A COMPARISON OF SEMIGLOBAL AND LOCAL DENSE MATCHING ALGORITHMS FOR SURFACE RECONSTRUCTION

E. Dall'Asta, R. Roncella

DICATeA, University of Parma, 43124 Parma (PR), Italy –
elisa.dallasta@studenti.unipr.it, riccardo.roncella@unipr.it

Commission V, WG V/1

KEY WORDS: DSM, Accuracy, Matching, Reconstruction, Comparison, Algorithms

ABSTRACT:

Encouraged by the growing interest in automatic 3D image-based reconstruction, the development and improvement of robust stereo matching techniques is one of the most investigated research topic of the last years in photogrammetry and computer vision.
The paper is focused on the comparison of some stereo matching algorithms (local and global) which are very popular both in photogrammetry and computer vision. In particular, the Semi-Global Matching (SGM), which realizes a pixel-wise matching and relies on the application of consistency constraints during the matching cost aggregation, will be discussed.
The results of some tests performed on real and simulated stereo image datasets, evaluating in particular the accuracy of the obtained digital surface models, will be presented. Several algorithms and different implementation are considered in the comparison, using freeware software codes like MICMAC and OpenCV, commercial software (e.g. Agisoft PhotoScan) and proprietary codes implementing Least Square e Semi-Global Matching algorithms. The comparisons will also consider the completeness and the level of detail within fine structures, and the reliability and repeatability of the obtainable data.

1. INTRODUCTION

Identifying depth information is an indispensable task to get detailed 3D models in different application areas (land survey and urban areas, remote sensed data, Cultural Heritage items, etc..). One of the most investigated research topics in photogrammetry and computer vision is the improvement of accurate stereo matching techniques, with respect to robustness against illumination differences, as well as efficiency and computational load. In both fields, several stereo matching methods have been developed and refined over the years. In particular in the last few years, in Computer Vision, an incredible amount of new matching algorithms has been developed even if, sometimes, without sufficient insight of the accuracy and metric quality of the results.
The Semi-Global Matching stereo method (SGM) (Hirschmuller, 2005 and 2008) is one of the many techniques spread in photogrammetry and Computer Vision, and its succesfully results have encouraged the algorithm implementation by many researchers and companies. It realizes a pixel-wise matching and relies on the application of consistency constraints during the cost aggregation. Combining many 1D constraints realized along several paths, symmetrically from all directions through the image, the method performs the approximation of a global 2D smoothness constraint which allows detecting occlusions, fine structures and depth discontinuities. In particular the regularity constraints allows using very small similarity templates (usually 1÷5 pixel) making the method particularly robust where shape discontinuities arise; on the other hand, traditional area based (template) matching techniques, using bigger templates to achieve good accuracies, are more prone to such issues.
In this paper, the results of some tests performed with different stereo matching application will be discussed, with the purpose to inspect the accuracy and completeness of the obtained three-dimensional digital models. In another section, the influence on the SGM algorithmic implementation, of several process variables involved during the cost aggregation step will be presented. Indeed, the choices of the cost function used for the stereo matching, the minimization method of this function, as well as the penalty functions which are used to enforce disparity continuity, are necessary connected with the regularity and reliability of the results.
Some of the images and models used in the study are extracted from well-established datasets used in scientific dense matching reconstruction benchmarking; being important to provide an error-free dataset, some artificial, computer generated 3D models, were used as well.

2. RELATED LITERATURE

The wide range of modern dense matching algorithms can be assigned to two big category: local methods, which evaluate the correspondences one point at a time, not considering neighbouring points/measures, and global methods where some constraint on the regularity of results are enforced in the estimation. The first approach can be very efficient and accurate making it particularly suitable for real-time applications; however, these algorithms are considerably sensitive to the presence of image regions characterized by sudden depth variations and occlusions and often can produce very noisy results in low contrasted textured regions. At the same time, these methods assume continuous or functionally variable disparities within the correlation window leading to incorrect results at discontinuities and detailed areas. The global methods, trying to overcome this limits, are less sensitive to ambiguous disparity values (occlusions, repeated patterns and uniform texture regions). In these algorithms the problem of finding the correspondences deals with the minimization of a global cost function extended (usually) to all image pixels: the use of global constraints is an additional support to the solving process and allows obtaining successfully results even with areas which may be difficult to compute with local methods. Finding the minimum of the global energy function can be performed with different approaches: for example scan-lines and two dimensional optimization. The first method can be efficiently dealt with, for example, Dynamic Programming algorithm (Birchfield & Tomasi, 1998; Van Meerbergen et. al., 2002) while the second mathematical optimization technique, can be implemented using Graph Cuts (Boykov et al., 2001; Kolmogorov & Zabih, 2001) or Belief Propagation (Sun et al., July 2003). Similarity cost functions, based generally on intensity differences of the images pixels, have a significant relevance on the disparity estimation. Correlation based stereo matching, like Normalized Cross Correlation (NCC) and Sum of Absolute Differences (SAD), have always been favorite in the past for dense depth maps reconstruction. Actually, many other tequinques can be used to perform the initial image correspondeces. Non-parametric costs like Rank and Census (Zabih & Woodfill, 1994), rely on the relative ordering of the local intensity values and not on the intensity values themselves, and are preferred for reducing outliers caused by non corresponding pixels. The original SGM formulation (Hirschmuller et. al., 2005) implements Mutual Information cost function which may foreseeably be used in a wide variety of imaging situations (Viola & Wells, 1997).

