

AUTOMATIC BUILDING RECONSTRUCTION FROM AERIAL IMAGES: A GENERIC BAYESIAN FRAMEWORK

Franck Taillandier^a, Rachid Deriche^b

^a Institut Géographique National/MATIS, 2-4 avenue Pasteur, 94165 Saint-Mandé - franck.taillandier@ign.fr

^b INRIA/ODYSSEE, 2004 Route des Lucioles - BP 93 - 06902 - Sophia-Antipolis Cédex, France - Rachid.Deriche@sophia.inria.fr

Commission III, WG III/4

KEY WORDS: Building, Reconstruction, Modeling, Aerial, Image, Automation, Systems, Photogrammetry.

ABSTRACT:

A novel system for automatic building reconstruction from multiple aerial images is presented. Compared to previous works, this approach uses a very generic modeling of buildings as polyhedral shapes with no overhang, in which external knowledge is introduced through constraints on primitives. Using planes as base primitives, the algorithm builds up an arrangement of planes from which a 3D graph of facets is deduced. In a so-called “compatibility graph” where the nodes are the initial facets of the 3D graph and edges between two nodes state that both facets belong to at least one common hypothesis of building, it is shown that maximal cliques supply all the hypotheses of buildings that can be deduced from the arrangement of planes. Among these hypotheses the choice is done through a bayesian formulation that balance data adequacy and caricature needs. Results are provided on real images and show the validity of the approach that remains very generic on the contrary to model-based methods while bringing external architectural information through geometric constraints, which generally lacks in data-driven algorithms.

1. INTRODUCTION

1.1 Context

Building reconstruction is of primary importance in many applications such as virtual tourism, fighting simulations, urban environment management. Unfortunately, manual reconstruction through usual photogrammetric techniques is a heavy burden and a lot of research tend towards automated solutions. Automatic building reconstruction from multiple aerial images is however a difficult problem due to the complexity of the roof structures present in urban or suburban areas and the presence of occlusions such as vegetation. Most automatic approaches have to deal with antagonistic aspects: on the one hand handling the tremendous complexity of roof structures and thus supplying a very generic modeling of building and on the other hand allowing simplifications of shapes when needed, which is often done in manual reconstruction for dormer windows, chimneys but also for small building recesses. The goal of an automatic algorithm is thus to provide an “acceptable” caricature of generic building; acceptable meaning that it meets some specifications. Furthermore, systems have to face errors of primitives detectors: under or over-segmentation, geometrical inaccuracy. To overcome these errors and to deal with mandatory simplifications, external knowledge must be introduced, either as models of buildings, which reduce the generality of the approach, either as constraints on primitives, which is the choice made in this article.

1.2 State of the Art

In the context of building reconstruction from multiple aerial images, systems can be classified as data-based or model-based. In the former one, authors (Baillard and al., 1999; Heuel and al., 2000; Ameri and Fritsch, 2000; Scholze and al., 2002) have chosen not to restrict the set of available shapes for roof structures. They often use only one kind of primitives (3D segments for (Baillard and al., 1999; Scholze and al., 2002), corners in (Heuel and al., 2000) and planar patches in (Ameri and Fritsch, 2000)) and handle under-segmentation or primitives errors with difficulty. On the contrary, the latter ones (Fuchs and Le-Men, 1999; Fischer and al., 1998) use some models of buildings to restrict the set of possible shapes. This external knowledge enables to overcome lack of detection and over detection. Although

these two approaches use different method to search for the best model, they both try to solve for the lack of generality inherent in these strategies either through heuristic rules in (Fischer and al., 1998) or through graph grammars in (Fuchs, 2001). They both provide promising results but, despite obvious efforts, are still limited to simple forms and thus can not handle all the shapes available in urban or suburban areas. Besides, the robustness of the approaches lies intrinsically in the small amount of models. Increasing the library of models would result in a increased complexity and a least robustness.

