A PLANE-SWEEP STRATEGY FOR THE 3D RECONSTRUCTION OF BUILDINGS FROM MULTIPLE IMAGES

C. Baillard and A. Zisserman
Dept. of Engineering Science, University of Oxford,
Oxford OX13PJ, England
{caroline,az}@robots.ox.ac.uk

ABSTRACT

A new method is described for automatically reconstructing a 3D piecewise planar model from multiple images of a scene. The novelty of the approach lies in the use of inter-image homographies to validate and best estimate planar facets, and in the minimal initialization requirements — only a single 3D line with a textured neighbourhood is required to generate a plane hypothesis. The planar facets enable line grouping and also the construction of parts of the wireframe which were missed due to the inevitable shortcomings of feature detection and matching. The method allows a piecewise planar model of a scene to be built completely automatically, with no user intervention at any stage, given only the images and camera projection matrices as input. The robustness and reliability of the method are illustrated on several examples, from both aerial and interior views.

1 INTRODUCTION

Automating reconstruction from images is one of the continuing goals in photogrammetry and computer vision. The special case of piecewise planar reconstruction is particularly important due to the large number of applications including: manufactured objects, indoor environments, building exteriors, 3D urban models, etc.

The target application of this paper is the 3D reconstruction of roofs of urban areas from aerial images, but the method is not restricted to this case. Recently, the massive development of telecommunication networks has even further increased the need for such urban databases. The difficulty of reconstruction in urban environments is mainly due to the complexity of the scene – the built-up areas are often very dense and involve very many types of buildings. Images of these areas are very complex and image boundaries often have poor contrast. All of these factors make automating reconstruction even more difficult.

One approach to reconstruction is to compute a dense Digital Elevation Model (DEM) using matching techniques based on cross-correlation (Berthod et al., 1995, Cord et al., 1998, Girard et al., 1998). The DEM is then segmented in order to provide a 3D delineation (boundaries) of the buildings (Weidner, 1996, Paparoditis et al., 1998, Baillard and Maître, 1999). However, the elevation maps resulting from stereo matching are generally insufficiently accurate or complete to enable the precise shape of buildings to be recovered. Thus most approaches have focused on the reconstruction of specific building models, using strong prior knowledge about the expected 3D shape: rectilinear shapes (McGlone and Shufelt, 1994, Roux and McKeown, 1994, Noronha and Nevatia, 1997, Collins et al., 1998), flat roofs (Berthod et al., 1995), or parametric models (Haala and Hahn, 1995, Weidner and Förstner, 1995). These models can obviously not cover all buildings present in a dense urban environment. More generic reconstruction can be achieved by employing simpler and less restrictive models but using using multiple high-resolution images (Bignone et al., 1996, Moons et al., 1998). These approaches generally rely on the detection and the grouping of neighbouring coplanar 3D lines computed from the images. However, due to the occurence of image boundaries with low contrast, feature detectors often fragment or miss boundary lines, and only an incomplete 3D wireframe can be obtained.

The problems caused by missing features in piecewise planar reconstruction are illustrated by the detail in figure 6a taken from the image set of figure 1. The correct roof model in this case is a four plane "hip" roof (Weidner and Förstner, 1995). However, the oblique roof ridges are almost invisible in any view, and certainly are not reliably detected by an edge or bar detector with only local neighbourhood support. Consequently, classical plane reconstruction algorithms which proceed from a grouping of two or more coplanar 3D lines will produce a flat roof, or at best a two plane "gable" roof if the central horizontal ridge edge is detected — however the two smaller faces will be missed.

This paper presents an approach to solving this problem, with a new method for computing planar facets starting from an incomplete wireframe of 3D lines. The key idea is that both geometric and photometric constraints should contribute from all images. The 3D planes defining the model are therefore determined by using both 3D lines (geometric features) and their image neighbourhoods over multiple views (photometric features). This is achieved through a *plane-sweep*

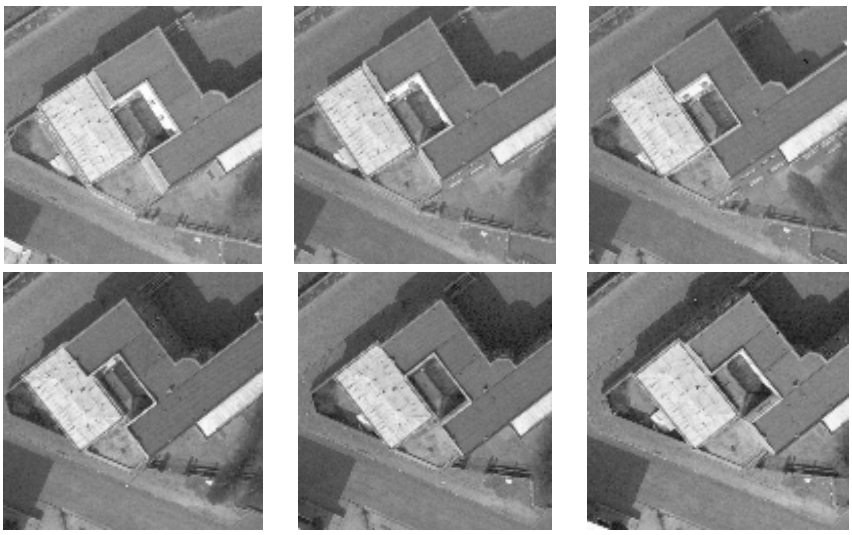


Figure 1: Six overlapping aerial views. The images are 600×600 pixels, one pixel corresponding to a ground length of 8.5cm.

strategy (Collins, 1996, Seitz and Dyer, 1997, Coorg and Teller, 1999). A set of planar facets constrained by the 3D lines is hypothetized in space and plausible plane hypotheses are identified by checking similarity over multiple images. The particular novelty of the approach is in the use of inter-image homographies (plane projective transformations) to robustly estimate the planar facets. These facets then enable both line grouping and, by plane intersection, the creation of lines which were missed during feature detection. The approach requires minimal image information since a plane is generated from only a line correspondence and its image neighbourhood. In particular two lines are not required to instantiate a plane. These minimal requirements and avoidance of specific object models facilitate the automatic reconstruction of objects with quite subtle geometry located within a complex environment.

This paper is organized as follows. Section 2 introduces the input data (image set and 3D lines) and overviews the method. Then the three stages of the piecewise planar reconstruction process are detailed in sections 3– 5 respectively, with results given in section 6 for a variety of scene types.

