EVALUATING MODEL FIDELITY IN AN AERIAL IMAGE ANALYSIS SYSTEM

F. Quint M. Sties
Institute for Photogrammetry and Remote Sensing
University of Karlsruhe
76128 Karlsruhe, Germany
quint@ipf.bau-verm.uni-karlsruhe.de

Commission III, Working Group 3

KEY WORDS: Colour, Aerial, Image, Model, Vision, Colour Aerial Image, Aerial Image Segmentation, Bayesian Model

ABSTRACT

The purpose of the system MOSES is the automatic recognition of objects in aerial images. In this system, a model based structural image analysis is performed. Specific models are gained through the analysis of digital maps. The models are stored in seantic networks. Image analysis is implemented as a search. To direct this search, one has to evaluate each state of the analysis process. One part of the computed valuations is the model fidelity, which is a measure for the goodness of match between the choosen image primitives and the specific model. We present in this article the procedures used to compute the model fidelity for line segments and polygons.

KURZFASSUNG

Das System MOSES dient der automatischen Erkennung von Objekten in Luftbildern. Es führt eine modellbasierte, strukturelle Bildanalyse durch, wobei spezifische Modelle der zu analysierenden Szene durch die Analyse von digitalen Karten gewonnen werden. Die Modelle werden in semantischen Netzen gespeichert. Der Analysevorgang ist ein Suchvorgang, zu dessen Steuerung Bewertungen des aktuellen Analysezustandes anzugeben sind. Ein Teil dieser Bewertungen ist die Modelltreue, die angibt, wie gut die ausgewählten Bildprimitiven zu dem vorgegebenen Modell passen. In diesem Artikel stellen wir die Prozeduren vor, mit denen die Modelltreue für Strecken und Polygone berechnet wird.

1 INTRODUCTION

Understanding of aerial images is one of the most challenging tasks in computer vision. Due to its complexity, a model based analysis has been found to be mandatory since several years, see e.g.(Agin, 1979), (Matsuyama and Hwang, 1990), (McKeown et al., 1985), (Nicolin and Gabler, 1987), (Sandakly and Giraudon, 1994), (Stilla, 1995). In our system MOSES (*M*ap *O*riented *SE*mantic image under*S*tanding) (Quint and Sties, 1995) we too perform a structural, model based analysis. We are interested in the recognition of objects in urban environment using large scale aerial images.

2 MOSES

One of the main characteristics of the system MOSES is that large scale topographical maps are used to automatically refine the models used for image analysis. Thus the object recognition process consists of three phases. The architecture of our system is shown in Fig. 1. The generative model contains domain independent, common sense knowledge the system designer has about the environment. The generic models in the map domain and in the image domain are specialisations of the generative model and they reflect the particularities of the representations of our environment in the map and image respectively. The models contain both declarative knowledge, which describes the structure of the objects, and procedural knowledge, which contains the methods used during the map and image analysis process. As a repository for the models semantic networks (Findler, 1979) are used, as implemented by the system ERNEST (Kummert et al., 1993).

The generative model and the generic models are that part of the system which is build by the system developer. The models and scene descriptions decribed in the sequel are automatically build in analysis processes.

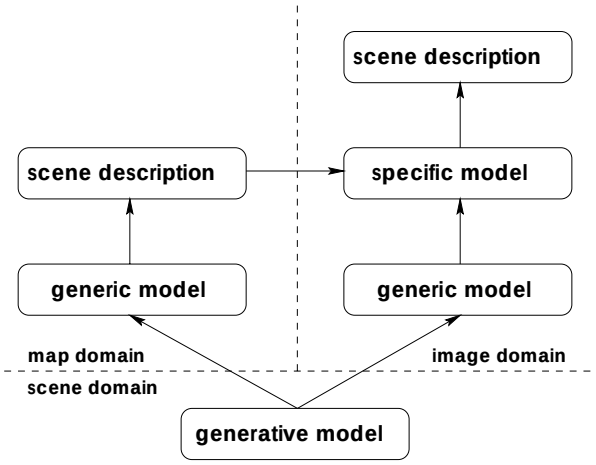

Figure 1: Architecture of the system MOSES

2.1 Map analysis

In the first phase, the generic model in the map domain is used to analyse the map, which is available as a list of digitized contours. The map analysis process is similar to the image analysis process which will be described in a following section. The result of the map analysis process is a description of the scene, as far as it can be constructed out of the map data. The scene description is stored in a semantic network. The nodes of the semantic network represent objects, parts and subparts of the scene. They are described with attributes, which in this case mainly contain their geometric properties. Links between the nodes represent relations between the corresponding objects or parts. The *part-of* relation describes the structure of the scene objects and along a *specialisation* link properties are inherited.

