SURFACE RECONSTRUCTION IN URBAN AREAS FROM MULTIPLE VIEWS OF AERIAL DIGITAL FRAME CAMERAS

Nicolas PAPARODITIS[*,**], **Christian Thom**[**], **Hassan Jibrini**[**]

[*] ESGT/CNAM - 1, boulevard Pythagore - 72000 Le Mans
nicolas.paparoditis@ign.fr
[**] MATIS/IGN - 2-4, avenue Pasteur - 94165 Saint-Mandé
Christian.Thom@ign.fr
Hassan.Jibrini@ign.fr

Working Group III/2

KEY WORDS : aerial images, aerial digital frame camera, multiple views, DSM, surface reconstruction, object space, multi-image matching, multi-dimensional correlation, self reliability indicator, hierarchical matching.

ABSTRACT

This paper describes an original approach and strategy for the automatic surface reconstruction of urban areas from digital aerial multiple views (across track as well as along track). The general reconstruction approach is based on a concept of multi-image matching guided from object space which provides the DSM and the corresponding orthoimage at the same time. The general idea is to match all the images at the same time with a multi-image correlation function instead of merging all elementary DSM computed on all stereopairs. This multi-image correlation function has very interesting properties, we can use very small templates (windows) in the image matching process, the correlation scores give a real reliability indicator of the matches and the false matches can be filtered with a deterministic threshold. All discarded areas are filled with a hierarchical matching strategy aiming at building the surface step by step.

1. INTRODUCTION

The reconstruction of Digital Surface Models from aerial stereopairs of urban landscapes has been a subject of constant attention from both photogrammetric and computer vision communities [Baillard & al 96][Cord & al 98]. DSM in urban areas are useful as such for a great number of applications: inter-visibility calculations for the optimisation of the location of telecommunication antennas, risk mapping (floods, landslides, avalanches, etc.), change detection for data base updating, mission planning, virtual reality, etc. Moreover, they are essential for the elaboration of other cartographic products such as orthoimages, 3D building models [Paparoditis & al 98][Fuchs & al 99][Jibrini & al 00], etc.

For the latter application, i.e. building reconstruction, which is a priority of many mapping laboratories, images of very high resolution (<20 cm) are required in order to bring out conspicuously the whole set of 3D structures which are necessary for the elaboration of a fine description of a building. At these levels of resolutions and in a context of dense urban tissue, there are a considerable number of hidden parts to the extent that a dense and thorough description of the scene by a stereo analysis becomes impossible. Besides in urban areas, the classical stereo processing admits some flaws because of all the matching ambiguities encountered due to: homogeneous surfaces, discontinuities, occlusions, mobile vehicles, non lambertian surfaces, and the poor image quality due to the image scanning process.

The acquisition of multiple views (a particular spot of the landscape is seen in more than two images) solves many problems. Actually, provided a sufficient overlapping exists, there will always be one or more couples among the whole set of couples in which a given point of the landscape is visible and where some of the problems described previously, i.e. mobile vehicles and non lambertian surfaces, will not appear. Ways of exploiting this data has been largely treated in recent literature [Okotumi & al 91][Gabet & al 94][Canu & al 95][Leloglu & al 98]. In most of the techniques, the general idea is to reconstruct the DSM by an *a posteriori* merging of all the elementary DSM calculated on all the possible stereopairs by a voting technique (e.g. based on the median value). This merging allows a densification and reliabilisation of the results to a large extent when the elementary DSM are complementary and when the results are valid in most of all the DSM.

In the contrary, the stereo matching process requires the use of non-negligible window sizes if one is looking for sufficiently reliable measures. One knows well that the morphological quality of the DSM, i.e. the ability to render discontinuities, slopes, slope breaks, surface microstructures, all of utmost importance for numerous applications, is directly dependent on the window size. The larger the windows and the more the “high frequencies” of the surface are smoothed out, the more the depth discontinuities are delocalised, and the more the matching of steep slopes is difficult (due to image deformations).

One way to reduce the window size is the signal to noise ratio of the images so as to reduce matching ambiguities. This can be obtained naturally by having an imaging system of higher quality, or artificially by increasing the number of observations. The originality of the work presented in this paper lies in the use of high quality of images acquired with IGN’s 12 bits panchromatic digital frame camera (SNR = 300) [Thom & al 99] and of highly overlapping images to improve the surface reconstruction problem with simple algorithms but really adapted to the nature and to the complementarity of the data.

