A Multi-Layer Strategy for 3D Building Acquisition*

A. Brunn, E. Gülch, F. Lang, W. Förstner

Institut für Photogrammetrie
Universität Bonn

Abstract: *In various projects we investigate on the extraction of buildings on different type and representation of data. This paper presents a strategy for 3D building acquisition which combines different approaches based on different levels of description. The approach consists of detection of regions of interest and automatic and semiautomatic reconstruction of object parts and complete buildings. We incorporate the approach in a global concept of interaction between scene and sensors for image interpretation.*

1 Introduction

Cities are the living place for more than 50% of the world population. Urban management requires up to date information one of which refers to buildings. Current maps mostly contain only the groundplan of buildings together with reference to various registers. 3D information on buildings appears to be urgently necessary for all types of the planning referring to sound emissions, air pollution, microclimatology or transmitter planning. Acquiring 3D information on buildings still is costly, only automatic or at least semi-automatic methods appear feasible in the long run.

This paper discusses strategies for 3D building acquisition from various sensor data, mainly, however, from aerial images. Building acquisition has to face a number of problems, which altogether appear as a challenge for automatic data interpretation. Some of these problems are:

* This research is supported in part by BMBF/ DLR under Grant 01 M 3018 F/6, by BMBF/ DARA GmbH under Grant 50 TT 9536, and by DFG under Grant 180-1/2. The views and conclusions in this document are those of the authors and should not be interpreted as representing the official policies, expressed or implied of those agencies.

- The *notion* "building" has many facets, economical, legal, architectural, social, technical ones. Sensor data only provide access to the geometry and the physics of a building in principle, allowing to derive the 3D scope and infer some of the semantics of its parts. Even with this restriction there is no commonly accepted and enough formalized model of a building which can be used in automatic interpretation. In [4] we argued that a volumetric representation, as e.g. constructive solid geometry (CSG), may be the right level of representation in order to facilitate a link to the meaning of the individual parts. As not all 3D shapes represent buildings constraints on sizes and regularities need to be incorporated in a building model as well as constraints on material, which implies appearance in the different sensor-data. The large variety of buildings up to now prevents automatic systems to yield successful data interpretation on larger data sets.

- There is various *sensor* yielded information on buildings. Images, especially aerial images, certainly are a main data source. Laser scanners are on the threshold of becoming an economical alternative if only geometry is of interest. But also maps on geoinformation systems can be regarded as sensors providing useful information, especially due to their links to all types of registers and the corresponding semantics provided this way.

  It is, however, by no means clear, which of these sensors has which use in building acquisition. This not only results from the different modes of the sensors, but much more from the great variety of scales or resolutions in which these sensor data are available. This has conceptual consequences on which aggregation level a building actually can be observed, is thus intimately linked with the complex aggregation hierarchy of buildings and their parts.

- *Sensor data* are projections, thus only show a specific portion of the scene. This projection leads to various defects, starting with the missing depth in images or GIS, occlusions both in images and laser data, local deformations of the 3D shapes, even if they are partially planar. Neighborhood relations, i.e. topology, usually are preserved, other invariants may be used to characterize the appearance, e.g. when describing the form of roof parts or the regularity of facades. The problem of automatically transferring 3D information about buildings to 2D-constraints has not yet been solved in a general manner, only specific solutions are knows, e.g. on invariants. The same holds for the physical appearance of buildings and their parts, be it in B/W or in color images.

- Finally, strategies for inverting the imaging process, inverting the projection are lacking. Many tools are available, such as feature extraction, segmentation, matching, grouping. All need to be adapted or evaluated with respect to their use in building extraction. However, this does not solve the problem really, as the impact of each of these tools on the solution remains unclear. General tools such as optimization, constraint satisfaction or heuristic search only provide a shell which needs to be filled.

Summarizing, acquiring 3D buildings from sensor data appears to be a challenge for automatic interpretation systems.

The goal of this paper is to discuss the aspect of processing within automatic interpretation more in depth, clarifying the role of algorithms as a link between sensors and scene and in this way earning the description of certain strategies. These examples of algorithms working on different layers within the analysis want to demonstrate the feasibility of the setup, but also the large number of yet unsolved problems.

2 Layers, models and strategy

2.1 Levels of description for data and models

We assume image analysis to be a task driven process to derive an application dependent description of the scene based on available sensor data. Both, scene and sensor data have a complex structure. Scenes are composed of objects, giving rise to a containment hierarchy for composite objects, to a specialization hierarchy for groups of objects and associations between objects for describing their mutual relations. Sensor data are complex, as we treat any type of derived sensor information also as sensor data, thus subsuming intelligent sensor systems under the notion sensor.[1] For discussing image analysis tasks we unify scene and sensor model as both can contain known and unknown parts. This allows to describe tasks, like detection, reconstruction, localization and interpretation on the one hand, and orientation and calibration on the other hand, in a uniform manner (cf. [7]).

Fig. 1 shows the principle of solving a specific image analysis task. An image analysis algorithm provides a link between a scene and a sensor. The scene provides the algorithm with the necessary and available knowledge, e. g. the class of the object to be detected, the structure of the object to be reconstructed, the representation in which the location is to be given, neighboring objects already detected together with the expected neighborhood relations. The sensor provides the algorithm with an iconic or symbolic description of the image on the level of aggregation adequate for the algorithm, and possibly information on its orientation, or other characteristics. The result of the algorithm generally is both, the desired completion of the scene description as well as of the sensor description.

We now want to break down scene and sensor models. In the most simple case the scene may consist of objects of different type. Each of it may be described by its containment and aggregation hierarchy. In fig. 2 the scene related task, the type of the unknown object and the internal structure are indicated, where task and type are lists, whereas the internal structure is a tree or

[1] At this point we integrate the model for the physical sensor and the derived 2D or 3D image, thus subsuming the image model under the sensor model.

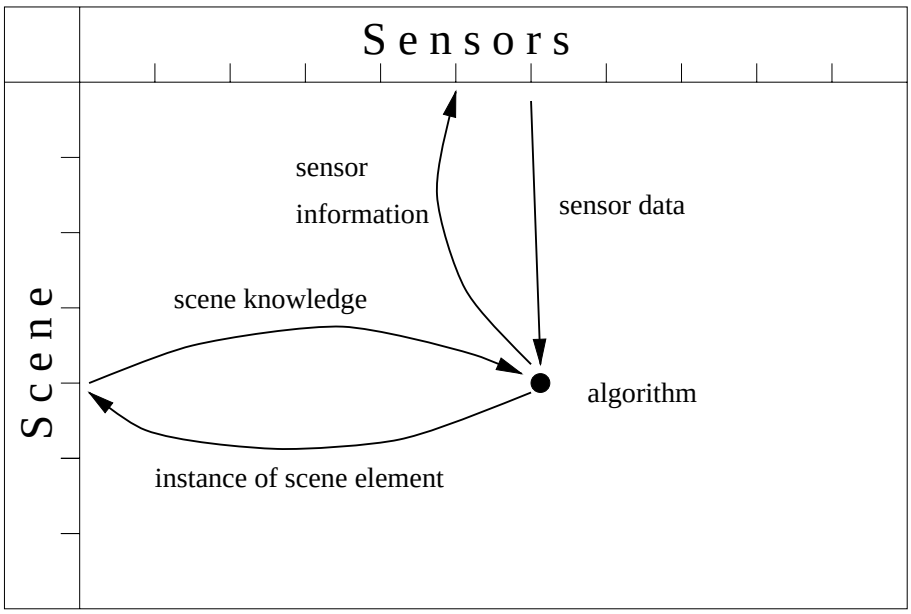

Figure 1: An algorithm links sensor and scene information.

a more complicated representation structure. Any task, type of object and internal structure may be combined leading to a specific scene related task with specific preknowledge fixing a point on the vertical axis. The task specifies which of the objects or structures are unknown. The situation is a bit more complicated for the sensors. Besides the list of the tasks, the list of the physical sensors and the internal structure of the possibly aggregated sensor information we need to distinguish between 2D and 3D, or even 4D, information derived from sensor data of lower dimensionality by matching or tracking algorithms. Also here, any combination of task, aggregation level, dimension or typ of physical sensor may lead to a specific sensor related task with specific sensor information fixing a point on the horizontal axis. The extremely high number of possible image analysis algorithms results from the pure fact, that in principle all combinations between scene and sensor related tasks, types, structure or dimensionality may be reasonable. The rectangle spanned by scenes and sensors thus qualitatively represents the space of possible vision algorithms. E. g. assuming four scene and two sensor related tasks, three types of scene objects with two components each, two types of sensors, two dimensionalities and three aggregation levels we already have to think of 576 different algorithms. This assumes *single* algorithms for each task, and does not take into account multiple goals or sensor fusion techniques. Of course not all algorithms may be meaningful. Therefore strategies for selecting appropriate algorithms and especially sequences of algorithms reveal to be necessary.

