AUTOMATIC DIGITAL TERRAIN MODEL GENERATION USING AERIAL IMAGES AND MAPS.

Michel Roux, Jaime Lopez-Krahe*, Henri Maître

Ecole Nationale Supérieure des Télécommunications
Département IMA, 46 rue Barrault
F 75634 Paris Cedex 13
email: mroux@ima.enst.fr
* now at Université Paris-8, Institut Intelligence Artificielle

Commission III, Working Group 3.

KEY WORDS: DEM/DTM, Vision, Cartography, Aerial Image Analysis, Urban Scene Interpretation.

ABSTRACT:

In this paper we present a new method for the generation of Digital Terrain Models (DTM) on urban areas using simultaneously a stereoscopic pair of aerial images and a scanned map. This method relies on the information given by the map on the position of the road network. Roads and crossroads are places where information on terrain elevation can be extracted directly from the aerial images.
The effectiveness of this approach is demonstrated with complex imagery on a urban area containing a large variety of different urban block types. The results are evaluated with respect to ground truth data.

1 Introduction

The automated generation of 3D cartographic databases from aerial images has become a major field of interest, with many applications in cartography, navigation, telecommunications, urbanism, etc [1]. However in the case of very complex scenes like dense urban areas the automatic process of the only aerial images leads to ambiguous solutions (many feasible interpretations can be made from the same elementary features). In order to help the image interpretation, existing maps provide useful information on the presence, shape and localization of various features in the scene [2, 3].

In this paper we present a new method for the generation of a Digital Terrain Model (DTM) of a urban scene using simultaneously a stereoscopic pair of aerial images and a scanned map. The map provides useful information on the places where the ground could be seen in the aerial images: mostly the road network and the surfaces without buildings (white surfaces of the map). We first describe briefly the analysis of the scanned map (road network extraction and urban block classification), then we depict in details our method for the generation of a DTM using the road network extracted from the scanned map. Results on a scene in the suburb of Paris are presented and compared to ground truth data.

This work was sponsored by the CNET-Belfort under Research Contract 94 PE 7216. Special thanks to MMs. René Mignone and Rafi Deryeghiyan from the CNET-Belfort.

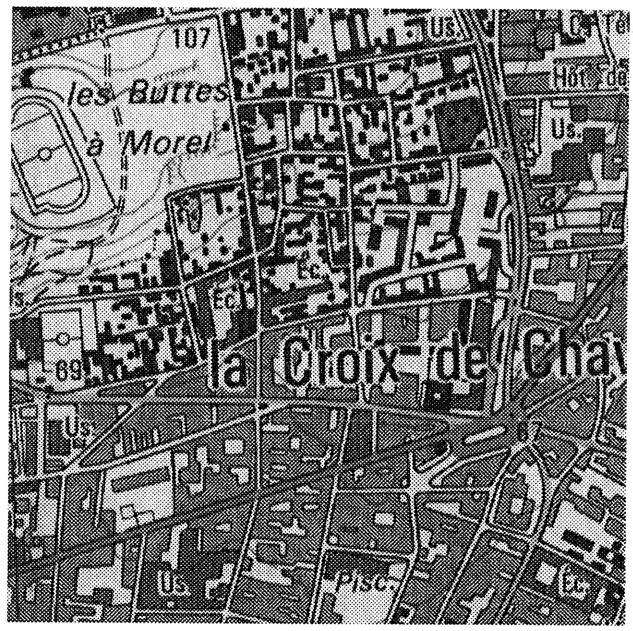

Fig. 1 - *Portion of the scanned map. ©IGN*
The complete scene is 2km x *2km.*

2 Map Analysis

For this application commercially available maps at the scale of 1:25000 have been purchased. They have been scanned at 300 dpi which gives a pixel approximatively equivalent to 2 meters on the ground. Because of the importance of color for the representation of various cartographic features (main roads, contour lines, forest, ...), maps were scanned in full color with 24 bit/pixel.

Information on various structures can be extracted from the map. For our application, the road network and the urban blocks (regions delimited by roads) are the most interesting features. Word extraction is also a source of information on natural or man-made features present in the scene. In the remaining of this section, road network extraction and urban block classification are briefly described. More details on road and crossroad extraction algorithms can be found in [4].

left image *right image*

Fig. 2 - *Portion of the stereo pair of aerial images.*

2.1 Road network

On IGN [1] maps at the scale of 1:25000 main roads are represented in red, while other roads remain in white. A simple colorimetric analysis allows the extraction of the main roads. A morphological opening removes contour lines represented with the same color, but with thinner lines.