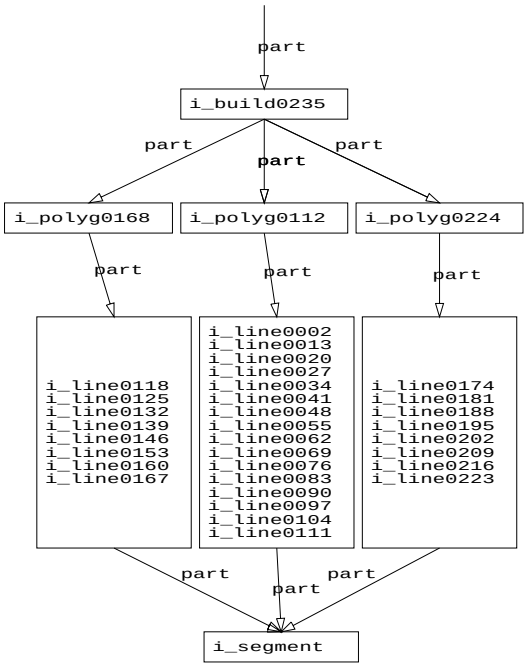

Figure 2: Detail of the part-of hierarchy of the specific model

2.2 Model building

In the second phase the scene description obtained after the map analysis is combined with the generic model in the image domain and results in the specific model in the image domain. A detail of the specific model, representing building nr. 0235 and its parts as far as they are given in the map, is given in Fig. 2. For each node (instance) in the scene description we create in the specific model a new node (concept), which is a specialisation of the corresponding concept in the generic model in the image domain. This new concept inherits the declarative and procedural knowledge of the concept in the generic model.

The values of the attributes in the scene description after map analysis are stored as restrictions for the corresponding attributes of the newly created concepts. They serve as initial estimates while verifying these values in the image data. The links between the instances in the scene description are transfered accordingly into links between the new concepts. Whilst the generic model in the image domain describes the representation of an arbitrary scene in an aerial image in a very general form, the specific model in the image domain describes in a detailed manner that part of the world, which is subject to the current analysis. The grade of detail depends of course from the contents of the map.

2.3 Image primitives

Prior to the model based image analysis primitives are extracted from the image data. We work with large scale color aerial images, which after digitization have a pixel size of 30 cm x 30 cm on the ground. As primitives serve line segments and regions. The line segments are extracted with a gradient based procedure (Quint and Bähr, 1994). The regions are gained by segmenting the aerial image using a Bayesian homogeneity predicate (Quint, 1996). The regions and the line segments are combined in an attributed undirected graph. The nodes of the graph are attributed with the regions. Nodes corresponding to neighbouring regions are connected with links. A link between two nodes is attributed with the line segment(s) which build the border between the corresponding regions. This feature graph is the database on which the model based image analysis operates.

2.4 Image analysis

In the third phase the specific model in the image domain is used to perfom the actual image analysis. The aim of this phase is to verify in the image the objects found after the map analysis and to detect and describe other objects of the scene which are not represented in the map. For the later, the context gained through the verification of the map objects will be helpfull.

The strategy followed in the analysis process is a general, problem independent strategy provided by the shell ERNEST. The analysis starts by creating a modified concept for the goal concept (expansion step). A modified concept is a preliminary result and it reflects constraints for the concept that have been determinated out of the context of the current analysis state.

Following top-down the hierarchy in the semantic network, stepwise the concepts on lower hierarchical levels are expanded until a concept on the lowest level is reached. Since this concept does not depend from other concepts, a correspondence between him and a primitive in the data base can be established and its attributes can be calculated. We call this instantiation. Analysis now moves bottom-up to the concept at the next higher hierarchical level. If instances have been found for all parts of this concept, the concept itself can be instantiated. Otherwise the analysis continues with the next uninstantiated concept on a lower level. Thus, in the analysis process top-down and bottom-up processing alternate. After an instantiation, the acquired knowledge is propagated bottom-up and top-down to impose constraints and restrict the search space. As well, expansion and instantiation alternate during the analysis.