In another context, with cadastral limits, (Jibrini, 2002; Flamanc and al., 2003) searches, from a set of planes for the best continuous polyhedral surfaces enclosed in the cadastral limits. This very generic modeling handles over-detection of planes. Our strategy inherits from this method with extensions to deal with vertical planes and to integrate external knowledge through geometric constraints in the choice of model representation and in the reconstruction step.

1.3 Global Algorithm

We decide to model a building as any polyhedral surface with no overhang whose external border is constituted of vertical planes. This definition remains very general and can represent almost any building seen from aerial images in urban area with exception of curved building. The algorithm focuses on individual building reconstruction with a focusing zone manually selected through a threshold on altitude, although automatic methods could be used instead. Multiple *calibrated* images (with a resolution of 25cm in our case) or products computed from these images such as DEM (Baillard and Dissard, 2000) are used. Figure 1 summaries the general scheme of the system that can be subdivided in three main steps: primitives detection, hypotheses extraction and choice of the best representation together with geometric refinement. In the primitives detection step, starting from correlation DEM and a rough focusing zone, the algorithm extracts planar patches and oriented portions of facades that represent the base primitives (Figure 7(a)). In a second step, from the arrangement of planes deduced from these base primitives, it builds up a 3D graph of facets and then a so-called “compatibility graph” where the nodes are the initial facets of the 3D graph and edges between two nodes state that both facets belong to at least one common hypothesis of

building. In our scheme, buildings are modeled, in a very generic way, as polyhedral volumes with no overhang and it is shown that maximal cliques in the compatibility graph supply all the hypotheses of buildings that can be deduced from the arrangement of planes. It thus extends the work of H. Jibrini to the cases where vertical planes may be present. Before effectively computing all the solutions by brute force search, some simplifications in the initial 3D and compatibility graphs are made, based on DEM and focusing mask, in order to reduce the crippling combinatory that arise from the maximal cliques problem. Finally, in the set of all the hypotheses extracted from the compatibility graph, the choice of the final model is done through a bayesian formulation that enables different kinds of observations to be taken into account (for instance 3D segments, images, focusing zone). Model complexity and a priori constraints on base primitives such as orthogonality, parallelism, symmetry, horizontality are also naturally introduced to balance caricature needs and data adequacy. Geometric refinement and constraints enforcement are finally performed on the reconstruction to provide the final result.

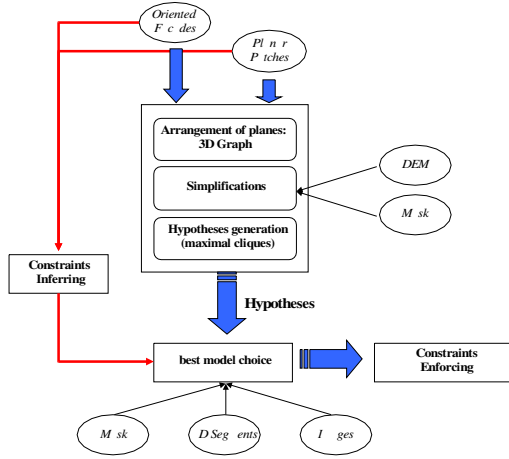


Figure 1: General overview.

2. 3D PRIMITIVES RECONSTRUCTION

As stated above, a focusing mask is provided to the algorithm along with a horizontal “ground” facet \mathcal{F}_S that lies on a plane $\mathbf{z} = \mathbf{z}_g$ (where \mathbf{z}_g denotes the altitude of the ground) and whose borders delineates the zone on which reconstruction takes place. Non vertical Planar patches extraction is performed on a DEM through a region growing algorithm whose details are given in (Taillandier and al., 2003). The main characteristic of the algorithm is its ability to integrate image constraints as well as 3D segments to better delineate the patches.

Vertical planar patches (facades) detection is then performed from the planar segmentation obtained at the previous step. Image and 3D segments constraints enable indeed to obtain a much better delineation of planar patches which is used as follows: each border between two patches that represent an important altitude gap is considered as a potential facade and polygonised. In the same way, the borders close to the focusing mask are considered as potential facades and polygonised. All facades are then oriented, normals pointing towards the outer part of the building.

