Automatic Extraction of Man-Made Objects from Aerial and Space Images,
Birkhäuser Verlag, Basel, 1995, pp. 231-240

Building Extraction from Stereo Pairs of Aerial Images: Accuracy and Productivity Constraint of a Topographic Production Line

Olivier JAMET, Olivier DISSARD, Sylvain AIRAULT

Institut Géographique National
Direction Technique, Laboratoire MATIS
2, avenue Pasteur ,94 160 Saint-Mandé, France
Ph.: +33-1-43988069, Fax: +33-1-43988581
e-mail: jamet, dissard, airault @ matis.ign.fr

Abstract

This paper presents the research that has been lead for a few years at IGN on automatic building extraction from aerial photographs. A description of the operational context and of the topographic application that is aimed at introduces the main constraints to be overcome. We give then a quick sketch of the building extraction process that has been worked out up to now, and develop its quality assessment on a test site. The analysis of the results underlines the part played by the planar approximation of disparities resulting of an area-based correlation process in the fulfilment of the accuracy requirements. Further research will focus on reliability improvement and on more complex shape handling.

1. Operational context

1.1. The Topographic Database

As a national mapping agency, the French National Geographic Institute (IGN) is mainly dedicated to the acquisition and updating of the national reference geographic data, including the geodetic network, a periodic medium scale photographic coverage of the territory, and the topographic mapping.

Till 1984, the reference topographic data has been the set of the 2000 1:25000 maps covering France, stored as paper map sheets as well as original films. In 1984, facing the increasing needs for geographic digital data, IGN decided to provide France with geographic databases at various scales. The smallest scale ones were to be produced from the digitisation of existing maps, while the topographic database (BDTopo®), that aimed at replacing the topographic reference, was to come from a new photogrammetric plotting of the territory.

This complete refection of our topographic data will provide its users with higher accuracy data than the former maps (1.5 m rms in XY on the average), each point being stored in 3D with a Z accuracy of 1m rms. The data model includes all information already present on the maps (contours lines, communication and hydrographic networks, buildings, vegetation...) in vector form, within a topological structure, each object supporting attributes and relations. The BDTopo® will thus provide a complete topographic data set, oriented, from its content, toward GIS applications, and dedicated to the scales 1:25000 up to 1:5000, according to its accuracy.

1.2. Evolution of the production capacity

The data input actually started in 1989, and very soon, the first analytical "integrated production lines" (LPI) were settled up. This LPI concept, gathering within a same team the photogrammetric data input, the complementary ground survey, and the GIS management, was designed in order to optimise productivity and to allow a better quality control along the process. Each LPI is composed of four analytical plotters, one data server, a 2D workstation for GIS correction and ground survey input, and a plotter for control drawings. It represents a staff of 15 people (two people per stereoplotter, topographers, controllers...), and can produce five to six 550 km^2 work units a year (a work unit corresponds to the area covered by two 1:25000 map sheets).

In march 1995, 6 LPI are currently working, and 6.5% of the territory (for 43% of the population) are available. This production capacity should extend up to 12 LPI with an average rhythm of 2 LPI a year, and will then correspond to an updating capacity of whole France every 8 years (on the average, the most evolutive areas being updated every 3 or 4 years). Within this production frame, the complete coverage of France will require 25 years, and represent 2 millions hours of stereoplotting. To meet faster the user needs, IGN aims at reducing this delay down to 15 years. If the main way to reach this ambition keeps promoting the use of the data, so as to fund part of the plotting task within private companies on commercial income, IGN also decided to invest in digital techniques. In 1991, a research program was started up in image analysis to provide semi-automatic helps to the plotting, while more recently (in 1994), an evaluation of the available softcopy stereoplotters was undertaken, and lead to the choice of digital plotters for the equipment of the forthcoming LPI.

1.3. Data input

The data acquisition is performed on 24 cm panchromatic aerial stereo pairs at the scale 1:30 000 in rural areas, and 1: 20 000 in urban areas. The focal lengths were chosen to fulfil the Z accuracy requirements (152 mm in rural areas and 210 mm in urban ones). The photographs are taken by strips with a 60% overlap (and 10% overlap between the strips). On each working unit, the LPI are supplied with the photographs and with the external orientations computed by a triangulation process. For the softcopy plotters, the photographs will be digitised at a step of 15 to 20 microns, and a quantification in 256 grey levels.

