ANALYSIS OF 3D RECONSTRUCTION ERROR IN THE CONTEXT OF COMPUTATIONAL STEREO IN REMOTE SENSING

C. Beumier and M. Idrissa

Signal & Image Centre, Royal Military Academy, Avenue de la Renaissance 30, 1000 Brussels, Belgium – (beumier, idrissa)@elec.rma.ac.be

KEY WORDS: Computational stereo, Disparity, Digital Surface Model, Photogrammetry, 3D reconstruction error, Database, Orthorectification

ABSTRACT:

The availability of stereo couples of satellite or aerial images allows for the extraction of the Digital Surface Model which is of high importance for earth observation. In the specific case of change detection of man made structures, the 3D information is a vital cue to distinguish buildings from flat objects (e.g. parking lots). Height variations are obtained from a disparity map, qualifying the magnitude of the parallax and henceforth the relative elevation. If available, the camera and flight parameters enable other types of applications thanks to the derivation of geographical 3-D coordinates. In particular, the National Geographic Institute of Belgium is interested in automatic estimation of 3D coordinates to support the nowadays time consuming manual procedure. In a first collaboration with NGI, we developed two disparity estimation procedures to detect buildings and look for related changes in the topographic database. Then flight and camera data were used to derive geographical 3D coordinates.
This paper first recalls how the different steps necessary to produce a DSM from stereo images were implemented, addressing epipolar rectification, disparity estimation and 3D reconstruction. The major contribution of the paper is to present a quantitative assessment of the quality of automatic 3D measures by comparison with available ground truth values. A separation between errors from artefacts in disparity estimation and precision limits in the reconstruction process has been considered.

RESUME:

La disponibilité de couples stéréo d'images aériennes ou satellites permet l'extraction de Modèles Numériques de Surface cruciaux dans l'observation de la terre. L'information 3D est en effet pertinente pour la détection du changement affectant le bâti. Les variations en hauteur s'observent sur la carte de disparité représentant l'amplitude de la parallaxe entre les deux images du couple stéréo. Si les données caméra et de prises de vue sont connues ou estimées, les coordonnées 3D peuvent être déduites de la carte de disparité, ouvrant la voie à de nombreuses applications. En particulier, L'Institut Géographique National de Belgique est intéressé par l'estimation automatique de données 3D pour assister la procédure manuelle de restitution. Dans une première collaboration avec l'IGN, nous avons développé deux méthodes d'estimation de la disparité pour détecter les bâtiments et chercher les changements relatifs à la base de données topographiques. Ensuite, les paramètres de vol et de caméra furent utilisés pour déduire les coordonnées géographiques 3D.
Cette publication présente d'abord les détails d'implémentation des étapes nécessaires à la production du MNS, soit la rectification épipolaire, l'estimation des disparités et la reconstruction 3D. La contribution majeure de l'article concerne l'évaluation de la qualité des mesures 3D automatiques par comparaison avec les données de référence disponibles. Nous avons considéré la séparation entre les erreurs à caractère aléatoire survenant dans la carte de disparité et l'erreur d'imprécision de la reconstruction.

KURZFASSUNG:

Anhand von Stereopaaren von Satelliten- oder Luftbildern können digitale Oberflächenmodelle erstellt werden, die für die Erdbeobachtung von großer Bedeutung sind. Insbesondere im Rahmen der Erfassung von Veränderungen künstlich geschaffener Strukturen ist die 3D-Information ein wichtiger Hinweis zur Unterscheidung von Gebäuden und flachen Objekten (z.B. Parkplätzen). Höhenunterschiede sind auf der Disparitätskarte erkennbar, die die Amplitude der Parallaxe zwischen den zwei Bildern des Stereopaars und damit die relative Erhebung darstellt. Falls verfügbar, ermöglichen Kamera- und Aufnahmeparameter eine Reihe weiterer Anwendungen, jeweils auf der Grundlage der Ableitung geografischer 3D-Koordinaten. Insbesondere das Belgische Nationale Geografische Institut ist interessiert an der automatischen Schätzung von 3D-Koordinaten, um das zeitaufwändige manuelle Verfahren zu unterstützen. In einer ersten Zusammenarbeit mit dem NGI haben wir zwei Verfahren zur Disparitätsschätzung entwickelt, um Gebäude zu erkennen und entsprechende Veränderungen in der topografischen Datenbank zu suchen. Dann wurden Flug- und Kameradaten verwendet, um geografische 3D-Koordinaten abzuleiten.
Die Veröffentlichung legt zunächst dar, wie die einzelnen Schritte zur Erstellung des digitalen Oberflächenmodells aus Stereobildpaaren umgesetzt wurden, und befasst sich hierbei mit epipolarer Rektifizierung, Disparitätsschätzung und 3D-Rekonstruktion. Der wesentliche Beitrag der Veröffentlichung besteht darin, die Qualität von automatischen 3D-Maßnahmen durch Vergleich mit verfügbaren Bezugsdaten quantitativ zu bewerten. Dabei wurde eine Trennung von zufälligen Fehlern bei der Disparitätsschätzung und Fehlern aufgrund Ungenauigkeiten bei der Rekonstruktion versucht.

