THE RESULTS OF THRESHOLD SETTINGS ON MODELBASED RECOGNITION AND PARAMETER ESTIMATION OF BUILDINGS FROM MULTI-VIEW AERIAL IMAGERY

Luuk Spreeuwers, Klamer Schutte* and Zweitze Houkes
Laboratory for Measurement and Instrumentation, Dept. of Electrical Engineering
University of Twente, P.O. Box 217, 7500 AE Enschede, Netherlands
*TNO Physics and Electronics Laboratory, Netherlands

KEY WORDS: modelbased, recognition, parameter estimation, hypothesis correspondence, multi-parameter segmentation

ABSTRACT

This paper describes a system for analysis of aerial images of urban areas using multiple images from different viewpoints. In this paper the emphasis is on the discussion of the experimental evaluation using segmented images obtained by applying 3 different parameters in the segmentation-process. The proposed approach combines bottom-up and top-down processing. To evaluate statistically the performance of the system, a set of 50 realisations of 5 images from different viewpoints was used, which was generated by combining real and ray-traced images. The experiments show a significant improvement of reliability and accuracy if multi-segmentation is used in multi-view imagery, instead of single-segmentation.

1 INTRODUCTION

The goal of this research is to design and evaluate a system capable of analysing aerial photographs of urban areas. The output of this process is a 3-D scene description which can be used to update a GIS. Basically, the process involves the recognition of objects present in the scene and estimation of the parameters describing the objects: position, size and orientation. If the camera model and parameters are known, in most cases, the 3-D parameters of objects can be estimated from a single image. The obtainable accuracy is, however, highly dependent on the viewpoint. Furthermore, from certain viewpoints objects may be difficult to recognise, because parts of them are invisible or blend with the background. Also if objects occlude each other, it may be impossible to reliably recognise the imaged objects or to obtain reliable estimates of object parameters. Stereo vision provides a more robust estimation for the object parameters but does for the general case not solve the recognition or occlusion problem. In the presented work, as we did in previous work, multiple images are used recorded from different viewpoints. From previous work (Spreeuwers et al., 1997) and in spite of the good results, it was concluded that the segmentation process is still a bottleneck. In cases where it is difficult to obtain a good segmentation, using only a single value of the segmentation parameters, it is profitably to use multiple values. The hypothesis corresponding process uses the fom to remove the hypotheses based on poor segmentation results and to preserve the hypotheses based on good segmentation results.

2 A MODEL BASED APPROACH

The proposed method combines top-down and bottom-up techniques. Figure 1 depicts the basic setup of the system.

The following six steps are distinguished:

Segmentation: region based segmentation of the images.

Shape-based segmentation correction: using knowledge about the expected shape of segments of man-made objects the segmentation is improved.

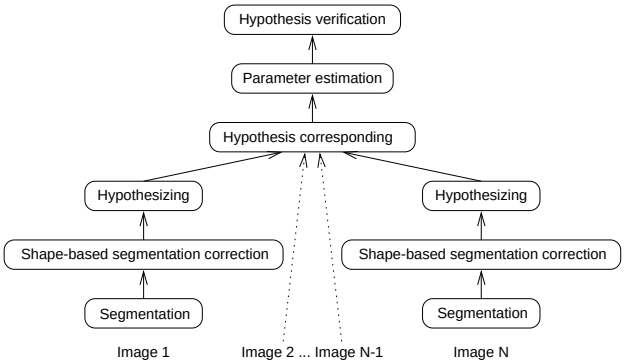

Figure 1: Setup for the proposed system

Hypothesising: using local evidence candidate scene descriptions are generated using a single image. For all images those scene descriptions are generated which have a sufficiently high likelihood.

Hypothesis corresponding: find out which hypotheses in the set of images correspond and thus refer to the same objects.

Parameter estimation: find the best set of parameters for all candidate scene descriptions using all the images, by predicting the segment shapes and selecting those parameters that result in the highest compatibility with the segmented images.

Hypothesis verification: Select from all candidate scene descriptions those that are most compatible with the measured images and do not contradict each other.

