Revised version

In Proc. Asian Conference on Computer Vision (ACCV), Jan. 13-16 2006, India

PROBABILISTIC MODELING FOR STRUCTURAL CHANGE INFERENCE

Wei LIU - Véronique PRINET

National Laboratory of Pattern Recognition (NLPR)
Sino-French research Laboratory in Information, Automation and Applied Mathematics (LIAMA)
Institute of Automation, Chinese Academy of Sciences (CASIA)
prinet@nlpr.ia.ac.cn

ABSTRACT

We view the task of change detection as a problem of object recognition from learning. The object is defined in a 3D space where the time is the 3rd dimension. We propose two competitive probabilistic models. The first one has a traditional regard on change, characterized as a 'presence-absence' within two scenes. The model is based on a logistic function, embedded in a framework called 'cut-and-merge'. The second approach is inspired from the Discriminative Random Fields (DRF) approach proposed by Ma and Hebert [KUMA2003]. The energy function is defined as the sum of an association potential and an interaction potential. We formulate the latter as a 3D anisotropic term. A simplified implementation enables to achieve fast computation in the 2D image space. In conclusion, the main contributions of this paper rely on : 1) the extension of the DRF to a 3D manifold ; 2) the *cut-and-merge* algorithm. The application proposed in the paper is on remote sensing images, for building change detection. Results on synthetic and real scenes and comparative analysis demonstrate the effectiveness of the proposed approach.

1. INTRODUCTION

This paper focuses on probabilistic modeling for object structural change inference.

Probabilistic modeling emerged as an increasingly efficient framework for segmentation and object recognition tasks. From Markov Random Fields [13] to recently introduced part-based modeling for recognition [4], the goal is to find a 'best' global configuration of a random variable associated to a label. Set as an energy minimisation problem, the basic challenges are : i) to define appropriately the functional to be optimised ; it is in general characterised by a data features term and a constraint or regularisation term; the latter is mainly derived from prior knowledge and enable to introduce smoothness or to sharpen edges. ii) to set efficient computational solution to reduce the combinatorial calculation burden [22].

The Remote Sensing (RS) image processing field has so far be little influenced by Pattern Recognition (PR) and Computer Vision (CV) works. However, the two communities deal with very similar key problems –restoration, segmentation and classification, flow vector field, reconstruction,... Far to be easy cases to handle, the complexity of the scenes and the high level of noise existing in RS images challenge the models developed from PR and CV, which often are not robust nor generic enough. RS images are characterised in particular by illumination changes, shadowing, projective distortion, occlusion, stochastic noise –thermal effect during the acquisition– and geometric noise. Geometric noise refers to all small size objects that appear in the scene and disturb the segmentation or detection process, such as cars, trees, etc. A contrary to man-made object detection in natural scenes, an RS image scene covering an urban area is mainly composed of objects which have similar structured and polygonal shapes: roads and buildings. To takle this problem, previous works on buildings detection in dense urban areas are mainly based on models which include high and restrictive constraints, therefore lacking of genericity and uncapable to work on a large set of images [6,11].

The present work was inspired from a recent publication from Kumar [8], in which he introduced a Discriminative Random Field (DRF) model, for man-made objects detection in natural scenes. This model formulates the constraint term of the Gibbs energy as a function of the global image.

We show in this paper how to generalise the concept of DRF modeling to a multi-dimensional space without increasing computational cost. Application is on building change retrieval from high resolution optical remote sensing images covering dense urban areas. The rest of this article is organized as follows: the next section introduces structural object change inference from a Kernel approach based on a so-called *cut-and-merge* algorithm and maximum likelihood. Section 3 presents the 3D anisotropic model derived from DRF and its practical implementation. Results are illustrated in section 4; the 5th section concludes the paper.

2. STRUCTURAL CHANGE RECOGNITION

2.1. Overview

The main idea of structural change recognition is to first perform object segmentation in each of the two images, then to analyse the probability of change, for each individual object. We do not perform a 'hard' segmentation, in the sense that an object can be detected even if its likelihood of being a building is low. It enables us to estimate the change as a cross product of probability functions, which appears to be much more powerful than computing a simple difference. The rest of this section gives the main cues of the approach.

