Knowledge Engineering in R.S. and Knowledge Based Systems in GIS

N. J. Mulder
ITC, Enschede, The Netherlands
K. Schutte
University of Twente, The Netherlands

Abstract

In this paper is explained by using the case study 'Knowledge based recognition of man-made objects' how model-based reasoning can be applied to the analysis of Remote Sensing images. The results are used for defining and updating a 3D GIS.

1 Introduction

In this paper a method for incorporating domain knowledge into the image analysis process is presented. The domain knowledge used is present as knowledge about the imaging process and its radiometric and geometric components. By using an iterative estimation algorithm this can be used to solve the inverse of the known imaging process, the image analysis process. An explanation of this method of problem inversion is done in section 2. A case study used as an example will be presented in section 3. Section 3.1 explains how new objects are added to the GIS. In section 3.2 is explained what in the presented case is optimized. Section 3.3 is about the set of used parameters. In section 4 are preliminary results presented, based upon a robot vision experiment. An extrapolation towards Remote Sensing also is presented. Final conclusions can be found in section 5.

2 Problem solving by solving the inverse problem

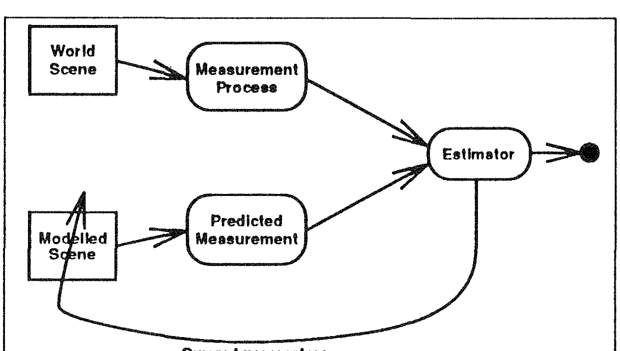

Figure 1: The general estimation process

In this section is explained how by problem inversion a problem can be solved. This technique can be used in situations where the problem A is to be solved, but solving this problem is not easy due to for example redundant and by noise corrupted measurements. When for a given A the inverse problem A^{-1} is well-known and easily solvable this can be used. In an iterative scheme A^{-1} is calculated for a given set of parameters $\vec{\alpha}$. This result Θ is compared to the measurements M. By varying the parameters $\vec{\alpha}$ the best fit between Θ and M is reached. The parameters $\vec{\alpha}$ with the best fit correspond to the best estimate. The general scheme can be seen in figure 1. The *predict measurement* process in this figure is equal to A^{-1}.

The forward modeling defines a reflection between parameters and measurements. In remote sensing the predicted measurements is a 2D projection of a 3D world onto the projection system of the imaging sensor. The forward model must contain parameters to specify the 3D geometry of the scene, as well as the radiometric properties of surfaces and transparent volumes plus an illumination model. The GIS showed contain at least information about 3D objects, their geometry and some radiometric properties. The goal is to update the GIS with RS data. As the update has to occur in a 3D, model inversion from a 2D image to a 3D (model of) reality is needed. Image analysis is essentially model inversion and parameter estimation. Knowledge engineering amounts to model based reasoning. The GIS should to be used together with a Markov state transition model to predict the 3D status and generate hypotheses for the RS image analysis. Classical expert knowledge plays a role in defining context dependent statistics in terms of conditional probabilities.

3 Knowledge based recognition of man-made objects

The knowledge based recognition of objects is based on modeling as mentioned in section 2. Because the modeling of man-made objects such as buildings is relatively easy, man-made objects were selected for a case study.