Generally, while performing an instantiation it is possible to establish several correspondences between a concept and primitives in the data base. However, only one of these correspondences leads to the correct interpretation. Since it usually is not possible to ultimately decide at the lower levels which correspondence is correct, all possible correspondences have to be accounted for.

Thus, the image analysis is a search process, which can be graphically represented by a tree. Each node of the tree represents a state of the analysis process. If in a given state several correspondeces are possible, the search tree is splitted: for each hypothesis a new node as successor of the current node is created.

The analysis process continues with that leaf node of the search tree, which is considered to be the best according to a problem dependent evaluation. It is know that the problem of finding an optimal path in a search tree can be solved by the A^*-algorithm (Nilsson, 1982). Its application is possible if one can evaluate the path from the root node to the current node and if one can give an estimate for the valuation of the path from the current node to the (not yet known) terminal node containing the solution.

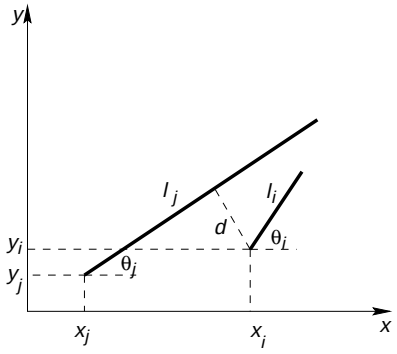

Figure 3: Parameters used to describe a line segment

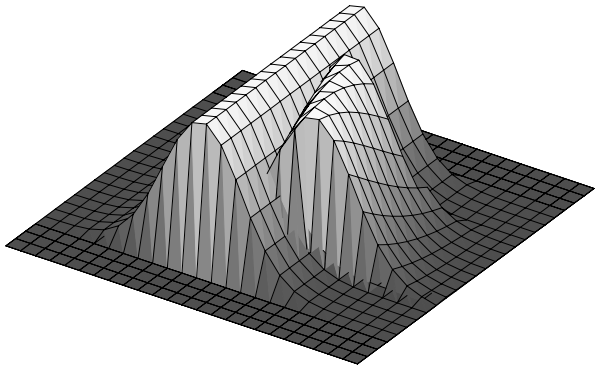

Figure 4: Neighbourhood function for the postion of line segments

3 VALUATIONS

The functions which evaluate the states of the analysis are very important since they are not only responsible for the efficiency of the search, but they are also decisive for the success or failure of the analysis. We relate the valuation of the search path to the valuation of the analysis goal in the given state of the analysis. The valuation of the goal is calculated considering the valuations of the instances and modified concepts already created and the estimates for the valuations of the instances and modified concepts which will be created in the path from the current node to the solution node.

When an instantiation is performed implicitly a hypotheses of match is established between the concept, for which the instantiation takes place and the chosen primitives from the data base. Since we can not ultimately decide at the moment the instantiation is performed, if it is the correct one, we are working under uncertainty and we have to quantify our uncertainty. Thus, at the level of each concept in the semantic network, we have a dichotomous frame of discernment with the events: the chosen primitives

- match
- do not match

to the concept (i.e. model).

The valuations computed for the instances and modified concepts in each state of the analysis are measures of our subjective belief in these hypotheses. We embed the valuations in the Dempster-Shafer theory of evidence (Shafer, 1976). The different valuations are combined and propagated in the hierarchy of the semantic network to result in the valuation of the analysis goal.

We impose the condition for a valuation to be a number between 0 and 1. The higher the valuation is, the higher is our subjective belief in the corresponding hypothesis. Since the valuations are used to compare different states of the analysis, there is no need for absolute exactness of their values, but only the relations in the ranking of the analysis states and the corresponding valuations have to be preserved.

We evaluate two aspects for our hypotheses of match: the compatibility and the model fidelity. The compatibility evaluates an analysis state considering the principles of perceptual grouping. It is calculated based on geometric, topologic and radiometric properties of the image primitives only. In this category belong for example the goodness of fit of several line segments extracted from the image data to form an edge of an object, the goodness of fit of several edges to form a polygon, the compatibility of the polarity of edges to form a polygon etc. The model fidelity measures the goodness of fit between the image primitives and the specific model gained through the analysis of the map. Portraying it in simplified terms, one can say that the compatibility is a measure for the ability of the chosen image primitives to form an object, whereas the model fidelity is a measure for the ability to form exactly that object, which is predicted by the map. We present in this article some of the measures used for the evaluation of the model fidelity.

