REMOTE SENSING MINEFIELD AREA REDUCTION: MODEL-BASED APPROACHES FOR THE EXTRACTION OF MINEFIELD INDICATORS

A. Katartzis[1], I. Vanhamel[1], J. C-W Chan[2], and H. Sahli[1,2]

[1]Vrije Universiteit Brussel, VUB-ETRO, Peinlaan, 2 - 1050 Brussels, Belgium, Belgium
[2]Interuniversitair Micro-Elektronica Centrum, IMEC-ETRO, Peinlaan, 2 - 1050 Brussels, Belgium, Belgium

ABSTRACT

The use of high resolution commercial satellite and airborne images for the survey of landmine suspected areas has been suggested recently in the context of Mine Action to (i) map the hazardous area (suspected minefield), and (ii) possibly reduce its extend. Minefields may be identified using methods that directly detect and confirm the location of landmines. Next to this approach, indirect indicators, closely related to the occurrence of the minefields themselves can be used. Such indicators correspond either to direct military activities, e.g. trenches, embankments, protection walls, bunkers, foxholes, fences, etc, or changes in the landscape, e.g. abandoned arable land, unused roads, foot paths and tracks through fields, etc. The present work investigates model-based approaches for the (semi-) automatic extraction of some of the indirect minefield indicators from high resolution airborne images.

Key words: Remote sensing, minefield indicators, image segmentation, labelling, Markov random fields, perceptual organization.

1. INTRODUCTION

In the last decade, Humanitarian Mine Action has acknowledged the role of remote sensing as a useful tool, able to enhance the productivity and safety of ground-based minefield area reduction methods (Maathuis 2001), (Batman & Goutsias 2003), (Pizurica et al. 1999), (van Kempen et al. 1999). Air- and space-borne platforms, equipped with multiple sensors, can quickly and safely scan large and inaccessible areas. Remote sensing can contribute to the mapping and identification of suspected areas, and eventually the determination of the minefield boundaries, by extracting from the acquired images suitable direct and indirect minefield indicators. The identification of image indicators, in combination with collected ancillary information, prior knowledge/intelligence, can assist in conventional General Mine Action Assessment and Technical Surveys prior to a clearance (demining) campaign and implicitly highlight areas of high risk. Minefield indicators are conventionally identified via visual image interpretation, by experts with in-depth field knowledge and experience. This process is carried out either through spontaneous recognition, through recollection of object features (such as color, texture, size, shape, shadow), or using a process of logical inference. Despite its effectiveness, visual interpretation can become intractable in the presence of huge volumes of data. In that case, automatic image analysis techniques could be an useful alternative. Systems for automatic image analysis will only be accepted by experts if two conditions are met: They have to detect visually perceivable image regions that carry semantic interpretation for an expert observer, and they must enable him to bear, rather than to oblige him to be familiar with digital image processing.

In this work we focus on the (semi-) automatic detection of indirect minefield indicators, namely, (i) perceptually uniform image regions, from which contextual land cover classification can be performed for the assessment of the land use, and (ii) meaningful topographic objects such as (a) linear structures corresponding to trenches, protection walls, roads, or foot paths and tracks through fields, and (b) buildings corresponding to houses or bunkers. The use of these indicators for minefield area reduction is described in details in (Sahli et al. 2004).

The paper is organized as follows. Section 2 presents the methodological framework used for image analysis. Section 3 describes the proposed hierarchical methods for image segmentation and labelling. Section 4 is dedicated to the model-based methods proposed for the detection of linear structures and buildings from remote sensing data. Finally, conclusions are given in section 5.

2. METHODOLOGY

Image segmentation/classification and topographic object extraction are critical for subsequent image analysis and further image understanding of remote sensing data, as they have to conform with the following facts: (1) remotely sensed imagery has multispectral and multiscale nature; (2) in contrast to other image modalities, remote sensing images contain various objects with heterogenous properties with respect to size, form, spectral behavior etc, so meaningful objects should be extracted at the appropriate scale; (3) model-based interpretation of remote sensing imagery is more difficult due to the heterogeneity of inherent object classes; and (4) sub-optimal solutions will probably not be considered for remote sensing applications because there is no need for real-time applications.