3. STEREO MATCHING COMPUTATION

Stereo matching is used to correlate points from one digital image of a stereo pair with the corresponding points in the second image of the pair. However, finding the best algorithms and parameters, is usually difficult, since different aspects can be considered: accuracy, completeness, robustness against radiometric and geometric changes, occlusions, computational efforts, etc. The Semi-Global matching is, actually, one of the best matching strategies used both in photogrammetry and computer vision, offering good results with low runtime.

Several considerations about the implemented matching cost functions (used to realize pixels correlation), the aggregation step that combine these costs and, finally, the choice of penalty functions which penalize depth changes, need to be evaluated.

3.1 Semi-Global Matching

The Semi-Global Matching method (Hirschmuller, 2008 and 2001) performs a pixel-wise matching allowing to shape efficiently object boundaries and fine details.

The algorithm works with a pair of images with known interior and exterior orientation and epipolar geometry (i.e. assumes that corresponding points lie on the same horizontal image line). It realizes the minimization of a global smoothness constraint, combining matching costs along indipendent one-dimensional paths trough the image.

The first scanline developments (Scharstein & Szeliski, 2002), exploiting a single global matching cost for each individual image line were prone to streaking effects, being the optimal solution of each scan not connected with the neighboring ones. SGM algorithm allows to overcome these problems thanks to the innovative idea of symmetrically compute the pixel matching cost through several paths in the image. With a a known disparity value, the costs extract by the each path are summed for each pixel and disparity value. Finally, the algorithm choses the pixel matching solution with the minimun cost, usually using a dynamic programming approach. The cost $L'_r(\boldsymbol{p}, d)$ (as defined in the oringinal Hirshmuller paper (Hirshmuller, 2005)) of the pixel p at disparity d, along the path direction r is defined as:

$$
\begin{aligned}
L'_r(p,d) = C(\boldsymbol{p},d) + \min(&L_r(\boldsymbol{p}-\boldsymbol{r},d), \\
&L_r(\boldsymbol{p}-\boldsymbol{r},d-1) + P_1 \\
&L_r(\boldsymbol{p}-\boldsymbol{r},d+1) + P_2 \\
&L_r(\boldsymbol{p}-\boldsymbol{r},d+1) + P_2 \\
&\min_i L_r(\boldsymbol{p}-\boldsymbol{r},i) + P_2) - min_k L_r(\boldsymbol{p}-\boldsymbol{r},k)
\end{aligned}
\qquad (1)
$$

where the first term is the similarity cost (i.e. a value that penalize, using appropriate metrics, solutions where different radiometric values are encountered in the neighbor area of the corresponding points) , whereas the second term evaluates the regularity of the disparity field, adding a penalty term P_1 for little changes and P_2 for all larger disparity change with respect to the previous point in the considered matching path. The two penalty values allow to describe curved surfaces and to preserve disparity discontinuities, respectively. Since the cost gradually increases during cost aggregationg along the path, the last term allows to reduce the final value subtracting the minimum path cost of the previous pixel from the amount.

Minimization operation is performed efficiently with Dynamic Programing (Van Meerbergen, et al., 2002) but, in order to avoid streaking effects, SGM strategy has introduced the innovative idea of computing the optimization combining several individual path, symmetrically from all directions through the image. Summing the path costs in all directions $\boldsymbol{r}$ and searching the disparity with the minimal cost for each image pixel $\boldsymbol{p}$, produce the final disparity map. The aggregated cost is defined as

$$S(\boldsymbol{p},d) = \sum_r L_r(\boldsymbol{p},d) \qquad (2)$$

and, for sub-pixel estimation of the final disparity solution, the position of the minimum is calculated fitting a quadratic curve through the cost values of the neighbours pixels.

Similar approaches, where the surface reconstruction is solved through an energy minimization problem has been evaluated in (Pierrot-Deseilligny & Paparoditis, 2006). He has implemented a Semi Global matching-like method identifying the formulation of an energy function $E(Z)$ described as:

$$E(Z) = \sum A\big(x,y,Z(x,y)\big) + \alpha * F(\vec{G}(Z)) \qquad (3)$$

where

- Z is the disparity function;
- $A\big(x,y,Z(x,y)\big)$ represents the similarity term;
- $F(\vec{G}(Z))$ is the positive function expressing the initial parameters which characterize the surface regularity;
- α represents the weight to permit the data adaptation to the image content (i.e. the weight of disparity regularization enforcement).

This formulation supposes the existence of an initial approximated solution (avoidable using combinatorial approaches - Pierrot-Deseilligny & Paparoditis, 2006).

3.2 Matching costs

Area based matching methods are the basic techniques to find corresponding pixels; however, correlation usually assumes equivalent depth values for all pixels of a correlation window even if this hypothesis is violated at depth discontinuities or with strong perspective changes between matching images. Using small templates in the matching stage can lead to noisy, low precision results; on the other hand using larger templates usually make constant depth hypothesis more inadequate and/or produce smoother results, losing information where small object shape details are present. In other words the size of the correlation window influences the results accuracy and completeness: small correlation windows improves object level of detail, but it can give a unreliable disparity estimation because it does not cover enough intensity variations (Kanade & Okutomi, 1994); on the other hand, big windows size don't allow estimating sudden depth changes leading to erroneous match pairs and produce smoother surfaces.

On the other hand, these methods are often used (especially by the Computer Vision community) because the image correlation is very fast and the whole computation demands low runtime and memory occupance, compared to other matching methods (e.g. Least Squares Matching - (Grun A., 1985)).
Some of the most popular parametric correlation techniques, used both in photogrammetry and computer vision, are Sum of Absolute Differences (SAD) and Normalized Cross Correlation (NCC) while, in the recent past, Zabih and Woodfill (Zabih & Woodfill, 1994), introduced non-parametric measures like Rank and Census.

