TOWARDS SEMI-AUTOMATIC CONTROL OF TOPOGRAPHIC DATA BASES FROM AERIAL IMAGES

Matthieu Cord, Michel Jordan, Jean-Pierre Cocquerez
ETIS - URA CNRS 2235
ENSEA
6, avenue du Ponceau
F 95014 CERGY-PONTOISE CEDEX
FRANCE
E-mail : {cord, jordan, cocquerez}@ensea.fr

Commission II, Working Group 6

KEY WORDS: Building detection, Image understanding, 3D Reconstruction, Aerial image, Cartographic feature extraction

ABSTRACT

We propose in this paper an original scheme to detect and localise buildings from a stereo-pair of high resolution aerial images. Our strategy is based on a close co-operation between 2D and 3D information.
A multi-resolution and adaptive template matching technique provides dense, reliable and depth discontinuity preserving DEM. DEM segmentation and raised structure extraction based on filtering and merging inside of an adjacency graph of the 3D regions are carried out.
Using an optimal edge detector adjusting, radiometric contours are extracted and properly localised. The 3D regions are then used for the contour grouping. A first grouping makes a contour selection for one region after the other. Inside each region, a second grouping process links up image contours along the region frontier.
Finally, tests are performed on a stereo pair of images covering a suburb of Paris, and results are compared to data from the manual IGN topographic data base. We then address the question of semi-automatic control of topographic data bases, which is an even more and more relevant question due to the great and greater need of accurate and up-to-date cartographic knowledge.

RÉSUMÉ

Dans cet article, nous proposons un schéma original pour la détection et la localisation des bâtiments à partir de couples stéréoscopiques d'images aériennes à haute résolution. Notre stratégie est basée sur une coopération étroite entre les informations 2D et 3D.
Un algorithme de mise en correspondance multi-résolutions à fenêtres adaptatives fournit un modèle numérique d'élévation dense, fiable et préservant les discontinuités altimétriques. Ce MNE est ensuite segmenté, et les régions élevées sont extraites par un filtrage et groupement dans un graphe d'adjacences.
Les contours sont extraits et bien localisés à l'aide d'un détecteur de contours paramétré de manière optimale. Les régions 3D sont utilisées pour grouper les contours. Une première étape sélectionne les contours région par région. A l'intérieur de chaque région, une seconde étape de groupement lie les contours le long de la frontière.
Des tests ont été effectués sur un stéréo-couple d'images couvrant la banlieue de Paris, et les résultats ont été comparés aux données de la base topographique de l'IGN. Enfin, nous abordons la question du contrôle semi-automatique des bases de données topographiques, qui reçoit de plus en plus d'intérêt du fait des besoins croissants de données cartographiques précises et actualisées.

INTRODUCTION

Nowadays, various application fields need accurate and up-to-date cartographic data and city model elaboration: urbanism, mobile telecommunications, virtual reality applications, *etc.* As a consequence there is many and many interest for automatic or semi-automatic cartographic processing's.
Our research takes part in these objectives, and we have a special interest in the localisation of man-made structures in urban areas. In this paper, we focus on building extraction and localisation from very high resolution aerial imagery — we call "very high resolution images" images having a ground resolution about 8cm per pixel.

In section 1, we shortly present some recent related works.
Figure 1 summarises the sequence of processing's that we have developed:

- **Digital elevation model** (DEM) **computation**: we present an original method for accurate and dense DEM from stereo pair of images computing; and

DEM segmentation in order to detect 3-D interesting areas (section 2).
- Edge and **linear segment detection** from the stereo pair of aerial images; only the edges, which are close to 3D regions of interest, are selected (section 3).
- **Building recognition and modelling**, using both radiometric segments and frontier lines of the 3D regions (section 4).

Section 5 presents some results and accuracy evaluation on tests performed on a stereo pair covering a suburb area of Paris, France.
We finally conclude and present some perspectives in section 6.