1.4. Research aims

The image processing research program that was settled up in 1991 aims at providing softcopy stereoplotters with semi-automatic tools in order to speed up the data input. The part to be played by IGN in this field corresponds to leading our own academic research as well as ensuring an interface between external research teams and our needs. It stays however within the research field (extended to validation studies): operational developments will be left to contracted industrial partners.

In first analysis, focus was given to three major topics: contour line production from automatic correlation techniques, road extraction, and building extraction. Starting from a small team in October 1991, the activity was to increase up to 15 people in 1995 within the laboratory "Méthodes d'Analyse et de Traitement d'Images pour la Stereorestitution" (MATIS, created in 1993), with a goal of 18 people in 1997. Keeping the same main topics of interest, the work is currently shared between long term research (5 PhD studies) and short term activity (evaluation of current algorithms in accordance with the production staff).

2. Constraints analysis

2.1. Data cost

The first constraint of this applicative context concerns of course the data cost. Within the frame of the BDTopo® initial acquisition, the *a priori* available information is restricted to the currently used aerial photographs. Any additional information input has to be confronted to a comparison between its cost and the productivity it brings. On the average, taking only into account the photographic and the digitisation processes, a single image represents from 2.5 to 5% of the cost of the photogrammetric plotting of one stereo pair. This means that doubling the image scale (using 4 times more images), for instance, would only justify under a productivity increase of 10 to 20%.

In the long term, expecting the use of digital cameras, this over-cost will be much lower - so we can't utterly ignore this kind of possibility. But in the short term, that is within the frame of initial acquisition, the likeliness for such improvements of the productivity to rely on changes in the used imagery is weak: the main limits of the road extraction or contour line extraction algorithms from stereo pairs at the scale 1:30000 rely more on the lack of high level understanding of the scene (interpretation of complex landscapes) than on the lack of the data available to simple algorithms; the only topic which would maybe request larger scale imagery would be the building extraction.

2.2. Work flow management

We can't expect to modify the general work flow within a LPI. Splitting the operator work into several phases is likely to lower the general efficiency in a greater proportion than all the improvement that is to be expected from automatism. This implies:

on the hypothesised buildings, based on the grouping techniques of Mohan and Nevatia (1988):
- first, in order to filter noise on the buildings boundaries; on the first hand, geometrical shapes have to be identified to describe topographic objects; on the other hand, the boundaries of the roofs cannot always result from the segmentation process (in presence of low contrast with the context);
- secondly, to give more clues for the acceptance or rejection of the building hypothesis.

Applied first on rectangular buildings with horizontal roofs, the method was complemented with a mixed surface approximation technique, and an edge selection before the grouping (Dang 1994). Further improvements were added by Olivier Dissard (1995), particularly through a study of the behaviour of the correlation process and of the grouping, according to the content of the scene (orientation of the shadows, relative position of the occlusions...).

A scheme illustrating the current process is presented below (Fig. 1).