3.2.1 Sum of Absolute/Squared Differences

Sum of Absolute Differences (SAD) is one of the simplest similarity measures commonly implemented for image correlation. It performs the absolute difference between each pixel of the original image and the corresponding pixel in the match image, using a search window to realize the comparison. Similarly, in Sum of Squared Differences (SSD) the differences between corresponding pixels are squared. Later, these differences are summed and optimized with the winner-take-all (WTA) strategy (Kanade et al., 1994).
SAD and SSD formulations have the following expression:

$$SAD = \sum_i \sum_j |f(i,j) - g(i,j)| \qquad (4)$$

$$SSD = \sum_i \sum_j (f(i,j) - g(i,j))^2 \qquad (5)$$

considering a block f centred in (x,y) position on the master image and the corresponding block on the slave shifted by $(\Delta x, \Delta y)$.

3.2.2 Normalized Cross Correlation

Normalized Cross Correlation (NCC) is more complex than both SAD and SSD (Sum of Squared Differences) but it is invariant to linear changes in image amplitude. Normalizing features vectors to unit length, the similarity measures between the features becomes independent to radiometric changes (Lewis, 1995).
The NCC finds matches of a reference template $f(j,i)$ of size m×n in a scene image $g(x,y)$ of size M×N and it is defined as:

$$\rho(i,j) = \frac{\sum_i \sum_j [(f(j,i) - \overline{f}) \cdot (g(j+\Delta x, i+\Delta y) - \overline{g}]}{\sqrt{\sum_i \sum_j \left[(f(j,i) - \overline{f})^2 \cdot (g(j+\Delta x, i+\Delta y) - \overline{g})^2\right]}} \qquad (6)$$

where $\overline{f}$ e $\overline{g}$ represent the corresponding sample means.
A unitary value of the NCC coefficient indicates a perfect matching window.

3.2.3 Census Transform

The census transform is a fairly new area-based approach to the correspondence images problem (Zabih & Woodfill, 1994); it realize a non-parametric summary of local spatial structure followed by a correlation method using an Hamming distance metric.
The transform maps the intensity values of the pixels within a square window W to a bit string where pixels intensity are compared to the window central pixel intensity p. The boolean comparison returns 1 when the pixel intensity is lesser then the central pixel, else 0. That is:

$$R(p) = \otimes_{p' \in W} \; \xi(I(p'), I(p)) \qquad \begin{array}{l} \xi(i,j) = 1, i < j \\ \xi(i,j) = 0, i > j \end{array} \qquad (7)$$

where ⊗ represent the concatenation.
Census transformed images values are then compared to perform the stereo matching: the two bit strings are evaluated identifying the number of pixels that change from one string to the other. In this way, the order of the local intensity does not change and all the radiometric variations between the images are irrelevant.
Census transform has been evaluated as the one of the most robust matching cost for stereo vision (Hirschmuller, 2001; Banks & Corke, 2001).

3.2.4 Rank Transform

The rank transform (Zabih & Woodfill, 1994) defines the number of pixels p' in a square window W whose intensity is less than the luminosity value of the central pixel p:

$$R(p) = \| \{ p' \epsilon \; W(p) | \; I(p') < I(p) \} \| \qquad (8)$$

In this case the function output is not an intensity value but an integer and the image correspondence is realized with SAD, SSD or other correlation methods on the rank transformed image. In other words, the images can be pre-filtered with the rank transform and then compared using one of the previous metrics. In our implementation Rank transform was used in conjunction with SAD metric.

4. IMAGE MATCHING STRATEGIES DESCRIPTION

Many image matching and surface reconstruction methods have been developed in recent years implementing, on one hand, well-established techniques like Least Square Matching (LSM - Grun A., 1985), on the other innovative global and semi-global matching methods.
Several algorithms and implementations are considered in this comparison study, including freeware software codes, commercial software and home variant of Least Square e Semi-Global Matching strategies.

4.1 DenseMatcher

DenseMatcher is a digital terrain model generation program developed at the University of Parma (Re et al., 2012). It implements NCC, LSM and Multiphoto Geometrically Constrained Matching (MGCM - Grun & Baltsavias, 1988) correlation algorithms and it uses a multi-resolution robust approach processing more levels of an image pyramid. Using known intrinsic and extrinsic orientation parameters, the algorithm perform the epipolar resampling of the image pair improving the efficiency of the process; then, using tie-points information (or an initial approximate depth map), it realizes the disparity data optimization with an additional NCC matching step (optionally at each level of the pyramid). The LSM proceeds to obtain the final correspondences with a parallel dense matching procedure.
The implemented codes allows to control several parameters such as the number of the pyramid levels, the template size, the correlation coefficient threshold, the correlation algorithm and many others variables involved in the matching process.
Actually the LSM module takes on input image sequences producing progressively DSM of consecutive image pairs but a multi-image matching extension is under development.

4.2 MicMac

APERO and MICMAC are two open source tools realized at IGN (Institut National de l'information Géographique et

Forestière, Paris) that allows to realize all the steps of a typical photogrammetry process, starting from Structure&Motion images processing up to dense point clouds and orthophotos generation. APERO is the orientation images software, which uses both computer vision approach for estimation of initial solution and photogrammetry for a rigorous compensation of the total error (Pierrot Deseilligny & Clery, 2011). It allows processing multi resolution images and, for each resolution level, it computes tie points extraction for all images pair performing finally a bundle block adjustment (Triggs et al., 2000).

