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

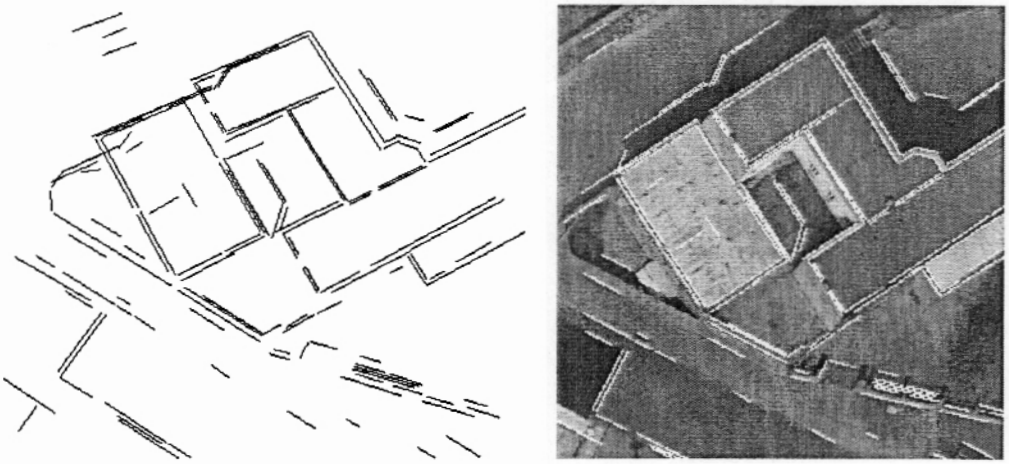
Automating reconstruction from images is one of the continuing goals in photogrammetry and computer vision. The special case of piecewise planar reconstruction is particularly important due to the large number of applications including: manufactured objects, indoor environments, building exteriors, 3D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



(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.

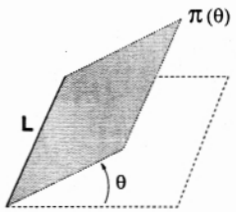


Figure 3: The one-parameter family of half-planes containing the 3D line **L**. The family induces a one-parameter family of homographies between any pair of images.

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).

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:

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

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 $\mathrm{H}^i(\theta)$. 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.

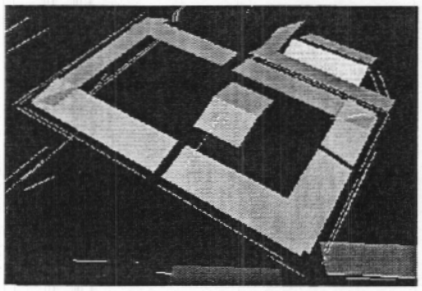


Figure 8: Detected half-planes over the interval $[-75^o; +75^o]$.

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).

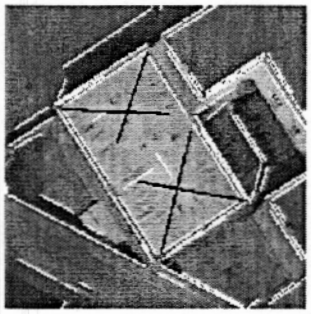


Figure 13: New lines (black) created by plane intersection.