Whereas previous primitives are considered as base primitives, 3D segments are important in the final choice of the best representation for the building thanks to their capabilities to enhance the structure of the scene. The algorithm used is detailed in (Taillandier and Deriche, 2002). Its main characteristic lies in the

use of a true multi view technique for matching between 2D segments and the propagation of uncertainty to validate matching hypotheses. These 3D segments suffer from few errors and some under-segmentation (roughly 30%) but are very precisely located. Planar hypotheses (vertical and non vertical) are base primitives for the following. For these primitives, emphasis is put on exhaustivity and geometric precision. During the generation of all building hypotheses, it is indeed essential that *all important planes equations* (non vertical and vertical ones) be present whereas planar delineation is of less importance.

3. MODELS GENERATION

3.1 Hypotheses

In the following, we will assume that all principal non vertical and vertical planes are detected with emphasis put on geometry. Erroneous planes are not critical insofar as this condition is met: they will be filtered out in the model choice step. Furthermore, all planes are oriented: normals pointing upward for non vertical planes and normals pointing towards the outer part of the volume for vertical planes.

Main steps of hypotheses generation and filtering as well as model evaluation are illustrated on a real example in figure 7.

3.2 3D Graph

A plane arrangement is deduced from the plane $\mathbf{z} = \mathbf{z}_g$ and all detected non vertical planes and facades, it builds up a 3D graph of facets. This graph is enclosed in a prismatic volume whose base is \mathcal{F}_S , delimited by \mathbf{z}_g and a maximum altitude. All facets of the graph are oriented from the planes they lie on (Figures 3(a) and 3(b)).

Two adjacent facets are *coherent* along their common edge s if, with the notations of figure 2(a):

$$[s; \mathbf{v}_1; \mathbf{n}_1] : [s; \mathbf{v}_2; \mathbf{n}_2] < 0 \quad (1)$$

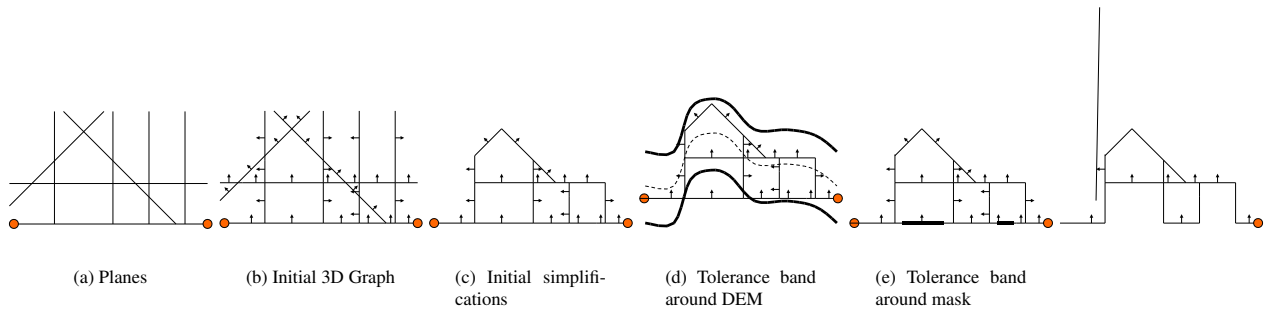
where $[a; b; c]$ stand for the triple scalar product. Intuitively, two facets are coherent if, assuming the normal points towards the “exterior” of a volume, the normal direction remains coherent when switching from one facet to the other one. Figure 2 shows some counterexamples.

We call *admissible surface* a continuous polyhedral surface made of coherent facets, whose horizontal projection on $\mathbf{z} = \mathbf{z}_g$ completely cover \mathcal{F}_S . An edge is said to be locally admissible if either it belongs to the borders of \mathcal{F}_S either there are at least two facets coherent along it. A facet is *locally admissible* if it is locally admissible along its bordering edges. A facet is *admissible* if it belongs to at least one admissible surface. Two facets are *compatible* if they belong to at least one common admissible surface. It is obvious that admissible facets are locally admissible. Once again, intuitively, an admissible surface keeps a coherent normals direction along its facets. It can also be shown that, given our context and especially normals orientations (normals pointing upwards for non vertical planes), an admissible surface is a continuous polyhedral surface *with no overhang*.