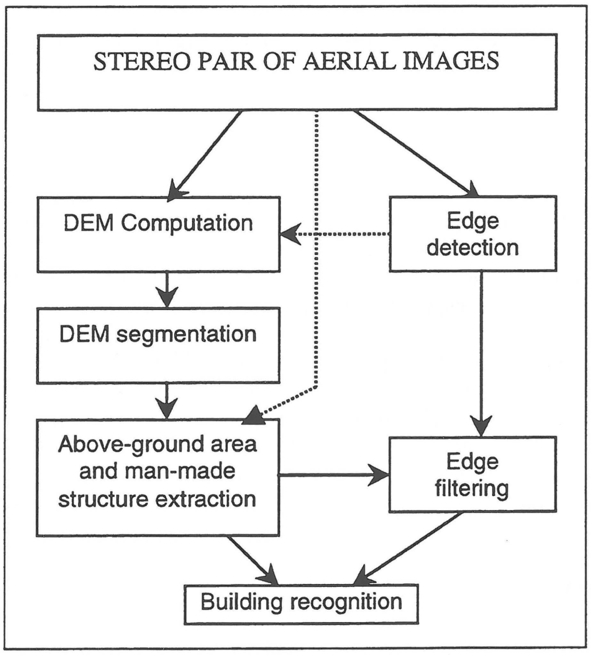


Figure 1 : Synoptic scheme

1 RELATED WORKS

Building detection from aerial images in urban areas is a rather complicated process, due to the great variability of man-made structures, the great number of details present in these images, and problems due to shadows and reflections.

Various solutions, using monocular or stereo approaches, have been investigated to deal with these difficulties. We just present here some recent interesting works.

Haala and Anders suggest the use of additional information, such as dense and accurate laser airborne DEM, or GIS information on ground plan (Haala and Anders, 1997).
A first method, combining grey-level images and DEM is developed: it consists in DEM segmentation for local maximal search and surface type labelling by combining Gaussian and mean curvature computation, and then reconstruction using linear segments extracted from the DEM segmentation and refined from stereo image pair linear segments extraction.
A second approach is then presented: DEM segmentation is performed from a planar ground plan provided by a digital cadastral map. Each building is represented by a general polyhedron, which is determined by the computation of an adjacency graph of the planar surfaces.

Lammi presents a similar system for 3D building information extraction from a set of aerial images and a two-dimensional basement data such as a cadastral map (Lammi, 1997). A heuristic search for edge finding is defined by projecting the polygon representing the building basement onto the images, and lifting it upwards within a chosen interval. The evaluation function is based on a gradient computation in rectangular windows along polygons. Subpixelar search is performed to refine the accuracy of the method.
This method is limited to buildings already presents in the cadastral map, but seems to provide accurate results.

The aim of the AMOBE project (Henricsson *et al.*, 1996) at ETH Zürich is the 3D reconstruction of man-made objects from 10cm per pixel resolution aerial images. A Digital Elevation Model is computed from a standard correlation algorithm with a grid spacing about 30cm. 3D blob detection is performed on the DEM and is combined with colour image information in order to extract building hypotheses. 2D linear segments are then extracted from one image and corresponding features are sought for in the other images by maximising an "edginess function" along the epipolar line. 3D segments are grouped giving sets of 3D planes, which are assembled to roof models.

Weidner developed an approach for extracting 3D building shapes from a DEM using constraints on building models (Weidner, 1995). Building detection is based on the fact that the buildings are higher than the topographic surface, which is estimated using mathematical morphology on the DEM.
Buildings are modelled depending on their complexity: two kinds of parametric models are used for simple buildings, having either flat or symmetric sloped roof; prismatic models are used for complex or connected buildings.

Frère, Vandekerckhove *et al.* developed a system for reconstructing house roofs from a set of 6 urban scenes colour images (Vandekerckhove *et al.*, 1997) (Frère *et al.*, 1997). Each image is segmented, and regions of interest are manually matched in the six images. Stereo matching of linear segments is performed between corresponding regions. Segments are then grouped in coplanar configurations, which are themselves grouped in polygonal hypotheses. Consistency of each polygon hypothesis is then tested and corrected with respect to the 3D reconstruction and the image data.

Girard *et al.* present a strategy for colour image matching, adapted to dense urban scenes (Girard *et al.*, 1998). A classical "coarse-to-fine" multilevel cross-correlation algorithm provides a sparse disparity map. This map is then interpolated on regions of radiometric colour segmentation. Robust estimation methods are iteratively applied to the map, giving high level information on the scene, especially on regions likely to appear as buildings.