Taking into account these requirements, in this study we propose the use of (i) multiscale/multispectral analysis for image segmentation and land cover classification using both deterministic and stochastic methods, and (ii) Markov random field (MRF) model-based approaches for topographic object detection. In the following two sections we present an overview of the two major components of our methodology.

2.1. Multiscale Image Analysis

Due to the hierarchical structure of images content, each object/structure can only be processed efficiently in its individual range of appropriate scales. Therefore multiscale approaches have been recommended (Koenderink 1984) as appropriate tools to extract as much visual information of an image as possible. While coarse scales reveal large objects, finer scales contain finer structural information. In a remote sensing image analysis system an image has to be decomposed into all its visually perceivable structures. Generally speaking, a visual decomposition consists of connected components (regions) on the image's pixels grid. To achieve a multiscale representation in a size-sensitive way, the generating regions have to be reduced successively and be scale-dependent. As a result, a multiscale analysis is a stack (tower) of partitions of an image, yielding regions that hold the connectivity criterion. When computing these partitions, it is important that each region of a smaller scale is completely contained in a region of a larger scale, otherwise there would be a translation of regions between scales and a violation of the paradigm of scale-space causality.

Several diffusion schemes have been proposed for generating a scale space representation of an image. Nonlinear approaches offer powerful operations to preserve region shapes and thus ensuring their stability along scales. In this work we apply a directional diffusion (Sapiro 2001), which is based on Euclidean shortening flow, augmented by an edge stopping function:

$$\begin{cases} \frac{\partial u}{\partial t} & = \operatorname{div}\left[g\left(\|\nabla u_{\sigma_r}\|\right) \frac{\nabla u}{\|\nabla u\|}\right] \\ u\,(t=0) & = f \end{cases} \tag{1}$$

where f is the original image, u the scale-space image, t the continuous scale parameter, g the edge stopping function, and σ_r is a regularization parameter to ensure the well-possedness of the scale-space. Furthermore, we use the general framework for vector-valued diffusion (Whitaker & Gerig 1994), where separate diffusion processes, corresponding to each image channel, evolve simultaneously. They are coupled with each other through the edge stopping function, which is estimated on the generalized gradient (DiZenzo 1986) (involving all the channels simultaneously).

The application of scale-space filtering entails the creation of a multiscale tower U, that consists of N discrete scale-space family members u_n. The latter is achieved by applying the natural scale-space sampling method (Koenderink 1984) to the anisotropic diffusion filter (Vanhamel et al. 2003). The finer scale (localization scale) $u_{n=0}$ is defined empirically as the scale that sustains all information in the image with a minimal amount of noise.

2.2. Markov Random Field Models

Contextual constraints are ultimately necessary in the interpretation of visual information. A scene is understood through the spatial and visual context of the objects in it; the objects are recognized in the context of object features at a lower level representation; the object features are identified based on the context primitives at an even lower level; and the primitives are extracted in the context of image pixels at the lowest level of abstraction. The use of contextual constraints is indispensable for a complex vision system.

Markov random field theory provides a convenient and consistent way of modeling context dependent entities (Li 1995). This is achieved through characterizing mutual influences among such entities using MRF probabilities. The theory tells us how to model the *a priori* probability of contextual dependent patterns. A particular MRF model favors its own class of patterns by associating them with larger probabilities than other pattern classes. In the following, we will briefly review the concept of MRF defined on graphs.

Let $\mathbf{G}$=$\{\mathbf{S}$,$\mathbf{A}\}$ be a graph, where $\mathbf{S}$=$\{S_1, S_2, ..., S_m\}$ is the set of nodes and $\mathbf{A}$ is the set of arcs containing them. We define a *neighbourhood system* on $\mathbf{G}$, denoted by:

$$\mathbf{N} = \{\ \mathbf{N}(S_1),\ \mathbf{N}(S_2), ..., \mathbf{N}(S_m)\}$$

where $\mathbf{N}(S_i), i = 1, 2, ..., m$ is the set of all nodes in $\mathbf{S}$ that are neighbors of S_i, such that:

i) $S_i \in \mathbf{N}(S_i)$, and
ii) if $S_j \in \mathbf{N}(S_i)$ then $S_i \in \mathbf{N}(S_j)$
Let $\mathbf{L}$=$\{L_1, L_2, ..., L_m\}$ be a family of random variables defined on $\mathbf{S}$, in which each random variable L_i takes a value l_i in a given set (the random variables L_i's can be numerical as well as symbolic, e.g. interpretation labels). The family $\mathbf{L}$ is called a random field. $\mathbf{L}$ is a MRF on $\mathbf{G}$, with respect to the neighbourhood system $\mathbf{N}$ if and only if