2. OUR APPROACH: A MULTI-IMAGE MATCHING GUIDED FROM OBJECT SPACE

A new matching process concept adapted to these data has therefore been developed. This process allows the matching of the images and the construction of the raster DSM, and the orthoimage all at the same time without having to go through the processing (epipolar resampling, image matching, 3D reconstruction, and 3D resampling) of all stereopairs and the merging of all elementary DSMs. This process is based on the concept of direct multi-image matching guided from object space (Figure 1).

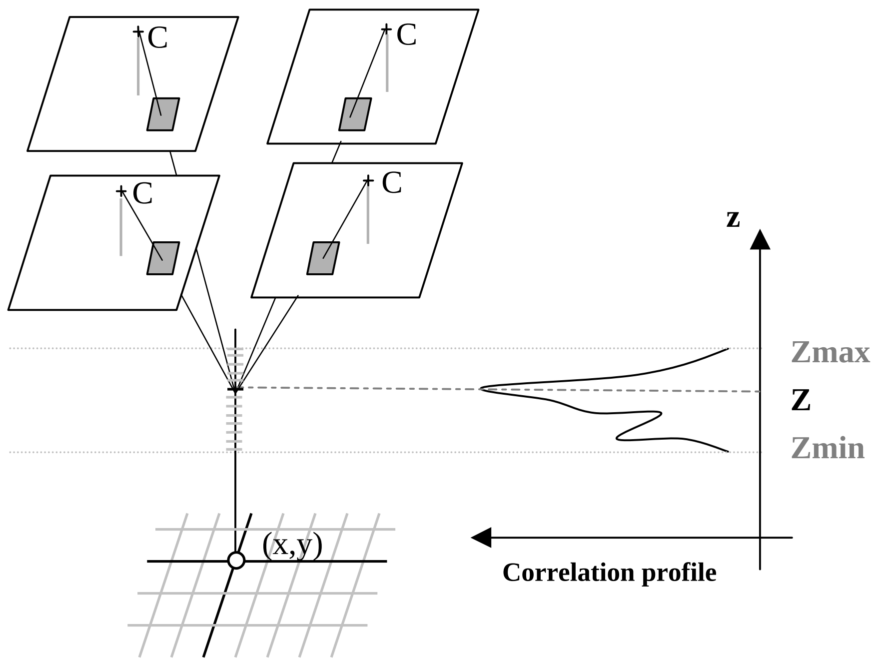

Figure 1: Principle of our multi-image matching system guided from object space

The process is based on the following algorithm. For every single node (x,y) of the DSM raster grid, a correlation profile is constructed gathering all the correlation scores calculated for each of the possible z through a plausible interval [Z_{min},Z_{max}] depending on an a priori knowledge of the scene (given by a map, a gross DTM, etc.). For each given z, one calculates in the whole set of images, the hypothetical corresponding (i_k,j_k) image co-ordinates in each image space k (where $0 \leq k \leq n$ and n is the number of images where this (x,y,z) point is seen). The likelihood of these hypotheses is given by a direct measurement of the similarity of the set of windows centred on the (i_k,j_k). The “estimated” Z value retained for a given (x,y) is the one for which the optimal value of the correlation profile is reached.

Besides, for this (x,y) and for this estimated Z we can calculate directly the corresponding grey level in the orthoimage from the set of associated (i_k,j_k) grey-levels by taking for instance the average or the median of the values.

In practice, one does not treat sequentially all the nodes of the grid. Indeed, the (i_k,j_k) have sub-pixellar positions. The corresponding windows are therefore obtained by resampling locally the images. In order to avoid the redundancy of resampling operations for neighbouring grid points in the DSM, one resamples for a given z value the whole set of images and calculates, for all the (x,y) and for this given z value, all the corresponding correlation scores (indeed if this z is the good one, the content of the resampled images must be locally alike and the correlation score optimal). Then one reiterates this process for the whole set of z values while keeping for each (x,y) the elevation associated to the best correlation score.