Bottom-up or data driven and top-down or goal driven processes generally need to be intimately linked. Aggregation, grouping or hypothesis generation tasks which necessarily need to be initiated and controled by high level knowledge, task or scene specific. E. g. the extraction of straight edge segments, a typical aggregation process, is based on the very generic knowledge the scene to contain man-made object. Such generic knowledge therefore needs to be encoded

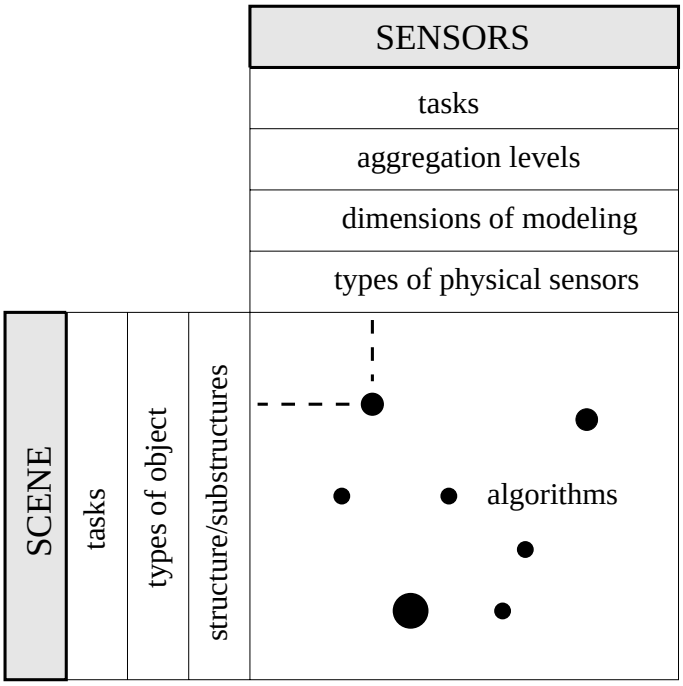

Figure 2: Tasks and structure of scene and sensor information. The points symbolize some algorithms. The different sizes indicate different gain and usability of the combinations.

in the scene model. On the other hand, the result of an algorithm may yield the hypothesis about a blob to represent a certain object part, e. g. a trapezoidal roof. This may initiate further hypothesis generation, e. g. neighboring blobs to be of the same type, as trapezoidal roofs unlikely appear alone. This may then trigger resegmentation.

The problem of control therefore has to solve several subtasks, which easily can be visualized in the diagram (cf. fig. 3)

1. Choose the best sensor for a given task, e. g. when choosing the level of aggregation for deriving 3D-information from 2D images.
2. Choose the best sensor combination for a given task. This is a problem of sensor fusion, e. g. when deciding which combination of gray level, color images and range images is best for building detection.
3. Choose the best hypothesis for a substructure of the scene for a given sensor. This is sensor driven hypothesis generation, e. g. when deciding which class of the possible image features is best to start reasoning with, or which of the given vertices is best for continuing search.
4. Choose the best hypothesis for a substructure of a scene for a given scene interpretation. This is internal reasoning. Examples are e. g. perceptual grouping, without going back into the image data, or focusing on relevant details before elaborate processing is initiated.

5. Decide on the completion of the task or detect failures, and give control to the operator. This is an algorithm internal decision when referring to a complex algorithm solving a highlevel task, i.e. located at the top level of the scene aggregation tree.

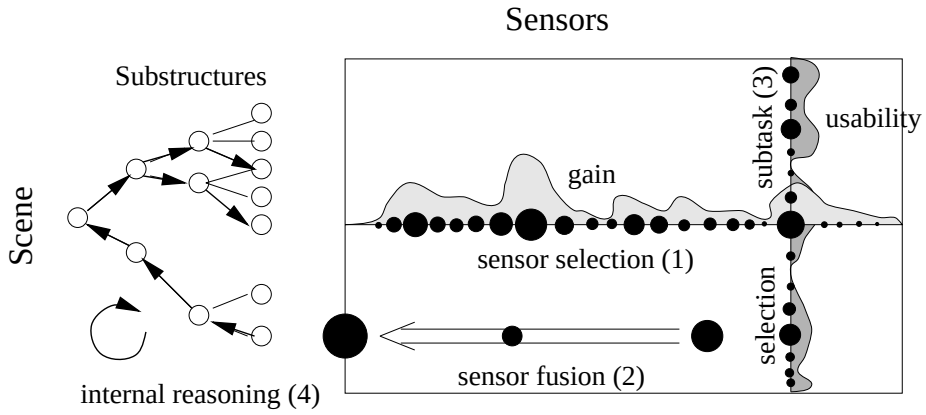

Figure 3: Different control schemes between scene and sensors.

The proposed scheme can be interpreted as a structured blackboard having different layers:

a) the layered structure of the scene referring to different levels of aggregation or abstraction

b) the layered structure of the sensors referring to different levels of aggregation

c) the layered structure of the algorithms referring to the different layers of aggregation

Obviously any vision algorithm which is meant to be used by the system needs documentation with respect to its use, covering at least the following aspects

- type of sensor data, specifying aggregation level dimension and characteristics of underlying physical sensor. This holds for both, given and unknown information
- type of scene information, specifying type of object and its internal structure, again for both, given and unknown information.

  This input–output description allows a formal check on the usability of an algorithm. It implicitely requires modeling of sensor and scene data at a level which is available by some algorithms. Thus scene and sensor modeling can not be performed without explicit reference to an algorithm linking both.

- performance characterization, specifying the expected performance for given input

  This is necessary to be able to compare different algorithms, sensors or hypotheses with respect to their usefulness. It implicitely requires a common language for describing performance. Performance may relate to precision, detection rate, likelihood of failure, computing time, memory resources etc. Documentation of performance of algorithms allow to use tools for optimization or constraint satisfaction for task solving, solely based on

the description of the algorithms and the (currently) available scene and sensor knowledge (cf. [20]).

This general scheme needs to be realized in steps. The problem of building acquisition seems to be of a complex enough nature to test the validity of the setup. We therefore will specialize and give some examples in the following.

2.2 Selected strategy for building acquisition

We describe a special strategy for 3D building acquisition that is based on our experiences with the applicability of different data sources and different algorithms for this specific task. Because reconstruction using aerial image data can not be solved in one step we split the task of 3D building acquisition into different subtasks using different types of sensor information. The subtasks are

- detection of regions of interest,
- 3D reconstruction by object parts, and
- 3D reconstruction of complete buildings.

This corresponds to a multiple usage of different paths in fig. 2. The output of one subtask will be the input to the next subtask. Up to now the strategy is fixed in order to exploit their usefulness. Other strategies need to be evaluated in order to be able to adaptively choose an optimal sequence depending on sensor data, quality of algorithm and specific task to be solved.

In the first step we use pixel based 3D sensor data from automatic generated Digital Elevation Models or laser scanners for extracting those regions which contain the desired objects (buildings) with a high probability. In this subtask buildings are searched in a high generalization level represented by low level pixel attributes (e.g. height). For these extracted building segments approximate structures of buildings can be derived depending on the resolution of the available 3D sensor data. The resulting segments of the elevation model are used as input for the following 3D reconstruction by object parts. This second step additionally uses a feature extraction of local structures in multiple monochrome or multichannel images. The automatically derived 3D structures are the input for the 3D reconstruction of complete buildings. This last step can be formulated on a lower aggregation level because of the strong impact of the integrated human operator even with only one stereo pair. During the third step prior results can be verified, corrected and improved. This step will be required for quite some time as a fully automatic acquisation of complex buildings can not be achieved due to the high variability of buildings. The amount of human interaction depends on the required detail of acquisation, the complexity of the buildings, and the availability and quality of sensor data. Partial solution neglecting one or more steps of our proposed procedure may be sufficient for special

requirements.

The algorithms of these three subtasks are described in the next chapters.

3 Detection

It is important for the process of reconstruction of 3D structures to reduce the vast amount of data provided by a couple of aerial images – working on grey value images means working on about 250 Mbyte data ro more for each image– and the result of the feature extraction (cf. sect. 4). Both methods described in the following sections need a priori focusing on the region of interest to reduce the search space. We apply a hierachical focusing techniques for grid based data. Additional to the presentation of a realization for focusing in this chapter we want to show that statistical influence diagrams in a specialization of bayesian networks can be used for low level, pixel based image interpretation.