1. $P(\mathbf{L} = l) > 0$, for all realizations l of $\mathbf{L}$;
2. $P(l_i|l_j, \forall S_j \neq S_i) = P(l_i|l_j, S_j \in\mathbf{N}(S_i))$

where $P(\mathbf{L} = l) = P(L_1 = l_1, L_2 = l_2, ..., L_m = l_m)$ (abbreviated by $P(l)$) and $P(l_i|l_j)$ are the joint and conditional probability functions, respectively. Intuitively, the MRF is a random field with the property that the statistics at a particular node depend mainly on that of its neighbors.

An important feature of the MRF model defined above is that its joint p.d.f has a general functional form, known as *Gibbs distribution*, which is defined based on the concept of *cliques*. A clique c, associated with the graph $\mathbf{G}$, is a subset of $\mathbf{S}$ such that it contains either a single node, or several nodes that are all neighbors of each other. If we denote the collection of all the cliques of $\mathbf{G}$, with respect to the neighbourhood system $\mathbf{N}$, as $\mathbf{C(G,N)}$, the general form of a realization of $P(l)$ can be expressed as the following Gibbs distribution:

$$P(l) = \frac{1}{Z} e^{-U(l)} \tag{2}$$

where $U(l) = \sum_{c \in \mathbf{C}} V_c(l)$ is called the *Gibbs energy function* and $V_c(l)$ the *clique potential functions* defined on the corresponding cliques $c \in \mathbf{C}(\mathbf{G}, \mathbf{N})$. Finally, $Z = \sum_{l \in \mathbf{L}} e^{-U(l)}$ is a normalizing constant called the *partition function*.

In case of a *labeling* problem, when we have both prior information together with knowledge about the distribution of our data, the most optimal labeling of the graph $\mathbf{G}$ can be obtained based on the maximum *a posteriori* probability (MAP)-MRF framework. According to the Bayes rule, the posterior probability of our system can be computed by using the following formulation:

$$P(l|d) = \frac{p(d|l)P(l)}{p(d)} \tag{3}$$

where $P(l)$ is the prior probability of labelings l, $p(d|l)$ is the conditional probability distribution function (p.d.f.) of the observations d, also called the *likelihood function* of l for d fixed, and $p(d)$ is the density of d which is a constant when d is given. By associating an energy function to $p(d|l)$ and $P(l)$ (denoted by $U(d|l)$ and $U(l)$ respectively), we can express the posterior probability as:

$$P(l|d) \propto e^{-U(l|d)} \tag{4}$$

where

$$U(l|d) = U(d|l) + U(l) \tag{5}$$

The most optimal labeling, given the observation field d, can be found by minimizing the energy function $U(l|d)$.

Due to its unique property of combining both global and local information, the MRF model-based approach, applied to image interpretation, provides potential advantages in knowledge representation, learning and optimization. In this work MRF models are used for both Multiscale Hierarchical Labelling and detection of topographic objects.