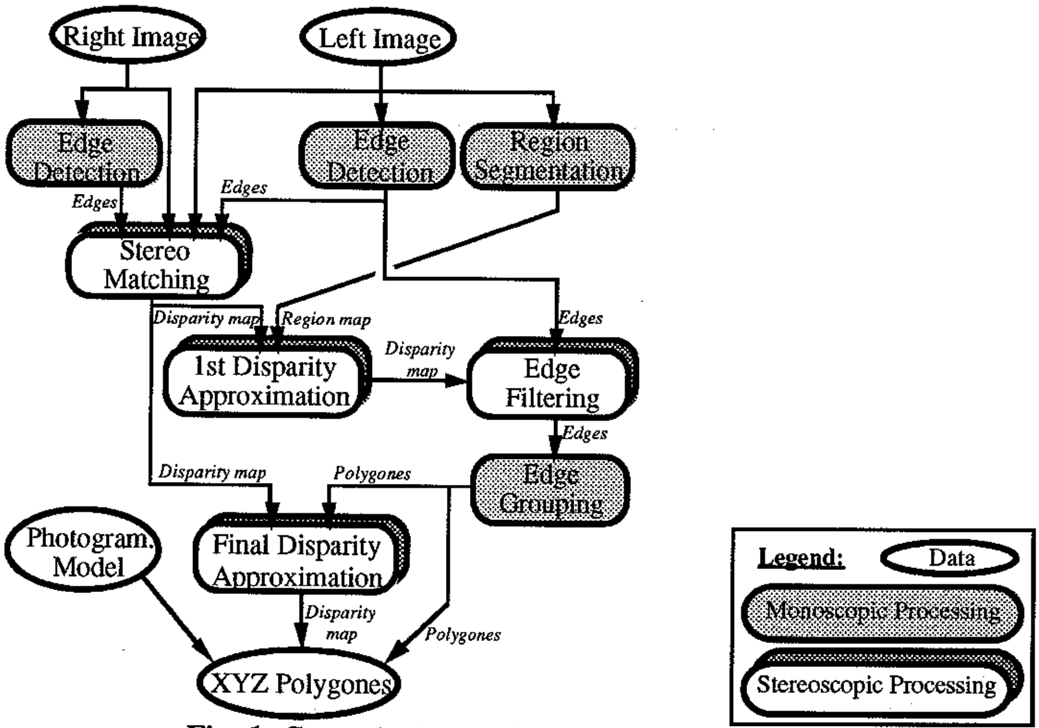

Fig. 1: General scheme of the building extraction process

4. Assessment of the process

4.1. Quality assessment

The method was evaluated on a test site located in the suburb of Paris. The photographs used for the BDTopo® acquisition were digitised at a step of 20 microns (40 cm ground resolution). The outer orientation was provided by the production department, and the vector data issued of the manual plotting was used as reference data. Forty one buildings of large size and rectangular shape were selected as test set. The results of the automatic extraction are illustrated by the examples provided on figure 2.

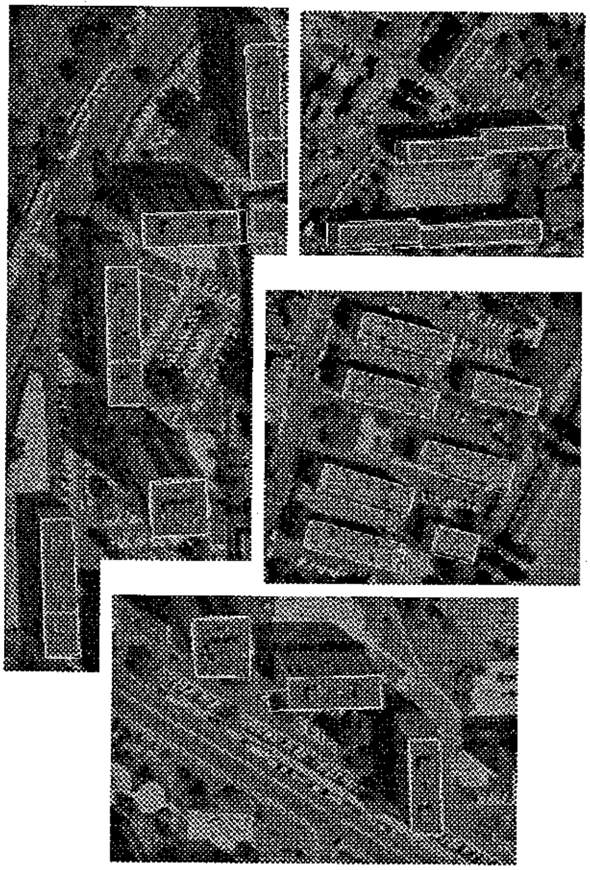

Fig. 2: Projection of the extracted buildings on the left image (extracts)

Accuracy measurements were performed on the corners of the detected buildings (according to the BDTopo® quality requirements), in XY on the monoscopic extraction in the image geometry, as well as in XYZ on the final results in geographic coordinates. This evaluation is summed up in the table 1.

As far as the exhaustivity is concerned, the test set keeps of course too small to provide definitive conclusions. 30 buildings out of the 41 were correctly plotted, 4 were rejected (unrecognised) and 7 failed.

Rejection correspond to misbehaviours of the stereoscopic part of the process, either due to segmentation errors (low contrast between the roof and its environment) or to correlation errors. In theses cases, the presence of a building is not detected (grouping rejection after filtering of the edges included in equal disparity regions).