1. INTRODUCTION

1.1 Problem definition

Geographical information received an increasing interest as applications like navigation entered public access. Even if 2D flat information was first proposed, current developments consider the 3D aspect as an important asset to increase the reality and correctness of representations.

In our rapidly changing world, the validity of the proposed information can only be maintained at the expense of frequent updates, requiring much human effort in the traditional procedures based on image observation. In order to keep update delays and costs to a reasonable level, geographical database producers are looking for alternative procedures based on software solution for (semi-) automatic change detection.

In a national project called C4-11, funded by the Belgian Ministry of Defence, the Signal & Image Centre of the Royal Military Academy in Brussels is looking for image processing solutions to offer support to the Belgian national mapping agency (NGI). This study, running from 2007 till 2011, contains two parts of great interest for NGI: the detection of changes in a database at 1:10000 from aerial images and the automatic estimation of 3D coordinates from a stereo couple of aerial images. We came to the conclusion that 3D information is particularly appropriate for change detection, so that the project mainly focuses on 3D estimation for elevation measurements and change detection, what a posteriori justifies the a priori meaningless study name C4-11: "see for elev'n" (elevation).

The general subject of the underlying work concerns 3D estimation of building and road elements from a typical stereo pair of images used by NGI (aerial with 0.3m resolution). This covers as many different topics as image pair rectification, disparity estimation and 3D reconstruction whose corresponding modules are summarised in section 2 since they were detailed in other publications. The objective of this paper is to present a quantitative precision analysis of automatic 3D measurements from a NGI image pair compared to ground truth 3D values obtained at NGI by restitution.

1.2 Related Works

Photogrammetry, the technique to estimate the relief of a scene from a pair of images, is nearly as old as photography. Its major field of interest has traditionally been cartography. The technique has been refined thanks to the improvement of photography and thanks to the introduction of computer processing, leading today to fully automatic solution.

In recent years, functions related to photogrammetry have been populating most Geographical Information Systems (GIS), either in commercial or open source solutions.

Epipolar rectification and 3D reconstruction are mature fields of research from which state of the art solutions were used in this work. The interested reader will consult section 2 which redirects to appropriate references detailing those developments.

Software solution for automatic disparity estimation is trickier. Although solutions have been proposed for a long time (survey of Scharstein, 2002), we decided to build our own algorithm to address two important aspects: robustness and computation time. Section 2 contains a summary of internal development for disparity estimation. Details are available from mentioned references.

Our work targets two different applications: change detection from 3D cues, and 3D measurement for 3D database verification or completion. Change detection can accommodate with imprecise 3D values as long as relative heights (or disparities) are coherent. On the contrary, database update needs precise 3D estimation to detect errors or provides for values not yet in the database.

Most works validate 3D estimations by presenting a RMS value of the difference between measures and reference values (Wolff, 2007). We believe that two different types of errors impair 3D measurements: faults due to a failure in disparity estimation and precision errors due to the limits of the mathematical model and image localisation. The formers introduce possibly large errors while the latters normally present a thin Gaussian–like distribution. We propose an error analysis based on histogram to separate the two error types and offer a deeper insight to find the roots and answers to those errors.

2. 3D ESTIMATION FROM STEREO IMAGES

2.1 Introduction

Computational stereo refers to the determination of 3D information from two or several images taken from different points of view. Refer to (Baillard, 1997 and Brown, 2003) for good surveys.

