PHOTOGRAMMETRY AND DENSE STEREO MATCHING APPROACH APPLIED TO THE DOCUMENTATION OF THE CULTURAL HERITAGE SITE OF KILWA (SAUDI ARABIA)

J. F. Hullo [a, *], P. Grussenmeyer [a], S. Fares [b]

[a] Photogrammetry and Geomatics Group, MAP-PAGE UMR 694 – INSA Strasbourg, Strasbourg, France - (jean-francois.hullo , pierre.grussenmeyer)@insa-strasbourg.fr
[b] HiSCANT Group, EA 1132, University of Nancy, France

KEY WORDS: Dense Stereo Matching, Photogrammetry, Archaeology, Epigraphy, 3D textured models, PhotoModeler Scanner©

ABSTRACT:

Since a couple of years, several commercial solutions of dense stereo matching have been developed. This process offers a really cheap, flexible and accurate solution to get 3D point clouds and textured models. The calibration of the camera allows a subpixellar correlation for correctly textured objects. In order to define the limits of such a process for cultural heritage applications, the Photomodeler Scanner© software has been tested on an archaeological site. The French project of archaeology and epigraphy in the region of Kilwa (Saudi Arabia) is aimed on two main works: prospecting and excavating. During the annual mission of one month a lot of information is collected, but the analysis of this data could not be done on the go. During prospecting in the area of 25km radius, epigraphs and carvings have been registered in a GIS with links to simple pictures in the database. But to describe, analyze and interpret those epigraphs, simple pictures are often not sufficient. Laser scanning cannot be used in this case for several reasons: difficulty of access and conditions (such as desert wind, or heat), location of epigraphs or cost. Therefore, solutions based on Dense Stereo Matching have been investigated. Two kinds of models are presented in this paper. The first application is the modeling of epigraphs from a single stereopair during prospection. Another application has been established on the site of Kilwa, where a building has been photographed and modeled. For this model, we got about 50 pictures. We inserted it on a Digital Elevation Model of the site obtained by tacheometry in order to meet the needs of the architects and archaeologists and to make assumptions about the original condition of the site. The main advantage of this photogrammetric methodology is to get at the same time a point cloud (resolution depends on the size of the pixel on the object), and therefore a mesh, and oriented high resolution images. After processing, we can use the data exactly as a laser scanning point cloud, with really better raster information for textures.

1. INTRODUCTION

Even though the theory of Dense Stereo Matching (called DStM in our paper) from images is known for a while, the hardware of a standard workstation did not allow this process easily. Both camera and computer developments which lead to the creation of new tools are discussed in this paper. Computer vision and photogrammetry are working in a parallel way, but with different objectives. Generally, Computer Vision community uses low resolution images for 3D reconstitution, with uncalibrated cameras. Calibration is on the go. Process is fast but accuracy is not the only aim. Literature on this subject really increased in the last few years. Algorithms based on projective geometry are now straight and can deal with a lot of different configurations. Photogrammetric approach is more strongly linked to the analogical and is mainly interested in precision, more than automation.
(Strecha *et al.*, 2008; Furukawa and Ponce, 2007) showed that we can use DStM really accurately with terrestrial high resolution images even if the camera intrinsic parameters are not well known. But very often, a photogrammetrist can calibrate his camera and is thus able to use these parameters as constants, reducing the correlation process time without losing accuracy. Several matching algorithms based on epipolar constraint are compared in (Seitz *et al.*, 2006).

(Fassi, 2007) compares DStM and laser scanner technologies and it is now clear that DStM offer more than an alternative. Lately, several commercial solutions have been developed and first experiences have been done in Cultural Heritage, e.g. (Reznicek and Pavelka 2008), (Remondino and Menna, 2008).
In this paper, we will discuss an approach for the application of DStM on archaeological excavations and epigraphist investigation. First, we will sum up the principles and theory, with a short state of the art, both for academic and commercial solutions. We will then focus on the parameters that influence the quality of the results. Then we will present our work, with results both for single and multi couples of stereo images. Finally, we will compare DStM to other actual technologies in order to propose the best applications for each one.
For the applications, we will use the PhotoModeler-Scanner© software from EOS Systems©.