3. A MULTI-IMAGE SIMILARITY FUNCTION

How can one measure the similarity of a set of image windows? The direct extension of the two window cross-correlation function is not possible because this provides the scalar product and therefore an information on the angle formed by both texture vectors whose components are the spatially ordered set of grey-levels inside the window. Therefore we have defined a new adapted Multi Image Correlation (MIC) function:

$$0 \leq \mathrm{MIC}(i_1, j_1, i_2, j_2 \ldots, i_n, j_n) = \frac{Var\left(\sum_{k=1}^{n} \overrightarrow{V_k}(i_k, j_k)\right)}{\sum_{k=1}^{n} Var(\overrightarrow{V_k}(i_k, j_k))} \leq n \qquad (1)$$

where:

$\overrightarrow{V_k}(i_k, j_k)$ is the texture vector associated to the window centred on the pixel (i_k, j_k) in image k.

and $Var(\overrightarrow{V_k}(i_k, j_k))$ the variance of the texture vector $\overrightarrow{V_k}(i_k, j_k)$, i.e. the grey levels inside the window centred on the pixel (i_k, j_k) in image k.

Why this correlation function? If the image texture windows are alike, i.e. the texture vectors are collinear, the similarity score is maximal. One can notice that if n=2 this function is very close to the classical centred cross-correlation function.

4. RADIOMETRIC WEIGHTING

In view of the great radiometric stability of the images of the digital aerial frame camera (which is not the case of scanned images) and under the assumption that most of the objects which describe the landscape have lambertian characteristics, we impose an additional constraint on the absolute radiometry of homologous neighbourhoods, in the shape of a weighting function. This weighting is necessary. Indeed, the MIC similarity function is “centred average”. This function measures the similarity of the textures of all neighbourhoods but not their radiometric similarity; this can give rise to mismatches. Our new similarity function called MICRA (Multi Image Correlation with Radiometric Attenuation) can be expressed in the following way:

$$\mathrm{MICRA}(i_1, j_1, \ldots, i_n, j_n) = \mathrm{MIC}(i_1, j_1 \ldots, i_n, j_n) \, . \exp - \left(\frac{\mathrm{Var}_k(I_k(i_k, j_k))}{k} \right). \qquad (2)$$

where :

Var is the variance,
I(i,j) the grey level of pixel (i,j),
and k a normalising coefficient.

The application of this weighting function permits to obtain a similarity criterion which characterises at the same time the texture and the radiometric resemblance of the neighbourhoods. Therefore the MICRA similarity function is more robust and discriminating (Figure 2).

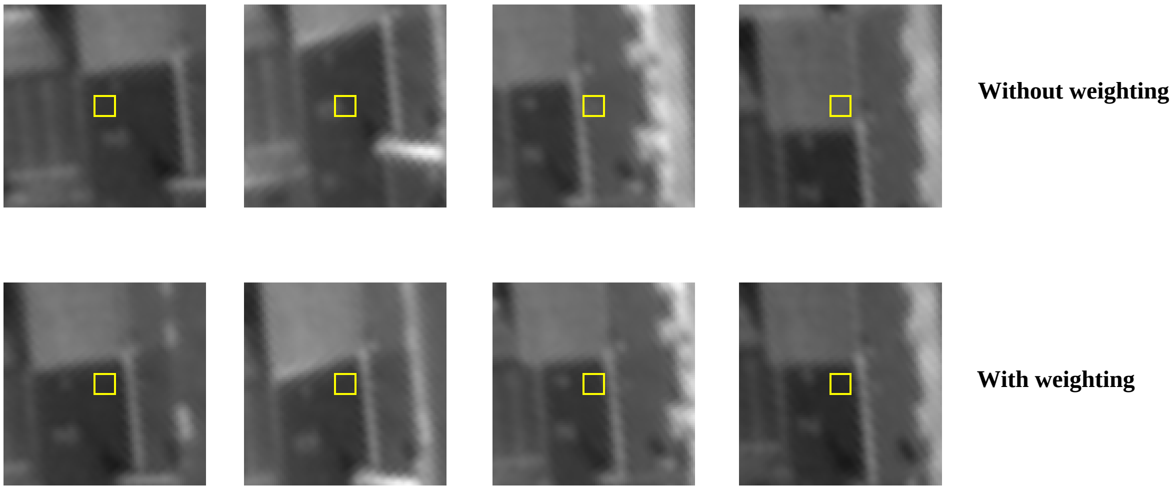

Figure 2: Comparison of homologous windows (3x3) found for a given (x,y) point without and with the radiometric weighting function.