2 OVERVIEW OF THE METHOD

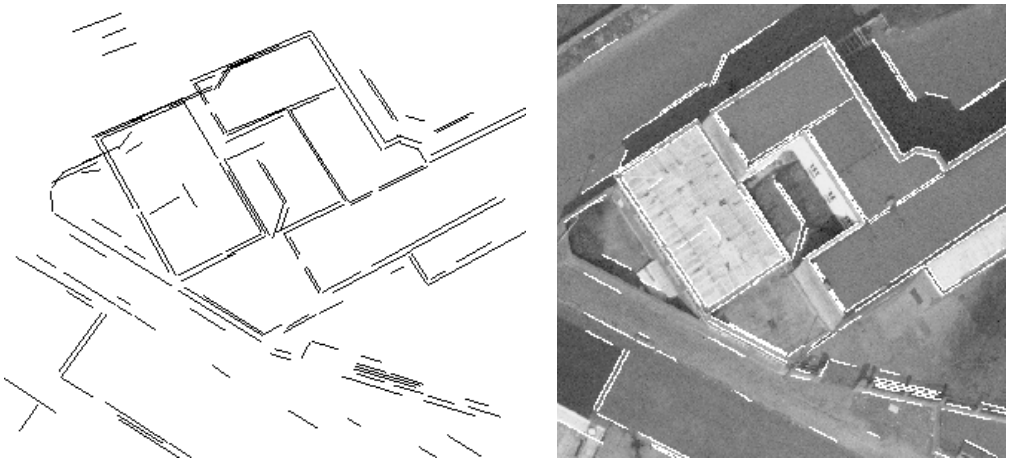
2.1 Input Data

Out target application is the automatic acquisition of 3D scene models of urban and suburban areas from aerial imagery. The typical input data consist of high resolution quasi-nadir images, with a resolution of $8.5 \times 8.5\ cm^2$ on the ground (see the example of figure 1). From these images only roof planes of buildings can be accurately extracted since vertical walls are generally not visible. At least 3 images taken from different positions are available of the scene. The camera projection matrix is known for each view. In this case the projection matrices are metric calibrated. However, the method requires only projective information, since it only involves image to image homographies induced by a plane, and the multiple view geometric relations given by the fundamental matrix and trifocal tensor (Spetsakis and Aloimonos, 1990, Shashua, 1994, Hartley, 1995).

Importantly, the scenes are of higher complexity than those that have traditionally been studied. In European urban and suburban areas buildings are often packed together and of variable types and irregular shapes. Figures 19 and 21 show typical examples of overlapping views. For clarity reasons, each step of the method will be illustrated on the smaller images shown in figure 1.

2.2 General Strategy

Although there is a wide diversity in the shapes of the individual buildings, almost all of them can initially be described as polyhedral structures. Hence, *coplanarity* of points and lines is the dominant criterion present in the grouping and modelling algorithms. Our general strategy is to first generate 3-dimensional information and perform all grouping and modelling operations in the 3-dimensional world. There is an abundance of geometric primitives in the individual images, and great care must be taken to generate reliable 3D information. In order to achieve this, our approach goes further than the traditional 2-view stereo approach and exploits the redundancy in such image sets to the full. Any 3D information is verified over multiple images with typically 4 to 6 images being used. In addition, both geometric and photometric clues



(a) (b)

Figure 2: Line matching. (a) 137 lines are matched automatically over 6 views. Their 3D position (shown) is determined by minimizing reprojection error over each view in which the line appears. (b) The lines projected onto the first image of figure 1.

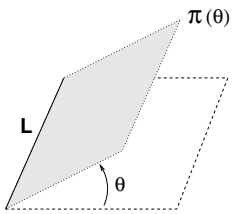


Figure 3: The one-parameter family of half-planes containing the 3D line **L**. The family induces a one-parameter family of homographies between any pair of images.

are exploited simultaneously. The method described here starts from multiple images and a set of 3D lines. Note, the production of 3D lines is out of the scope of this paper, and only the key ideas of the algorithm are summarized in the next subsection.

2.3 Production of 3D lines.

The 2D image lines are obtained by applying a local implementation of the Canny edge detector (with subpixel accuracy), detecting tangent discontinuities in the edgel chains, and finally straight line estimation by orthogonal regression. Then lines in 3D are generated by using an implementation of the line matching algorithm for 3 views described in (Schmid and Zisserman, 1997). Matches are disambiguated by a geometric constraint over 3 views (using epipolar geometry and trifocal geometry), together with a photometric constraint based on line intensity neighbourhoods. In addition, fragmented lines can be joined and extended when there is photometric support over the views. Here the line matching has been extended to six views (Baillard et al., 1999). Figure 2 shows the result of the line matching on the data set of figure 1. Note that some of the scene lines are missing, and some of the recovered lines are fragmented.

2.4 Method for producing piecewise planar models.

The overall algorithm consists of three main stages, which will be illustrated on the building of figure 6a:

1. *Computing reliable half-planes* defined by one 3D line and similarity scores computed over all the views (section 3). This is the most important and novel stage of the algorithm.

2. *Line grouping and completion* based on the computed half-planes (section 4). This involves grouping neighbouring 3D lines belonging to the same half-plane, and also creating new lines by plane intersection.

3. *Plane delineation and verification* where the lines of the previous stage are used to delineate the plane boundaries (section 5).

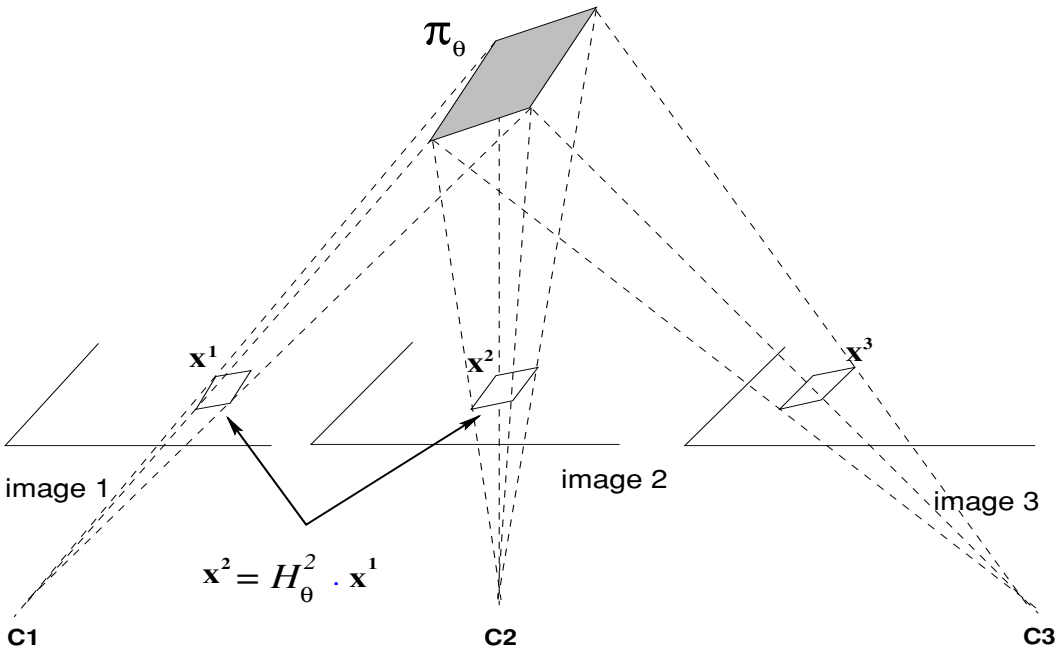


Figure 4: Geometric correspondence between views. Given θ, the homography $\mathtt{H}^i(\theta)$ determines the geometric map between a point in the first image and its corresponding point in image i.

3 COMPUTING HALF-PLANES

3.1 Principles and objectives

Given a 3D line, there is a one-parameter family of planes $\boldsymbol{\pi}(\theta)$ containing this line (see figure 3). As each plane defines a (planar) homography between two images, the family also defines a one-parameter family of homographies $\mathtt{H}(\theta)$ between any pair of images. Note, each side of the line can be associated with a different half-plane. Our objective is therefore to determine for each line side whether there is an attached half-plane or not, and if there is we want to compute a best estimate of θ. For that purpose we use a plane-sweep strategy: for each angle θ, we are hypothesizing a planar facet attached to the line, and verifying or refuting this model hypothesis using image support over multiple views. We wish to employ only the minimal information of a single 3D line and its image neighbourhood.