2. THEORY AND FRAMEWORKS

For the theoretical aspects of DStM, we will consider the case of a single pair of images. When using more than one stereopair, a bundle adjustment is computed for the whole block, but each couple is matched separately. However, since we generate point clouds, a number of precautions is discussed in the experience part.

2.1 Idea and principles

The equations of stereophotogrammetry allow getting the 3D position of homologous points located on both images. The

* Corresponding author.

main idea is: "For each pixel in a reference image, find the homologous position in the other one." The first idea, the simplest, would be to take the intensity value of one pixel in an image, and to locate in the other one a pixel with the same value. Firstly, we are here in a typical *ill-posed* problem: the solution may not be unique (the ambiguity is caused by several pixels with the same value), may not exist (because of noise), and/or be unstable (because of changing radiometry). Secondly, with high resolution images, a raw matching creates a *combinational* explosion. So, to deal with those two problems, we have to use constraints (Tikhonov, 1963). Kasser and Egels (2001) detail two constraints.

A first constraint consists in rethinking the matching tool. In order to decrease the ambiguity, we consider the neighborhood of the pixel. Thus, we do not work with a single value but with a vector including the values of the pixels located in the matching region. Moreover, the non-existence of the solution implied by the noise has to be considered. We use a correlation score to describe the similarity. The correlation score is usually computed with the scalar product of the two vectors. Finally, we normalize the cross-correlation in order to take into account a global radiometry variation. Now the problem can be considered as a *well-posed* one, since existence, uniqueness and stability have been improved.
The cross correlation is defined by the ratio of the covariance of the two vectors and the products of their standard deviation. Consequently, 1 is the perfect correlation factor.

$$C((i_1, j_1), (i_2, j_2)) = \frac{\overrightarrow{V_1}(i_1, j_1).\overrightarrow{V_2}(i_2, j_2) - \mu_1.\mu_2}{\sigma_1.\sigma_2} \qquad (1)$$

where $C((i_1,j_1), (i_2,j_2))$ = correlation score
$V_n(i_n,j_n)$ = neighbourhood intensity value vector
μ_n = average of V_n
σ_n = standard deviation of V_n

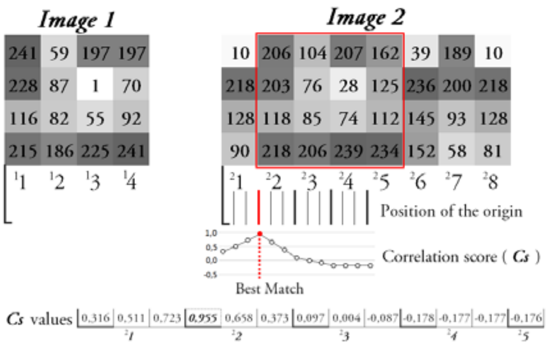

Fig. 1. Subpixellar ($^1/_3$ pixel interval) matching computation

In order to counter the combinational explosion, we have to use a constraint that reduces the number of matching combinations. Epipolar geometry offers one. For each point observed in one image, the same point must be observed in the other image on a known epipolar line. Consequently, the search interval is along a line in the image plane. But this interval can be reduced again; if we have information about the global shape of the object, and define a min/max distance to a medium plane or surface model, the interval has min/max values on the epipolar line. So the problem has become a one dimensional one.

Epipolar geometry can be determined by two approaches from photogrammetry or computer vision. In the first, the position of the two epipoles (projection of the centre of projection of one image in the other) can be computed since the relative orientation is known.

