STATISTICAL SNAKES FOR BUILDING EXTRACTION FROM STEREOSCOPIC AERIAL IMAGES

H. Oriot

[a] ONERA, DTIM/IOR BP 72, 29 avenue de la division Leclerc 92322 Châtillon Cedex, France - (helene.oriot@onera.fr)

Commission III, WG III/2

KEY WORDS: 3D Reconstruction, Building, Stereoscopy, Aerial images, Statistical active models.

ABSTRACT:

3D modelling of urban objects is widely used for Geographic Information System, telecommunication or defence applications. The most classical technique to obtain such models consists in using several aerial images of the same area. In this paper we propose a new semiautomatic method to delineate buildings which is based on statistical active models. This method aims at limiting the number of images used to extract buildings –it only uses a stereoscopic pair of images-, and has been thought to limit the number of interactions in order to be easily upgraded into an automatic approach. An operator selects an area of interest with one building. The building shape is then automatically estimated by segmenting the disparity space into two areas: building and background. This is done by finding a regular polygon, close to the intensity image edges that minimises a correlation cost on the building roof, a correlation cost on the ground and that defines an occluded area with poor correlation costs. Good results are shown on a stereoscopic pair of images of Amiens (France). This algorithm is easy to parametrize, it finds solutions far away from the initialisation and buildings are described accurately when there are no trees close to the building.

RÉSUMÉ:

Les modèles 3D en site urbain sont largement utilisés dans les systèmes d'information géographiques, les télécommunications et les systèmes de défense. La technique la plus classique pour obtenir ces modèles consiste à utiliser plusieurs images aériennes de la même scène. Dans cet article, nous proposons une nouvelle méthode semi-automatique à base de modèles déformables statistiques pour délimiter les bâtiments. Cette méthode permet de limiter le nombre d'images utilisées –un couple stéréoscopique uniquement-, et limite aussi le nombre d'intéractions de l'utilisateur afin de pouvoir facilement être rendue automatique. Un opérateur sélectionne une région d'interêt conprenant un bâtiment. La forme du bâtiment est ensuite estimée automatiquement par une segmentation de l'espace des disparités en deux zones : bâtiment et sol. On recherche un polygone régulier, proche des contours de l'image d'intensité qui minimise un coût de corrélation sur le toit du bâtiment, un coût de corrélation sur le sol et qui définit une zone d'occultation sur laquelle le coût de corrélation est médiocre. De bons résultats sont montrés sur un couple d'images stéréoscopqiues d'Amiens. L'algorithme se paramétrise facilement, permet de trouver des solutions éloignées de l'initialisation et décrit précisment les bâtiments tant que ceux-ci sont éloignés de bosquets d'arbres d'altitude proche.

1. INTRODUCTION

3D modelling of urban objects is still very important for several kinds of applications (Geographic Information Systems (GIS) for urban planing or telecommunication, mission planning). The most classical technique to obtain such models consists in using several aerial images of the same area. Despite constant efforts from the image processing community, an automatic system for building extraction using only images is still not available (Forstner, 1999). In this paper we propose a new semiautomatic method based on statistical active models to delineate buildings. This method aims at limiting the number of images used to extract buildings –it only uses a stereoscopic pair of images-, and has been thought to limit the number of interactions in order to be easily upgraded into an automatic approach. After having described related work (section 2), we present our method and its implementation (section 3), we then show results on a real scene (section 4) and conclude on its performances and potentialities.

2. PREVIOUS WORK

Structure extraction from one or more aerial images has been studied for more than 20 years. Methods differ according to the nature of the data used: structures can be extracted from images (Baillard, 2000), from Digital Surface Models (DSM) (Weidner, 1995), or using both DSM and images (Paparoditis, 1998). Methods also differ depending on the prior knowledge used: geometrical constraints, building models which range from rectangular models (Weidner et al., 1995, Oriot et al., 1998, …), to polyhedral shapes (Scholtze et al., 2002), or agglomeration of rectangular parts (Haala et al., 1999).
C. Baillard (Baillard et al., 2000) presents a data-driven approach. She uses 3D segments found by a matching algorithm on several images and retrieves half 3D planes passing through these lines using a correlation algorithm. Half-planes are then grouped to complete roofs. S Scholtze (Scholtze et al., 2002) uses the same input data (3D segments) and retrieves roof structures using a bayesian approach. The reconstruction algorithm deals with inaccurate and imprecise input data. A.