3.2 Point to point correspondence

The existence of an attached half-plane and a best estimate of its angle is determined by measuring image support (defined in the next subsection) over multiple views according to the geometry illustrated in figure 4. Given θ, the plane $\boldsymbol{\pi}(\theta)$ defines a point to point map between the images. If the plane is correct then the intensities at corresponding pixels will be highly correlated.

In more detail, the point to point correspondences are computed over the image set according to the homographies defined by θ. Given the plane $\boldsymbol{\pi}(\theta)$ there is a homography represented by 3×3 matrix $\mathtt{H}^i(\theta)$ between the first and ith view, so that a point $\mathbf{x}$ in the first image is mapped to $\mathbf{x}^i$ in the ith image according to:

$$\mathbf{x}^i = \mathtt{H}^i \mathbf{x}, \tag{1}$$

where $\mathbf{x}$ and $\mathbf{x}^i$ are represented by homogeneous 3-vectors.

The homography matrix is obtained from the 3×4 camera projection matrices for each view. For example, if the projection matrices for the first and ith views are $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$ and $\mathtt{P}^i = [\mathtt{A}^i \mid \mathbf{a}^i]$ respectively, and 3D points $\mathbf{X}$ on the plane satisfy $\boldsymbol{\pi}^\top \mathbf{X} = 0$, where the plane is represented as a homogeneous 4-vector $\boldsymbol{\pi}$ in the world frame, then (Luong and Viéville, 1996):

$$\mathtt{H}^i = \mathtt{A}^i + \mathbf{a}^i \mathbf{v}^\top \text{ where } \mathbf{v} = -\frac{1}{\pi_4} (\pi_1, \pi_2, \pi_3)^\top$$

provided $\pi_4 \neq 0$. Note, $\mathbf{v}$ is independent of the view i.

3.3 Similarity score function

The correlation of the image patches mapped by the homographies $H^i(\theta)$ is assessed by the following similarity score function:

$$Sim(\theta) = \sum_{I^i \text{valid},\, 1 \leq i \leq n} \int_{\text{POI}_L^0} w_L(\mathbf{x}) Cor^2(\mathbf{x}, \mathtt{H}^i(\theta)\mathbf{x}) d\mathbf{x} \tag{2}$$

and ranges between $(0, 1)$. This function has been designed to be selective, and also robust to occluded portions and irrelevant points. The design of this equation is explained below.

First, correlation is computed only in the neighbourhood of textured points. The set of points of interest in the i^{th} image is determined with respect to the reference view I^0, as the image of POI_L^0 by the homography $\mathtt{H}^i(\theta)$. The set POI_L^0 is computed by applying an edge detector with a very low threshold on gradient (an example of detection is given in figure 6). The edges are then linked and regularly sampled over a topological neighbourhood $\mathcal{V}_L$ of the line $\mathbf{L}$ projected in the image. This neighbourhood is determined using a Delaunay triangulation constrained to fit the projected line segments.

Since no particular view should have a special role, the reference view is automatically selected for each line side. A set of points of interest is detected in each image, then the most textured image (i.e., providing the largest number of points of interest) is selected as the reference. The use of the largest number of textured points produces a selective and discriminating similarity function of θ - when the intensity of the image is locally homogeneous, correlation between images is similar for any θ. However, at locally textured regions this problem will not arise.

The role of the weighting factor $w_L(\mathbf{x})$ is to take into account the likelihood that a point $\mathbf{x}$ from POI_L^0 actually describes the planar facet. This is necessary since the topological neighbourhood $\mathcal{V}_L$ over which POI_L^0 is defined is not guaranteed to exactly correspond to the planar facet. Thus $w_L(\mathbf{x})$ has been defined as:

$$w_L(\mathbf{x}) = \frac{1}{D_L^0(\mathbf{x}, \mathbf{L})},$$

where $D_L^0(\mathbf{x}, \mathbf{L})$ is the distance of the point $\mathbf{x}$ from the line $\mathbf{L}$ projected onto the reference view. This weighting gives more weight to points which are closer to the line, and consequently more likely to belong to the considered plane. Additional robustness is provided by only including *valid* views in the summation. Valid views are those which have a sufficient number of high correlation scores at points of interest, thereby rejecting views where the plane might be occluded.

Figure 5 shows two typical examples of score functions. Averaging the scores over views exploits the complementarity of the short and wide baseline separations (see figure 7) in the data set.

Probabilistic interpretation. The similarity score function (2) may be thought of as a log likelihood function on θ: an evaluation of the function at a particular POI is equivalent to a likelihood for that point. Each of these evaluations may be treated as independent, so that the overall likelihood of θ is the product of the likelihoods of each point. Taking logs in the usual manner then results in the summation in $Sim(\theta)$.

3.4 Optimization

The optimal angle θ is the one which maximizes the function $Sim(\theta)$ over the range $[\theta_{min}; \theta_{max}]$ (chosen as $[-75^o; +75^o]$ for our application). This maximum is determined using the Newton's method in order to gain time (correlation over multiple views is quite expensive). First the similarity function is evaluated over a regular sample of values located within the range, which gives a coarse estimate of the optimal angle. The similarity is then refined around this value by removing outsiders (points providing very bad scores), which is important in case of occluded portions. Then a parabolic estimation around the maximum computed score is performed, leading to a new best estimate of θ. The operation is iterated until uncertainty about the optimal angle is less than a predefined threshold (practically chosen as 1^o).

The corresponding half-plane hypothesis is accepted or rejected as valid according to the characteristics of $Sim(\theta)$, as shown in figure 5: the maximum value of $Sim(\theta)$ and the absolute value of the estimated second derivative around the maximum must be above a certain threshold (chosen quite low to keep many hypotheses). For example, an occluding edge would not have a half-plane attached on the occluded side. The line side is thus classified as supporting or not supporting a half-plane, with a number of reliability indicators: maximum value (confidence), second derivative around the maximum (accuracy), as well as number of points of interest used for computation.

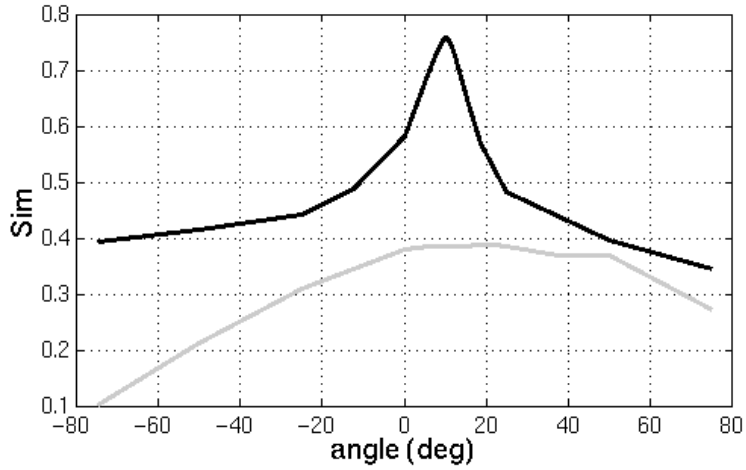
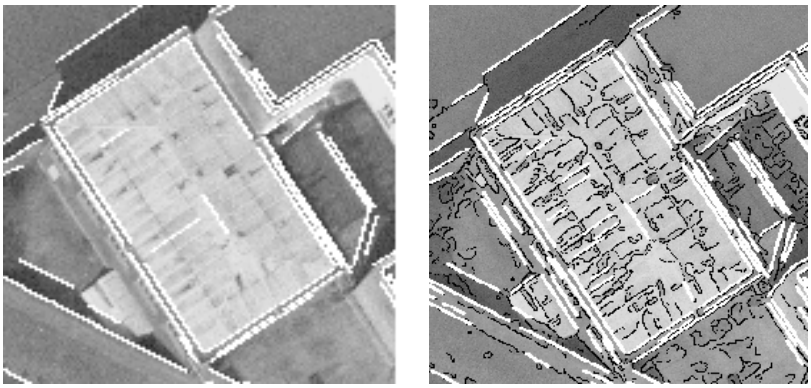