3.3 Graph simplifications

In the 3D graph, facets that are locally non admissible are recursively suppressed. These are facets hanging on the lateral and top parts (not the base that is part of the 3D graph) of the virtual prismatic volume. Non vertical facets touching the plane $\mathbf{z} = \mathbf{z}_g$ are then also recursively removed.

From this point, on any admissible surface, *all connected sets of facets that are not included in plane $\mathbf{z} = \mathbf{z}_g$ are one building as defined previously. the search for all building models derived from data boils down to the search for all admissible surfaces*. An admissible surface can however include more than one building



Using Bayes theorem, it comes:

$$\mathbf{P}(\mathbf{M} \mid \mathcal{D}) = \frac{\mathbf{P}(\mathcal{D} \mid \mathbf{M}) \cdot \mathbf{P}(\mathbf{M})}{\mathbf{P}(\mathcal{D})} \quad (3)$$

where $\mathbf{P}(\mathcal{D})$ does not depend on the model \mathbf{M} , thus

$$\hat{\mathbf{M}} = \underset{\mathbf{M} \in \mathcal{M}}{\operatorname{argmax}} \mathbf{P}(\mathcal{D} \mid \mathbf{M}) \cdot \mathbf{P}(\mathbf{M}) \quad (4)$$

We use images \mathbf{I} , 3D segments \mathbf{S} and focusing mask \mathbf{M}_a as observations \mathcal{D} although any additional observation could be used (laser points clouds, manual selection, cadastral maps for instance). Assuming independence between these observations leads to

$$\hat{\mathbf{M}} = \underset{\mathbf{M} \in \mathcal{M}}{\operatorname{argmax}} \mathbf{P}(\mathbf{S} \mid \mathbf{M}) \cdot \mathbf{P}(\mathbf{I} \mid \mathbf{M}) \cdot \mathbf{P}(\mathbf{M}_a \mid \mathbf{M}) \cdot \mathbf{P}(\mathbf{M}) \quad (5)$$

The model probability $\mathbf{P}(\mathbf{M})$ depends on the model complexity. As recalled in section 1.1, one looks for the simplest model, given the observations. The model probability is linked to the minimum description length through:

$$\mathbf{P}(\mathbf{M}) = \mathbf{C} \exp \frac{-\mathcal{L}(\mathbf{M})}{\beta} \quad (6)$$

where \mathbf{C} is a normalization factor common to all models of \mathcal{M} that will be omitted in the following. β tunes the level of caricature desired. A study on the influence of this parameter is done in section 5.

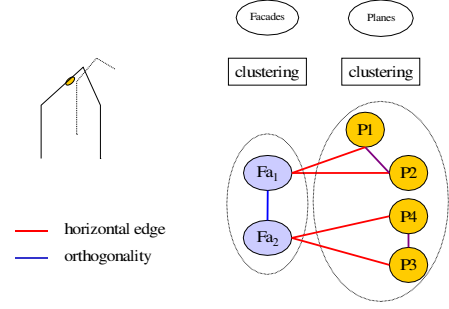
4.2 Model Formulation

Constraints Inferring As stated above, constraints on primitives represent the external knowledge brought in the decision process to guide the choice of $\hat{\mathbf{M}}$. That way, we can also integrate architectural knowledge to favor some types of architecture when ambiguities remain in the choice of the model.

Constraints are first inferred on normals of the base primitives. We adopt the principles of the system described in (Grossmann, 2002) since this system is used for enforcing constraints in the reconstruction. The algorithm uses one threshold θ . Normals are first clustered by angular proximity, thus grouping approximately parallel normals. For each cluster, a direction, average of the clustered normals, is defined. This process ensures that minimum angular distance between two directions is θ and handles parallelism constraints in the model.