The final DSM generation phase is performed with the MICMAC tool which produces the depth maps, and consecutive 3D models, from the oriented images. This step is performed using the semi-global approach which solves the surface reconstruction problem under the form of the minimization of an energy function (Pierrot-Deseilligny & Paparoditis, 2006).

The software is very interesting for the photogrammetric community because it provides statistical information of the data and allows detailed analysis of the photogrammetric processing results. Moreover, all the parameters and the results of the orientation and matching step are stored in XML files which can be adapt whenever the user needs to impose certain settings and values at the processing parameters.

4.3 Our implementation of SGM

As described in section 3.1 (and in previous section also), the development of semi-global matching techniques is very important for computing 3D models from stereo image acquisition The implementation of this methods requires the introduction of many parameters and their evaluation is fundamental to have good performances and accurate results. For this reason, the development of a proprietary code (still improved work in progress by our research group) enabled the evaluation of the best variables values and formulations of the matching cost which are involved in the disparity map generation. For instance, the application of different cost function (Rank, Census, SAD, NCC, etc.) in the stereo matching step is instrumental in the computation of the depth maps.

In regards to the considered area based matching methods, the size of the template used to compute the correspondences have also been considered. The ideal block size to perform the stereo matching depends on the chosen function and the evaluated object: in analogy with other techniques for DTM generation in close range (e.g. LSM), there seems to be an optimal range for template size value according to object features (Re, 2012).

Finally, as described in (Hirschmuller, 2008) and in 3.1 paragraph of the paper, the implemented penalty functions are closely related with the intensity changes. Adapting P_2 formulation and value to the image information, we can improve the algorithm performances to enforce the disparity continuity and penalize sudden disparity changes.

In order to perform an accurate and efficient stereo matching, the developed software implements a multi-resolution approach using image pyramids and a coarse-to-fine disparity map evaluation.

In the following section the results will highlight the influence of these different variables on the final reconstructed digital surface models with the aim to identify the best strategy and parameters combination that allows the most accurate description of different object typologies.

4.4 Photoscan

Agisoft PhotoScan is a commercial software, developed by Agisoft LLC company. It has a very simple graphical user interface and, as MicMac, it is able to perform both the orientation and the following dense stereo matching steps using a multi-image approach. Initially the software defines the images orientation and refines calibration camera parameters (the geometry of the images sequence allows to estimate a set of interior orientation parameters for each camera, whether these are not previously assigned); in a second step, it proceeds to the DSM generation. Differently to MicMac, PhotoScan doesn't display the statistical results of the photogrammetric processing, being a sort of "black-box" software. All the photogrammetric process is performed with a high level of automation and the user can decide the desired points cloud density and the 3D modelling quality. The workflow is therefore extremely intuitive being an ideal solution for less experienced users. Due to commercial reasons very few information about the used algorithms are available: some details can be recovered from the Photoscan User forum (PhotoScan, 2014). Apparently except a "Fast" reconstruction method, selectable by the user before the image matching process starts, that use a multi-view approach, the depth map calculation is performed pair-wise (probably using all the possible overlapping image pairs) and merging all the results in a single, final, 3D model. In fact, a multi-baseline matching extension is more robust with regard to occlusions detection and wrong matches, realizing the fusion of disparity information given by all the match images and producing smoother results.

Anyway in all the following example only stereo-pair were considered.

4.5 OpenCV libraries

OpenCV (Open Source Computer Vision Library: http://opencv.org) is an open-source BSD-licensed library written in C, C++, Python and Java that offers high computational efficiency and a simple use of Computer Vision and Machine Learning infrastructures. The library, developed by Intel in 1998, is cross-platform, running on Windows, Linux, Mac OS X, mobile Android and iOS; it contains several hundreds of optimized algorithms for image processing, video analysis, augmented reality and many more, providing the tools to solve most of computer vision problems.

Using IPP (Intel Performance Primitives) it provides an improvement in processing speed and optimization that are very useful for real time applications (Bradsky & Kaehler, 2008). In 2010 a new module that provides GPU acceleration was added to OpenCV and, right now, the library is expanding day-by-day. In order to perform the image matching strategies comparison, the open library for computing stereo correspondence with semi-global matching algorithm was used. The method executes the semi-global block matching (SGBM) (by Hirschmuller, 2008) on a rectified stereo pair, introducing some pre and post processing steps of the data. Several matching parameters may be controlled and set to a custom value but, in order to isolate only the matching step contribution, these additional processing parameters were disabled.

OpenCV version of the SGM strategy is focused on speed and, in contrast to our implementation of SGM (more close to the Hirschmulller implementation), calculates the disparity following the sub-pixel implementation described in (Birchfield et al., 1998), using less path directions to calculate the matching costs (8 paths, instead of 16).

doi:10.5194/isprsarchives-XL-5-187-2014

5. IMAGE DATA: THREE TEST CASES

In order to understand the main performance differences between the different strategies/implementation, three tests on real and synthetic images have been performed. An exhaustive description about the different datasets and three-dimensional digital surface models (DSM) used as reference data in the comparisons is presented in the next sections.

5.1 Synthetic images of 3D computer-generated basic shapes

First of all, in order to evaluate the performance and the best parameters combination used by the different stereo matching approach, 3D simple scenery was created using 3D modelling software. Spherical and rectangular objects were located on a wavy surface creating significant depth changes in the scene (as it is visible in Figure 1).
An ideal, well-contrasted texture was draped to the objects favouring the matching algorithms performance. Two virtual cameras were located in the scene taking two nadiral synthetic images. Optimum illumination conditions were simulated, producing light and shadows useful for a simpler (human) depth identification.