2.2. Cut-and-merge algorithm

The cut-and-merge algorithm performs binary segmentation without explicit thresholding. The *cut* task will blindly binarize the input image and create a set of black-and-white images, without knowledge of the object's type we are seeking for. The *merge* task will make use of prior knowledge and fuse this segmented set of data such as to retrieve the regions which maximise a given 'criteria'.

Cut : A band-pass filter is convolved several times with the original image [18]. The filter is characterised by the min and max values of the band. All pixels x with intensity level $I(x)$ satisfying $min < I(x) < max$ are retained. From one convolution to an other, min and max values are progressively incremented, thus generating a set of binary images. In each of these images, aggregated pixels define regions which closed contours are extracted.

Merge : The likelihood of a each region and its closed contour R_i to be a object-building is given by $p_i = p(R_i)$ (see section 2.3). For two overlapping regions R_i and R_j resulting from two different pass-band filters, we calculate $p_{ij} = p(R_i \bigcap R_j)$. We then merge the overlapping regions by retaining the one R_k that verifies :

$$k = arg \max_{k \in \{1, j, ij\}} (p_i, p_j, p_{ij})$$

In an iterative procedure, we can therefore eliminate all regions unlikely to correspond to our searched object.

2.3. Functional and features

The problem is to determine whether a segmented area R_i –called "element"– is likely to be a object-building or not. The label assigned to R_i is represented by a random value x_i. Assuming that events are independent, finding the configuration X that maximizes the probability P over the image is equivalent to maximizing the probability density function at each site R_i : $P(X|y) = \Pi_i p(x_i|y_i)$ where y_i is the features vector computed at element R_i. We define $p(x|y)$ by a logistic function :

$$p(x_i|y_i) = \sigma(x_i w^T f(y_i))$$

w is the unknown parameters vector and f is an application that transforms the data features to a higher dimensional space: $f : y_i \rightarrow [1, y_i, y_i y_i^T, \tilde{y}]$, where $\tilde{y}$ is defined as the product of each feature with an other. A total of eight individual feature are computed —it includes : region entropy, edges points, intensity mean, standard deviation, gradient direction moment and their difference, shadow parameter–, leading to a parameter vector size of 45 degrees of freedom. Parameters are retrieved by maximizing the log-likelihood of P via a ICM module.
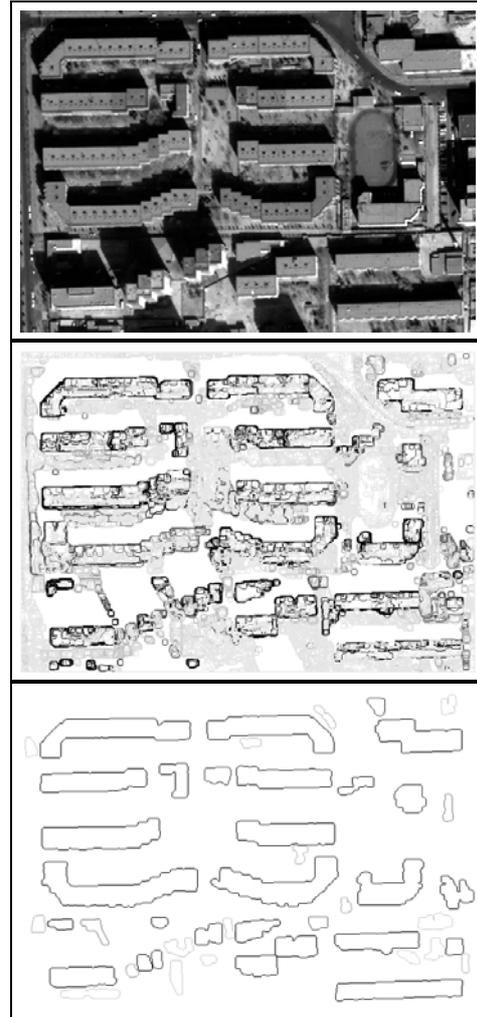

Fig. 1. Illustration of the *cut-and-merge* algorithm. Top: original image. Middle: superimposition of the contours extracted during the *cut* step; Bottom: building candidates resulting from the *merge* step. The grey level of the contours is inversely proportional to the probability to delineate a building-like object.