In general, a distinction has to be made between processing aerial and satellite images. Aerial images are usually shot with a digital camera, so that all pixels are acquired at the same time. The projective (pinhole) model is applicable, what generally leads to the chain of processing described in this paper. On the contrary, satellite images are often captured with pushbroom cameras, one line at a time. An adaptation of the presented procedure for 3D reconstruction should thus be made to handle most satellite images.
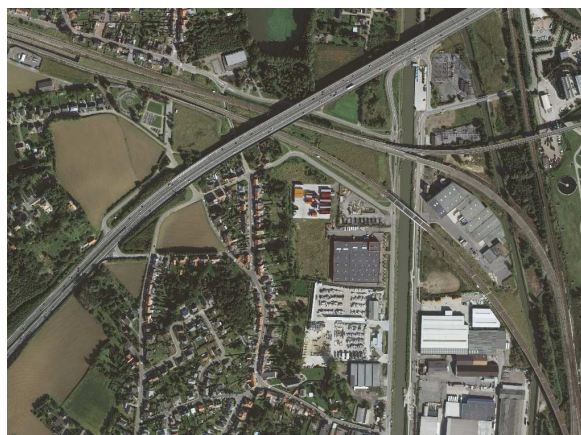

Figure 1: Part of an aerial image (0.3m) acquired for NGI in the vicinity of Leuven (Belgium)

This section describes the processing chain for 3D reconstruction from a stereo couple of aerial images, adopting

the pinhole model. The following three subsections detail epipolar rectification, disparity estimation and 3D reconstruction. The last subsection presents the quantitative analysis of automatically obtained 3D values by computational stereo compared to ground truth.

2.2 Epipolar Rectification

In a remote sensing campaign, the flight and camera information is necessary to geo-reference the image data. If not available, this information can be derived from a set of Ground Controlled Points of known geographical coordinates marked in the images.

In our collaboration with NGI, the camera specifications are available and aero-triangulation is performed by NGI to refine imprecise image references offered by the image provider. However, since the stereo matching algorithms are much simpler when applied to epipolar images (defined below), we developed a module to transform any stereo couple into epipolar images.

When two images of a stereo couple are acquired at differed times as in airborne or satellite acquisition campaigns, object image positions may be displaced in any direction, dependent of relative object motion. Said differently, due to perspective projection of the cameras, to an image position in the left image corresponds an epipolar line in the right image with possible projection points. By changing the geometry, it is possible that the right epipolar lines of any left point are horizontal and vice versa. In this case, the image couple is said to be epipolar and the search to pair each right point with a correspondent left point simplifies to a 1-D search instead of a 2-D search. Fortunately, stereo images captured in a given geometry can be transformed a posteriori to become epipolar.

The approach followed to transform both images into the epipolar geometry consists first in estimating the fundamental matrix F, characterizing the cameras and the relative points of view of the two image captures. The "8-point algorithm" has been applied to estimate F thanks to a collection of points automatically selected in the left and right images by SIFT local feature detection (Lowe, 2004). The transformations (homographies) to be applied on both images to make the stereo couple epipolar are derived from F (Hartley, 2003).

2.3 Disparity Estimation

Two different approaches for disparity estimation have been conceived and implemented for the case of epipolar images, usually adopted for computational stereo thanks to its reduced search complexity.

First, a correlation method based on multiple window sizes has been developed. The compromise between the precision achieved for small correlation windows and the robustness obtained with larger window sizes has been tackled by estimating the disparity for all window sizes in a user-selectable range and delaying the final disparity choice to a relaxation procedure optimizing a smoothness criterion by simulated annealing (Idrissa, 2009). A dense disparity map is produced.

Secondly, in an attempt to reduce computation time, a feature based approach has been developed by considering the horizontal gradient as basic features to find candidate left-right matches. The candidate sets are reduced by considering the

disparity range limits, the intensity level and the gradient orientation similarity. Robustness is achieved by an aggregation phase which elicits the mode of all disparity candidates in a small spatial area. This aggregation phase serves also the purpose of delivering a dense disparity map (Beumier, 2009).
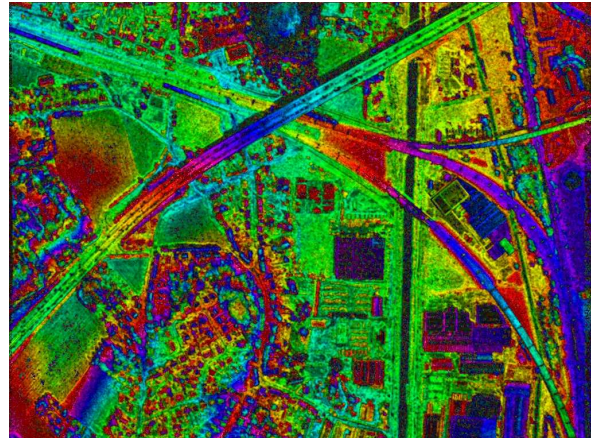

Figure 2: Disparity represented in false colours corresponding to area of Figure 1

2.4 3D Reconstruction

3D reconstruction consists in converting image positions (x,y) and associated disparity (2.3) into geographical X, Y, Z coordinates. Although this transformation is a simple application of well known mathematical formulas, the process requires calibration from camera and flight information.