2.1 Segmentation

The segmentation process (Schutte, 1994) consists of a region growing process (Schutte, 1993) and a segmentation improvement step in which a priori shape knowledge is used.

In the shape based segmentation correction we used a set of procedures for incorporating such knowledge into the segmentation process, similar to the rule bases proposed by

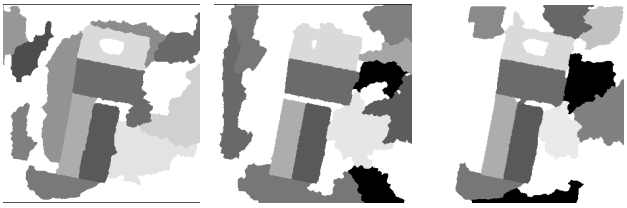

Figure 2: Results of the segmentation for 3 different and increasing values of the segmentation parameter.

Nazif and Levine (Nazif and Levine, 1984). The shape knowledge used is based on the use of polyhedra to describe man-made objects. The projections of the polyhedra on the image plane are polygons. Also we use the fact that the polygons tend to have few corners and a certain minimum area. Figure 2 shows the effect of a variation of the segmentation parameter, which determines whether a pixel or a region should belong to a segment, on the segmentation results.

2.2 Hypothesising: from regions to parametric object models

The input to the hypothesis generation is a description of the regions found in the image. Such a description is noisy, due to the nature of the images and the segmentation process. This means that some regions are found which do not correspond to visible object faces, and vice versa. The method should recognise the object, even if not all of the faces of the object correspond to a region. The hypotheses to be found consist of parametric object models. The models used are volumetric objects, such as a block, representing an office building, house etc. The output of the hypothesising method should include initial estimates needed by the parameter estimation procedure. Erroneous hypotheses generated by the hypothesis generator will be discarded by either the hypothesis corresponding or the final hypotheses verification process.

The hypothesising method consists of 4 steps. The first step (detection) comprises the extraction relational graphs from the segmentation. The second step is a relaxation process to find the best match with precalculated graphs of object models (*aspects*), stored in a database. Bipartite matching ensures unambiguity. In the last step the graph descriptions are transformed into parametric object models. A full description of the hypothesising method can be found in (Schutte and Boersema, 1993).

The model data base, shown in figure 3 consists of the various objects which are of interest and can be expected in the scene. The objects currently defined are BlockShapedBuilding and House. For each object a set of aspects exists in the database.

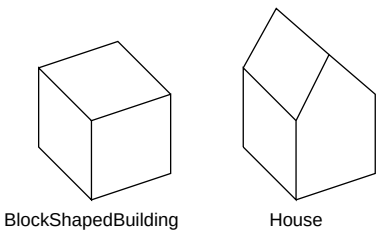

Figure 3: The objects in the model database

2.3 Hypothesis correspondence in multiple view imagery

After the hypothesis generation stage on single images, for each image there is a list of hypotheses, containing for each hypothesis the object class and initial estimation of position, orientation and size parameters.

The objective of the hypothesis corresponding stage is firstly to find correspondences between hypotheses for the different images and reduce the total number of hypotheses by creating hypothesis groups with corresponding hypotheses. Secondly, unreliable hypotheses (e.g. that occur only in a single view) are discarded. Thirdly, not corresponding hypotheses that occupy the same space are marked *mutually exclusive*, since they cannot be valid simultaneously. Finally hypotheses that are *close* and do not correspond are marked, because they may cause occlusion. In order to determine whether two hypotheses i and j correspond, are close or mutually exclusive, three distance measures are defined:

$D(i,j)$ geometrical distance between the centres of gravity of the two hypotheses i and j

$O(i,j)$ measure of *overlap*, i.e. how much space is shared by the ground planes of the hypotheses i and j

$M(i,j)$ feature match quality, i.e. how well the hypothesised objects i and j resemble, taking into account: object class, size, orientation

Correspondence is defined as:

$$(O(i,j) \geq O_{min}) \textbf{ and } (M(i,j) \geq M_{min}) \qquad (1)$$

so for correspondence there must be a certain minimum of overlap between the hypotheses and the hypotheses must resemble sufficiently. Two hypotheses are marked mutually exclusive if:

$$(O(i,j) \geq O_{min}) \textbf{ and } (M(i,j) < M_{min}) \qquad (2)$$

i.e. the hypotheses occupy the same space, but do not resemble each other, hence it is impossible that both are correct. Finally two hypotheses are *close* if:

$$(D(i,j) < D_{max}) \qquad (3)$$

In the above formulas 1-3, the constants O_{min}, M_{min} and D_{max} depend on (among others) the size of the buildings, the flight height and viewing angles.

