# Automatic Modelling and 3D Reconstruction of Urban House Roofs from High Resolution Aerial Imagery

Theo Moons[1], David Frère[1], and Jan Vandekerckhove[1] and Luc Van Gool[1]

Katholieke Universiteit Leuven, ESAT / PSI,
Kard. Mercierlaan 94, B-3001 Leuven, BELGIUM
Firstname.Lastname@esat.kuleuven.ac.be

**Abstract.** Many tasks in modern urban planning require 3-dimensional (3D) spatial information, preferably in the form of 3D city models. Constructing such models requires automatic methods for reliable 3D building reconstruction. House roofs encountered in residential areas in European cities exhibit a wide variety in their shapes. This limits the use of predefined roof models for their reconstruction. The strategy put forward in this paper is, first, to construct a polyhedral model of the roof structure, which captures the topology of the roof, but which might not be very accurate in a metric sense; and then, in a second step, to improve the metric accuracy by fitting this model to the data. This decoupling of topology extraction from metric reconstruction allows a more efficient roof modelling involving less criteria. And, restricting the processing, at all stages, to one or just a few roof structures, by using a colour-based segmentation of the images, allows to use constraints that are not very tight. The approach has been tested on a state-of-the-art dataset of aerial images of residential areas in Brussels.

## 1 Introduction

Automatic generation of 3D models of buildings and other man-made structures from aerial images has become a topic of increasing importance. Although man-made objects generally exhibit quite a bit of regularity in their geometry, extracting and reconstructing buildings from aerial images is hampered by the lack of a generic 'building template'. In industrial areas or sites for official use, many flat roof and gable roof buildings are encountered. In the literature much attention has been paid to the extraction of such structures (see e.g. [3, 12–14]). House roofs encountered in residential areas in urban sites, on the other hand, show a much wider variety in their shapes. Many roofs neither are flat nor are composed of simple rectangular shapes. Model-based reconstruction now critically depends on the selection of the correct building model. Different strategies have been proposed in the literature, ranging from model selection by a human operator [7] to indexing in a model database [17]. In [5] it is observed that quite a variety of building models can be generated from a relatively small class of predefined building parts. Reconstruction is then performed by a 2-step process of hypothesis generation and verification: First, 3D feature points are grouped and related to building part primitives; and, in a second stage, these parts are combined into a building model. An alternative is to define very generic and free-form roof primitives that can be generated from the image data. In this respect, the observation that the 3D geometry of a house roof can be described as collections of line segments which tend to combine into planar structures, was used in [2] to model a house roof as a set of planar *polygonal patches*, each of which encloses a compact area with consistent photometric and chromatic properties, and that mutually adjoin along common boundaries. Since there are no constraints on the number of edges of the constituting polygons, nor on their lengths or angles, such an approach allows to model both simple as well as complicated roof structures. The strategy presented in this paper combines ideas from both approaches: First, we build, from the images, a polyhedral model of the roof structure — which captures the topology of the roof, but which might not be very accurate in a metric sense — using an hypothesis generation and verification procedure; and then, in a second step, we improve the metric accuracy by fitting this model to the data.

For the roof modelling step, we adopt the same methodology as in [2], but our implementation of the individual parts differs from that in [2] in the following respects: Firstly, it starts by delineating in the images regions that correspond to house roofs or house roof structures. Matching line segments across different views is simplified by restricting the search space to corresponding regions. Only line segments that are matched across three or more views are used for reconstruction. Moreover, coplanar grouping and polygonal patch formation also are initialized from 3D line segments that are reconstructed from those regions. In particular, polygon hypotheses are formed in 3D and verified both in 3D and by back-projecting in the images; and, if necessary, corrected accordingly. By processing one region at the time, we are able to keep the combinatorics under control. Secondly, in this paper all available views equally contribute to the reconstruction process, both on the level of 3D reconstruction itself as for polygon generation. In a next stage the different polygons are glued together into a roof model. As in [2], none of the constraints used for grouping and consistency verification are very tight. But, because the processing is restricted to regions that correspond to roof structures, we are able to retain valid object candidates. Moreover, in contradistinction to [2], the emphasis during the modelling stage is on extracting the correct topology of the roof structure, rather than on the metric accuracy of the reconstruction. This allows a more efficient roof modelling involving less criteria. Metric accuracy is obtained in an additional step by back-projecting the recovered (wireframe) model of the roof structure onto the images and minimizing the total reprojection error. This approach has been tested on a state-of-the-art dataset of aerial images of residential areas in Brussels.

The paper is organized as follows: Section 2 gives a brief overview of the different stages in the reconstruction process. Sections 3, 4 and 5 describe each of these parts in more detail. Each step is also illustrated on a real example.

Planned improvements for the implementation and possible extensions of the method are discussed in Section 6.

## 2 Method Overview

The roof modelling process is formulated as a feed-forward scheme in which 4 stages can be recognized: *2D edge detection and region selection*, *line segment matching and 3D reconstruction*, *3D grouping and polygonal patch formation*, and finally, *roof model generation and model fitting*. Observe that the first part and half of the second are purely 2D, whereas the other parts are purely 3D in nature. Obviously, these parts are not completely separate entities, but the 2D components mutually exchange data and attribute to the 3D modules. Each part will be described in more detail in the subsequent sections.

