A solution to the correspondence problem in multi-view imagery

L.J. Spreeuwers
Measurement Laboratory
University of Twente
Netherlands

K. Schutte
Pattern Recognition Group
Delft University of Technology
Netherlands

Z. Houkes
Measurement Laboratory
University of Twente
Netherlands

1 Introduction

This paper concerns the analysis of multi-view imagery in order to obtain a scene description and, specifically, the correspondence problem that occurs in this analysis. The required scene description in this case consists of the classes of the objects present in the scene and their parameters: position, size and orientation. The images are aerial photographs and the objects in the images are man-made objects, such as buildings, roads etc. The goal of the complete system is the fully automatic analysis of aerial photographs of urban areas. The output of the system is a scene description that can be used to generate or update a GIS (Geo Information System). Up to now a system has been developed that performs this analysis on a single image [2, 3]. The advantage of using multi-view imagery compared to using single images is that (partly) occluded buildings may still be recognized, because they can be more clearly visible in other images acquired from a different viewpoint. However, the use of multi-view imagery complicates the analysis, because the objects in the different images have to be corresponded to each other. In this paper a solution to this correspondence problem is presented on object hypothesis level. First the image analysis system for single images is described in short. Then the system is extended for multiple images and the method for corresponding object hypotheses is presented. Finally, experiments and conclusions are given.

2 The analysis system for single images

The basic setup of the single image analysis system is shown in figure 1. The segmentation process consists of a segmentation based on region growing, followed by a shape-based segmentation correction process [4], which results in well defined segments. These segments are fed into the hypothesis generation stage [5], which generates hypotheses of the classes of objects (in our case different types of buildings) and a rough estimate of the parameters (size, position, orientation). The hypothesis generation uses so-called aspects [6], or views

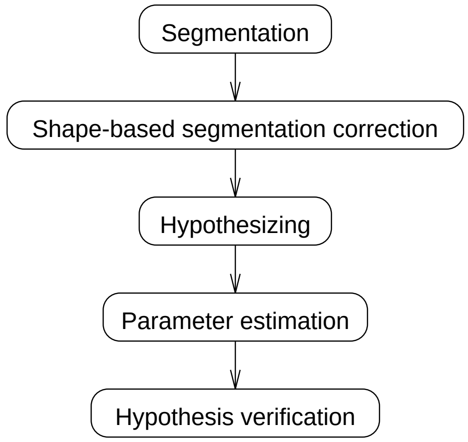


Figure 1: Basic setup of the single image analysis system

of the objects, which are stored in an object model database. It applies a relaxation procedure to obtain the most likely object, i.e. the object of which a certain aspect matches a combination of segments best. The hypotheses are fed to the parameter estimation stage, where, based on the object hypothesis and the rough parameter estimates, the parameters of the objects are estimated [1]. For this estimation procedure an iterative estimator, based on a Gauss-Newton optimisation procedure, is used which optimises the match between the object model and the segmented image by varying the object parameters. The optimisation procedure is performed for each object separately. The final stage is the verification stage, where the residue obtained from the optimisation process is compared with a certain threshold, to determine whether the initial hypothesis is acceptable. The output of the analysis consists of the accepted hypotheses.

3 Extension to multi-view imagery

The analysis on a single image has as a major disadvantage that buildings that are occluded by other buildings cannot be recognised well (depending on how much of a building is occluded, partly occluded buildings can sometimes be recognised though.) By using multiple images, acquired from different viewpoints, this problem can be circumvented. The use of multiple images results in more hypotheses which may refer to the same object. Therefore, a processing stage is required to determine which hypotheses possibly refer to the same object. In our approach, each image is handled separately up to and including the hypothesis generation step in figure 1. After this step, a hypothesis correspondence stage has been introduced to compare the hypotheses found in one image to hypotheses found in the other images, and to find out which of the hypotheses correspond to the same object. There is another problem, related to the correspondence problem. The parameter estimation process can be obscured if in an image buildings occlude each other. The estimation of parameters of buildings that occlude each other in one or more images can be done more accurately if the estimation is performed for these objects simultaneously. The hypothesis corresponding stage provides the proximity information, required to decide for the simultaneous estimation of parameters of more than one object. The parameter estimation process uses the combined hypothesis information and the proximity information to estimate the parameters from the multi-view imagery (i.e. all images are used simultaneously in the estimation process). The setup for the multi-view imagery analysis system is shown in figure 2.