Fischer (Fisher et al., 1997) uses different kinds of input data (points, lines, regions) and several levels of abstraction to describe a building (feature aggregate, building parts, building). An algorithm based on hypotheses generation and verification is used to reconstruct the buildings.
U. Weidner (Weidner, 1995) from the Bonn institute of Photogrammetry, presents a mixed approach (data-driven and model-based) using only a DSM. His method is divided into 2 steps: ground/above-ground segmentation and planimetric delineation of buildings using either a polygonal model obtained from segmentation and simplification of the DSM, or parametric models. He chooses between several parametric models and the data-driven extraction using the Maximum Description Length (MDL) method.
Most work deals with roof extraction from DEM and differentiate the detection and extraction steps. Let us mention M. Ortner (Ortner, 2002) who presents an alternative approach: his method detects and extracts buildings at the same time. A point process is introduced to deal with the generation, optimisation and destruction of a building hypothesis. So far only rectangular buildings with a symmetrical roof are considered.
In another context, The Fresnel Institute proposes a segmentation method using statistical active models (Chesnaud, 2000). C. Chesnaud segments an object and the background by modelling the statistics of the image. The segmentation is viewed as an estimation problem of the object shape in the image. The object edge is modelled as a polygon and the shape is estimated using statistical techniques. The algorithm developed is based on the calculation of the statistics of the inner and the outer regions (defined by the snake) (Chesnaud, 1999). In his PhD thesis, C. Chesnaud gives a description of a rapid implementation of the algorithm. We propose to adapt the stochastic segmentation problem to the disparity segmentation under photometric constraints in order to find precisely building planimetry using only a stereoscopic pair of images.

3. METHODOLOGY

We suppose we have a stereoscopic pair of images sampled in epipolar geometry and that we have selected an area of interest with a rough initialisation of the building shape. This initialisation may be obtained using the interactive method described in (Michel, 1998) or using a ground/above-ground segmentation (Baillard, 1997). We also suppose that we know the disparity interval of the building roof and an estimation of the disparity level of the ground around the building. Again, this data may be directly obtained interactively using (Michel, 1998) or automatically using (Baillard, 1997).
The building shape is then estimated by segmenting the disparity space into two areas: building and background. This is done by finding the polygon that minimises an energy defined in the correlation coefficient space over the area of interest. This energy is explained in the next paragraph.

3.1 Energy

Let I_l be a small image centered on the building of interest. Let I_r be the same area on the second image.
Let Δd_b be the disparity interval of the roof and Δd_g the disparity interval of the ground. Let C be a polygon initialised near the building of interest, and S_C the ordered set of vertices that compound C. $n_{x,y}$ is the normal of C at point (x,y).

Let C_{ext} be the edge of the area compounded of the building roof, the visible walls on the image and the occluded part on the other image.
T_C is the area inside C, O_C the area between C and C_{ext} and E_C the area outside C_{ext}.
Figure 1 summarises these notations

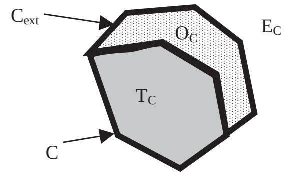

Figure 1. Notations

We are looking for a polygon C_{min} which minimises F :

$$
\begin{aligned}
F = & \sum_{(x,y)\in T_c} \min_{d\in\nabla d_t} \left(Crit\left(I_l(x,y), I_r(x+d,y)\right)\right) \quad \textbf{(term 1)} \\
& + \sum_{(x,y)\in E_c} \min_{d\in\nabla d_g} \left(Crit\left(I_l(x,y), I_r(x+d,y)\right)\right) \quad \textbf{(term 2)} \\
& + \sum_{(x,y)\in O_c} No\_matching\_cost \quad \textbf{(term 3)} \\
& - \lambda \sum_{(x,y)\in C} \vec{grad}(I_l(x,y)).\vec{n}_{x,y} \quad \textbf{(term 4)} \\
& - \lambda \sum_{(x,y)\in C} \vec{grad}(I_r(x+d_{x,y},y)).\vec{n}_{x,y} \quad \textbf{(term 5)} \\
& + \beta \sum_{s_i\in S_C} R(s_{i-1},s_i,s_{i+1}) \quad \textbf{(term 6)}
\end{aligned}
\tag{1}
$$

where:
- Crit is a correlation cost; We call “correlation cost” a correlation measure (sum of square differences, cross-correlation coefficient, …) scaled in such a way that the best match is produced by a 0 correlation coefficient.
- “No_matching_cost” is a threshold that will be defined in paragraph 3.1.2.
- lambda and beta are two regularising parameters called respectively edge parameter and rigidity parameter.
- R is a regularising function of the angle of the C polygon. This function favours 180° and 90° angles as shown in paragraph 3.1.4.

