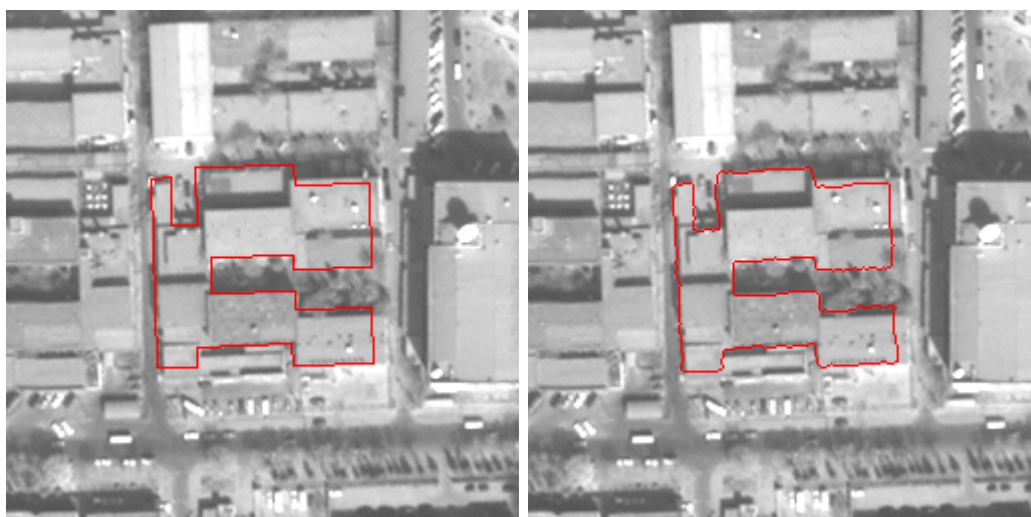
illustrate results of experiments performed with cartographic objects containing local shape errors of low amplitude. We can see that for each of these experiments, the result of fine matching improves data consistency, although local errors persist and cannot be lessened. The presence of these errors reduces the Hough voting score. Indeed, some parts of the refined cartographic object do not coincide with the gradients of the image and thus cannot receive the segments contribution for the estimation of μ_{Hough} . In the present case, local errors are in a minority and are of low amplitude. They do not impair the Hough voting to the extent of making the change decision indeterminate or incorrect ($p_{NC} = 0.66$ in figure 7.19, $p_{NC} = 0.71$ in figure 7.20 and 7.21). These examples illustrate the robustness of the proposed approach with respect to minor local shape errors present in the map. This robustness is manifest both in the matching process and during the Hough poll.



(a) Initial map-to-image superimposition without active contours matching

(b) After active contours matching

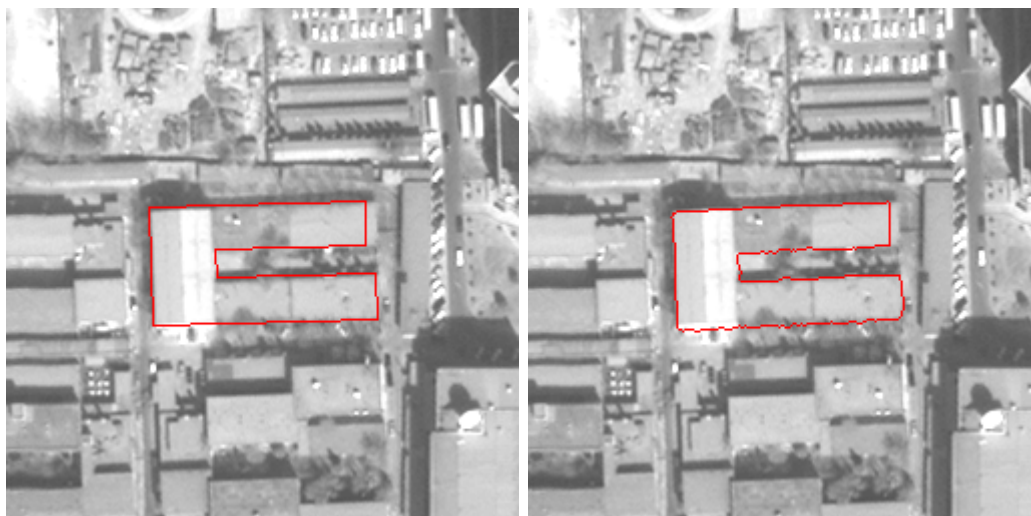
Figure 7.18: Detection of non-change with a matching error that can be modeled by a global transformation. Non-change is confirmed with $s_{Hough} = 0.88$, $\mu_{\mathbf{Hough}} = (0,0)$, $E_{geom} = 0.03$ and $p_{NC} = 0.83$.



(a) Initial map-to-image superimposition without active contours matching

(b) After active contours matching

Figure 7.19: Detection of non-change with local shape errors. Non-change is confirmed with $s_{Hough} = 0.62$, $\mu_{\mathbf{Hough}} = (1,0)$, $E_{geom} = 0.02$ and $p_{NC} = 0.66$.



(a) Initial map-to-image superimposition without active contours matching

(b) After active contours matching

Figure 7.20: Detection of non-change with local shape errors. Non-change is confirmed with $s_{Hough} = 0.73$, $\mu_{Hough} = (1, 0)$, $E_{geom} = 0.03$ and $p_{NC} = 0.71$.



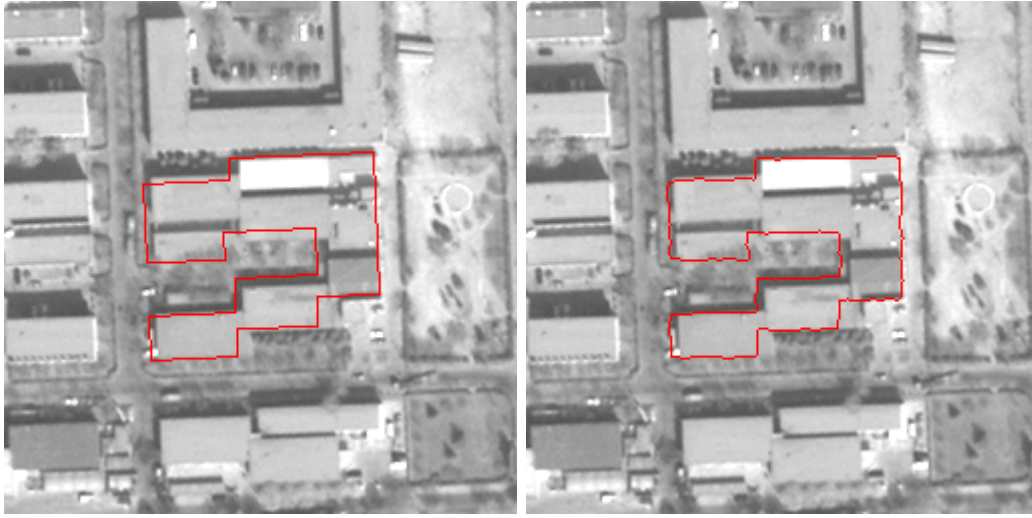
(a) Initial map-to-image superimposition without active contours matching

(b) After active contours matching

Figure 7.21: Detection of non-change with local errors of shape. Non-change is confirmed with $s_{Hough} = 0.75$, $\mu_{Hough} = (1, 0)$, $E_{geom} = 0.05$ and $p_{NC} = 0.71$.

7.5.2.3 Ambiguity and false alarm with a map marred by local errors

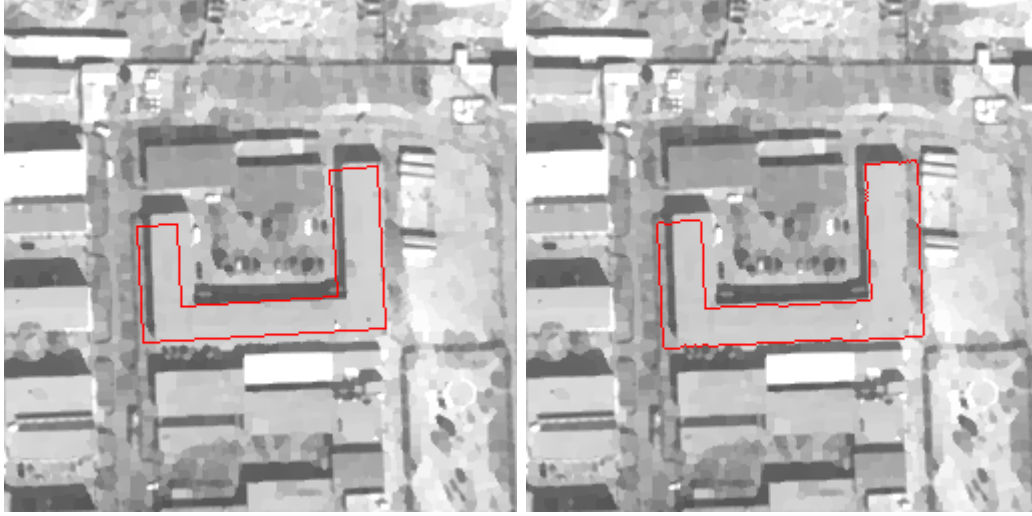
When local shape errors in the map are more severe, final decision-making becomes either ambiguous (see figure 7.22 with a probability close to 0.5, $p_{NC} = 0.55$), or false (see figure 7.23 where $p_{NC} = 0.39$ whereas no change has occurred). In case more significant shape errors exist in the map, map/image inconsistencies will remain after the active contours matching and will be too numerous to allow a significant Hough score. Hough scores for the experiments of figures 7.22 and 7.23 are equivalent ($s_{Hough} = 0.45$ and $s_{Hough} = 0.47$) and not as good as those obtained with minor shape errors (see figures 7.19-7.21). While in the experiment of figure 7.22, the final decision is indeterminate, in the second case, an erroneous detection of change is produced due to a more significant geometrical variation ($E_{geom} = 0.19$).



(a) Initial map-to-image superimposition without active contours matching

(b) After active contours matching

Figure 7.22: Ambiguous detection (indeterminate case). $s_{Hough} = 0.45$, $\mu_{Hough} = (0, 0)$, $E_{geom} = 0.02$ and $p_{NC} = 0.55$.



(a) Initial map-image superimposition without active contours matching

(b) After active contours matching

Figure 7.23: Erroneous change detection due to local shape errors on the map. $s_{Hough} = 0.47$, $\mu_{\mathbf{Hough}} = (-1, -1)$, $E_{geom} = 0.19$ and $p_{NC} = 0.39$.

These two cases illustrate the limits to the proposed model and its inability to overcome too significant local shape errors. This is inherent to the fine matching process that does not allow significant local variations of the active contour with respect to the shape derived previously from the map. The works exposed in chapter 6 are an attempt to alleviate this problem and should thus allow a high Hough score to be obtained despite local errors, as long as they can be modeled by parallel variations with respect to the reference shape.

7.5.2.4 Summary of results

Table 7.1 summarizes the quantities calculated for the detection of change/non-change in the experiments shown.

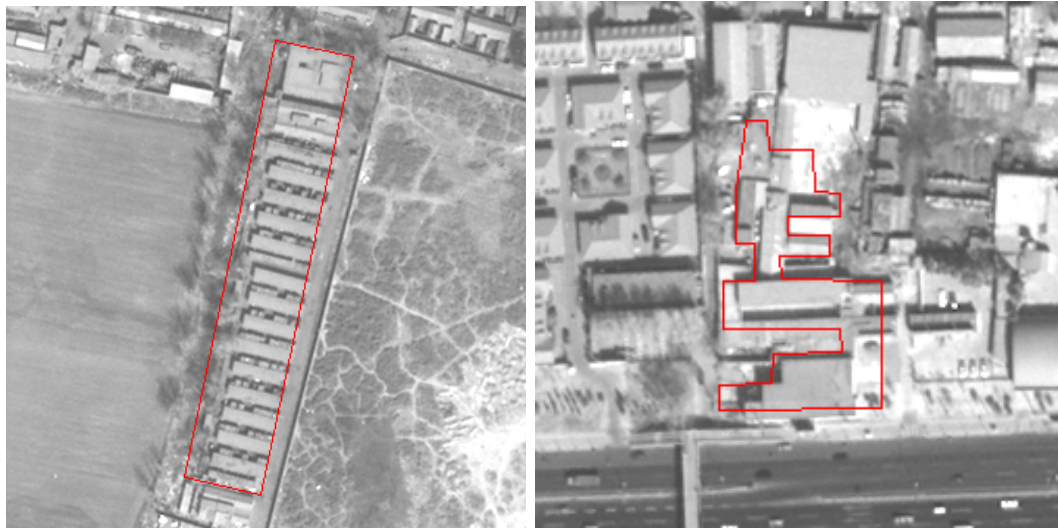
| | <i>Experiment</i> | Before matching | | After matching | | | |
|---------------|-------------------|-----------------|-------------|----------------|---------------|------------|----------|
| | | s_{fusion} | s_{Hough} | s_{Hough} | μ_{Hough} | E_{geom} | p_{NC} |
| Change | figures 7.10.a-b | 1 | / | / | / | / | / |
| | figure 7.11 | 0 | 0.22 | 0.91 | (0, -1) | 0.54 | 0.30 |
| | figure 7.12 | 0 | 0.12 | 0.27 | (-6, -26) | 0.13 | 0.00 |
| Non-change | figure 7.13 | 0 | 0.34 | 0.76 | (1, 0) | 0.1 | 0.63 |
| | figure 7.14 | 0 | 0.30 | 0.72 | (1, 0) | 0.08 | 0.63 |
| | figure 7.15 | 0 | 0.54 | 0.81 | (0, 0) | 0.04 | 0.76 |
| | figure 7.16 | 0 | 0.69 | 0.84 | (-1, 0) | 0.01 | 0.83 |
| | figure 7.17 | 0 | 0.74 | 0.86 | (0, -1) | 0.12 | 0.68 |
| | figure 7.18 | 0 | 0.78 | 0.88 | (0, 0) | 0.03 | 0.83 |
| | figure 7.19 | 0 | 0.33 | 0.62 | (1, 0) | 0.02 | 0.66 |
| | figure 7.20 | 0 | 0.35 | 0.73 | (1, 0) | 0.03 | 0.71 |
| | figure 7.21 | 0 | 0.52 | 0.75 | (1, 0) | 0.05 | 0.71 |
| Indeterminate | figure 7.22 | 0 | 0.35 | 0.45 | (0, 0) | 0.02 | 0.55 |
| False alarm | figure 7.23 | 0 | 0.41 | 0.47 | (-1, -1) | 0.19 | 0.39 |

Table 7.1: Quantification of change/non-change scores. p_{NC} is the probability of non-change. s_{Hough} is the normalized Hough score.

We observe that the decrease of exogenous variabilities thanks to the active contours matching is effective. Indeed, the s_{Hough} score is always improved after the deformable models' processing. In the case of non-change, the improvement varies between 10% and 40%.

7.5.3 Statistical analysis

In this section, we analyze the success of the proposed method over a greater number of cases of changes/non-changes from available cartographic data. A first set of tests was conducted with a good quality map. What we understand by "good quality" concerns a low rate of local shape errors for the cartographic objects, as well as only slight effects of generalization and simplification. A second set of tests was performed with a medium quality map. In this map, severe local shape errors exist. They are either due to poor delineation of the buildings seen in the remote sensing images that were used to generate the map, or to significant simplification/generalization effects (see figure 7.24).



(a) Generalization effect

(b) Generalization, delineation errors and presence of partial changes

Figure 7.24: Illustration of delineation errors and of generalization effect in the map of 1996. The cartographic object is superimposed on the satellite image of 2002.

To perform active contours matching, we chose the edge-based data attachment model (GVF, see section 4.3.2.2). Indeed, this is the model capable of processing the largest number of cases due to its insensitivity to the homogeneity of roofs and the generalization effect of the map. We use the spatio-temporal shape constraint with $\lambda_{min} = 1$, $\lambda_{max} = 3$ and $d_0 = 2$ (see section 5.4).

7.5.3.1 Results with a good quality map

The map used was produced by the Beijing Institute of Surveying and Mapping (BISM) from the Quickbird satellite image of 2002 used in the experiments. This led to enhanced map/image consistency although effects of simplification, generalization and error were present. In addition, cases of change must not be detected with such a map. We conducted a first set of tests with 40 buildings aimed at quantifying the method's performance with respect to non-changes. Secondly, we translated the map by 30 meters in the eastern direction. The map and the image then no longer matched, so we could detect changes only (out of 37 cases). The results were then brought together and classified in accordance with the following nomenclature:

1. Correct non-change detection (*true positive*)
2. Correct change detection (*true negative*)
3. Erroneous change detection (the method considers the building to be changed whereas in reality it remains unchanged): *false negative*

4. Erroneous non-change detection (false alarm: the method considers the building to be unchanged, whereas it has disappeared from the image): *false positive*
5. Indeterminate case

The results derived from the 2002 map are shown in table 7.2. We observe that the rate of indetermination relating to the method is reasonably high. Thus, 87% of cases are assigned a decision of change/non-change. Among all cases considered, 5.2% were wrongly interpreted as being changes. These cases of missed non-change detections were small in number and can be explained by the presence of local shape errors in the map. As we saw in section 7.5.2.3, the presence of severe errors weakens the Hough voting score to the extent of making the decision indeterminate or erroneous. These errors are responsible for missed detections and indeterminate cases. The remainder of cases represents 81.8% of correct detections of changes and non-changes.

| | |
|--|---|
| Correct non-change detection: 42.8% | Erroneous non-change detection: 0% |
| Erroneous change detection: 5.2% | Correct change detection: 39% |
| Indeterminate cases: 13% | |

Table 7.2: Quantification of change detection results with the 2002 map. Results are expressed as a percentage over a total of 77 cases.