5. RESULTS AND COMMENTS

The increase in the number of observations, allows, on the one hand to reduce, sizeably, matching ambiguities met with in the classical stereo-processing, and thus to increase the reliability of the process. On the other hand, it allows, also thanks to the high image quality of our digital frame camera, to use sizes for the correlation windows of 3x3 (Figure 3). The results show clearly a dense DSM, a good rendering of the relief microstructures, and a good localisation of the discontinuities (especially the building edges). However the very small size of the windows is at the origin of false matches in the very homogeneous areas which do not appear with a 5x5 window (Figure 4). On the contrary, a 5x5 window has a less accurate rendering of microstructures and discontinuities.

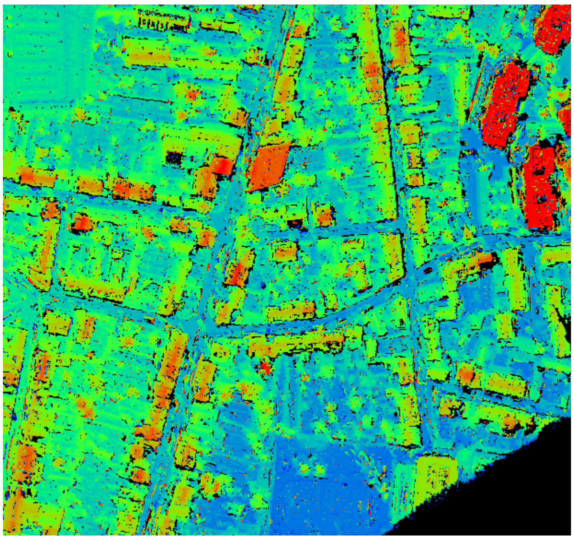

Figure 3 : DSM generated with 4 images and 3x3 windows

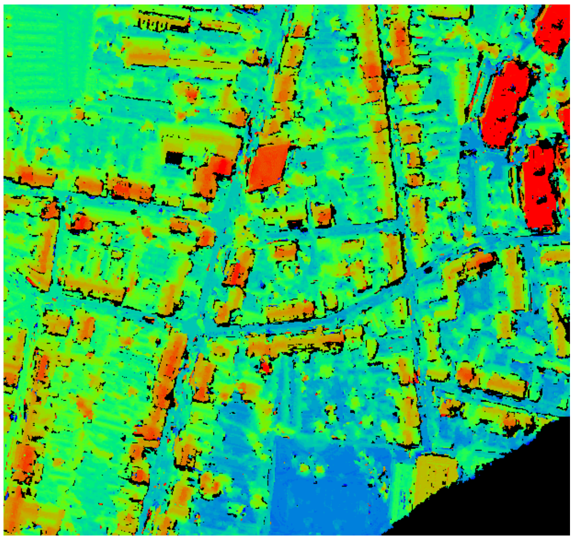

Figure 4: DSM generated with 4 images and 5x5 windows

Figure 5: Orthoimage corresponding to the DSM of Figure 3
40 cm ground pixel size digital survey on the french town of Le Mans with a 50% across and along track overlap

Working from object space offers a good number of advantages. One can work in a transparent way on images of different resolutions, and the parameters are expressed in a metric form. Also we avoid the blind resampling of all the 3D samples corresponding to all image matches in order to generate a regular grid DSM in object space. Indeed, these samples when the matching process is guided from image space follow, although their distribution is regular in image space, an irregular spatial distribution in object space. Finally, its major advantage is that it allows, on the one hand to treat N images in a natural and transparent way, and on the other hand, to construct at the same time the DSM and the corresponding orthoimage (Figure 5).

6. MATCHING SELF EVALUATION AND AUTOMATIC FILTERING OF FALSE MATCHES

As expected, the multi-image correlation score does not supply a probability, but a good self-indicator of the measurement reliability. The results (in Figure 6 and Figure 7) show that DSM aberrant points have much weaker correlation scores.