There are already several focusing techniques presented in the literature of image processing and interpretation each using special data sources. In contrast to those methods provided by Ackermann et al. ([1]) using image pyramids and Ravela et al. ([25]) using a scale space algorithm we follow a strategy based on bayesian networks to find regions of interest (ROI). Statistical methods support the fusion of different data sources and integrate uncertainty and errorness of observed data. Especially bayesian networks supply a close connection to pyramidal image interpretation. Terzopoulos (cf. [29]) showed that hierachical approaches for segmentation tasks reduce the complexity of search. We just want to mention the approach of Bouman et al. ([3]) who used a multigrid relaxation method. Our algorithm differs because we use a simpler, straight forward propagation to reduce the computational effort.

Our method consists of two steps, the generation of the feature pyramid and the goal oriented statistical evaluation. In the next subsection we will describe the algorithm, and then we compare it with the bayesian approach. Some extensions will also be described.

3.1 General strategy

The step of focusing consists of two parts: generating a feature pyramid and evaluating the branches of the pyramid. Let a feature vector field $\boldsymbol{f}(r,c), r \in \{0,\ldots R-1\}, c \in \{0,\ldots C-1\}$ each of n observations be given[2]. Thus each grid point of this vector field should carry information about its interest for a specific task coded in the assigned feature vector. During the preprocessing we build up an image pyramid using techniques from multichannel image

[2]Vectors are printed in fat and non capital, matrixes in fat and capital letters. Stochastic variables are underscored.

processing (cf. e. g. [19]). Let $\boldsymbol{f}_0(r,c) := \boldsymbol{f}(r,c)$ then

$$\boldsymbol{f}_{l+1}(r,c) = \boldsymbol{M} \odot \boldsymbol{f}_l(2r,2c) \quad \text{with} \quad l \in \{0,\ldots,L-1\} \tag{1}$$

holds when L is the number of layers. The amount of grid points decreases by the factor $\frac{1}{4}$ in each reduction step. $\boldsymbol{M}$ is an information preserving multidimensional filter. We tested several morphological and statistical filters. Which filter should be applied is very problem dependent according to the type and amount of noise in the data (some experiences with different filters are described in the following).

After creating the feature pyramid we start the evaluation of the feature vectors. We use every feature vector to determine if that grid position is of interest for the task. For this purpose we treat the feature vectors as observations of a measuring process and assume that – due to the central limit theorem– they are normally distributed. Because it is not possible to estimate any statistical parameter from one observation we take the observation representative for the mean value and use a covariance matrix given a priori.

$$\underline{\boldsymbol{f}}(r,c) \sim \mathcal{N}(\boldsymbol{f}(r,c), \boldsymbol{C}) \tag{2}$$

At this point we start to build a dynamical generated pyramid of random variables $v_l(r,c)$ that can hold the two statements "The grid point is member of a region of interest" and "The grid point is not member of a region of interest" for each pixel in each layer. The **Pr**obability that a grid point is member of a region **O**f **I**nterest (**PrOI**) will be denoted with $p(v(r,c)=t)$ where 't' means true, $p(v(r,c)=f)$ resp. It is calculated by the probability that a feature vector belongs to a region of interest Θ in the feature space a priori defined.

$$p(v(r,c)=t|\Theta) = \frac{1}{\sqrt{2\pi}\det \boldsymbol{C}} \int_{\Theta} e^{-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{f}(r,c))\boldsymbol{C}^{-1}(\boldsymbol{\theta}-\boldsymbol{f}(r,c))} \quad d\boldsymbol{\theta} \tag{3}$$

These boundary information of Θ can be obtained by estimation from a training set. We use a multidimensional open box with the edges $\boldsymbol{f}_b = f_{bi}, i \in \{0\ldots n-1\}$ and infinity as subspace Θ. That means that high values of the feature vector are more interesting than low ones. Therefore the integral could be written as

$$p(v(r,c)=t|\boldsymbol{f}_b) = \frac{1}{\sqrt{2\pi}\det C} \int_{f_{b0}}^{\infty} \ldots \int_{f_{bn-1}}^{\infty} \frac{1}{\sqrt{2\pi}\det \boldsymbol{C}} e^{-\frac{1}{2}(\boldsymbol{f}_l(r,c)-\boldsymbol{\theta})^T \boldsymbol{C}^{-1}(\boldsymbol{f}_l(r,c)-\boldsymbol{\theta})} \quad d\boldsymbol{\theta} \tag{4}$$

If the observations in the feature vector are uncorrelated the integral can be calculated as a product from the normal distributions of each observation. The probability of the complementary outcome is calculated by

$$p(v(r,c)=f) = 1 - p(v(r,c)=t) \tag{5}$$

Starting from the highest pyramid level we perform a depth-first search to obtain the ROI. We evaluate all grid points on the L-th level and sort them in descending order.

$$p(v(r,c)_{max}=t), \quad \ldots, \quad p(v(r,c)_{min}=t) \tag{6}$$

In the next level we start with those pixels that during the pyramid generation had been combined to the most promising element of the highest level. In all levels, except the highest one, we use both information, the actual information at the current level and the information aggregated in the next higher, i.e. the previous level, taking the conditional probabilities into account (cf. fig. 4). We neglect the correlations between the two sources due to the feature

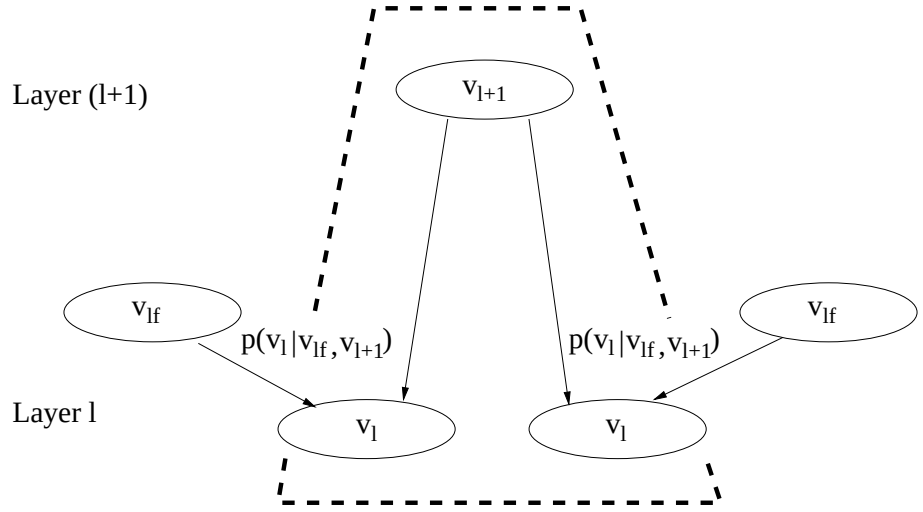

Figure 4: Propagation of the probabilities of interest in the pyramid.

pyramid generation. Therefore we have to distinguish between the PrOI on the information only provided by the feature vector $p(v_{lf}(r,c))$ and the total, a posteriori PrOI $p(v_l)$ at each level of the pyramid. In this refined notation holds

$$p(v_l(r,c)=t) \quad \propto \sum_{v_{lf}(r,c), v_{l+1}([r/2],[c/2])} p(v_l(r,c)|v_{l+1}([r/2],[c/2]), v_{lf}(r,c)) \\ p(v_{l+1}([r/2],[c/2]))p(v_{lf}(r,c)) \qquad (7)$$

when $[x]$ means the highest integer that is smaller or equal x. We do not introduce special knowledge in the conditional probabilities: when both information sources support each other we strengthen that content otherwise we delay the decision.

This procedure is done for each level of the pyramid up to a given pixel resolution depending on the task. The pixel extracted by this depth-first search defines a first element of the ROI and can be used as a focus for the next step of interpretation (cf. sec. 4). By backtracking in the pyramid other elements of ROIs are found. Up to this moment we have no experiences where to stop backtracking because classification against a ROI can not be done at higher levels. Fine structures might appear in lower levels in dependence of the filter generating the feature pyramid. Only when using a maximum filter this decision can be done on a higher level assuming the noise to be negligible.

This algorithm for focusing onto the relevant parts of data, in particular the evaluation step, can be interpreted as a dynamical bayesian network. Each pixel with its probability of interest

represents a random variable in the bayesian network and the conditional probabilities are associated with the edges of the graph[3]. This configuration makes two possible extensions obvious: on the one hand grid points that are far apart should influence one another by incorporating more model knowledge about neighborhood relations between ROI for that task. Furthermore, if one region is a priori known it should influence the PrOI of other, related regions. Thus solving data restoration and data interpretation in a common algorithm becomes possible. That would mean bottom-up propagation in the sense of bayesian networks. Therefore the network in left part of the fig. 3 could be moved into the interior of the rectangle, changing from a control network to an interpretation network (cf. [2]). So bayesian networks seem to be a feasible tool for hierarchical image interpretation, not only for grouping as [26] and [13] showed. A extension of the detection algorithm on irregular pyramids (cf. [22] and [14]) will be done in future. Thus the geometrically blind pyramid generation can be improved because this modification provides the building of task dependent neighborhoods.