Note that two hypotheses can at most have *one* of the above described relations: they *either* correspond *or* are exclusive *or* are close *or* have none of the relations.

2.4 Parameter estimation

The scene descriptions resulting from the hypothesis corresponding stage, consist of a list of hypothesis groups each with corresponding initial estimates of the parameters (position, size and orientation). For each hypothesis group the estimation process predicts the segments in all the images and adjusts the parameters for maximum compatibility with the segmented images. The setup of the estimation process is shown in fig.4.

The scene model consists of the hypothesised objects and the illumination and the camera models. The camera model used is a pinhole projection and the scene is illuminated by

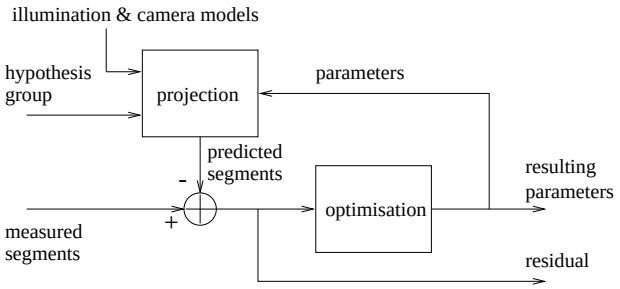

Figure 4: Setup of the iterative parameter estimation process

the sun. The buildings are modelled using primitive volumes, like block, roof etc., each with their own set of parameters. The set of parameters used consists of the internal parameters of the primitives, such as their shape and the parameters describing the pose.

The measurement vector $\underline{\vec{m}}$, in our case a scan line code representation of the segments in the images (Schutte, 1994, Schutte and Hilhorst, 1993), is related to the parameters via a nonlinear model $\vec{f}$, which involves the acquisition by the camera, segmentation and conversion to scan line code representation:

$$\underline{\vec{m}} = \vec{f}(\vec{p}) + \underline{\vec{\nu}} \quad (4)$$

Where $\underline{\vec{\nu}}$ is a noise term representing the stochastic properties of the process and the model errors and $\underline{x}$ means x is stochastic. Using the model $\vec{f}$ a prediction $\hat{\vec{m}} = \vec{f}(\hat{\vec{p}})$ of the measurement $\underline{\vec{m}}$ can be obtained for an estimate $\hat{\vec{p}}$ of the parameter vector $\vec{p}$. An optimal estimate $\hat{\vec{p}}$ for $\vec{p}$ is obtained if for the realisation $\vec{m}$ the residual function:

$$R(\hat{\vec{p}}) = (\vec{m} - \hat{\vec{m}})^T(\vec{m} - \hat{\vec{m}}) = (\vec{m} - \vec{f}(\hat{\vec{p}}))^T(\vec{m} - \vec{f}(\hat{\vec{p}})) \quad (5)$$

is minimised. There is no general direct solution for this problem, because of the non-linearity of $\vec{f}(\hat{\vec{p}})$. An optimum for $R(\hat{\vec{p}})$ is found using the Levenberg-Marquardt method (Gill et al., 1981).

2.5 Verification: towards a consistent scene description

In the verification stage the best hypothesis group is chosen in case of exclusive hypothesis groups and it is determined if a hypothesis group is sufficiently compatible with the images to be accepted.