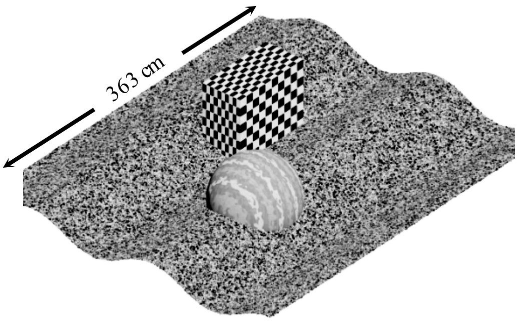

Figure 1: Computer-generated 3D primitives dataset.

5.2 Synthetic images of a 3D reference model

Using the same 3D modelling program, a 3D model of an architectural element was imported and covered with an ideal texture. The chosen object is a Corinthian capital 1.33 meter high and with a circular base of 90 cm diameter, characterized by complex architectural details.

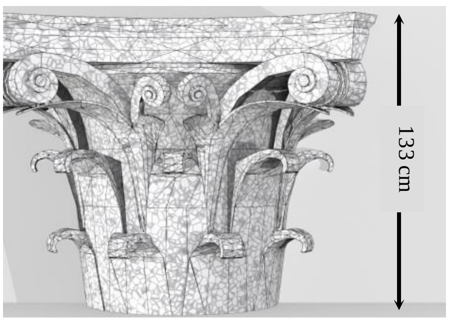

Figure 2: Synthetic image of a 3D reference model.

As in previous case, two virtual nadiral cameras were created, with known interior and exterior orientation parameters. To simulate a photorealistic scenario, illumination sources were located using photometric virtual lights. Finally, images rendered through a raytracing algorithm by the cameras can be generated and exported (a nadiral image is shown in Figure 2).

5.3 Real images and reference DSM

The third case is an image pair extracted from a sequence of a 5 meter high richly decorated fountain from the cvlab dataset (Stretcha, von Hansen, Van Gool, Fua, & Thoennessen, 2008). The dataset consists of known interior orientation parameters, distortion removed images and a reference laser scanning DSM. The exterior orientation parameters were estimated through a Structure from Motion procedure, followed by a bundle adjustment step using some features extracted from the DSM as Ground Control Point (GCP). The availability of a laser scanning surface reference model of the fountain has allowed to validate the results of the surface reconstructions. Figure 3 illustrates one of the two convergent images used in the stereo matching.

Figure 3: Real image of the fountain.

6. RESULTS

6.1 DSM generation and comparison procedure

As mentioned before, the comparison will be focused on shape differences between the reconstructed and the reference digital surface models. The evaluated matching methods implement different strategies for generating the final DSM. In order to obtain comparable solutions, the models were generated using known internal and external orientation parameters, making the results as independent as possible from the automatic orientation procedure. In fact, both PhotoScan and MicMac software are able to perform, beside the DSM generation, also the automatic orientation of the image block; however, different orientation solutions can produce unwanted DSM deformation. On the other hand OpenCV library expects to work with rectified images (i.e. corresponding points in the stereo pair lay on the same horizontal image line) and produce a disparity map. Differently from MicMac and PhotoScan, a subsequent triangulation stage, to be carried out externally, was required to produce the final DSM.
Each test case was performed varying image matching method and parameters, which have relevant influence on the digital model accuracy. In order to ensure the correct evaluation of the DSM precision, no post-processing steps were performed.

6.2 Matching costs optimization

The proprietary implementation of a Semi-Global Matching method (Hirschmuller, 2008) has allowed studying the influence of some matching variables on the final DSM accuracy data.
The first study case, simulating significant depth changes, appeared as an optimal dataset for characterizing the contribution of the variables involved in the DSM generation process.
Dynamic programming method requires the implementation of cost function for computing the disparity estimation. The results of four different cost functions implementation, by changing the template windows size, are shown below in Table 1; the

accuracy data are expressed in term of standard deviation of the points distances with respect to the reference model.

	NCC	SAD	CENSUS	RANK
Block Size 1	0.523	0.534	-	-
Block Size 3	0.523	0.510	0.526	0.534
Block Size 7	0.512	0.516	0.530	0.535
Block Size 11	0.471	0.495	0.517	0.535
Block Size 21	0.468	0.500	0.504	0.546

Table 1: Accuracy of the reconstructed DSM (in cm) for all implemented cost functions.

Table 1 shows interesting data: while NCC (and less clearly SAD) accuracy improves with the increase of template size (with an overall 10% gain), other cost functions (Census and Rank) don't show the same trend and a significant improvement is not manifest. Previous experiences (Re, et al., 2012) by the authors stressed the importance of the block size for the reconstruction accuracy; in this case the influence is less evident, most likely because the regularity constraint limits the measurement noise.
It's worth noting, anyway, that the improved large block size results obtained using NCC cost function are probably not sufficiently satisfying if the computational efficiency of the processing is considered.
Also the role of the implemented penalty function is a crucial aspect in depth evaluation step. Currently, two penalty function are implemented in the code: the costs penalization proposed by Hirshmuller, which introduces the two penalty parameters P_1 and P_2 for adapting to image contents (eq. 1.), and a different penalization method, characterized by a linearly increasing penalty with the disparity difference of neighbour pixels. The P_1 and P_2 optimal values are closely related with image data, the chosen cost function and the block size, since different metrics are characterized by different cost ranges, and some additional tests were performed to identifying this relationship. In other words, as a first level of understanding, the analysis was important to identify the correlation between the penalty coefficients, the image content and cost functions.
On the other hand, at this stage, the results of these elaborations haven't produced significant information about the best cost or penalty function to use: the evaluated functions have shown solutions with similar accuracy values. Further analysis would be beneficial to optimize our SGM implementation in the next future.