4 MODEL FIDELITY

4.1 Model fidelity for line segments

At the level of line segments we define the model fidelity with help of a distance function between the image primitves and the contours stored in the specific model after map analysis. The distance functions results from a metric defined with help of a set of square integrable funtions on the parametric space for line segments.

We describe a line segment with the coordintes of its starting point, its length and the angle between the line and positive x–axis (see Fig. 3). Thus, a line segment s_i is represented in the space $S = (x, y, l, \theta)$ by the point $s_i = (x_i, y_i, l_i, \theta_i)$. The coordinates of a line segment take values $(x, y) \in \mathbb{R}^2$, the lenght of line is in $l \in \mathbb{R}_+$ and the angle is in $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2}]$. The space (x, y, l, θ) is the cartesian product of the before mentioned subspaces and is different from $\mathbb{R}^n$. For this reason we do not use the euclidean distance between two points in this space to calculate the distance between two line segments, but use instead a metric defined on an isomorphic space of functions.

We define an isomorphism by attaching each point s_i in the space S a function $n_i(x, y, l, \theta)$ from the space of square integrable functions $\mathcal{L}^2(S)$. We call this function *neighbourhood function*. As a distance between two line segments s_i and s_j we now use the distance defined on the family of functions n_i. It is well known, that a distance defined with the expression:

$$d_{ij} = \left[\int_S \big(n_i(x, y, l, \theta) - n_j(x, y, l, \theta) \big)^2 dx\, dy\, dl\, d\theta \right]^{\frac{1}{2}} \quad (1)$$

induces a metric on $\mathcal{L}^2(S)$. If we choose the functions $n_i(x, y, l, \theta)$ such, that their norm in the above given met-

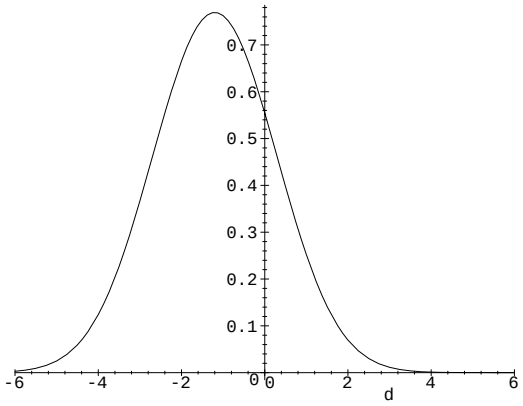

Figure 5: Position fidelity as a function of d (Fig. 3)

ric is equal to 1, i.e.

$$\int_S \big(n_i(x,y,l,\theta)\big)^2 dx\,dy\,dl\,d\theta \stackrel{!}{=} 1, \tag{2}$$

the expression (1) simplifies to:

$$d_{ij} = \left[2 - 2\int_S n_i(x,y,l,\theta) n_j(x,y,l,\theta) dx\,dy\,dl\,d\theta\right]^{\frac{1}{2}}. \tag{3}$$

The distance d_{ij} decreases when the integral in eq. (3) increases. If the neighbourhood functions are positive functions the integral in eg. (3) takes values between 0 and 1.

We have formulated our search problem using as evaluations of the nodes in the search tree merit functions and not cost functions. The reason for this is pragmatic: it is more natural to evaluate the goodness than the badness of a match. Thus, we will not use the distance as given by equation (3) as a measure of the model fidelity, but only the integral in equation (3):

$$m_{ij} = \int_S n_i(x,y,l,\theta) n_j(x,y,l,\theta) dx\,dy\,dl\,d\theta \tag{4}$$

This integral equals to the cosinus of the angle between the two versors n_i and n_j in the vector space $\mathcal{L}^2(S)$ and can be thought of as a correlation measure between these two versors.