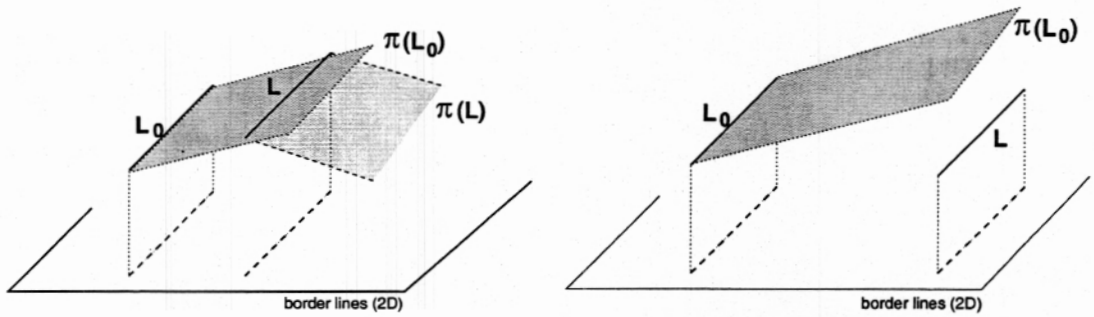


Figure 14: Border line computation for plane delineation. (a) The line $\mathbf{L}$ lies in the plane $\pi(\mathbf{L}_0)$ but has an attached plane which is not consistent with it, therefore it is stored as a border line; (b) The line $\mathbf{L}$ does not belongs to the plane $\pi(\mathbf{L}_0)$ but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

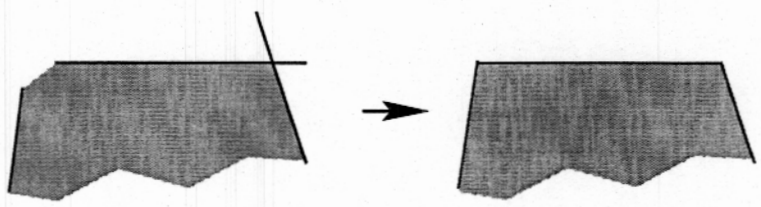
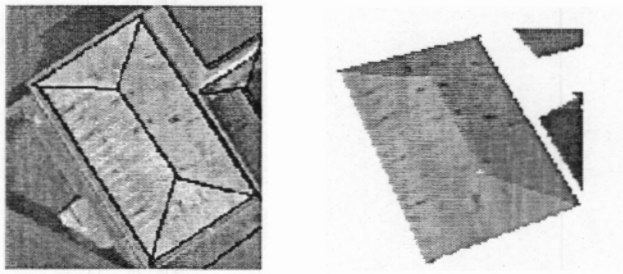
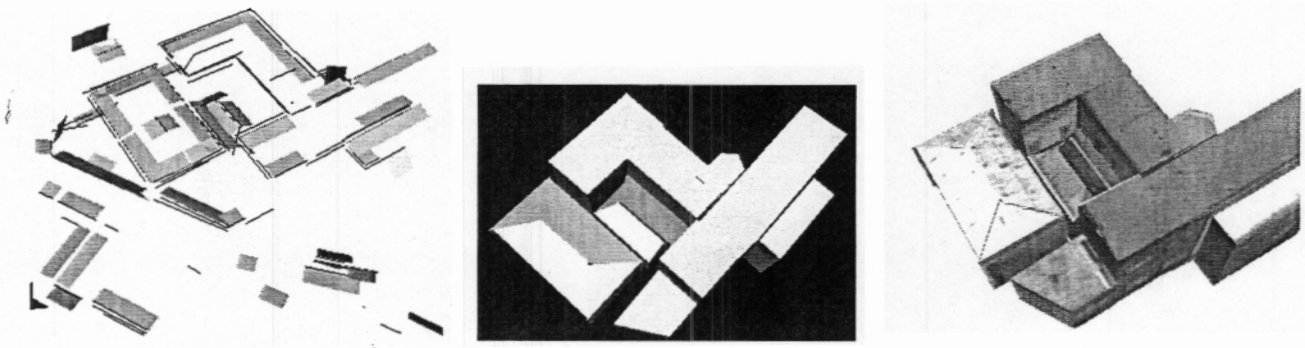


Figure 15: Updating end points of the border lines: any end point outside the region of interest is moved into it; two close endpoints are replaced by the intersection of the two lines



(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.