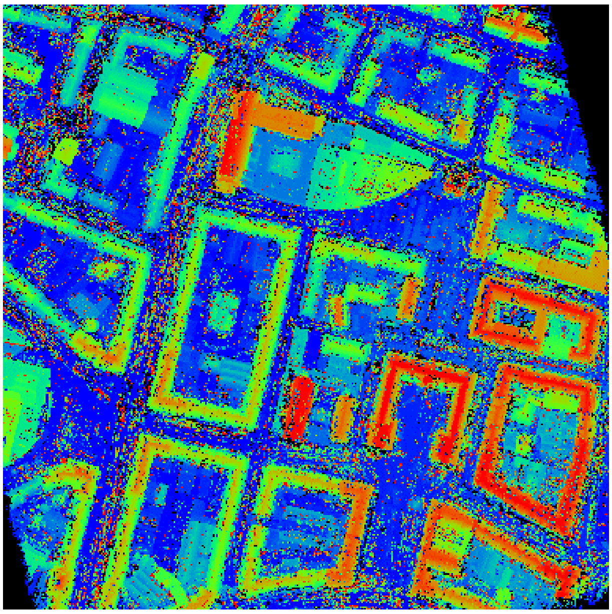

Figure 6: 2000x2000 Raw DSM obtained with our process on 4 images and 3x3 windows

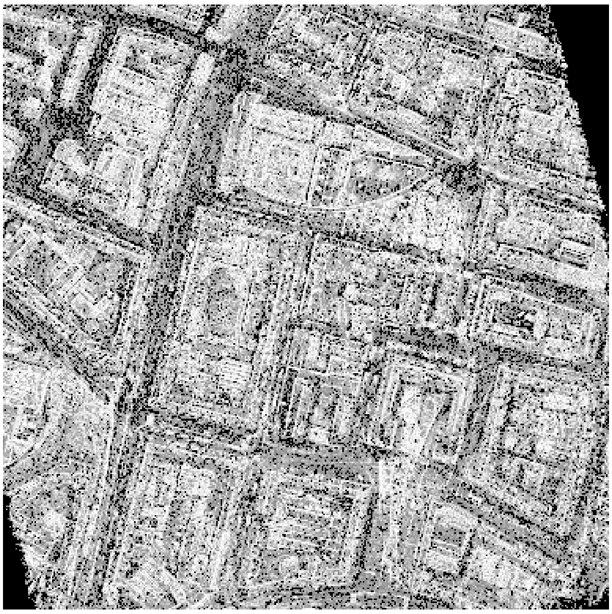

Figure 7 : Correlation scores corresponding to Figure 6

20 cm ground pixel size digital survey on the french town of Amiens with a 60% across and along track overlap

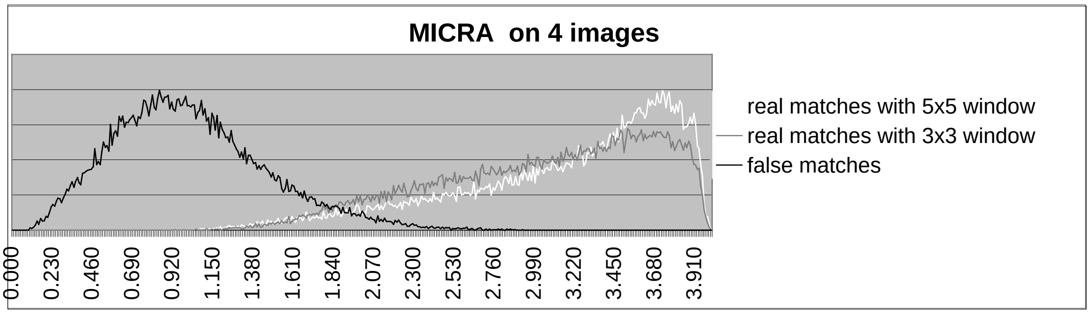

Figure 8 : Correlation score histograms

A simple experiment of analysis of the correlation score distributions show this very clearly. Let us consider real and aberrant score distributions. We call real scores those found with our process, and aberrant scores those obtained by the same process with the proviso that one deliberately stands within an altimetric search interval of the same amplitude but not containing the surface. One can observe that the two distributions are but slightly mixed (Fig. 8). From an inspection of the curves, it can be said that the points having correlation scores above 2.9 are almost 100% sure.