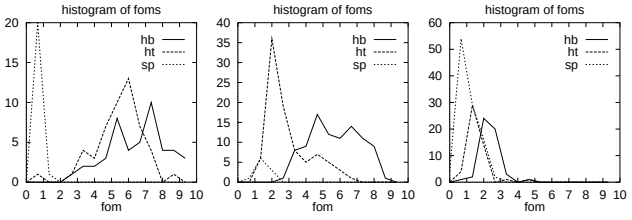

Figure 5: Histograms of the foms for multi-view and a) multi-segmentation and b) single-segmentation and for single-view c) x-200, y-200

A figure of merit (fom) is defined, based upon the mean residual $\overline{r}$ of the estimation, the number of images in which the hypotheses in the hypothesis group are detected n_{views} (which may be lower than the total number of views) and furthermore the number of regions $n_{regions}$ and hypotheses $n_{hypotheses}$ of a hypothesis group. The fom for a certain hypothesis group is defined as:

$$fom = \frac{n_{views} \cdot n_{regions}}{\overline{r} \cdot n_{hypotheses}} \quad (6)$$

A high value for $\overline{r}$ means a large difference between prediction and measurement, therefore the fom should decrease with the mean residual $\overline{r}$.

The reliability of the recognition increases with the number of views and regions involved. In order to avoid the fom to increase if a single object is detected more than once for a single view (e.g. because of symmetry), we have to divide by the number of hypotheses in the hypothesis group. The mean residual $\overline{r}$ is defined by:

$$\overline{r} = \frac{R}{N} \quad (7)$$

with N the number of measurements, and R defined as in eq.(5).

The use of the mean residual instead of the residual itself ensures that the fom is not sensitive to scaling.

In fig.5 the histograms of the foms for multi-view and a) multi-segmentation and b) single-segmentation are depicted. The segmentation parameter for case b is the middle one of the 3 values used in case a. The figures clearly indicate that the multi-segmentation approach offers a much better separation between the detected houses - both bottom and top house - and the spurious detections. The 3^{rd} histogram also shows the problem to separate the correct detections from the spurious detections in the single-view situation.

3 EXPERIMENTS

3.1 Test images

In order to experimentally evaluate the proposed method a set of aerial images is required of an urban area from different viewpoints. Furthermore, the parameters of the camera and the buildings in the scene must be accurately known. To evaluate the accuracy and reliability of the system in a statistical sense, a very large set is required. However, sets of images like this are hard to obtain. An example of a small set of images is described in (Mason et al., 1994). This set contains four images from different viewpoints. In order to obtain a sufficiently large set of images, we generated images based on the images of (Mason et al., 1994). The houses are replaced by ray-traced houses and the textures of the original houses were mapped onto the roofs and the walls. In this way very realistic images of arbitrary views can be generated and the scene parameters are known exactly. Furthermore, an estimate of the image noise in the original recordings was made ($\sigma_{noise} = 3$ [grey levels]) and the original image (that serves as the background) was low pass filtered to suppress the noise. After the ray-tracing process, Gaussian noise with the same standard deviation was added again to the image. In this way different realisations can be generated for a single viewpoint. Figure 6 shows the set of five different views used in the experiments. In two of the five views occlusion occurs, while in the case that the camera looks straight down (fig. 6.e) the walls are invisible. Of this set of five views 100 different realisations were generated. A pixel in the test images measures about 0.15x0.15 [m^2] on the surface and the images represent a surface area of 45x45 [m^2].

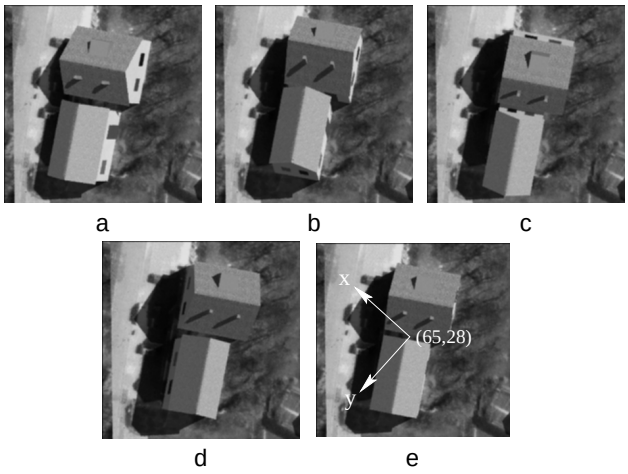

Figure 6: Test set of 5 different views. The respective camera x and y positions were (in [m]): a: (-200,-200), b: (-200,200), c: (200,-200), d: (200,200) and e: (0,0). The z-position of the camera was 700 [m] in all cases. In e) also the correspondence between the image and the world coordinate system is shown. A point (65,28) [m] on the ground in world coordinates is projected onto the centre of the image. The projections of the x and y directions of the world coordinate system are as shown by the white arrows.