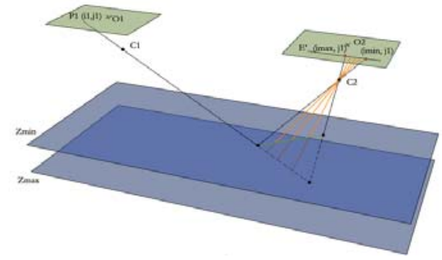

Fig. 2. Epipolar geometry and linear constraint

To solve the epipolar geometry in Computer Vision, a fundamental matrix is used (Hartley and Zisserman 2004). Once epipolar geometry is known, correlation can be done. In order to get aligned and parallel epipolar lines, we apply an epipolar rectification. Rectification determines a transformation of each image plane, such that pairs of conjugate epipolar lines become collinear and parallel to one of the image axes. Two virtual sensors are created, and the matching process is applied to those two new images.

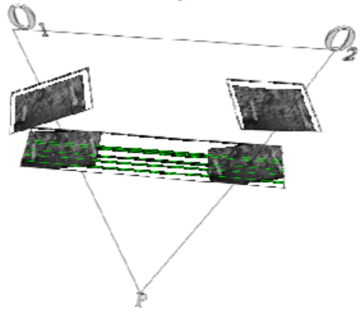

Fig. 3. : Epipolar rectification

Fig. 4. General Dense Stereo Matching framework

2.2 Parameters' influence

Since real cameras differ from a pinhole camera model, error sources have to be located and estimated. We have different kinds of perturbation: the noise, the accuracy of positioning and the accuracy of the visual rendering. Those sources can be classified by either the moment when they are created or when they influence the result. In order to help the user, we list them here by considering their origins.

During the acquisition of the images of the object, the operator has to take into account several parameters.
Firstly, the location and the orientation of the images have to respect a *geometric constraint.* The computation of the relative orientation requires having clear intersections between rays, so the ratio R =base/height has to be greater than $^1/_{20}$. A smaller R would lead to a poor geometry case with increasing errors on depths. For the multi stereopairs case, geometry might need additional oblique images to consolidate the block, following the 3x3 CIPA rules. A comparison of orientation techniques are detailed in (Grussenmeyer and Al Khalil 2002).
Secondly, even if it is obvious in a single couple case, it is really important to ensure an *overlay* for each point that needs to be measured. This is also true for targets or control points.

Thirdly, the *object and its environment* have to be considered. To ensure a correct matching, we need well textured images. Depending on the kind of material (sandstone, wood...), the size of the projection of the pixel on the object has to be adjusted by varying the distance to object and/or the focal length. If texture is not contrasted enough, the correlation score will not have a strong maximum, and the choice of the homologous interval will not be true. The result will be a noisy point cloud. Lights can also be a source of texture: with a light well positioned, the shadows can be really useful (depending on the shape and the granularity). However, the R ratio has to be correctly chosen in order to minimize the deformation of the projection of the objet on the two images (i.e. object on the two images must look identical) in order to be able to process the matching. Generally, R<0.4 is a valid case (Table 1.). Finally, the radiometry does not need to be too different; a moving light source or contrast enhancing or equalization could make the matching impossible.

R values(B/H)	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0
Geometry											
Recovery											
Correlation											
Optimal R											

Table 1. Determination of the optimal R (base/height ratio): darker areas represent favorable value of the parameter; the optimal R for DStM requires the validation of the 3 parameters (Geometry, Recovery and Correlation) at the same time.

During the orientation, epipolar geometry is determined. Thus several parameters have to be defined. Firstly, the number of homologous points has to be sufficient. Moreover redundancy always ensures better quality. The other major parameter is the calibration. Since optics are not ideally made, distortions have to be measured for each one. The quality of the calibration is really important since theoretical epipolar lines become curves through optic distortions (Zhang, 1996). Thus, to ensure a good epipolar rectification, images have to be undistorted.
We notice that all parameters have a final pixel influence since DStM returns a non-scale model. Thus, the size of the object does not matter during DStM process; this is an important point.