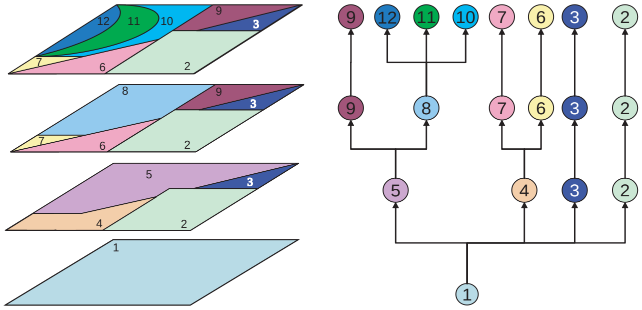
3. HIERARCHICAL IMAGE ANALYSIS

3.1. Hierarchical Region Adjacency Graph

Using the scale-space tower of Section 2.1, a hierarchical region adjacency graph can be extracted by tracking the regional minima of the generalized gradient image, through the scale-space. Initially, the watershed transformation is performed at the localization scale in order to identify the position of all the contours in the image. Then, the duality between the regional minima of the gradient and the catchment basins of the watershed (Jackway 1996) is exploited to form a robust region-based parent-child linking. The linking is performed in a bottom-up fashion, from the finer scale (localization scale) to the coarsest one. However, for the clarity of description, the top down approach is given. Starting with the complete image as the coarsest region of a scale space, each region is considered as an attributed graph node. In the succeeding finer scales all the partitions of the initial region are detected and so forth for each scale until the finest layer. A new node is added as a successor node to an existing region only if it breaks up to at least two subregions. Otherwise this region is taken as a successor node. Within this graph structure, the nodes represent unique regions and the edges their topological descriptive scale inclusion. Figure 1 illustrates the construction of the hierarchical region adjacency graph. This hierarchical structure contains a set of regions with visually perceived contours, linked by their topological adjacency and full inter-scale relations. Figure 2 depicts the segmentation of the multiscale tower of a mosaic airborne image.

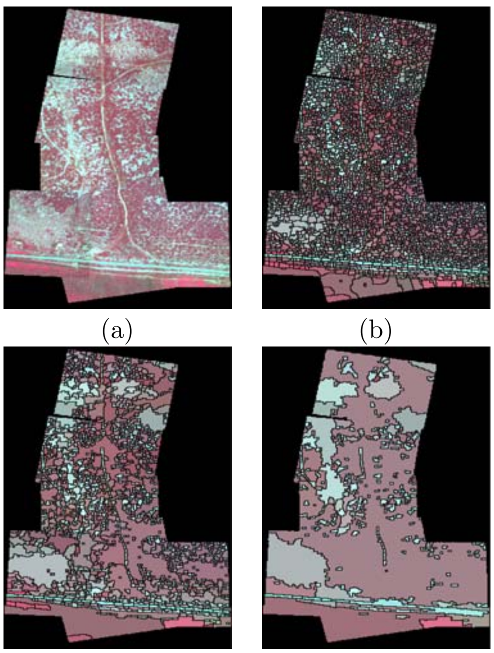
3.2. Multiscale Hierarchical Segmentation

We propose a region-based segmentation scheme that involves scale-space theory, causality of object contours obtained with morphological watersheds, and iterative merging of neighboring regions, producing a hierarchy of region adjacency graphs (Vanhamel et al. 2003). Besides the topological relations of regions, their interpretation in an 'object-oriented'



(a) (b)

Figure 1. Hierarchical Region Adjacency Graph. (a) scale space tower, and (b) linkage list.



(a) (b) (c) (d)

Figure 2. Segmented scale space tower. (a) original image, (b) finest scale 0, (c) scale 9, and (d) scale 17.

sense is given by descriptive attributes, grouped into (1) scale-dependent information, like hierarchical level, merging homogeneity, etc, and (2) region properties such as shape, size, color or texture.

Figure 3 gives a schematic diagram of the proposed hierarchical segmentation scheme for multispectral images. It consists of two basic modules. The first module (*Salient Measure Module*) is dedicated to attribute a saliency measure to each contour arc at the finest scale taking into account the whole scalespace tower. The entire process to retrieve the saliency measure for the generalized gradient watersheds requires three steps: (i) scalespace stack generation by vectorvalued nonlinear diffusion filtering (see Section 2.1); (ii) Region-based parent-child linking (see Section 3.1); (iii) Contour valuation by downward projection. The dynamics of contours in scalespace are used to valuate the contours detected at the finest scale. The second module (*Hierarchical Levels Retrieval Module*) identifies the different hierarchical levels through a hypothesis testing criterion. Starting from the watershed segmentation at the finest scale, a successive merging operation is performed until a stopping criterion is satisfied. The merging of adjacent regions is based upon a color similarity measure, while the merging sequence is designated by the contour valuation measure (dynamics of contours in scalespace). Figure 4 shows the obtained segmentation results using the scale scale space tower of Figure 2.