Other roads are represented with two parallel black lines. Their extraction is performed using the road network connexity. Linear elements with a high probability to be road portions (i.e., white pixels bordered with black pixels) are detected on the scanned map. Links between these elements are created if it is possible to find a path from one to an other composed of only white pixels (at the crossroad level it is not possible to use the border to follow a road).

Both networks are merged, and only large connex components are kept in order to remove elements detected between black objects representing buildings. A linear approximation of the network gives a vectorial representation. Collinear segments are grouped to form roads.

At that point, an interactive tool allows us to correct the road network automatically detected. A small part of the errors are some lines detected inside the building blocks. Other errors correspond to missing road parts due to letters or city limit lanes overlapping the roads on the paper map.

Crossroads are detected by grouping the road junctions (X, T or L shaped junctions): a crossroad could be defined as only one junction, or could gather several junctions if they are close enough. Each road is then decomposed into sections: a section is the portion of a road between two consecutive crossroads. Crossroads and sections which are not in a cycle are then eliminated.

1. IGN: Institut Géographique National, the French national agency for cartography.

2.2 Word extraction and urban block classification

The first step for word extraction consists in the detection of black features, which represent letters or large buildings. Size and shape criteria allow the elimination of some elements representing buildings. Finally the grouping of horizontally aligned black features gives the words present on the map.

Except roads and letters, one can distinguish 4 different surface features on the scanned map: large buildings in black, administrative buildings (schools, city halls, ...) in dark grey, apartment building in light grey and building-free surfaces in white. Once letters and roads have been removed from the map image, it is possible to separate the different surfaces just by considering the pixel grey levels. One can notice that only 3 different tones (black, grey and white) were used for the representation of these regions: dark grey features are composed of black and grey pixels, and light grey features are composed of grey and white pixels. The separation of these 3 tones is done by a k-means algorithm that performs a thresholding operation on the image grey level histogram and divides it into 3 classes. Morphological operations (opening and closing) allow the grouping of the pixels of these 3 classes to form the 4 different kinds of surface features.

3 Digital Terrain Model

For this application a pair of stereoscopic images covering the same scene as the scanned map have been acquired. The resolution of both images is 1m/pixel.

Fig. 3 - *The road network extracted from the map is superimposed on the aerial left image.*

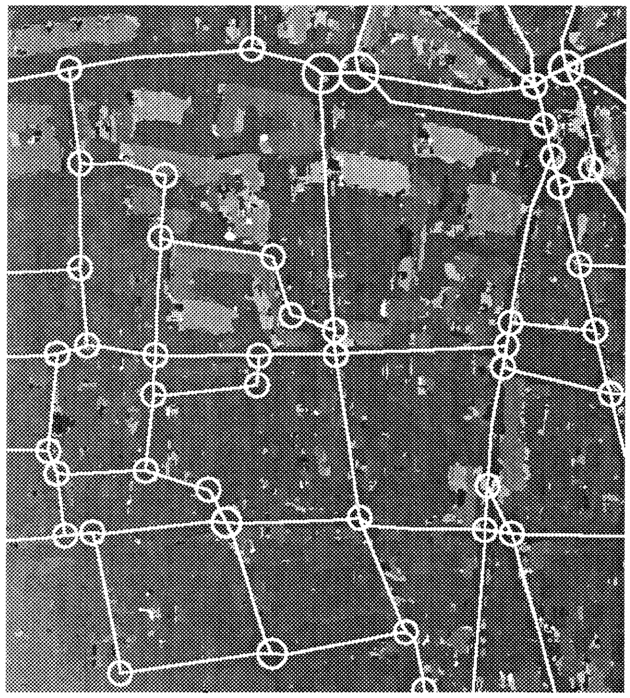

Fig. 4 - *Road and crossroads extracted from the map are superimposed on the disparity image.*

First the images are corrected into an epipolar geometry (see figure 2). Then geometrical relationship between both aerial images and the scanned map is determined: a polynomial transform of degree 2 is calculated using manually selected control points. These transforms are only used to transfer the vectorial representation of the road network extracted from the map into both aerial images as presented on figure 3, but no registration has been performed directly on the images.