Revised version

In Proc. Asian Conference on Computer Vision (ACCV), Jan. 13-16 2006, India

2.4. Object change detection

We consider two images I^1 and I^2 acquired at t^1 and t^2 respectively. We estimate the probability density function p.d.f. at each of the elements i of R^1 and j of R^2: $p_i^1 = p(y_i^1)$ and $p_j^2 = p(y_j^2)$ respectively.

The probability function that an object appears in the two images simultaneously is given by $p_{nc} = p_i^1 p_j^2$, while the probability function for an object to appear in one of the image only is : $p_c = p_i^2(1 - p_i^1) + p_i^1(1 - p_j^2)$. Selecting candidate contours that verify $p_c > 0.5$ provides us with the changed structured objects.

3. CHANGE INFERENCE FROM 3D DRF

3.1. The model

We first recall some basic notations and define the main concept of 3D DRF modeling. A image is represented as a graph $G = (V, E)$, where V are the vertices/nodes of the graph and E are the edges between vertices. In order to create G, the image is divided into regular patches. Each patch is a node and two mutually connected nodes within a n-neighborhood determine a clique. Each node $v \in V$ is characterised by the data it encompasses – it can also be features computed from the data– : y_v. The label associated to each node is a random variable $x_{v \in V} = \{-1, 1\}$, where value 1 stands for 'true' (i.e. "there is a building-like object in this patch") and -1 for 'false' ("no building-like object here").

Considering the set of multi-temporal images as a 3D data, G is defined in the 3D space (2 spatial dimensions + 1 temporal dimension). G is the union of k 2D spatio-subgraphs $G^{s,i \in \{1,..,k\}}$, linked by temporal edges E^t, where s and t denote the spatial and temporal indices respectively. Then $G = \{\bigcup_i G^{s,i}, \bigcup_j E^{t,j}\}$, $i \in \{1,..,k\}, j \in \{1,..,k-1\}$. Then, in the specific case of two images: $G^{s,i=\{1,2\}} = (E^{s,i}, V^{s,i})$ with $i = \{1, 2\}$. Pair-wise cliques associated to each node v cover a 5-neighborhood N_v characterised by its four edges in the image domain, augmented with a unique edge in time domain.

The probability distribution to retrieve the configuration of the random variable X given the input data features $Y = \{y_v\}$ is expressed by :

$$P(X|Y) = \frac{1}{Z} e^{-U}$$

where $U = U(X, Y)$ is the total Gibbs energy and Z is the partition function. In [8], Kumar defines U as the sum of an association potential A and an interaction potential I, such that :

$$U(X, Y) = \sum_{v \in G} \gamma_v A(x_v, y_v) + \sum_{v \in G} \sum_{v' \in N_v} \beta_{v,v'} I(x_v, y_v, x_{v'})$$

N_e is the five-neighboring. $A(x_v, y_v)$ is the association potential; the interaction term $I(x_v, y_v, x_{v' \in N_v})$ is a smoothness factor. It determines how much a site is similar or not to its associated neighboring sites. A and I are defined as parametric logistic functions which exact formulation can be found in [8].