3.2 Using Digital Elevation Models (DEMs) for focusing and getting approximate building structures

The aggregation of the dataset focusing is based should be very low cost or fully automatically to enlarge its advantage. So we use digital terrain models for focusing either directly scanned with an airborne laser scanner or automatically generated DEMs based on correlation or matching techniques.

Change detection: For demonstration the usefulness of the approach we took two DEM from an area in Alter-Oedekoven. Both were derived by image matching. One dated back to 1982 the other to 1993. The ground resolution was $1m^2$ (cf. fig. 5). The images[4] with an image scale 1:12000 do not allow a generation with a finer grid with more surface information. We use the height difference as the feature for determining the regions of interest to update a map from the earlier datum. We used the height variance $\sigma_h = 2m$ for the statistical distribution and we define a ROI in feature space if the difference in height differs by $4m$ from zero which is motivated from the usual height of a building storey. For the purpose of visualization we show the probabilities of all pixels of the presented layers. Because of the coarse DEM grid it is not possible to stop at a higher level than the lowest one. In fig. 6 the result of focusing is shown. Dark regions are classified as regions of change. The result shows some blurring due to effects from the automatic generation of the DEM and to changed vegetation in eleven years. So further data have to be integrated, especially colour. In contrast to a simple thresholding of the difference DEM this method reduces random noise of the difference DEM because of the used neighborhood information from previous layers. Thus it reduces splitting of segments.

[3]More basics of bayesian networks can be found in [23]

[4]The images were kindly provided by the LVA NRW (Bonn, Bad Godesberg).

Figure 5: a) DEM of 1982; b) DEM of 1993; c) difference DEM; For visualization the three DEMs are spread on 256 grey values. (Bright/Dark pixel mean large/low height.)

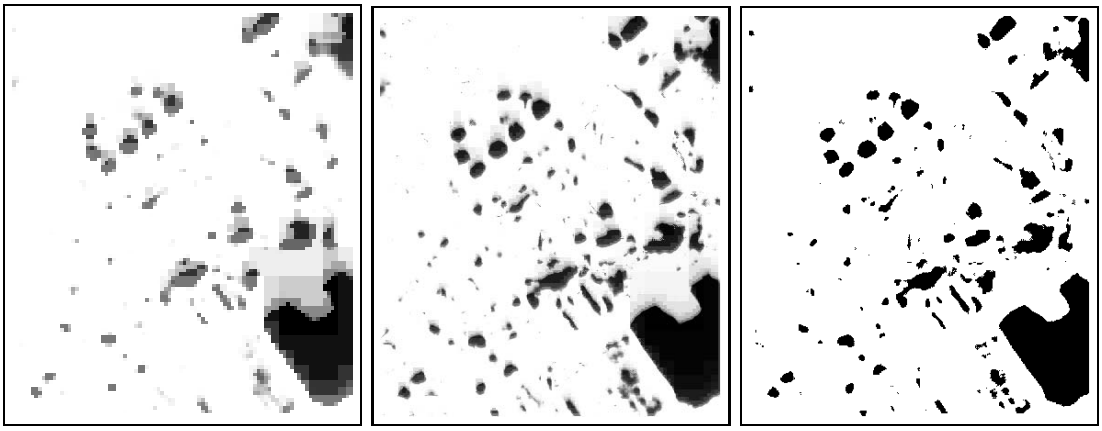

Figure 6: a) Third layer and b) lowest layer of the probability pyramid. Dark means high probability for changes. From layer to layer downwards the segments get more specialized. Large segments in the lowerest layer can already be recognized in higher layers; c) Result of classification in the lowest layer which equals the result of a binarization with a probability treshold of 0.5. Small disturbances due to changed vegetation and due to the automatic DEM generating process caused by the high variability of the surface exist in the resulting classification image.

Furthermore the depth-first search leads the following 3D acquisition to the ROI in descending order of interest; dominant buildings can be reconstructed at first. At the moment those extracted regions are classified by the building extraction technique presented in [30]. In future we want to improve the detection in context of building extraction by extending the feature vector and introducing color and texture observations in the focusing step.

Approximation of building structures: If dense DEM information is available the shape of complex buildings can be extracted already from the DEM. Therefore we segment roof planes and want to incorporate knowledge as e. g. symmetries, coplanarities, collinearities and parallelities by MDL-techniques (cf. [31] and [5]). The expressiveness of a dense DEM can be

seen in fig. 7[5]. The extracted 3D information about planes and vertices will be used as a priori information for the next step of interpretation, the 3D reconstruction.

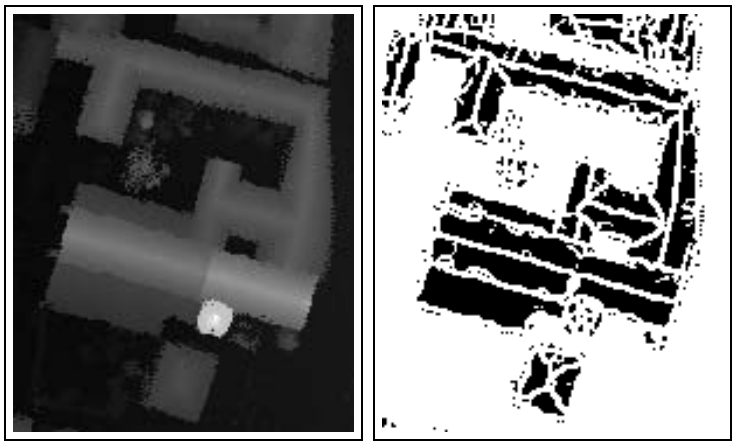

Figure 7: a) Laser scanning data from Ravensburg with a ground resolution of 0.36 m^2; b) result of the surface reconstruction as a priori information for the 3D reconstruction based on multiple images. Even roof structures can be provided.

4 3D Reconstruction by Object Parts

In this chapter we present a procedure for fully-automatic 3D reconstruction which infers local 3D structures in space from its observed local 2D structures in multiple images. The result of the low-level pixel based image interpretation (cf. chapter 3) is used to restrict the aerial image to small regions of interest which serve as input for the subsequent process of 3D reconstruction. The orientation data of the images are assumed to be known, leading to geometric restriction during the matching procedure.

Notation:

Image features and aggregates are denoted by capital sens serif letters $\mathsf{F}_{ij}, \mathsf{A}_{ij}, ...$, i standig for the image, j standing for the individual feature. Sets of these features are denoted by capital calligraphic letters $\mathcal{F}_i, \mathcal{A}_i, ...$ Relations between features are denoted by slanted sens serif letters, e. g. $\mathsf{R}_{jj'}$. Object parts are denoted by capital $\mathsf{C}_{oj}, \mathsf{E}_{oj}, ...$ Vectors and matrices are denoted with bold face type letters $\boldsymbol{x}, \boldsymbol{\Sigma}, ...$ Estimates are denoted with a hat '$\hat{}$', true values with a tilde '$\tilde{}$'.

[5]The data was kindly provided by TOPOSYS (Ravensburg)

4.1 General aspects

The following aspects are decisive for characterizing our procedure :

Feature extraction and feature aggregation: Based on a polymorphic feature extraction (cf. [8], [9]) we derive a rich symbolic image description consisting of attributed points $\mathcal{P}$, lines $\mathcal{L}$ and regions $\mathcal{R}$ together with their mutual relations R that are contained in a feature adjacency graph (FAG).

Analysing this relational image description, especially the FAG, we are able to derive sets of basic aggregates $\mathcal{A}^{\mathsf{F}}{}_i = \{\mathsf{A}_{ij}\}$ by a bottom up process. These are point, line and region induced structures namely vertices $\mathcal{V}_i = \mathcal{A}^{\mathsf{P}}{}_i$, wings $\mathcal{W}_i = \mathcal{A}^{\mathsf{L}}{}_i$, cells $\mathcal{C}_i = \mathcal{A}^{\mathsf{R}}{}_i$ containing all neighbors and possibly indirect neighbouring features (cf. fig. 8). These basic aggregates serve as starting point for our reconstruction.

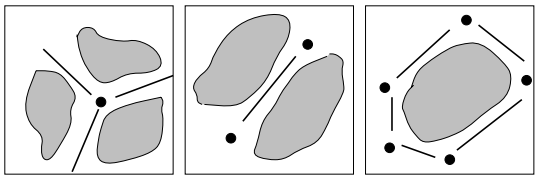

Figure 8: point-, line- and region-induced local aggregates $\mathsf{A}^{\mathsf{F}}_{ij}$ derived form the feature extraction, especially vertices V_{ij}, wings W_{ij} and cells C_{ij}, which are used for reconstruction.