7.5.3.2 Results with a medium quality map

The medium quality map we considered dates from 1996. As the time difference between the map and the satellite image (2002) was fairly significant, we did not resort to simulation of changes as for the previous tests. The results obtained from 105 cases are shown in table 7.3. Compared with the results derived from the 2002 map, we can see a slight increase in indeterminate cases (14.3%). However, the biggest increase recorded was the rate of missed detections of non-change that reached 17%. This high rate is consistent with the experimental conditions: the used map included more severe errors than that from 2002. Awkward cases of non-change were not detected and instead of switching to the category of indeterminate cases, they were wrongly interpreted as being changes. In addition to the map errors, analyzed cases from this second set of tests was a more delicate matter: the shadow cast by high buildings could completely mask the building of interest, thus making it undetectable; the canopy could totally cover some smaller buildings. As with the case of the 2002 map, erroneous non-change detections remained negligible. They arose when segments of the image were consistent with the geometry of the building in the map, which is no longer present in the image. Such cases occurred but were extremely rare. Among all the considered cases, the rate of correct change/non-change detections came to 66.7%, which amounts to a rate of 77.8% among cases where there was decision-making.

| | |
|--|---|
| Correct non-change detection: 50.5% | Erroneous non-change detection: 2% |
| Erroneous change detection: 17% | Correct change detection: 16.2% |
| Indeterminate cases: 14.3% | |

Table 7.3: Quantification of change detection results with the 1996 map. Results are expressed as a percentage over a total of 105 cases.

7.6 Conclusion

We introduced a methodology for change detection between a map and remote sensing data. This approach was broken down into two stages and took advantage both of data fusion and reducing exogenous map/image variabilities using the active contours technique. The first phase involved calculating and merging change indices respectively based on comparison of the map with a Digital Surface Model (DSM) and a multi-spectral image. The methods implemented were simple and quick but only permitted detection of an unequivocal change in the urban construction fabric.

When this was incapable of detecting the absence of a considered cartographic building, a second phase compared it with its representation in a panchromatic satellite image. Map-to-image superimposition errors were first reduced by fine matching that was performed using active contours. A subsequent comparison between the panchromatic image and the refined cartographic object was made using two measurements: one measurement by Hough voting (that quantified consistency between the segments of the object and those extracted from the image) and another measurement of geometrical deformation between the cartographic object before and after fine matching. Combining these two measurements with a Gibbs energy, a probability density for non-change was formulated and evaluated for each building of the map.

Experimental results illustrated the method's efficiency on cases where local errors in the shape of the cartographic objects were minor. A statistical analysis demonstrated that in these circumstances, 82% of cases of change/non-change were correctly processed. We were thus able to observe that the active contours fine matching by could significantly increase map/image consistency to improve the level of confidence ascribed to the non-change decision. The presented results also highlighted limitations in the proposed approach: the presence of severe errors in the map meant the failure of the fine matching process and did not allow for a significant Hough voting. Using statistical analysis, we showed that in this case, the rate of missed detections of non-change increased significantly (+12%).

In conclusion, the proposed approach allowed the detection of "obvious" building changes or to measure the similarity between the cartographic object and its representation in the image. A low level of similarity characterizes a difference without us being able to determine whether it derived from a real change or an error in the cartographic

plot.

Chapter 8

Conclusion

8.1 Summary

The goal of this thesis was to develop a method for analyzing changes between a digital urban map of buildings and multi-source high-resolution remote sensing data. We have focused particularly on resolving the problem of exogenous variabilities between the cartographic objects and their representations in the image. These variabilities are due either to map-to-image registration imprecision, or to errors when plotting the cartographic object, or to the maps simplification and generalization effects. These exogenous differences are independent from an effective change and must be attenuated in order to guarantee a reliable change detection. The correction of these map/image inconsistencies is made difficult because of the great number of artifacts present in the images and which is specific to our application focused on urban environments.

The insertion of a shape constraint derived from the map in the active contour models has shown its effectiveness to overcome some of these artifacts. This prior knowledge extracted from the map is also specific to the considered scene and enables to free ourselves from the geographic place under study. We have experimentally demonstrated that in the presence of shape constraints, the active contours were sensitive to an initialization far from the target object. This has led us to propose two solutions aiming to drive the active contour more surely towards the solution and avoid the problem of local minima. The first one consists in merging the altitude information from a digital surface model (DSM) with the one from a panchromatic satellite image. The complementarity of these two representations has proved to be effective. The second solution consists in spatially and dynamically relaxing the shape constraint during the active contour's convergence process. An approximate and non-regular solution is first reached while the constraint is spatially relaxed. The later and spatially uniform restoration of the constraint enables to regularize the contour, thus avoiding the urban artifacts. We have also proposed new - linear and quadratic - shape constraint energies authorizing a class of deformations of the active contour with respect to the prior shape. These models enable the contour to undergo parallel variations with respect to the segments of the polygonal cartographic object and aims to correct the object's local shape errors. Although encouraging results are obtained with synthetic images, this latter technique is not yet mature and requires a deeper understanding of the models and of their imple-

mentations in order to make them efficient on real images.

Lastly, we have proposed a method for analyzing changes between the map and different remote sensing data. Change indexes, calculated simply from the DSM and from a multispectral satellite image, are merged to carry out the detection of unequivocal changes of buildings. In most cases where it has not been possible to identify a change in an evident way, a second more complex phase is carried out to characterize the degree of resemblance between the cartographic object and its representation in a panchromatic satellite image. This processing takes advantage of the fine matching by active contours and enables to formulate a non-change probability for each element of the map. This probability is derived from a Gibbs energy. It is calculated from the geometric deformation due to the fine matching and a Hough voting designed to measure the resemblance of the image segments and the refined cartographic object. A quantitative analysis has shown that the fine matching improves the robustness of the non-change decision. Tests carried out on a greater number of buildings have demonstrated statistically that the method's performances are satisfactory when the map only contains a small number of local shape errors.

This thesis work highlights the following points:

- * The spatio-temporal variation of the shape constraint weight has shown its efficiency for overcoming local minima due either to a uniform and high shape constraint, or to alterations of the image. It enables to move the critical choice of a constant shape constraint weight to a family of parameterized functions. The result of the segmentation is then less sensitive to the value of those parameters than to that of a constant value. However, our formulation of the spatio-temporal constraint is based on empirical observations and lacks theoretical foundations. Nevertheless, it may show the way for a more rigorous formulation leading to the determination of the spatio-temporal function according to the information derived from the image and from the prior shape and not by the definition of ad hoc profiles whose parameters setting represents a limitation.
- * The most innovative contribution of this work is certainly the one concerning the formulation of new energies aiming to reduce the local shape errors of the map. Nevertheless, numerous aspects relative to the numerical stability and the influence of the quadratic energy are still poorly understood. The results of this approach, which is not yet mature, are however encouraging.
- * The simplex optimization algorithm has shown its usefulness in two situations: the calibration of the camera for the generation of the DSM and the similarity parameters estimation making the active contours' shape constraint invariant from rotation, translation and scaling. In both cases, this method has been found to be efficient and robust. Furthermore, the simplex enables to avoid the linearization (calibration) or the derivation (gradient descent) of the cost function to be minimized. So, its use can be transposed to all types of functional, assuming that the number of parameters to be estimated is not too great. Lastly, we have demonstrated that the simplex, contrary to intuition, converges more rapidly than

the gradient descent. This fact can be explained by the simultaneous estimation of the parameters and thanks to the reduction of the calculation complexity by means of the narrow band technique.

- * The use that we make of active contours is rather paradoxical: whereas these tools are known for their ability to deform themselves, we have imposed a shape constraint on them inhibiting this flexibility property. So, the matching by active contours would be similar to a rigid matching which could have been completed, for example, by the presented Hough voting technique. However, despite a sufficiently strong shape constraint to compensate for the urban artifacts, the active contours keep a very local flexibility which finally makes them better performing than the rigid matching techniques.
- * The change indexes proposed in this thesis, based on bottom-up approaches, are relatively easy to calculate. The Hough score used to measure the similarity between the cartographic object and the image is robust and complementary to the top-down approach of active contours.
- * Lastly, we have been able to verify that the active contours are flexible and powerful segmentation tools, capable of integrating generic and specific geometric constraints and of accomplishing data fusion.

8.2 Discussion of the limitations

8.2.1 Limits of the shape constraint model

The most constrictive limitation of the proposed approach is certainly the inability of the active contours to free themselves from the prior shape derived from the map in order to correct the local shape errors that it contains. The shape constraint models used are invariant from direct plane similarity and thus enable to correct global inaccuracies of rotation, translation and scaling modulo local variations errors of very low amplitude depending on the weight allocated to the constraint with respect to the image information. In this study we have shown that the invariance by similarity was not capable of resolving the problem of local errors, which has led us to propose a more flexible way of imposing the shape constraint with the possibility of local and parallel deformations of the contour with respect to the prior shape. This flexibility is nevertheless restricted to a class of deformations: any attempt to incorporate a greater degree of flexibility into the model must be constrained and must obey certain rules failing which the active contour will be sensitive to the image artifacts.

Our choice of conferring a degree of freedom on the active contour that is restricted and controlled by the parallelism reflects our prior knowledge of the local shape errors that are most often found in the map. However, this knowledge is clearly insufficient for modeling all of the map-to-image exogenous variabilities. Indeed, these variabilities are either due to plotting imprecision or to the map simplification and generalization effects. In most cases, these variabilities obey the photo-interpreter's subjectivity more than they obey well-defined rules, and they therefore seem to be impossible to model

with non-heuristic methods. So, the active contour models proposed in this thesis will never be in a position to totally eradicate the cartographic objects local errors.

8.2.2 Limits of the change analysis

The cartographic local shape errors, when they are of low amplitude, do not have a great impact on the proposed change analysis method. We have nevertheless shown quantitatively that the presence of severe errors significantly increases the rate of erroneous change detections. This is the direct consequence of the active contours' inability to free themselves from local errors of great amplitude. So, in the case of notable cartographic errors, the comparison between the "refined" cartographic object and the image's segments by Hough voting produces a mediocre similarity score without being able to state whether this lack of resemblance is due to a change or to errors in the map.

8.2.3 Limits of the active contours

The utilization of constrained active contours requires the setting of weights determining the influence of the attachment to data and of the shape constraint. This setting is empirical which represents a limitation in the totally automatic nature of the method. We have nevertheless noted that a single setting makes it possible to obtain satisfactory results whatever the place analyzed on the image.

The choice of attachment to data model, however, appears to be more delicate. In this study we have shown that the region-based models had better performances, but only when applied to buildings with a homogeneous roof. The edge-based active contours are more sensitive to initialization but can be applied to all types of building. Being able to choose automatically between one or the other of the models would represent a gain in robustness and efficiency. This issue remains open, nevertheless, since determining in advance whether a building is homogeneous or not is a difficult task insofar as the initial map-to-image registration is imprecise.

The active contours are known for their sensitivity to initialization and their long convergence times. The first limitation obliged us to process only the low to medium height buildings in the image. The second one represents an obstacle to the operational character of the approach proposed in this thesis. Although the procedures had been optimized, the processing of a building varies from a few seconds to a few minutes; the calculation times obtained using the quadratic models are prohibitive.

8.3 Perspectives

Before closing this document, we would like to list some possible ways for future works:

- * We have just seen that the robustness of the map-to-image changes analysis depends on the ability of the active contours to successfully overcome the local shape errors in the cartographic data. We have also noted that an exhaustive correction of the errors is illusory. However, the total elimination of the errors is not essential for increasing the robustness of the change analysis. Indeed, the reduction of the

error's amplitude would already be sufficient for reducing the rate of erroneous detections of change. So, the work undertaken on the subject of quadratic energies should be pursued. Alternatively, it would be interesting to see to what extent the incorporation of a statistical model of the shape constraint (based on learning from errors) could improve the technique.

- * The non-change energy criteria formulated in our approach are global on the scale of the cartographic object. Indeed the Hough score is the result of the total contribution of the segment that took part in the poll; the geometric variation energy is an integral calculation on the image domain, and is therefore global. It would be judicious to carry out a more localized change/non-change analysis in order to validate the parts of the refined cartographic object which are coherent with the image. The incoherent parts, which are either due to errors in the map or to effective partial changes, would then be submitted to the photo-interpreter's examination.
- * It should be noted that the shadow information, characteristic of buildings, has not been used for change detection. This should be considered in order to increase the robustness of the final change/non-change decision.
- * With a view to improving the robustness of the active contours, it is conceivable to use the multispectral satellite image before the panchromatic image (with a resolution four times lower than that of the multispectral image). This would represent a multi-scale processing likely to reduce the calculation time. The color information can be used directly in the Bayesian model according to the works of M. Rousson [102]. Although color appears to be a characteristic that only moderately discriminates the building from the rest of the scene, this is an aspect that deserves investigation.
- * Lastly, the Hough voting technique can be used to place the cartographic object on the roof of tall buildings (great perspective effect) before applying the finer and more precise matching of the active contours. The approach based on Hough voting is non-local and fast, and would enable to generalize the proposed method to all types of buildings. It would nevertheless have to be extended to the invariance from rotation and scaling to take into account the initial map-to-image registration imprecision.

We hope that this thesis work will have made a modest contribution to the (semi-)automated analysis of urban scenes for supporting and helping photo-interpreters in their map updating task. Beyond any immediate application, we hope that this work will be of benefit to the society and to the sustainable development of urban areas which are in the process of bringing together nearly half of the world's population.

Appendix A

Data presentation and pre-processing

In this section we propose to describe the cartographic and remote sensing data used for this PhD study. These data represent the whole city of Beijing at different dates. The remote sensing images are optical and are of two types. The first one concerns stereoscopic pairs of aerial images from which we will generate a Digital Surface Model (DSM) of the scene. The other type of remote sensing data concerns very high resolution satellite imagery. We will present the characteristics of the images acquired by the Quickbird satellite which have the best resolution at the present time in the civil sector. Lastly, the cartographic data used are available in digital form and are taken from a geographic information system (GIS). We will only describe the data relative to constructions.

A.1 Data presentation

A.1.1 Remote sensing data

A.1.1.1 Aerial images

We have three images of Beijing covering the zone of the future village for the 2008 Olympic Games (figure A.1). The images are analog and represent a scene at a scale of 1:10000. They were then scanned by the Beijing Institute of Survey and Mapping (BISM) at a pitch of 21 microns. The resulting digital images therefore have a resolution of 21 cm per pixel. The other characteristics of these images are as follows:

- * acquisition date: autumn 1999
- * acquired by an analog "RC 30 Leica Systems Camera"
- * conical geometry optical sensor
- * multispectral images (3 bands: RGB)
- * image dimensions: 11,000 by 11,000 pixels (i.e. a zone covered of $2.3 \times 2.3 \text{ km}^2$)

- * the optical system ensuring photo acquisition has negligible distortion
- * no GPS or gyroscopic geo-referencing system for determining the image acquisition parameters.



Figure A.1: Aerial image shown at 5% of its original size. The future Olympic village will be situated in the north-east of the image.

Several remarks must be made further to the description of these characteristics. The first one concerns the large size of the images. With a view to using them with the 3D reconstruction software to generate a DSM, they will have to be subsampled with the risk of being confronted with prohibitive calculation times (a DSM generated by the software results from the correlation of two stereographic images). Time is not the

only argument concerning the choice of subsampling. Indeed it will enable to reduce the noise inherent to the scanning of the photograph and the level of detail which is a source of error when generating a DSM with a high level of detail.

The reconstruction software imposes another constraint concerning the aerial images. They must be single-band images. We have chosen to convert the original three-band (RGB) images by calculating the mean of the three bands for each pixel considered. The resulting single-band image is thus as close as possible to the panchromatic spectrum. This conversion is carried out before the subsampling of the images.

A.1.1.2 Satellite images

We have two raw Quickbird images (DigitalGlobe “basic bundle” product) acquired in 2002 and 2003 which have the following characteristics:

- * Acquisition date: March 2002 / September 2003.
- * Radiometric corrections only.
- * Panchromatic image at 0.6 m/pixel.
- * Multispectral four-band image (three conventional red, green, blue channels with a fourth band in the near infrared) with 2.8 meter pitch on the ground.
- * Latitude/longitude of the rectangular zone covered:
 - ↪ North-west corner: Latitude: 40.0667° (2002) / 40.0808° (2003); Longitude: 116.2608° (2002) / 116.259° (2003)
 - ↪ South-east corner: Latitude: 39.9475° (2002) / 39.919° (2003); Longitude: 116.4247° (2002) / 116.455° (2003)
- * Each spectral channel is coded on 16 bits.
- * The raw image does not contain any geocoded information.
- * Angle with respect to the nadir: 5° (2002) / 9° (2003).

A.1.2 Cartographic data

We have three sets of cartographic data made in 1996, 2001 and 2002 by the BISM. The 2001 and 2002 data were obtained by photo-interpretation of Ikonos (1m/pixel) and Quickbird images respectively, the 1996 data were obtained by photo-interpretation of pairs of aerial photographs. These GIS type data have the following characteristics:

- * Scale: 1:10000.
- * A map is divided into 4 km (north-south) \times 5 km (east-west) tiles.
- * Organized in layers: road, constructions, rivers, lakes, green spaces (figure A.2).