Knowing that object changes are characterised by continuity in spatial neighborhood and discontinuity in temporal neighborhood, we consider an anisotropic formulation of U given by :

$$\beta_{v,v'} = \beta_{v,v''} \quad iif \quad (v, v', v'') \in G^{s,i}$$

$$\beta_{v,v'} \neq \beta_{v,v''} \quad otherwise$$

The anisotropic constraint makes it possible the detection of 3D objects which structured shape –i.e. building-like shape– is observed in one image but not in the other. We call $C = \{c_v\}_{v \in G^1}$ the hidden variable defined such as : $c_v = x_v^1 x_{v'}^2$, for which v is connected to v' by $E(v, v') \subset E^t$. Then, in a straightforward manner:

$$P(C|Y) \equiv P(X|Y)$$

$P(C|Y)$ is the probability of structural changes, defined at each node of each of the 2D images taken individually. Note that C is defined over the *projection* of G_1 and G_2 in the spatial dimension and has 'lost' the 3rd temporal dimension.

3.2. Computational issues

In order to fasten the computation, it is possible to implement the model in its 2D formulation, while modifying the choice of the features, such that new features by them selves are charactistics of a structural change. The simplest way is to define new features' vector as the difference of features computed from bi-dimensional images.

Parameters are estimated by maximizing the pseudo-likelihood, using a large training image set which has been manually labelled. On the testing images set, the optimal configuration C is computed using ICM.

4. RESULTS ANALYSIS

The two proposed methods have been implemented and validated on composite and real remote sensing images. Composite images were artificially created by mapping locally some small textured patches onto real images: it will validate the method without illumination change. Remote sensing images are from Quickbird satellite (resol. 0.6m/pixel, panchromatic, acquired in 2002 and 2003 ; covering the area of Beijing city). Ground truth is given by manual segmentation. Note that Beijing area is particularly interesting to study because of the rapid undergoing changes in preparation of the 2008 Olympic games.

We used, for the two models, a training set of nearly 2000 object sites manually detected from 10 sub-images. The Kernel computation (section 2) takes about 4mn on Pentium4. Cost comes from contour feature calculation while the optimization per see only takes less than 10 seconds.

Figure 1 illustrates the principle of the *cut-and-merge* algorithm and building candidates selection. Results from the object change detection Kernel approach are shown in figures 2 and 4. Tables 1 and 2 give statistical analysis of the results. We recall that the ultimous purpose is not to delineate precisely the new/old buildings, but to locate the changed objects, as indicated by the crosses. The kernel approach gives a precise counting of the changed objects.

Good performance obtained from 3D DRF (section 3) on a toy picture without noise validate the model definition and features choice (fig.5). Note that only the changes of structured objects are detected. Figures 3 and 4 illustrate the approach and compare it to the kernel method. The DRF model acts as a region-of-interest detector, by giving rough location of areas where structural changes have appeared.

5. CONCLUSION

We proposed in this paper two probabilistic framework for structural change inference in complex scenes. The first one, closely related to classical differential methods, computes the changes based on a likelihood function, which makes the approach very robust and enable to decrease erroneous detections rate. The second model is derived from 3D DRF modeling, where the third dimension is the temporal component. Its fast implementation in 2D space makes it extremely efficient. One may notice however that the proposed model formulation is not invariant by symetry with respect to temporal axis. From a practical point of view, the two approaches differ by the output they provide: 3D DRF gives a rough location of structural changes ; the minimal area is set by the size of the patches used for the computation ; a contrario, in the *cut-and-merge* algorithm, the segmentation step delineates the buildings' shape and makes possible the counting of changed objects. Extension of this work could be on object tracking from video sequence.

GT	TP	FP	DR	CR
62	45	22	0.72	0.67

Table 1. Statistical figures on change detection (section 2). GT=Nber of changed-objects given by the ground truth; TP=True positive; FP=False positive; DR=Detection rate=TP/GT ; CR= Correctness rate=FP/(FP+TP).