2.5 Quality Estimation

NGI is one of the Belgian providers of vector databases for geographical information. In their recent version called ITGI (Inventaire Topogéographique - Topogeographische Inventaris), a Z value corresponding to elevation is part of the data associated to each polygon vertex. Z values will serve as references to assess the quality of our automatically extracted Digital Surface Model (DSM).

Differences between automatically obtained 3D coordinates of the DSM and database values may come from different origins. The DSM may suffer from the limits of the mathematical models, inaccuracies in image localisation and wrong disparity estimation from bad matching (due to occlusion, weak texture or confusion). Database coordinates may contain errors due to human distraction and inaccuracies due to generalisation rules followed to simplify data input during human restitution. In the specific case of NGI, the recent ITGI database is in the update process after being initialised from its predecessor Top10v-GIS. In particular, Z values, not part of Top10v-GIS, were initialised thanks to the restitution phase used to get correctly geo-referenced X, Y values for Top10v-GIS.

Our concern is that the 3D quality is generally estimated by the Root Mean Square of the difference between estimated and reference values. For us, this measure has little meaning if large discrepancies due to unrealistic circumstances hide the precision achievable by the DSM extraction approach. We will try to get rid off unrealistic errors due to bad disparity estimation or database errors.

2.6 Results
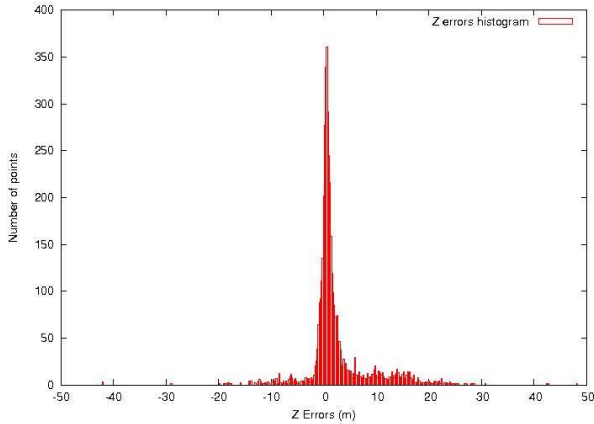
2.6.1 Global analysis for roads

Figure 3: Histogram of Z errors for points of the road network

We first analysed the Z differences from a histogram as shown in Figure 3. To set up this histogram, all the Z values of the points of road segments of the vector database were subtracted from the Z values of the DSM estimated by our algorithm on a aerial image in the vicinity of Leuven (about 4900x13800 pixel). This figure shows the relative importance of the different levels of elevation errors. The central peak mainly consists of errors up to +/- 4m. The tails however contain errors up to +/- 20m. These important differences, mainly due to the presence of trees above the roads and some wrong disparity estimation from image matching.

We introduced the threshold T to distinguish between two types of errors. Errors above T are expected to originate from spurious circumstances like wrong disparity estimation or road occlusion (by trees) generally leading to unexpectedly large Z error values. Errors below T are used to derive a RMS value. This measure mainly includes deviations from the camera or mathematical model and small imprecision in image localisation and disparity estimation. The value T is picked from the histogram, separating the Gaussian-like part due to imprecision and the more uniform distribution parts due to spurious disparity estimation or database error. As listed in Table 1, the RMS value is 1.21 m for T = 4 m.

T (m)	2	3	4	5	100
RMS (m)	0.85	1.05	1.21	1.36	5.8
Num Pts	3115	3361	3473	3547	4381