6.3 Relative accuracy and reliability

The relative accuracy of the DSM reconstructed models with synthetic and real images is summarized in Table 2 where, for each test case, statistics of the distances between reconstructed and reference DSM are presented. Disparity map comparison (i.e. evaluating the algorithm accuracy in finding the corresponding point on image space) were considered as well but, for some methods the parallax field are not directly accessible and are hard to be computed. Also, the influence on the final reconstruction accuracy was considered more interesting.
To make the results independent of the total size of the object all the distances standard deviations are normalized with respect to the best value. At the same time, some methods, present in some areas of the model very evident gross errors that must be removed from the relative accuracy computation. On the other hand, it's important to highlight which algorithm produces more reliable results (in terms of inlier percentage): for each model, tolerance ranges were selected based on some assumption about image matching (and consequent reconstruction) a-priori precision and on the actual performance of the best method. In particular the ranges (1 cm for 3D Shape, 3 mm for Capital and 3 cm for the fountain case study) are selected considering that at least one algorithm must produce a 90% in-tolerance 3D model: in this way reconstruction accuracy is related to a sort of quality completeness for each method.

	DM	OpenCV	PS	SGM	MicMac
3D Shapes	100 %	92 %	91 %	83 %	94 %
	86.9 %	91.4 %	88.7 %	83.2 %	80.9 %
Capital	100 %	85 %	83 %	71 %	81 %
	77.5 %	92.7 %	89.27	70.1 %	66.8 %
Fountain	90 %	97%	100 %	89 %	90 %
	84.6 %	88.7%	97.5 %	86.5 %	88.1 %

Table 2: Relative accuracy of the reconstructed DSM.

The table shows that, for each test case, the different matching algorithms produce results that are not dramatically different. The general trend is similar, though not identical, in particular for computer-generated data. Indeed, analysing the first two tests, we can identify that the best solution were obtained by LSM with DenseMatcher, followed by OpenCV, PhotoScan and our implementation of Semi-Global matching. It's worth noting that, the LSM estimate a local affine transformation between corresponding areas in the two images for every measured point, trying removing the perspective changes; on the other hand, all the other algorithms consider a similarity function that is invariant just to feature translation. The capital test case, presents the higher base-length to distance ratio, and higher perspective effects can be expected: in this case unsurprisingly the LSM achieve the highest accuracy. On the contrary, in the Capital and Fountain test cases, some area of the images lack of a well-contrasted pattern: while the semi global methods achieve good quality results also in these areas, the LSM algorithm cannot always produce complete results (see figure 8, for instance).

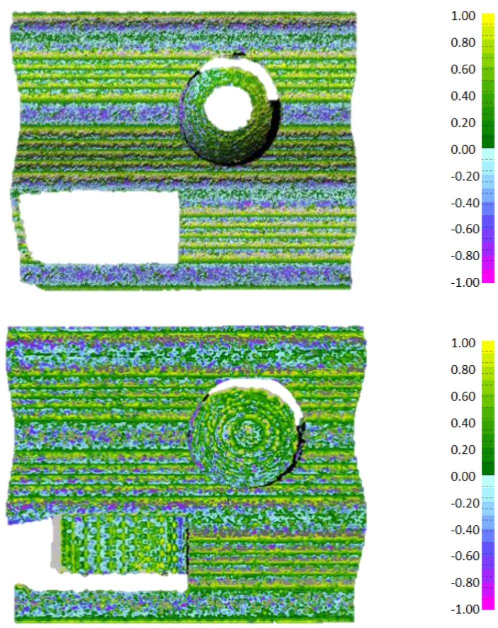

Figure 4: 3D Shape error maps (cm).
Top: PhotoScan. *Bottom*: OpenCV.

This contribution has been peer-reviewed.
doi:10.5194/isprsarchives-XL-5-187-2014

Comparing PhotoScan results, which is the only software code where smoothing and filtering procedure cannot be disabled, with the other methods, two important aspects must be analysed: the models completeness (specifically, the ability of the method to produce a complete digital model, without holes) and the smoothing effect consequences. These aspects cannot be ignored since they are connected with the evaluated data accuracy.
Figure 4 shows the error map for the PhotoScan and OpenCV DSM. The two solutions have the same metric accuracy but the same cannot be said for models completeness; two big holes, in correspondence to high depth changes, are clearly visible in PhotoScan DSM. The same can be highlighted in figure 5-*Top* and *Center*.