The neighbourhood functions are chosen regarding the physics of the image formation process and some heuristics motivated by experience. We construct the function $n_i(x,y,l,\theta)$ as a product of three functions defined on $\mathbb{R}^2, \mathbb{R}_+$ and $(-\frac{\pi}{2}, \frac{\pi}{2}]$ respectively:

$$n_i(x,y,l,\theta) = f_i(x,y)\, g_i(l)\, h_i(\theta)$$

Since the parameters of the camera and the position of the airplane at the moment the aerial image was taken are known, we can determinate the transformation between the image coordinates and the coordinates in the specific model (map coordinates). Using this we transform the image primitives into the map coordinate system. Assuming that the corresponding contours are depicted in the map, there are several error sources which are responsible for the fact that the line segments extracted from the image will not overlap with the map contours. These are for example inaccuracies in:

- the extraction of line segments from the image,
- the determination of the transformation parameters,
- the aquisition and digitization of the map data.

Subsuming all these effects, we can safely assume that the position of the image primitives is normally distributed arround their "true" position as given by the specific model.

For this reason we use as a neighbourhood function $f_i(x,y)$ for the position of the line segments a Gaussian shaped function. However, since we do not want to evaluate differently the situations when a short line segment lies in the middle of its model line or closer to the endpoints, our function is constant along the length of the line. We choose for the neighbourhood function $f_i(x,y)$:

$$f_i(x,y) = K_{xy} \exp\left(-\frac{\big((x-x_i)\sin\theta_i - (y-y_i)\cos\theta_i\big)^2}{2\sigma^2}\right)$$

for positions (x,y) between the endpoints of a line, i.e. $\{(x,y) \mid (x-x_i)\cos\theta_i + (y-y_i)\sin\theta_i \geq 0 \;\&\&\; (x-x_i)\cos\theta_i + (y-y_i)\sin\theta_i \leq l_i\}$, and $f_i(x,y) = 0$ otherwise. The neighbourhood functions $f_i(x,y)$ and $f_j(x,y)$ for the example of the line segments in Fig. 3 are displayed in Fig. 4. The variance of the Gaussian is chosen equal to the residual mean square error of the transformation.

For the part of the neighbourhood function, which depends from the length of the line, we choose a function, which is proportional to the square root of the lengh "inside" the line and 0 "outside":

$$g_i(l) = \begin{cases} K_l\sqrt{l} & \text{if } l \in [0, l_i] \\ 0 & otherwise \end{cases}$$

As we will see later, this choise penalizes image primitives proportional to the ratio of their length and the lenght of the model contour.

The considerations regarding the uncertainty of the position of line segments applies also for small deviations of the angle. Thus, the neighbourhood function for the angle is chosen following similar reflections. But, because the domain of definition of the angle is an interval and because we want a stronger penalization of large deviations of the angle, we use a trigonometric function instead of the Gaussian shaped function:

$$h_i(\theta) = K_\theta \cos(\theta - \theta_i)$$

The constants K_{xy}, K_l and K_θ are calculated imposing normalization for each of the partial neighbourhood functions. Thus we also assure the fulfillment of condition (2).

With this choise of neighbourhood functions, the integral for the model fidelity is separable into three terms: the position fidelity, the length fidelity and the angle fidelity. The integral over the product of the neighbourhood functions for the position, i.e. the position fidelity can generally not be expressed in a closed form. However, if the angle between the two lines is small or the parameter σ is is in the same order of magnitude as the mean geometric distances between the two line segments, which can be safely assumed in our situation, then a good approximation is given by:

$$\int_{\mathbb{R}^2} f_i(x,y) f_j(x,y) dx\,dy = \frac{\sqrt{\pi}\sigma}{l_i \sin\Delta\theta} \times$$
$$\left(\operatorname{erf}\left(\frac{u_1 \sin\Delta\theta - A}{\sigma\sqrt{2+2\cos\Delta\theta^2}}\right) - \operatorname{erf}\left(\frac{u_2 \sin\Delta\theta - A}{\sigma\sqrt{2+2\cos\Delta\theta^2}}\right)\right)$$

where $\Delta\theta = \theta_j - \theta_i$, $A = -(x_i - x_j)\sin\theta_j + (y_i - y_j)\cos\theta_j$ and u_1 and u_2 are the u-coordinates of the endpoints of line l_i in a coordinate system with origin in the starting point of line l_j and where the u-axis is the line l_j. For the situation shown in Fig. 3 the position fidelity varies with a parallel displacement of a line as a function of d as shown in Fig. 5.