- * Data that are only two-dimensional and geocoded in the cartographic system specific to Beijing with a precision of less than or equal to one meter.
- * Vector format (at the level of the constructions):
 - ↪ each building is represented by a simple polygon (no intersection of edges) resulting from the outlining of the object by an operator. It is possible that one or more polygons may be included in another polygon to represent courtyards or superstructures.
 - ↪ The polygons do not intersect each other partially.
 - ↪ The outlining of the building by the operator varies according to the perspective effect (visible facades). In the case where the building is not very high (negligible perspective distortion for the considered scale), the contour of the roof is entered and inserted in the map. When the perspective effect is great (the roof and the building's footprint do not match), the roof is outlined and then translated to the footprint.
 - ↪ The GIS data can be handled in the form of DXF-type ASCII files (format created by the developers of AUTOCAD).
 - ↪ The file relative to a tile contains a list of polygons. Each polygon contains a sub-list of points designating the polygon's vertices. This list is cyclic (the last point in the list is identical to the first one), each point is expressed in the Beijing coordinates system.
- * Simplification when creating the data:
 - ↪ **Generalization.** It is difficult to outline each building when they are in clusters, are numerous and of a small size. The solution adopted by the BISM consists of outlining the block made up from the buildings. The blocks are separated by the main roads. Figure 2.4 illustrates this generalization effect.
 - ↪ **Individual simplification.** Certain buildings have a complex shape that is not worth transcribing precisely in the map for the scale considered. The shape is therefore simplified in a more or less arbitrary way by the operator.

A.2 Pre-processing of the remote sensing data

A.2.1 Subsampling of the aerial images

The subsampling of an image is a basic application in image processing. There are many methods that can be used to do this (nearest neighbor, bilinear, bicubic, etc.). A subsampling operation can be formalized by an ideal low-pass filtering [14] :

$$\hat{I}(i, j) = \sum_{k, l \in \mathbb{Z}} \text{sinc}(\nu_e^c i - k) \text{sinc}(\nu_e^l j - l) I(k, l) \quad (\text{A.1})$$

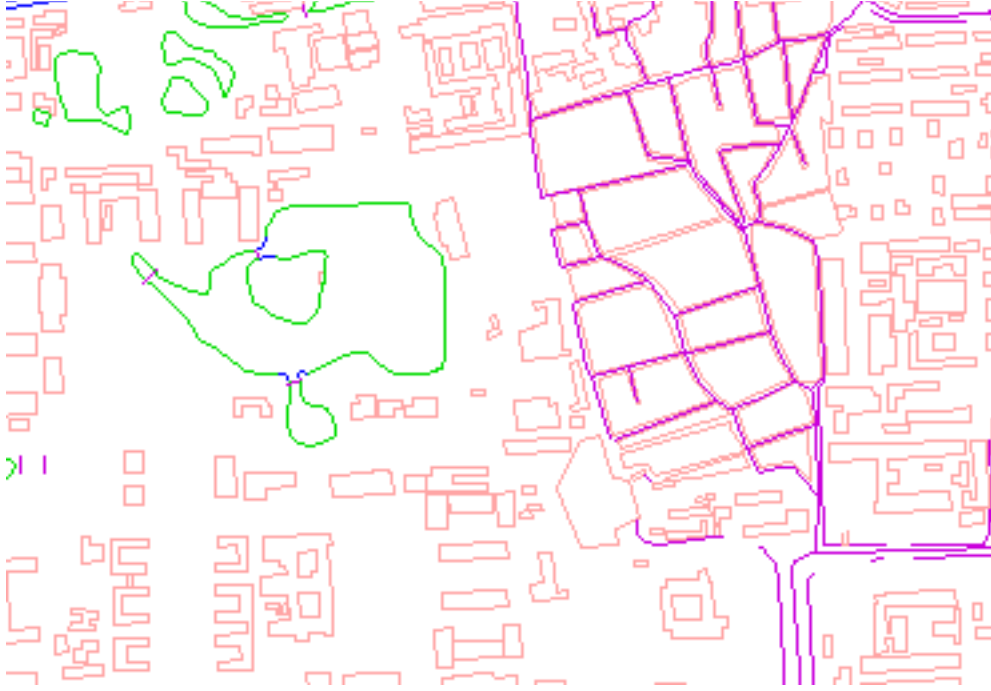


Figure A.2: Beijing city: extract from the GIS data at a 1:10000 scale. The constructions are in pink, the road network in purple, the rivers and canals in blue, the lakes in green.

where: ν_e^c and ν_e^l are the sampling frequencies on image I per the columns and lines respectively.

This type of filtering is carried out in the spatial domain by a convolution with a cardinal sine on an infinite scale. In practice, there are several methods for rendering the spatial windowing finite. The most brutal way consists of multiplying the signal by a rectangular window, thus generating Gibbs effects (this being due to this window's infinite slope at its edges). Multiple windows with finite derivatives are then proposed (Blackman, Hanning, Hamming, Kaiser, etc.). [50] is the reference article concerning these apodization windows.

In our case, particular attention will be paid to the preservation of the buildings' contours since they are preserved by the correlation technique developed in the 3D reconstruction software [85, 86]. In most cases, a building's contour is materialized by a difference of luminance in the image. It is a step or ramp type transition characterized by high frequencies in the spectral domain. In [14], Philippe Blanc has carried out a comparative study of the low-pass filters associated with the different types of sampling. It came out that the so-called apodized Shannon interpolation subsampling is the technique that enables to best keep an image's high frequencies. Consequently, it is this method that we have chosen for the under-dimensioning of the Beijing aerial photographs. The interpolation by apodized Shannon kernel is a convolution of the image by a cardinal sine windowed by the apodization functions discussed in [50]. We

have chosen to subsample the images by a factor of three in order to obtain a resolution of 0.63 meters per pixel which is the best resolution achieved to date for observation satellites in the civil sector (Quickbird 2). This subsampling factor also makes it possible to reduce the complexity of the images and consequently the error rate when they are correlated. It has been possible to reduce the noise introduced by the scanner characterized by high frequencies by low-pass filtering of the subsampling. The apodized Shannon convolution was carried out on a domain of 97×97 pixels with a Blackman window.

A.2.2 Rectification of the satellite images

The satellite images were rectified from the terrain altimetric variations at a later stage by a DTM. This processing, carried out by the BISM, enables the geocoding of the data in the Beijing cartographic system. Rectification is a critical phase because its precision will impact the quality of the updated map. Indeed, we will consider that the satellite images are the recent reference data from which we will update the map. Fifty Ground Control Points were used by the BISM to assess the geocoding quality of the Quickbird images. The overall achieved precision is 0.65 pixel, i.e. 0.4 m.

Appendix B

Simplex algorithm

The simplex method consists in finding a set of optimum parameters that minimize a cost function. This is a 0-order technique not requiring the calculation of the cost function's gradient. In this method, which we owe to Nelder and Mead [76], we look for the values of the cost function on a polyhedron with $N + 1$ vertices in an N -dimension space (a triangle in a plane, a tetrahedron in space, etc.). The first vertex is the evaluation of the cost function with the initial parameters, the other vertices are constructed by varying these parameters.

At each iteration, we calculate a new simplex which will enable to efficiently come closer to the minimum. To do this, we start by classifying the vertices of the current simplex according to the values of the cost function at these vertices. Let \mathbf{m} be the best point, \mathbf{p} the worst, \mathbf{a} the penultimate immediately better than \mathbf{p} . To build a new simplex, the general direction must correspond to getting afar from \mathbf{p} . We will choose the direction defined by the straight line $\{\mathbf{p}, \mathbf{c}\}$, \mathbf{c} being the center of the face opposite to \mathbf{p} (its coordinates are the mean of the other points coordinates). On this straight line, we try the point \mathbf{e} symmetrical to \mathbf{p} with respect to \mathbf{c} :

$$\mathbf{e} = \mathbf{c} + (\mathbf{c} - \mathbf{p}) = 2\mathbf{c} - \mathbf{p}$$

If \mathbf{e} is better than \mathbf{m} , we can try to go even farther in that direction:

$$\mathbf{e}' = \mathbf{c} + \gamma(\mathbf{c} - \mathbf{p}), \gamma > 1$$

If \mathbf{e}' is better than \mathbf{e} , the new simplex will be obtained by replacing \mathbf{p} with \mathbf{e}' (*expansion* of the simplex, figure B.2). Otherwise, we will replace \mathbf{p} with \mathbf{e} (*reflection* of the simplex see figure B.1)

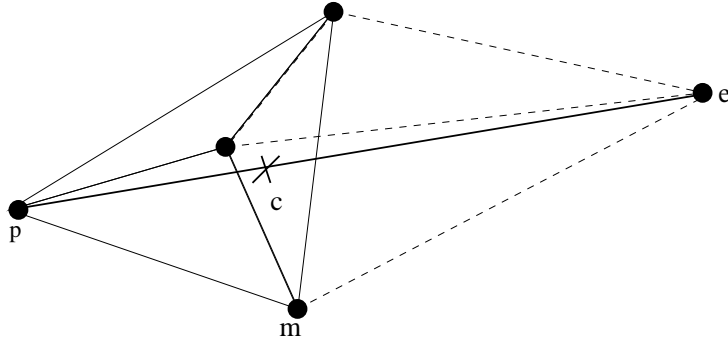


Figure B.1: Initial simplex and reflection.

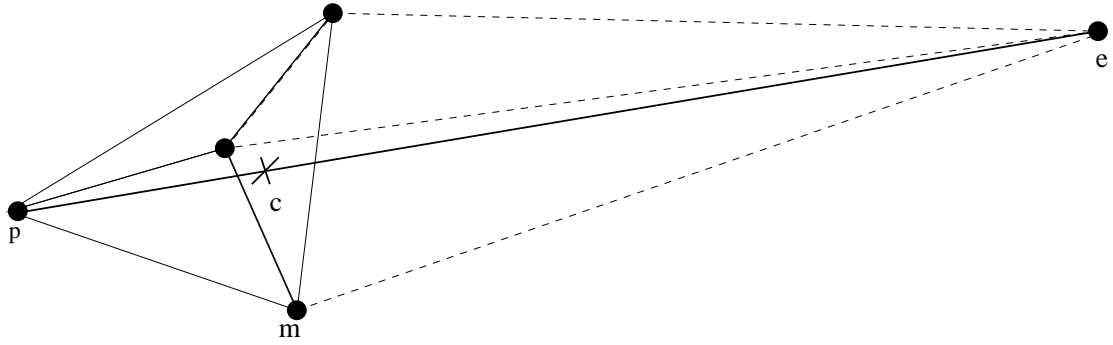


Figure B.2: Expansion

If \mathbf{e} is not better than \mathbf{m} we try \mathbf{e}'' and \mathbf{e}''' situated on the straight line $\{\mathbf{c}, \mathbf{p}\}$ on either side of \mathbf{c} according to:

$$\mathbf{e}'' = \mathbf{c} + \frac{1}{2}(\mathbf{c} - \mathbf{p})$$

$$\mathbf{e}''' = \mathbf{c} - \frac{1}{2}(\mathbf{c} - \mathbf{p})$$

Otherwise, if the best of these points is better than \mathbf{a} , we will choose it to replace \mathbf{p} (*contraction* of the simplex, figure B.3)

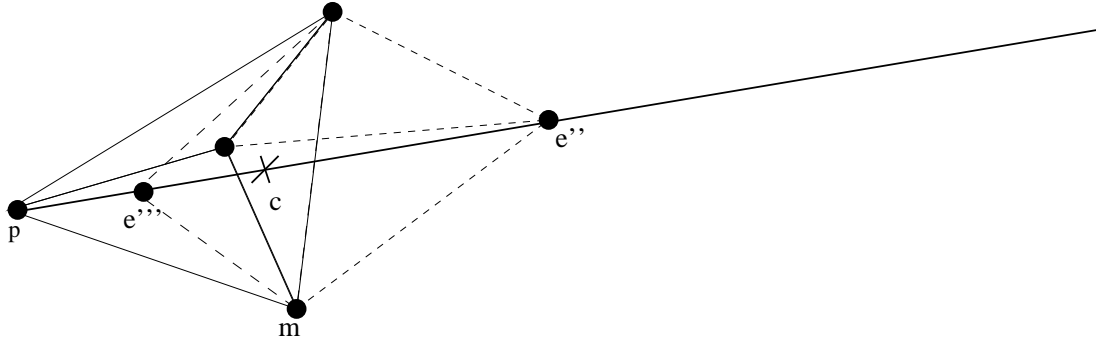


Figure B.3: Possible contractions

Otherwise, it means that the point \mathbf{m} is really close to the minimum: we then shrink the simplex keeping \mathbf{m} and replacing all the points \mathbf{s}_i by the middles of the segments $[\mathbf{s}_i, \mathbf{m}]$.

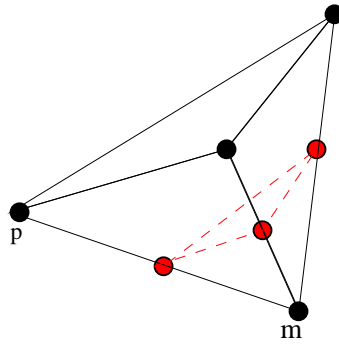


Figure B.4: Shrinking

The method is very simple and converges with great robustness. The iterations stop when the dimension of the simplex is sufficiently small.

Appendix C

Direct and inverse localization modeling

C.1 Inverse modeling

This modeling consists in determining the position in the image of a point initially expressed in three-dimensional coordinates in a reference frame \mathfrak{R} . Whether \mathfrak{R} is a geo-centered reference frame, or associated with a cartographic projection, this will not affect the writing of the inverse modeling equations. We only assume that \mathfrak{R} is an ortho-normed direct reference.

1. Let $(X_{sol}, Y_{sol}, Z_{sol})$ be the coordinates of a point in the scene expressed in the reference frame \mathfrak{R} . The coordinates of the viewing vector in \mathfrak{R} are $(X_{sol} - X_{cam}, Y_{sol} - Y_{cam}, Z_{sol} - Z_{cam})$. The coordinates indexed “cam” are those of the position of the camera’s optical center in \mathfrak{R} .
2. This same vector expressed in the camera reference is

$$\begin{pmatrix} U \\ V \\ W \end{pmatrix} = M \begin{pmatrix} X_{sol} - X_{cam} \\ Y_{sol} - Y_{cam} \\ Z_{sol} - Z_{cam} \end{pmatrix}$$

where M is the rotation matrix that makes the transfer from reference \mathfrak{R} to the camera reference. The expression of M is given by:

$$M = \begin{pmatrix} \cos \alpha_z & -\sin \alpha_z & 0 \\ \sin \alpha_z & \cos \alpha_z & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos \alpha_y & 0 & -\sin \alpha_y \\ 0 & 1 & 0 \\ \sin \alpha_y & 0 & \cos \alpha_y \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha_x & -\sin \alpha_x \\ 0 & \sin \alpha_x & \cos \alpha_x \end{pmatrix} \quad (\text{C.1})$$

where $(\alpha_x, \alpha_y, \alpha_z)$ are the rotation angles of the camera reference with respect to the (0X), (0Y) and (0Z) axes of the reference frame \mathfrak{R} .

3. The coordinates expressed in the camera reference, centered on the Principal Point of Symmetry (PPS), without distortion, of the corresponding point are:

$$\begin{pmatrix} -\frac{U}{W}f \\ -\frac{V}{W}f \end{pmatrix}$$

where f is the optical system's focal length.

4. By ignoring the distortion, and re-centering in the camera reference (direct and ortho-normed), these coordinates become:

$$\begin{cases} x = X_{PPS} - \frac{U}{W}f \\ y = Y_{PPS} - \frac{V}{W}f \end{cases}$$

The transition of the coordinates expressed in the camera reference to the image's pixel coordinates uses the affinity parameters:

$$\begin{cases} line = T_{lig} + a_{00}x + a_{01}y \\ column = T_{col} + a_{10}x + a_{11}y \end{cases}$$

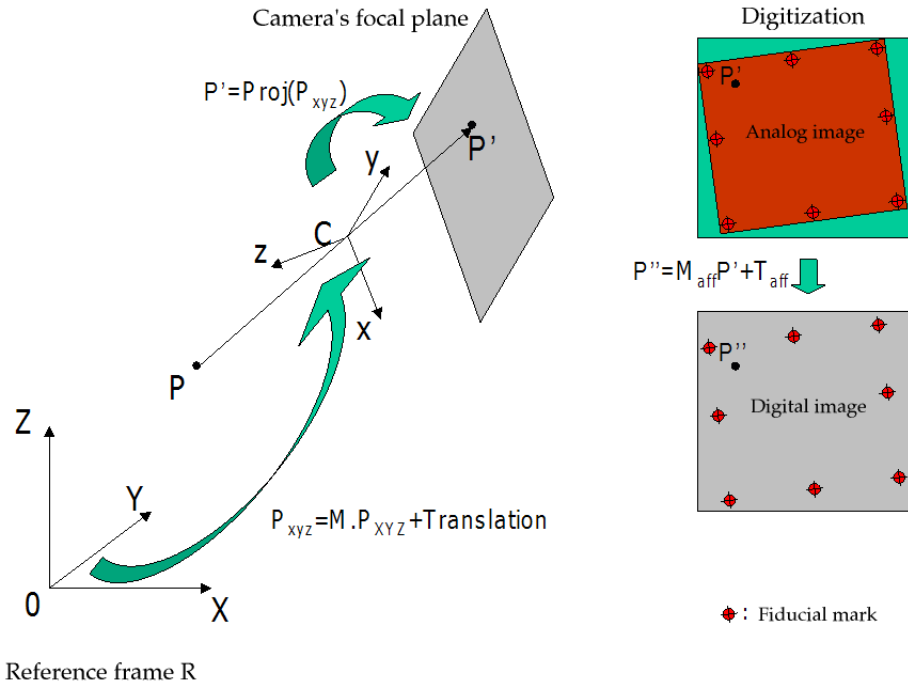


Figure C.1: Diagram illustrating inverse modeling. The terrain and camera references are direct.

C.2 Direct modeling

Direct modeling enables to calculate from a point in the image its three-dimensional position in \mathfrak{R} at a given altitude h above an ellipsoid.