The strategy presented in this paper is tested on a state-of-the-art dataset, produced by Eurosense Blefotop n.v.. It consists of high resolution colour images of residential areas in Brussels. The image characteristics are: 1:4000 image scale and geometrically accurate film scanning with 20 microns pixel size, four-way image overlap, and precise sensor orientation. The image overlap guarantees that each building is visible in 4 to 6 images. Our method requires at least 3 views of the scene to be present.

## 3 From Images to 3D Line Segments

### 3.1 2D Edge Detection and Region Selection

The image features used in this approach are straight line segments, which are extracted from the images by running the Canny edge detector followed by straight line fitting. Due to the nature of the scene and the large size (typically $3\,K \times 5\,K$ pixels) of the aerial images, an enormous number of straight line segments is found in every image. For an efficient further processing, it is desirable to segment the image(s) or to divide the set of line segments into relatively small parts containing only a few buildings. In the literature, this is done in many different ways, ranging from manual delineation [2], possibly combined with epipolar constraints and flight information [4, 14], via perceptual grouping techniques [3, 13], to the introduction of external knowledge which, in most cases, is available from Geographical Information Systems (GIS) or digital cadaster maps [10, 15]). For the moment, we solve this problem by navigating through a constraint triangulation network which has been constructed from the line segments extracted in the image(s). Such a constraint triangulation has the advantage that the extracted line segments coincide with edges of the triangles and that the colour content of a triangle's interior is fairly homogeneous. As a house roof generally is constructed from the same roofing material, roof structures are likely to correpond to image regions with a fairly homogeneous colour distribution. Therefore, image segmentation is performed by selecting a triangle in the triangulation and growing a region from it by merging adjacent triangles that have the same mean colour vector as the selected triangle. In our implementation, the generating triangle is selected manually (with a mouse click) in each image, but this process can be automated by selecting a triangle in one image and using epipolar and trifocal constraints to delineate a window in the other images, in which a triangle with a similar mean colour vector is to be found. Since at this stage only corresponding regions are needed, it is not crucial to find the exact triangle in the other images that corresponds to the selected one. These triangles will probably not be part of the triangulation of the other images anyway. A house roof generally is composed of more than one region.

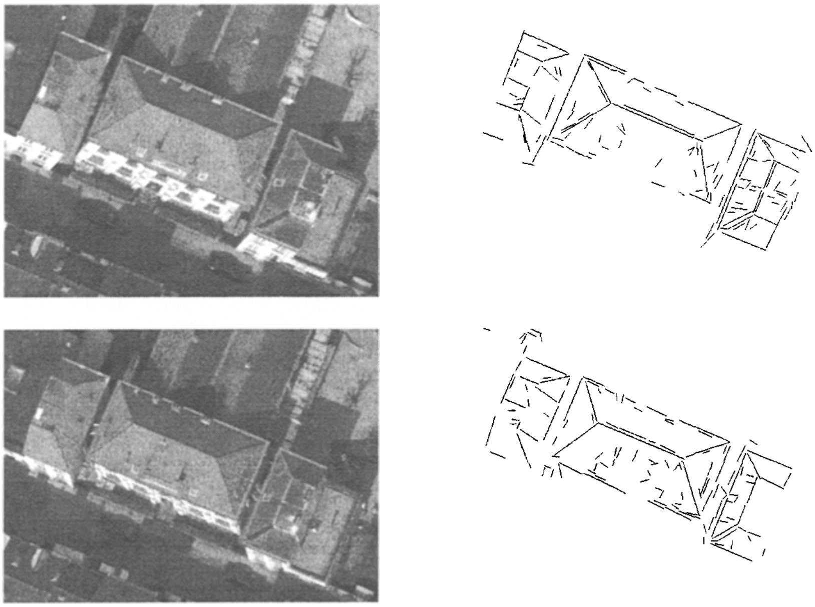

**Fig. 1.** Left : *A detail from two images in the dataset.* Right : *The edges contained in the selected regions.*

Fig. 1 (left) shows a small part of two aerial images of the same area in Brussels. The right views show the line segments which are contained in the regions that were generated by selecting 5 triangles (2 for the left building and 3 for the right one) in each image. These line segments will be used for matching and possible reconstruction in the next steps of the algorithm. Observe that the extracted regions correspond well to the actual roof surfaces and are quite stable over the images.

### 3.2 Line Segment Matching and 3D Reconstruction

Once a region of interest is constructed in every image, the longer ones among the line segments in each region are identified. As the regions are relatively small, only a small number of line segments will be selected in a region; and, because a constraint triangulation has been used, most line segments will occur at the region boundaries. Using epipolar geometry and flight information, matching line segments between corresponding regions in different images is rather easy. A complicating factor, however, is that a relatively long line segment in one image may correspond to a number of relatively short edges in another one. So, a line segment in one image must be allowed to correspond to more than one line segment in the other image. Most mismatches are then ruled out by using the trifocal constraints [11, 16] that must hold between any three views of a stationary scene. For all line segments that are matched across 3 or more views, a 3D reconstruction is computed by bundle adjustment [4] using the flight information and the calibration data of the camera.

## 4 3D Grouping and Polygonal Patch Formation

### 4.1 Coplanar Grouping

Next, the reconstructed 3D line segments are to be grouped into coplanar configurations. As the selected regions correspond to roof structures in the images, the grouping process can be restricted to the 3D line segments from one (or a few adjacent) region(s). Starting with the longest ones, 2 line segments are selected in the region(s). If the orthogonal distance between the corresponding reconstructed lines is small, a plane is constructed that fits to the line segments in a least-squares sense. In that case, the other line segments in the region(s) are tested for coplanarity with the hypothesized plane. Coplanarity is assumed if the line segment is almost parallel to the plane and if both its end points are close to the plane. All segments that satisfy these constraints are included in the defining set of the plane and the plane's equation is updated. This process is repeated until no more plane hypotheses can be formed form the selected region(s).