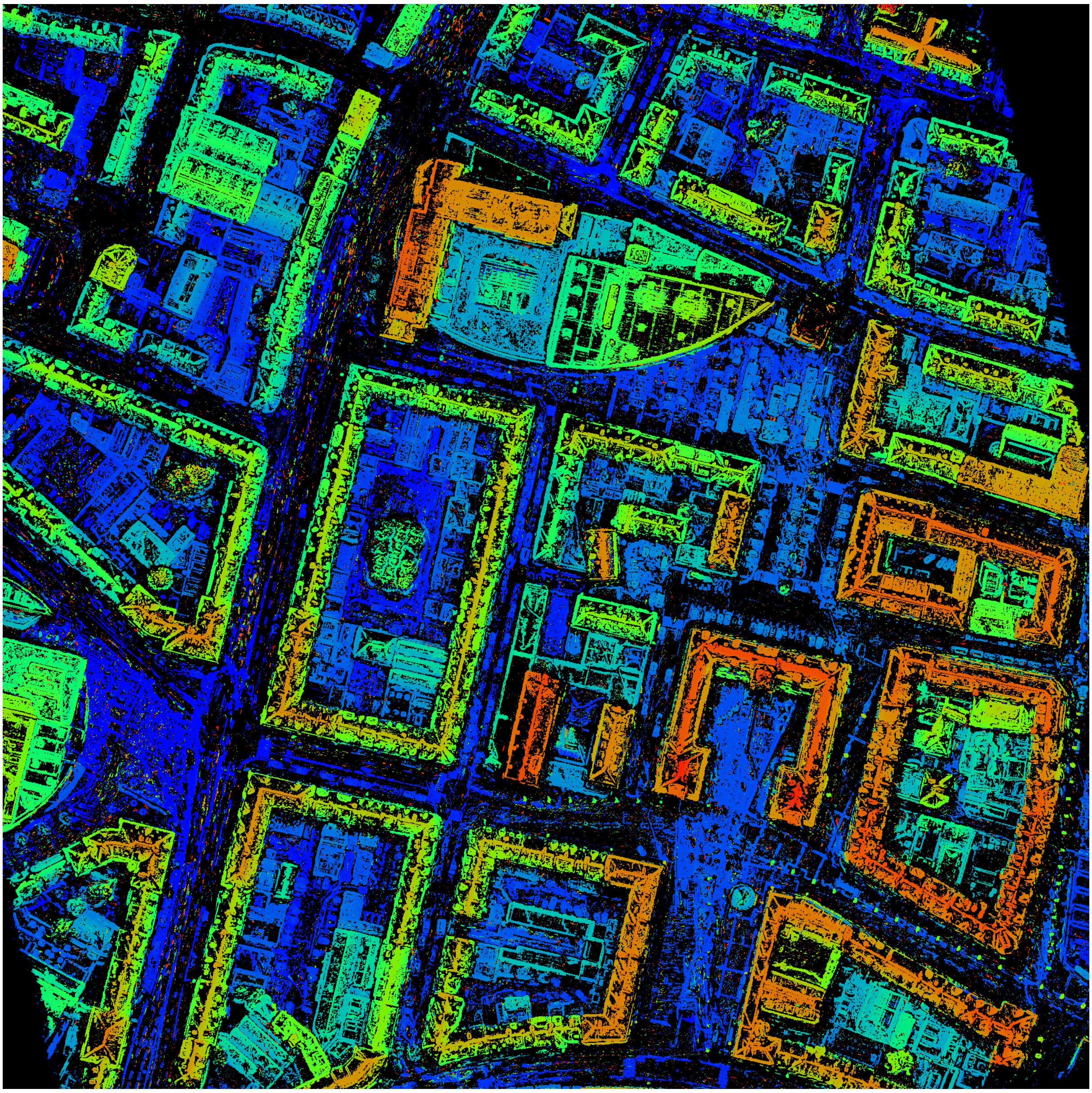

Figure 9 : DSM of Figure 6 filtered with a 2.9 threshold on the correlation scores of Figure 7.

The relative distribution separation in the case of multi-image matching allows to define a valid criterion for false points filtering (Fig. 9). It can be noticed that greater the number of images, the easier is the separation of the two distributions. In the case of stereo-processing, the mixing of the distributions is considerable. This explains the difficulty in and the impossibility of defining a reliability criterion and a satisfying rejection threshold on the correlation score values. A low threshold retains a great number of aberrant points, while a high one rejects a very great number of valid points.

7. GENERAL STRATEGY FOR THE GENERATION OF A DENSE AND HIGH QUALITY DSM

As expected, the previously described radiometric weighting function penalises those areas which are not seen (hidden areas) or rather which are "well seen in the same way" only in a subset of images (non lambertian surfaces, mobile vehicles, etc.). Our matching process tends voluntarily to determine solely the landscape points seen in a rather similar way on all the images. As a result, the filtering process will remove all other points and leave missing areas inside the DSM. In consequence, our general strategy is to complete in a progressive and hierarchical way the DSM starting from the most reliable measures (with the filtering process) in order to avoid the propagation of errors.