1. Let (l, c) be the line/column coordinates of the pixel to be localized. The coordinates (x, y) of this same point in the camera reference are given by:

$$\begin{cases} x = \frac{(l-T_{lig})a_{11}-(c-T_{col})a_{01}}{a_{00}a_{11}-a_{01}a_{10}} \\ y = \frac{-(l-T_{lig})a_{10}+(c-T_{col})a_{00}}{a_{00}a_{11}-a_{01}a_{10}} \end{cases}$$

where: (T_{lig}, T_{col}) are the affinity parameters in translation; (a_{ij}) are the rotation/scaling affinity parameters .

2. The coordinates (x', y') in this same reference but centered on the PPS are

$$\begin{cases} x' = x - X_{PPS} \\ y' = y - Y_{PPS} \end{cases}$$

3. By ignoring the distortion, we can then express the viewing direction V in the camera reference:

$$\begin{pmatrix} \frac{x'}{f} \\ \frac{y'}{f} \\ -1 \end{pmatrix}$$

4. This same direction expressed in \mathfrak{R} is

$$V' = M^{-1}V$$

where M is the rotation matrix of equation (C.1).

The 3D point of height h in \mathfrak{R} is determined by the intersection of the viewing direction with the ellipsoid of revolution with radii $(a + h, b + h)$ (a : major radius, b : minor radius). The ellipsoid of revolution's equation modeling the surface of the Earth and expressed in \mathfrak{R} is equivalent to:

$$\frac{x^2}{(a+h)^2} + \frac{y^2}{(a+h)^2} + \frac{z^2}{(b+h)^2} = 1 \quad (\text{C.2})$$

The intersection point I between the viewing direction V' and the ellipsoid verifies:

$$I = C + uV'$$

where C is the position of the camera's optical center expressed in \mathfrak{R} ; u is a scalar. Finding the intersection I is therefore equivalent to determining the scalar u . The previous equation can also be written in the form:

$$\begin{cases} x_I = x_C + uV'_x \\ y_I = y_C + uV'_y \\ z_I = z_C + uV'_z \end{cases} \quad (\text{C.3})$$

Injecting the equations of system (C.3) into the equation of the ellipsoid (C.2) enables to find the roots of the second degree polynomial:

$$\alpha u^2 + \beta u + \gamma = 0$$

with:

$$\begin{cases} \alpha = (a+h)^2 V_z'^2 + (b+h)^2 (V_x'^2 + V_y'^2) \\ \beta = 2 \left\{ (a+h)^2 z_c V_z'^2 + (b+h)^2 (x_c V_x' + y_c V_y') \right\} \\ \gamma = (a+h)^2 z_c^2 + (b+h)^2 (x_c^2 + y_c^2) - (a+h)^2 (b+h)^2 \end{cases} \quad (\text{C.4})$$

The two roots of the second degree polynomial correspond to the two intersections of the viewing direction with the ellipsoid. We are looking for the closest intersection to the camera's optical center, i.e. the one corresponding to the smallest root. Since the quantity α is strictly positive, the smallest root is:

$$u_0 = \frac{-\beta - \sqrt{\beta^2 - 4\alpha\gamma}}{2\alpha}$$

Intersection I of altitude h is therefore

$$\begin{cases} x_I = x_C + u_0 V_x' \\ y_I = y_C + u_0 V_y' \\ z_I = z_C + u_0 V_z' \end{cases} \quad (\text{C.5})$$

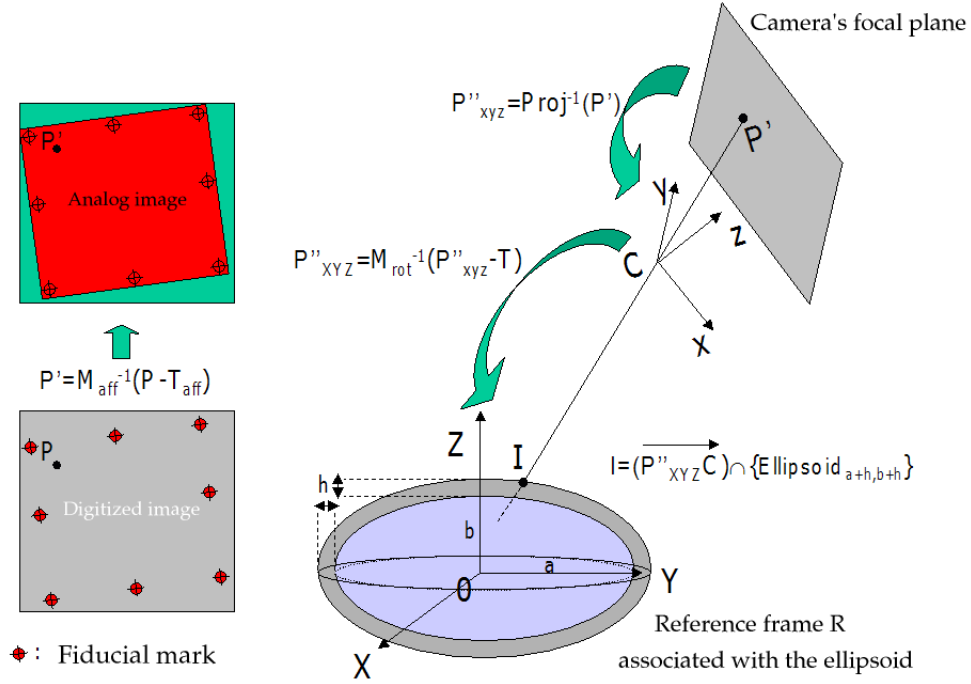


Figure C.2: Diagram illustrating the direct modeling. The terrain and camera references are direct.

Appendix D

Fast calculation of the mean and of the variance inside and outside an active contour represented by a levels set

As proposed by M. Rousson in [102], it is useless to integrally calculate the mean and the variance of the pixels in the image inside and outside the active contour at each iteration. Indeed, this operation must be performed once at time $t = 0$. For the later estimation of the statistical quantities, it is sufficient to identify the pixels in the image that enter and leave the region delimited by $\mathbf{C}(t)$ in order to correct the values of the mean and of the variance obtained beforehand. Let $\Omega_{out,in}$ be the set of pixels excluded from the contour at time t and included in the contour at $t + 1$. The cardinal of this set is $n_{out,in}$. Let $\Omega_{in,out}$ be the dual of $\Omega_{out,in}$ with cardinal $n_{in,out}$. By definition, the mean inside the contour at time $t + 1$ is:

$$c_{in}(t+1) = \frac{1}{n_{in}(t+1)} \int_{\Omega_{in}(t+1)} I(\mathbf{x}) d\mathbf{x}$$

This function is expressed according to $c_{in}(t)$ in the following way:

$$c_{in}(t+1) = \frac{1}{n_{in}(t) + n_{out,in} - n_{in,out}} \left\{ n_{in}(t) c_{in}(t) + \int_{\Omega_{out,in}} I(\mathbf{x}) d\mathbf{x} - \int_{\Omega_{in,out}} I(\mathbf{x}) d\mathbf{x} \right\} \quad (\text{D.1})$$

Likewise, we find the following expression of the mean on the outside:

$$c_{out}(t+1) = \frac{1}{n_{out}(t) - n_{out,in} + n_{in,out}} \left\{ n_{out}(t) c_{out}(t) - \int_{\Omega_{out,in}} I(\mathbf{x}) d\mathbf{x} + \int_{\Omega_{in,out}} I(\mathbf{x}) d\mathbf{x} \right\} \quad (\text{D.2})$$

By remarking that the variance of a random variable X is:

$$\sigma^2(X) = E(X^2) - E(X)^2$$

we then have

$$\sigma_{in}^2(t+1) = \frac{1}{n_{in}(t+1)} \int_{\Omega_{in}(t+1)} I^2(\mathbf{x}) d\mathbf{x} - c_{in}^2(t+1)$$

which we express according to $\sigma_{in}^2(t)$ and to $c_{in}(t)$ thanks to (D.1):

$$\begin{aligned} \sigma_{in}^2(t+1) = \frac{1}{n_{in}(t) + n_{out,in} - n_{in,out}} & \left\{ n_{in}(t) [c_{in}^2(t) + \sigma_{in}^2(t)] + \int_{\Omega_{out,in}} I^2(\mathbf{x}) d\mathbf{x} \right. \\ & \left. - \int_{\Omega_{in,out}} I^2(\mathbf{x}) d\mathbf{x} \right\} - c_{in}^2(t+1) \end{aligned} \quad (D.3)$$

The dual expression $\sigma_{out}^2(t+1)$ is analog:

$$\begin{aligned} \sigma_{out}^2(t+1) = \frac{1}{n_{out}(t) - n_{out,in} + n_{in,out}} & \left\{ n_{out}(t) [c_{out}^2(t) + \sigma_{out}^2(t)] - \int_{\Omega_{out,in}} I^2(\mathbf{x}) d\mathbf{x} \right. \\ & \left. + \int_{\Omega_{in,out}} I^2(\mathbf{x}) d\mathbf{x} \right\} - c_{out}^2(t+1) \end{aligned} \quad (D.4)$$

The advantage of this type of technique is that it only operates on the $\Omega_{in,out}$ and $\Omega_{out,in}$ spaces whose cardinals are very low compared with those of Ω_{in} and Ω_{out} . The reduction in calculation time is significant since the calculation complexity becomes proportional to the length of the active contour. In [102], M.Rousson experimentally demonstrates the saving made in the computational time.

Appendix E

Calculation of the shape constraint and quadratic energy variations

E.1 Evolution equation derived from the shape constraint functional by difference of normals

Let us consider the shape constraint energy term based on the difference between the unit normal of the active contour implicitly represented by the level set ϕ and that of a reference shape represented by ψ :

$$J_{Shape}(\phi, \psi) = \int_{\Omega} |\nabla \phi(\mathbf{x}) - \nabla \psi(\mathbf{x})|^2 (H(\phi(\mathbf{x})) - H(\psi(\mathbf{x})))^2 d\mathbf{x} \quad (\text{E.1})$$

The evolution equation consists of finding the values of ϕ that minimize the functional $J_{Shape}(\phi, \psi, t)$. To achieve this, we adopt the classic local minimization by gradient descent scheme:

$$\frac{\partial J_{Shape}(\phi, \psi, t)}{\partial \phi} = -\phi_t(\mathbf{x}, t) \quad (\text{E.2})$$

The expression of $\frac{\partial J_{Shape}(\phi, \psi, t)}{\partial \phi}$ is deduced from the Gâteaux derivative of $J_{Shape}(\phi, \psi, t)$ with respect to ϕ in direction $\tilde{\phi}$:

$$\left(\frac{\partial J_{Shape}(\phi, \psi, t)}{\partial \phi} \right)_{\phi=\tilde{\phi}} = \left\langle J_{Shape}(\phi, \psi, t), \tilde{\phi} \right\rangle = \int_{\Omega} \frac{\partial J_{Shape}(\phi, \psi, t)}{\partial \phi} \tilde{\phi}(\mathbf{x}, t) d\mathbf{x} \quad (\text{E.3})$$

The Gâteaux derivative of the functional is expressed in the form:

$$\left(\frac{\partial J_{Shape}(\phi, \psi, t)}{\partial \phi} \right)_{\phi=\tilde{\phi}} = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \left(J_{Shape}(\phi(\mathbf{x}, t) + \varepsilon \tilde{\phi}(\mathbf{x}, t), \psi, t) - J_{Shape}(\phi(\mathbf{x}, t), \psi, t) \right) \quad (\text{E.4})$$

Let us detail the calculation of the equation (E.4):

$$J_{Shape}(\phi(\mathbf{x}, t) + \varepsilon \tilde{\phi}(\mathbf{x}, t), \psi, t) =$$

$$\int_{\Omega} \left\langle \nabla \phi(\mathbf{x}) + \varepsilon \nabla \tilde{\phi}(\mathbf{x}, t) - \nabla \psi(\mathbf{x}), \nabla \phi(\mathbf{x}) + \varepsilon \nabla \tilde{\phi}(\mathbf{x}, t) - \nabla \psi(\mathbf{x}) \right\rangle \cdot \left(H(\phi(\mathbf{x}) + \varepsilon \tilde{\phi}(\mathbf{x}, t)) - H(\psi(\mathbf{x})) \right)^2 d\mathbf{x} \quad (\text{E.5})$$

noting that:

$$\delta_a(\phi(\mathbf{x}, t)) = \lim_{\varepsilon \rightarrow 0} \frac{H_a(\phi(\mathbf{x}, t) + \varepsilon \tilde{\phi}(\mathbf{x}, t)) - H_a(\phi(\mathbf{x}, t))}{\varepsilon \tilde{\phi}(\mathbf{x}, t)} \quad (\text{E.6})$$

and by developing equation (E.5), we have:

$$\begin{aligned} J_{Shape}(\phi(\mathbf{x}, t) + \varepsilon \tilde{\phi}(\mathbf{x}, t), \psi, t) &= J_{Shape}(\phi(\mathbf{x}, t)) \\ &+ 2\varepsilon \int_{\Omega} \tilde{\phi}(\mathbf{x}, t) \delta_a(\phi(\mathbf{x}, t)) (H(\phi(\mathbf{x})) - H(\psi(\mathbf{x}))) \langle \nabla \phi(\mathbf{x}) - \nabla \psi(\mathbf{x}), \nabla \phi(\mathbf{x}) - \nabla \psi(\mathbf{x}) \rangle d\mathbf{x} \\ &+ 2\varepsilon \int_{\Omega} (H(\phi(\mathbf{x})) - H(\psi(\mathbf{x})))^2 \langle \nabla \phi(\mathbf{x}) - \nabla \psi(\mathbf{x}), \nabla \tilde{\phi}(\mathbf{x}, t) \rangle d\mathbf{x} \end{aligned} \quad (\text{E.7})$$

By integrating by parts the latter term of equation (E.7), we obtain:

$$\begin{aligned} J_{Shape}(\phi(\mathbf{x}, t) + \varepsilon \tilde{\phi}(\mathbf{x}, t), \psi, t) - J_{Shape}(\phi(\mathbf{x}, t)) &= \\ &+ 2\varepsilon \int_{\Omega} \tilde{\phi}(\mathbf{x}, t) \delta_a(\phi(\mathbf{x}, t)) (H(\phi(\mathbf{x})) - H(\psi(\mathbf{x}))) \cdot \\ &\quad \left(|\nabla \phi(\mathbf{x})|^2 + |\nabla \psi(\mathbf{x})|^2 - 2 \langle \nabla \phi(\mathbf{x}), \nabla \psi(\mathbf{x}) \rangle \right) d\mathbf{x} \\ &+ 2\varepsilon \left\{ \left[(H(\phi(\mathbf{x})) - H(\psi(\mathbf{x})))^2 (\phi_x(\mathbf{x}, t) + \phi_y(\mathbf{x}, t) - \psi_x(\mathbf{x}, t) - \psi_y(\mathbf{x}, t)) \tilde{\phi}(\mathbf{x}, t) \right]_{\Omega} \right. \\ &+ \int_{\Omega} \tilde{\phi}(\mathbf{x}, t) [2(H(\phi(\mathbf{x})) - H(\psi(\mathbf{x}))) ((\delta_a(\phi(\mathbf{x}, t)) + \delta_a(\psi(\mathbf{x}, t))) \langle \nabla \phi(\mathbf{x}), \nabla \psi(\mathbf{x}) \rangle \\ &\quad + \delta_a(\phi(\mathbf{x}, t)) |\nabla \phi(\mathbf{x})|^2 - \delta_a(\psi(\mathbf{x}, t)) |\nabla \psi(\mathbf{x})|^2 \\ &\quad \left. + (H(\phi(\mathbf{x})) - H(\psi(\mathbf{x})))^2 (\Delta \psi(\mathbf{x}) - \Delta \phi(\mathbf{x})) \right] d\mathbf{x} \Big\} \end{aligned} \quad (\text{E.8})$$

The term $\left[(H(\phi(\mathbf{x})) - H(\psi(\mathbf{x})))^2 (\phi_x(\mathbf{x}, t) + \phi_y(\mathbf{x}, t) - \psi_x(\mathbf{x}, t) - \psi_y(\mathbf{x}, t)) \tilde{\phi}(\mathbf{x}, t) \right]_{\Omega}$ is null since it is always possible to find a rectangular space Ω that can include the shapes represented by ϕ and ψ . So, far from the level 0 of the level sets, we have $H(\phi(\mathbf{x})) = 0 = H(\psi(\mathbf{x}))$. The difference of energies becomes:

$$\begin{aligned} J_{Shape}(\phi(\mathbf{x}, t) + \varepsilon \tilde{\phi}(\mathbf{x}, t), \psi, t) - J_{Shape}(\phi(\mathbf{x}, t)) &= \\ 2\varepsilon \int_{\Omega} \tilde{\phi}(\mathbf{x}, t) \Big\{ (H(\phi(\mathbf{x})) - H(\psi(\mathbf{x}))) \left(\delta_a(\phi(\mathbf{x}, t)) (|\nabla \psi(\mathbf{x})|^2 - |\nabla \phi(\mathbf{x})|^2) \right. \right. \\ &\quad \left. \left. - 2\delta_a(\psi(\mathbf{x}, t)) |\nabla \psi(\mathbf{x})|^2 + 2\delta_a(\psi(\mathbf{x}, t)) \langle \nabla \phi(\mathbf{x}), \nabla \psi(\mathbf{x}) \rangle \right) \right\} d\mathbf{x} \end{aligned}$$

$$+ (H(\phi(\mathbf{x})) - H(\psi(\mathbf{x})))^2 (\Delta\psi(\mathbf{x}) - \Delta\phi(\mathbf{x})) \} d\mathbf{x} \quad (\text{E.9})$$

The energy gradient then comes directly, and noting that the normals are unit ($|\nabla\psi(\mathbf{x})|^2 = 1 = |\nabla\phi(\mathbf{x})|^2$):

$$\begin{aligned} \frac{\partial J_{Shape}(\phi, \psi, t)}{\partial \phi} &= 4 (H(\phi(\mathbf{x})) - H(\psi(\mathbf{x}))) \{ \delta_a(\psi(\mathbf{x}, t)) (\langle \nabla\phi(\mathbf{x}), \nabla\psi(\mathbf{x}) \rangle - 1) \} \\ &+ 2 (H(\phi(\mathbf{x})) - H(\psi(\mathbf{x})))^2 (\Delta\psi(\mathbf{x}) - \Delta\phi(\mathbf{x})) \end{aligned} \quad (\text{E.10})$$

The evolution equation is deduced by change of sign according to equation (E.2).