The houses shown are described with 6 parameters: position (x, y) orientation (γ), size (width w and length l) and height of the house (h).

3.2 Operation of the multiple view approach

The results of the different stages in the process are shown in figures 7a-e. Fig.7a shows the result after the split and merge step. Fig.7b shows the result after shape based segmentation correction. In fig.7c all the hypotheses generated by the hypothesis generation stage for one realisation for all the 5 images, a total of 19 hypotheses, are projected on one image. Fig.7d shows the hypotheses groups that were formed by the hypothesis corresponding process. The minimum number of views an object must be observed by was set to 2. Six hypotheses groups remain after this stage: only a single hypothesis group for the top house and five competing hypotheses groups for the bottom house. Figure 7e shows the final result after estimation and verification. Since all five hypotheses groups of the bottom house were mutual exclusive, only one remained after verification.

3.3 Robustness and accuracy of single and multiple view approaches

In order to determine the reliability and accuracy of the single and multiple view approaches, the building recognition and estimation process was carried out for both approaches. First the recognition and estimation process was carried out for every realisation of each of the 5 views separately, thus using a single view. This yields 5 sets of 100 recognition and parameter estimation results. Next, the recognition and estimation process was carried out for the multiple view approach, combining the 5 views and again yielding a set of 100 recognition and estimation results.

To evaluate the performance of the multiple and single view approaches, the detections were classified in correct and spurious detections. A spurious detection is a hypothesis of which the parameters deviate too much from the actual

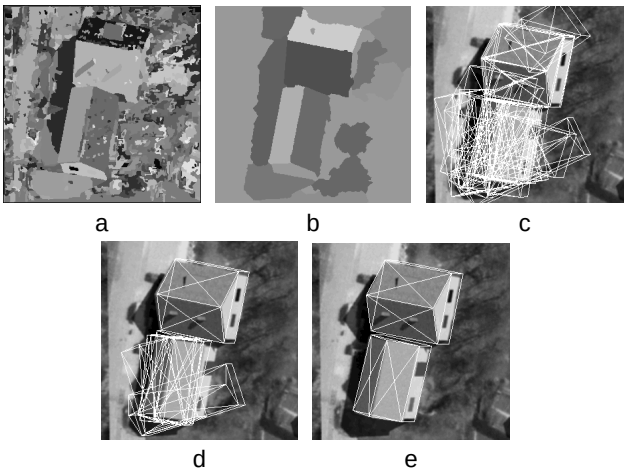

Figure 7: a) result after the split & merge step of the segmentation b) result after shape based segmentation correction; c) hypotheses generated by hypothesis generation step; d) hypothesis groups formed in hypothesis correspondence step; e) final result after estimation and verification.

parameters of the house. Only the position and orientation parameters were taken into account. A hypothesis is considered spurious if:

$$(x - \hat{x})^2 + (y - \hat{y})^2 + 250(\gamma - \hat{\gamma})^2 \geq 10 \qquad (8)$$

This allows a maximum displacement of $\sqrt{10}$ [m] or a maximum orientation error of 0.2 [rad]. Based on occlusion and visibility, a prediction can be made about the expected performance for the different views. E.g. in fig.6a the walls are bright and no occlusion occurs, which should yield good results. The expected performances are summarised in table 1. The detection and spurious rates depend on the choice

viewpoint	view dir.	fig.	comments	hb	ht
(-200,-200)	↖	6.a	no occlusion walls light	+	+
(-200,200)	↗	6.b	hb occludes ht walls dark	+/-	-
(200,-200)	↙	6.c	ht occludes hb walls light	-	+
(200,200)	↘	6.d	no occlusion walls dark	+/-	+/-
(0,0)	•	6.e	no occlusion no walls visible	+/-	+/-
multiple	all			+	+

Table 1: Expected performance for different viewpoints. ht and hb are the top resp. bottom house.

for the threshold for the fom, used in the verification stage to discard unreliable hypotheses. The detection rates and spurious rates were determined for a range of thresholds, see fig.8.