TP	FP	FN	AreaTrueR
0.834	0.058	0.166	0.861

AreaErrorR	AreaLosetR	Peri Rate	AreaRate
0.060	0.139	0.244	0.189

Table 2. Statistical figures on building detection from cut-and-merge. FN=False negative; AreaTrueRate: rate of the surface size of the object that have been correctly recovered; AreaError-Rate: rate of the surface area which is out side true building; AreaLostRate: rate of the surface area which has not been detected.
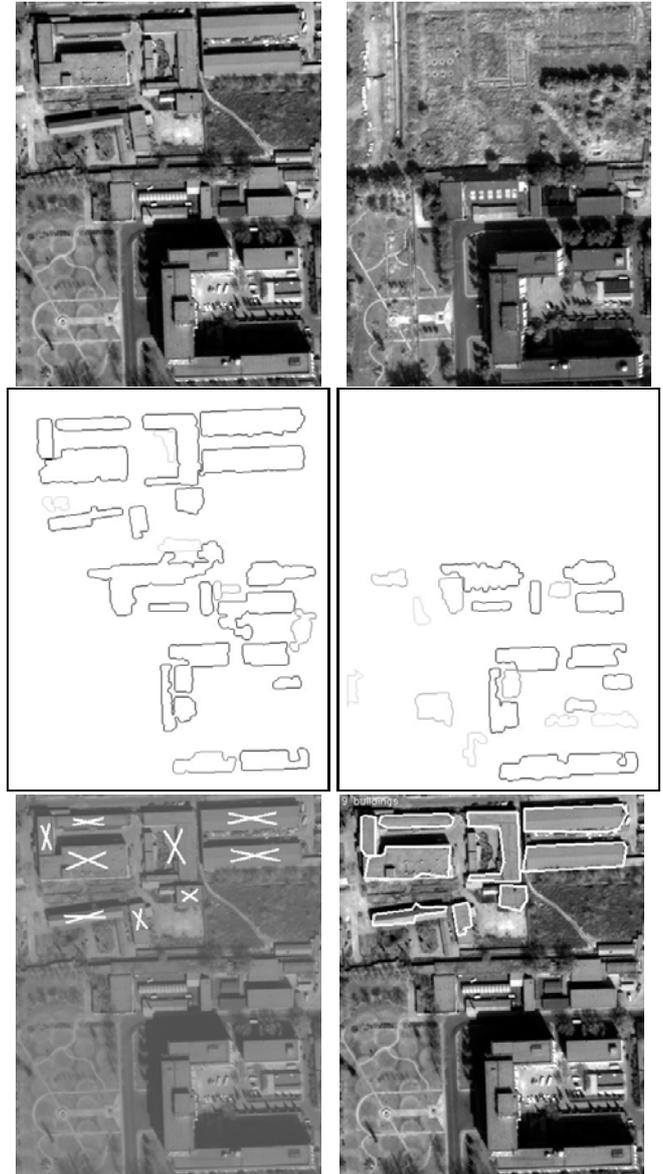

Fig. 2. Top: Original images. Center: candidate buildings' contours retrieved from *cut-and-merge*; Bottom: building-object changes represented by polygonal approximation of the contours (right) or marked with a cross (left) –black for new building, white for disappeared.