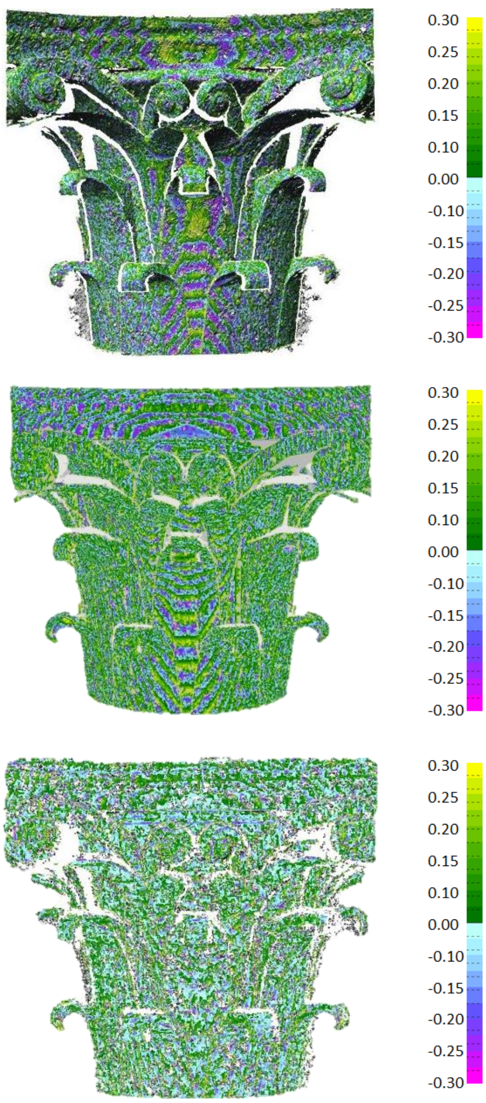

Figure 5: Corinthian Capital error maps (cm).
Top: PhotoScan. *Center*: OpenCV. *Bottom*: DenseMatcher

Therefore, low values of discrepancies from the reference model, derived from standard deviation information, cannot be the only indication of the digital model reliability: the completeness and surface distribution of the points must be evaluated. In this regard, the spatial distribution of the differences must be taking also into account, highlighting the importance of results visual assessment. As can be seen in Figure 5-*Bottom*, the error map of DenseMatcher usually shows more noisy data, due to its pointwise estimation approach: while semi-global-like methods constrains (with different degree of enforcement) the regularity of the disparity field, every point in LSM methods are considered and evaluated individually. On the other hand, the reconstructed DSM reveals the LSM ability to produce reliable results, as shown not only from standard deviation (see Table 2) but also from spatial distribution of the distance values; the disparity regularity constraints (and smoothing filtering procedure – e.g. those implemented in the PS workflow) can generate erroneous systematic surface reconstruction, if image noise, occlusions, repeated pattern influence a whole matching path.
The same behaviour is clearly visible in the last case study, which considers real images; figure 6 shows the comparison between the results obtained with our implementation of SGM, MicMac and PS. SGM and MicMac produced more noisy results but, at the same time, if the smaller object features are considered, captured finer details; the smoothing effect implemented in PhotoScan, on the contrary, produced an apparently more appealing results but with some local discrepancies (see for instance the yellow regions), flattening some detailed areas; on flat, low contrasted areas, the smoothing and filtering procedures, probably allow acquiring better results with an overall higher completeness level. Finally DM, suffers of lack of contrasted texture and higher noise due to the real image quality, producing a 3D model with higher error levels in some region.

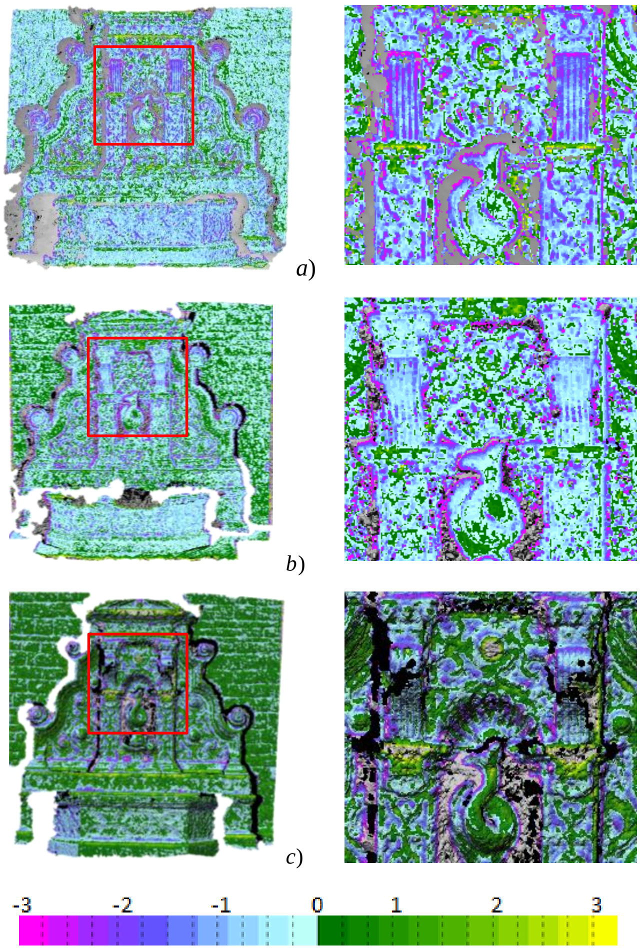

Figure 6: Fountain Error map (cm): (*a*) SGM, (*b*) MicMac, (*c*) PS - with a zoom (on the right) on a delimited area.

CONCLUSIONS

The paper presented some tests, executed in order to study the performance of different stereo matching algorithms and check the accuracy of a new, proprietary semiglobal matching software code. Unfortunately, from this first series of tests, our implementation resulted the less accurate and complete, but we expect that the algorithm can be further optimized. Overall, the compared matching strategies have shown similar accuracy values, but the indication of the data tolerance percentage, as well as error map analysis, were necessary in order to understand the reliability of these information. Despite some models presented holes and inaccurate areas, the completeness of the DSM is usually good and error maps analysis has allowed to explain their quality (in term of discrepancies distribution and noise). LSM are still the most accurate approach (at least pointwise), but presents noisy data in low-contrast or blurred regions, where semi-global matching provides better results.
As a final consideration, for future development, semiglobal matching strategies allow a much simpler multi-view approach and pre- and post- filtering implementation (not considered in this paper) that, probably, would provide in the next future higher level of improvement also in terms of accuracy.