The problem was studied before [1,2], but the reasoning process was controlled by the data in a bottom-up fashion. A problem with such an approach is that a single detail of the image might correspond to several 3D interpretations. Where in some earlier approaches this problem this problem is solved by using multiple views [3,4], here the problem is attacked by starting with a single view (monoscopic reasoning). So one is forced to check out many pseudo solutions before selecting the most likely one. Many of these pseudo solutions can respond to physically impossible solutions. When it is possible to suppress searching for these solutions the search tree will be greatly reduced. The problem is an inversion of the well-known (graphics) problem of rendering a 3D model into a 2D image. This graphics problem has already been solved for well behaving surfaces.[1] The 3D model as generated is a set of descriptions of 3D objects. These objects consist of primitive objects such as blocks, cylinders etc. Associated with each object is a set of parameters describing the actual shape, position and attitude of the objects. So the solution of the problem consists of identifying which objects do exist and what are the correct parameter values for each object. The strategy followed is:

1. For each object class, find evidence for each class of objects to exist in the image, $P(H_{objectclass}|evidence)$.
2. Make a first estimation for the associated parameters of each object found, $Max_{param.}P(param.|objectclass, evidence)$.
3. Improve the parameters of the object iteratively by rendering the 3D model and comparing the resultant image to the source image.
4. When no real correspondence between the rendered and source image can be found there is the possibility of having searched for the wrong class of object in step 1. Having detected this situation, it is clear that starting at step 1 the procedure has to be repeated with a different hypothesis about *objectclass*.
5. This has to be repeated until the desired level of correspondence has been achieved,

The result of this procedure will be a bag of simple objects. These objects will have to be combined to get the resultant 3D description.

3.1 Hypothesis generation or how to make the initial guess

The initial guess cannot always be the result of reasoning in the 3D model because there doesn't always exist a 3D model (GIS) yet. Fortunately the initial guess has not to be very accurate due to corrections in the following stages.

In the other case of updating an existing 3D GIS, the prior probabilities for class membership of (partial) objects and parameter values are derived from the $GIS(t)$ predicting the status $P(class, parameters)$, $GIS(t+1)$.

In the initial guess two things have to be detected: which object is present and what are its parameters. The detection of which object is present must be the result of some bottom-up low level image processing. Preferably domain knowledge is used to choose between all the possible solutions. When not guessing the first object in a scene it is possible to use domain knowledge to give information about likely clustering of primitive objects (e.g. when encountering a block whose size $\approx$ size of a house then it is likely to encounter a triangular object with approximate equal base size, this rule representing the domain knowledge that many houses have roofs.) The value of the parameters does not have to be accurate. The criterion for the necessary accuracy is that the initial values as starting values in the estimator algorithm (see section 3.2) will lead to converge of the estimation.

In essence this step is all that is done in some other approaches to the same problem [1]. Basic advantages from this approach over the others are:

- The precision achieved in this step does not have to be high, since it is only a first estimate. This generally means that faster algorithms can be applied.
- The penalty of making a wrong initial identification of an object is not too high. Here the only result is a search for parameter values which will not converge sufficiently for that object, which will lead to rejection of the object. Other approaches might end up with a wrong final result.
- Because of the relative ease of hypothesis generation (3D) and evaluation of this stage it is not necessary to apply any algebra here. In other approaches rather complicated algebra on 2D features has to be applied to solve problems like 3D edge interpretation and occlusion [5].

3.2 Parameter estimation

The 3D model as generated is a set of objects which are built up from primitives. The objects (and thus also the primitives) are fixed to a given actual shape and position by a set of parameters. Now let us define a function $\mathcal{E}(S, R(\Theta))$ with S the measurement data and $R(\Theta)$ the predicted data. In the problem described here S is depending on the image (pre) processing performed:

1. The raw image data. In this case $R(\Theta)$ is a rendered image.
2. The segmented input image. Here $R(\Theta)$ is a prediction of the segments.
3. A set of shape descriptors of the found segments. Than $R(\Theta)$ is the prediction of those shape descriptors.

The best choice between these levels is still an open question. The raw data has the advantage of giving access to the full data, in any image processing step information could be lost. Working on segments has the advantage of not introducing errors due to errors in the used illumination model (but a correct segmentation must be present). The main advantage of using a set of shape descriptors is that this might involve a