Fig. 3. Illustration of the 3D DRF on real Quickbird images acquired in 2002 (right) and 2003 (left). Patches indicate areas detected as changes (patch size=64x64pixels). New buildings are properly recovered but false detections also appear.
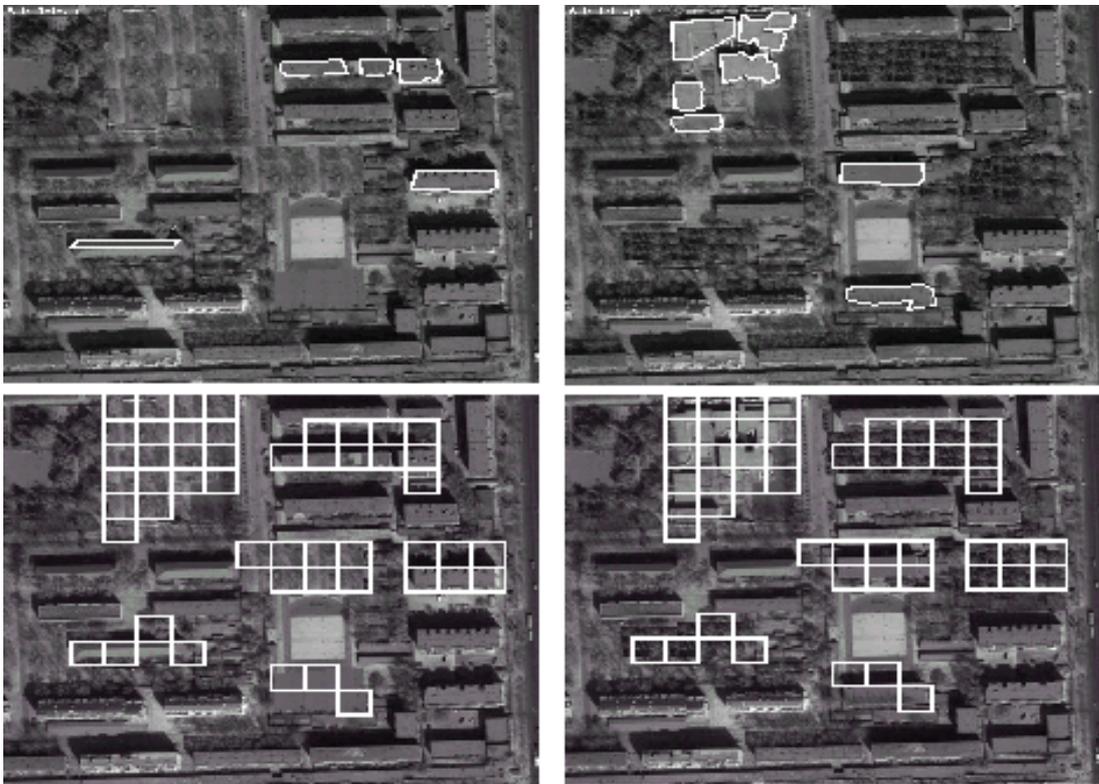

Fig. 4. Results comparison from 3D DRF (bottom) and kernel based approaches (top), on a composite image. Top: white lines indicate the changed objects retrieved as destroyed and new buildings ; Bottom: White squares indicate the location of structural changes, as detected by the DRF model.

6. REFERENCES

[1] J. Besag. "On the statistical analysis of dirty pictures". Journal of Royal Statistical Soc., B-48:259-302, 1986

[2] L. Garcin, X. Descombes, J. Zerubia, and H. Le Men. "Building extraction using a Markov point process". In Proc. ICIP, Greece, 2001.

[3] L.M.T. Carvalho, L.M.G.Fonseca. "Digital change detection with the aid of multiresolution wavelet analysis". Int. J. Remote. Sens., 2001, vol.22, no.18, pp.3871-3876.

[4] D. Crandall and P. Felzenszenzwald. "Spatial priors for part-based reco-gnition using statistical methods". In Proc.CVPR, SanDiego, June 2005

[5] S. Geman and D. Geman. "Stochastic relaxation, Gibbs Distribution and Bayesian Restoration of Images", IEEE PAMI vol.6, no.6, Nov. 1984.

[6] R.B. Irvin and D.M. McKeown. "Method for exploiting the relationship between buildings and their shadows in aerial imagery". IEEE Trans. Syst. Man Cybern, vol. 19, pp.1564-1575, 1989.

[7] K. Kulschewski. "Building recognition with Bayesian Networks". Workshop on Semantic Modelling for the Acquisition of Topographic Images and Maps, Bonn, Germany, 1997.

[8] S. Kumar and M. Hebert. "Discriminative Random Fields: A Discriminative Framework for Contextual Interaction in Classification". In Proc. ICCV, 2003.

[9] S. Kumar and M. Hebert. "Man-Made Structure Detection in Natural Images using a Causal Multiscale Random Field". In Proc. CVPR, 2003.

[10] S. Krishnamrchari and Ag. "Delineating Buildings by Grouping Lines with MRFs". IEEE Trans. on Image Processing, vol. 5, NO. 1., 1996.