41 Buildings / 178 corners	
Rejection rate (% number of corners)	9.0 %
Failure rate (% number of corners)	7.9 %
X rms	0.75 m
Y rms	0.71 m
XY rms	1.02 m
Z rms	0.60 m

Table 1: Quality measurements

We call failures extracted buildings presenting shape errors. Theses errors are mainly due to the fusion of a visible vertical face with the roof during the grouping process (enlargement of the roof: see the top right extract on fig.2), and affect generally only one side (even sometimes only one corner) of the building. The same effect can be observed

in presence of vegetation in the near neighbourhood of the building (see the bottom extract on fig.2). More seldom, some building can be shortened, equally during the grouping phase (closure of a rectangle with pieces of edges within the roof).
The accuracy results provided on table 1 were computed after exclusion of the failures (in fact, we only excluded the misinterpreted corners of the failed extractions). The 3D accuracy shows compatible with BDTopo® requirements:
- the XY accuracy keep low (2.5 pixels rms), but satisfies the 1.2 m that was aimed at;
- the Z accuracy is pretty satisfying (less than 1 disparity pixel rms).

4.2. Comments on the exhaustivity of the process

This study for an automatic process for building extraction was first of all motivated by the need to evaluate the performances of the algorithms before any analysis of what could be the real operational use of them (that is the user interface).

We can however consider this first method as a possible preprocessing doing part of the plotting, the results of which would be presented to the operator during the interactive plotting phase. If the accuracy results quite fulfil the BDTopo® requirements, exhaustivity keeps rather low. The rejection rate isn't by itself the main weakness of the method, since exhaustive extraction isn't presupposed. Failures may cause much more supplementary work: the interactive correction of objects as simple as rectangular buildings will probably cost more time than their entire manual plotting, and their simple deletion can even be heavy if they are connected to neighbouring objects.

A reduction of the failure rate (even to the "benefit" of a higher rejection rate) which could be reached by working out a confidence measure from complementary image processing (for instance by using the building shadows as a way to verify the consistency of the extraction, as proposed by Irvin and McKeown in 1989), would, in our application context, bring a serious improvement to the method.

4.3. General conclusions on building extraction techniques

The proposed methods will not, of course, be implemented "as is" on our softcopy plotters. First of all, the range of handled building shapes is still narrow, and the consequent low productivity improvement doesn't make software developments worthwhile. Secondly, as discussed in the previous section, progress is still to be done as far as the automatic control of the results is concerned. This work brings however some interesting pieces of information for the orientation of our further works.

Area-based vs feature-based matching. The first conclusions concern the use of area based matching techniques on this kind of imagery. The scale constraint, which bounds the resolution to values upper than 30 cm (40 cm on our test site), associated with a rather low B/H ratio (0.46 in the case of our test site), make the aimed Z accuracy a real challenge. Many authors have proposed building reconstruction techniques based on feature matching techniques. On our test site, supposing the feature location errors

independent on both images, a one metre Z accuracy would suppose a feature position accuracy better than 0.7 pixel. Knowing that the features of interest are *in fine* the corners of the buildings, such a quality seems difficult to reach (even visually, the correct position of a building corner can appear ambiguous). This very Z accuracy was reached by an estimate of the elevation based on area-based correlation and surface models.
As a confirmation of this analysis, we performed two complementary tests. First, we computed the 3D building rectangles issued of the two monocular groupings on the left and the right image. Despite the imperfection of the process (the double monoscopic grouping is of course less exhaustive than a 3D grouping procedure), the accuracy of the obtained building is representative of the accuracy we can expect of a feature matching process (table 2): the difficulty for locating correctly the edges of the building comes mainly from the need for grouping neighbouring parallel edges to cope with unavoidable ambiguities in presence of multiple detection (as in the case of buildings bordered by a thin white wall on fig.2), or of perturbation in the surrounding (especially vegetation).

Estimation technique	Z rms
Matched rectangles (17 buildings)	2.98 m
Surface model on manual plotting	0.54 m
Surface model on automatic plotting	0.60 m

Table 2: Z accuracy according to the estimation technique

Secondly, we computed the elevation resulting from the surface model approximation within the manually plotted rectangles, in order to have an estimation of the sensitivity of the approximation to an incorrect delineation of the buildings. The accuracy of this new Z estimate is slightly better (table 2), but the low difference with the accuracy obtained with the automatic rectangle extraction shows the robustness of the surface model technique.

Error propagation handling in the correlation phase. The correlation algorithm makes use of the same classical geometric constraints as many others: matches are performed along the epipolar lines (which are known with a subpixelar accuracy under the production conditions), with an order constraint, and a figural continuity constraint based on edges (Hoff 1989). The behaviour of this kind of algorithm on the building roofs highly depends on the direction of the computation (right to left or reverse along the epipolar line), due to the propagation effect induced by the order constraint (the match for a point is to be looked for after the match of its predecessor).