The integrals over the neighbourhood functions for the length and the angle of the line segments can be expressed in closed form and result to:

$$\int_{\mathbb{R}_+} g_i(l)g_j(l)dl = \frac{\min(l_i, l_j)^2}{l_i l_j}$$

and

$$\int_{-\pi/2}^{\pi/2} h_i(\theta)h_j(\theta)d\theta = cos(\theta_i - \theta_j)$$

The length fidelity amounts thus to the ratio of the lenght of the shorter line to the lenght of the longer line. The angle fidelity is the cosinus of the angle difference of the two lines. The total model fidelity for line segments is given by the product of the three components.

Usually, due to noise influence the visible contour of an object in the image is broken and thus several line segments will form that contour. In this case, the contour is constructed step by step by adding another line segment until the contour is completed. The A^*-algorithm requires also an optimistic estimate of the merit for future instantiations. Given a partially instantiated contour an optimistic prediction for the future instantiations is obtained when one elongates the already instantiated contour until the model is completed. The estimate of the model fidelity for line segment for the future instantiations is also computed with the above described procedure for the predicted contours.

4.2 Model fidelity for polygons

A different approach for the model fidelity is used at the hierarchical level of polygons. Whilst at the level of line segments the similarity in positon and orientation between the selected image primitives and the model contour has been evaluated, we evaluate at the level of polygons the similarity between the shape of the polygon created by the image primitives and the shape of the model polygon.

The corner points of the polygon in the image domain are obtained as intersections of the choosen image primitives. In the case where several image primitives form an edge of an object, these primitives are replaced for the purpose of the corner point calculation with a regression line. The error produced by the approximation with the regression line is taken into account in the valuations of the compatibility. In the case where no correspondence could be established between an edge of an object and an image primitive we make a wildcard assignment to the current edge. In this case the correspondig corner points are choosen to be the end point of the image primitive assigned to the edge previous to and the starting point of the image primitive assigned to the edge after the wildcard-assigned edge. The wildcard assignments however lead to a a penalization in the model fidelity of the line segments.

To not include position and orientation errors in our measure we first transform the polygon in the image domain on the model polygon. We take a similarity transformation between the corresponding corner points of the two polygons and calculate the transformation parameters such that the residual mean square error is minimal. Since the scale of the image and the map are known, we fix the scale parameter in the similarity transformation to the known value.

The resulting minimal mean square error is a measure for the similarity of the shapes of the two polygons. We gain our subjective belief in the hypotheses of match between the image polygon and the model polygon with help of a fuzzy function:

$$p_{ij}(r) = \exp\left(-\frac{r^2}{\sigma_r^2}\right)$$

where r is the residual mean square error after the transformation and σ_r is a parameter whose value is determinated by experiment. As experiments have shown the image analysis process is robust with respect to this parameter.

5 Conclusion

We presented a method to derive a merit function for guiding search in a model based image analysis system. The Dempster-Shafer theory of evidence serves as a theoretical background. To propagate valuations calculated at different levels of the hierarchical approach we have extended proposals found in the literature to suit our needs.

The derived merit function gives a common ground for the comparison of paths developed further with paths abandoned earlier in the search tree. The main difficulty in finding a merit function for informed search methods is to give an estimate for the merit of the yet unknown path from the current node to the solution node of the search tree. An important property of the derived merit function is, that it is not necessary to assign valuations to yet unknown instances and modified concepts. By explicitly modeling the lack of knowledge with the methods offered by Dempster-Shafer theory, our formalism provides in a natural way the required overestimate for the merit of the yet unknown path from the current node to the solution node of the search tree.

The experiments have shown that our merit function can be used successfully to guide search with an ε-A^*-algorithm. The merit function is robust with respect to the parameter ε and leads to a good solution for values of ε up to a problem dependent upper bound. Higher values of the parameter ε lead to a considerable speed up and smaller memory requirement of the analysis process. Several other factors also contribute to the success of the analysis process, i.e. the valuations computed for the instances and modified concepts at the different levels of the hierarchical model, although they are not in the scope of this paper. For defining these valuations we take advantage of having a specific model for the objects to be recognized in the image. This specific model is automatically build by our system through the analysis of the available map of the scene. We plan to extend out system to recognize objects in the image, which are not represented in the map and for which a specific model is thus not available.

