Computing Invariants for Structural Change Detection in Urban Areas

FengFeng Tang and Véronique Prinet
National Laboratory of Pattern Recognition (NLPR)
Sino-French Lab. in Computer Science, Automation and Applied Mathematics (LIAMA)
Institute of Automation, Chinese Academy of Sciences, Beijing 100080, P.R. China
{fftang,prinet}@nlpr.ia.ac.cn

***Abstract*— This paper introduces a new approach to compute similarity/dissimilarity between images or regions from images. It borrows concepts from object recognition and content based image retrieval. The main idea is to build a bag-of-features where the feature clusters are learnt from a single pair of images. Normalised Cut is used to segment the image into regions which exhibit homogeneous and regular *structural pattern* distribution. Histograms of feature types are compared using the χ^2 distance. The application is on very high resolution optical satellite images, for urban development analysis.**

I. Introduction

Our objective is to propose and to demonstrate the limit and potential of a completely new framework for structural change analysis from two co-registered remote sensing images, with application on urban areas and VHR optical data.

A survey of change detection techniques can be found in [10]. Thought it is a quite thorough review, it does not give much insight of most recent and innovative approaches. Kumar and Hebert introduced Discriminative Random Fields (DRF) to detect structured objects (i.e. polygonal objects) from natural images. Liu Wei extended and adapted DRF model to the case of structural change analysis [6]. Boyer and Sarkar developed in [11] and [14] an approach based on graph spectra and eigencluster in order to segment the image into developed and less developed areas, where the word "developed" refers to the lines organisation in the image. Li and Hu [5], in a similar fashion, proposed to first partition the image, then to analyse the changes at the level of the sub-graphs.

The approach we introduce in this paper is originally motivated by the need to find a correct definition of "what is a change", from remote sensing VHR images. Figure 2 shows examples of different types of change in a scene, from structural (top-left), to object (top-right), or texture (bottom-left). The bottom-right image pair illustrates the difficulty to differentiate between "apparent changes" in the image and "real changes" of the real scene. "Apparent changes" can be generated by shadows, acquisition view angle, occlusion, seasonal effect, or by the introduction of temporary objects in the scene, such as people, car, bird, ... Bottom-right images at figure 2 accumulates the effects of shadow, projection, and apparition of birds. Indeed, we are interested to analyse only the "real changes", or long term changes which happen in the real word. We address the issue of *structural change*. A structural change is characterised by the apparition or disparition of a compact block of similar geometric objects in the scene. One can observe a typical structural change on figure 1, localised at the centre of the images. The presence of buildings forms a *structural pattern* or *structural texture* in the image.

Fig. 1. Original panchromatic Quickbird (0.61m/pixel) pair of images over Beijing city. Top: Year 2001. Bottom: Year 2003.

To tackle the problem of structural change, we are guided by two main principles: both global pattern and local features are important; it is easier to formalise the notion of *similarity* rather than the notion of change itself. The approach we

1-4241-4244-0712-5/07/$20.00 © 2007 IEEE

Fig. 2. Different types of change in a scene : structural -top left, texture —bottom left, object —top right right, and apparent —bottom right. The two former characterise global change —at the scale of the scene or regions of the scene— while the two latter are local —at the scale of the object.

propose takes root in the works that emerged since ten years about in the field of object-categories recognition ([7], [8]). We adapted this concept to the problem of *change recognition*.

The rest of the paper is organised as follows: section II introduces the principle of structural change computation; we first give a general overview, then we explain in details the segmentation from Normalised Cut, the generation of a bag of invariant features and the quantitative evaluation of change. In section III, an illustration of the approach is shown from concrete examples on a VHR optical image pair taken on Beijing City. The paper ends with a conclusion in section IV.

II. Structural change from bag-of-features

A. Approach overview

The general principle is to assume that a given scene in the real world can be represented by a limited number of specific local appearance features (lets call them *feature types* at this stage) computed from the image, independently of the image acquisition conditions. We proceed at first to a rough segmentation of the image into structurally consistent regions (section II-B). At the scale of the images, we constitute a dictionary of words (section II-C). Each region is then described with a vocabulary taken from this dictionary. Then, the change analysis between two associated regions extracted from a pair of images is performed by computing the similarity within their vocabularies (section II-D).

B. Graph-based image segmentation

Our approach for image segmentation into regions of different structural patterns is mainly inspired by the works from Boyer and Sarkar [11], [13].