A constraint graph is then deduced from these directions that are nodes of the graph, whereas edges represent constraints between directions: orthogonality, horizontal cross product (stating that the intersection of both planes is horizontal, which is usual in urban environment) and vertical symmetry. An edge is created if the relation is verified in the angular tolerance θ . Each edge \mathbf{q} is valued with a weight $\mathbf{C}(\mathbf{q})$ related to the number of degree freedom that the constrain suppress on normal coordinates (Figure 6). This will be of primary importance to integrate constraints in the choice process (see next).

Notations \mathbf{F}_M , \mathbf{E}_M and \mathbf{V}_M define respectively the number of facets, edges and vertices for a given model \mathbf{M} . \mathbf{P}_M , \mathbf{R}_M represent the number of non-vertical directions and vertical directions used in the models (after the clustering process) and $\mathbf{D}_M = \mathbf{P}_M + \mathbf{R}_M$ the total number of directions. Finally \mathbf{C}_M is the number of constraints on directions in the model (the number of edges in the constraints graph related to \mathbf{M}). The basic idea for complexity computations follows the principles edicted in (Kolbe, 1999) based on the transmission of information related to the model: topological, geometrical and constraints.



By noting $\mathbf{f}(\mathbf{x}; \mathbf{y})$ the altitude given by a facette $\mathbf{f} \in \mathcal{H}(\mathbf{M})$ at location $(\mathbf{x}; \mathbf{y})$, the score is:

$$\mathbf{P}(\mathbf{I} | \mathbf{M}) = \sum_f \sum_{(M)(x,y)} \frac{\mathbf{C}(\mathbf{x}; \mathbf{y}; \mathbf{f}(\mathbf{x}; \mathbf{y}))}{|\mathbf{SP}(\mathbf{M})|} \quad (11)$$

where, in multi-image context, the correlation score is the one proposed in (Paparoditis and al., 2000) normalized by the number of images \mathbf{n} :

$$\mathbf{C}(\mathbf{x}; \mathbf{y}; \mathbf{z}) = \frac{\text{Var}(\sum_i \mathbf{v}_i(\mathbf{u}_i; \mathbf{v}_i))}{\mathbf{n} \cdot \sum_i (\text{Var}(\mathbf{v}_i(\mathbf{u}_i; \mathbf{v}_i)))} \in [0::1] \quad (12)$$

where each correlation window centered on $(\mathbf{u}_i; \mathbf{v}_i)$ in image \mathbf{I}_i is represented by a vector $\mathbf{v}_i(\mathbf{u}_i; \mathbf{v}_i)$. $(\mathbf{u}_i; \mathbf{v}_i)$ is the projection of $(\mathbf{x}; \mathbf{y}; \mathbf{z})$ through known projection matrix from image \mathbf{I}_i . Note that the cube of correlation scores can be precomputed which speeds up evaluations.

3D segments 3D segments bring important information on structure of the scene. Being reconstructed independently of the planes, they can give very good evidence on presence of some edges. For each 3D segment \mathbf{s} , an edge \mathbf{a} is matched if angular deviation is lower than a threshold θ_s and if distance deviation is lower than another threshold \mathbf{d}_s , it will be noted $\mathbf{a} \in \mathcal{R}_s$. For each matched edge, the overlap score $\mathbf{r}(\mathbf{a}; \mathbf{s})$ of \mathbf{s} by \mathbf{a} is used for probability computation: but errors of the 3D segment detector must be taken into account and thus a default value for fake matching (which is assumed to be the case when a 3D segment is matched to no edge) θ_s is thus attributed, leading:

$$\mathbf{P}(\mathbf{s} | \mathbf{M}) = \min \left(\max \left(\sum_{\mathbf{a} \in \mathcal{R}_s} \frac{\mathbf{r}(\mathbf{a}; \mathbf{s})}{\|\mathbf{s}\|}; \theta_s \right); 1 \right) \quad (13)$$

For the set of 3D segments, assuming they are independent, it follows:

$$\mathbf{P}(\mathbf{S} | \mathbf{M}) = \prod_{\mathbf{s} \in \mathbf{S}} \mathbf{P}(\mathbf{s} | \mathbf{M}) \quad (14)$$