References

Agin, J. (1979). Knowledge–based detection and classification of vehicles and other objects in aerial images. In *Proceedings of the DARPA Image Understanding Workshop*, pages 66–71, Palo Alto, CA.

Findler, N. (1979). *Associative Networks*. Academic Press, Orlando.

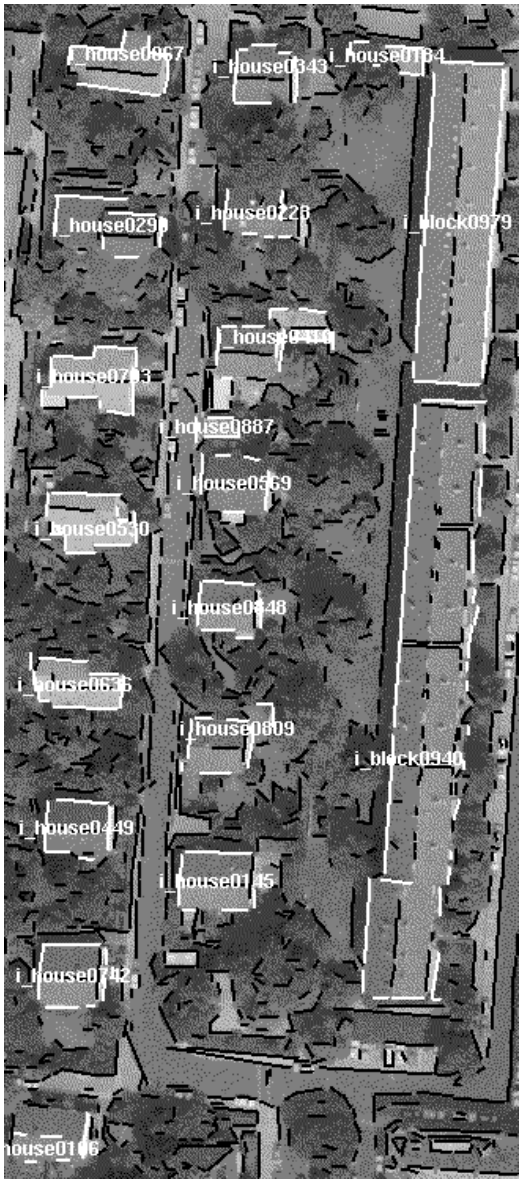

Figure 6: Result

Kummert, F., Niemann, H., Prechtel, R., and Sagerer, G. (1993). Control and explanation in a signal understanding environment. *Signal Processing*, 32:111–145.

Matsuyama, T. and Hwang, V. (1990). *SIGMA: A Knowledge-Based Aerial Image Understanding System*. Advances in Computer Vision and Machine Intelligence. Plenum Press, New York, London.

McKeown, D., Harvey, W., and McDermott, J. (1985). Rule based interpretation of aerial imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(5):570–585.

Nicolin, B. and Gabler, R. (1987). A knowledge–based system for the analysis of aerial images. *IEEE Transactions on Geoscience and Remote Sensing*, 25(3):317–329.

Nilsson, N. (1982). *Principles of artificial intelligence*. Springer-Verlag, Berlin.

Quint, F. (1996). Colour aerial image segmentation using a bayesian homogeneity predicate and map knowledge. In *Proceedings of the 18th ISPRS-Congress*, Vienna.

Quint, F. and Bähr, H.-P. (1994). Feature extraction for map based image interpretation. In Shi, X., Du, D., and Gao, W., editors, *Third Internartional Colloquium of LIESMARS: Integration, Automation and Intelligence in Photogrammetry, Remote Sensing and GIS*, pages 1–8, Wuhan, China.

Quint, F. and Sties, M. (1995). Map–based semantic modeling for the extraction of objects from aerial images. In Grün, A., Kübler, O., and Agouris, P., editors, *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, pages 307–316. Birkhäuser, Basel.

Sandakly, F. and Giraudon, G. (1994). Multispecialist system for 3D scene analysis. In Cohn, A., editor, *11th European Conference on Artificial Intelligence, ECAI 94*, pages 771–775. John Wiley & Sons, Ltd.

Shafer, G. (1976). *A mathematical theory of evidence*. Princeton University Press.

Stilla, U. (1995). Map-aided structural analysis of aerial images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 50(4):3–10.