Man made objects –buildings, roads, railways,...– in an abandoned rural area, will be sparse and have random directions; a contrario, man-made objects in a highly developed region, for example an inhabitant suburban area, are dense and regularly spaced in the scene, most often parallel or orthogonal to each others. The level of spatial organisation of the objects in a real scene is characterised in the image by the degree of *spatial organisation of edge lines*. Highly regular spatial pattern of edge lines, or, at the opposite, fully random distribution of edge lines, characterises the *"structural pattern"* of a region in the image.

Hence, in order to segment the image into regions of similar structural pattern, we build up a graph $G = (V, E)$, where the nodes/vertexes V are the line segments, extracted from the image using support region method as described in [14]; the weight w_{ij} of each edge $e_{ij} \in E$ linking pairwise nodes (i, j) (i.e. straight lines) is defined as a function of the distance between these two connected nodes and of their relative orientation:

$$\omega_{ij} = |\cos(2\alpha_{ij})| \exp(-\frac{d_{ij}^2}{\beta^2}) \quad (1)$$

where α_{ij} is the angle between the normals of line i and of line j; d_{ij} is the distance between line i and line j; β is a scale constant that controls the maximum distance in which the lines' pair will affect each other; and $|.|$ denotes the absolute value. The relation matrix defined by $W(i, j) = \omega_{ij} = \omega_{ji}$ is symmetric and positive definite.

Normalized Cuts (Ncut) [12] is used to partition graph G into sub-graphs G_1, G_2. The cost function to minimise in order to generate sub-graphs with maximum intracohesivness (with respect to the definition of the weight function), is defined by:

$$Ncut(G_1, G_2) = \frac{cut(G_1, G_2)}{assoc(G_1, G)} + \frac{cut(G_1, G_2)}{assoc(G_2, G)} \quad (2)$$

where $assoc(G_1, G) = \sum_{i \in G_1, j \in G} \omega_{ij}$, $cut(G_1, G_2) = \sum_{i \in G_1, j \in G_2} \omega_{ij}$.
Equation 2 has an analytical global minimum solution which can be computed by solving the following eigenequation [12]:

$$D^{-\frac{1}{2}}(D - W)D^{-\frac{1}{2}}\mathbf{x} = \lambda\mathbf{x}$$

where D is a diagonal matrix with each item on the diagonal is the sum of all the items in the corresponding row in W. The eigenvector $\mathbf{x}$ with the second smallest eigenvalue λ determines the optimal bi-partitioning of the graph.

This bi-partition process is performed iteratively. Intuitively, we wish the partitioning process to stop when the intra-cohesiveness within each resulting region is high enough. Simultaneously we will wish to favour the partitioning at the beginning of the iterations, and to penalise it while iterations go on. Based on these remarks, we propose the following Iteration Stopping Criteria (ISC) to be applied within each region i:

$$\frac{assoc(G_i^l, G^{l-1})}{conn(G_i^l)} > T_{ISC} \exp(-\frac{l}{\delta}) \quad (3)$$

where $i = \{1, 2\}$, G_i^l is the sub-graph i at iteration level $l > 0$, $conn(G_i^l)$ is the number of connections in sub-graph G_i^l —i.e. $card(V_i^l)$—, and T_{ISC}, δ are constants which control the maximum threshold and the decreasing speed respectively. G^0 is the initial graph formed by all segments in the image. The number of partitions at the end of the iterations is thus determined by the image itself –and the parameters values indeed. Partitions are identified by their index k, $1 \leq k \leq N$.

The graph partitioning however does not segment the image itself —since $G = (V, E)$ is built on the edge lines and not the image pixels. Image segmentation is performed by back projecting the vertexes V onto the image. Pixels are assigned the index of the partition they are the closest to. We end up with an image segmented into geometrically structured regions.

We propose to evaluate the quality of the resulting segmentation by comparison with a manually created ground truth —based on our own (subjective) understanding of what is a structurally consistent region. We define a similarity factor Sim between $Ncut$-segmentation S_{Ncut}, and ground truth segmentation S_g by:

$$Sim(S_g, S_{Ncut}) = \exp(-|N_g - N_{Ncut}|) \prod_k \frac{A_{maxinter}(P^k_{Ncut}, P^h_g)}{A(P^k_{Ncut}) + A(P^h_g)} \quad (4)$$

where $A_{maxinter}(P^k_{Ncut}, P^h_g)$ is the maximum intersection area between the $Ncut$ region k, P^k_{Ncut}, and its corresponding overlapping ground truth region indexed h, P^h_g; $A(P^k_{Ncut})$ is the area of the $Ncut$ region k; N_{Ncut}, N_g are respectively the total number of regions resulting from the $Ncut$ segmentation and from ground truth segmentation. The manually created segmentation is shown in figure 3.