Object Modeling: For introducing scene knowledge into the reconstruction process, the 3D object model must be placed at the same representation level like the image aggregates $\mathsf{A}^{\mathsf{F}}_{ij}$ which we propose for reconstruction. Transferring the 2D aggregates $\mathsf{A}^{\mathsf{F}}_{ij}$ into 3D results in a local boundary representation of *parts* $\mathtt{P}_{oj}$ of the object of three different classes being corners $\mathtt{C}_{oj}$, edges $\mathtt{E}_{oj}$ and faces $\mathtt{F}_{oj}$ of the object, where the index o stands for object space. These object parts are instances in the hierarchically structured 3D scene model (cf. [4]). In addition to a purely geometric/physical model of the scene, we introduce thematic information in order to exploit the generic scene knowledge about the object. This semantic modeling in a first step consists of class-labels for the 3D aggregates, which give the assignment to the underlying object hierarchy.

Strategy: The reconstruction is done by following the *hypothesize and verify* paradigm: The first step consists in the data driven generation of a set of corresponding aggregates leading to a set of hypotheses for 3D aggregates $\mathcal{A}^{\mathsf{F}}{}_{oj} = \{\mathsf{A}^{\mathsf{F}}_{oj}\}$. The second step is the model driven verification of the hypotheses by integrating 3D scene knowledge of a building specific model. Each hypothesis is semantically interpreted and classified to belong to one or several parts $\mathtt{P}_{oj}$ of the object. The reconstructed object parts can be connected by a grouping in space which

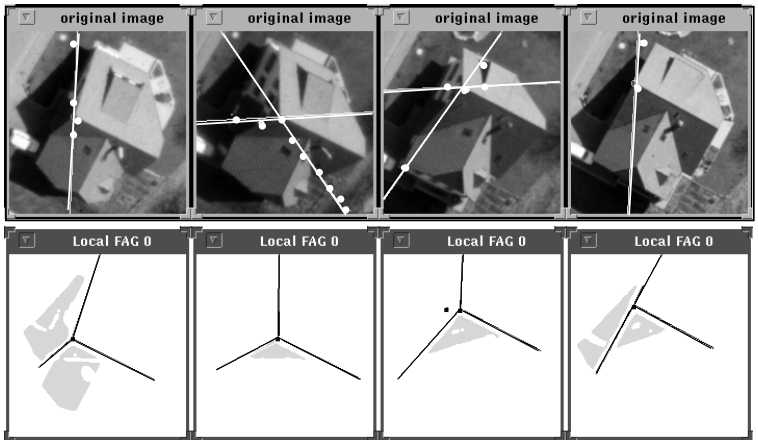

Figure 9: Selected basic aggregates are automatically reconstructed in 3D. The first row shows the original images, selected vertices and their corresponding epipolar lines. The second row shows a set of corresponding vertex aggregates $\{\mathsf{V}_{ij}\}$ being the basis for the transition to a 3D vertex V_{oj} in object space.

finally leads to a complete 3D surface description or which can be used as an input for our semiautomatic procedure (cf. chapter 5).

4.2 Hypotheses generation

We start the reconstruction based on point-induced local aggregates namely vertices $\mathcal{V}_i = \mathcal{A}^{\mathsf{P}}{}_i$. We found vertices to be most appropriate for matching, as for these local aggregates the imaging geometry leads to sharp constraints during the correspondence analysis. Moreover they contain neighboring edges (cf. fig. 8) which carry strong geometric information and help both for linking and for geometrically consistent reconstruction.

a. Finding vertex hypotheses:
The vertex reconstruction is initiated in order to come to a first 3D description of the object. The first step consists in the data driven generation of a set of corresponding vertices $\mathcal{V}_i$ being the basis for deriving 3D vertices $\mathcal{V}_o$. Here weak domain specific knowledge is used. For each vertex structure V_{ij} we perform a classification which serves for selecting strong vertices, i. e. vertices which can be found with high probability in the other images. This classification leads to a priority list of vertices. The search for corresponding vertices starts with selecting a vertex V_{ij} in one of the images by following this list. Restrictions for corresponding vertices are given due to the epipolar geometry. The priority list is also used for controling the search for correspondences out of the set of vertices which fulfil the constraints of imaging geometry and show structural similarity.

Additional heuristics for the selection of the next vertex structure to be analysed are used after the previous vertex has been successfully reconstructed. E. g. following edges or segments of previously reconstructed vertices can restrict the set of vertices $\mathcal{V}_i$ in the different images, which shall be analysed in the next step.

b. Finding corner hypotheses:
The second task within hypothesis generation is the model driven interpretation of the set of corresponding vertices $\{\mathsf{V}_{ij}\}$ referring to one hypothesized 3D vertex V_{oj} by integrating the object model of building specific corners $\mathtt{C}_{oj}$. Therefore a preliminary reconstruction is executed which solves the one-to-one mapping of the features $\mathcal{F}$ by relational matching of the local vertex aggregates $\{\mathsf{V}_{ij}\}$. This results is a preliminar estimation $\widehat{\mathsf{V}}_{oj}$ of a 3D vertex consisting of the 3D point, 2 or 3 neighboring edges and the face(s) between the edges.

Each hypothesis $\widehat{\mathsf{V}}_{oj}$ is now semantically interpreted and classified to be a corner $\mathtt{C}_{oj} = (\mathsf{V}_{oj}, \omega_{oj}^{C})$ of class ω_{oj}^{C}. The corner model contains not only the basic features $\mathcal{F}$ and their relations R, but also their domain dependent type and geometric constraints. For classification we analyse the line orientation in space discriminating 5 types of orientation, namely { `horizontal (h)`, `oblique+ (o+)`, `oblique- (o-)`, `vertical+ (v+)`, `vertical- (v-)` }. The sign symbolizes a positive/negative slope related to the position of the corner point (cf. [11]). In addition, symmetries to the ridge and planes being vertical are analysed to obtain a more detailed division into subclasses. The corner classes we have modeled up to now are covering 4 building types, namely { `flat_roof` , `non_orthogonal_flat_roof`, `gable_roof` and `hip_roof` }. The corresponding corners are shown in fig. 10. As far as an unambigous classification can not to be reached, all possible classifications have to be analyzed during the verification step.

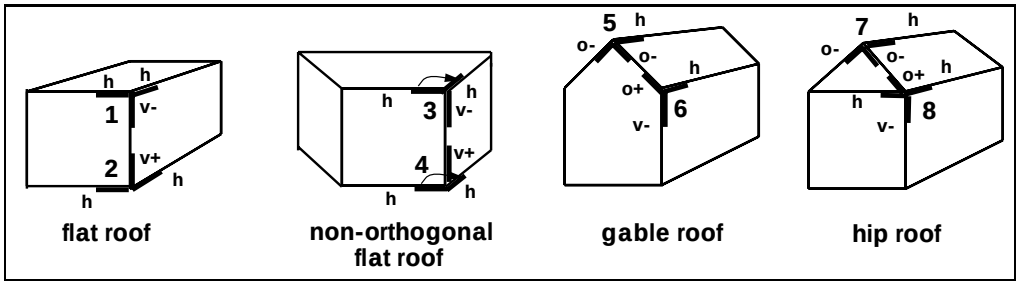

Figure 10: shows the 8 corner classes ω_{oj}^{C} we are using up to now, which are sufficient for the part-of description of 4 building types, namely { `flat_roof`, `non_orthogonal_flat_roof`, `gable_roof` and `hip_roof` } .