2.3 Detailed framework

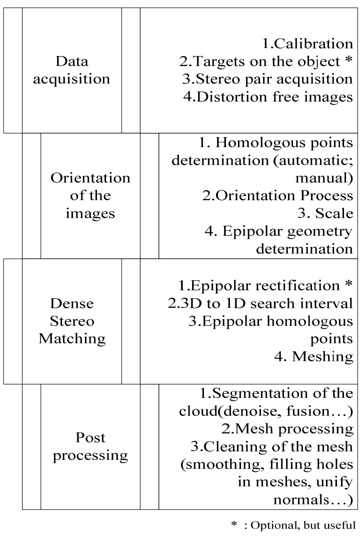

Fig. 5. Detailed Dense Stereo Matching framework

3. EXPERIMENTS

3.1 Presentation of the site and the problematic

The mission of Kilwa[1] brings together specialists in both fields: epigraphy and archaeology. Researches are taking place in a 25km radius zone around the central site.

Fig. 6. Location of Kilwa (Saudi Arabia)
Source: http://maps.google.com

In this central area, architects, draftsmen and archaeologists are working together. Two buildings (settling tanks) partly in ruins are under study. Many other cells are more damaged. Traditionally, the survey of those buildings combines topographical measurements and architectural drawings. But those techniques require several days or weeks, for really partial information. Therefore, a multi stereopair project has been realized on one of them (Fig. 13. Photogrammetric block of the external part of the building). The aim is to get a 3D textured model, with enough details for architectural interpretation afterwards. Transport difficulties, high price of renting and requirement of high quality textures did not allowed the use of a Terrestrial Laser Scanner.

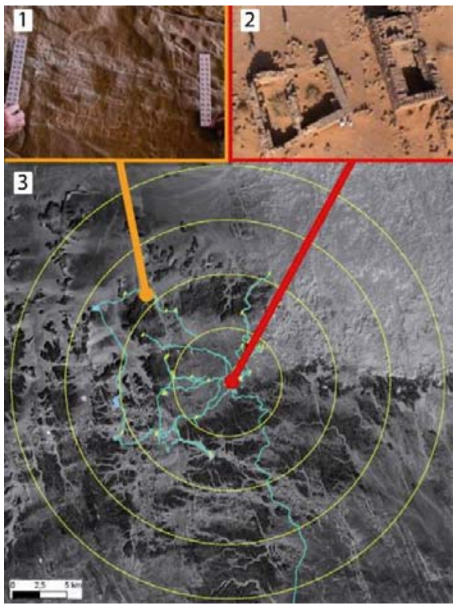

Fig. 7. Prospecting and excavating zones
Sources: 1. Terrestrial Reflex camera image 2. Kite images from Geofil© 3.Quickbird© image.

[1] French archaeological mission of KILWA (Saudi Arabia), directed by Saba FARES, co-financed by the sub-directorate of Human sciences and archaeology, the French ministry of foreign affairs, the Saudi Commission for Tourism and Antiquities and Nancy-2 university (France)

Beside this, epigraphists are prospecting in a large zone to discover, register and analyse carvings of several periods. A GIS is collecting information (period, type, GPS location...) and images. Many of the carvings can only be understood and analyzed by coupling them. This analysis work is mainly done at the office after the field work, since the month on site is used to collect data (Meyer *et al*. 2006). Generally, the carvings are made on a quasi planar stone, and a single couple model offers great information for analysis. The main advantage is that only two images are necessary; thus the survey time is really small, and the creation of a 3D model does not exceed 15 minutes for processing the couple. DStM has been tested on a ten of carvings with a single stereopair for each one.
For the processing, we use PhotoModeler Scanner© (EOS Systems©) with a Canon EOS 5D reflex camera. Used focal lengths for the project are 28, 50 and 85mm.

3.2 Single stereopair applications

First of all, before the acquisition of the images on the site, we need to calibrate the camera (body and optics). The knowledge of the intrinsic parameters of the camera (sensor size and resolution, principal point coordinates, focal length, optical distortions) is directly linked with the accuracy of the final results. In PhotoModeler Scanner© (PMS©), the distortion model is based on a two polynomial model respectively for radial and tangential distortions.

Fig. 8. Example of stereopair of an arabic inscription

Following the theoretical survey detailed in the first part of this paper (Table 1.), it is important to ensure correct values of the parameters during the acquisition of the stereopair: R ratio (base over height), recovery of the targets in both images and lighting. It is also recommended to take a triplet of images, in order to prevent errors during the acquisition.