Focusing mask One measures the mask overlap by $\mathcal{H}(\mathbf{M})$ compared to the union of mask and planimetric surface, which leads to

$$\mathbf{P}(\mathbf{M}_a | \mathbf{M}) = \frac{|\mathbf{SP}(\mathbf{M}) \cap \mathbf{M}_a|}{|\mathbf{SP}(\mathbf{M}) \cup \mathbf{M}_a|} \quad (15)$$

4.4 Building Extraction and Geometric Refining

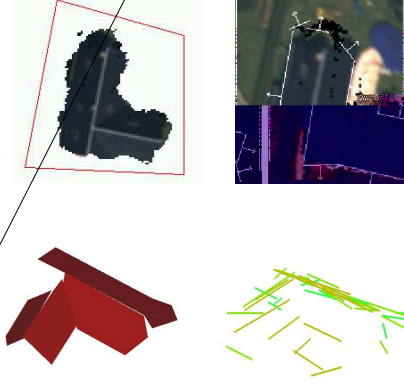
From the chosen admissible surface, it is trivial to extract connected sets of facets not touching the plane $\mathbf{z} = \mathbf{z}_g$, thus extracting only roofs structures. Let us emphasize that several buildings can be present on one focusing area as well as roofs integrating altimetric discontinuities.

The set of all admissible surfaces hypotheses is build up from an arrangement of planes in which it is difficult to handle 4 planes intersection. As a post-processing step, topological inconsistencies are corrected by a simple snapping algorithm.

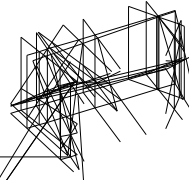
Another post-processing step enables also to enforce constraints in the real reconstruction so as to give a much more regularized shape. This important step based on (Grossmann, 2002) will not be detailed here due to lack of space.

5. RESULTS AND DISCUSSIONS

Figure 7 shows the main steps of the algorithm on an example. The projection of the result with and without enforcing constraints proves the gain of this step in the reconstruction. Figure 8 shows the results on 45 buildings with 6 images at resolution 25cm and ≈ 75 . By visual inspection, 75% of reconstructions are “acceptable”, meaning they perfectly fit the reality or the given caricature is an acceptable generalization of the reality. Right part of



(a) Focusing mask, facades, planar patches and 3D segments.



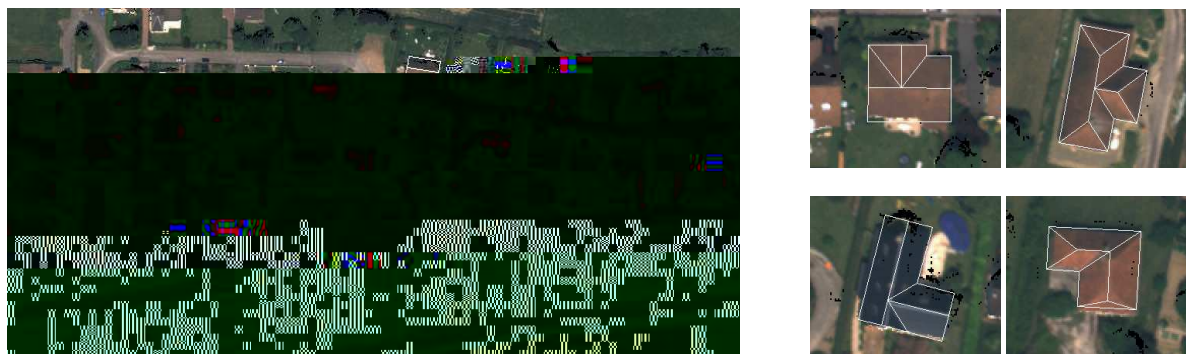


Figure 8: Some results projected on an orthophoto.



Figure 9: Influence of β .

the figure proves the ability of the algorithm to reconstruct complex buildings.

is a critical parameter. It tunes the ratio between data adequacy and caricature needs. Figure 9 shows the influence of beta on the model given by the algorithm. This parameter is to be set according to the degree of caricature desired for an application.