In (Spreeuwers et al., 1997) it was shown that the multi-view approach resulted in a significant higher detection rate and lower spurious rates than any of the single-view approaches. In fig.8 the graphs of the detection rates for the bottom and top house and the spurious rate are shown for the multi-view approach and single segmentation together with the graphs for the multi-segmentation approach. From the graphs it can be seen that the detection rate of the bottom house was already high and doesn't change very much, while the detection rate for the top house is improved significantly. The graphs

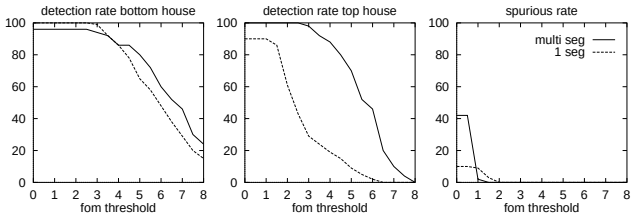

Figure 8: Detection and spurious rates for a range of thresholds for the fom

reflect the expectations from the table well. In the single-segmentation approach, an optimal threshold of 1.8 was found for the multiple view approach. In this case no spurious detections were left and the detections rates for the bottom and top house were 100% resp. 78%. In the multi-segmentation appraoch the detection rates for the bottom and top house were 96% resp. 100%, for a threshold on the fom varying between 2.1 and 2.9.

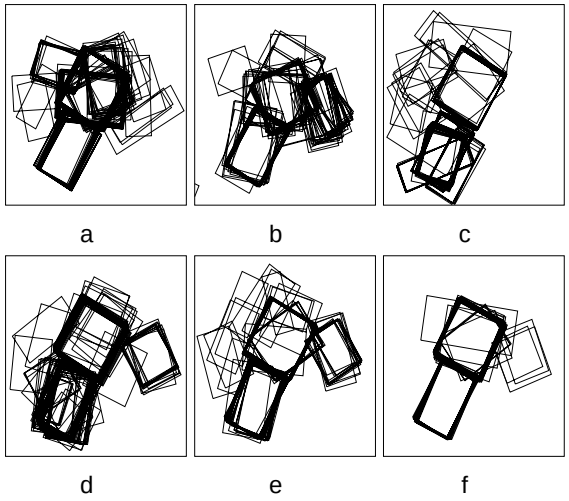

Figure 9: Estimations of ground-planes for single view approaches: a) (-200,-200), b) (-200,200), c) (200,-200), d) (200,200), e) (0,0) and for multi view approach f)

The difference in performance between multiple and single view approaches is also illustrated in figure 9, where the estimated ground-planes of the two houses for all 50 realisations are drawn in a single figure (no threshold on the fom is used here).

The number of detected houses and the spurious detections are shown for a range of fom thresholds in table 2 for the bottom house, and in table 3 for the top house. Both tables show clearly that the choice of the threshold on the fom isn't very critically. This is a great difference with the multi-view/single segmentation approach. The "best" threshold on the fom (no spurious detections!) for this approach is found to be 1.8 exactly. The last column of both tables give the results for this approach. The tables 4 and 5 show the results for the best single-view obtained with the multi-segmentation approach. Both tables show clearly that increasing the fom-threshold decreases not only the spurious detections, but also the correct detections. It should be remarked that for the evaluation we made use of the 'known' parameters. This is not the case in the operational situation, where only a threshold for the fom is available to discard unreliable hypotheses. As can be seen from fig.8, the multi-view/multi-segmentation approach in this case clearly outperforms all single view approaches.