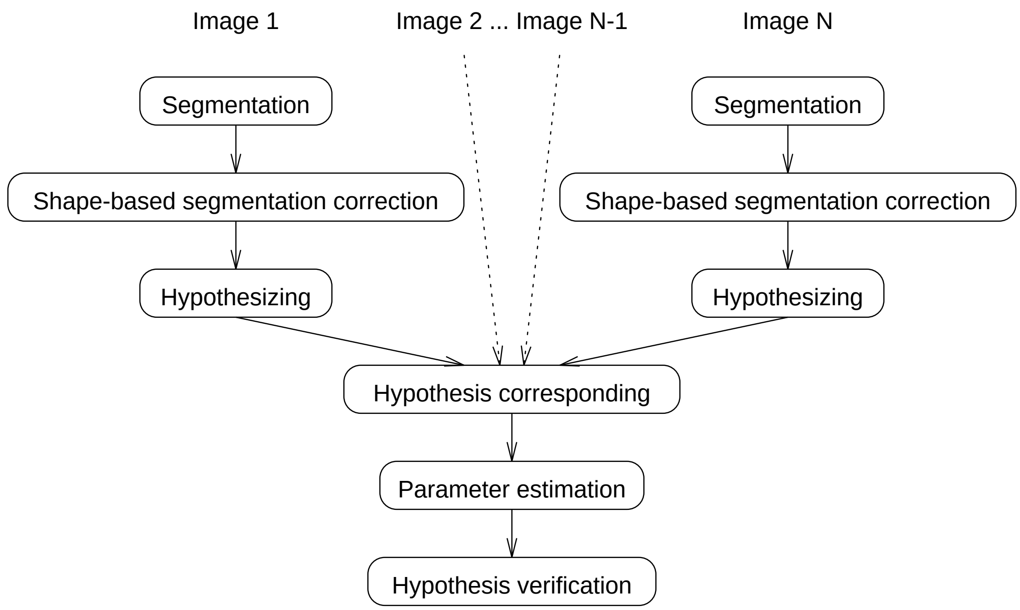


Figure 2: Setup of a multi-view imagery analysis system.

4 The hypothesis corresponding method

4.1 Distance measures and correspondence

The objective of the corresponding stage is threefold:

1. Reduction of the number of hypotheses. All corresponding hypotheses are grouped into a hypothesis group. From this group a new hypothesis is formed that replaces the original corresponding hypotheses.

2. Hypotheses that result in occlusion in one or more images must be marked, so that their parameters can be estimated simultaneously.

3. Different hypotheses (i.e. not corresponding) that occupy the same space should be marked, because they can not be true at the same time. These will be referred to as "mutually exclusive".

After the hypothesis generation stage on single images, for each image there is a list of hypotheses, containing the following information:

1. Object class

2. Initial estimation of position

3. Initial estimation of size
4. Initial estimation of orientation
5. Reliability of the hypothesis

For testing occlusion, a simplification was used, i.e. if the distance between buildings is below a certain threshold, they are marked as "close" and may cause occlusion in one or more images. For mutually exclusive hypotheses, after the estimation process, the most likely hypothesis is selected.

In order to determine if two hypotheses i and j correspond, are "close" or mutually exclusive, three distance measures are defined:

$D(i,j)$ geometrical distance between 2 hypotheses

$O(i,j)$ measure of "overlap", i.e. how much space is shared

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

Correspondence can now be defined as:

$$(O(i,j) \geq Omin) \textbf{ and } (M(i,j) \geq Mmin) \tag{1}$$

i.e. for correspondence there must be a certain minimum of overlap between the hypotheses and the hypotheses must resemble each other enough.

Two hypotheses are marked mutually exclusive if:

$$(O(i,j) \geq \ Omin) \textbf{ and } (M(i,j) < Mmin) \tag{2}$$

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 can possibly cause occlusion if:

$$(D(i,j) < Dmax) \tag{3}$$

In the above definitions 1-3, the constants $Omin$, $Mmin$ and $Dmax$ depend on (among others) the size of the buildings, the flight height and viewing angles.

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

4.2 The hypothesis corresponding algorithm

The hypothesis corresponding algorithm consists of the following steps:

- Put all hypotheses of all images into a list.
- Combine the hypotheses that correspond according to eq.1 into groups. Note that one hypothesis may occur in several hypothesis groups, for example if hypothesis A corresponds to B and C and hypotheses B and C do not correspond, the groups AB and AC are formed).
- Calculate the "average" of the parameters of the hypotheses in the hypothesis groups and assign these to the hypothesis groups. The list of hypothesis groups is now in fact a new list of hypotheses. The "average" could be a weighted average that takes into account the reliability of the individual hypotheses in the group. Currently, an unweighted average is used.
- Determine which pairs of hypothesis groups are mutually exclusive.
- Determine which pairs of hypothesis groups are 'close'.