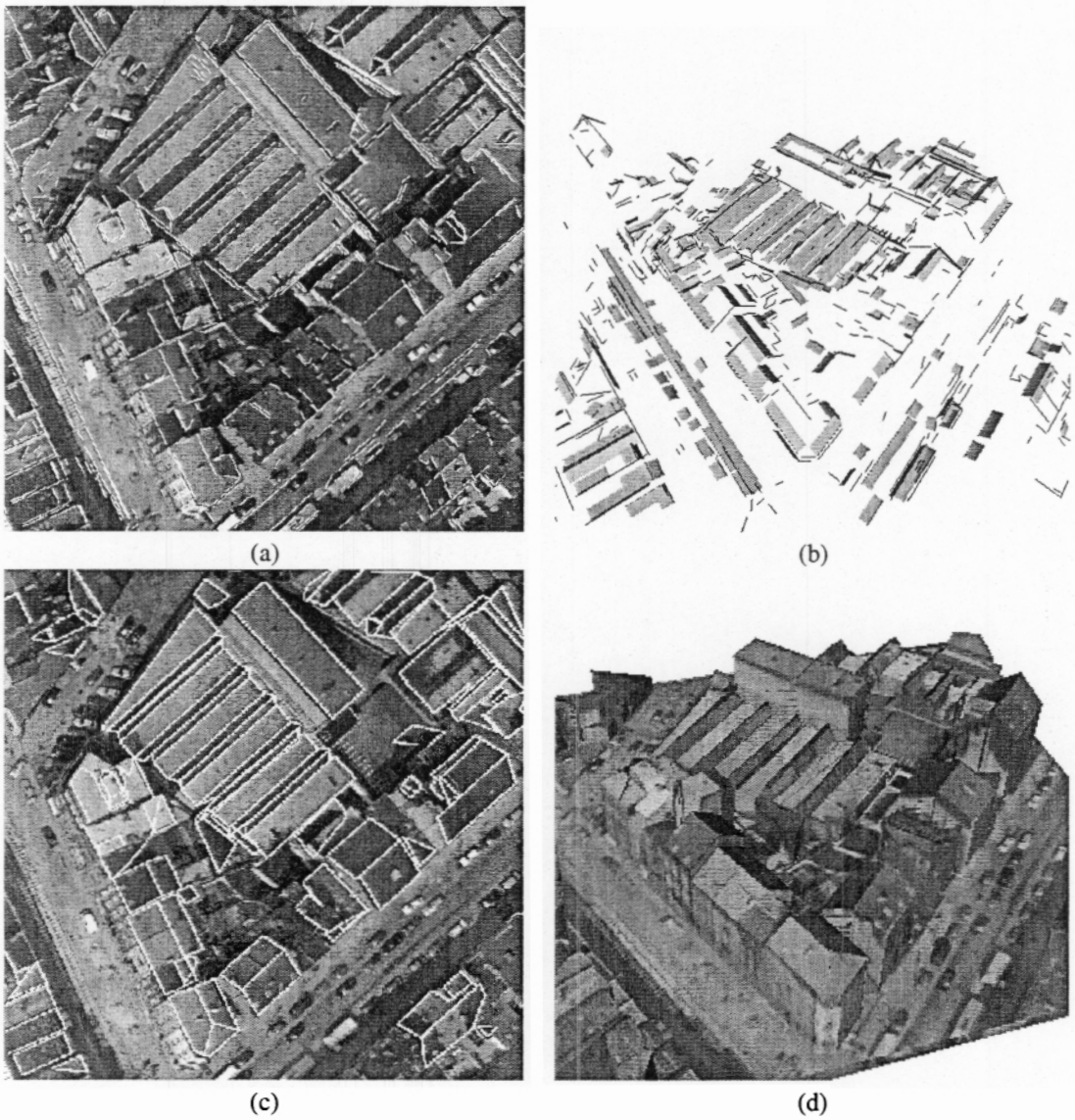


Figure 20: Results on the image set of figure 19. (a) One of the 6 images and the 739 projected 3D lines. (b) Detected half-planes (267). (c) final delineation of the planar facets (180 roof planes) (d) 3D model of the scene, with texture mapping.