ACKNOWLEDGEMENTS

This research, is supported by the Italian Ministry of University and Research within the project "FIRB - Futuro in Ricerca 2010" – Subpixel techniques for matching, image registration and change detection (RBFR10NM3Z).

REFERENCES

Banks , J., & Corke, P. (2001). Quantitative evaluation of matching methods and validity measures for stereo vision. *Int. J. Robotics Research*, vol. 20, pp. 512-532.

Birchfield , S., & Tomasi, C. (1998). Depth discontinuities by Pixel-to-Pixel Stereo. *Proc. Sixth IEEE Int'l Conf. Computer Vision*, (p. pp.1073-1080).

Boykov, Y., Veksler, O., & Zabih, R. (2001). Efficient approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (p. 23(11):1222–1239).

Bradsky, G., & Kaehler, A. (2008). *Learing OpenCV: Computer Vision with the OpenCV Library.* Mike Loukides.

Grun A. (1985). Adaptive Least Squares Correlation: A Powerful Image Matching Technique. *South African J. Photogramm. Remote Sensing and Cartography*, (p. 14(3): 175-187).

Grun, A., & Baltsavias, E. (1988). Geometrically constrained multiphoto matching. *PE&RS*, (p. pp. 633-641).

Hirschmuller, H. (2001). Improvements in Real-Time Correlation Based Stereo Vision. *Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision*, (p. pp. 141-148). Kauai, Hawaii.

Hirschmuller, H. (2008). Stereo Processing by Semiglobal Matching and Mutual Information. *IEE Trans. Pattern Analysys and Machine Intelligence 30(2)*, (p. pp. 328-341).

Hirschmuller, H., Scholten, F., & Hirzinger, G. (2005). Stereo Vision Based Reconstruction of Huge Urban Areas from an Airborne Pushbroom Camera (HRSC). *Lecture Notes in Computer Science: Pattern Recognition, Proceedings of the 27h DAGM Symposium*, (p. Volume 3663, pp.58-66). Vienna, Austria.

Kanade, T., & Okutomi, M. (1994). A stereo matching algorithm with an adaptive window: Theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (p. vol. 16, p. 920).

Kanade, T., Kano, H., & Kimura, S. (1994). Development of a video-rate stereo machine. *Image UnderstandingWorkshop*, (p. pp. 549–557). Monterey,CA.

Kolmogorov, V., & Zabih, R. (2001). Computing visual correspondence with occlusions using graph cuts. *International Conference for Computer Vision*, (p. pages 508-515).

Lewis, J. (1995). Fast Normalized Cross-Correlation. *Vision interface*, Vol. 10 (1), pp. 120-123.

PhotoScan. (2014). Algorithms used in Photoscan (last accessed 05.09.14). http://www.agisoft.ru/forum/index.php?topic=89.0

Pierrot Deseilligny, M., & Clery, I. (2011). Apero, An Open Source Bundle Adjusment Software For Automatic Calibration And Orientation Of Set Of Images. *In IAPRS* Vol. XXXVIII-5/W16, Trento .

Pierrot-Deseilligny, M., & Paparoditis, N. (2006). A Multiresolution And Optimization-Based Image Matching Approach: An Application To Surface Reconstruction From Spot5-Hrs Stereo Imagery. *In IAPRS vol XXXVI-1/W41 in ISPRS Workshop On Topographic Mapping FromSpace (With Special Emphasis on Small Satellites).* Ankara, Turquie.

Re, C., Roncella, R., Forlani, G., Cremonese, G., & Naletto, G. (XXII ISPRS Congress, 25 August – 01 September 2012). Evaluation Of Area-Based Image Matching Applied To Dtm Generation With Hirise Images. *ISPRS Annals*, (p. Volume I-4, 2012). Melbourne, Australia.

Re, C., Cremonese, G., Dall'Asta, E., Forlani, G., Naletto, G., & Roncella, R. (2012). Performance evaluation of DTM area-based matching reconstruction of Moon and Mars. SPIE Remote Sensing- International Society for Optics and Photonics., pp. 85370V-85370V.

Scharstein , D., & Szeliski, R. (2002). A taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision*, 47 (1/2/3), pp. 7-42.

Sun, J., Shum, H., & Zheng, N. (July 2003). Stereo matching using belief propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (p. 25(7):787–800).

Triggs, B., McLauchlan, P., Hartley, R., & Fitzgibbon, A. (2000). Bundle Adjustment -- A Modern Synthesis. *Lecture Notes in Computer Science*, Vol 1883, pp 298—372.

Van Meerbergen, G., Vergauwen, M., Pollefeys, M., & Van Gool, L. (Apr.-June 2002). A Hierarchical Symmetric Stereo Algorithm Using Dinamic Programming. *Int'l J. Computer Vision, vol. 47, nos. 1/2/3*, (p. pp. 275-285).

Viola, P., & Wells, W. (1997). Alignment by maximization of mutual information. *International Journal of Computer Vision*, (p. 24(2):137–154).

Zabih, R., & Woodfill, J. (1994). Non-parametric Local Transforms for Computing Visual Correspondence. *Proceedings of the European Conference of Computer Vision*, (p. pp. 151-158). Stockholm, Sweeden.