4.3 Hypotheses verification

After hypotheses generation we start the model driven verification of the hypotheses $\mathcal{V}_o = \{\mathsf{V}_{oj}\}$ by integrating the 3D scene knowledge of the building specific corner model $\mathtt{C}_{oj}$ of class ω_o^{C},

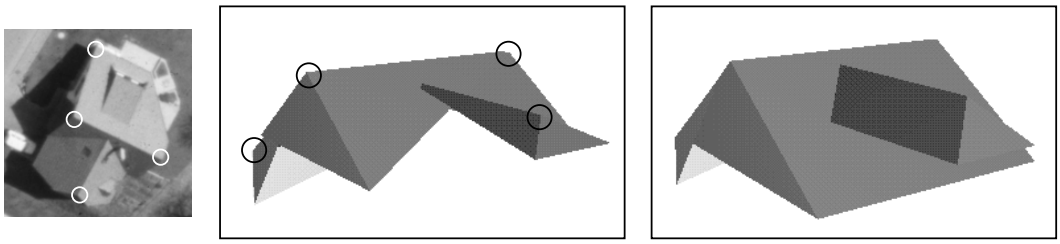

Figure 11: shows the result of the automatic 3D reconstruction: **a.** original image; **b.** partially reconstructed roof in 3D composed by different corners, which are denoted by circles; **c.** result of a 3D grouping process, which connects the reconstructed 3D parts of the building by analysing corner point and line identity and symmetry properties.

which is assigned to be the semantic interpretation of the reconstructed 3D vertex V_{oj}. The interpretation is evaluated by maximization of the probability for a given hypothesis set j.

$$P(\mathsf{C}_{oj} \mid \{\mathsf{V}_{ij}\}) = \frac{P(\{\mathsf{V}_{ij}\} \mid \mathsf{C}_{oj})P(\mathsf{C}_{oj})}{P(\{\mathsf{V}_{ij}\})} \tag{8}$$

For evaluating the geometry G, e. g. the geometric parameters $\boldsymbol{g}$, the classical modeling techniques of observation errors can be applied. Assuming the observations $E(\boldsymbol{y}) = f(\boldsymbol{g})$ being a function of the geometric parameters $\boldsymbol{g}$, the evaluation can be derived from the residuals $\boldsymbol{y} - \widehat{\boldsymbol{y}}$ of the optimal estimation $\widehat{\boldsymbol{y}} = f(\boldsymbol{g})$ using the *probability density function* $p(\boldsymbol{g})$ in case the feature exists and has been successfully matched to the model.

$$p(\boldsymbol{g}(\mathsf{F})|\exists(\mathsf{F}), \text{matched}, \mathsf{C}_o) = \frac{1}{(2\pi)^{n/2}(\det\boldsymbol{\Sigma}_{yy})^{1/2}} \quad e^{(-\frac{1}{2}(\boldsymbol{y}-f(\widehat{\boldsymbol{g}}))^T\boldsymbol{\Sigma}_{yy}^{-1}(\boldsymbol{y}-f(\widehat{\boldsymbol{g}})))}, \tag{9}$$

where n is the number of unknowns in $\boldsymbol{g}$. The evaluation of the existence of the features an their relations has been studied earlier (cf. [10]).

Figure 11 shows an example of the fully automatic 3D vertex reconstruction. More details of this approach are described in [17].

4.4 Grouping of object parts

The vertex reconstruction conceptually works in parallel on all vertices found in the images. After the vertex reconstruction is performed, a subsequent grouping of interpreted 3D vertices takes place where neighboring vertices, that share a common edge in at least one of the images, are linked using the part-of hierarchy of our object model. The information achieved so far may be incomplete, as some vertices or mutual relations between vertices may not be found. But due to the interpretation step we are able to establish a link to the modeling tools from CAD, namely **C**onstructed **S**olid **G**eometry (CSG) being the basis for our One-eye Stereo System

for Semiautomatic Building Extraction (cf. [16]). The interpretation during the automatic reconstruction, that assigns the 3D vertices $\mathtt{V}_{oj}$ to object corners $\mathtt{C}_{oj}$ of our hierarchical object model, leads to a partial instantiation of the type of the object and of different parameters of the basic shape primitives. An example for the instantiation of a gable roof shape primitive after reconstruction and interpretation of 5 corners is shown in fig. 12. Thus the remaining user interaction reduces to quality control of the result on the one hand and to completion of not yet instantiated parameters on the other hand.

We finally can reach global consistency of the result by a parameter estimation of complete CSG objects or parametrized object primitives as presented in [18].

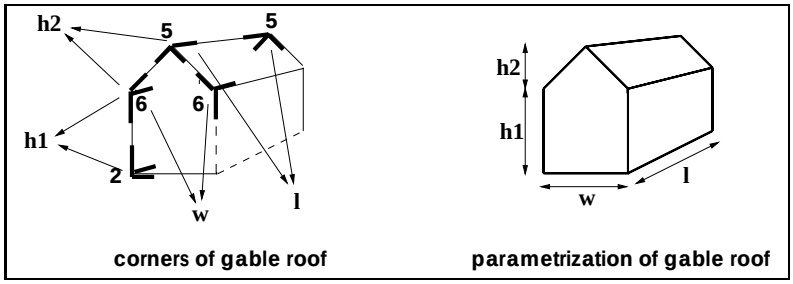

Figure 12: shows how the corner reconstruction and interpretation serves for deriving the parameter instances of a parametrized `gable_roof` building. Note that 5 corners are sufficient to define the 4 shape parameters of the object. The absolute position of the parametrized model can be fixed by the 3D coordinates of one reconstructed corner point. The orientation of the gable is for example defined by the horizontal ridge line of corner class number 5.

5 3D Reconstruction of Complete Buildings

The human operator is on the top level of our overall strategy and is required to at least finalize or validate the reconstruction of complete buildings as combinations of building parts. This is especially expected for very complex buildings with a polygonal ground plan and a detailed roof shape. Thus they are in general combinations and variations of basic building types. Higher degrees of details may contain also canopies, dormers, oriels, chimneys, and overhanging eaves. We have to cope with the fact the automated procedures described above may not be able to produce a sufficient level of detail or may produce incomplete results or may fail in some cases, especially if only two images are available.

To be able to reliably perform or complete the reconstruction we propose **semi-automatic procedures**, related to early work done at SRI [24]. Here the emphasis is on using the potential to incorporate the automatically derived parametric or prismatic building models to substantially reduce the workload for the operator. Previous work on semi-automatic procedures at our Institute has been performed by [18]. Their approach has been extended using

the Constructive Solid Geometry [12] for the three-dimensional modeling of complex buildings [6]. This work is done in cooperation with the Institute of Computer Sciene III, University of Bonn. The modeling process is done in monocular mode assisted by various supporting and automating tools for the form and pose adaptation of a large amount of CSG primitives.

In the following we describe the procedures and the stages in the workflow where information on the object structure derived by the automated procedures described above can be used to advantage.

5.1 Workflow and connection to overall strategy

The 3D acquisition process for buildings (cf. Figure 13) is divided into two phases, based on the assumption that the interior and exterior orientation data of the images are known. First, in the **navigation phase** the operator may zoom down into the aerial image and focus his interest on a particular building in one image. This step can be replaced by the automated detection alone or in combination with the automated 3D reconstruction of nodes. The **modeling phase** is performed by a semi-automatic form and pose adaptation of 3D models. This step can be substantially speeded up, if the type of parametric model or even the approximate parameters are already known, up to a simple confirmation of the automatically derived form and pose of the building. One homologous point has to be measured in the images in order to compute 3D world coordinates. This single step has not to be performed manually if the automated reconstruction of parts has been successful. The result of the building acquisition process is a 3D building description, which can also contain several buildings. For further data analysis and visualization a boundary representation (B-rep) is derived.

In the following we describe the Constructive Solid Geometry applied to 3D building acquisition, we briefly specify the supporting tools and describe then the automated tools for the form and pose adaptation, without support from previous automated detection and reconstruction of models. An extended description of the current system can be found in [6] and [16].

5.2 Constructive Solid Geometry

Within this approach buildings are reconstructed by combining a series of 3D atoms, so-called CSG primitives (short: primitives), until the complete building has been modeled. For the combination of primitives different commutative and distributive CSG operations are provided: union, intersection, and difference. As primitives we are using box, chock, cone, cylinder, half-chock, pyramid, and tetrahedron (cf. Figure 14) and three **combined primitives**: saddleback roof building, hip roof building, and lop-sided saddleback roof building. A parameterized description of these combined primitives can be found in [18].

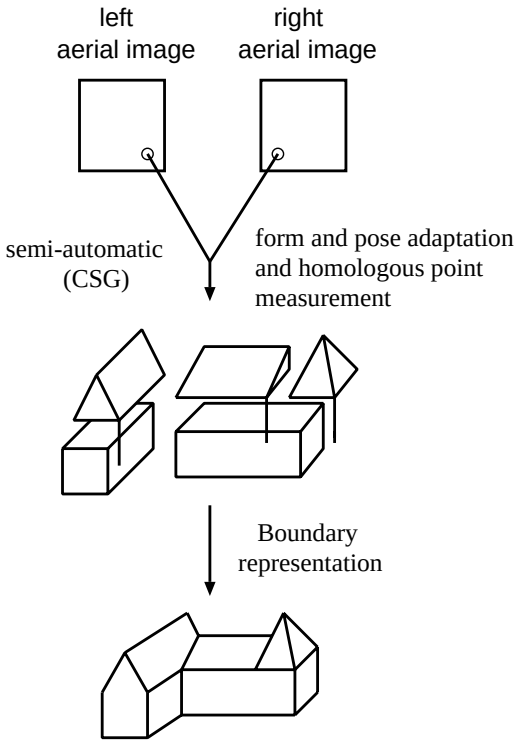

Figure 13: 3D building acquisition from stereo aerial images. CSG primitives are adapted in one image resulting in a binary tree.