[11] C.G. Lin and R. Nevatia. "Building Detection and Description from a Single Intensity Image". CVIU, vol 72, no2, Nov.1998, pp101-121.

[12] J. Lafferty, A. McCallum, and F. Pereira. "Conditional random fields: Probabilistic models for segmenting and labeling sequence data". In Proc ICML, 2001.

[13] S. Z. Li. Markov Random Field Modeling in Computer Vision. Springer-Verlag, Tokyo, 2001.

[14] D. Marr. Vision. Ed. Freeman and Company, NY, 1982.

[15] Hans-Gerd Maas. "Closed solutions for the determination of parametric building models from invariant moments of airborne laserscanner data". IAPRS, vol. XXXII (B3), 1999.

[16] Helmut Mayer."Automatic Object Extraction from Aerial Imagery: A Survey Focusing on Buildings". CVIU, vol. 74 no.2, pp. 138-149, 1999.

[17] T. P. Minka. "Algorithms for Maximum-Likelihood logistic Regression". Statistics Tech Report 758, Carnegie Mellon University.

[18] Open CVcourses, CVPR01 course "(http://sourceforge.net/projects/opencvlibrary)".

[19] N. Paparoditis, M. Cord, M. Jordan, and J.-P. Cocquerez. "Building Detection and Reconstruction from Mid- and High-Resolution Aerial Imagery". CVIU, vol 72 no2, Nov.1998.

[20] V. Venkateswar and R. Chellappa. "A hierachical approach to detection of buildings in aerial images". Tech. Rep. CAR-TR 567, Univ. of Maryland, 1991.

[21] Special section on Syntactic and structural pattern recognition, IEEE pattern analysis and machine intelligence, vol.27 no.7, July 2005

[22] Y. Boykov. "Discrete Optimization Methods in Computer Vision", SanDiego, In Proc. CVPR 2005.
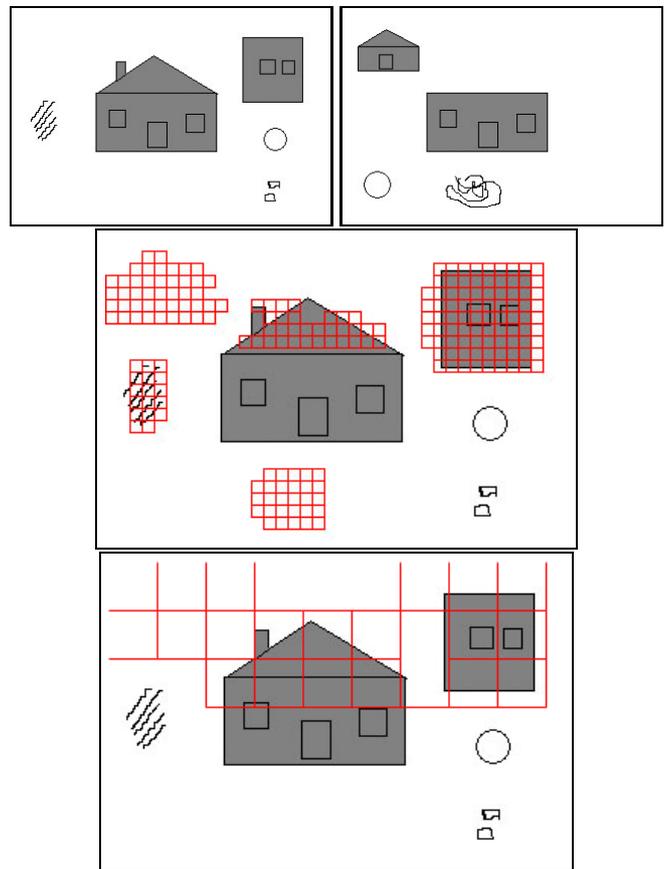

Fig. 5. Structural changes from 3D DRF - Application to a toy picture. Top: original images. Middle and bottom: patches –computing size = 8x8 (middle) and 32x32pixels (bottom) indicating the detected changes are superimpose to one of the input image . Note that only the structured object changes are retrieved.