It is important to note from these different approaches that building detection is rarely carried out just using aerial images. Laser airborne DEM, cadastral map, initial manual region matching, *etc.*, are so many heuristics to regularise the detection problem and to provide satisfactory results. Our work only uses high resolution aerial images as data source. Our strategy rests on the obtaining of dense and reliable DEMs. In fact, we believe that such an altimetric information is decisive in the building detection process.

2 DEM COMPUTATION AND 3D SEGMENTATION

2.1 DEM Computation

The process of DEM computation is detailed in another paper in these archives (Cord *et al.*, 1998).

The matching algorithm is based on a cross-correlation template matching to provide the denser depth maps as possible. In this algorithm, template windows are not rectangular, they are landscape adaptive. An internal validation of the disparity measurements based on the study of symmetrical correlation coherence enhances the reliability of the process but therefore leads to sparser maps.
To obtain denser maps and to accelerate the matching process especially on very high resolution images where the disparity search intervals for the points to be matched can be very wide an thus lead to a combinatorial explosion, our single scale matching process is integrated in a multi-resolution matching strategy.
This processing scheme stands very good results on a set of complex and various urban scenes images.
DEMs are dense, reliable and depth discontinuity preserving.

Disparity values are then transformed in absolute altitude values, knowing all the orientation data of both images of the pair.
All subsequent processes are based on these altitude values. This conversion allows us to have a fine tuning of parameters, related to building physical dimension and characteristic knowledge.

2.2 DEM Segmentation

The DEM characteristics are decisive for the segmentation.
In fact, the density allows us to apply a region grower segmentation technique. The region grower merging is limited to the known elevation grid points, which are large and spread out on the entire image.
The depth discontinuity preserving allows us to use a local and simple merging criterion. It is based on the local elevation difference. Given a 4 connected neighbouring pixel *(k,l)* of a *(i,j)* pixel belonging to the region *R*, and *z(k,l)* the *(k,l)* elevation grid point, we have:

$$(k,l) \in R \Leftrightarrow \begin{cases} z(k,l) \neq \varnothing \\ |z(i,j) - z(k,l)| \leq thr \end{cases}$$

The threshold *thr* is fixed to 0.2m that corresponds to the elevation resolution. Such 3D segmentation does not depend on the initial seeds. In fact, the merging criterion is symmetrical in *(k,l),(i,j)*. Any pixel belonging a given region *R* could be chosen to initialise the growing process. Using a line scanning, each time a no labelled pixel is encountered, it is chosen as seed to initialise a new region. The main problems of this kind of segmentation are the over/under-segmentation and the irregularity of the frontiers. We prefer to choose a very low merging thresholding value. That allow us to avoid under-segmentation problems. The following region merging process will correct the over-segmentation problems.

2.3 Above-ground area extraction

From the 3D segmentation map, the above-ground binary map may be extracted. Different strategies have been investigated. Paparoditis (Paparoditis *et al.*, 1998), Gabet *et al.* (Gabet *et al.*, 1994) inspect the immediate region neighbourhood to compute the region relative elevation, whereas Baillard (Baillard *et al.* 1996) use a first estimated digital terrain model (DTM) to compute the relative elevation of the regions. Then, the classification process may be iterative: the first above-ground map is used to refine the DTM estimation, which is used in turn to a new above-ground estimation.
Moreover, to correct the DTM bad estimation effect, they use a Markov model associated to the binary classes.

As the ground of the processing scenes is flat or few sloping and the covered area not very extensive because of the image resolution, we have implemented a simple scheme of classification. From the 3D segmentation regions, the lowest one provides the first estimation of the ground elevation of the scene. In fact, as the ground elevation is very few varying on the entire scene, the computed value is considered as valid for the entire image. Then, a simple thresholding provides a good initial classification of the 3D segmentation regions.
Then, using an adjacency graph representation of the 3D regions, a merging process based on altimetric criterion is carried out to correct the effect of over-segmentation. This merging has to be adjusting according to the search precision of the raised structure description.
Moreover, some errors resulting from the DEM computation generate regions with aberrant elevation. There are isolated or connected regions but their global surface never exceeds few square meters. That is why we filter them from the description of the above regions.
From the graph, all the groups of connected regions are formed, and all hotchpotch of connected 3D above regions, which is too small to constitute a real 3D above object, is removed.