In the current implementation, the geometrical distance measure $D(i,j)$ is the Euclidean distance in 3-D between the centres of gravity of the two hypotheses (see fig.3.)

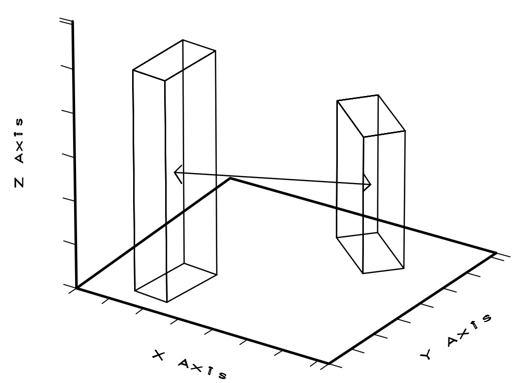


Figure 3: Distance between two hypotheses

For the overlap $O(i,j)$, currently only the overlap in the ground plane is used. This is illustrated in fig.4. The overlap of the ground plane is defined as the ratio between the area of the intersection of the two ground planes and the larger of the two ground planes:

$$O(i,j) = \frac{A_i \wedge A_j}{max(A_i, A_j)} \tag{4}$$

Where A_i and A_j are the areas of the ground planes of the hypotheses i and j. This yields a number between 0 and 1 which is 0 for no overlap at all and 1 if the ground planes of the hypotheses i and j coincide.

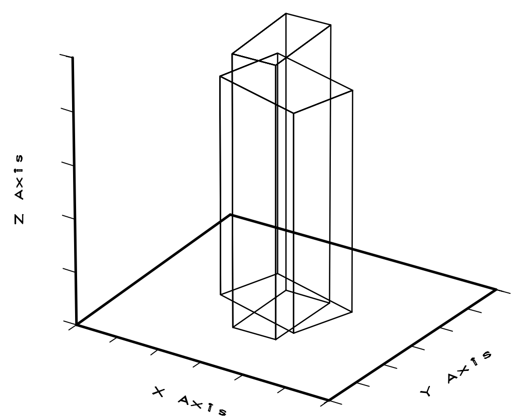


Figure 4: Overlap of two hypotheses

In the feature match $M(i,j)$ the class, size and orientation of the hypotheses are compared.

In the current implementation the class match is 0 for hypotheses of different classes and 1 for equal classes.

The size match is obtained by aligning the hypotheses (i.e. discarding the rotations and the translations) and performing a volume intersection. The size match is defined as the ratio of the shared volume and the joint volume:

$$M_{size}(i,j) = \frac{V_i \wedge V_j}{V_i \vee V_j} \tag{5}$$

Where V_i and V_j are the volumes of the aligned hypotheses i and j. This again yields a number between 0 and 1. The shared and total volumes are illustrated in fig.5.

The orientation match is defined as:

$$M_\gamma(i,j) = 1 - \frac{\mid \gamma_i - \gamma_j \mid}{\pi} \tag{6}$$

Where γ_i and γ_j are the rotation angles of hypotheses around the vertical axis. The range of the orientation match is 0 for a $\frac{\pi}{2}$ radians angle difference to 1 for a 0 or π radians angle difference between the hypotheses (the hypotheses are assumed to be symmetric for the two main axes). In the alignment process symmetry problems like for instance two cubes that differ $\frac{\pi}{2}$ radians in orientation are detected and corrected.

The feature match is constructed by multiplying the class, size and orientation matches:

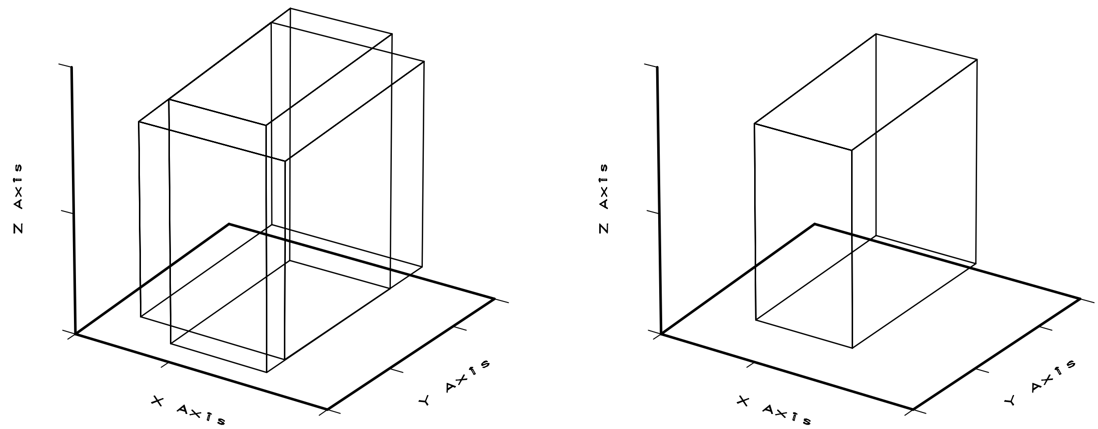


Figure 5: Two aligned hypotheses (left) for size comparison and their common volume (right)

$$M(i,j) = M_{class} * M_{size} * M_{\gamma} \tag{7}$$

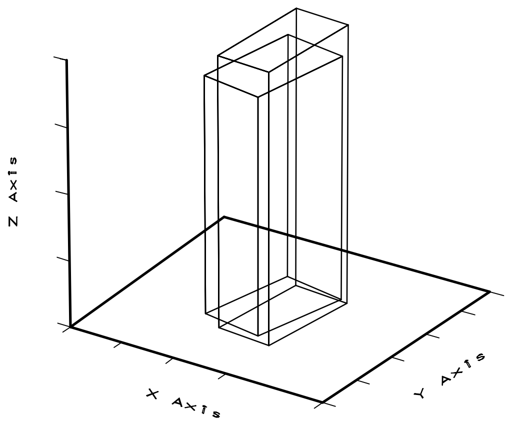
And since the three range from 0 to 1 the combined feature match also ranges from 0 to 1.

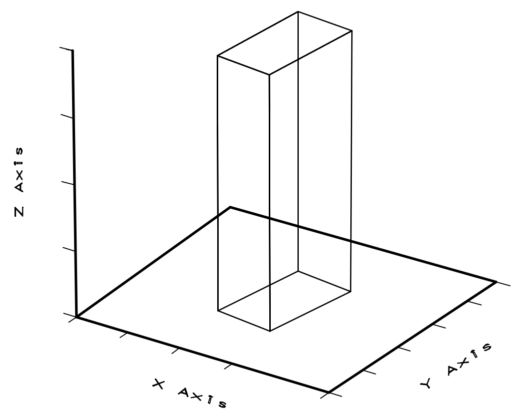
5 Experiments

A number experiments was conducted to evaluate the algorithm and to check whether the output of the hypothesis corresponding algorithm makes sense. The hypothesis lists that form the input of the hypothesis correspondence algorithm were generated by hand. Four characteristic scenes were generated:

- 2 corresponding hypotheses
- 2 exclusive hypotheses
- 2 'close' hypotheses
- a scene with several different corresponding, exclusive and close hypotheses

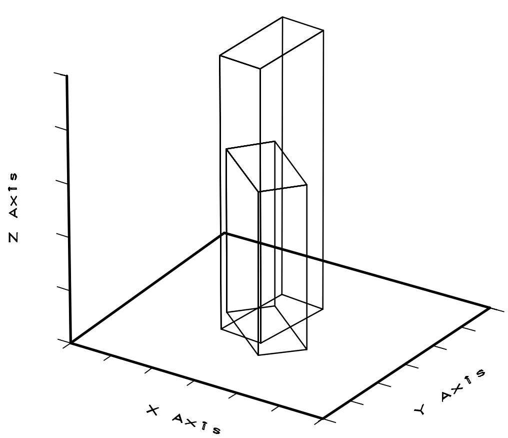
The results of the simulated scenes are shown in the figures 6 to 9.