Figure 5: Example of similarity score functions $Sim(\theta)$. The black curve corresponds to a valid plane, whereas the grey one is rejected. The following validity criteria are used: maximum value of the function $Sim(\theta^{max}) \geq 0.4$, absolute value of the estimated second derivative around the maximum $|Sim''(\theta^{max})| \geq 4.0$.



(a) (b)

Figure 6: (a) Detail of figure 1a with projected 3D lines (white). This building is used to illustrate the reconstruction method. The correct reconstruction is a four plane hip roof. (b) Detected edges (black) after applying an edge detector with a very low threshold on gradient. These edges provide the points of interest.

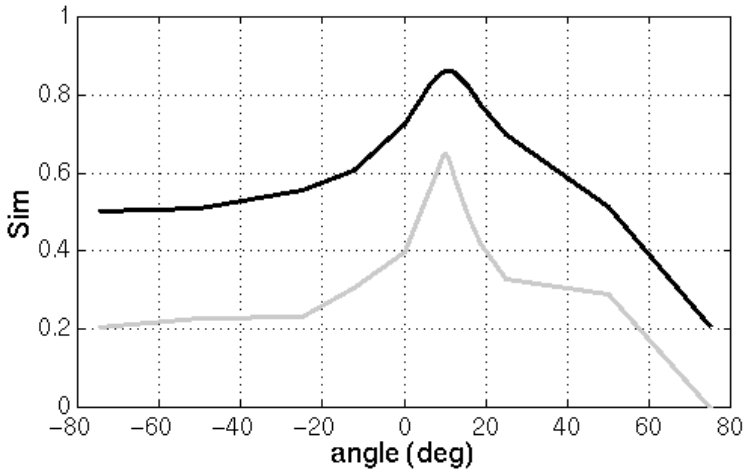


Figure 7: Effect of the baseline on 2-view similarity scores. The black curve corresponds to a short baseline between views (e.g. views 1 & 2), the grey curve to a wide one (e.g. 1 & 4). The curves apply to the same half-plane. A short baseline leads to high maxima (low distorsion between the images), but often located with a poor accuracy (wide peak); in contrast, a wide baseline is more likely to produce accurate maxima, but with a lower score. However, the maxima generally differ by less than 5^o.

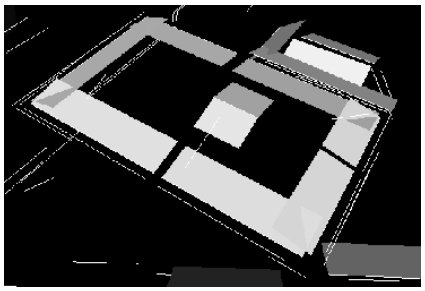


Figure 8: Detected half-planes over the interval $[-75^o; +75^o]$.

3.5 Results of half-plane detection

Figure 8 shows all the half-planes which are hypothesised on the example building. All parts of the roof of the main building are detected, whereas no valid planes are detected for the walls within the considered angle interval (we are not aiming to reconstruct vertical walls). Occasionally erroneous half-planes arise at shadows, but these are removed in the subsequent stages.

4 GROUPING AND COMPLETION OF 3D LINES BASED ON HALF-PLANES

At this stage of the process, we have produced a set of independent half-plane hypotheses, each characterized by:

- one 3D line and one side (defined in a reference coordinate system),
- an infinite plane (containing the 3D line by construction),
- the reliability indicators mentionned in section 3.4.

These half-planes are now used to support line grouping and the creation of new lines. In some of these operations, the order of processing can have an effect on the result, therefore all hypotheses are first sorted by decreasing reliability, using the indicators mentionned above.

Importantly, in grouping operations, thresholds on distances are avoided by using the topological neighbourhood between projected lines, defined in section 3.3. The neighbourhood of a plane is defined via the neighbourhood of the lines belonging to it. The planes and the lines are therefore represented within a graph structure, which enables quick access to neighbours.

Collinear grouping. First two collinear lines which have attached coplanar half-planes are merged together (see figure 9). The optimal plane angle is recomputed for the merged line, again using the score function $Sim(\theta)$ as described in section 3. This is more accurate than, for instance, averaging angles, because planes are more reliable when defined over a long line (more points of interest available). The result of the collinear grouping of half-planes of figure 8 is shown in figure 11a.

Coplanar line and half-plane grouping. Any line which is neighbouring and coplanar with the current plane is associated with it. Besides, if this line has also an attached but consistent half-plane (see figure 10), then the two plane hypotheses are merged into a new unique plane. In both cases, the new plane is computed by orthogonal regression to a regular point sampling of the lines belonging to it. Note, the planes defined by two (non-collinear) lines or more are thus estimated using geometrical support (3D lines) rather than photometric support (similarity function), because it usually provides more accuracy. However, if the resulting angle differs too much from the original one (it can happen when the new line is too short or too close to the initial one), then the new line is rejected. An example of coplanar grouping is shown in figure 11b).

Creating new lines by plane intersections. New lines are created when two neighbouring planes intersect in a *consistent* way, i.e, when the intersection line segment belongs to each half-plane (see figure 12). This process is very important as it provides a mechanism for generating additional lines which may have been missed during image feature detection (see the example of figure 13).

Since these lines are "virtual" (artificially created from two plane hypotheses), they rely on the validity of the two original planes. It is therefore necessary to keep trace of this pair of planes for further update operations (see next section).

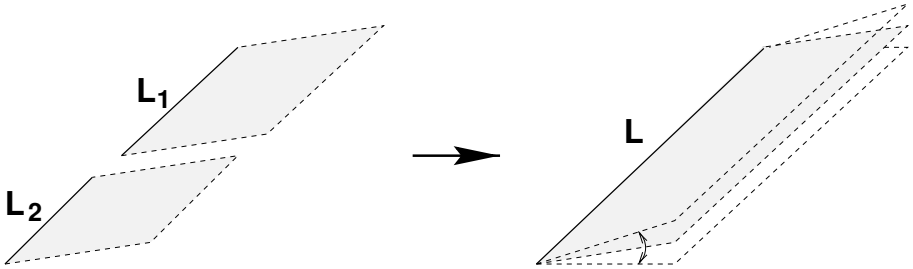


Figure 9: Collinear grouping. Collinear lines with coplanar half-planes are merged together, and the optimal plane angle is recomputed using $Sim(\theta)$.