Finally, above regions are filtered using an erosion and dilation mathematical morphology filters. Thanks to this step, the contours of segmented regions, sometimes irregular, are rectified.

2.4 Man-made structure extraction

The above extracted regions correspond generally to the building themes because of the minimal size imposed to these regions, but may contain some areas corresponding to vegetation themes.
We then filter according to an entropy criterion (Baillard *et al.*) to extract the raised structures corresponding to the building themes. This criterion is directly applied on the regions and does not allow to divide into two new regions possible mixed regions. However, because of the depth discontinuity preserving DEM, these mixed regions rare and the following segment grouping partly corrects this problem.

At the end of the processing, the man-made structures are extracted. However, the limits of these regions are not systematically properly localised.
That is why we propose to use radiometric segments in order to refine the building localisation.

3 EDGE AND LINE DETECTION

3.1 Edge detection filter Adjusting

As detection and localisation errors usually result from the nearness of several contours, constraints are introduced and limit values are determined in order to detect close contours without shifting. These values depend on both distance and relative amplitude between the contours and can be directly extracted from the image. The application of these constraints to Deriche filter allows to set up limit values of the scale parameter α.
As an alternative solution to the classical gradient norm thresholding, we propose in (Cord *et al.* 1997) sorting out the true contours by way of geometric and topological attributes of the edges. These attributes are computed on the edge itself and on its neighbourhood, determined by a Voronoï diagram. Its leads to a post-processing more suited than the global thresholding on the gradient norm. This ensures the obtaining of contours along regions of interest, even having poorly contrast.

3.2 Edge filtering using 3D regions

We reduce the set of 2D contour points to a restrained set of interest contour points corresponding to the frontiers of the 3D regions.
Edges are then polygonalised in linear segments, using a classical algorithm.

4 BUILDING RECOGNITION

The scheme that we propose is based on perceptual grouping techniques, as presented since many years (Mohan & Nevatia, 1992). Unlike these approaches, we include in the grouping process criteria on 3D information from the DEM.

4.1 Segment merging

Each region of the 3D segmentation (§ 2.2) is processed separately. Only linear segments and edges lying in the neighbour of the 3D region contour are considered, as explained in § 3.2.

Linear segments are associated taking into account their geometrical relationships (collinearity, vicinity) and the altimetric and radiometric properties of their neighbourhood.

For each region, the first step is the collinear segments merging.
Two collinear neighbouring segments are associated if they pass both conditions:
- radiometric neighbour coherence;
- altimetric neighbour coherence.

Radiometric and altimetric neighbour coherence are evaluated by giving a score depending on distances between radiometric (*resp.* altimetric) histograms in two fixed-size windows on each segment side. The window-size may be tuned according to *a priori* knowledge on the scene (expected building dimensions, and pixel ground resolution).

A vicinity parameter is set on the segment extremities, in order to limit the search combinatory. This parameter may also be tuned according knowledge on the scene.

This step allows us to remove short segments and to merge them in longer and more significant features, and because it relies on both radiometric and 3D data, it is more robust and well suited to our problem.

4.1 Segment grouping

New links between selected segments are carried out not using perceptual grouping or geometric rules but using the line of the 3D region contours. Segments are grouped along the line.
By this way, we can set up building hypotheses. If several close paths corresponding to different building hypotheses are set up, we select the hypothesis minimising inside segment covering and maximising the 3D contour line covering. When no segment exists in a large part of the 3D contour line, the line itself is used to complete the building description.

5 RESULTS

5.1 Image Data

We performed tests of the above described processing on several epipolar stereo pairs extracted from the same views provided by the french IGN (Institut Géographique National). These images, covering a part of the city of Colombes, near Paris, were digitised so that their ground resolution is about 8cm per pixel, giving two views having a common part about 10.000 × 6.000.

We then computed several 1.000 × 1.000 epipolar pairs from these views. Figure 2 shows two of these pairs.