[1] Rendering has problems with partial reflecting surfaces

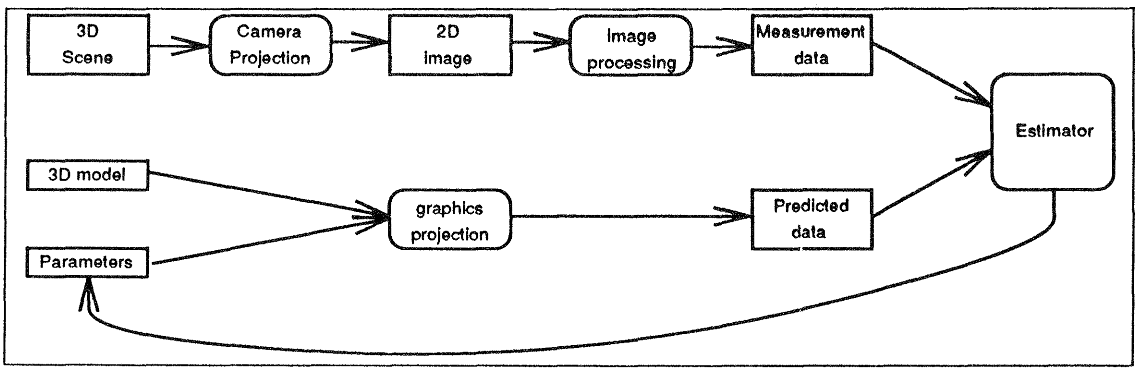

Figure 2: Diagram describing the iteration loop of the parameter estimation, applied to image data.

limited data set, and thus might prove to be an efficient approach. A detailed description about this trade-off can be found in [6]. The parameters Θ are the parameters of the 3D model. Typical model parameters are object position, size and orientation. When the estimation level is at raw image data, such as in item 1 mentioned above, these parameters also include radiometric properties such as diffuse and specular spectral reflection. This function $\mathcal{E}(S, R(\Theta))$ describes the 'difference' between the source and the rendered image in terms of image features. The value of this function will be high when the two images differ a lot; its value will be low when the two images correspond. In this case the problem of estimating the correct parameters is reduced to the minimization of a multivariable function, where the function is $\mathcal{G}(\Theta) = \mathcal{E}(S, R(\Theta))$ and its parameters Θ are the parameters which we want to know for defining or updating a 3D GIS.

3.3 The parameters

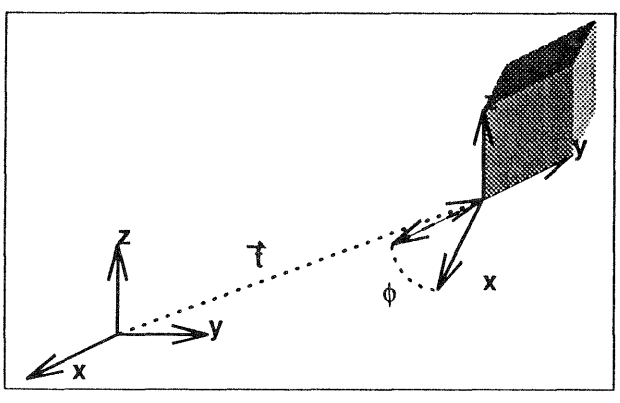

Figure 3: Position and orientation parameters for a body:
a vector $\vec{t}$ which describes the translation relative to the origin and three rotational parameters (only one drawn). The rotational parameter drawn is the only interesting one for buildings.

The parameters which we want to know can be split in two groups: the position and orientation parameters (see figure 3) and the parameters describing properties the object, such as size, shape and the reflectance properties of the surfaces of the object. In the case study diffuse reflectance in a single spectral band is assumed when the estimation level is the raw image data. Adding to this unknown the three position parameters, three size/scale parameters and one orientation parameter gives a total of eight dimensions in parameter space, for each object involved. For other estimation levels the reflectance model is not used, and is parameter thus irrelevant.

3.4 Convergence of the iteration process

The iteration of rendering the image and comparing this with the source image will, given a proper error function, converge to some minimal value for the error function. When this value is low, the rendered image will correspond to the source image. In this case it will be likely that the hypothesized object with its parameter value gives a correct description of the real world. A high error value can be seen as a signal that the scene description is not (yet) a correct description of the real world. Two different reasons can cause this. First it is possible that the basic assumption of the object is correct but the model needs to be refined. This can be something like the need to add additional detail to the description. The other case is that the basic assumption is wrong. In this case a new hypothesis has to be made. An indication for the first case is that only a small part of the rendered image will differ from the source image. In the second case the rendered image is likely to differ over the whole region of interest (R.O.I.) from the source image.