C. Invariant features

We argue that similar scenes will have similar feature-types (so-called *bag-of-features*, or *words dictionary*, see for example [8], [15]) and that *defacto* it is possible to "recognise" the similarity between two scenes based on the comparison of their feature-types computed from images.

We compute local features with directional gradient histograms over small patches (typically 16x16 pixels window size). Patches are centred on key points extracted with a corner detector (e.g. Harris detector), or randomly distributed over the image [8], [9]. These SIFT-like operators [8], [15] have proved to be very robust features to characterise the geometry of the object locally; they are the most popular descriptors for object recognition from static images or video [1], [4]. The classical approach is to compute SIFT features from a large database of positive and negative image samples, and to cluster them into a certain number of classes or feature-types, characterised by their centre (i.e. the average of all features belonging to a cluster) and possibly their variance (in the feature space). This is for the learning phase. The recognition phase is achieved by classifying the test image as positive (the object exists) or negative (there is no such object in the image) [15].

Features should be robust to illumination changes and to differences in view angle during the image acquisition process. To make directional gradient features *invariant to affine transformation* of intensity level, it can be shown that it is sufficient to normalise the gradient directions by the local average intensity level. *Invariance to rotation* has to be treated more carefully. We detect the predominant direction in the local patch and compute its associated orientation angle α.

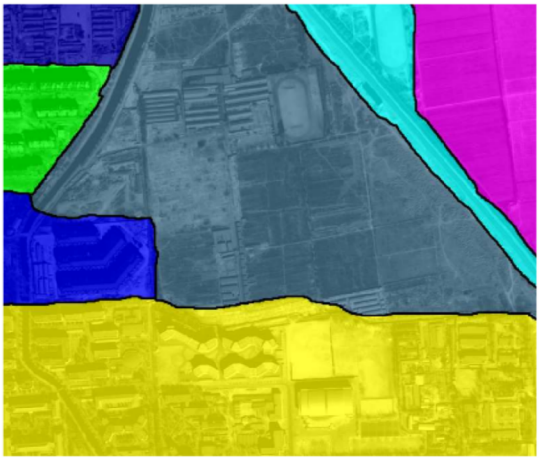


Fig. 3. Ground truth segmentation generated from the top image of fig. 1 .

All gradient directions computed in the patch are then turned with an angle α. It results that, the "normalised" predominant direction is identical for all the patches.

The problem of *change recognition* needs to be tackled slightly differently than for "pure" object recognition – although based largely on the main concepts– since we do not have access to a large data-base for training. The purpose is not to describe a scene uniquely but to compare the similarity between a pair of images in term of feature types (namely "words"). We do not have positive and negative samples. Corner detector and SIFT descriptor are applied on each of the two images. Clustering is performed and generate a dictionary of words. To each patch is assigned the closest word in the feature space. Thus, each of the two images are described with a set of words taken from the same dictionary. Because we deal only with local appearance patches, two different scenes which have similar local geometrical properties —e.g. a scene of residential buildings in dense urban area— will be described with a similar vocabulary.

Structural change analysis will then be achieved by comparing the set of words that describe each image.

D. Change analysis

The most simple and straightforward way to express — and to quantify— changes between two images or regions in images, is in term of a distance of local features. It is the approach we choose in this paper.

In practice, we carry out the segmentation step (section II-B) in one image only. The resulting segmentation, i.e. delineation of contours that separate two or more regions, is projected onto the second image. For each region of each of the two images, a histogram of words, $hist$, is built (section II-C).