On these areas, we also have some reference data, provided by the IGN from its topographic database. These reference data consist in the description of the limit of buildings, manually seized on mid-resolution images (figure 4.a).
A second database, unfortunately not available on the whole area, is the "TRAPU" database, also provided by IGN. These data describe buildings more accurately; each building is modelled by a set of planes representing façades, roof and significant superstructures (figure 5.a).

Figure 2 : Two epipolar pairs covering a part of the city of Colombes, France.

5.2 Results

DEM were computed on both pairs, and results are presented on figure 3, left column, DEM images are in the geometry of the left images on figure 2.
White pixels are those which have no correlation score (either belonging to hidden parts or not present on both images). The darker a pixel is, the higher its altitude value is.
The initial disparity intervals were 130 pixels (stereo pair 1, top row) and 100 pixels (stereo pair 2, bottom row) wide, and we use a 3-level matching strategy.

Both figures show a quite dense and accurate disparity map. In particular, one can note the fact that depth discontinuities are well preserved.

3D above-ground regions are then extracted on these altitude maps. Results are shown on figure 3, right column.
All the buildings or building groups totally present on the scenes are detected, and one can even note that connected buildings, but having different altitudes, are divided (stereo pair 1, top row).
As a drawback, 3D regions contours are not very accurate, and have to be refined on the building image boundaries, as explained in §4.

Final results are presented on figures 4.b (stereo pair 1), and 5.b (stereo pair 2).
Building boundaries can be compared to those from the database data. One can notice that linear edges coincide with building boundaries.
Nevertheless, not all boundaries are correctly extracted, and it remains some inaccuracies where we could not obtain linear segments from radiometric edges and had to use 3D region contours to close the building description. This may be corrected by introducing linear segments from both images, which could be back projected in the same image.

6 CONCLUSION

We have presented a set of procedures in order to automatically detect and localise buildings from a stereo-pair of aerial images. The image ground resolution is about 8cm per pixel, and the images are registered in epipolar geometry.
We compute DEM by mean of a multi-resolution and adaptive template matching algorithm.
DEM is then segmented and man-made raised structures are extracted from an adjacency graph of the 3D regions.
Radiometric edges and linear segments are extracted in one image. 3D regions are then used for contour grouping. A first grouping step select contours for each 3D region. Inside each region, a second grouping process links up linear segments when available, and image edges along the region frontier.

Tests on various 1.000 × 1.000 stereo pairs of aerial images show the efficiency of the method. All buildings present in the scene are detected, and their boundaries are generally well-localised. This allows us to envisage at this time the introduction of our procedures in a semi-automatic process of data bases control and constitution.

ACKNOWLEDGMENTS

Most of this work was done within a research contract between ETIS and the french Institut Géographique National "MATIS" department. We particularly thank Olivier Jamet (IGN-MATIS) for supplying images and reference data that we used in this paper.

REFERENCES

(Baillard et al., 1996) Baillard C., Dissard O., Jamet O. & Maître H., 1996. Extraction and characterization of above-ground areas in a peri-urban context. In: *Proc. of the IAPR/TC-7 workshop on methods for extracting buildings, roads and other man-made structures from images*, Graz.

(Cord *et al.* 1997) Cord M., Huet F. & Philipp S., 1997. Optimal Adjusting of Edge Detectors to Extract Close Contours. In: *Proc. of Scandinavian Conference on Image Analysis SCIA'97 Conference*, Lappeenranta, Finland, June 1997.

(Cord *et al.* 1998) Cord M., Paparoditis N. & Jordan M., 1998. Dense reliable depth discontinuity preserving DEM reconstruction from very high resolution aerial urban images. In: *ISPRS Commission II Symposium*, Cambridge, July 1998. (These proceedings)

(Frère *et al.*, 1997) Frère D., Hendrickx M., Vandekerckhove J., Moons T. & Van Gool L., 1997. On the reconstruction of urban house roofs from aerial images. In: *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Birkhaüser Verlag, Basel, pp. 87-96.