4 Preliminary results

A prototype implementing this approach will be finished during 1992. Some preliminary results already are available. These results are obtained for an experiment in the robot vision field. This experiment is about determination the position and orientation of a single cube. In [6] is demonstrated that the wanted accuracy (about pixel accuracy, for segments with a size of about 25 pixels) are obtained for all estimation levels. Also is shown that convergence is reached in a few iterations (2-10,

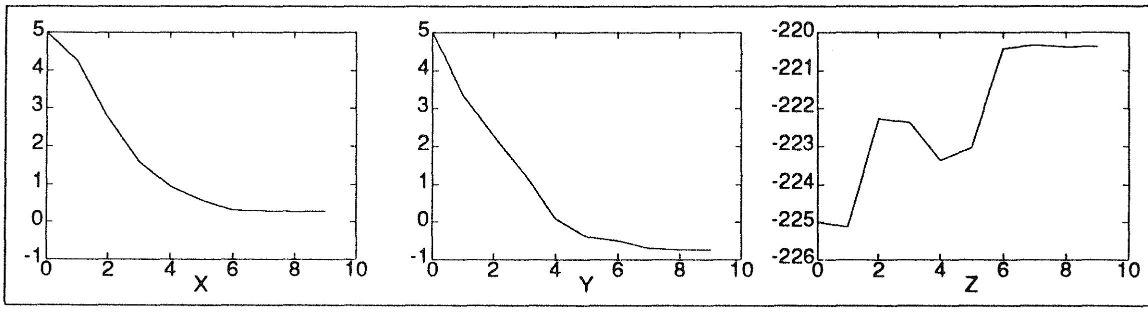

Figure 4: Results of the estimation process. Calibrated values are 0 mm for x and y, -220 mm for z. View size is approximately 5 cm, and camera distance 20 cm.

depending on estimation level). The needed time on a Sun SS2 workstation is in the range of 0.2-12 seconds, depending on the estimation level. The used image size is 64 x 64 pixels. In figure 4 are the results shown for the estimation process. The raw image data comparison level is shown overhere.

Extrapolating these results to the presented case study, convergence of the estimation process is likely in a few seconds, using a normal workstation, when estimating at the level of segmented images. When estimating at raw image data longer run-times are anticipated, because R. S. images generally have a higher resolution, and the calculation of the individual pixel values is the time limiting step. Expected precision of the 3D parameters is at pixel accuracy for the associated 2D shifts.

5 Discussion

The approach to apply model based knowledge engineering to hypotheses generated from a 3D GIS and to use RS data as evidence for the updating of the (likelihoods of) 3D GIS in terms of object classes and parameters has been shown to be a robust one. Like in object oriented software, the method hides "information" about the actual representation of 3D objects. Only the object parameters are accessible (exported) for each instantiation of a class of primitive objects.

Further work is needed for the determination the computational feasibility of the method by applying the method to complex problems like building a GIS of the University of Twente campus.

References

[1] Detecting Building in Aerial Images, A. Huertas and R. Nevatia.
Computer vision, graphics, and image processing 41, 131-152 (1988)

[2] Towards Automatic Cartographic Feature Extraction, David M. McKeown, Jr.
Mapping and Spatial Modeling for Navigation, NATO ASI Series, Vol F 65, Springer-Verlag Berlin Heidelberg 1990

[3] Model Construction and Shape Recognition from Occluding Contours, Chiun-Hong Chien and J.K. Aggarwal.
IEEE transactions on pattern analysis and machine intelligence, vol11, no 4, April 1989

[4] Using Perceptual Organization to extract 3-D Structures, Rakesh Mohan and Ramakant Nevatia.
IEEE transactions on pattern analysis and machine intelligence, vol11, no 11, November 1989

[5] A Spatial Structure Theory in Machine Vision and its Application to Structural and Textural Analysis of Remotely Sensed Images, He Ping Pan.
Ph.D. Thesis, Enschede 1990, ISBN 90-9003757-8

[6] Evaluation of comparison levels in iterative estimators, A.J. de Graaf, K. Schutte.
Proceedings ICPR11, August 30 - September 3, 1992, The Hague (to be published).