Table 1: RMS values for different threshold values

Figure 4 shows the localisation of road points for a part of the image considered for quality estimation. Red points relate to road vertices whose errors exceed threshold T. Many of them correspond to road parts hidden by trees. Some correspond to database errors. We noticed the case of a road passing under the motorway, but with the Z level of the motorway. We plan to develop an automatic procedure based on the detection of unlikely local slopes along roads to handle those database errors.
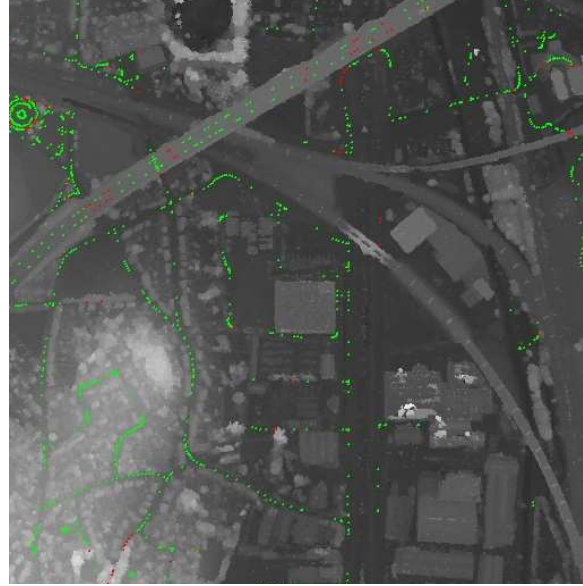

Figure 4: Localisation of database road points. Red points have an error above T

2.6.2 Fine analysis for roads

In order to estimate a reliable level of error in the Z estimation from our stereo computation algorithm, we had to remove the most possible cases of errors explaining the tails in Figure 3. The level of errors affecting real road points is likely to be low since many roads are large and continuous objects (if not hidden), well handled by disparity estimation. Road points of the database correspond to the road axis which has less chance to be influenced by elevated objects at the road sides.

For a finer analysis, we considered the plot of Z profiles along road segments. Figure 5 depicts the case of a segment corresponding to a motorway. Red plus signs are ITGI Z values while green crosses are estimated DSM Z values.
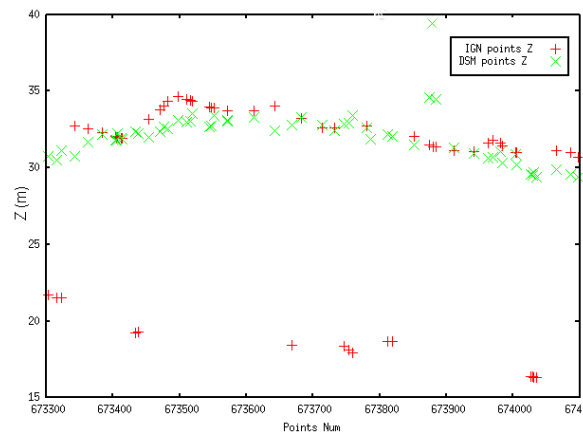

Figure 5: Estimated (green) and reference (red) Z profiles along the motorway

We notice in this specific case a number of database errors due to the fact that the motorway lies about 20m above the terrain level. As said earlier ITGI is not fully updated yet and received some Z values from interpolation, possibly leading to errors

where Z varies much (bridges, overlapping roads). We also see in the figure an abrupt Z variation in the DSM values. This originates from wrong disparity estimation due to a truck. The Z difference is so large (10m) because of the truck motion between the two captures of the stereo pair.

To reject these large errors, we applied the threshold T (4m) defined in the previous subsection and obtained RMS values in the range of 0.5 and 1.5 m for several road segments chosen for their minimal number of points. Of course, we did not consider roads occluded by trees. These RMS levels appear to us as a superior limit because the quality of Z from the database is not perfect.

The approach of error rejection from threshold T gives a way to support NGI in its updating process of ITGI, localising points or segments where restitution should be applied to get better Z values. Our algorithm for DSM extraction could be used for this as well.

2.6.3 Global analysis for buildings

The quality estimation of Z measured for buildings is less direct as building database vectors do not superpose perfectly to their image, even after ortho-rectification, due to elevation error and delineation rules. As the 3D values of the building database concern polygon vertices, a small displacement has a large chance to lead to a point out of the building roof in the DSM. Much more, in Belgium, a majority of roofs are gable, although the database only encodes Z at the gutter level, and generalises the elevation for a block of houses of different Z circled by a single polygon
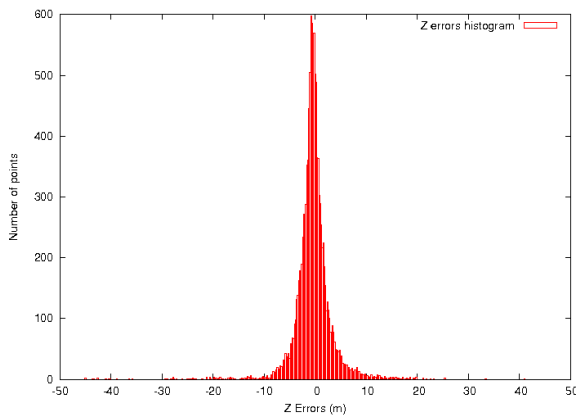

Figure 6: Histogram of Z errors for points of the building layer

Figure 6 depicts the distribution of Z differences between the Digital Surface Model and the Z values of building vertices contained in the database. To account for the X, Y localisation imprecision, the DSM was dilated, meaning that higher local Z values were propagated to their close neighbourhood (up to 1 pixel).