E.2 Derivative of the quadratic correction energy

The energy introduced into equation (6.7) can be split according to:

$$J_{correction}(\mathbf{C}) = J_1(C) - J_2(C) \quad (\text{E.11})$$

with

$$J_1(C) = \oint \oint dp dp' \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 |\dot{\mathbf{C}}|^{-1} |\dot{\mathbf{C}}'|^{-1} \Psi(R) \quad (\text{E.12})$$

and

$$J_2(C) = \oint \oint dp dp' \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^4 |\dot{\mathbf{C}}|^{-3} |\dot{\mathbf{C}}'|^{-3} \Psi(R) \quad (\text{E.13})$$

E.2.0.1 Energy $J_1(C)$

$$J_1(C + \delta\tilde{\mathbf{C}}) = \oint \oint dp dp' \langle \dot{\mathbf{C}} + \delta\dot{\tilde{\mathbf{C}}}, \dot{\mathbf{C}}' + \delta\dot{\tilde{\mathbf{C}}}' \rangle^2 |\dot{\mathbf{C}} + \delta\dot{\tilde{\mathbf{C}}}|^{-1} |\dot{\mathbf{C}}' + \delta\dot{\tilde{\mathbf{C}}}'|^{-1} \Psi(\tilde{R}) \quad (\text{E.14})$$

The quantity $\Psi(\tilde{R})$ is rewritten according to:

$$\Psi(\tilde{R}) \approx \Psi(R) + \langle \tilde{\mathbf{R}}, \delta\tilde{\mathbf{C}} - \delta\tilde{\mathbf{C}}' \rangle \Psi'(R) \quad (\text{E.15})$$

and $|\dot{\mathbf{C}} + \delta\dot{\tilde{\mathbf{C}}}|^{-1}$ is:

$$|\dot{\mathbf{C}} + \delta\dot{\tilde{\mathbf{C}}}|^{-1} \approx |\dot{\mathbf{C}}|^{-1} \left(1 - \langle \dot{\mathbf{C}}, \delta\dot{\tilde{\mathbf{C}}} \rangle |\dot{\mathbf{C}}|^{-2} \right) \quad (\text{E.16})$$

By developing the square of the scalar product, while replacing $\Psi(\tilde{R})$ with (E.15), and $|\dot{\mathbf{C}} + \delta\dot{\tilde{\mathbf{C}}}|^{-1}$ with (E.16) and ignoring the infinitesimal variations of the order two, we have:

$$\begin{aligned} J_1(C + \delta\tilde{\mathbf{C}}) &= J_1(C) + 2 \oint \oint dp dp' \Psi'(R) \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 \langle \tilde{\mathbf{R}}, \delta\tilde{\mathbf{C}} \rangle |\dot{\mathbf{C}}|^{-1} |\dot{\mathbf{C}}'|^{-1} \\ &- 2 \oint \oint dp dp' \Psi(R) \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 |\dot{\mathbf{C}}|^{-3} |\dot{\mathbf{C}}'|^{-1} \langle \dot{\mathbf{C}}, \delta\dot{\tilde{\mathbf{C}}} \rangle \end{aligned}$$

$$+4 \oint \oint dpdp' \Psi(R) \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle |\dot{\mathbf{C}}|^{-1} |\dot{\mathbf{C}}'|^{-1} \langle \dot{\mathbf{C}}', \delta \dot{\mathbf{C}} \rangle \quad (\text{E.17})$$

By integrating the last term by parts, we have¹:

$$\begin{aligned} J_1(C + \delta \tilde{\mathbf{C}}) - J_1(C) &= 2 \oint \oint dpdp' \Psi'(R) \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 |\dot{\mathbf{C}}|^{-1} |\dot{\mathbf{C}}'|^{-1} \langle \tilde{\mathbf{R}}, \delta \tilde{\mathbf{C}} \rangle \\ +2 \oint \oint dpdp' \left[\Psi(R) \left(2 \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle \langle \ddot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle |\dot{\mathbf{C}}|^{-3} |\dot{\mathbf{C}}'|^{-1} - 3 |\dot{\mathbf{C}}|^{-5} |\dot{\mathbf{C}}'|^{-1} \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 \langle \ddot{\mathbf{C}}, \dot{\mathbf{C}} \rangle \right) \right. \\ &\quad \left. + \Psi'(R) \langle \dot{\mathbf{C}}, \tilde{\mathbf{R}} \rangle \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 |\dot{\mathbf{C}}|^{-3} |\dot{\mathbf{C}}'|^{-1} \right] \langle \dot{\mathbf{C}}, \delta \tilde{\mathbf{C}} \rangle \\ +2 \oint \oint dpdp' \Psi(R) \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 |\dot{\mathbf{C}}|^{-3} |\dot{\mathbf{C}}'|^{-1} \langle \ddot{\mathbf{C}}, \delta \tilde{\mathbf{C}} \rangle \\ -4 \oint \oint dpdp' \left[\Psi(R) \left(\langle \dot{\mathbf{C}}', \ddot{\mathbf{C}} \rangle |\dot{\mathbf{C}}|^{-1} |\dot{\mathbf{C}}'|^{-1} - \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle \langle \ddot{\mathbf{C}}, \dot{\mathbf{C}} \rangle |\dot{\mathbf{C}}|^{-3} |\dot{\mathbf{C}}'|^{-1} \right) \right. \\ &\quad \left. + \Psi'(R) \langle \dot{\mathbf{C}}, \tilde{\mathbf{R}} \rangle \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle |\dot{\mathbf{C}}|^{-1} |\dot{\mathbf{C}}'|^{-1} \right] \langle \dot{\mathbf{C}}', \delta \tilde{\mathbf{C}} \rangle \end{aligned} \quad (\text{E.18})$$

E.2.0.2 Energy $J_2(C)$

$$J_2(C + \delta \tilde{\mathbf{C}}) = \oint \oint dpdp' \langle \dot{\mathbf{C}} + \delta \dot{\mathbf{C}}, \dot{\mathbf{C}}' + \delta \dot{\mathbf{C}}' \rangle^4 |\dot{\mathbf{C}} + \delta \dot{\mathbf{C}}|^{-3} |\dot{\mathbf{C}}' + \delta \dot{\mathbf{C}}'|^{-3} \Psi(\tilde{R}) \quad (\text{E.19})$$

with

$$|\dot{\mathbf{C}} + \delta \dot{\mathbf{C}}|^{-3} \approx |\dot{\mathbf{C}}|^{-3} \left(1 - 3 \langle \dot{\mathbf{C}}, \delta \dot{\mathbf{C}} \rangle |\dot{\mathbf{C}}|^{-2} \right)$$

By developing the scalar product by injecting the result of (E.15) and ignoring the terms of order two, we have:

$$\begin{aligned} J_2(C + \delta \tilde{\mathbf{C}}) - J_2(C) &= 2 \oint \oint dpdp' \Psi'(R) \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^4 |\dot{\mathbf{C}}|^{-3} |\dot{\mathbf{C}}'|^{-3} \langle \tilde{\mathbf{R}}, \delta \tilde{\mathbf{C}} \rangle \\ -6 \oint \oint dpdp' \Psi(R) \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^4 |\dot{\mathbf{C}}|^{-5} |\dot{\mathbf{C}}'|^{-3} \langle \dot{\mathbf{C}}, \delta \dot{\mathbf{C}} \rangle \\ +8 \oint \oint dpdp' \Psi(R) \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^3 |\dot{\mathbf{C}}|^{-3} |\dot{\mathbf{C}}'|^{-3} \langle \dot{\mathbf{C}}', \delta \dot{\mathbf{C}} \rangle \end{aligned} \quad (\text{E.20})$$

By integrating by parts, we have:

$$\begin{aligned} J_2(C + \delta \tilde{\mathbf{C}}) - J_2(C) &= 2 \oint \oint dpdp' \Psi'(R) \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^4 |\dot{\mathbf{C}}|^{-3} |\dot{\mathbf{C}}'|^{-3} \langle \tilde{\mathbf{R}}, \delta \tilde{\mathbf{C}} \rangle \\ +2 \oint \oint dpdp' \left[\Psi(R) \left(12 \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^3 \langle \ddot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle |\dot{\mathbf{C}}|^{-5} |\dot{\mathbf{C}}'|^{-3} - 15 |\dot{\mathbf{C}}|^{-7} |\dot{\mathbf{C}}'|^{-3} \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^4 \langle \ddot{\mathbf{C}}, \dot{\mathbf{C}} \rangle \right) \right. \\ &\quad \left. + 3 \Psi'(R) \langle \dot{\mathbf{C}}, \tilde{\mathbf{R}} \rangle \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^4 |\dot{\mathbf{C}}|^{-5} |\dot{\mathbf{C}}'|^{-3} \right] \langle \dot{\mathbf{C}}, \delta \tilde{\mathbf{C}} \rangle \end{aligned}$$

¹This result is obtained by injecting: $\frac{\partial \Psi}{\partial p} = \Psi'(R) \langle \tilde{\mathbf{R}}, \dot{\mathbf{C}} \rangle$

$$\begin{aligned}
& +6 \oint \oint dp dp' \Psi(R) \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^4 |\dot{\mathbf{C}}|^{-5} |\dot{\mathbf{C}}'|^{-3} \langle \ddot{\mathbf{C}}, \delta \tilde{\mathbf{C}} \rangle \\
& -4 \oint \oint dp dp' \left[\Psi(R) \left(6 \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 \langle \dot{\mathbf{C}}', \ddot{\mathbf{C}} \rangle |\dot{\mathbf{C}}|^{-3} |\dot{\mathbf{C}}'|^{-3} - 6 \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^3 \langle \ddot{\mathbf{C}}, \dot{\mathbf{C}} \rangle |\dot{\mathbf{C}}|^{-5} |\dot{\mathbf{C}}'|^{-3} \right) \right. \\
& \quad \left. + 2 \Psi'(R) \langle \dot{\mathbf{C}}, \tilde{\mathbf{R}} \rangle \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^3 |\dot{\mathbf{C}}|^{-3} |\dot{\mathbf{C}}'|^{-3} \right] \langle \dot{\mathbf{C}}', \delta \tilde{\mathbf{C}} \rangle \quad (\text{E.21})
\end{aligned}$$

E.2.1 Total variation of the energy $J_{\text{correction}}(C)$

By grouping the results of (E.18) and (E.21), we have:

$$\begin{aligned}
J_{\text{correction}}(\mathbf{C} + \delta \tilde{\mathbf{C}}) - J_{\text{correction}}(\mathbf{C}) &= \oint \oint dp dp' \Psi(R) \{ \\
& 2 \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 |\dot{\mathbf{C}}|^{-3} |\dot{\mathbf{C}}'|^{-1} \left(1 - 3 \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 |\dot{\mathbf{C}}|^{-2} |\dot{\mathbf{C}}'|^{-2} \right) \langle \ddot{\mathbf{C}}, \delta \tilde{\mathbf{C}} \rangle \\
& + 4 \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle \langle \ddot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle |\dot{\mathbf{C}}|^{-3} |\dot{\mathbf{C}}'|^{-1} \left(1 - 6 \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 |\dot{\mathbf{C}}|^{-2} |\dot{\mathbf{C}}'|^{-2} \right) \langle \dot{\mathbf{C}}, \delta \tilde{\mathbf{C}} \rangle \\
& - 6 \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 \langle \ddot{\mathbf{C}}, \dot{\mathbf{C}} \rangle |\dot{\mathbf{C}}|^{-5} |\dot{\mathbf{C}}'|^{-1} \left(1 - 5 \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 |\dot{\mathbf{C}}|^{-2} |\dot{\mathbf{C}}'|^{-2} \right) \langle \dot{\mathbf{C}}, \delta \tilde{\mathbf{C}} \rangle \\
& - 4 \left(1 - 6 \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 |\dot{\mathbf{C}}|^{-2} |\dot{\mathbf{C}}'|^{-2} \right) |\dot{\mathbf{C}}|^{-1} |\dot{\mathbf{C}}'|^{-1} \left[\langle \ddot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle - \langle \ddot{\mathbf{C}}, \dot{\mathbf{C}} \rangle \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle |\dot{\mathbf{C}}|^{-2} \right] \langle \dot{\mathbf{C}}', \delta \tilde{\mathbf{C}} \rangle \\
& + \oint \oint dp dp' \Psi'(R) \left\{ 2 \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 |\dot{\mathbf{C}}|^{-1} |\dot{\mathbf{C}}'|^{-1} \left(1 - \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 |\dot{\mathbf{C}}|^{-2} |\dot{\mathbf{C}}'|^{-2} \right) \langle \tilde{\mathbf{R}}, \delta \tilde{\mathbf{C}} \rangle \right. \\
& \quad + 2 \langle \dot{\mathbf{C}}, \tilde{\mathbf{R}} \rangle \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 |\dot{\mathbf{C}}|^{-3} |\dot{\mathbf{C}}'|^{-1} \left(1 - 3 \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 |\dot{\mathbf{C}}|^{-2} |\dot{\mathbf{C}}'|^{-2} \right) \langle \dot{\mathbf{C}}, \delta \tilde{\mathbf{C}} \rangle \\
& \quad \left. - 4 \langle \dot{\mathbf{C}}, \tilde{\mathbf{R}} \rangle \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle |\dot{\mathbf{C}}|^{-1} |\dot{\mathbf{C}}'|^{-1} \left(1 - 2 \langle \dot{\mathbf{C}}, \dot{\mathbf{C}}' \rangle^2 |\dot{\mathbf{C}}|^{-2} |\dot{\mathbf{C}}'|^{-2} \right) \langle \dot{\mathbf{C}}', \delta \tilde{\mathbf{C}} \rangle \right\} \quad (\text{E.22})
\end{aligned}$$

By noting that $\langle \ddot{\mathbf{C}}, \dot{\mathbf{C}} \rangle = 0$, certain terms cancel each other out. By assuming that p is the arclength, this last equation can be simplified and is equivalent to (6.9).

Appendix F

3D representation of an urban scene

F.1 Objective and methodology

The results of the chapter 3 showed that we were able to generate an orthorectified Digital Surface Model (DSM) of an urban scene. This altitude information can be transferred to the 2D building digital map in order to add a third dimension. This chapter shows a middle level 3D representation of the buildings which was coded in VRML (*Virtual Reality Modeling Language*). This work was achieved by And  ol Ayzac, intern at the LIAMA from April to August 2004.

F.2 Modeling principle

Buildings are modeled by parallelepipeds whose polygonal bases are derived from the map. The height assigned to each building is the mean altitude value derived from the DSM within the cartographic polygon. Such modeling is clearly inadequate for the sole goal of a realistic representation (virtual visit of a city). Nevertheless, it is fully automatic and sufficient to model noise or atmospheric pollution.

In order to make the represented scene more realistic than a heap of “extruded boxes”, we wrapped an orthoimage (jointly generated with the DSM) and some fa  ades textures. The fa  ade texture is a collection of identical patterns which density (per pixel) is a parameter of the program. The patterns are either pictures of fa  ades taken in Beijing, or synthetic images. There exists approximatively ten classes of patterns randomly placed on the buildings. A last option of the program enables to model the terrain variations if a Digital Terrain Model (DTM) is available. The DTM is modeled by B-splines which smoothing effect is not a limitation as urban relief is seldom hilly.

F.3 Results over Beijing city

Figure F.1 illustrates the 3D representation of the stadium built for the Asian Games held in Beijing in 1990.