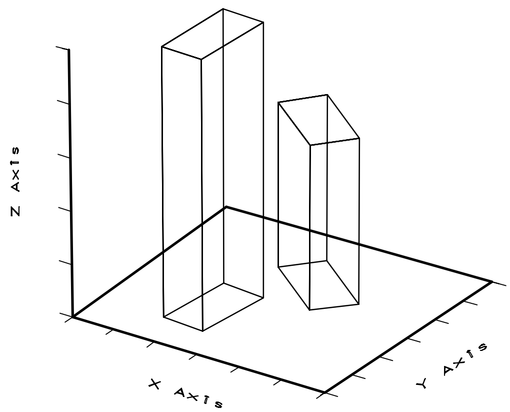
class match	1.0
size match	0.739094
orientation match	0.968169
feature match	0.715568
overlap	0.628244
distance	0.360555

Figure 6: Typical case of two corresponding hypotheses before (left) and after (right) the hypothesis corresponding process and their distance measures



class match	1.0
size match	0.6
orientation match	0.690986
feature match	0.414592
overlap	0.493782
distance	0.538516

Figure 7: Typical case of two exclusive hypotheses and their distance measures



class match	1.0
size match	0.6
orientation match	0.690986
feature match	0.414592
overlap	0.0
distance	2.33238

Figure 8: Typical case of two 'close' hypotheses and their distance measures

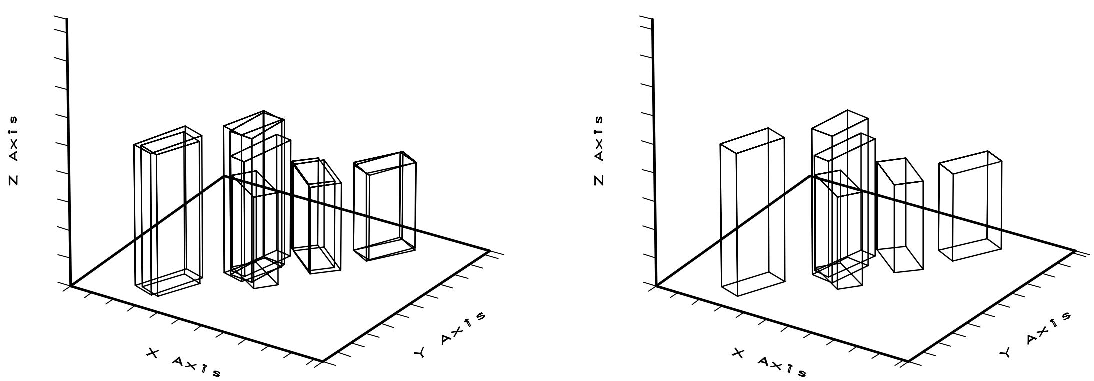


Figure 9: Typical case of multi-view hypotheses (left) and the result after hypothesis corresponding (right). The three hypotheses in the middle are marked mutually exclusive

6 Conclusions

A method for solving the correspondence problem of multi-view imagery on hypothesis level is presented. Three distance measures are defined: feature match, overlap and geometrical distance. These distance measures are used to determine whether hypotheses are corresponding, mutually exclusive or close. The result of the hypothesis corresponding is a list of hypothesis groups. Within each group all hypotheses correspond. Furthermore mutual exclusive pairs of groups and 'close' groups are marked. The operation of the hypothesis corresponding method is demonstrated with a number of typical scenes. The next step is to apply the method on real data. We plan to use a dataset that has been made available on the internet by the Swiss Federal Institute of Technology (ETH) in Zürich, Switzerland.

References

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

[2] N.J. Mulder and K. Schutte, "Knowledge Engineering in R.S. and Knowledge Based Systems in GIS", *International archives of photogrammetry and remote sensing*, Volume XXIX, part B3, pp. 965-968, 1992

[3] Klamer Schutte, "Knowledge Based Recognition of Man-Made Objects", *PhD Thesis*, University of Twente, ISBN 90-9006902-X, 1994

[4] Klamer Schutte, "Region growing with planar facets", *Proceedings of The 8th Scandinavian Conference on Image Analysis, Tromso, May 25-28, 1993*, Volume 2, pp. 719-725, 1993

[5] Klamer Schutte and Gerard Boersema, "Hypothesizing a 3-D scene from a segmented aerial photograph", *Second Conference on Optical 3-D Measurement Techniques, Zurich, October 4-7, 1993*, ISBN 3-87907-254-X, pp. 452-459, Wichmann, Karlsruhe, 1993

[6] Olivier Faugeras, Joe Mundy, Narendra Ahuja, Charles Dyer, Alex Pentland, Ramesh Jain, Katsushi Ikeuchi and Kevin Bowyer, "Why Aspect Graphs Are Not (Yet) Practical for Computer Vision", *CVGIP: Image Understanding*, Vol. 55, no. 2, March, pp. 212-218, 1992