A disparity map is calculated for the complete scene with a classical cross-correlation algorithm using a square window of 13x13 pixels. The result, as it can be seen on figure 4, is very noisy because of the presence of numerous hidden parts and because of the very large disparity range required for the complete scene (80 pixels).

The generation of the DTM is then split up into 3 steps. First disparity is calculated for each crossroad of the network. Then disparity is calculated along the road sections joining the crossroads in order to validate the disparity at the crossroads. Finally a dense map of disparity is calculated using the validated crossroads.

3.1 Disparity at the crossroads

In order to calculate the disparity at a crossroad we consider a large window (30x30 pixels) centered on each crossroad. A large window is required because of the lack of precision on the crossroad localization that can be caused by several factors: precision of the map itself, displacement during the road extraction operation and accuracy of the polynomial transform between the map and the aerial images (see figure 4).

The disparity histogram is calculated on this large window and its most significant peaks are considered (see figure 5). The peak given by the lowest disparity is selected as the ground point representative. Other peaks are usually provided by points at the top of

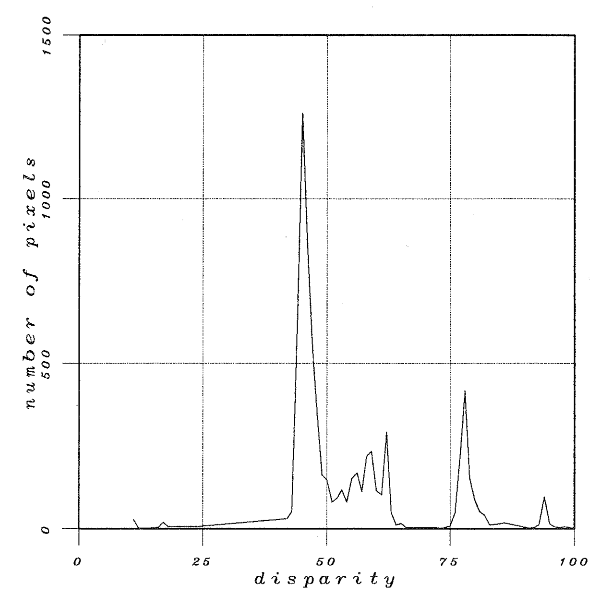

Fig. 5 - *Disparity histogram at a crossroad.*

buildings. The disparity of the selected peak is then recorded as the disparity of the crossroad.

3.2 Disparity along the road sections

Since crossroad disparity was calculated using only local information, we would like to validate this result with more global information, in order to remove wrong crossroads and to correct wrong values of disparity. The graph of roads provides the consistency required for this validation step.

Fig. 6 - *Paths in the disparity map using a dynamic programming technique.*

A road section is the portion of road between two consecutive crossroads. On the disparity map calculated previously we consider a large band along each road section. If it is possible to find a regular path in this band between the two extremities of the road section, then the road section is validated as well as the two crossroads.

The search for a regular path is performed by a dynamic programming procedure[5], that finds the best path from one crossroad to the other one. The cost of an elementary displacement in the disparity map between two neighboring pixels is equal to the absolute value of the disparity difference. The cost of a path from one crossroad to the other one is then the sum of these elementary costs. The dynamic programming approach allows to find the best path with respect to that global cost with $O(nm)$ operations where n is the length of the band and m its width. The best path is validated if it does not contain any disparity gradient.

This validation step is applied twice:

- in a first iteration, a constraint is applied in order to force the different paths to begin and to end at pixels with the same disparity as the two extremity crossroads;

- in a second iteration, the previously non-validated crossroads and road sections are examined again without this constraint on the extremities. This allows to correct wrong values of disparity at some crossroads.