In other words, one seeks a regular polygon, close to the image edges, which minimises a correlation cost on the building roof, minimises a correlation cost on the ground and which defines an occluded area with poor correlation costs.

3.1.1 Disparity interval on the roof and on the ground (term #1 and #2): The correlation cost inside and outside the building must be minimised. Correlation cost are computed on each point: it is the minimal correlation cost in the corresponding disparity interval (Δd_b inside the building, Δd_g outside the building). Δd_b has been chosen during the initialisation stage, Δd_g is the complement of Δd_b over the disparity interval.

3.1.2 Occluded areas (term #3): As soon as the calibration parameters of the stereo pair are known, one can define on one image the area that belongs either to a wall or to an occluded part using the roof edge and the disparity values inside and outside the building. Correlations in this area do not have any sense: pixels in occluded area have no correspondent in the other image and pixels on walls do not satisfy the correlation hypothesis –translation between the two correlation windows-. It is therefore necessary to define a threshold on this area that corresponds to a no-matching cost. Several methods have been tried in order to define this cost (no-matching cost equals to the mean cost of the other parts of the image, no-matching cost defined from the histogram of the correlation coefficients). We have chosen to use the second method. The no-matching cost is defined as a threshold computed on the cumulative distribution function of the correlation coefficients. The operator roughly estimates the pourcentage of pixels being occluded on the image, the no-matching cost is considered to be the correlation cost associated with this pourcentage on the cumulative distribution of the correlation coefficient.

3.1.3 Intensity images (terms #4-5): The correlation criterion defined as the first three terms of equation (1) are not sufficient to delineate building roof adequately. This criterion is based on the correlation on small windows. On roof edges, this correlation is not always reliable. Consequently, we favour polygons that coincide to high gradients on the two intensity images. One difficulty is to determine the translation between the edges of the two images when roof edges are not located at the same altitude. In this case the disparity $d_{x,y}$ of term 5 in eq (1) is not constant over the building edges. Since the disparity map is nor reliable on building edges, we have chosen to estimate $d_{x,y}$ independantly on each pixel as the disparity of the nearest pixel lying on the initial contour.
Notice that we use the amplitude of the projection of the gradient normal to the polygon in order to favour edges of the same direction as the polygon. Small very contrasted edges like chimneys are eliminated this way.

3.1.4 Regularising term (term #6): The regularising term is applied to introduce a priori on the building shape. We have chosen to model buildings using polygons and more specifically we want to favour 90° angle buildings. The regularising term:

– Is computed on the vertex angles of the polygon.
– It favours 180 and 90 degree angles against other angles
– It penalises 0° angle
– It must be easy to compute.

The chosen function can be written analytically as:

$$\begin{aligned} R(\theta) &= 2-|\sin 2\theta| \quad \textbf{when} \quad \cos\theta \geq |\sin\theta| \\ R(\theta) &= |\sin 2\theta| \quad \textbf{when} \quad \cos\theta \leq |\sin\theta| \end{aligned} \tag{2}$$

This function is plotted on Figure 2.

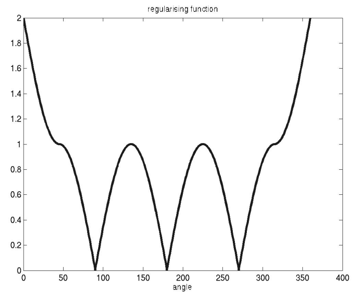

Figure 2. Regularising term for angles varying between 0 and 360°.

90 ,180 and 270° angles are not penalised. The more penalised angle is the 0° angle. The curve is concave around 90, 180 and 270°angles which favours few sharp angles against several angles almost flat. This function may be written using only sine and cosine which are rapid to compute if one knows the position of the vertex and its neighbours.
One can fear that the minimisation of such a non convex function might be difficult but the other terms are also non convex and the minimisation can only be performed on a approximated way. Using a non convex regularising function does not increase the algorithm complexity.