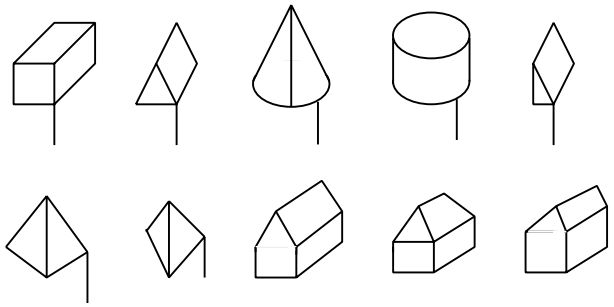

Figure 14: Primitives (with flagpole): box, chock, cone, cylinder, half-chock, pyramid, tetrahedron. Combined primitives: saddleback roof building, hip roof building, and lop-sided saddleback roof building.

5.3 Semi-automatic pose and form adaption of models

The **degree of generalization** determines which details of a building are reconstructed, e.g. if chimneys have to be modeled. The choice of the degree of generalization has to be made by the operator. The modeling process results in a CSG tree, whose interior nodes contain operations and the leaves contain instantiated primitives and attributes, e.g. form and pose parameters. There are several possibilities to describe and acquire a building. The operator has the freedom of choice to determine the construction of the CSG tree.

During the modeling phase the operator has to choose a primitive which will be projected as a wire frame model (with removed hidden lines) into the focussed image region, adapt the parameters of the wire frame model by clicking with the mouse onto its edges and pulling them to the correct size of the modeled building part. These steps have to be repeated for each primitive until the whole building is described. All these adaptations are performed in monocular mode.

During the adaptation process the system supports the operator by diverse tools. A 3D rendering of the currently acquired building description is used to check for completeness and correctness. The inheritage of parameters of the previously acquired primitive allows efficient aquisition in areas with similar building types, like row houses and reduces the need for homologous point measurements of primitives belonging to the same building. The flagpole principle allows the operator to efficiently adjust primitives above ground level, like canopies, dormers, smoke stacks, chimneys, etc. All primitives (except the combined ones) are equipped with a pole along which they are moved up or down by the operator in order to adjust their height above ground. The height of the lower end of the flagpole is the ground height inherited from the previously acquired primitive.

Substantial reduction of user interaction is provided by the following automated modules:

- Docking of primitives. Describing a building by the combination of primitives or combined primitives requires a precise "docking" of the primitives. This docking is supported by matching and gluing facilities. The former allows to match at least two edges of different primitives and the latter matches and glues exactly two faces of different primitives together. These functions are based on a user-defined radius (cf. Figure 15) and are thus extended to three dimensions. The matching and gluing are performed automatically with already instantiated primitives and even with invisible lines, respectively faces. This enables the operator to dock the current primitive easily to neighboring primitives.

- Line extraction for the acquisition of prismatic building models. For the extraction of prismatic building models, i.e. buildings with a polygonal ground plan and a constant height the operator simply has to identify automatically extracted straight line segments in the image belonging to the ground plan, instead of measuring the vertices. The intersections of lines are automatically calculated. With a manual definition of the height of the building and the measurement of one homologous point in the other image the prismatic building model can be reconstructed in 3D.

- Multi-image matching and fine-tuning by clustering techniques. The measurement of homologous points for single primitives can be replaced by an automated matching procedure. It is based on automatically extracted straight line segments in multiple images. The lines of the form and pose adapted parametric building model in one image (cf. above) are compared to the extracted line segments in the other images. Robust pose clustering

techniques ([28], [27]) are used to determine the height in 3D. In a final fine-tuning step, a robust spatial resection, using all line segments in all images provides an optimal fit of the selected model to the image data. More details can be found in [18], [15].

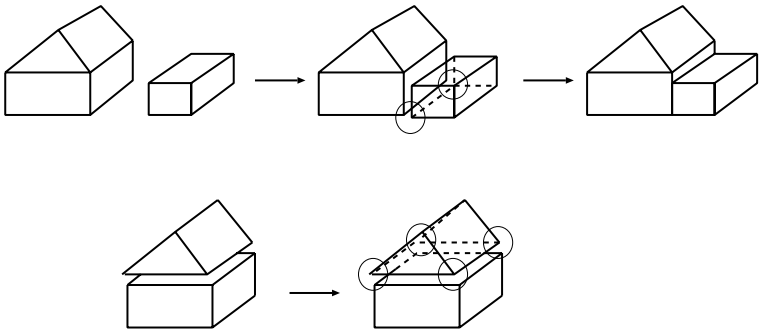

Figure 15: Automatic matching process of two edges (first row) and automatic gluing process of two faces (second row). The circles show the user-defined radius of a sphere, where matching and gluing is performed.

Before storing the corresponding CSG tree of a 3D building description, the operator may add useful knowledge to the building description. This extended CSG structure can be further analyzed by a conversion to Drawing Interchange File (DXF) format for standard CAD systems or visualized and animated by tools, like the Virtual Reality Modeling Language (VRML).

5.4 Experimental results with the semi-automatic system

The semi-automatic system is a practical tool for the acquisition of complex 3D building descriptions in extended scenes. The test field OEDEKOVEN has been chosen to acquire data for an OEEPE[6] test on 3D city models. The image scale of the B/W aerial imagery is 1:12000, the focal length is 153 mm. A stereo pair had been digitized with a pixel size of 12.5 μm in the image or 15 cm on the ground. The scene of about 3 km^2 had been divided into two parts for two operators. The task was to acquire detailed roof and building structures, which refers to a very low level of generalization.

5.4.1 Primitives used

1870 buildings or building-blocks (CSG trees) have been extracted with an average amount of 2.9 primitives per CSG tree resulting in 5499 primitives. As primitives occurred 55% boxes, 0.5% chocks and half-chocks, 31% saddleback roof buildings, 2.5% hip roof buildings, 10% lop-sided saddleback roof buildings, and 1% others. There are single buildings, building groups, garages, churches, farms, and plants. Figure 16 visualizes a small subset of the acquired buildings. The

[6]Organisation Européenne d'Etudes Photogrammétriques Expérimentales

Figure 16: A visualized part of the acquired extensive scene OEDEKOVEN.

terrain between the buildings is automatically triangulated from the measured ground heights of the buildings. An optional texture extraction process fully automatically provides each 3D face with the texture from one of the images, if visible there.

5.4.2 Performance

The gross and net times for data acquisition are given in Table 1. The gross time contains the modeling time, the internal navigation and the external navigation and organization. The times per building primitive are given as **mean values**.

The modeling time contains the form adaptation, the specification of operations, the measurement of homologous points and for complex buildings a 3D visualization. Due to some very complex buildings the mean value for the modeling (47.8 seconds) is higher than the median value. The median modeling time per building primitive is below 40 seconds. Independent from the complexity of the building or the amount of primitives in a CSG tree we are observing a more or less constant modeling time per primitive. The local navigation time contains the navigation through the pyramid and the selection of primitives. The short local navigation time of 10.3 seconds per primitive indicates the near optimality of this acquisition step. The global navigation and organization (31.5 seconds) covers the navigation through the project area, checks of completeness, editing and 3D visualization.

The gross times of this test are about 25% shorter compared to an earlier version of the system [21] tested on comparable image material, but with much less buildings. The higher performance here is in addition combined with a more detailed building acquisition. On of the operators was inexperienced, but reached the same performance for the modeling as the experienced operator,

Time/primitive [seconds]	Total
Modeling time	47.8
Local navigation	10.3
Global navigation, organization	28.3
Gross time	86.4

Table 1: Average acquisition times OEDEKOVEN (status May, 1996) for all primitives (5499).

after one week training only. This shows the potential of this system, for usage as a modeling tool by a non-photogrammetrists in practice.

The accuracy of the system depends on many factors, like image scale, pixel size, orientation, film processing, scanning, selection and measurement of models and homologous points, the definition uncertainty of building corners, and the generalization level. A preliminary accuracy check of few buildings based on the differences of measured roof point coordinates and ground truth yielded an external accuracy of $\sigma_{X,Y} \approx 25$ cm and $\sigma_Z \approx 35$ cm. This corresponds to the accuracy of analytical photogrammetric methods.

From succesfull applied automated procedures described above we expect a substantial reduction for (a) the modeling time, by preliminary form adaption and measurement of homologous points for b) the local navigation by a preliminary selection of primitives and c) for the global navigation by a guided navigation through the project area and a focusing on each building. Having more than two images availabe we further expect an improvement in accuracy and the reliability of the model and homologous point measurement.