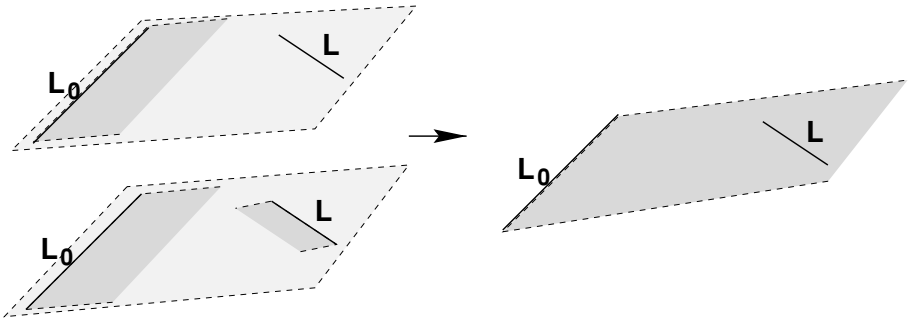
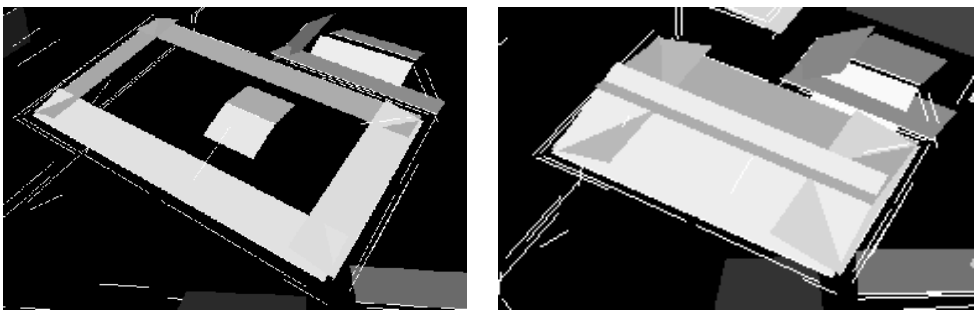


Figure 10: Coplanar line and half-plane grouping. In the top case, $\mathbf{L}$ belongs to the half-plane $\pi(\mathbf{L}_0)$, and a new plane is computed by orthogonal regression to a regular point sampling of $\mathbf{L}_0$ and $\mathbf{L}$. In the bottom case, $\mathbf{L}$ has an attached but consistent half-plane, therefore the two plane hypotheses are merged into a new unique plane, also computed by orthogonal regression.



(a) (b)

Figure 11: 3D line grouping. (a) Collinear grouping reduces the 9 planes prior to grouping to only 6. (b) Coplanar grouping and plane merging reduces the number of planes further so that only 4 remain. These are the correct four planes which define the roof, but at this stage the plane boundaries are not delineated.

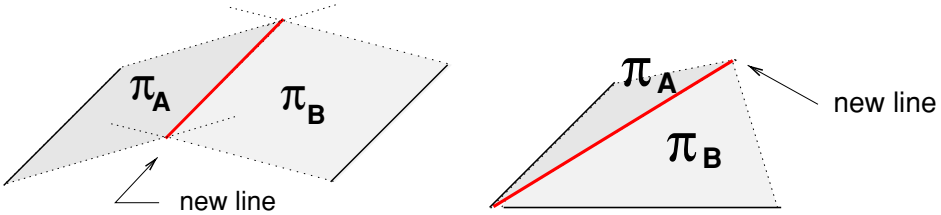


Figure 12: Creation of new lines when two planes intersect.

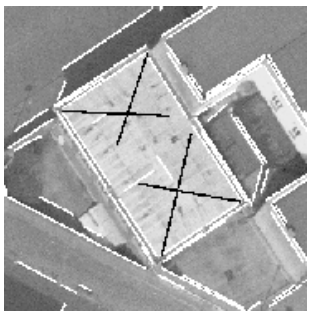


Figure 13: New lines (black) created by plane intersection.

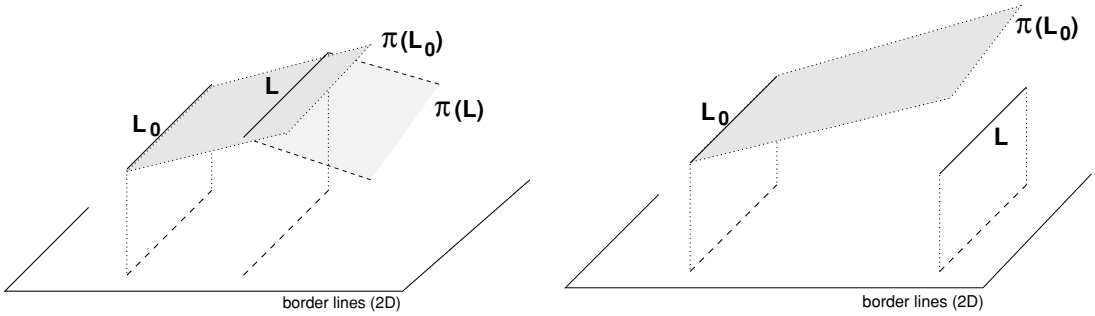


Figure 14: Border line computation for plane delineation. (a) The line $\mathbf{L}$ lies in the plane $\pi(\mathbf{L}_0)$ but has an attached plane which is not consistent with it, therefore it is stored as a border line; (b) The line $\mathbf{L}$ does not belongs to the plane $\pi(\mathbf{L}_0)$ but it is stored as a border line.

5 PLANE DELINEATION AND VERIFICATION

Plane delineation. In order to produce a piecewise planar model of the scene a closed delineation is required for each plane. For this purpose, it is necessary to determine a set of *border lines*, which will define the final boundaries of the face.

The initial support line of a planar facet is a natural border line. Additional border lines are provided by the following features (see examples of figure 14a):

- 3D lines belonging to the current plane but attached to a different one,
- virtual lines which were created by intersection,
- neighbouring and *reliable* 3D lines not belonging to the plane (it is necessary to take only reliable lines into account here since there is no way to verify them in the subsequent stages - length is the reliability criterion currently used).

A closed delineation can then be computed by using heuristic grouping rules (Weidner and Förstner, 1995, Noronha and Nevatia, 1997, Moons et al., 1998) to associate border lines. For instance the end points of the border lines are updated when lines intersect or have close end points (see figure 15). Then the convex hull of the border line end points is computed. The final delineation is derived from the convex hull and the lines defining the plane according to simple perceptual rules (see an example in figure 16). Whenever a patch is added or removed, intensity similarity over the views is verified.

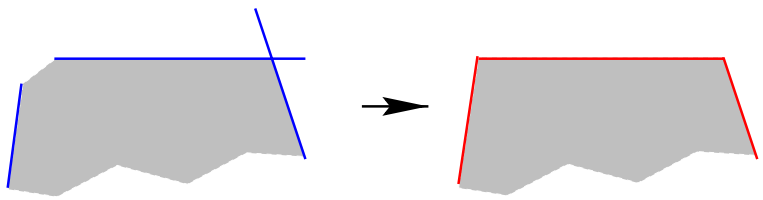


Figure 15: Updating end points of the border lines: any end point outside the region of interest is moved into it; two close endpoints are replaced by the intersection of the two lines

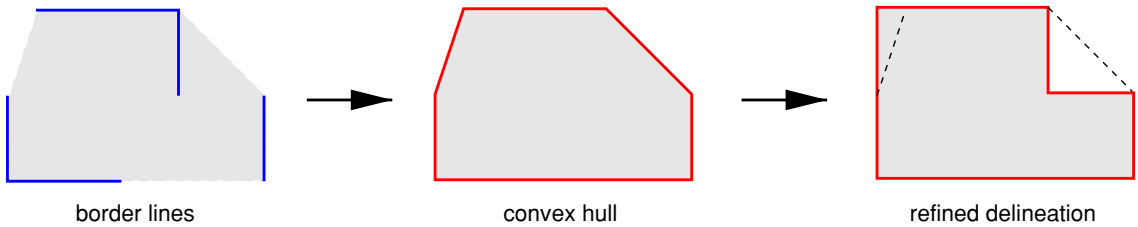


Figure 16: Delineation refinement. The final delineation is derived from the convex hull and the lines defining the plane according to simple perceptual rules. Whenever a patch is added (respectively removed), the similarity of the corresponding pixels over the views must be high (resp. low) enough.