3.2 Optimisation algorithm

The implementation is largely inspired from our anterior work on rectangular roof extraction (Michel, 1998) and on C. Chesnaud (Chesnaud, 2000) studies on statistical snakes. We summarise here the approach and indicate only changes between our implementation and C. Chesnaud's implementation.
As usually in image processing, such an optimisation is difficult: this criterion is non convex and cannot be differentiated.
We have chosen a deterministic algorithm in order to simplify the stop criterion. The optimisation is an iterative sequential process including insertion/updating/deletion of polygon vertices according to the scheme displayed on Figure 3.

– 1st step: The initial building edge is polygonised with a 16 pixel interval between two consecutive vertices.

– 2nd step : each vertex is displaced until stabilisation of the criterion: In order to do so, we compute *F* (see eq 1) for different positions of each vertex and we accept the displacement that lowers *F*. This step is performed successively on all vertices until stabilisation.
– 3rd step: a deletion test is performed on each vertex. In order to do so, we compute *F* by deleting one vertex and we accept the solution if *F* is lowered. This step is performed successively on all vertices until stabilisation.

The 2nd and 3rd step are repeated until stabilisation.

– 4th step : vertices are added to the polygon so that there is a 8 pixel interval between consecutive pixels, both step #2 and #3 are repeated until stabilisation.

If we compare this algorithm with the ones presented in [8], we have added a deletion step. This step is important for two reasons. First, we want to have a description of the building using few vertices. Secondly, this deleting step helps us in eliminating algorithm problems due to small segments. This algorithm is not optimal, it always converges since it modifies the contour polygon only when *F* lowers but we have no guaranty that the solution is optimal.

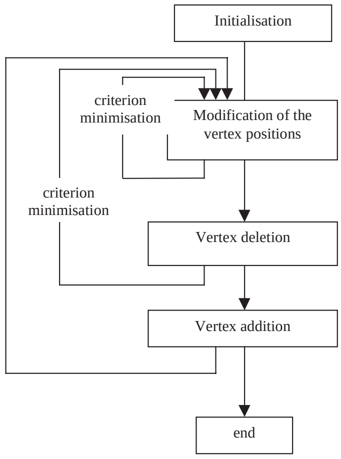

Figure 3. Process flow chart

A pre-integration of the correlation coefficient is used to transform the surface integral of terms 1,2 and 3 into a contour integral (Chesnaud, 2000). This trick considerably speeds up this algorithm.

4. RESULTS

4.1 Pre processing:

Amiens images have been provided by IGN (Figure 4). They have been taken with a digital camera, have a 25 cm resolution and are precisely calibrated.

Figure 4. Amiens stereoscopic images (amiens4_322.tif@IGN01, amiens4_323.tif@IGN01)

The first step consists in computing a disparity map from the two images. Because of the high quality of such images, a simple processing is efficient. Figure 4 shows the disparity map obtained with a dynamic programming algorithm and a fixed occlusion correlation cost. This cost is defined according to the cumulative distribution function of the correlation coefficient.

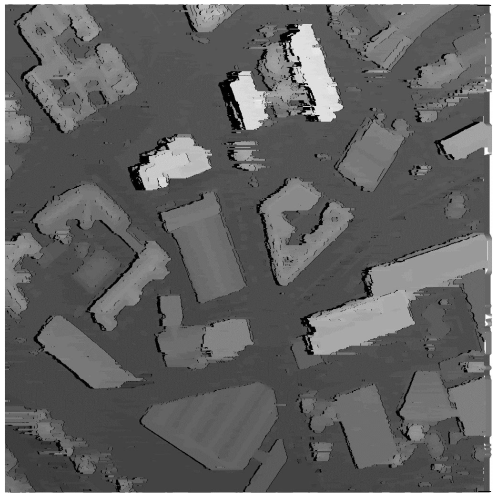

Figure 5. Disparity map

4.2 Result analysis

All buildings have been extracted using the same parameters (90% of non occluded pixels, rigidity parameter : 15, edge parameter :5). Generally, we have good results, buildings are extracted precisely using simple polygons.

Let us first examine buildings 1 to 4. Even if these buildings are quite different, these extractions fit the image. Notice that edges of buildings 1 and 2 are not contrasted at all on the image. The algorithm succeeds in extracting them using mainly the correlation information. Building 3 is homogenous and correlation is not discriminating but edges are precise enough to delineate the building. Building #4 is well delineated even if it is compounded of different textural zones which would have made a feature based algorithm ineffective.