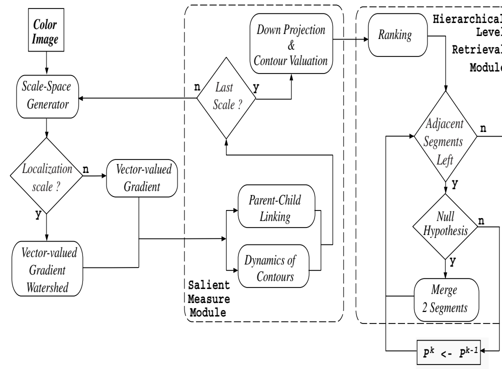
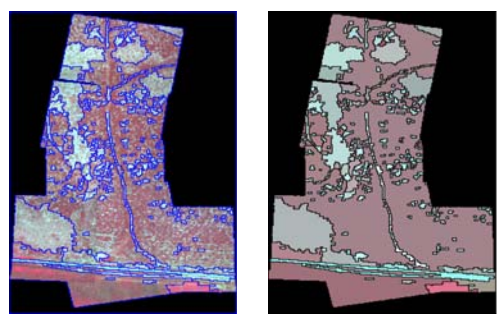


Figure 3. Multiscale Hierarchical Segmentation flow chart.



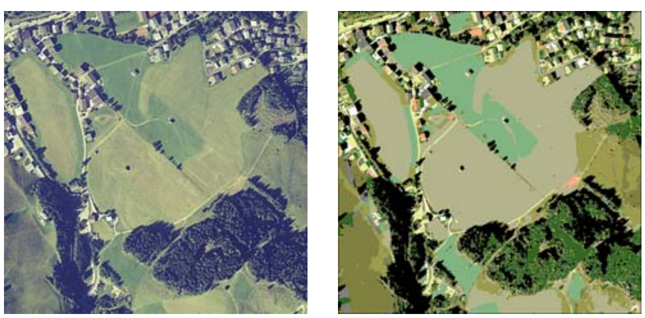
(a) (b)

Figure 4. Multiscale hierarchical segmentation results. (a) overlay of watershed lines, (b) labelled image.

3.3. Multiscale Hierarchical Labelling

One of the requirements of an air- or space-borne minefield area reduction system is the characterization of the land cover of the surveyed area. Its major aim is the discrimination between areas of human activity (e.g. residential areas, agricultural fields) and natural undisturbed environments (areas with high and dense vegetation). In the case of high resolution airborne images, the defined land cover classes have a non homogeneous appearance with respect to spectral and textural responses, something that hampers the classification efficiency of conventional pixel-based classifiers (like Maximum Likelihood, fuzzy c-means clustering, neural networks). Taking into account this constraint, a variety of classification approaches based on the framework of Markov Random Field theory have been proposed in the literature, in order to capture contextual information, in terms of spatial consistency. These methods involve either non-causal energy-based models defined on the image lattice (Geman & Geman 1984) or hierarchical, multiresolution models defined on pyramidal representation of the image (Bouman & Shapiro 1994).

In our analysis for land cover classification, we have used an hierarchical Markovian model defined on the hierarchy of multiscale region adjacency graph, described in Section 3.1. The image classification is treated as a hierarchical labelling problem, applied to the multiscale region adjacency graph, using a finite set of interpretation labels (e.g. land cover classes) (Katartzis et al. 2003). We associate to the nodes of our hierarchy a causal label field in the form of a coarse-to-fine Markov chain and an observation field, which represents the signature of each region in the original image. The data likelihoods of the field of observations are expressed in terms of a non-parametric dissimilarity measure, and have the advantage to be model-free, in the sense that the underlying probability distributions are not assumed to belong to a parametric model class. The nature of the proposed multiscale region adjacency graph allows us to incorporate efficient, non-iterative schemes for Bayesian inference, which have been inspired from the state-of-the-art developments of hierarchical Markovian models defined on quadtrees. An example of the multiscale hierarchical approach for land cover classification is presented in Figure 5.



(a) (b)

Figure 5. Land cover classification. (a) original image, (b) classification map.

4. DETECTION OF TOPOGRAPHIC STRUCTURES

4.1. Linear structures

Minefield indicators of this type include elongated structures related to warfare activities (e.g. trenches), roads and paths that designate safe passage areas, protection walls, rivers and geological erosions. We have developed a model-based approach for the identification of linear structures, which follows certain assumptions concerning their geometric and radiometric properties (Katartzis et al. 2001b). The flow chart of the proposed approach is given in Figure 6.