(Gabet *et al.*, 1994) Gabet L., Giraudon G. & Renouard L., 1994. Construction automatique de modèle numérique de terrain haute résolution en zone urbaine. *Bulletin de la Soc. Française de Photogrammétrie et de Télédétection*, n°135, pp. 9-25. In french.

(Girard *et al.*, 1998) Girard S., Guérin Ph., Maître H., Roux M., 1998. Design of a dense disparity map for building recognition in aerial images. In: *Proc. of RFIA'98 Conference*, Clermont-Ferrand, Jan. 1998, pp. 319-328. In french.

(Haala and Anders, 1997) Haala N. & Anders K-H., 1997. Acquisition of 3D urban models by analysis of aerial image, digital surface models and existing 2D building information. In: *Proc. of SPIE Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision III*, Orlando, pp. 212-221

(Henricsson *et al.*, 1996) Henricsson O., Bignone F., Willuhn W., Ade F., Kübler O., Baltsavias E., Mason S., Grün A., 1996. Project AMOBE: strategies, current status and future work. In: *International Archives of Photogrammetry and Remote Sensing*, Wien, pp. 321-330.

(Lammi, 1997) Lammi J., 1997. Automatic building extraction using a combination of spatial data and digital photogrammetry. In: *Proc. of SPIE Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision III*, Orlando, pp. 223-230.

(Mohan & Nevatia, 1992) Mohan R. & Nevatia R., 1992. Perceptual organization for scene segmentation and description. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, no. 6, pp. 616-635.

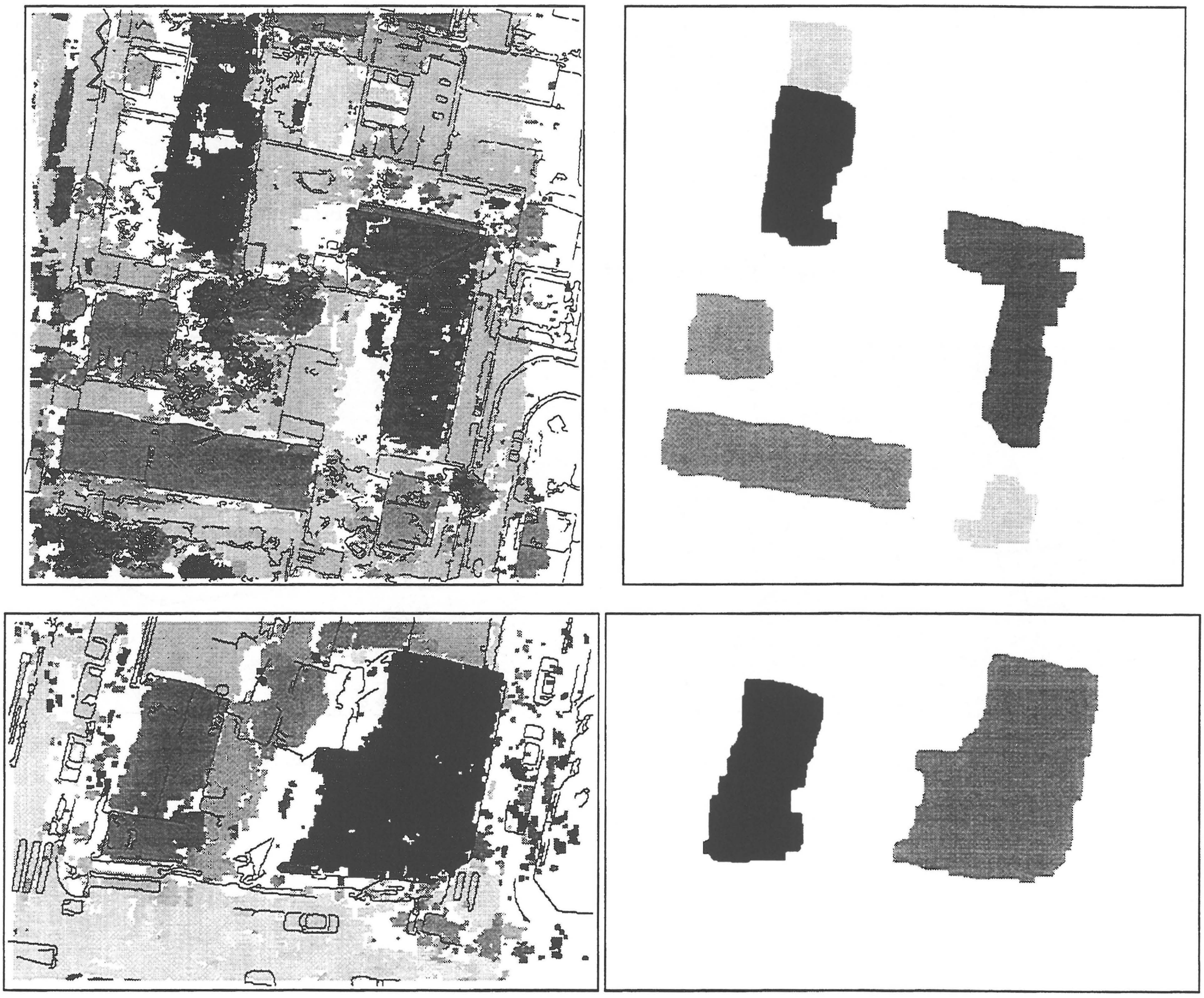
(Nevatia *et al.*, 1997) Nevatia R., Lin C. & Huertas A., 1997. A System for Building Detection from Aerial Images. In: *Automatic Extraction of Man-Made*