### 4.2 Polygon Hypothesis Generation

Subsequently, every line segment in the defining set of an hypothesized plane is projected orthogonally onto that plane. As the regions correspond well to roof structures, most plane hypotheses will correspond to planar patches of the roof structure. Thus, polygonal patch hypotheses can be formed directly from the projected segments. Unfortunately, the polygonal shapes encountered in the roof structures of urban and suburban houses — and consequently also the extracted regions — seldomly are convex, due to roof structures such as chimneys, attics, dormer-windows, etc.. So, great care should be taken when constructing the polygons. In particular, the polygon construction algorithm must try to involve as many 3D line segments in the region as possible. An initial hypothesis is formed by constructing the convex hull of the (projected) line segments. Then an initial boundary is generated by adapting the hull to include the line segments of which one of the end points contributes to the convex hull. In a next step it

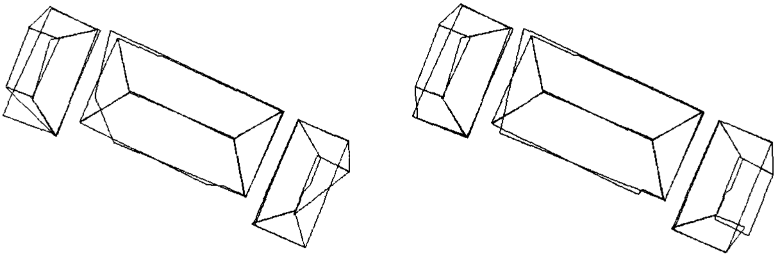

**Fig. 2.** *Initial polygon hypotheses* (left) *and after consistency verification* (right) *of selected houses in Fig. 1.*

is investigated whether this initial boundary can be adapted to incorporate line segments in the region's interior, one of whose end points is close to the boundary.

Fig. 2 (left) shows the result of coplanar grouping and the initial polygon hypotheses that are generated for two of the houses in Fig. 1. Five polygons are found for the middle house in Fig. 1: 2 triangular ones and 2 trapezia corresponding to the actual roof structure, and a horizontal plane that is formed by (parts of) the gutters of the building. Apart from a relatively small defect in the lower left corner of the horizontal plane — caused by short edge segments that could not be reconstructed properly (cf. Fig. 1 (lower right)) — the extracted polygons correspond quite well to the actual shape of the building. For the left building, also 5 polygons are constructed: 2 horizontal and 3 slanted ones. Clearly, 4 of the 5 polygons are correct, but the triangular shape of the lower side patch does not correspond to reality. The reason is that, only 1 of the 4 images used for reconstruction contains relatively complete edge information for that patch. Hence, no 3D line segments could be reconstructed for the lower left part of the patch, and the polygon construction algorithm was forced to insert wrong edge hypotheses. Similarly, only 3 of the 5 polygons that are constructed for the building on the right are correct. The 2 triangular shapes do not correspond to reality. Again, this is caused by the fact that, except in the upper image in Fig. 1, edges for the upper right corner in the other 3 views are either missing or to short to be reconstructed properly (cf. Fig. 1 (lower right)). Before glueing the polygons together, their edges should be compared with the data, and erroneous edges should be corrected. This is done in the next step.

### 4.3 Consistency Verification and Edge Correction

For each polygon that is constructed, every one of its edges is subjected to a consistency verification with respect to the 3D and 2D data. If the edge is one of the original (projected) 3D line segments, then the edge hypothesis is accepted. Otherwise, it is investigated whether the hypothesized edge is close to some reconstructed 3D line segments and has significant overlap with them. If the outcome is negative again, the edge hypothesis is back-projected into the images, and one looks for supporting information in at least 1 image. With supporting information, we mean original (possibly small) 2D line segments that are close to the projected edge and that cover at least half of the extent of the edge. It should be emphasized that the error tolerances in none of the above constraints may be very tight, in order not to rule out simple (but possibly correct) polygonal patches such as triangles, etc. too early in the process. Also observe that, since the processing is restricted to relatively small regions that correspond to actual roof structures, the number of polygons is kept under control at every stage of the reconstruction process.

Next, one tries to replace the edge hypotheses that failed for the previous consistency tests by better ones. For the moment, new hypotheses are only generated in case of a non-consistent edge for which both adjacent edges were found consistent with the data. In that case, the algorithm tries to construct a right angle. Depending on the geometric configuration of the adjacent edges, different possibilities occur, as can be seen in Fig. 3. For each possibility, its consistency with the data is verified by back-projection into the images as before, and the possibility that gets the most 2D edge support is accepted. The edges that fail for all of the previous tests finally are labeled as *doubtful.*

Fig. 2 (right) shows the result of consistency verification and edge correction of the initial polygon hypotheses of the houses in Fig. 1. Recall that the erroneous triangular shapes of the original polygon hypotheses for the left and the right roofs in Fig. 2 (left) were caused by the lack of 3D edge information among the reconstructed line segments. On the other hand, the upper image in Fig. 1 yields clear edges — and consequently, also sufficient 2D edge support — for both triangular patches of the right roof to correct the polygon hypotheses by the described procedure. And, similarly, the lower image in Fig. 1 together with another view yield sufficient 2D edge support to correct the erroneous lower left patch. Note also that the shape of the horizontal polygon of the middle house has been corrected, at least qualitatively. The orientation of the leftmost edge still deviates from reality, due to the lack of information about the position of the leftmost corner and because this edge actually connects the outer ridge of the gutter at the back of the house with the inner ridge of the gutter in the front. This, however, will be corrected in the last step of the method (cf. Section 5.3), when metric accuracy is retrieved.