The edge found on building #5 is not completely satisfying: a corner has not been positioned adequately. In the particular case, the correlation coefficient does not help to choose between ground and building; the photometric information is very local and the algorithm has converged to a local minimum. This minimum could have been avoided with a more sophisticated algorithm that would have modified two vertices at the same time.

Buildings #7,8, and 10 show the main drawback of this method: buildings are close to trees of approximately the same height and the algorithm can not distinguish buildings from trees, trees are therefore considered as part of the buildings. The only way to improve this situation would be to segment the image into natural or artificial above-ground areas using both the intensity and disparity images.

Extaction of buildings # 6 8 9,11,12 and 13 show that this algorithm may be used on sloped roofs.

The algorithm described in this paper cannot deal with inner courtyards that is the reason why building #14 and 15 are not very well delineated. However, notice the differences between the initialised contours and the final contours showing that this algorithm can cope with solutions far away from the initialisation.

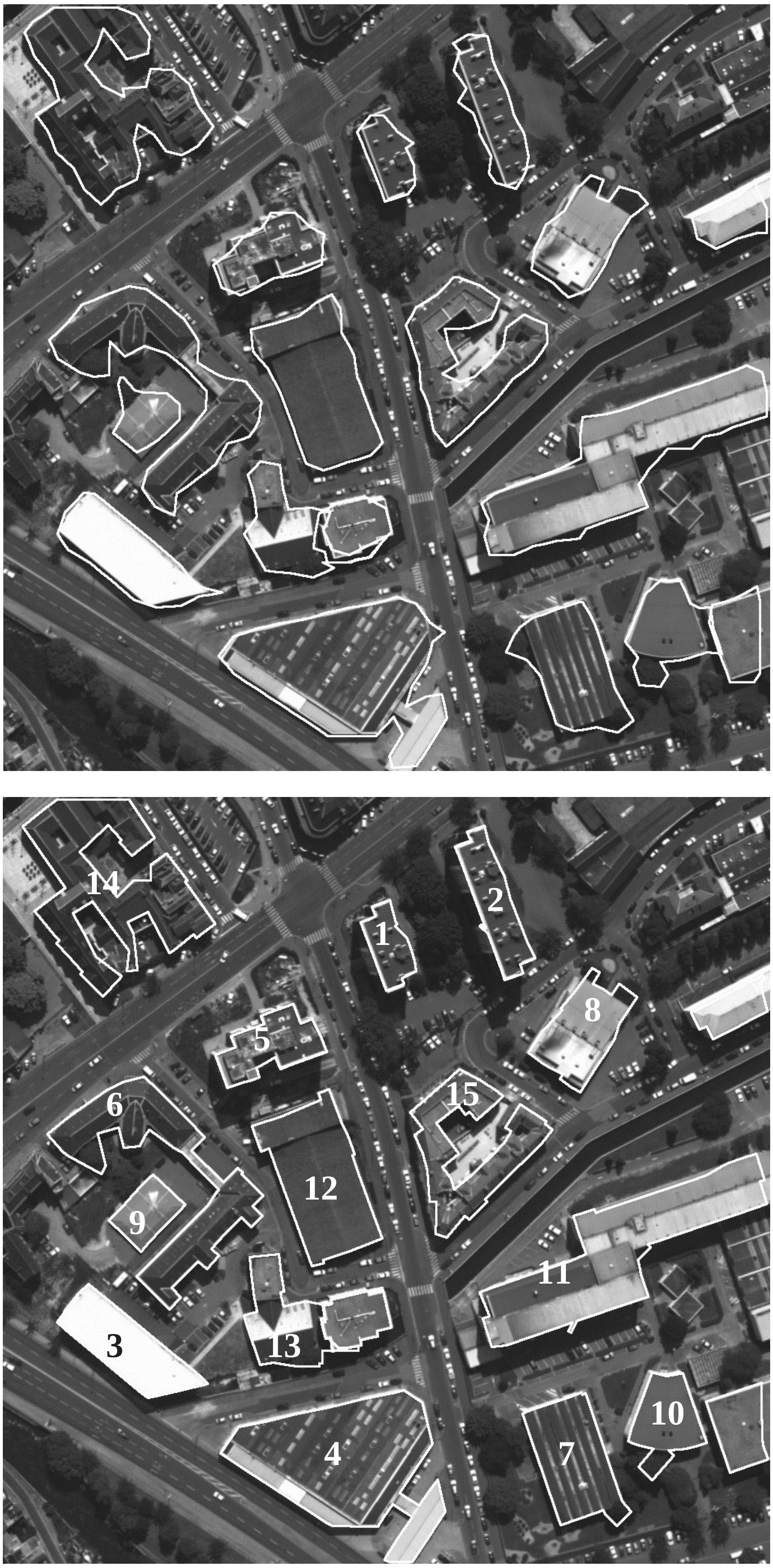

Figure 6. Initialisation and extracted buildings

5. CONCLUSION

In this paper, we have presented a new method to precisely delineate buildings. This method gives good results on several kinds of buildings and is easy to parametrize. The potentiality of this method has not been completely exploited and the different terms of the minimisation criterion could be refined.

The initialisation step should be automated in order to have an automatic building delineation algorithm. This could be done using an automatic segmentation of ground/above-ground algorithm as the one described in (Baillard, 1998). A natural/artificial above ground segmentation should also be performed to avoid grouping buildings and trees during extraction.

Notice, that this technique only delineates building edges. One has to estimate the roof shape to have a complete description of the buildings. The algorithm presented in this paper should make this step easier since it produces a complete set of reliable segments delineating buildings. It could be used as an alternative to 3D segments, extendingin this way the application domain since our method requires only need a stereoscopic pair of images.

BIBLIOGRAPHY