Figure 6 shows several paths found by this dynamic programming procedure, white lanes correspond to validated paths and black lanes to non-validated paths. One can notice that some paths not validated during the first iteration because of a wrong disparity value at the crossroads are validated during the second iteration (lower left part of the figure). Wrong disparity values for these crossroads were mainly due

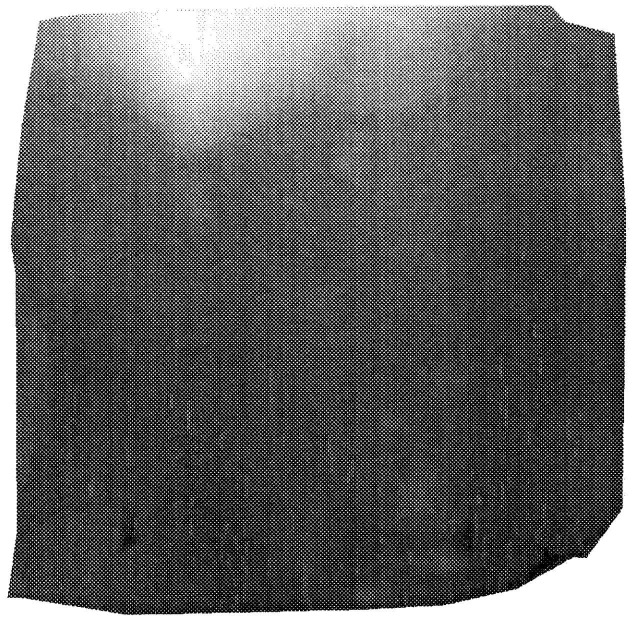

Fig. 7 - *First digital terrain model before the road section validation.*

to misplacement of the road network extracted from the map.

3.3 Calculation of a dense image of disparity

At this point of the process, disparity is known only for a few points: the validated crossroads and road sections. In order to obtain a dense map of disparity, a 2D Delaunay triangulation is performed on the validated crossroads. The disparity at each point is then interpolated considering that each triangle represents a planar surface: the plane equation associated with each triangle is calculated with the 3 coordinates $(u, v, disparity)$ of its corners.

A more sophisticated approach would consider complex 3D surface models like Bernstein-Bézier patches in order to obtain G^1 continuous surfaces[6].

4 Test scene

Results are presented on a scene in the suburb of Paris. This scene presents a large variety of construction density and buildings of various sizes and shapes. A portion of the scanned map is presented on figure 1, while figure 2 shows a small but significant part of the stereo pair of aerial images.

4.1 Results

The road network extraction provides, after the manual correction, 369 crossroads and 612 road sections. A polynomial transform of degree 2 between the map and the left aerial image has been calculated with a least mean square approximation on manually selected control points. On figure 3 the road network extracted from the scanned map has been superimposed on the left aerial image.

Disparity at crossroads was calculated as described in section 3.1. The DTM calculated without the validation process along the road sections is presented on figure 7. One can notice some obvious errors given