Plane verification. At this stage, there are still several planar faces which do not correspond to any real face, and some others which are locally incorrect (case of some non convex faces for instance). They must therefore be detected and corrected/removed from the data set.

First, at least two lines (one of them can be virtual) are required to accept a plane, and this provides a very efficient culling mechanism for removing erroneous half-planes with one line only.

Each delineated 3D face is then segmented into several patches, defined by the lines belonging to the plane. Each planar patch is verified by assessing intensity similarity over the complete image set, at corresponding points within the projected delineation. The similarity must be verified by either SSD or cross-correlation (two complementary measures are used to guarantee that only wrong patches are removed). This verification step removes fallacious planes, for example those which erroneously bridge two buildings. Figure 17 shows both the 2D delineation and a 3D view of the roof produced for the building of figure 6a.

Conflict management. Finally, occlusion prediction is used to signal and resolve conflicts between inconsistent plane hypotheses. A conflict occurs between two facet hypotheses when their projections onto an image substantially overlap, i.e. when one of them is occluded by the other. Where conflicts between plane hypotheses arise, there are two possibilities, depending on the coplanarity of the conflicting planes.

If the conflicting planes are coplanar, then the planes are merged together. The position of the new single plane is computed using lines belonging to both merged planes, and its delineation is obtained by merging the previous delineations together.

On the contrary, if the the conflicting planes are not coplanar, then the conflict is resolved using a confidence score $S(\mathcal{P})$, which denotes the quality of the face $\mathcal{P}$:

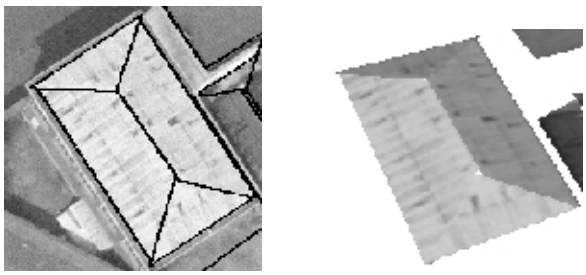
$$S(\mathcal{P}) = \mathcal{A}(\mathcal{P}) \times \tau_{real}(\mathcal{P}),$$

where $\mathcal{A}(\mathcal{P})$ is the area of the face, and $\tau_{real}(\mathcal{P})$ the ratio of the input 3D lines to the delineation. The plane hypothesis with the lowest confidence score is removed. Note, this score based on geometric support has proved much more reliable for selecting the best plane than any score based on photometric similarity support.

Plane update. Whenever a plane is updated or removed, all the neighbouring planes must be updated in order to keep consistency. In particular, all the virtual lines related to the current plane must also be updated, as well as the delineation of the second plane related to this virtual line.

6 RESULTS

Figure 18 shows the 3D reconstruction related to the image set of figure 1. Figure 20 shows the result on the larger and more complicated images of figure 19. Note that intricate and unusual roofs (for example the factory in the upper part of the image) have been completely recovered. Only two roofs are missed in the entire scene. Figure 21 shows the reconstruction of a scene starting from $2K \times 2K$ images (from which the small images of figure 1 were taken). Again, only 2 roofs are missed (if we do not include roof planes only partly visible in the images). In particular, even roofs with virtually homogeneous intensity have been entirely retrieved. This also demonstrates how little photometric texture is required by the method. Finally, the versatility of the approach is demonstrated on figure 22 where the images are of an indoor scene, at a different scale and under differing photometric conditions to those of the aerial images. The main planes of the scene are correctly retrieved.



(a) (b)

Figure 17: Example of reconstructed roof. (a) Delineation of the validated roofs projected onto the first image; (b) 3D view with texture mapping.

(a) (b) (c)

Figure 18: Results on the full example scene. (a) 49 detected half-planes from 137 3Dlines (b) Delineation of the final roofs projected onto the first image; (c) 3D model of the scene, with texture mapping (12 roof planes). The vertical walls are produced by extruding the roof's borders to the ground plane.

Performance analysis. Of the three stages of the method, the half-plane detection stage is the most robust and is also the most expensive. The adaptive selection of views and points of interest in the similarity cost function provides robustness to partial and total occlusions. In addition, the adaptive selection of the reference view as the most textured enables the recontruction of roofs with very little photometric texture. Finally, this stage requires very few key parameters to be specified. The use of a topological neighbourhood is very important since it avoids thresholds on distances. When a face is well textured (as in the case of the example building roof of figure 6), the angle of the initial half-plane is estimated to an accuracy of better than 2^o. When there is little texture, the accuracy can decrease to 5^o, but a higher accuracy is determined during the coplanar grouping stage.

The grouping and delineation stages are robust to a proportion of missing and erroneous lines because mechanisms are included to generate new lines by plane intersection, and to cull erroneous lines with their associated half-planes, in the final verification stages. Consistency verification is a key point of the process. However, these stages depend on internal thresholds defining geometric properties like collinearity, coplanarity, etc. These parameters are currently fixed and empirically based, although identical for all data sets. It would be preferable if their definition was also adaptive (locally determined for each feature), for instance through a statistical model involving uncertainty about line and plane location. The plane delineation stage is the least robust to changes in scene type because it involves heuristic grouping rules.

Finally, the quality of the reconstruction is governed by the completeness and correctness of the input line set. The method is robust to a proportion of missing and erroneous lines, but the performance is improved if too many, rather than too few, lines are supplied. This is because a line is the only mechanism for instantiating a plane hypothesis, and if lines are missing then entire planes may be missed. Besides, there are so many ways of culling wrong plane hypotheses than erroneous lines are unlikely to generate a facet in the final model.

7 CONCLUSION

The results demonstrate the efficiency of the method for automatically constructing piecewise planar models of scenes from multiple images using quite minimal information. The models are of very reasonable quality given the complexity of the original scenes.

Figure 19: Six overlapping aerial views. The images are about 1200×1200 pixels, one pixel corresponding to a ground length of 8.5cm. The images are acquired at a height of 1300m in two triplets: the camera approximately translated by about 300m between successive views in images 1-3, and between images 4-6. The first set acquired on the outward flight, and the second set on the approximately parallel return flight.

The originality of this approach lies in the use of photometry (over the image set) to generate plane hypotheses rather than 3D geometric features only. In addition, great care is taken that all plausible hypotheses are kept until the latest stages of the process (no early thresholding). Erroneous hypotheses are detected using several loose thresholds rather than a single but tight one. The use of redundancy between views provides robustness to occlusions and missing features. Consistency constraints are exploited to detect and manage conflicts, providing a very efficient way of culling hypotheses. The underlying assumption is that the complexity and the density of the scene guarantee a reduced number of possible solution.