As the database was not updated thanks to the processed image, consistency errors may be present in the results, but they are statistically limited as the database version and the image are both recent.

In comparison with roads, the database contains much more building vertices, explaining the larger histogram values. The distribution of the central peak is larger due to the more important variability of elevation around buildings, should it be due to the relief reality, the infidelity of disparity estimation or inconsistencies of the database due to approximation rules concerning building restitution.

The RMS value of the central peak (T = 10m) for buildings is 3.4 m (4 % building points rejected out of 12000).

3. CONCLUSIONS

This paper presented the analysis of errors of 3D values extracted from a pair of stereo aerial images (0.3m). The approach for automatic DSM extraction has been first described. Then, automatically extracted X, Y, Z values and coordinates contained in the NGI vector database ITGI were compared for road and building points.

The presented work mainly consists in the determination of elevation (Z) errors. For the horizontal direction (X, Y: planimetry), no error analysis was performed other than through observation. These coordinates were sufficiently precise as roads have limited slopes and as building vertices were compared after DSM dilation, making the planar localisation less sensitive. Should the precision in X and Y be improved, a few reference points of known coordinates would suffice to perform a finer registration.

The analysis of 3D values concentrated on elevation Z as one of the two goals of our study is to analyse if automatically extracted DSM can replace or complement the current restitution work for the ITGI database. For points situated on the road network, the comparison of database elevation Z with corresponding DSM values led to a RMS value of 1.2 m, where differences larger than a threshold T (defined from the Z difference histogram) were discarded as a way to reject unexpected values due to wrong disparity estimation or database errors. The error level for building points is higher.

The present quality analysis demonstrated the general correctness of all the steps involved in the automatic estimation of the DSM (epipolar rectification, disparity estimation and 3D reconstruction). The global analysis through histogram of differences enabled the rejection of most spurious errors due to either occlusion or database errors. The finer analysis indicated that DSM values should have a sufficient quality for ITGI, considering that the residual errors seem to be partly inherent to the current database values. Trials with Z values of higher quality will be needed to confirm that point.

As a consequence of the finer analysis performed on road segments, we can propose an automatic solution for the detection of database inconsistencies like the local Z discontinuities due to road overlapping or bridges.

In the future, we plan to refine the disparity estimation approaches in order to reduce the global error and inconsistencies discovered in these experiments. We also intend to create a graphical interface that will integrate the implementation of the different software pieces into a user-friendly environment.

REFERENCES

Wolff K., Gruen A., DSM Generation from early ALOS/PRISM Data using SAT-PP, *ISPRS Hannover Workshop 2007: High-Resolution Earth Imaging for Geospatial Information*.

Scharstein, D., Szeliski, R., A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms, *International Journal of Computer Vision*, Vol. 47, Apr-Jun 2002, pp. 7-42.

Baillard, C., 1997. Analyse d'images aériennes stéréoscopiques pour la restitution 3D des milieux urbains. Détection et caractérisation du sursol (In french), Thèse Signal et Images, TSI, Télécom Paris [ENST].

Brown, M., Burschka, D. and Hager, G., Advances in Computational Stereo, *IEEE PAMI*, Vol. 25, NO 8, Aug 2003, pp. 993-1008.

Lowe, D. G., 2004, Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, 60, 2, pp. 91-110.

Hartley, R. and Zisserman, A., 2003, *Multiple View Geometry in computer vision*, Cambridge University Press.

Idrissa, M, Lacroix V, 2009, A Multiresolution-MRF Approach for Stereo Dense Disparity Estimation. In *IEEE-GRSS/ISPRS Joint Urban Remote Sensing Event*, May 20-22 2009, Shanghai China.

Beumier, C., 2009, Fast Dense Disparity Estimation of Stereo Couples from Image Gradient, *The 5th Int. Conf. on Signal-Image & Internet-based systems (SITIS09),* Marrakech, Maroc.