fom threshold	true values	0	1.5	2.0	2.5	1.8
#det.	50	48	48	48	48	100
#sp det.	0	22	1	1	0	
$\hat{x}$	62.0	61.9	61.9	61.9	61.9	62.1
$\sigma_{\hat{x}}$		0.1	0.1	0.1	0.1	0.1
$\hat{y}$	36.0	35.9	35.9	35.9	35.9	35.9
$\sigma_{\hat{y}}$		0.1	0.1	0.1	0.1	0.1
$\hat{\gamma}$	0.53	0.49	0.49	0.49	0.49	0.48
$\sigma_{\hat{\gamma}}$		0.02	0.02	0.02	0.02	0.03
$\hat{w}$	10.5	10.4	10.4	10.4	10.4	10.3
$\sigma_{\hat{w}}$		0.2	0.2	0.2	0.2	0.2
$\hat{l}$	16.6	16.3	16.3	16.3	16.3	16.5
$\sigma_{\hat{l}}$		0.1	0.1	0.1	0.1	0.2
$\hat{h}$	8.8	9.1	9.1	9.1	9.1	8.8
$\sigma_{\hat{h}}$		0.2	0.2	0.2	0.2	0.4

Table 2: Average and standard deviations of estimated parameters of the bottom house for a range of the fom threshold. All parameters are in [m] except γ, which is in [rad].

fom threshold	true values	0	1.5	2.0	2.5	1.8
#det.	50	50	50	50	50	90
#sp det.	0	22	1	1	0	
$\hat{x}$	69.0	68.8	68.9	68.9	68.9	68.9
$\sigma_{\hat{x}}$		0.04	0.04	0.04	0.04	0.1
$\hat{y}$	20.9	20.8	20.8	20.8	20.8	20.9
$\sigma_{\hat{y}}$		0.04	0.04	0.04	0.04	0.3
$\hat{\gamma}$	2.1	2.1	2.1	2.1	2.1	2.1
$\sigma_{\hat{\gamma}}$		0.01	0.01	0.01	0.01	0.03
$\hat{w}$	14.9	15.1	15.1	15.1	15.1	15.3
$\sigma_{\hat{w}}$		0.2	0.2	0.2	0.2	0.4
$\hat{l}$	14.3	14.0	14.0	14.0	14.0	14.1
$\sigma_{\hat{l}}$		0.1	0.1	0.1	0.1	0.2
$\hat{h}$	9.8	9.8	9.8	9.8	9.8	9.3
$\sigma_{\hat{h}}$		0.1	0.1	0.1	0.1	1.0

Table 3: Average and standard deviations of estimated parameters of the top house for a range of the fom threshold. All parameters are in [m] except γ, which is in [rad].

4 CONCLUSIONS AND SUGGESTIONS

A complete system for the recognition of buildings from multiple images is described and experimentally evaluated. The performance of the system was evaluated using a set of 100 artificially generated realisations of 5 images, acquired from different viewpoints, of a scene containing 2 houses. The experiments show that using multiple images drastically improves performance. Detection rates are improved considerably compared to a single view approach. Simultaneously an increase in the accuracy of the estimated parameters is obtained. The improvement in the detection rates is the result of applying a hypothesis corresponding step, which removes incompatible hypotheses, generally caused by segmentation errors. The accuracy of the estimated parameters is increased by using more independent data from multiple images.

ACKNOWLEDGEMENTS

This work was supported by the Foundation for Computer Science in the Netherlands (SION) and the Dutch Organisation for Scientific Research (NWO). The authors wish to thank the Institute for Geodesy and Photogrammetry, Swiss Federal Institute of Technology (ETH) for making available photogrammetric test data (Mason et al., 1994).

fom threshold	true values	0	1.5	2.0	2.5	1.8
#det.	50	50	49	43	17	96
#sp det.	0	105	25	5	2	?
$\bar{\hat{x}}$	62.0	62.3	62.3	62.3	62.5	62.2
$\sigma_{\hat{x}}$		0.3	0.3	0.3	0.5	0.3
$\bar{\hat{y}}$	36	36.2	36.2	36.2	36.4	36.1
$\sigma_{\hat{y}}$	0.3	0.3	0.3	0.3	0.3	0.3
$\bar{\hat{\gamma}}$	0.54	0.51	0.51	0.51	0.51	0.51
$\sigma_{\hat{\gamma}}$		0.03	0.02	0.02	0.02	0.02
$\bar{\hat{w}}$	10.5	10.1	10.1	10.1	10.2	10.1
$\sigma_{\hat{w}}$		0.3	0.3	0.4	0.3	0.5
$\bar{\hat{l}}$	16.6	16.3	16.3	16.3	16.3	16.3
$\sigma_{\hat{l}}$		0.2	0.2	0.2	0.3	0.2
$\bar{\hat{h}}$	8.8	8.3	8.2	8.2	7.7	8.3
$\sigma_{\hat{h}}$		0.9	0.8	0.9	1.2	0.8