Objects from Aerial and Space Images, Birkhaüser Verlag, Basel, pp. 77-86.

(Paparoditis et al., 1998) Paparoditis N., Cord M., Jordan M. & Cocquerez J-P., 1998. Building Detection and Restitution from High Resolution Aerial Imagery, *Computer Vision and Image Understanding*, 1998, Accepted for publication.

(Vandekerckhove *et al.*, 1997) Vandekerckhove J., Frère D. & Moons T., 1997. Reconstructing House Roofs from High Resolution Aerial Images of Urban Scenes. Technical Report, Katholieke Universiteit Leuven, Belgium, no. KUL/ESAT/MI2/9703.

(Weidner, 1995) Weidner U., 1995. Building Extraction from Digital Elevation Models. Technical report, Uni. Bonn, Institut für Photogrammetrie, Germany.



a. DEM.

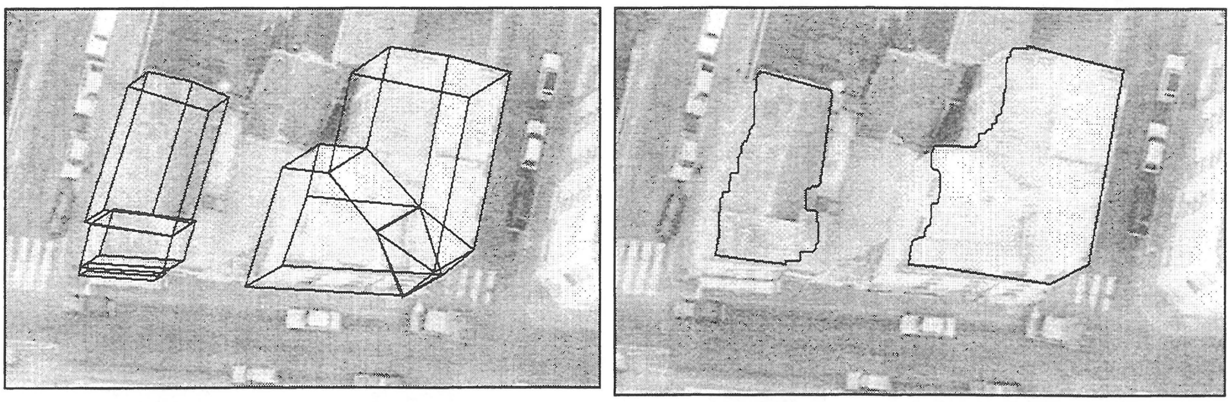
b. Man-made strcuctures.

Figure 3 : DEM computed on figure 2; 3D regions recognized as buildings

a. Reference from the IGN "BD Topo" database

b. Final results

Figure 4 : Reference and final results on stereo pair 1.



a. Reference from the IGN "Trapu" database

b. Final results

Figure 5 : Reference and final results on stereo pair 2.