We use χ^2-distance to quantify the similarity/dissimilarity between regions. The degree of change DG of a given region k between image 1 and image 2 is then given by:

$$DG_k = \chi^2(hist^1(k), hist^2(k)))$$

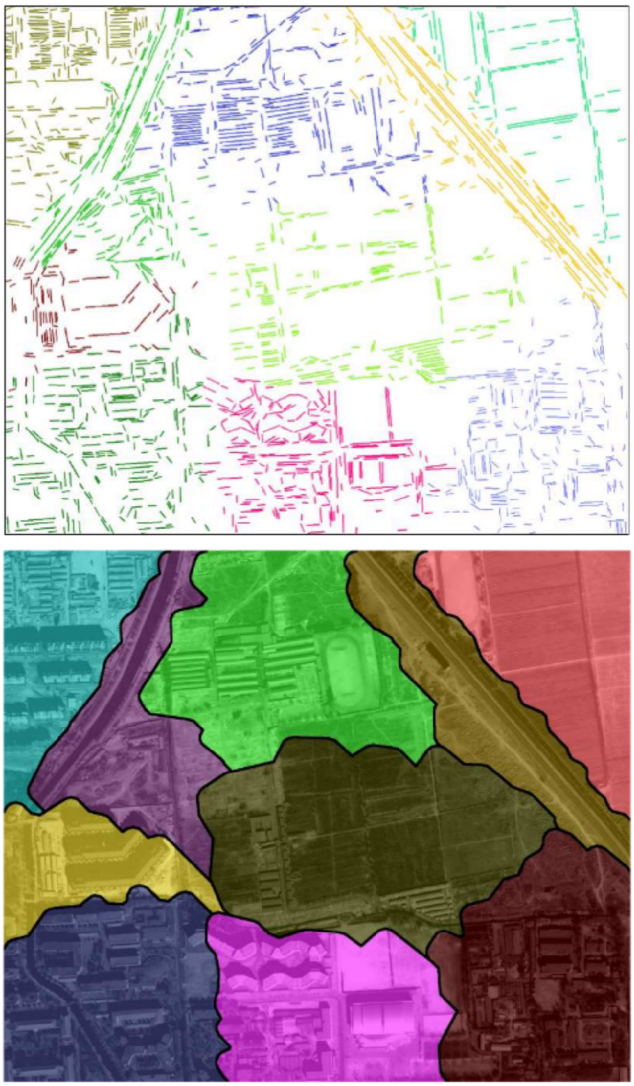


Fig. 4. Graph final partitions (top) and image segmentation (bottom).

where $hist^1(k)$ and $hist^2(k)$ are the words' histograms computed in region k from image 1 and 2 respectively. High χ^2-distance value corresponds to very different scenes and vice-versa.

For comparison with a manually estimated "ground truth" of the changes, we threshold the DG_ks and quantify them to $\{0, 0.5, 1\}$, associated respectively to the classes "no-change", "partial-change", "total-change". The overall quality index of the changes estimation between two images is then defined by:

$$Q = \frac{1}{N} \sum_k \left(\frac{1}{2}\, round(2\ DG_k) - gt(k) \right)^2$$

where the function $round(a)$, $a \in \Re^+$, is the closest integer approximation; $gt(k)$ is the ground truth for region k (ground truth takes values belonging to $\{0; 0.5; 1\}$; and N is the total number of regions.

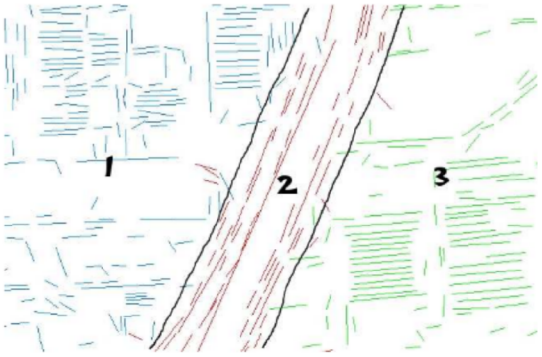


Fig. 5. Zooming of the top-left part of the graph partition at figure 4. Colours represent partition index. The straight dark lines delineate the partitions indexed $1, 2$ and 3.

III. Results

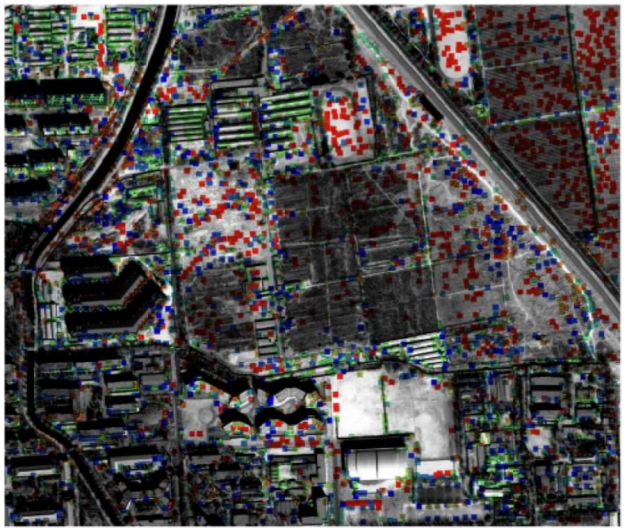
Input data used for experimentation and validation are a pair of Quickbird panchromatic images (resolution 0.61metre/pixel), acquired in 2001 and 2003, and covering the area of Beijing. An example of input images is shown in figure 1. The image size is 1625 x 1318 pixels (about 1000x800m^2 on the ground), and the grey level spreads over 8 bits.