Baillard C., 1997. Analyse d'images aériennes stéréoscopiques pour la restitution 3D des milieux urbains, *Thèse de doctorat de l'ENST*, France.

Baillard C., Zisserman A., 2000. A plane-sweep strategy for the 3D reconstruction of buildings from multiple images. *IASPRS*, Vol XXXIII, Part B2, pp 56-62, Amsterdam.

Brenner C., 2000. Towards fully automatic generation of city models, *IASPRS*, Vol XXXIII, Part B3/I, pp. 85-92, Amsterdam.

Chesnaud C., Réfrégier P., Boulet V., Statistical Region Snake-Based Segmentation Adapted to Different Physical Noise Models, In: *Pattern Analysis and Machine Intelligence*, Vol 21, n° 11, Nov 1999, pp 1145-1157.

Chesnaud C., 2000. Techniques statistiques de segmentation par contour actif et mise en oeuvre rapide, *Thèse de doctorat de Université de droit, d'économie et des sciences d'Aix-Marseille*, France.

Fisher A., Kolbe T. H., Lang F., 1997. Integration of 2D and 3D Reasoning for Building Reconstruction Using a Generic Hierarchical Model, In: *Proceedings of the Workshop on Semantic Modeling for the Acquisition of Topographic Information from Images and Maps SMATI'97*, pp. 159-180.

Forstner W, 1999, 3D-City Models: Automatic and Semiautomatic Acquisition Methods, *Photogrammetric Week*'99, Fritsh, D. and Spiller, R. (eds.), Wichmann, Karlsruhe.

Haala N., Brenner C., 1999. Extraction of buildings and trees in urban environments, *ISPRS Journal of Photogrammetry and Remote Sensing* 54, pp. 130-137.

Michel A., Oriot H., Goretta O., 1998, Extraction of rectangular roofs on stereoscopic images – an interactive approach, *IASPRS*, Vol XXXII, Part III/I, Columbus, pp. 360-366.

Oriot H., Michel A.,2002, Reconstruction 3D de sites urbains par stéréoscopie optique haute résolution, *Bulletin SFPT* n°166(2002-3) pp. 19-26.

Ortner M, Descombes X, Zerubia J, 2002. Building Extraction from Digital Elevation Model, *Rapport INRIA N°4517*, INRIA, France.

Paparoditis N., Cord M., Jordan M., Cocquerez J.-P., 1998, Building Detection and Reconstruction from Mid- and High-Resolution Aerial Imagery, *Computer Vision and Image Understanding*, vol 72, n° 2, pp. 122-142.

Scholze S., Moons T., Van Gool L.,.2002 A probabilistic approach to roof extraction and reconstruction, Photogrammetric Computer Vision, Graz, Austria, , In: *IASPRS*, Commission III, pp B 231- 237.

Weidner U., Forstner W., 1995. Towards Automatic Building Extraction from High Resolution Digital Elevation Models, *ISPRS Journal*, 50(4), pp.38-49.

ACKNOWLEDGMENTS

We want to thank IGN for the stereoscopic pair of images on the city of Amiens.