6. CONCLUSION AND PERSPECTIVES

We have presented a novel algorithm for automatic building reconstruction from multiple aerial images. Compared to model-based methods, it uses a very generic modeling of buildings, not restraint to a small number of shapes. Compared to data-driven approaches, external knowledge is introduced through favored constraints on primitives. Some promising results are presented that show the validity of the approach. Future work will include integration of cadastral maps in the system and evaluation of the relative importance of each type of observation in the choice of the final model. The need for heuristics to counter combinatorial explosion of brute force search will also become mandatory when dealing with bigger and more complex buildings.

REFERENCES

- Ameri, B. and Fritsch, D., 2000. Automatic 3D building reconstruction using plane-roof structures. In *ASPRS*, Washington DC.
- Baillard, C. and Dissard, O., 2000. A stereo matching algorithm for urban digital elevation models. *Photogrammetric Engineering and Remote Sensing*, 66(9):1119–1128.
- Baillard, C., Schmid, C., Zisserman, A., and Fitzgibbon, A., 1999. Automatic line matching and 3D reconstruction of buildings from multiple views. In *Proc of ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery, IAPRS*, volume 32, pages 69–80.
- Bron, C. and Kerbosch, J., 1973. Algorithm 457: Finding all cliques of an undirected graph. *Communications of the ACM*, 16:575–577.
- Fischer, A., Kolbe, T., Lang, F., Cremers, A., Förstner, W., Plümer, L., and Steinhage, V., 1998. Extracting buildings from aerial images using hierarchical aggregation in 2D and 3D. *CVIU*, 72(2):163–185.
- Flamanc, D., Maillet, G., and Jibrini, H., 2003. 3d city models: an operational approach using aerial images and cadastral maps. In Ebner, H., Heipke, C., Mayer, H., and Pakzad, K., editors, *Photogrammetric Image Analysis*, pages 53–58, Munich, Germany. ISPRS.
- Fuchs, F., 2001. *Contribution à la reconstruction du bâti en milieu urbain, à l'aide d'images aériennes stéréoscopiques à grande échelle. Étude d'une approche structurale*. PhD thesis, Université de René Descartes.
- Fuchs, F. and Le-Men, H., 1999. Building reconstruction on aerial images through multi-primitive graph matching. In *2nd IAPR Workshop on Graph-based Representations*, Vienna, Austria.
- Grossmann, E., 2002. *Maximum Likelihood 3D Reconstruction From One or More Uncalibrated Views Under Geometric Constraints*. PhD thesis, Universidade Tecnica de Lisboa.
- Heuel, S., Frstner, W., and Lang, F., 2000. Topological and geometrical reasoning in 3D grouping for reconstructing polyhedral surfaces. In *ISPRS*, volume XXIII, pages 397–404, Amsterdam.
- Jibrini, H., 2002. *Reconstruction automatique des bâtiments en modèles polyédriques 3D à partir de données cadastrales vectorisées 2D et d'un couple d'images aériennes à haute résolution*. PhD thesis, Ecole Nationale Supérieure des Télécommunications.
- Kolbe, T.-H., 1999. *Identifikation und Rekonstruktion von Gebäuden in Luftbildern mittels unsharfer Constraints*. PhD thesis, Hochschule Vechta.
- Paparoditis, N., Thom, C., and Jibrini, H., 2000. Surface reconstruction in urban areas from multiple views of aerial digital frames. In *IAPRS*, volume XXIII, Amsterdam.
- Scholz, S., Moons, T., and Van-Gool, L., 2002. A probabilistic approach to building roof reconstruction using semantic labelling. In *Proceedings of the DAGM Conference*, pages 257–264, Zürich.
- Taillandier, F. and Deriche, R., 2002. Reconstruction of 3D linear primitives from multiple views for urban areas modelisation. In *PCV'02*, volume 34:3B, pages 267–272.
- Taillandier, F., Guigues, L., and Deriche, R., 2003. A framework for constrained multi-scale range image segmentation. In *ICIP*, Barcelona.