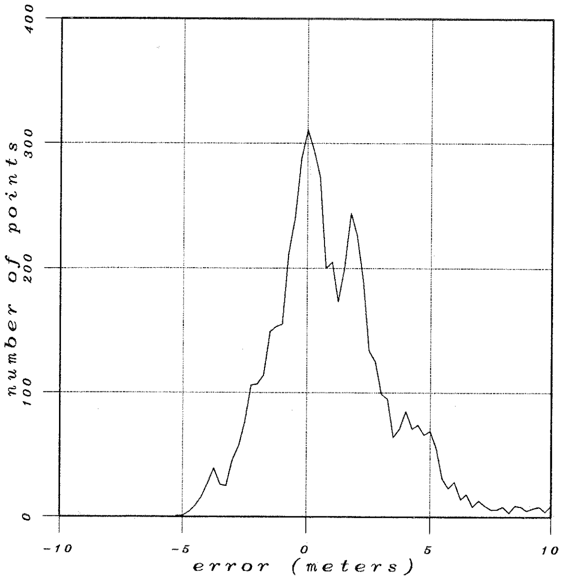

Fig. 8 - *Error histogram between ground truth and generated DTM.*

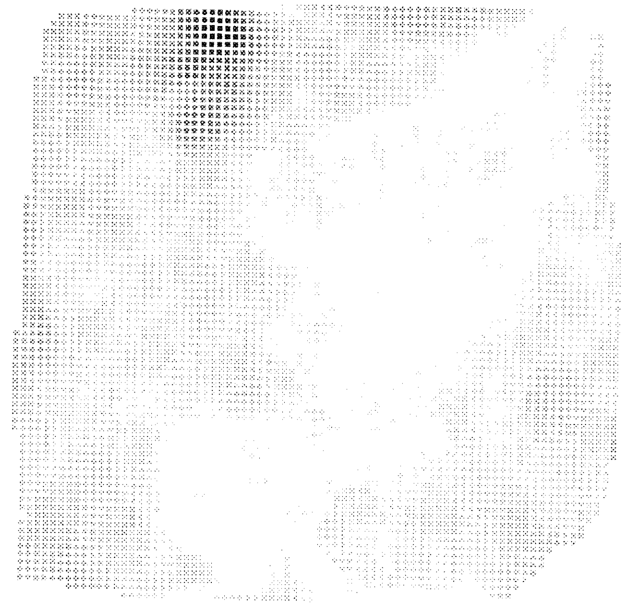
Fig. 9 - *Error distribution in the scene, dark points indicate large errors.*

by wrong disparity values at crossroads. These errors generate pitches or holes in the DTM.

The road section validation process confirmed the disparity for 330 crossroads after the first iteration and for 348 crossroads after the second iteration. Figure 10 presents the final DTM on which red features of the map (main roads and contour lines) have been superimposed. One can notice that the terrain model fits very well the contour lines, and that the errors previously seen on figure 7 have been removed.

4.2 Performance analysis

Evaluation of these results has been made against a ground truth given on grid of 25m x 25m resolution. The left image in which referential the DTM is calculated has been calibrated with these ground truth data. Each point in the ground truth is given by its coordinates (X, Y, Z) in a cartographic referential. The calibration gives the coordinates (u, v) of the corresponding pixel in the left aerial image. It is then possible to compare the ground truth elevation Z of the point with the elevation Z_{dispa} given by the calibration of the pixel disparity. The error at a point is then defined as the value of the difference between these two elevations:

$$e = Z_{dispa} - Z$$

This error has been calculated for the 5489 common points of the ground truth and the generated DTM. The mean of the error absolute value is equal to 2.10m. Figure 8 presents the error distribution for the scene. As it could be expected this distribution presents a peak centered on zero, but one can notice that other significant peaks are present for positive error values. This indicates that our elevation estimation method tends to over-estimate the altitude of points.

At this point of our research further considerations should be undertaken in order to identify these errors more precisely. Figure 9 presents the error distribution in the scene. We can first notice that errors occur on the sides of the scene where crossroad disparity is validated with less road sections (see 3.2). But largest errors appear on the highest region of the scene where less crossroads are present, because of a building-free area. Future developments for our DTM generation system consist in the analysis of the white areas of the map in order to compensate the lack of crossroads in these regions.

5 Conclusion

In this paper we describe a new method for automatic digital terrain model generation using information provided by scanned maps and stereo pairs of aerial images. Scanned maps provide useful information on the location where the terrain could be directly seen on aerial images, mainly roads, crossroads and building-free areas. We demonstrated theses ideas with complex urban scenes presenting a large variety of construction density and buildings of various sizes and shapes.

Further developments are under progress in order to improve the quality of the DTM:

- use of Bernstein-Bézier patches in order to obtain a G^1 continuous surfaces,
- finer disparity analysis in building-free areas in order to compensate the lack of crossroads.

DTM generation is the first part of a more global system dedicated to the construction of 3D cartographic databases using simultaneously aerial images and scanned maps. Further developments are dealing with the generation of building 3D models, for which scanned maps are also a rich source of information on the presence, shape, size and localization of buildings in a dense urban scene.