## 5 Roof Model Generation and Model Fitting

The last step in the roof modelling process consists of combining the extracted polygons into roof models. A logical way to proceed would be to glue any 2 polygons together that have an edge which is the projection of one and the same reconstructed 3D line segment. Unfortunately, this is not feasible in practice. Indeed, because the polygons are constructed one region at the time, line segments at region boundaries may contribute to polygon hypotheses that are formed at different iterations of the reconstruction process. Hence, it would require a complicated data structure to record the history of every consistent polygon edge and to group the polygons accordingly. Furthermore, there not necessarily is

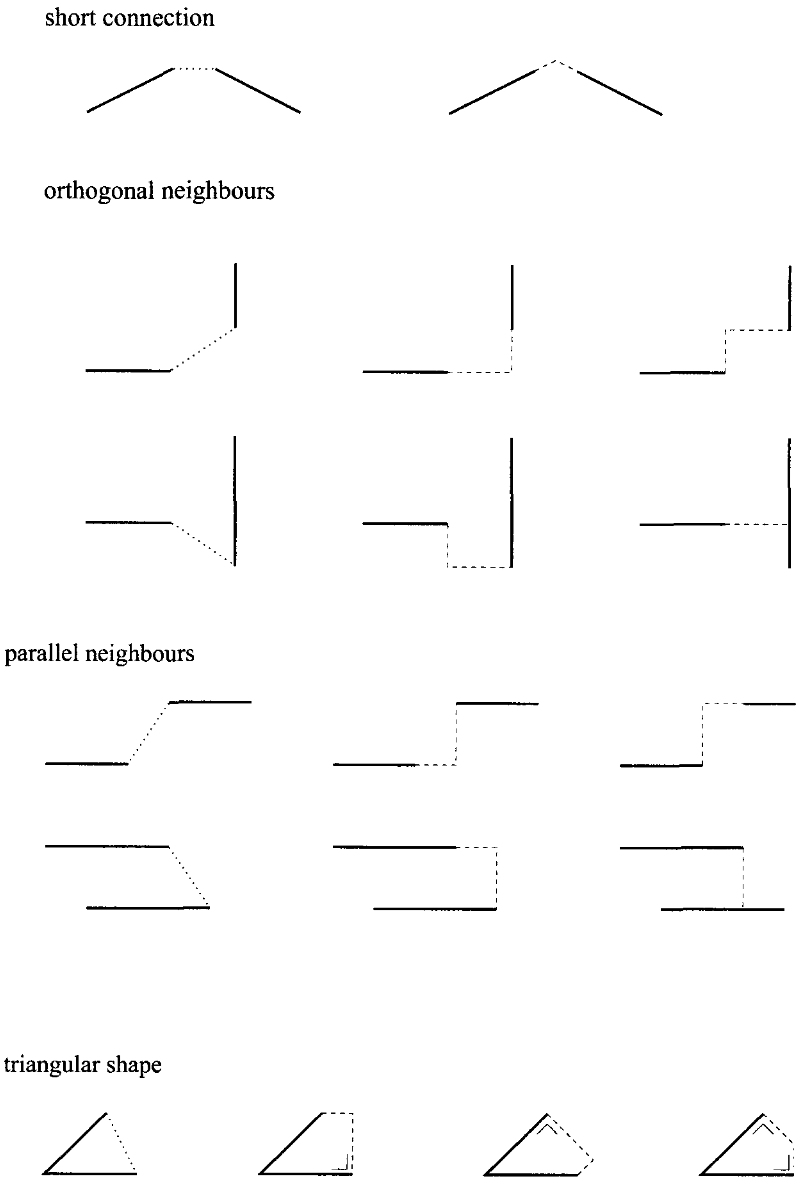

**Fig. 3.** *If an edge hypothesis is not accepted, it is replaced by new ones. The dotted line is the original edge hypothesis, full lines are its (accepted) neighbours, dashed lines are the alternative edge hypotheses.*

a reconstructed 3D line segment for every consistent edge that is constructed during the replacement of a non-consistent one. Therefore, we adopt another strategy that starts from the set of polygons *per se*. The key idea is to provide each edge of the polygons with a functional label, and to link polygons along edges bearing the same label and which nearly coincide.

### 5.1 Labeling the Edges of a Polygon

To each edge of the polygon, a label is attached according to its 3D position in the polygon. Edges whose end points have a significant difference in height above the ground level or edges that have a significant angle with the ground plane are called *slanted*. The remaining (horizontal) edges are divided into ridges and gutters. A *ridge* is a horizontal edge whose adjacent edges are both slanted and pointing downwards; whereas a *gutter* is a horizontal edge both whose adjacent edges are slanted and pointing upwards. If the plane of the polygon is horizontal, then all edges are labeled both *gutter* as well as *ridge*. Finally, a horizontal edge that does not satisfy one of the previous constraints is labeled as *ridge* (resp. as *gutter*) in case the average of the heights of its end points is greater (resp. smaller) than the average of the heights of the vertices of the polygon. If the labeling of the edges is completed, the topology of the labeled polygons is simplified by replacing any 2 adjacent edges that have the same label and are nearly collinear by one edge connecting the outer end points and having the same label.