The restitution of homogeneous areas (where 3x3 windows did not succeed) is obtained by reiterating the same process (multi-image matching, radiometric weighting, filtering) with larger sizes of windows (5x5 and 7x7). In order to complete still missing areas, the same strategy is locally applied but restricted to a deterministic subset of images. As for the hidden parts, we determine for each missing point, with the help of all the DSM samples previously computed and of a z-buffer algorithm, the image subset upon which our DSM grid point is really seen and we apply to these points the matching process on this image subset. As for all the remaining areas, i.e. the non lambertian surfaces and the mobile vehicles, the median value of all the correlation scores calculated on all the subsets of n minus one images is taken. The completion of the hierachical matching strategy is under current investigation.

8. CONCLUSION

We have described a very original approach for the automatic surface reconstruction of urban areas from highly overlapping (across track as well as along track) aerial images. This system with all its simplicity in use and strategy, supplies extremely promising results with particularly very interesting properties for orthoimage production, relief micro-structure detection, and especially for building reconstruction. On the one hand, the generated DSMs are dense, reliable, accurate, and correctly reproduce all the relief morphology (discontinuities, steep slopes, slope breaks, microstructures, etc.) thanks to the very small window sizes we use. On the other hand, a reliability indicator for each DSM sample is obtained, which can be used either for a direct back-tracking on our processes and a real hierarchical strategy, or to point out "uncertain" areas so that other processes based on the analysis of the DSM may take these into account. The morphological quality of the DSMs produced with our image processing approach and strategy on high quality digital images should be better than that obtained with LASER ranging techniques, and the accuracy should be very close.

REFERENCES

[Baillard &al 96] Baillard C., Dissard O., Jamet O., Maître H. Extraction and characterization of above-ground areas in a peri-urban context, *Proccedings of the IAPR/TC-7, Workshop on Methods for Extracting and Mapping Buildings, Roads and Other Man-made structures from Images*, pp. 159-174, Graz, Austria, 2-3 September, 1996.

[Canu & al 95] D. Canu, N. Ayache, J.A. Sirat, « Accurate and robust stereovision with a large number of aerial images », *SPIE*, Vol. 2579, pp. 152-160, Paris, September, 1995.

[Cord & al 98] M. Cord, N. Paparoditis, M. Jordan. « Dense, reliable and depth discontinuity preserving DEM computation from. H.R. V. images», proceedings of ISPRS Commission II symposium, Data Integration : Systems ans techniques, vol 32 n°2, pp. Cambrige, Grande-Bretagne, July, 1998.

[Fuchs & al 99] F. Fuchs, H. Le Men, « Building reconstruction on aerial images through multi-primitive graph matching», Proceedings of the 2nd IAPR workshop on graph-based representations, 1999.

[Gabet & al 94] L. Gabet, G. Giraudon, L. Renouard, « Construction automatique de Modèles Numériques de Terrain à haute résolution », Bulletin SFPT n°135, 1994.

[Jibrini & al 00] H. Jibrini, , N. Paparoditis, M.P. Deseilligny, H. Maître, Building reconstruction from very high resolution digital stereopairs using cadastral ground plans, *International Archives of Photogrammetry and Remote Sensing, ISPRS, Geo-Information for all*, Amsterdam, Pays-Bas, July, 2000.

[Leloglu & al 98] U.M. Leloglu, M. Roux, H. Maître. « Dense urban DEM with three or more high resol ution aerial images », ISPRS Symposium on GIS, Between Vision and applications, Stuttgart, Septembre 1998.

[Okotumi & al 91] M. Okotumi and T. Kanade, « Multiple Baseline Stereo », IEEE transactions on Pattern Analysis and Machine Intelligence, vol 15 n°4, Avril 1993.

[Paparoditis & al 98] N. Paparoditis, M. Cord, M.Jordan, J.P. Cocquerez. Building detection and Reconstruction from Mid to High Resolution Images. *Computer Vision and Image Understanding*, 72(2), pp. 122-142, November 1998

[Thom & al 99] C. Thom, J.P Souchon, « IGNs digital aerial frame camera: work in progress », *In proceedings of the Photogrammetric Week*, Stuttgart, September 1999.