Table 4: Average and standard deviations of estimated parameters of the bottom house for a range of the fom threshold. The obtained results are for the best single view (x-200,y-200) and multi-segmentation. All parameters are in [m] except γ, which is in [rad].

fom threshold	true values	0	1.5	2.0	2.5	1.8
#det.	50	46	22	7	2	64
#sp det.	0	105	25	5	2	?
$\bar{\hat{x}}$	69.0	69.3	69.2	69.3	68.9	69.2
$\sigma_{\hat{x}}$		0.3	0.3	0.3	0.2	0.3
$\bar{\hat{y}}$	20.9	21.1	21.0	21.1	20.9	21.1
$\sigma_{\hat{y}}$		0.3	0.3	0.3	0.2	0.2
$\bar{\hat{\gamma}}$	2.11	2.09	2.09	2.10	2.10	2.09
$\sigma_{\hat{\gamma}}$		0.02	0.02	0.01	0.003	0.01
$\bar{\hat{w}}$	14.9	15.2	15.4	15.3	15.9	15.3
$\sigma_{\hat{w}}$		0.6	0.7	0.8	1.5	0.5
$\bar{\hat{l}}$	14.3	13.9	14.0	13.9	13.9	14.0
$\sigma_{\hat{l}}$		0.1	0.1	0.1	0.1	0.1
$\bar{\hat{h}}$	9.8	8.9	9.1	8.9	9.9	9.1
$\sigma_{\hat{h}}$		0.9	0.9	0.9	0.6	0.9

Table 5: Average and standard deviations of estimated parameters of the top house for a range of the fom threshold. The obtained results are for the best single view (x-200,y-200) and multi-segmentation. All parameters are in [m] except γ, which is in [rad].

REFERENCES

Gill, P. E., Murray, W. and Wright, M. H., 1981. Practical Optimization. Academic Press, London. ISBN 0-12-283950-1.

Mason, S., Baltsavias, M. and Stallmann, D., 1994. High precision photometric data set for building reconstruction and terrain modelling. Technical report, Institute for Geodesy and Photogrammetry, Swiss Federal Institute of Technology (ETH) Zurich, Switzerland.

Nazif, A. and Levine, M., 1984. Low level image segmentation: An expert system. IEEE Transactions on Pattern Analysis and Machine Intelligence 6(5), pp. 555–577.

Schutte, K., 1993. Region growing with planar facets. In: Proceedings of The 8th Scandinavian Conference on Image Analysis, Vol. 2, Tromso, pp. 719–725.

Schutte, K., 1994. Knowledge Based Recognition of Man-Made Objects. PhD thesis, University of Twente. ISBN90-9006902-X.

Schutte, K. and Boersema, G., 1993. Hypothesizing a 3-D scene from a segmented aerial photograph. In: Second Conference on Optical 3-D Measurement Techniques, Wichmann, Karlsruhe, Zurich, pp. 452–459.

Schutte, K. and Hilhorst, G., 1993. Comparison levels for iterative estimators for model-based recognition of man-made objects in remote sensing images. In: Proc. IS&T/SPIE 1993 Symposium on Electronic Imaging: Science and Technology, SPIE, Vol. 1904, San Jose, pp. 222–228.

Spreeuwers, L. J., Schutte, K. and Houkes, Z., 1997. A model driven approach to extract buildings from muti-view aerial imagery. In: ISPRS XVII, Commission III, International Archives of Photogrammetry and Remote Sensing, Vol. ???, pp. ??–??