6 Discussion and Conclusions

The paper wants to discuss strategies for 3D-building acquisitions by giving a frame of reference for describing the role of algorithms and for making control tasks explicit. Three examples illustrate what type of problems we encounter when automating image interpretation for building acquisition.

It seems useful to discuss the role of the three examples within the complete - though by no means completed - process.

1. *Focussing* on areas of interest necessarily requires low level, i.e. raster models of the objects to be found. Thus the appearance of buildings in the raster data needs to be available. In our example digital surface models, possibly complemented by colour images appear to be a natural choice. The use of Bayesian sets for modelling the context, stored in the higher levels of a feature pyramid, seems to be promising. Focussing thus is interpreted

as generating hypotheses on regions for which a likelihood is given, which can be used for controlling the next steps.

2. *3D-reconstruction* is shown to be feasible on the mid- and the high level. As geometric and semantic models are integrated into the matching process. Again the result are hypotheses on corners of buildings. Their likelihood can be used in the grouping process. Obviously the verification step which uses a classification procedure integrates geometric knowledge, information on defects of the feature extraction and semantic knowledge and requires an optimal geometric reconstruction as a prerequisite.

3. Semiautomatic procedures will be necessary as long as the variety of buildings is larger than covered by the models used in the automatic procedures. The building models used in the two previous examples are too coarse and not general enough. Semiautomation always faces the problem to reasonably divide the work between computer and operator, here by performing the measurement tasks to the computer and leaving the decision steps to the operator. The CSG-modelling has shown to be quite effective. The achieved acquisition times always need to be seen as reference for automatic procedures, which claim to be operational.

The chosen strategy in all three cases was an interpretation of bottom-up with top-down processes: grouping hypotheses were generated data driven, these hypotheses were then verified using higher level models, where the final verification was left to the operator.

This report on ongoing work leaves enough open problems. Without going into the details of the described algorithms a few should be mentioned:

- The current exploration of finding good algorithms for investing the observability of building features by various sensor data should be continued. Their performance should be clearly documented in order to make these algorithm available to other research groups.

- In order to achieve this goal an attempt should be made to find basic models for buildings, levels of such models, parts of buildings and some standardized notions. This would ease comparisons on a conceptual level, i.e. on standard data sets.

- The general problem of performance characterization of algorithms may use building acquisition as a test bed [7]. Here a convention on how to describe the quality of an output needs to be achieved, e.g. requiring covariance matrices, confusion matrices and probabilities of failure and success.

- Only after this stage different strategies, i.e. sequences of processes/algorithms can be compared in case these sequences are not meant to be fixed. The comparison then may refer to both, the theoretical performance, i.e. the performance prediction, and the empirical performance. Here standard tests are helpful.

[7] where buildings may easily be replaced by 3D-man-made objects, including office scenes or industrial parts

As a consequence of these open problems we can expect progress to be slow but steady. This avoids unnecessary hurry and allows time to discuss the real interesting questions, why algorithms actually work.

References

[1] F. Ackermann and M. Hahn. Image pyramids for digital photogrammetry. In Heipke Ebner, Fritsch, editor, *Digital Photogrammetric Systems*. Wichmann-Verlag, 1991.

[2] J. M. Agosta. *The structure of Bayes networks for visual recognition*, pages 397–405. 1990.

[3] Charles Bouman and Bede Liu. Multiple resolution segmentation of textured images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-13(2):99–113, February 1991.

[4] C. Braun, T.H. Kolbe, F. Lang, W. Schickler, V. Steinhage, A.B. Cremers, W. Förstner, and L. Plümer. Models for photogrammetric building reconstruction. *Computer & Graphics*, 19(1), 1995.

[5] A. Brunn, U. Weidner, and W. Förstner. Model-based 2d-shape recovery. In G. Sagerer, S. Posch, and F. Kummert, editors, *Mustererkennung 1995*, pages 260–268. DAGM, Springer, 1995.

[6] R. Englert and E. Gülch. One-eye stereo system for the acquisition of complex 3D building descriptions. *GIS*, 4, 1996.

[7] W. Förstner. Image analysis techniques for digital photogrammetry. In *Proc. of 42nd Photogrammetric Week, Stuttgart, Sept. 1989*. Schriftenreihe des Instituts für Photogrammetrie, Heft 13, 1989.

[8] W. Förstner. A framework for low level feature extraction. In J.-O. Eklundh, editor, *Computer Vision - ECCV '94, Vol. II*, pages 383–394. Lecture Notes in Computer Science, 801, Springer-Verlag, 1994.

[9] C. Fuchs and W. Förstner. Polymorphic grouping for image segmentation. In *5th ICCV '95, Boston*, pages 175–182. IEEE Computer Society Press, 1995.

[10] C. Fuchs, F. Lang, and W. Förstner. On the noise and scale behaviour of relational descriptions. In Ebner, Heipke, and Eder, editors, *ISPRS Vol. 30, 3/2*, pages 257–267. SPIE, 1994.

[11] E. Gülch. A knowledge-based approach to reconstruct buildings in digital aerial imagery. In *Proceedings of 17th ISPRS Congress, Washington D. C., BII*, pages 410–417, 1992.

[12] C.M. Hoffmann. *Geometric and Solid Modeling*. Morgan Kaufmann, Palo Alto, CA, USA, 1989.

[13] K. R. Koch. Bildinterpretation mit Hilfe eines Bayes-Netzes. *Zeitschrift für Vermessungswesen*, 120(6):277–285, Juni 1995.

[14] W. G. Kropatsch and Annick Montanvert. Irregular pyramids. Technical Report PRIP-TR-5, Dept. f. Pattern Recognition and Image Processing, TU Wien, 1992.

[15] T. Läbe and K.-H. Ellenbeck. 3D-wireframe models as ground control points for the automatic exterior orientation. In *Proceedings ISPRS Congress, Comm. II, Vienna*, 1996.

[16] F. Lang and W. Förstner. 3D-city modelling with a digital one-eye-stereo system. In *ISPRS Congress, Comm. IV, Vienna*, 1996.

[17] F. Lang and W. Förstner. Surface reconstruction of man-made objects using polymorphic mid-level features and generic scene knowledge. In *Proceedings ISPRS Congress, Vienna*, 1996.

[18] F. Lang and W. Schickler. Semiautomatische 3D-Gebäudeerfassung aus digitalen Bildern. *Zeitschrift für Photogrammetrie und Fernerkundung*, 5:193–200, 1993.

[19] T. S. Lee, D. Mumford, and A. Yuille. Texture segmentation by minimizing vector-valued energy functionals: The coupled membrane model. In *Computer Vision – ECCV '92, Proceedings*, pages 165–173. Springer, 1992.

[20] C.-E. Liedtke, Th. Schnier, and A. Bloemer. Automated learning of rules using genetic operators. In *Proc. ICAP 93*, 1993.

[21] T. Löcherbach. System performance for semiautomatic building reconstruction. In *Second Course in Digital Photogrammetry, Institute of Photogrammetry, University of Bonn, Bonn, Germany*, 1995.

[22] P. Meer. Stochastic image pyramids. *CVGIP*, 45(3):269–294, March 1989.

[23] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann Publishers, 1988.

[24] L. Quam and Th. Strat. SRI image understanding research in cartographic feature extraction. In C. Heipke H. Ebner, D. Fritsch, editor, *Digital Photogrammetric Systems*, pages 111–121. Wichmann, Karlsruhe, 1991.

[25] S. Ravela, R. Manmatha, and E. M. Riseman. Image retrieval using scale-space matching. In Bernard Buxton and Roberto Cipolla, editors, *Computer Vision-ECCV'96*, pages 273–282, 1996.

[26] S. Sarkar and K. Boyer. Integration, Inference, and Management of Spatial Information Using Bayesian Networks: Perceptual Organization. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 15(3):256–274, Mar. 1993.

[27] W. Schickler. Feature matching for outer orientation of single images using 3-D wireframe controlpoints. In *Internat. Archives for Photogrammetry, B3/III, Washington*, pages 591–598, 1992.

[28] G.C. Stockman. Object recognition and localization via pose clustering. *Computer Vision, Graphics, and Image Processing*, 40:361–387, 1987.

[29] D. Terzopoulos. Image analysis using multigrid relaxation methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(2):129–139, 1986.

[30] U. Weidner and W. Förstner. Towards automatic building extraction from high resolution digital elevation models. *ISPRS Journal*, 50(4):38–49, 1995.

[31] Uwe Weidner. MDL-basierte Formrekonstruktion zur Gebäuderekonstruktion. In *DGPF Jahrestagung 4.-6.10.1995*, Hannover, 1995.