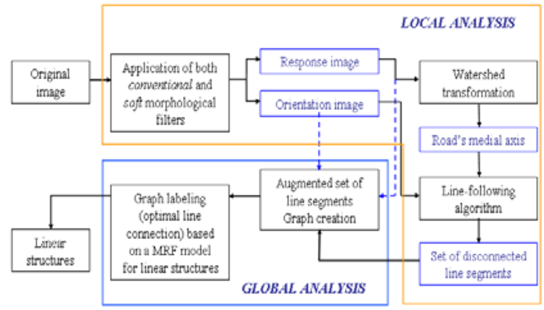
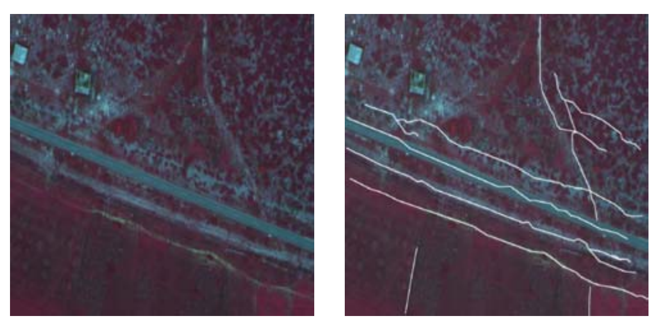


Figure 6. Linear structures detection flow chart.

During a local analysis step, the detection of elongated structures is performed by applying a series of morphological filters. The main axis of the extracted elongated structures is determined by applying the watershed transformation on the response of the morphological filtering. The response values along the watershed lines, together with information about orientation, is then used as an input to a line-following algorithm that produces a set of line segments. During a global analysis step, the produced line segments, together with an additional set of segments that correspond to all possible connections between them, are organized as a graph. The nodes of the graph are associated with an observation field and a dedicated Markov Random field that describes the geometrical properties of the linear structures of interest. Using a binary set of interpretation labels, the final result of optimal connected configurations (optimal graph labelling) is then extracted based on a maximum a posteriori probability criterion. An example is given in Figure 7. The main advantage of our approach is its high detection performance in heavily textured environments and its ability to identify elongated structures of different size.



(a) (b)

Figure 7. Linear structures detection results. (a) original VNIR image, (b) detected linear structures.

4.2. Buildings

The presence of buildings, which correspond either to residences or parts of an industrial infrastructure, can be considered as a major indicator of low risk areas with human activity. We have investigated the possibility of identifying building rooftops from a single remotely sensed image, without the use of digital terrain models or stereo vision. Our approach is based on an image interpretation model, which combines both 2-D and 3-D contextual information of the imaged scene (Katartzis et al. 2001a). The building rooftop hypotheses are extracted using a contour-based grouping hierarchy using the principles of perceptual organization (Schluter et al. 2000). We use levels of increasing abstraction to represent intermediate results and thus successively bridging the gap between raw image data and the 3-D objects of interest. The different steps of the method are depicted in Figure 8.

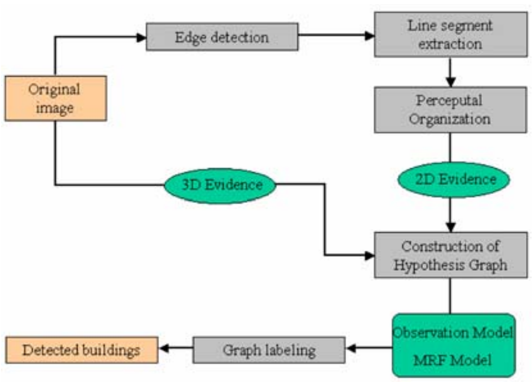


Figure 8. Building detection flow chart.

The hypothesis verification step is performed with a MRF-based labelling scheme for contour grouping. Initially, we apply a color edge detection operator, followed by an edge-linking step in order to extract the building boundaries. The resulting edge points are grouped in a hierarchical way into line segments, junctions, parallel lines and closed contours (polygons with pairs of parallel segments). The grouping is based on both geometrical similarities and color region attributes. The highest level of the hierarchy consists of the hypothesized rooftops. Regarding each grouping as an object hypothesis (token), we can create a hypothesis graph with certain node attributes and relations. Figure 9 describes the generation of the hypothesis graph.