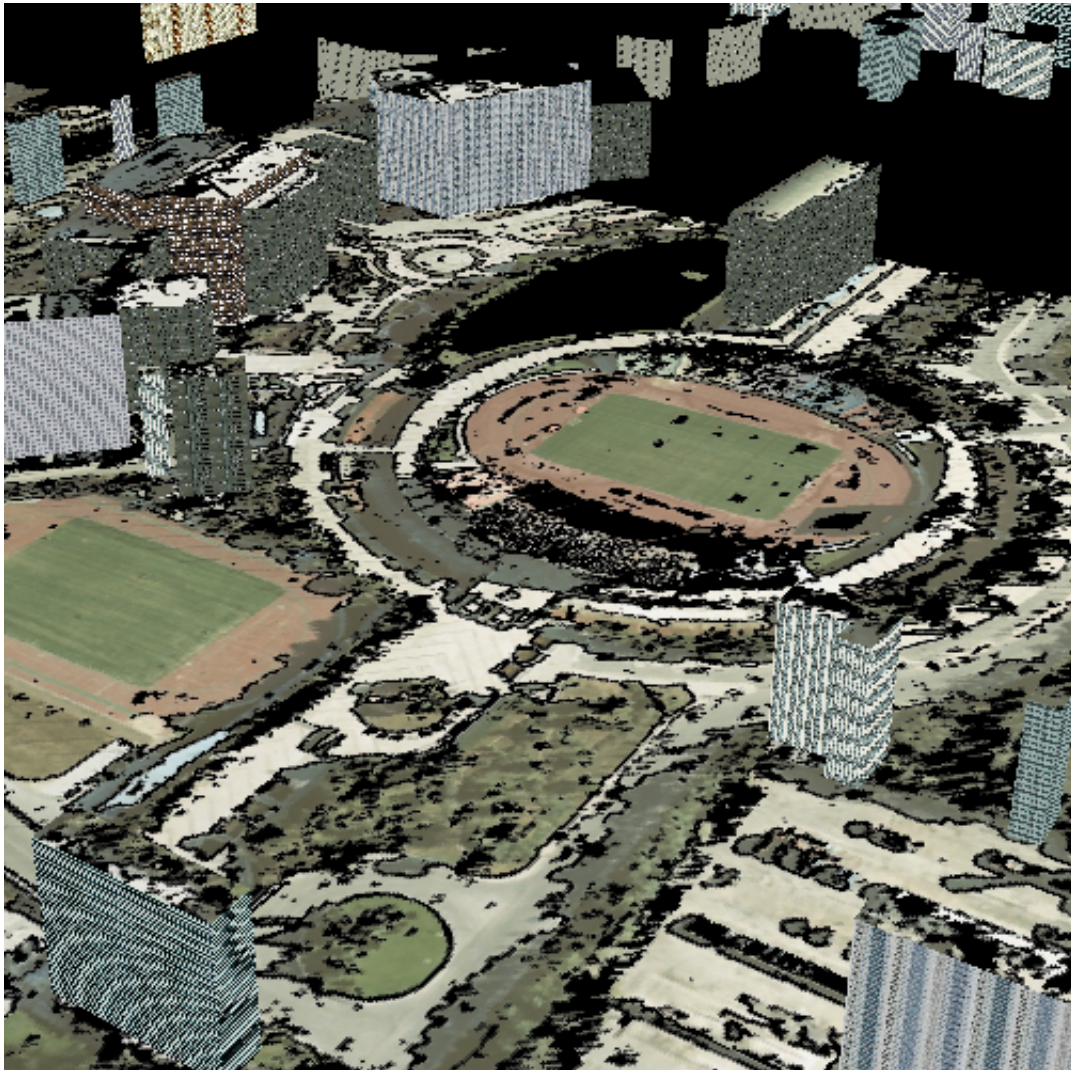


Figure F.1: 3D visualization of Beijing city - with textures and no DTM.

Bibliography

- [1] D. Adalsteinsson and J. Sethian, *The fast construction of extension velocities in level set methods*, Journal of Computational Physics **148** (1999), 2–22.
- [2] P. Agouris, K. Beard, G. Mountrakis, and A. Stefanidis, *Capturing and modeling geographic object change: a spatiotemporal gazetteer framework*, Photogrammetric engineering & remote sensing **66** (2000), no. 10, 1224–1250.
- [3] P. Agouris, G. Mountrakis, and A. Stefanidis, *Automated spatiotemporal change detection in digital aerial imagery*, Aerosense 2000, SPIE Proceedings Vol. 4054, pp. 2-12 (2000).
- [4] P. Agouris, A. Stefanidis, and S. Gyftakis, *Differential snakes for change detection in road segments*, Photogrammetric Engineering and Remote Sensing, Vol. 67, No. 12, pp. 1391-1399 (2001).
- [5] G. Ball and D.J. Hall, *A novel method of data analysis and pattern classification*, Tech. report, Standford Research Institute, Menlo Park, CA, USA, 1965.
- [6] E. Baltsavias, S. Mason, and D. Stallmann, *Use of dtms/dsms and orthoimages to support building extraction*, Proc. Workshop on "Automatic extraction of man-made objects from aerial and space images" (A. Gruen O. Kuebler, P. Agouris, ed.), Birkhauser Verlag, 1995.
- [7] E. Baltsavias, L. Zhang, and H. Eisenbeiss, *DSM generation and interior orientation determination of ikonos images using a testfield in Switzerland*, Proc. ISPRS Hannover Workshop, 2005.
- [8] E. P. Baltsavias, A. Grün, and L. Van Gool (eds.), *Automatic extraction of man-made objects from aerial and space images*, vol. 3, A.A. Balkema, 2001.
- [9] E.P. Baltsavias, *Object extraction and revision by image analysis using existing geospatial data and knowledge: State-of-the-art and steps towards operational systems*, International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences **34** (2002), no. 2, 13–22.
- [10] J. Barron, D. Fleet, and S. Beauchemin, *Performance of optical flow techniques*, Int. J. Comput. Vis **12** (1994), no. 1, 43–77.
- [11] M. Baümker and F.J. Heimes, *New calibration and computing method for direct georeferencing of image and scanner data using the position and angular data of*

- an hybrid inertial navigation system*, Proc. OEEPE-Workshop Integrated Sensor Orientation (2001).
- [12] L. Bentabet, S. Jodouin, D. Ziou, and J. Vaillancourt, *Road vectors update using sar imagery: a snake based method*, IEEE Transactions in Geoscience and Remote Sensing **41** (2003), no. 8.
- [13] A. Blake and M. Isard, *Active contours*, Springer, 1998.
- [14] P. Blanc, *Développement de méthodes pour la détection de changement*, Ph.D. thesis, Ecole des Mines de Paris, 1999.
- [15] D. Boldo and H. Le Men, *Remote sensing model adaptation to very high resolution digital images of urban areas*, IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas, pp. 20-24 (2001).
- [16] A. Busch, *Revision of built-up areas in a GIS using satellite imagery and GIS data*, Proc. ISPRS Symposium on GIS - Between visions and applications, vol. 32/4, 1998.
- [17] V. Caselles, F. Catte, T. Coll, and F. Dibos, *A geometric model for active contours*, Numerische Mathematik **66** (1993).
- [18] V. Caselles, R. Kimmel, and G. Sapiro, *On geodesic active contours*, International Journal of Computer Vision (1995).
- [19] T. Chan and L. Vese, *Active contours without edges*, IEEE Trans. on Image Processing **10** (2001), no. 2, 266–277.
- [20] ———, *A level set algorithm for minimizing the Mumford-Shah functional in image processing*, IEEE Computer Society Proceedings of the 1st IEEE Workshop on Variational and Level Set Methods in Computer Vision (2001), 161–168.
- [21] T. Chan and W. Zhu, *Level set based shape prior segmentation*, Tech. report, UCLA, 2003.
- [22] Y. Chen, F. Huang, H. Tagare, M. Rao, D. Wilson, and E. A. Geiser, *Using prior shape and intensity profile in medical image segmentation*, Proc. ICCV'2003 (2003).
- [23] Y. Chen, H. Tagare, S. Thiruvenkadam, F. Huang, D. Wilson, K. Gopinath, R. Briggs, and E. Geiser, *Using prior shapes in geometric active contours in a variational framework*, International Journal of Computer Vision 50(3): 315-328 (2002).
- [24] Y. Chen, S. Thiruvenkadam, H. Tagare, F. Huang, D. Wilson, and E. A. Geiser, *On the incorporation of shape priors into geometric active contours*, IEEE 1st Workshop on Variational Framework and Level Sets methods (2001).
- [25] Y. Chen, S.R. Thiruvenkadam, K.S. Gopinath, and R.W. Briggs, *Fmr image registration using the mumford-shah functional and shape information*, Intl Conference on Information Systems, Analysis and Synthesis (2002).

- [26] D. L. Chopp, *Computing minimal surfaces via level set curvature flow*, Journal of Computational Physics, 106(1):77-91 (1993).
- [27] L. D. Cohen, *On active contour models and balloons*, Computer Vision, Graphics, and Image Processing. Image Understanding **53** (1991), no. 2, 211–218.
- [28] T.F. Cootes, D. Cooper, C.J. Taylor, and J. Graham, *Active shape models - their training and application*, Computer Vision and Image Understanding. **61** (1995), no. 1, 38–59.
- [29] D. Cremers, *Statistical shape knowledge in variational image segmentation*, Ph.D. thesis, Department of Mathematics and Computer Science, University of Mannheim, Germany, 2002.
- [30] D. Cremers, S. Osher, and S. Soatto, *Kernel density estimation and intrinsic alignment for knowledge-driven segmentation: teaching level sets to walk*, DAGM'04: 26th Pattern Recognition Symposium (2004).
- [31] D. Cremers and S. Soatto, *A pseudo-distance for shape priors in level set segmentation*, Proc. ICCV'2003 (2003).
- [32] D. Cremers, N. Sochen, and C. Schnörr, *Towards recognition-based variational segmentation using shape priors and dynamic labeling*, Intl. Conf. on Scale-Space Theories in Computer Vision (2003).
- [33] D. Cremers, F. Tischhauser, J. Weickert, and C. Schnorr, *Diffusion snakes : introducing statistical shape knowledge into the mumford-shah functional*, ICCV (2002).
- [34] E. P. Crist and R. C. Cicone, *A physically-based transformation of Thematic Mapper data - The TM tasseled cap*, IEEE Trans. Geosci. Remote Sens. **22** (1984), no. 2, 256–263.
- [35] M. de Gunst, *Knowledge based interpretation of aerial images for updating of road maps*, Ph.D. thesis, Technical University of Delft, 1996.
- [36] M.E. de Gunst and J.E. den Hartog, *Knowledge-based updating of maps by interpretation of aerial images*, Proc. ICPR (Jerusalem, Israel), 1994, pp. 811–814.
- [37] Y. Dufournaud, C. Schmid, and R. Horaud, *Image matching with scale adjustment*, Computer Vision and Image Understanding **93** (2004), 175–194.
- [38] Ch. Eidenbenz, Ch. Kaeser, and E. Baltsavias, *ATOMI - Automated reconstruction of topographic objects from aerial images using vectorized map information*, International Archives of Photogrammetry and Remote Sensing, vol. 23, 2000.
- [39] EUROCONTROL and IfEN, *WGS84 implementation manual, version 2.4*, 1998.
- [40] A. Foulonneau, *Une contribution à l'introduction de contraintes géométriques dans les contours actifs orientés région*, Ph.D. thesis, Université Louis Pasteur - Strasbourg I, December 2004.

- [41] A. Foulonneau, P. Charbonnier, and F. Heitz, *Geometric shape priors for region-based active contours*, IEEE Int. Conf. Image Processing, ICIP 2003 (2003).
- [42] M. Gerke and A. Busch, *Verification of a digital road database using IKONOS imagery*, Proc. ISPRS Hannover Workshop, May 2005.
- [43] J. Gomes and O. Faugeras, *Reconciling distance functions and level sets*, Tech. report, INRIA Report no 3666, 1999.
- [44] A. Grün, *Adaptative least square correlation : a powerful image matching technique*, South African Journal of Photogrammetry, Remote Sensing and Cartography, Vol. 14, No. 3, pp. 175-187 (1985).
- [45] A. Grün, E.P. Baltsavias, and O. Henricsson (eds.), *Automatic extraction of man-made objects from aerial and space images*, vol. 2, Birkhauser Verlag Basel, 1997.
- [46] T. Guo and Y. Yasuoka, *Combining high resolution satellite imagery and airborne laser scanning data for generating bare land dem in urban areas*, Proc. International Workshop on Visualization and Animation Of Landscape (2002).
- [47] ———, *Snake-based approach for building extraction from high-resolution satellite images and height data in urban areas*, Proc. of the 23rd Asian Conference on Remote Sensing (2002).
- [48] N. Haala and V. Walter, *Automatic classification of urban environments for database revision using LIDAR and color aerial imagery*, IAPRS, vol. 32, 1999.
- [49] S. Haker, S. Angenent, A. Tannenbaum, and R. Kikinis, *Nondistorting flattening maps and the 3D visualisation of colon ct images*, Proc. MICCAI (2000), 358–366.
- [50] F.J. Harris, *On the use of windows for harmonic analysis with the discrete fourier transform*, Proc. IEEE, vol. 66, pp. 51-88 (1978).
- [51] J.J Helmsen, *A comparison of the three-dimensional photolithography development methods*, Ph.D. thesis, EECS, University of California, Berkeley, 1994.
- [52] P.V.C. Hough, *Machine analysis of bubble chamber pictures*, International Conference on High Energy Accelerators and Instrumentation, CERN, 1959.
- [53] A. Huertas and R. Nevatia, *Detecting changes in aerial views of manmade structures*, Image Visi. Comput. **18** (2000), no. 8, 583–596.
- [54] O. Jamet, *Comparaison vecteurs-images pour la détection de changements d'une base de données topographiques*, Ph.D. thesis, ENST, 1998.
- [55] R. Jeansoulin, *Les images multi-sources en télédétection*, Ph.D. thesis, Université Paul Sabatier, Toulouse, France, 1982.
- [56] S. Jehan-Besson, *Modèles de contours actifs basés régions pour la segmentation d'images et de vidéos*, Ph.D. thesis, Université de Nice-Sophia Antipolis, 2003.

- [57] J. Jensen, *Introductory digital image processing, a remote sensing perspective*, Upper Saddle River, NJ: Prentice-Hall, 1996.
- [58] M. Kass, A. Witkin, and D. Terzopoulos, *Snakes : active contour models*, 1st International Conference on Computer Vision, pp. 259-268 (1987).
- [59] R. Keck, *Reinitialization for level set methods*, Ph.D. thesis, University of Kaiserslautern, 1998.
- [60] R. Keriven, *Equations aux dérivées partielles, évolutions de courbes et de surfaces et espaces d'échelle: Applications à la vision par ordinateur*, Ph.D. thesis, Ecole Nationale des Ponts et Chaussées, December 1997.
- [61] T. Knudsen and A. Nielsen, *Detection of buildings through multivariate analysis of spectral, textural, and shape based features*, Proc. IGARSS, 2004.
- [62] T. Knudsen and B. Olsen, *Photogrammetric Engineering and Remote Sensing* **69** (2003), no. 11, 1289–1296.
- [63] O. Kuebler, P. Agouris, and A. Grün (eds.), *Automatic extraction of man-made objects from aerial and space images*, Birkhauser Verlag, 1995.
- [64] F. Laporterie, *Evaluation of the quality of panchromatic / multispectral fusion algorithms performed on images simulating the future pleiades satellites*, Proc. 2nd GRSS/ISPRS joint workshop on remote sensing data fusion over urban areas (URBAN 2003), May 2003.
- [65] J. Leitloff, S. Hinz, and U. Stilla, *Vehicle queue detection in complex urban areas by extraction and analysis of linear features*, Proc. ISPRS Hanover Workshop, May 2005.
- [66] M. Leventon, E. Grimson, and O. Faugeras., *Statistical shape influence in geodesic active contours*, Comp. Vision and Patt. Recon. (CVPR) (2000).
- [67] R. Lillestrand, *Techniques for change detection*, IEEE Trans. Comput. **21** (1972), no. 7, 654–659.
- [68] R. Malladi and J.A. Sethian, *Level set methods for curvature flow, image enhancement, and shape recovery in medical images*, Proc. of Conf. on Visualization and Mathematics (1995), 329–345.
- [69] R. Malladi, J.A. Sethian, and B.C. Vermuri, *Shape modeling with front propagation: a level set approach*, IEEE Trans. on PAMI **17** (1995), no. 2, 158–175.
- [70] C.R. Maurer, R. Qi, and V. Raghavan, *A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions*, IEEE Transactions on Pattern Analysis and Machine Intelligence **25** (2003), no. 2, 265–270.
- [71] H. Mayer, *Automatic object extraction from aerial imagery - A survey focusing on buildings*, Comput. Vis. Image Underst. **74** (1999), no. 2, 138–149.

- [72] T. McInerney and D. Terzopoulos, *T-snakes: Topology adaptive snakes*, Medical Image Analysis **4** (2000), no. 2, 73–91.
- [73] J. B. Mena, *State of the art on automatic road extraction for gis update: a novel classification*, Pattern Recogn. Lett. **24** (2003), no. 16, 3037–3058.
- [74] D. Mumford and J. Shah, *Boundary detection by minimizing functionals*, Proc. CVPR'85 (1985), 22–26.
- [75] S. Negahdaripour, *Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis*, IEEE Trans. Pattern Anal. Mach. Intell. **20** (1998), no. 9, 961–979.
- [76] J.A. Nelder and R. Mead, *A simplex method for function minimization*, Computer Journal **7** (1965), no. 4, 308–313.
- [77] R. Nevatia, C. Lin, and A. Huertas, *A system for building detection from aerial images*, Workshop on "Automatic extraction of man-made objects from aerial and space images", Birkhauser Verlag Basel, 1997, pp. 77–86.
- [78] M. Niederöst, *Reliable reconstruction of buildings for digital map revision*, International Archives of Photogrammetry and Remote Sensing (Amsterdam), vol. 23, 2000.
- [79] ———, *Automated update of building information in maps using medium-scale imagery (1:15,000)*, Proc. workshop on "Automatic Extraction of Man-Made Objects from Aerial and Space Images" (A. Gruen E. P. Baltsavias and L. Van Gool, eds.), vol. 3, A.A. Balkema, 2001.
- [80] S. Noronha and R. Nevatia, *Detection and description of buildings from multiple aerial images*, DARPA image understanding workshop, 1996, pp. 469–478.
- [81] K. Novak, *Rectification of digital imagery*, Photogrammetric Engineering & Remote Sensing, vol. 58, no 3, pp. 339–344 (1992).
- [82] B.P. Olsen, *Automatic change detection for validation of digital map databases*, ISPRS XXth congress, IAPRS, vol. 24, 2004, pp. 120–125.
- [83] H. Oriot, *Statistical snakes for building extraction from stereoscopic aerial images*, Proc. of the ISPRS Workshop 'Photogrammetric image analysis' (2003).
- [84] S. Osher and J.A. Sethian, *Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations*, Journal of Computational Physics, **79**, pp. 12–49 (1988).
- [85] N. Paparoditis, *Reconstruction tridimensionnelle de paysages urbains en imagerie stéréoscopique satellitale haute résolution*, Ph.D. thesis, Université de Nice-Sophia Antipolis, 1998.
- [86] N. Paparoditis, M. Cord, M. Jordan, and J-P. Cocquerez, *Building detection and reconstruction from mid- and high-resolution aerial imagery*, Computer Vision and Image Understanding, Vol. 72, Issue 2, pp. 122–142 (1998).