The values of the parameters fixed manually are: $\beta = 30$ (equ. 1), $T_{ISC} = 0.06, \delta = 10$ (equ. 3). K-means algorithm stops when less than 1% of points are reassigned to a new cluster, with a total number of clusters to 17.

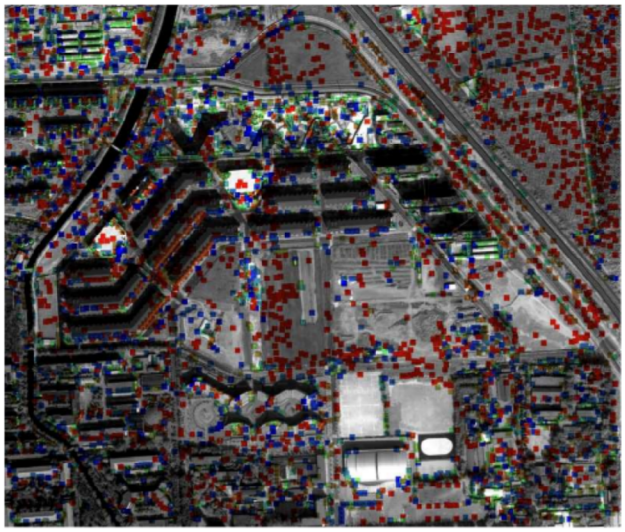
We illustrate the results from figures 4 to 7. Graph partitioning and image segmentation results are given at figure 4, with a zoom of the graph shown at figure 5. The partitioning clearly separates regions composed of edge lines having different orientations and organisation levels. The drawback of graph-partitioning based on edge lines is that an image region without edge lines (e.g. grass land) cannot be identified as an individual part. The resulting segmentation at figure 4 is however rather good and corresponds well to our perception of structural segmentation.

The spatial distribution of feature-types (words) on each of the two images is shown on figure 6. One can notice (from the colour image) how well different words (identified by different colours) are associated to different local geometrical features. For example red-word is localised mainly in the low textured land areas, while green-word covers mainly the buildings.

The final classification of each region into "structural change", "no-structural change", or "partial-structural change", is given in figure 6(b) and compared to the manual classification on 6(a). Here two regions are mis-classified. If we take a look at the central (oval) region, we can notice that there exist clear modifications in this area, but that there are not of structural type, and therefore have not been detected as structural change. This show the power of computing local geometrical features, rather than global texture characteristics.

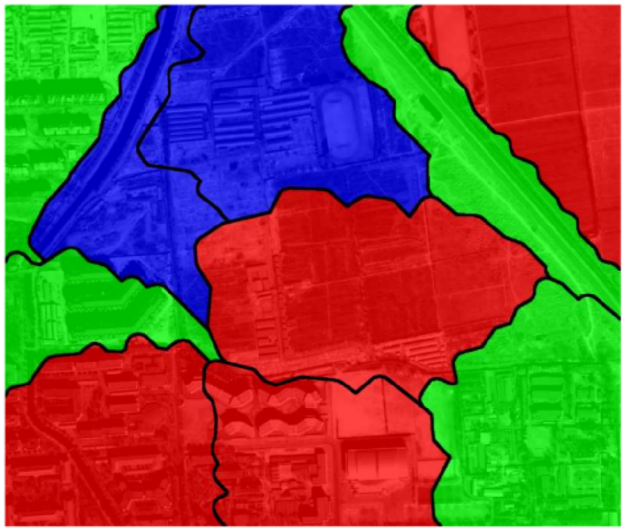


(a)