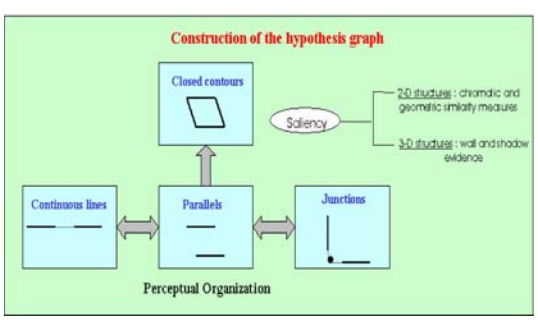


Figure 9. Perceptual organization - construction of hypothesis graph.

The attributes of the graph reflect our a priori knowledge about the shape of the 3-D structure we are looking for, and are dependent on domain specific knowledge like the type of camera model used and evidence from the presence of shadows and walls. The node relations, which can be either supporting or competing, describe the dependencies between hypotheses with regard to a globally consistent interpretation of the image. We associate a MRF to the grouping graph in order to find the global optimal configuration for the locally interacting grouping hypothesis. The random variables associated to each node of the graph represent the evidence of the hypothesis being a correct interpretation of the image data. Finally, a hypothesis verification step (optimal graph labelling) is carried out via a stochastic labelling scheme, based on a MAP criterion. Illustrative results are given in Figure 10.

Figure 10. Building detection results.

5. CONCLUSIONS

In this paper, we illustrate the potential of computer vision techniques for the (semi-) automated extraction of minefield indicators in rural areas with heavily cluttered environments. For this purpose, we have developed a series of model-based methods for the detection of regions/objects/structures with a versatile appearance in terms of size, shape and spectral/textural signatures.

ACKNOWLEDGMENTS

The authors gratefully acknowledge support from the EC-IST-INFSO project IST-2000-25300 "Airborne minefield area reduction", and the Fund for Scientific Research - Flanders (FWO, Belgium) project FWOAL264 "Multiscale Stochastic Models for Image Denoising and Segmentation".

REFERENCES

Batman S., Goutsias B., 2003, IEEE Trans. on Image Processing, 12(5), 509

Bouman C.A., Shapiro M., 1994, IEEE Trans. on Image Processing, 3 (2), 162

DiZenzo S., 1986, Computer Vision, Graphics and Image Processing, 33, 116

Geman S., Geman D., 1984, IEEE Trans. Pattern Anal. Mach. Intell., 6, 721

Jackway P.T., 1996, IEEE Transactions on Image Processing: Special Issue on Nonlinear Processing, 5

Katartzis A., Sahli H., Nyssen E., Cornelis J., 2001a, In: IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2001), vol. (6), 2832–2834, Sydney, Australia

Katartzis A., Sahli H., Pizurica V., Cornelis J., 2001b, IEEE Trans. on Geoscience and Remote Sensing, 39 (9), 2073

Katartzis A., Vanhamel I., Sahli H., 2003, In: IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data (WARDS 2003), NASA Goddard Center, Greenbelt Maryland, USA

van Kempen L., Katartzis A., Pizurica V., Cornelis J., Sahli H., 1999, In: Euroconference on Sensor systems and signal processing techniques applied to the detection of mines and unexploded ordnance (MINE99), 48–53, Firenze, Italy

Koenderink J., 1984, Biol. Cybern., 50, 363

Li S.Z., 1995, Markov random field modeling in computer vision, Springer Verlag

Maathuis B., 2001, Remote Sensing based detection of landmine suspect areas and minefields, PhD Thesis, Department of Geosciences, University of Hamburg

Pizurica V., Katartzis A., Cornelis J., Sahli H., 1999, In: 2nd International Symposium Operationalization of Remote Sensing, 399–406, Enschede, Nederland

Sahli H., Busto F., Katartzis A., Vanhamel I., Chan J.C.W., 2004, In: ESA-EUSC 2004: Theory and Applications of Knowledge driven Image Information Mining, with focus on Earth Observation, Madrid, Spain

Sapiro G., 2001, Geometric Partial Differential Equations and Image Processing, Cambridge University Press

Schluter D., Wachsmuth S., Sagerer G., Posch S., 2000, In: Proc. IEEE International Conference on Image Processing, vol. II, 100–103

Vanhamel I., Pratikakis I., Sahli H., 2003, IEEE Trans. on Image Processing, 12 (6), 617

Whitaker R., Gerig G., 1994, In: ter Haar Romeny B. (ed.) Geometry-Driven Diffusion in Computer Vision, 93–134, Kluwer, Dordrecht