### 5.2 Combining Polygons into Roof Surfaces

The final step in the roof modelling process is the combination of the polygons into roof models. To this end, the polygons are ordered by their area and linking starts with the polygons having the largest area. The underlying philosophy is that larger polygons generally consist of more and larger edges; and thus are more likely to correspond to actual scene structures. Two polygons will be glued together if they are not parallel, and if in both polygons an edge can be found with the same label and such that these two edges nearly coincide and have a significant mutual overlap. The actual glueing of the polygons is done by replacing these 2 edges by a segment of the line of intersection of the planes of the 2 polygons. Generally, the orthogonal projections of the end points of the 2 edges on the intersection line will not coincide. Therefore, the longest line segment that can be formed from these projections is taken as the common joint of the 2 polygons (cf. Fig. 4). This process is repeated until no more polygons can be glued together. However, there is an important geometric constraint to be taken into account, viz. each polygon can be glued *only once* to another polygon *along a given edge*. Indeed, it is unlikely for large roof structure to contain three planar faces intersecting in the same line segment. Since the larger — and hence more reliable — polygons are glued first, this constraint allows to rule out small polygons, which mostly correspond to polygonal roof structures that only have partially been recovered. Finally, all polygons that have not been joined to

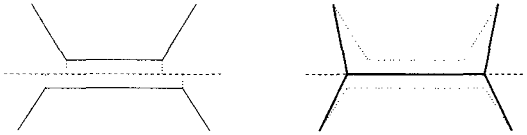

**Fig. 4.** *Two polygons are glued together along the (dashed) line of intersection of the planes of the polygons. The resulting wireframe (bold) is shown on the right.*

another one, are removed; except for large isolated horizontal planes, which may correspond to a flat roof. Their analysis will be discussed in a subsequent paper.

Fig. 5 shows the roof models that are obtained by applying this procedure to the polygon hypotheses depicted in Fig. 2 (right). The resulting wireframe models clearly are topologically correct.

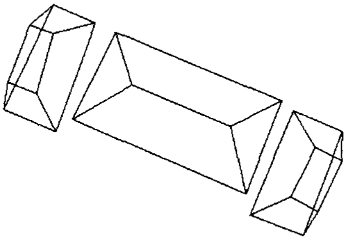

**Fig. 5.** *The roof models obtained by combining the polygonal patches shown in Fig. 2.*

### 5.3 Fitting the Wireframe Model to the Image Data

During the 3D grouping and the roof modelling stages of the algorithm, the metric accuracy of the reconstruction has been neglected. Indeed, a 3D line segment that originally was reconstructed from the images by bundle adjustment, first is orthogonally projected onto one or more planes in order to form polygonal patches, and then in the glueing process it is replaced by the line of intersection of the planes of the polygons to which it contributes. Moreover, the strategy expounded above fully exploits the polyhedral structure of the observed scene, an assumption that may not strictly be satisfied in reality. As a consequence, the computed roof model generally will deviate from the actual roof structure in the scene. This can clearly be observed when back-projecting the roof model into the images. For example, the left image in Fig. 6 shows the back-projection of one of the models in Fig. 5 into the lower image of Fig. 1. It immediately catches the eye that the projected gutters of the model do not coincide with those of the house in the image. But there also is a deviation in the position of the projected ridge and in the orientation of the projected rafters at the back of the roof model

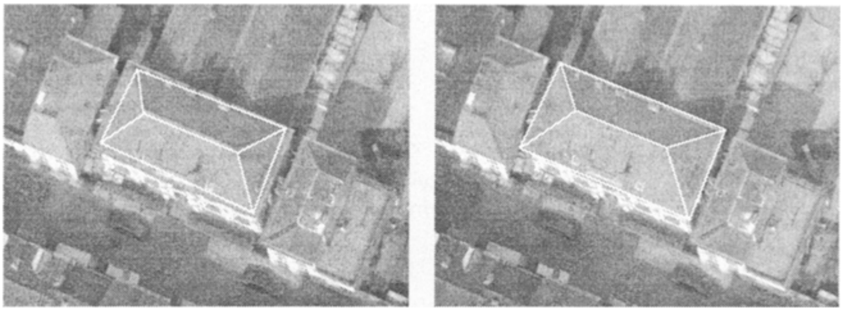

**Fig. 6.** Left: *Back-projecting the roof model of Fig. 5 into the original images illustrates the loss of metric accuracy during the 3D grouping and roof modelling process.* Right: *Metric accuracy can be restored by changing the wireframe model so as to minimize the reprojection error in all the available images.*