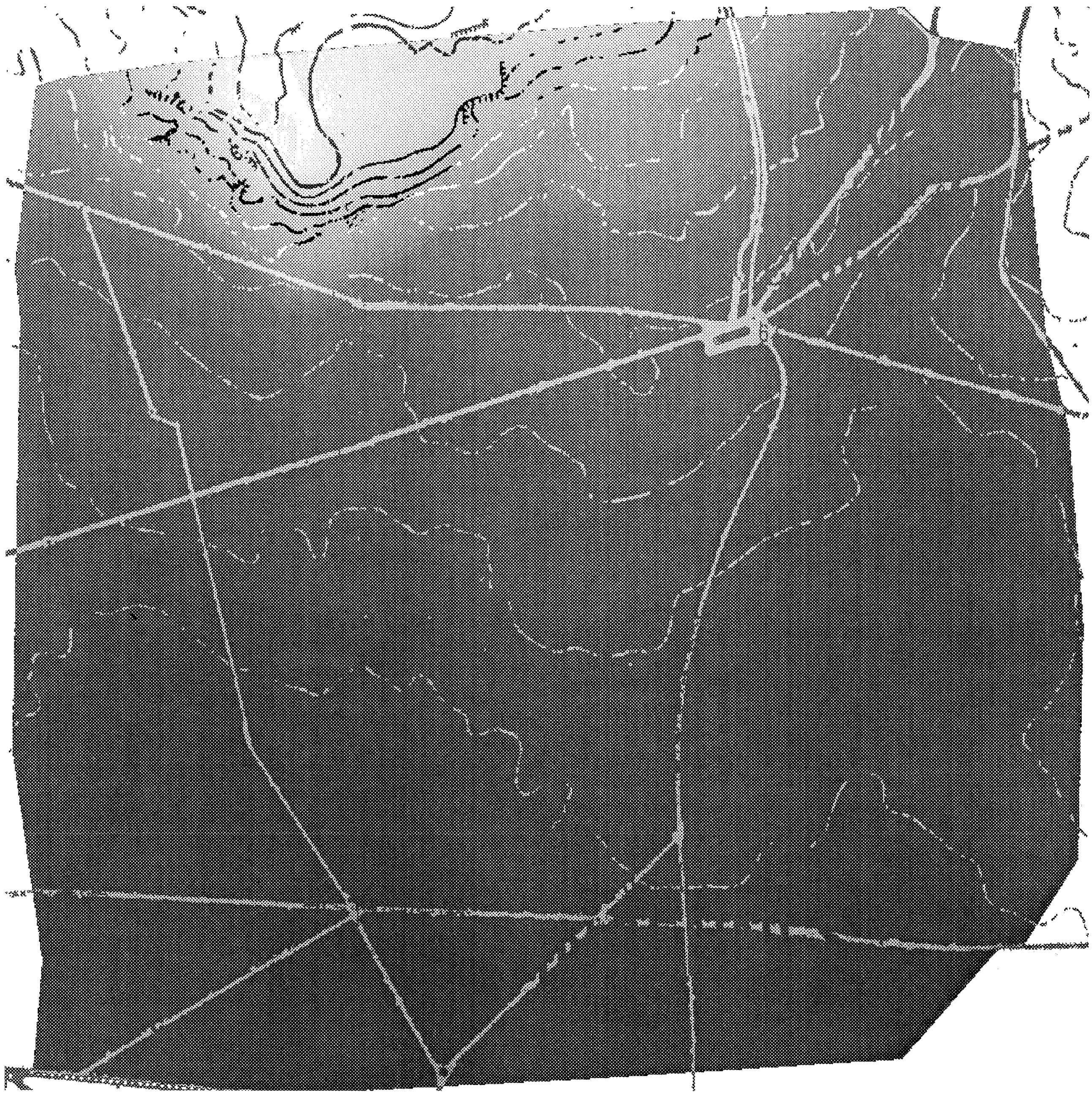

Fig. 10 - *Final digital terrain model overlapped with the contour lines extracted from the map.*

References

[1] A. Gruen, O. Kuebler, and P. Agouris. *Automatic Extraction of Man-Made Objects from Aerial and Space Images*. Birkhäuser Verlag, 1995. Ascona Workshop Proceedings.

[2] H. Moissinac, H. Maître, and I. Bloch. Graph based urban scene analysis using symbolic data. In *SPIE: Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision*, volume 2486, pages 93–104, Orlando, April 1995.

[3] H. Moissinac, H. Maître, and I. Bloch. Markov random fields and graphs for uncertainty management and symbolic data fusion in a urban scene interpretation. In *SPIE: Image and Signal Processing for Remote Sensing II*, volume 2579, pages 298–309, Paris, September 1995.

[4] M. Roux, J. Lopez-Krahe, and H. Maître. Recalage image SPOT / carte routière. In *XVIIe Congrès ISPRS, Washington, International Archives of Photogrammetry and Remote Sensing, Commission III*, volume XXIX, pages 384–391, August 1992.

[5] N. Merlet and J. Zerubia. New prospects in line detection for remote sensing images. In *ICASSP, Adelaïde, Australia*, April 1994.

[6] G. Farin. Triangular Bernstein-Bézier patches. *Computer Aided Geometric Design*, 3:83–127, 1986.