- [87] N. Paragios, *Geodesic active regions and level set methods : contributions and applications in artificial vision*, Ph.D. thesis, Université de Nice-Sophia Antipolis, January 2000.
- [88] N. Paragios and R. Deriche, *A PDE-based level-set approach for detection and tracking of moving objects*, Tech. report, INRIA, 1997.
- [89] ———, *Geodesic active contours and level sets for the detection and tracking of moving objects*, IEEE Transactions on Pattern Analysis and Machine Intelligence archive **22** (2000), no. 3, 266–280.
- [90] ———, *Geodesic active regions: A new paradigm to deal with frame partition problems in computer vision.*, Journal of Visual Communication and Image Representation, **13** (2002), no. 1/2, 249–268.
- [91] N. Paragios, O. Mellina-Gottardo, and V. Ramesh., *Gradient vector flow fast geometric active contours*, IEEE Transactions on Pattern Analysis and Machine Intelligence **26** (2004), no. 3, 402–407.
- [92] N. Paragios, M. Rousson, and V. Ramesh, *Matching distance functions: A shape-to-area variational approach for global-to-local registration*, European Conference in Computer Vision (2002).
- [93] T Pavlidis, *Algorithms for graphics and image processing*, Computer Science Press, Inc., 1982.
- [94] R. Péteri and T. Ranchin, *Extraction and update of street network in urban areas from high resolution satellite images*, Proc. of conference ISPRS (2002).
- [95] ———, *Multiresolution Snakes for urban road extraction from IKONOS and Quickbird images*, 23rd EARSeL Annual Symposium "Remote Sensing in Transition" (Ghent, Belgium) (Tomas Benes, ed.), Millpress, Rotterdam, Netherlands, 2-5 June 2003.
- [96] ———, *Urban street mapping using quickbird and ikonos images*, Proc. IGARSS 2003, Volume III, pp. 1721 - 1723 (2003).
- [97] R. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, *Image change detection algorithms: A systematic survey*, IEEE Transactions on Image Processing **14** (2005), no. 3, 294–307.
- [98] T. Riklin-Raviv, N. Kiryati, and N. Sochen, *Unlevel sets: Geometry and prior-based segmentation*, Proc. 8th European Conference on Computer Vision (ECCV'2004) (2004), 50–61.
- [99] M. Rochery, I.H. Jermyn, and J. Zerubia, *Higher order active contours and their application to the detection of line networks in satellite imagery*, Proc. IEEE Workshop on Variational, Geometric and Level Set Methods in Computer Vision (at ICCV, Nice, France), 2003.

- [100] M. Rochery, I.H. Jermyn, and J. Zerubia, *Étude d'une nouvelle classe de contours actifs pour la détection de routes dans des images de télédétection*, Proc. GRETSI Symposium on Signal and Image Processing (Paris, France), 2003.
- [101] ———, *Higher-order active contours*, International Journal of Computer Vision (2005), to appear.
- [102] M. Rousson, *Cue integration and front evolution in image segmentation*, Ph.D. thesis, Université de Nice-Sophia Antipolis, December 2004.
- [103] M. Rousson and R. Deriche, *A variational framework for active and adaptative segmentation of vector valued images*, Tech. report, INRIA, 2002.
- [104] M. Rousson and N. Paragios, *Shape priors for level set representations*, Proc. ECCV (Lecture Notes in Computer Science vol. 2351, ed.), vol. 2, 2002, pp. 78–92.
- [105] M. Roux, *Recalage d'images multi-sources. application au recalage d'une image spot et d'une carte*, Ph.D. thesis, ENST, 1992.
- [106] M. Roux and H. Maître, *Three-dimensionnal description of dense urban areas using maps and aerial images*, Proc. Workshop "Automatic extraction of man-made objects from aerial and space images", Birkhauser Verlag Basel (1997).
- [107] M. Roux and D. McKeown, *Feature matching for building extraction from multiple views*, DARPA image understanding workshop, 1994, pp. 331–349.
- [108] G. Sapiro, *Geometric partial differential equations and image analysis*, Cambridge University Press, 2001.
- [109] S. Servigne, *Base de données géographiques et images aériennes : de l'appariement à la mise à jour*, Ph.D. thesis, INSA Lyon, 1993.
- [110] J.A. Sethian, *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision and materials science*, Cambridge University Press, 1999.
- [111] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, New Jersey, USA, 1976.
- [112] Kaleem Siddiqi, Benjamin B. Kimia, and Chi-Wang Shu, *Geometric shock-capturing eno schemes for subpixel interpolation, computation and curve evolution*, Graph. Models Image Process. **59** (1997), no. 5, 278–301.
- [113] L. H. Staib and J. S. Duncan, *Boundary finding with parametrically deformable models*, IEEE Transactions on Pattern Analysis and Machine Intelligence **14** (1992), no. 11, 1061–1075.
- [114] C. Steger, W. Eckstein, and C. Wiedemann, *Update of roads in GIS by automatic extraction from aerial imagery*, Proceedings of the Second International Airborne Remote Sensing Conference and Exhibition, vol. 3I, Environmental Research Institute of Michigan, 1996, pp. 308–317.

- [115] M. Sussman, E. Fatemi, P. Smereka, and S. Osher, *An improved level set method for incompressible two-phase flows*, Journal of Computers and Fluids **27** (1998), no. 5-6, 663–680.
- [116] Mark Sussman and Emad Fatemi, *An efficient, interface-preserving level set re-distancing algorithm and its application to interfacial incompressible fluid flow*, SIAM J. Sci. Comput. **20** (1999), no. 4, 1165–1191.
- [117] A. Tsai, A. Yezzi, and A.S. Willsky, *Curve evolution implementation of the mumford-shah functional for image segmentation, denoising, interpolation, and magnification*, IEEE Transactions on Image Processing **10** (2001), no. 8, 1169–1186.
- [118] M. S. Ullstad, *An algorithm for estimating small scale differences between two digital images*, Pattern Recognit. **5** (1973), 323–333.
- [119] S. Vinson, L.D. Cohen, and F. Perlant, *Extraction of rectangular buildings using dem and orthoimage*, Proc. of Scandinavian Conference on Image Analysis (SCIA'01) (2001).
- [120] G. Vosselman and M. de Gunst, *Updating road maps by contextual reasoning.*, Proc. Workshop on "Automatic Extraction of Man-Made Objects from Aerial and Space Images" (A. Gruen, E.P. Baltsavias, and O. Henricsson, eds.), Birkhäuser Verlag, 1997, pp. 267–276.
- [121] L. Wald, *Definitions and terms of reference in data fusion*, IAPRS (Valladolid, Spain), vol. 32, June 1999.
- [122] V. Walter, *Automatic classification of remote sensing data for gis database revision*, IAPRS, vol. 32, 1998, pp. 641–648.
- [123] ———, *Automatic change detection in GIS databases based on classification of multispectral data*, Proc. IAPRS, vol. 23, 2000.
- [124] V. Walter and D. Fritsch, *Automatic verification of gis data using high resolution multispectral data*, IAPRS, vol. 32, 1998, pp. 485–490.
- [125] Z. Wang, D. Ziou, C. Armenakis, D. Li, and Q. Li, *A comparative analysis of image fusion methods*, IEEE Transactions on Geoscience and Remote Sensing **43** (2005), no. 6, 1391–1402.
- [126] D.J. Williams and M. Shah, *A fast algorithm for active contours and curvature estimation*, Image Understanding **55** (1992), no. 1, 14–26.
- [127] B. Xie, V. Ramesh, and T. Boulton, *Sudden illumination change detection using order consistency*, Image Vis. Comput. **22** (2004), no. 2, 117–125.
- [128] C. Xu and J. Prince, *Gradient vector flow: A new external force for snakes*, Proceedings of Computer Vision and Pattern Recognition (CVPR '97) (1997), 66–71.

- [129] G. Xu and Z. Zhang, *Epipolar geometry in stereo, motion and object recognition - a unified approach*, Computational Imaging and Vision, vol. 6, Kluwer Academic Publishers, 1996.
- [130] Z. Xue, S. Li, and E.K. Teoh, *Bayesian shape model for facial feature extraction and recognition*, Pattern Recognition (2003).
- [131] Zhong Xue, Stan Z. Li, and E.K. Teoh, *AI-EigenSnake: an affine-invariant deformable contour model for object matching*, Image and Vision Computing **20** (2002), no. 2, 77–84.
- [132] S. Yui, K. Hara, H. Zha, and T. Hasegawa, *A fast narrow band method and its application in topology-adaptative 3-d modeling*, 16th Intl Conf. on Pattern Recognition (ICPR'02) Vol. 4 (2002).
- [133] C. Zhang, E. Baltsavias, and A. Gruen, *Knowledge-based image analysis for 3d road reconstruction*, Asian Journal of Geoinformatics **1** (2001), no. 4, 3–14.
- [134] ———, *Updating of cartographic road databases by image analysis*, Proc. workshop on "Automatic Extraction of Man-Made Objects from Aerial and Space Images" (A. Gruen E. P. Baltsavias and L. Van Gool, eds.), vol. 3, A.A. Balkema, 2001.
- [135] H.K. Zhao, T. Chan, B. Merriman, and S. Osher, *A variational level set approach to multiphase motion*, Journal of Computational Physics **127** (1996), no. 167, 179–195.
- [136] L. Zhu, T. Sasagawa, K. Tachibana, and Y. Fukuzawa, *Triplet image matching for airborne digital sensor ADS40*, Proc. ISPRS XXth ISPRS Congress, 2004.
- [137] Song Chun Zhu and Alan L. Yuille, *Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation*, IEEE Transactions on Pattern Analysis and Machine Intelligence **18** (1996), no. 9, 884–900.

List of publications

1. T. Bailloeul, J. Duan, V. Prinet and B. Serra, *Urban digital map updating from satellite high resolution images using GIS data as a priori knowledge*, in Proc. 2nd GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion on Urban Areas, 2003.
2. T. Bailloeul, V. Prinet, B. Serra, P. Marthon, P. Chen and H. Zhang, *Digital building map refinement from knowledge-driven active contours and very high resolution optical imagery*, in Proc. ISPRS Hannover Workshop, 2005.
3. T. Bailloeul, V. Prinet, B. Serra, P. Marthon, P. Chen and H. Zhang, *Urban building land use change mapping from high resolution satellite imagery, active contours and Hough voting*, in Proc. ISPMSRS, 2005.
4. T. Bailloeul, V. Prinet, B. Serra and P. Marthon, *Spatio-temporal prior shape constraint for level set segmentation*, in Proc. EMMCVPR, LNCS 3757 Springer-Verlag Berlin Heidelberg, pp. 503-519, 2005.

List of Figures

| | | |
|------|---|----|
| 2.1 | Contour delineation errors: different segmentations of the same object by several photo-interpreters. The diagram is taken from [79]. | 17 |
| 2.2 | Generalization effect due to scale. The same scene is represented at different scales: 1:25000 / 1:50000 / 1:100000 (from left to right). The diagram was taken from [79]. | 18 |
| 2.3 | Superposing of a map of the constructions (yellow polygons) with a satellite image. The white arrows point to the maps local shape inexactitudes. | 20 |
| 2.4 | Superposing of a map of the constructions (yellow polygons) with a satellite image illustrating the generalization effect: a single cartographic polygon encompasses numerous adjacent buildings. | 20 |
| 3.1 | Example of a pair of stereoscopic images. The images are satellite data simulated from aerial images. | 25 |
| 3.2 | For the stereoscopic restitution, two images are acquired from two different viewpoints. One point P in the scene is projected onto each image at the intersection of the image plane and of the straight line joining P to the optical center C_A or C_B | 26 |
| 3.3 | For two conjugated points P_A and P_B , the relative orientation enables to calculate the position of point P in the coordinates system linked to the pair of images. P is defined as the center of the smallest segment joining the straight lines (P_A, C_A) and (P_B, C_B) | 28 |
| 3.4 | Example of correlation surface. | 30 |
| 3.5 | Pair of images in any geometry. | 31 |
| 3.6 | Pair of images in epipolar geometry. | 31 |
| 3.7 | Adaptive correlation mask. | 32 |
| 3.8 | Error in the case of occulted zones with a fixed correlation window. (a) a null disparity pixel is considered. (b) its homologous point in the conjugated image is not found because part of the blue building is taken into account in the correlation window. | 33 |
| 3.9 | Application of the mask on the context window. The part of the blue building is withdrawn from the correlation window thanks to the mask adaptive geometry. The homologous pixel is found. | 33 |
| 3.10 | Disparity image generated from the pair in figure 3.1: (a) Conventional correlation (b) Correlation by adaptive masks. | 34 |
| 3.11 | Summary of the 3D reconstruction process. | 35 |

| | | |
|------|--|-----|
| 3.12 | Definition of the attitude angles (roll, pitch, heading) of a camera on-board an aerial platform (assuming that the aircraft and the camera rotation axes are identical). The diagram is taken from [11]. | 36 |
| 3.13 | Illustration of the angles $(\varphi, \omega, \kappa)$ used in photogrammetry. These angles express the rotation between the camera reference (x^B, y^B, z^B) and the terrain reference (x^E, y^E, z^E) associated with a cartographic projection. The diagram is taken from [11]. | 37 |
| 3.14 | Eight fiducial marks represented in the camera's focal plane. The focal plane's reference mark is centered on the optical system's principal point of symmetry (PPS). | 37 |
| 3.15 | Algorithmic summary of the optimization of the image acquisition parameters of n aerial images. | 43 |
| 3.16 | Pair of stereoscopic images rectified in epipolar geometry. | 45 |
| 3.17 | Generated DSM in epipolar geometry. Unfilled pixels are in black; intensity is proportional to the altitude. | 45 |
| 3.18 | Orthoscopic grid associated with the cartographic projection. | 49 |
| 3.19 | Projection of a triangle of the epipolar DSM to the orthoscopic grid. | 49 |
| 3.20 | Multiple triangles projected into the orthoscopic grid and intersecting with the same pixel. | 50 |
| 3.21 | DSM orthorectified with a pitch of $R = 0.65$ m and $\varepsilon_{corr} = 0.4$ | 52 |
| 3.22 | Orthorectified DSM with different correlation thresholds. | 53 |
| 3.23 | Orthoscopic DSM made with a disparity image filtered with a median filter. | 54 |
| 3.24 | Ortho-DSM made with a disparity image filtered with a rejection filter of size 7×7 | 55 |
| 3.25 | Merged DSM: $DSM_1 + DSM_2$ | 58 |
| 3.26 | Enlargement of the red rectangular zone in figure 3.25. | 59 |
| 4.1 | Under-sampling of an explicitly represented contour that is parameterized by its arc length. | 64 |
| 4.2 | Representation in three dimensions of the intersection of a level set with the image plane (zero level). The contour implicitly represented is that of figure 4.1. The caption indicates the signed Euclidean distance of the contour. | 65 |
| 4.3 | Example of the extraction of segments (a) and of the associated gradient vector flow (b). | 86 |
| 4.4 | Circular neighborhoods used for the construction of the narrow rapid band. The caption gives an approximation of the pixel distance to the center. | 89 |
| 4.5 | Linear approximation of a contour | 90 |
| 4.6 | Evolution diagram of the shape constrained active contour | 92 |
| 5.1 | Different representations of the same scene: satellite image and 3D reconstruction | 97 |
| 5.2 | Spatio-temporal variation of the function weighting the shape constraint. $d_0 = 3$, $t_1 = 100$, $t_2 = 400$, $\lambda_{min} = 1.25$ et $\lambda_{max} = 2$ | 101 |