with respect to those in the image. However, if the roof models constructed by the above procedure are primarily considered as wireframe models capturing the topological structure of the building, then a metric 3D reconstruction of the scene is to be obtained by fitting these models to the data. Generally speaking, there are two alternatives for fitting a 3D model: either the model is fitted directly to a 3D reconstruction of the scene, or the model is projected into the images and adjusted to fit the image data. Since our roof model has been constructed from the reconstructed 3D line segments, fitting must be done with respect to the original image data in order to use maximally independent information. Therefore, the roof model, as computed in the previous sections, is converted into a *wireframe model*, meaning that it is represented by a set of 3D points, called *vertices*, that are connected by straight line segments, called *edges*. Recall that the roof model constructed above actually has a polyhedral structure. Therefore, for each vertex not only a list of all the edges ending in that vertex is constructed, but also a list of all the planes of the polygonal faces that contain the given vertex is provided. This wireframe is then projected into all the images by using the flight information and the calibration data of the camera. The idea now is to adapt the 3D position of the vertices of the wireframe so as to bring the projected edges in as good as possible accordance with the observed image data. To this end, an *energy functional* is introduced, which measures the average image gradient along the projected edges of the wireframe. More precisely, for each vertex $V_j$ of the wireframe, let $\mathcal{E}_j$ be the set of edges of which $V_j$ is an end point, and let $\mathcal{P}_j$ denote the set of the planes of the polygonal faces that contain $V_j$. Furthermore, let $L_i$ denote the intensity distribution (luminance function) of the $i$th image ($i = 1, \ldots, m$). Then the *total mean gradient energy* $G(V_j)$ of the vertex $V_j$ is given by

$$G(V_j) = \frac{1}{m}\sum_{i=1}^{m}\left\{\frac{1}{|\mathcal{E}_j|}\sum_{E\in\mathcal{E}_j}\frac{1}{\ell(\pi_i(E))}\int_{\pi_i(E)}\|\nabla L_i\|\right\} , \qquad (1)$$

where $|\mathcal{E}_j|$ is the number of elements in the set $\mathcal{E}_j$ (i.e. the number of edges ending in $V_j$), $\pi_i(E)$ is the projection of the line segment $E$ in the $i$th image, $\ell(\pi_i(E))$ is the length of the projected line segment $\pi_i(E)$, and $||\nabla L_i||$ is the norm of the gradient $\nabla L_i$ of the image intensity function $L_i$. To meet the image data, the vertices of the wireframe model should be placed at those 3D positions for which the total mean gradient energy $\frac{1}{n}\sum_{j=1}^{n} G(V_j)$ is maximal. Here $n$ is the number of vertices in the wireframe model. An important observation here is that, as each vertex is free to move in any 3D direction, the resulting wireframe that maximizes the total mean gradient energy generally will not define a polyhedral roof structure anymore. Theoretically, this need not to be a disadvantage, as most house roofs probably are not strictly polyhedral either. From a practical point of view, however, it is not advisable to allow the wireframe to deviate too much from a polyhedral shape; the reason being that in case a vertex of the wireframe happens to be occluded in one or more views, the 3D position of that vertex may drift away in order to bring the projected edges closer to (non-corresponding) image contours. One way to avoid this undesired behaviour is to use the median instead of the mean in the definition of the gradient energy $G(V_j)$. However, in practice this results in fitting the wireframe model to just one or two images, whereas our strategy is to use information from as many image as possible. Instead, we introduce an extra energy functional which measures the deviation of the vertex position from its polyhedral position. In particular, the *distance energy* $D(V_j)$ of the vertex $V_j$ is defined as the sum of the distances $d(V_j, P)$ of $V_j$ to each of the planes $P \in \mathcal{P}_j$ to which it originally belonged:

$$D(V_j) = \frac{1}{|\mathcal{P}_j|} \sum_{P \in \mathcal{P}_j} d(V_j, P) \ . \tag{2}$$

Another possibility is to update the equation of the planes $\mathcal{P}$ at each step of the iteration. This is computationally more demanding, and, in practice, yields the same results.

Together, $G$ and $D$ yield the following *energy functional* $E$ that has to be maximized in order for the constructed wireframe to fit the image data:

$$E(V_1, V_2, \ldots, V_n) = \frac{1}{n} \sum_{j=1}^{n} \{ G(V_j) - \beta\, D(V_j) \} \ , \tag{3}$$

where $0 < \beta < 1$ is a real number that regulates the strength of the distance energy $D$ in the energy functional $E$. To find the maximum of the energy functional $E$, the maximization problem is replaced by a dynamical system whose equilibrium states coincide with the critical points of the energy functional $E$. The critical points of $E$ are the point positions $(V_1, V_2, \ldots, V_n)$ at which the gradient $\nabla E$ vanishes: $\nabla E(V_1, V_2, \ldots, V_n) = 0$; or equivalently, those points $V_j$ for which

$$\nabla G(V_j) - \beta \nabla D(V_j) = 0 \qquad (j = 1, \ldots, n). \tag{4}$$

The same equations give the equilibrium states of the dynamical system

$$\frac{\partial V_j}{\partial t} = \alpha\, [\nabla G(V_j) - \beta \nabla D(V_j)] \qquad (j = 1, \ldots, n), \tag{5}$$

with $\alpha$ a positive real number. Since by construction the vertices of the wireframe computed in the previous sections are relatively close to a maximum of the energy functional $E$, starting the dynamical process (5) from these vertex positions will lead the vertices of the wireframe to those positions that correspond to the desired maximum of $E$. The main advantage of this approach is that it can be implemented as an iterative process by using a finite difference approximation of the system (5). At each iteration step, the vertex positions $V_j^{(p)}$ are all updated to new positions

$$V_j^{(p+1)} = V_j^{(p)} + \alpha \left[ \Delta G_j^{(p)} - \beta \, \Delta D_j^{(p)} \right] \ , \tag{6}$$

where $\Delta G_j^{(p)}$ and $\Delta D_j^{(p)}$ are the finite difference approximations of the gradients $\nabla G$ and $\nabla D$ at vertex position $V_j^{(p)}$. Note that the computation of $\Delta G_j^{(p)}$ involves the evaluation of the functional $G$, and hence, the calculation of the line integrals $\int_{\pi_i(E)} \|\nabla L_i\|$ (cf. eq. (1)). These integrals are computed by sampling the gradient image along the given line segment and using Gaussian interpolation for obtaining sub-pixel accuracy.