The quality might be further improved further if additional features were provided for matching. For example, a bar detector would gain extra lines in the example building of figure 6. A second improvement would be the incorporation of other, albeit more restrictive, 3D grouping mechanisms. The approach of this paper is not an alternative to roof generation systems based on the grouping of at least two neighbouring coplanar lines (Bignone et al., 1996, Moons et al., 1998), but is complementary to such systems. An efficient and robust approach would use both, grouping and removing from consideration coplanar lines and using the described single 3D line method for the remainder. This requires an architecture for a cooperating strategy to be developed. The robustness of the recontruction could finally be further improved if the definition of geometric properties had a statistical foundation. In particular, thresholds about collinearity and coplanarity should depend on uncertainty about line and plane location (Bittner and Winter, 1999). The reliability indicators provided by our system could also be used for that purpose.

Acknowledgements. We are grateful to Dr A. Fitzgibbon for his assistance with software and helpful discussions. Financial support for this work was provided by the EC Esprit Project IMPACT and the UK EPSRC IUE Implementation Project GR/L05969.

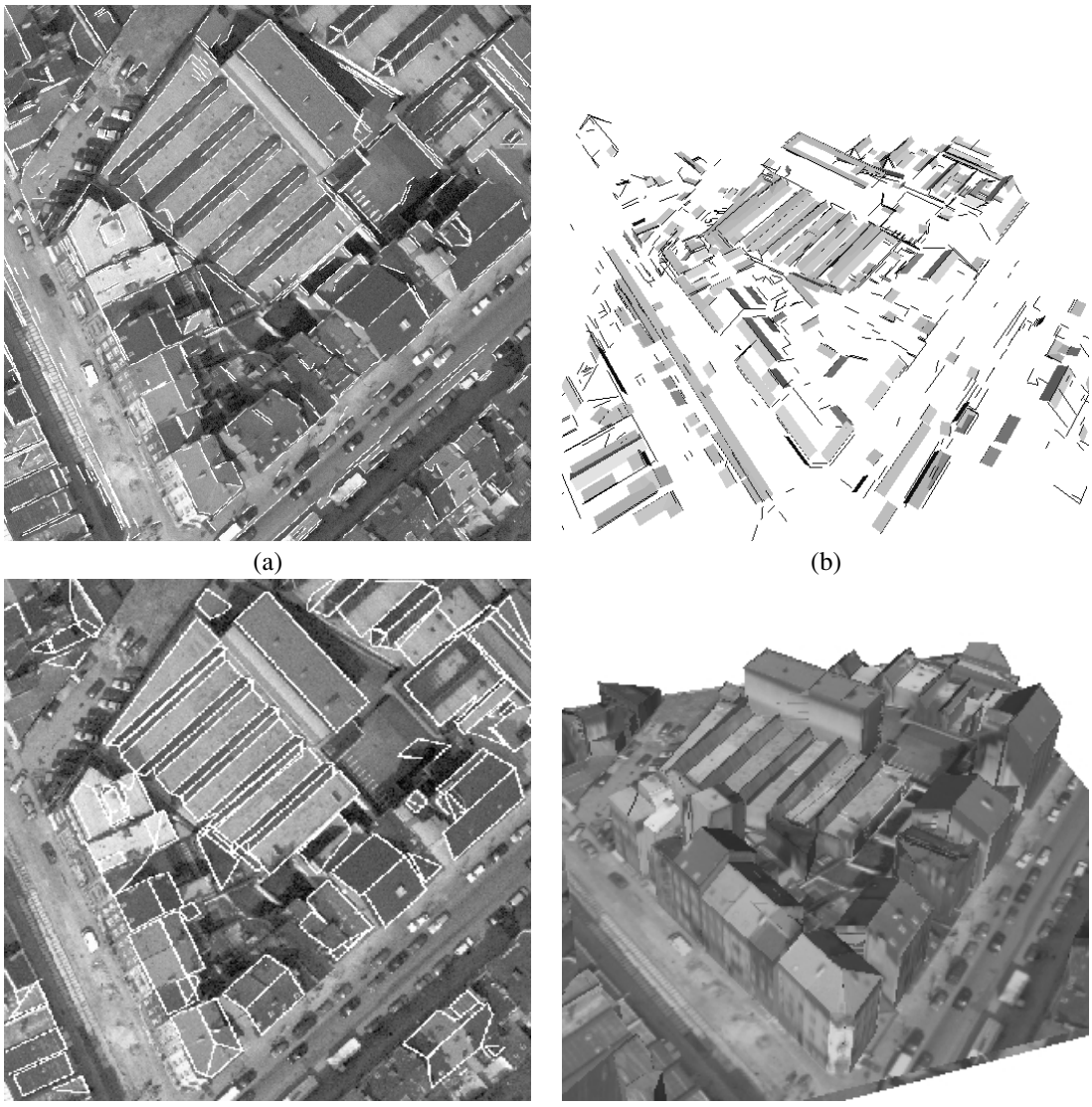
REFERENCES

Baillard, C. and Maître, H., 1999. 3D reconstruction of urban scenes from aerial stereo imagery: a focusing strategy. Computer Vision and Image Understanding pp. 244–258.

Baillard, C., Schmid, C., Zisserman, A. and Fitzgibbon, A., 1999. Automatic line matching and 3D reconstruction of buildings from multiple views. In: ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery, IAPRS Vol.32, Part 3-2W5, pp. 69–80.

Berthod, M., Gabet, L., Giraudon, G. and Lotti, J. L., 1995. High-resolution stereo for the detection of buildings. In: A.Grün, O.Kübler and P.Agouris (eds), Automatic Extraction of Man-Made Objects from Aerial and Space Images, Birkhäuser, pp. 135–144.

Bignone, F., Henricsson, O., Fua, P. and Stricker, M., 1996. Automatic extraction of generic house roofs from high resolution aerial imagery. In: Proc. 4th European Conference on Computer Vision, Cambridge, pp. 85–96.



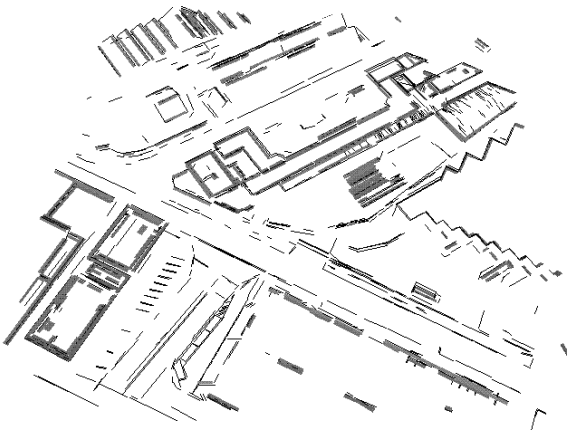
(a) (b)

(c) (d)

Figure 20: Results on the image set of figure 19. (a) One of the 6 images and the 739 projected 3D lines. (b) Detected half-planes (267). (c) final delineation of the planar facets (180 roof planes) (d) 3D model of the scene, with texture mapping.