This is illustrated by the figure 3: one direction shows more favourable to high elevation surfaces (Fig.3-d), while the other one to the ground and the low elevation surfaces (Fig. 3-c). This behaviour is rather deterministic. The correlation matches quite well the shadows (and thus the ground) when they are first examined, and jumps more easily on the roofs otherwise.

Up to now, we cope with this problem by merging both results (Fig. 3-e). From an operational point of view, this default keeps bearable on large building, if the results we obtained on our test site are confirmed on larger areas. Small buildings seem however out of reach of this correlation process: even after the combination of the two results, the noise keeps too high to handle small surfaces. This limitation of the process can raise up the interest we may have in feature matching techniques.

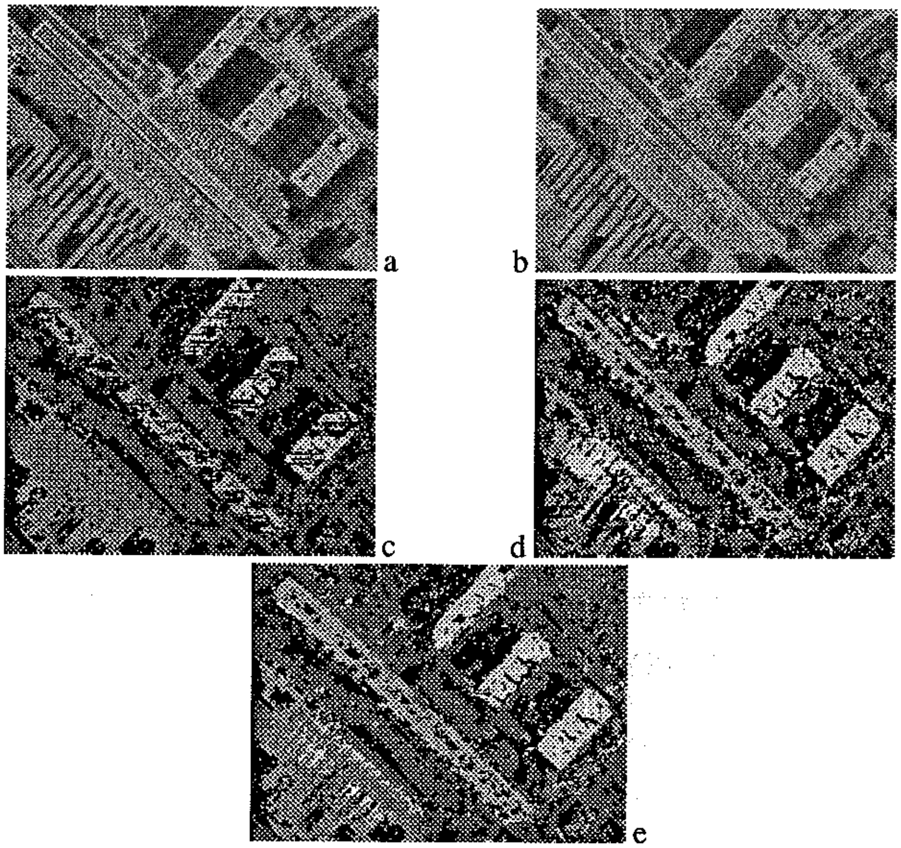

Fig. 3: Behaviour of the correlation process according to the propagation direction
a,b: aerial stereo pair; c,d: correlation results; e: merging of c and d

Non linearity in the disparity estimation. The median filter proposed by T.Dang (1994) in the disparity approximation phase proved to be efficient. In the current implementation of the method, surface model approximation is first tried by least square adjustment, in order to allow non horizontal plane handling; median approximation is tested afterwards, in case of failure. On the test data, median filtering was finally used most of the time, due to its lesser sensitivity to non gaussian noise. During the first disparity approximation, segmentation errors on the delimitation of the regions induce non gaussian noise in the disparity map within a region. Besides, even after the edge grouping, non gaussian noise can still occur within the roofs (due to the presence of staircase accesses to the roof, of chimneys...).

This latter effect is inherent to any object recognition problem. The object model is necessarily a simplification of the real world, suppressing all the details that will be too small, at the imagery resolution, to be correctly extracted, but that can affect any basic detector (such as an automatic correlation process) at the pixel level. A research effort