The right image in Fig. 6 illustrates the effect of applying this procedure to one of the wireframes in Fig. 5. The resulting 3D reconstruction agrees much better with the image data, as can be seen in Fig. 6 (right) which shows the back-projection of the optimized wireframe into the second image of Fig. 1. In particular, the horizontal edges of the reconstruction coincide quite well now with the outer edges of the gutters in the image(s); and the projections of the ridge and the rafters of the model coincide with those in the image(s) as well.

From the reconstructed roofs a 3D model of the buildings can be generated by adding artificial vertical walls. This is done by constructing vertical line segments that connect the end points of the gutters with the ground plane. The ground plane is extracted automatically by selecting regions (i.e. triangle(s)) in the images that correspond to the street level, reconstructing the line segments contained in it as described in Section 3, and performing a coplanar grouping as outlined in Section 4.1. Fig. 7 illustrates the result of this operation by showing two views from different viewing directions of the houses in the images of Fig. 1, whose roofs were reconstructed by the method described above. The texture of the roof surface, as seen in the first image of Fig. 1, has been mapped onto the roofs in the reconstruction.

Fig. 8 (upper left) shows another detail of (one of) the aerial image(s). Observe the a-symmetrical shape of the corner building. Its roof is reconstructed from three regions: one large region covering the roof patches along the street sides, and two small ones corresponding to the upper right faces. Note that the patches contained in the large region actually belong to 3 different planes. Moreover, no 3D line segments could be recovered for the triangular patch at the corner, because its projection in the images consists of short edge segments only. So, initially, only two polygons are constructed: a trapesium for the left part of the roof, and a triangle for the lower right patch at the street side. The scattered edge information of the corner patch, however, provides sufficient support for the

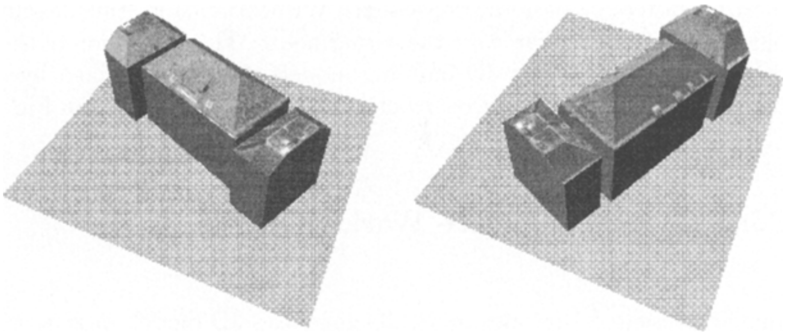

**Fig. 7.** *Two views of the houses reconstructed from the images in Fig. 1.*

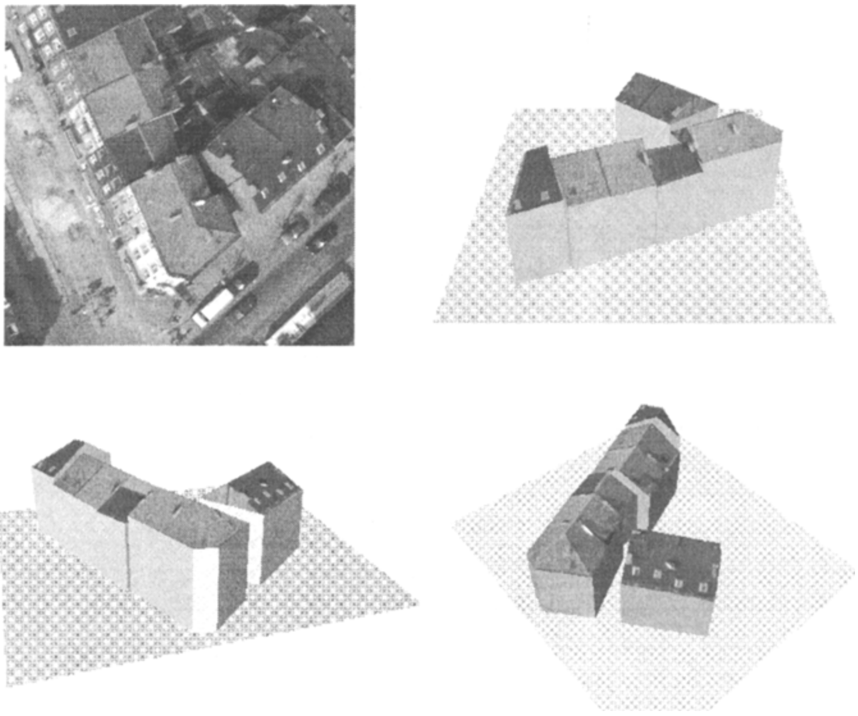

**Fig. 8.** *Another detail from an aerial image in the dataset and three views of the re·constructed houses.*

hypothesized rafters to pass the consistency verification step, thus leading to a topologically correct wireframe for the corner house. The extraction of the other roofs is straightforward. The 3D building models that are obtained by adding artificial vertical walls to the reconstructed roofs can also be seen in Fig. 8.