(a)

(b)

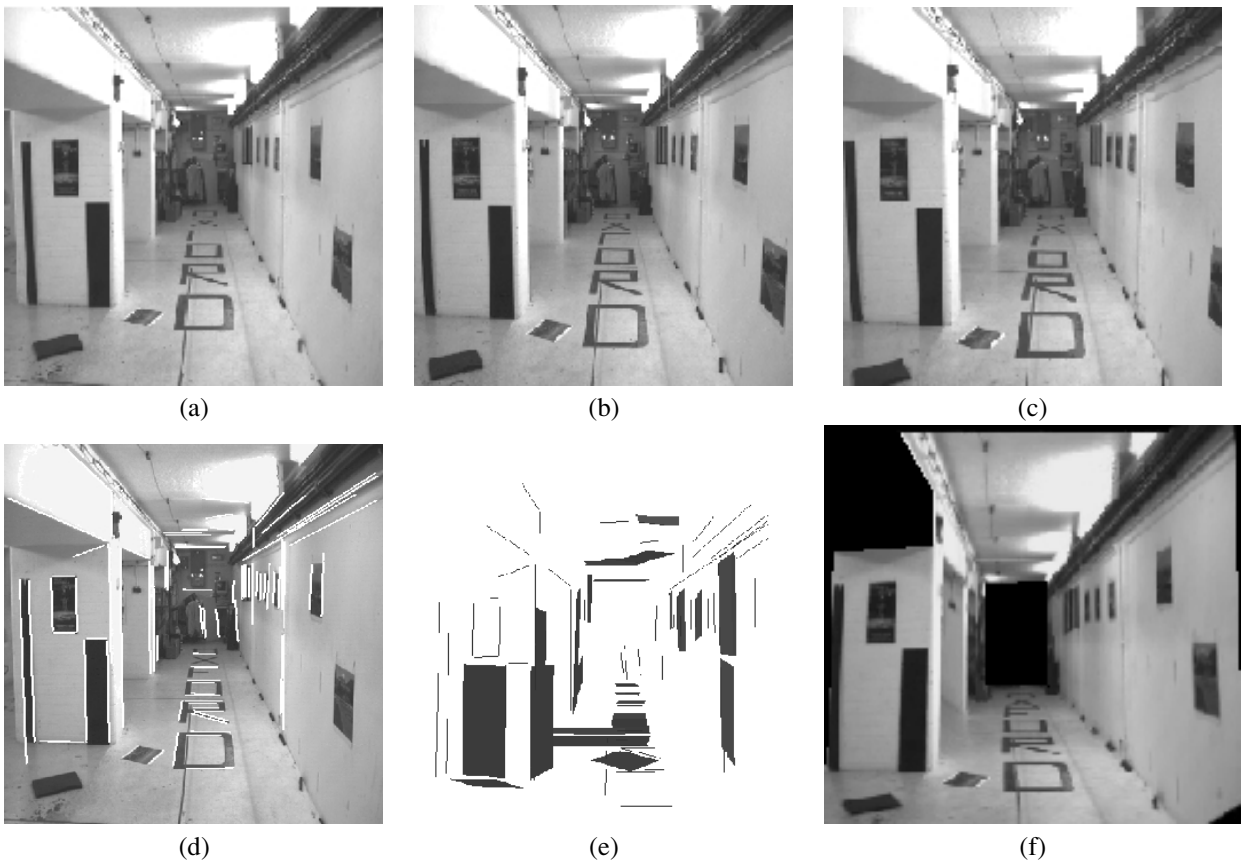


(c)

(d)

(e)

Figure 21: Results on a 2K $\times$ 2K image set. (a,b) : Two of six overlapping views. (c) 3D model of the 3D lines with the detected half-planes (998 lines and 373 half-planes). (d) final delineation of the planar facets (87 roof planes). (e) textured 3D model of the scene.



(a) (b) (c)

(d) (e) (f)

Figure 22: Results on an indoor scene. (a,b,c) 3 of the 6 input images (size 512×512 pixels) (d) First image and the 87 projected 3D lines. (e) Detected half-planes (30). (f) 3D model of the scene, with texture mapping (5 planes).

Bittner, T. and Winter, S., 1999. On ontology in image analysis. In: International Workshop on Integrated Spatial Databases, Portland, ME, USA. Lecture Notes in Computer Science 1737, pp. 168–191.

Collins, R., 1996. A space-sweep approach to true multi-image matching. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 358–363.

Collins, R., Jaynes, C., , Cheng, Y.-Q., Wang, X., Stolle, F., Riseman, E. and Hanson, A., 1998. The ascender system: Automated site modeling from multiple images. Computer Vision and Image Understanding 72(2), pp. 143–162.

Coorg, S. and Teller, S., 1999. Extracting textured vertical facades from controlled close-range imagery. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 625–632.

Cord, M., Paparoditis, N. and Jordan, M., 1998. Dense, reliable, and depth discontinuity preserving dem computation from very high resolution urban stereopairs. In: ISPRS Symp., Cambridge (England).

Girard, S., Guérin, P., Maître, H. and Roux, M., 1998. Building detection from high resolution colour images. In: SPIE Europto Image and Signal Processing for Remote Sensing IV (S. Serpico Ed.), Vol. 3500, Barcelona (Spain), pp. 278–289.

Haala, N. and Hahn, M., 1995. Data fusion for the detection and reconstruction of buildings. In: Automatic Extraction of Man-Made Objects from Aerial and Space Images, Birkhäuser, pp. 211–220.

Hartley, R. I., 1995. A linear method for reconstruction from lines and points. In: Proc. International Conference on Computer Vision, pp. 882–887.

Luong, Q. T. and Viéville, T., 1996. Canonical representations for the geometries of multiple projective views. Computer Vision and Image Understanding 64(2), pp. 193–229.

McGlone, J. and Shufelt, J., 1994. Projective and object space geometry for monocular building extraction. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 54–61.

Moons, T., Frère, D., Vandekerckhove, J. and Van Gool, L., 1998. Automatic modelling and 3D reconstruction of urban house roofs from high resolution aerial imagery. In: Proc. 5th European Conference on Computer Vision, Freiburg, Germany, pp. 410–425.

Noronha, S. and Nevatia, R., 1997. Detection and description of buildings from multiple images. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico, pp. 588–594.

Paparoditis, N., Cord, M., Jordan, M. and Cocquerez, J.-P., 1998. Building detection and reconstruction from mid- and high-resolution aerial imagery. Computer Vision and Image Understanding 72(2), pp. 122–142.

Roux, M. and McKeown, D. M., 1994. Feature matching for building extraction from multiple views. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition.

Schmid, C. and Zisserman, A., 1997. Automatic line matching across views. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 666–671.

Seitz, S. and Dyer, C., 1997. Photorealistic scene reconstruction by voxel coloring. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 1067–1073.

Shashua, A., 1994. Trilinearity in visual recognition by alignment. In: Proc. 3rd European Conference on Computer Vision, Stockholm, Vol. 1, pp. 479–484.

Spetsakis, M. E. and Aloimonos, J., 1990. Structure from motion using line correspondences. International Journal of Computer Vision 4(3), pp. 171–183.

Weidner, U., 1996. An Approach to Building Extraction from Digital Surface Models. In: XIX Congress of ISPRS, Comm. III, Int. Archives of Photogrammetry and Remote Sensing Vol. 31, Vienne, pp. 924–929.

Weidner, U. and Förstner, W., 1995. Towards automatic building extraction from high-resolution digital elevation models. ISPRS j. of Photogrammetry and Remote Sensing 50(4), pp. 38–49.