| | | |
|------|--|-----|
| 5.3 | Prior shapes (in white, $\{\mathbf{x}/\psi_0(\mathbf{x}) = 0\}$) and their associated level sets (gray levels, $ \psi_0(\mathbf{x}) $). (a) toy (b) L-shaped building. | 103 |
| 5.4 | Segmentation without insertion of prior shape information and $\xi_0 = \mathbf{0}$. (a) Chan and Vese model (b) Bayesian model. | 105 |
| 5.5 | Segmentation with insertion of prior shape information: (a) Chan and Vese model and $\lambda = 1$ (b) Bayesian model and $\lambda = 10$ | 106 |
| 5.6 | Segmentation with an edge-based active contour (GVF) and $\xi_0 = (s_0 = 1.1, \theta_0 = 0.1, \mu_{x,0} = -4, \mu_{y,0} = 1)$: (a) result without any prior shape constraint. (b) result with prior shape constraint ($\lambda = 5$). | 107 |
| 5.7 | Sensitivity to translation: segmentation with the Chan and Vese model incorporating a shape constraint. The initial active contour of experiments (a), (b) and (c) has been transformed by a direct plane similarity of parameters $\xi_0 = (s_0 = 1.0, \theta_0 = 0.0, \mu_{x,0} = 5, \mu_{y,0} = 5)$ | 109 |
| 5.8 | Sensitivity to rotation: segmentation with the Chan and Vese model incorporating a shape constraint. The initial active contour of experiments (a), (b) and (c) has been transformed by a direct plane similarity of parameters $\xi_0 = (s_0 = 1.0, \theta_0 = 0.3, \mu_{x,0} = 0, \mu_{y,0} = 0)$ | 110 |
| 5.9 | Sensitivity to the scaling factor: segmentation with the Chan and Vese model incorporating a shape constraint. (a) $\xi_0 = (s_0 = 1.5, \theta_0 = 0.0, \mu_{x,0} = 0, \mu_{y,0} = 0)$; (b) $\xi_0 = (s_0 = 0.8, \theta_0 = 0.0, \mu_{x,0} = 0, \mu_{y,0} = 0)$; (c) $\xi_0 = (s_0 = 1.5, \theta_0 = 0.0, \mu_{x,0} = 0, \mu_{y,0} = 0)$ | 111 |
| 5.10 | Sensitivity to translation: segmentation with the Bayesian model and shape constraint. The initial active contour of experiments (a), (b) and (c) has been transformed by a direct plane similarity of parameters $\xi_0 = (s_0 = 1.0, \theta_0 = 0.0, \mu_{x,0} = 5, \mu_{y,0} = 5)$ | 113 |
| 5.11 | Sensitivity to rotation: segmentation with the Bayesian model and shape constraint. The initial active contour of experiments (a), (b) and (c) has been transformed by a direct plane similarity of parameters $\xi_0 = (s_0 = 1.0, \theta_0 = 0.3, \mu_{x,0} = 5, \mu_{y,0} = 5)$ | 114 |
| 5.12 | Sensitivity to the scaling factor: segmentation with the Bayesian model and the shape constraint. (a) $\xi_0 = (s_0 = 1.5, \theta_0 = 0.0, \mu_{x,0} = 0, \mu_{y,0} = 0)$ (b) $\xi_0 = (s_0 = 0.8, \theta_0 = 0.0, \mu_{x,0} = 0, \mu_{y,0} = 0)$ (c) $\xi_0 = (s_0 = 1.5, \theta_0 = 0.0, \mu_{x,0} = 0, \mu_{y,0} = 0)$ | 115 |
| 5.13 | Sensitivity to initialization: segmentation with an active contour based on the contour information (GVF) and with the shape constraint. (a) $\xi_0 = (s_0 = 1.1, \theta_0 = 0.1, \mu_{x,0} = -4, \mu_{y,0} = 1)$ (b) $\xi_0 = (s_0 = 1.0, \theta_0 = -0.1, \mu_{x,0} = -1, \mu_{y,0} = -2)$ (c) $\xi_0 = (s_0 = 1.0, \theta_0 = 0.1, \mu_{x,0} = 0, \mu_{y,0} = 2)$ | 117 |
| 5.14 | Sensitivity to a distant initialization: segmentation with an active contour based on the contour information (GVF) and with the shape constraint. (a) $\xi_0 = (s_0 = 1.1, \theta_0 = 0.25, \mu_{x,0} = -4, \mu_{y,0} = 1)$ (b) $\xi_0 = (s_0 = 1.0, \theta_0 = -0.25, \mu_{x,0} = -2, \mu_{y,0} = -3)$ (c) $\xi_0 = (s_0 = 1.0, \theta_0 = 0.1, \mu_{x,0} = -2, \mu_{y,0} = 2)$ | 118 |

| | | |
|------|--|-----|
| 5.15 | Successful matching with a distant initialization and insertion of the exogenous DSM in the energy functional. (a) Chan and Vese model, (b) Bayesian model. | 120 |
| 5.16 | Failed matching with a distant initialization and the Chan and Vese model (no incorporation of the DSM). | 120 |
| 5.17 | Successful matching with a distant initialization and insertion of the exogenous DSM in the energy functional (GVF model for the image, Chan and Vese model for the DSM). | 121 |
| 5.18 | Failed matching with a distant initialization and the Chan and Vese model (no incorporation of the DSM). | 121 |
| 5.19 | Successful matching with a distant initialization and insertion of the exogenous DSM in the energy functional (GVF model for the image, Chan and Vese model for the DSM). | 122 |
| 5.20 | Failed matching with a distant initialization and the GVF model. | 122 |
| 5.21 | Segmentation with a spatio-temporal variation of the shape constraint and the Bayesian model ($\lambda_{min} = 3$, $\lambda_{max} = 30$, $d_0 = 7$). | 123 |
| 5.22 | Segmentation with a spatio-temporal variation of the shape constraint and the Bayesian model ($\lambda_{min} = 27$, $\lambda_{max} = 70$, $d_0 = 7$). | 124 |
| 5.23 | Segmentation with a constant shape constraint weight and the Bayesian model ($\lambda = 50$). | 124 |
| 5.24 | Spatio-temporal variation of lambda with the GVF model and $\lambda_{min} = 1$, $\lambda_{max} = 5$, $d_0 = 5$: (a) $\xi_0 = (s_0 = 1.1, \theta_0 = 0.25, \mu_{x,0} = -4, \mu_{y,0} = 1)$ (b) $\xi_0 = (s_0 = 1.0, \theta_0 = -0.25, \mu_{x,0} = -2, \mu_{y,0} = -3)$ (c) $\xi_0 = (s_0 = 1.0, \theta_0 = 0.1, \mu_{x,0} = -2, \mu_{y,0} = 2)$ | 125 |
| 5.25 | Segmentation with spatio-temporal variation of the shape constraint weight (Bayesian model): (a) $\lambda_{min} = 15$, $\lambda_{max} = 100$, $d_0 = 15$ (b) $\lambda_{min} = 50$, $\lambda_{max} = 100$, $d_0 = 10$ | 126 |
| 5.26 | Failed matching with a constant shape constraint weight (Bayesian model): (a) $\lambda = 20$ (b) $\lambda = 70$ | 126 |
| 5.27 | Failed matching (Bayesian model) with: (a) single temporal variation of the shape constraint with $\lambda_{min} = 30$, $\lambda_{max} = 70$ (no spatial relaxation); (b) spatio-temporal variation of the shape constraint with a spatial profile defined in equation (5.11) and $\lambda_{min} = 27$, $\lambda_{max} = 70$, $d_0 = 7$; (c) spatio-temporal variation of the shape constraint with Heaviside step functions for $\lambda_a(t)$ and $d(t)$ | 128 |
| 5.28 | Superimposition of the ground truth (grayed) and a locally erroneous cartographic polygon. | 129 |
| 5.29 | Final convergence of the active contour with an erroneous prior shape. | 130 |
| 5.30 | Matching by GVF-based active contour and with a non-erroneous prior shape. | 130 |
| 5.31 | Matching by GVF-based active contour and with an erroneous prior shape derived from the map. | 131 |
| 6.1 | Modified Bessel functions of the second kind. | 140 |
| 6.2 | Diagram describing the sampling mode with a Bessel function for profile Ψ | 141 |

| | | |
|------|--|-----|
| 6.3 | Synthetic image with the initial contour shown in red. | 142 |
| 6.4 | (a) Prior shape and its associated level set (b). | 142 |
| 6.5 | Result with the conventional shape constraint. | 143 |
| 6.6 | Segmentation with shape constraint authorizing parallel variations: (a) $\lambda_{para} = 5$; (b) $\lambda_{para} = 10$ | 143 |
| 6.7 | Segmentation with shape constraint authorizing parallel variations: (a) $\lambda_{para} = 15$; (b) $\lambda_{para} = 30$ | 144 |
| 6.8 | Segmentation with the Chan and Vese attachment to data model and the shape constraint authorizing parallel variations: (a) without quadratic correction and $\lambda_{para} = 20$ at instant n_0 ; (b) with quadratic correction, $\lambda_{para} = 20$, $\lambda_{quad} = 8$ and $a = 0.5$ | 145 |
| 6.9 | Representation of the quadratic force at instant n_0 . The blue zones correspond to a positive force (dilation effect), and the green zones represent a negative force (compression effect). | 145 |
| 6.10 | Segmentation with the Chan and Vese model and the shape constraint authorizing the parallel variations: (a) without quadratic correction and $\lambda_{para} = 50$ at instant n_0 ; (b) with quadratic correction, $\lambda_{para} = 50$, $\lambda_{quad} = 7$ and $a = 0.5$ | 146 |
| 6.11 | Representation of the quadratic force at instant n_0 . The blue zones correspond to a positive force (dilation effect), and the green zones represent a negative force (compression effect). | 146 |
| 6.12 | Experiment with a real image, the prior shape is the result of the matching with the conventional shape constraint. The segmentation is carried out with the Chan and Vese model, the shape constraint authorizing the parallel variations and with quadratic correction $\lambda_{para} = 4$, $\lambda_{quad} = 0.5$ and $a = 0.5$ | 147 |
| 7.1 | Proposed change detection methodology: (A) calculation and fusion of change indices. (B) attenuation of map/image exogenous variabilities (C) analysis of final changes. Legend: Qm and Qp are respectively the multispectral and panchromatic satellite images. | 150 |
| 7.2 | Registration of the building map (shown in yellow) with the DSM using geocoding information. The red arrow points to a change: building replaced by a terreplein. | 153 |
| 7.3 | Result of change detection between the map and the DSM. Legend: red (change), yellow (ambiguous), blue (undecidable). Parameters: $thres_{filled} = 10\%$, $thres_{flat} = 1\text{ m}$, $thres_{build} = 3\text{ m}$ | 154 |
| 7.4 | The method's success in relation to $thres_{filled}$. The rate of buildings correctly detected as having changed is shown in red. The blue curve represents the rate of buildings detected as having changed with respect to the total number of buildings in the map. $thres_{flat} = 1\text{ m}$ and $thres_{build} = 3\text{ m}$ | 155 |
| 7.5 | The method's success in relation to $thres_{flat}$. The rate of buildings correctly detected as having changed is shown in red. The blue curve represents the rate of buildings detected as having changed in relation to the total number of buildings on the map. $thres_{filled} = 50\%$ and $thres_{build} = 3\text{ m}$ | 156 |

| | | |
|------|---|-----|
| 7.6 | The method's success in relation to $thres_{build}$. The rate of buildings correctly detected as having changed is shown in red. The blue curve represents the rate of buildings detected as having changed in relation to the total number of buildings on the map. $thres_{filled} = 65\%$ and $thres_{flat} = 1.5m$ | 156 |
| 7.7 | Infra-red band, real colors (RGB), and $NDVI$ | 159 |
| 7.8 | Estimation of the maximum μ_1 and minimum μ_2 translation between a map segment [AB] and an image segment [A'B']. | 163 |
| 7.9 | 2D Hough accumulator (μ_x, μ_y) . Accumulation between minimum and maximum translation is enlarged in a band of l pixels. | 163 |
| 7.10 | Detection of unequivocal changes. Changes were detected by comparison with the $NDVI$ image. The figures illustrate cartographic objects superimposed on a more recent satellite image. | 168 |
| 7.11 | Change detection due to strong geometrical variation: $E_{geom} = 0.54$ and $p_{NC} = 0.30$. Change is simulated. | 168 |
| 7.12 | Change detection due to a high μ_{Hough} translation. $s_{Hough} = 0.27$, $\mu_{Hough} = (-6, -26)$ and $p_{NC} = 0.00$. Change is simulated. | 169 |
| 7.13 | Detection of non-change with a perfect cartographic object transformed by similarity of parameters $\xi_0 = (s_0 = 1, \theta_0 = 0.1, \mu_{x,0} = 1, \mu_{y,0} = 1)$. Non-change is confirmed with $s_{Hough} = 0.76$, $\mu_{Hough} = (1, 0)$, $E_{geom} = 0.1$ and $p_{NC} = 0.63$ | 170 |
| 7.14 | Detection of non-change with a perfect cartographic object transformed by similarity of parameters $\xi_0 = (s_0 = 1.05, \theta_0 = 0.075, \mu_{x,0} = 4, \mu_{y,0} = -1)$. Non-change is confirmed with $s_{Hough} = 0.72$, $\mu_{Hough} = (1, 0)$, $E_{geom} = 0.08$ and $p_{NC} = 0.63$ | 171 |
| 7.15 | Detection of non-change with a perfect cartographic object transformed by similarity of parameters $\xi_0 = (s_0 = 1.0, \theta_0 = -0.02, \mu_{x,0} = -2, \mu_{y,0} = 1)$. Non-change is confirmed with $s_{Hough} = 0.81$, $\mu_{Hough} = (0, 0)$, $E_{geom} = 0.04$ and $p_{NC} = 0.76$ | 171 |
| 7.16 | Detection of non-change with a perfect cartographic object transformed by similarity of parameters $\xi_0 = (s_0 = 1.0, \theta_0 = 0.0, \mu_{x,0} = 5, \mu_{y,0} = 5)$. Non-change is confirmed with $s_{Hough} = 0.84$, $\mu_{Hough} = (-1, 0)$, $E_{geom} = 0.01$ and $p_{NC} = 0.83$ | 172 |
| 7.17 | Detection of non-change with a matching error that can be modeled by a global transformation. Non-change is confirmed with $s_{Hough} = 0.86$, $\mu_{Hough} = (0, -1)$, $E_{geom} = 0.12$ and $p_{NC} = 0.68$ | 173 |
| 7.18 | Detection of non-change with a matching error that can be modeled by a global transformation. Non-change is confirmed with $s_{Hough} = 0.88$, $\mu_{Hough} = (0, 0)$, $E_{geom} = 0.03$ and $p_{NC} = 0.83$ | 174 |
| 7.19 | Detection of non-change with local shape errors. Non-change is confirmed with $s_{Hough} = 0.62$, $\mu_{Hough} = (1, 0)$, $E_{geom} = 0.02$ and $p_{NC} = 0.66$ | 174 |
| 7.20 | Detection of non-change with local shape errors. Non-change is confirmed with $s_{Hough} = 0.73$, $\mu_{Hough} = (1, 0)$, $E_{geom} = 0.03$ and $p_{NC} = 0.71$ | 175 |
| 7.21 | Detection of non-change with local errors of shape. Non-change is confirmed with $s_{Hough} = 0.75$, $\mu_{Hough} = (1, 0)$, $E_{geom} = 0.05$ and $p_{NC} = 0.71$ | 175 |

| | | |
|------|---|-----|
| 7.22 | Ambiguous detection (indeterminate case). $s_{Hough} = 0.45$, $\mu_{\mathbf{Hough}} = (0, 0)$, $E_{geom} = 0.02$ and $p_{NC} = 0.55$ | 176 |
| 7.23 | Erroneous change detection due to local shape errors on the map. $s_{Hough} = 0.47$, $\mu_{\mathbf{Hough}} = (-1, -1)$, $E_{geom} = 0.19$ and $p_{NC} = 0.39$ | 177 |
| 7.24 | Illustration of delineation errors and of generalization effect in the map of 1996. The cartographic object is superimposed on the satellite image of 2002. | 179 |
| A.1 | Aerial image shown at 5% of its original size. The future Olympic village will be situated in the north-east of the image. | 189 |
| A.2 | Beijing city: extract from the GIS data at a 1:10000 scale. The constructions are in pink, the road network in purple, the rivers and canals in blue, the lakes in green. | 192 |
| B.1 | Initial simplex and reflection. | 195 |
| B.2 | Expansion | 195 |
| B.3 | Possible contractions | 196 |
| B.4 | Shrinking | 196 |
| C.1 | Diagram illustrating inverse modeling. The terrain and camera references are direct. | 198 |
| C.2 | Diagram illustrating the direct modeling. The terrain and camera references are direct. | 200 |
| F.1 | 3D visualization of Beijing city - with textures and no DTM. | 209 |

List of Tables

| | | |
|-----|--|-----|
| 2.1 | Classification of urban artifacts according to the different universes. . . | 18 |
| 2.2 | Possible changes of a building in $U_{objective}$ | 19 |
| 3.1 | Results in pixels of the absolute precision quantification of the image acquisition parameters for the three aerial images. | 39 |
| 3.2 | Results in pixels of the quantification of the relative precision of the image acquisition parameters for each of the pairs of aerial images. | 40 |
| 3.3 | Results of the optimization of the image acquisition parameters on the images with a resolution of 0.21 m/pixel. | 44 |
| 3.4 | Results of the optimization of the image acquisition parameters on the images subsampled by a factor of three. | 44 |
| 3.5 | Absolute error between the GCPs and the 3D reconstructions of the DSM_1 . ΔN : error in the northern direction, ΔE : error in the eastern direction, ΔZ altimetric error, error in the plane: $\Delta Plane = \sqrt{\Delta N^2 + \Delta E^2}$. 47 | 47 |
| 3.6 | Absolute error between the GCPs and the 3D reconstructions of the DSM_2 . ΔN : error in the northern direction, ΔE : error in the eastern direction, ΔZ altimetric error, error in the plane: $\Delta Plane = \sqrt{\Delta N^2 + \Delta E^2}$. 47 | 47 |
| 5.1 | Absolute error between ξ_0 and its estimate $\hat{\xi}_0$ by the gradient descent algorithm and calculation time. | 104 |
| 5.2 | Absolute error between ξ_0 and its estimate $\hat{\xi}_0$ by the simplex algorithm and calculation time. | 104 |
| 5.3 | Ratio between the convergence times without and with DSM merging, $ratio = \frac{t_{without\ fusion}}{t_{with\ fusion}}$ | 123 |
| 5.4 | Ratio between the convergence times without and with spatio-temporal shape constraint: $ratio = \frac{t_{uniform\ constraint}}{t_{spatio-temporal\ constraint}}$ | 128 |
| 7.1 | Quantification of change/non-change scores. p_{NC} is the probability of non-change. s_{Hough} is the normalized Hough score. | 178 |
| 7.2 | Quantification of change detection results with the 2002 map. Results are expressed as a percentage over a total of 77 cases. | 180 |
| 7.3 | Quantification of change detection results with the 1996 map. Results are expressed as a percentage over a total of 105 cases. | 181 |