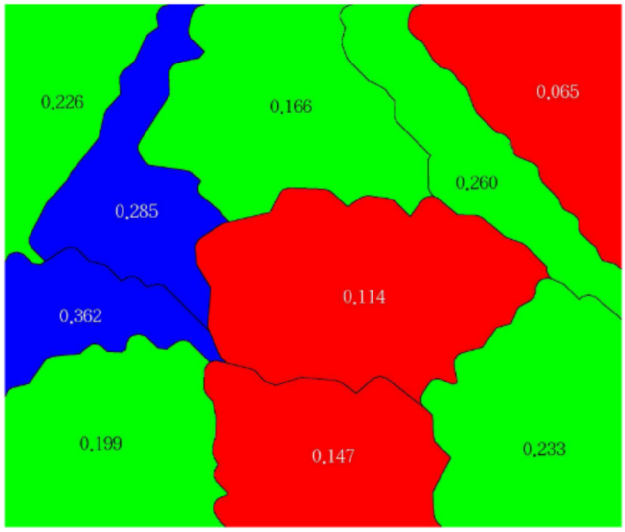


(b)

Fig. 6. Extracted patches and feature descriptors: colours represent the word index to which the local patch is assigned. (a) Year 2001, (b) Year 2003.



(a) Change labelled by hand



(b) Change quantification

Fig. 7. Structural change is computed at each region using χ^2 distance (indicated by the numerical value), then quantified. Colour code is: red: no change, green : partial change, blue : change.

IV. Conclusion

In this paper, we proposed a novel framework for structural change detection and analysis from roughly registered very high resolution remote sensing image pair. The main idea is that a scene can be described in term of a limited number of local appearance feature types (words). The distance between words histograms is an indicator of the similarity/dissimilarity between two scenes. A rough image segmentation achieved at first, makes it possible to localise the areas of the changes/no change.

Future work will focus on the reformulation of the weights of the relation matrix by adding concepts borrowed from Gestalt theory, the integration of a multi-scale approach, and the development of a probabilistic model of change.

References

[1] A. Agarwal and B. Triggs, *Hyperfeatures-Multilevel Local coding for visual recognition*, in Proc. European Conference on Computer Vision, 2006.

[2] R. Fergus, P. Perona and A. Zisserman, *Object class recognition by unsupervised scale invariant learning*, in Proc. Intern. Conf. on Computer Vision and Pattern Recognition, 2003.

[3] S. Kumar and M. Hebert, *Discriminative Random Fields: A discriminative framework for contextual interaction in classification*, in Proc. Intern. Conf. on Computer Vision, 2003.

[4] I. Laptev, *Improvement of object recognition using boosted histograms*, in Proc. British Machine Vision Conference, 2006.

[5] W.M. Li and Z.Y. Hu, *A novel framework for urban change detection using VHR satellite images*, in Proc. Intern. Conf. on Pattern Recognition, 2006.

[6] W. Liu, *Building recognition and change inference using probabilistic modeling from VHR images*, Master Thesis, June 2005 (in Chinese) (downloadable from http://kepler.ia.ac.cn/).

[7] D. Lowe, *Distinctive image features from scale-invariant keypoints*, International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.

[8] K. Mikolajczyk and C. Schmid, *A performance evaluation of local descriptors*, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 10, Oct. 2005.

[9] E. Nowak, F. Jurie, B. Triggs, *Sampling strategies for Bag-of-Features image classification*, in Proc. European Conf. on Computer Vision, 2006.

[10] R.J. Radke, O. Andra, O. Al-Kofahi and B. Roysam, *Image change detection algorithm: A systematic survey*, IEEE Trans. on Image Processing, vol. 14, no. 3, March 2005

[11] S. Sarkar and K.L. Boyer, *Quantitative measures of change based on feature organisation: eigenvalues and eigenvectors*, Computer Vision and image Understanding, vol. 71, no. 1, pp 110-136, 1998.

[12] J.B. Shi and J. Malik, *Normalized Cuts and image segmentation*,IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 8, 2000.

[13] C. Unsalan and K.L. Boyer, *Classifying land development in high-resolution panchromatic satellite images using straight-line statistics*, IEEE Trans. Geoscience and Remote Sensing, vol. 42, pp. 907-919, April 2004

[14] C. Unsalan and K.L. Boyer, *A theoretical and experimental investigation of graph theoretical measures for land development in satellite imagery*, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 4, April 2005.

[15] J.G. Zhang, M. Marszalek, S. Lazebnik, C. Schmid, *Local features and kernels for classification of texture and object categories: an in-depth study*, Inria Research Report RR-5737, Oct. 2005.