## 6 Conclusion and Future Work

A method is presented that automatically generates 3D models of generic house roofs from aerial images of residential areas in urban sites. Describing the 3D geometry of a house roof as collections of line segments which tend to combine into planar structures allows to model both simple as well as complicated roof structures by the same approach. Crucial to the method is the possibility of delineating regions in the images that correspond well to actual roof structures. This can be automatically done by introducing external information which often is available from Geographical Information Systems (GIS) or digital cadaster maps (cf. [15]). In our current implementation, colour-based image segmentation is performed by navigating through a constraint triangulation network that has been constructed from the line segments extracted in the image(s). Restricting the processing to relatively small regions allows at all stages of the algorithm to use constraints that are not very tight, and, at the same time, to keep the combinatorics under control. Furthermore, all modelling is done by reasoning in 3D. By adopting a strategy of hypothesis generation and verification, we not only are capable of using all the available image data at every step in the algorithm, but also to treat all views at the same level. Moreover, by decoupling the retrieval of the topology of the roof structure from the metric accuracy of the reconstruction, it is possible to generate and test combinations which otherwise would have been ruled out by more tight constraints.

The method is implemented and the first tests are performed on a state-of-the-art dataset containing aerial images of residential areas in Brussels. As can be seen from the examples above, it yields acceptable results. Of course, there still is a lot of work to be done. First of all, attention will be paid to optimizing and finetuning each of the different steps in the algorithm. On the theoretical level, more sophisticated techniques which have proven successful in a lot of other applications will be considered. Examples include the use of the Minimum Description Length principle for coplanar grouping and the exploitation of observed symmetry relations between different polygons in the glueing process. Another important issue that will be investigated is the formulation of a set of consistent geometric adjacency relations both for the 3D model and its projections in the images in order to obtain a quantitative assessment of the correctness and completeness of the constructed roof model. But the final test, of course, is an extensive evaluation of the accuracy and the completeness of the reconstructed house roof models with respect to the ground truth. Such a test is planned for the near future.

### Acknowledgements

This work is support by the EU ESPRIT ltr 20.243 'IMPACT' project. The authors thank Eurosense Belfotop n.v. (Wemmel, Belgium) for providing the dataset of aerial images. Special thanks also goes to Marc Proesmans for valuable discussions and advice on the implementation of the wireframe fitting algorithm.

## References

1. E.P. Baltsavias, W. Eckstein, E. Gülch, M. Hahn, D. Stallmann, K. Tempfli, and R. Welch (eds.), *3D Reconstruction and Modelling of Topographic Objects*, International Archives of Photogrammetry and Remote Sensing, Vol. **32** Part **3-4W2**, International Society of Photogrammetry and Remote Sensing, Sydney, 1997.
2. F. Bignone, O. Henricsson, P. Fua and M. Stricker, Automatic extraction of generic house roofs from high resolution aerial imagery, in : B. Buxton and R. Cipolla (eds.), *Computer Vision - ECCV'96*, Lecture Notes in Computer Science **1064**, Springer-Verlag, Berlin, 1996, pp. 85–96.
3. R.T. Collins, A.R. Hanson, M.R. Riseman and H. Schultz, Automatic extraction of buildings and terrain from aerial images, pp. 169–178, in [8].
4. O. Faugeras, S. Laveau, L. Robert, G. Csurka and C. Zeller, 3-D reconstruction of urban scenes from sequences of images, pp. 145–168, in [8].
5. A. Fischer, T.H. Kolbe and F. Lang, F., Integration of 2D and 3D Reasoning for Building Reconstruction Using a Generic Hierarchical Model, pp. 159–180, in [6].
6. W. Förstner, F. and L. Plümer, *Semantic Modeling for the Acquisition of Topographic Information from Images and Maps*, Birkhäuser-Verlag, Basel, 1997.
7. E. Gülch, Application of semi-automatic building acquisition, pp. 129–138, in [9].
8. A. Grün, O. Kübler and P. Agouris, *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Birkhäuser-Verlag, Basel, 1995.
9. A. Grün, E.P. Baltsavias and O. Henricsson, *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, Birkhäuser-Verlag, Basel, 1997.
10. N. Haala, C. Brenner and K.-H. Anders, Generation of 3D city models from Digital Surface Models and 2D GIS, pp. 68–76, in [1].
11. R. Hartley, Lines and points in three views — a unified approach, *Proc. ARPA Image Understanding Workshop* (IUW '94), Monterey, CA, 1994, pp. 1009–1016.
12. J. Mc Glone and J. Shuffelt, Projective and object space geometry for monocular building extraction, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition* (CVPR '94), Seattle, Washington, June 1994, pp. 54–61.
13. R. Nevatia, C. Lin and A. Huertas, A system for building detection from aerial images, pp. 77–86, in [9].
14. M. Roux and D. Mc Keown, Feature matching for building extraction from multiple views, *Proc. ARPA Image Understanding Workshop (IUW'94)*, Monterey, CA, 1994, pp. 331–349.
15. M. Roux and H. Maître, Three-dimensional description of dense urban areas using maps and aerial images, pp. 311–322, in [9].
16. A. Shashua, Algebraic functions for recognition, *IEEE Trans. on Pattern Analysis and Artificial Intellingence* (T-PAMI), Vol. **17** (1995), no. 8, pp. 779–789.
17. F. Stolle, A. Hanson, C. Jaynes, E. Riseman, and H. Schultz, Scene Reconstruction research — towards an automatic system, pp. 33–42, in [9].