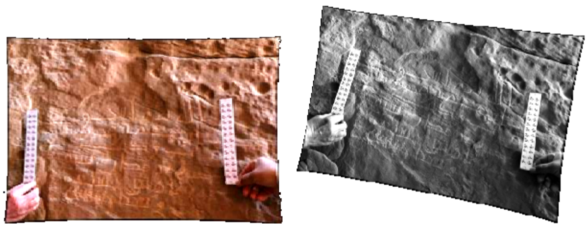

Fig. 9. Undistorted and epipolarly rectified image

Then, the first step of the computation is the orientation of the couple. The best couple for DStM computation is chosen from the position and orientation of the cameras. In this step, epipolar geometry is (implicitly) calculated. Thereafter, we will prepare the data set for matching. Firstly, we create two new images, free of distortions, by applying the inverse distortion equations. Once we have created those new "idealized" images, and corresponding camera, epipolar lines are straight, but not parallel. Consequently, the epipolar rectification is the next step. In PMS©, this epipolar rectification is made during the DStM computation, after the setting of the sampling rate. Two bitmap images are created in the project folder. Then, after adjusting parameters (depth range above and below global model, sub pixel factor, matching region radius, texture type), matching is processed. Both images are used as a reference in order to minimize wrong matching and two disparity maps are computed as bitmap files in the project folder too. Then, 3D points are triangulated and the point cloud is created.
The length of the process, from data acquisition to export, a single stereopair does not generally require more than one hour of computation on recent workstations or laptops; and it only takes 15 minutes for the operator to create the model.

Fig. 10. Final 3D textured model of the inscription

Thereafter, the point cloud can be denoised, segmented, meshed and textured. 3D data can be exported in many standard 3D formats. In this project, we used both OBJ and VRML. OBJ data has been converted into a 3D PDF that allows faster and easier measuring and navigation than VRML explorers (as Cortona© for example)

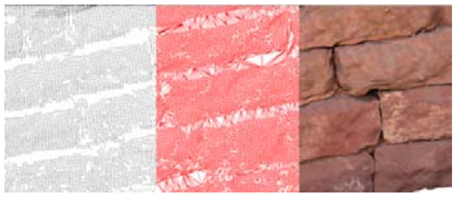

Fig. 11. (a)point cloud (b)wireframe (c)textured model of a wall

3.3 Multi stereopair application

For the buildings, a single point of view does not ensure a total coverage of the object. Thus, we have to multiply the stereopairs in order to model the whole building (Fig. 14). But to compute a bundle adjustment for the block, we need additional images. We followed the 3x3 CIPA rules (http://cipa.icomos.org). Many targets were placed all around the building, and finally, 50 images were used in the project. To scale the block, 20 control points measured by total station have been used.

Fig. 12. View of one of the two settling tanks

After the block orientation, the best stereopairs were chosen to apply DStM separately for each of them. Parameters were set up differently for each couple depending on the size, the illumination and other criteria. After denoising the whole model, we merged all the point clouds via the Merge function of PMS©.

Thereafter, we processed the merged point cloud classically: meshing, filling holes in meshes and smoothing in PMS©.

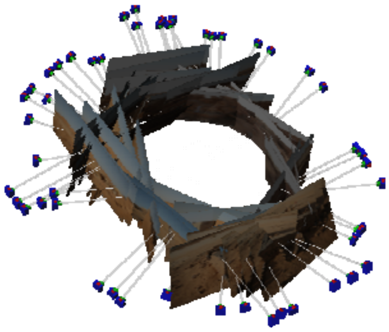

Fig. 13. Photogrammetric block of the external part of the building

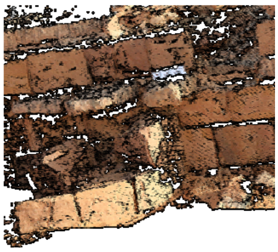

Fig. 14. Detail of the southern wall of the building

It is an important fact, because in most 3D works, several software have to be used for the different part of the process, increasing cost, but also quickly increasing complexity of the workflow.

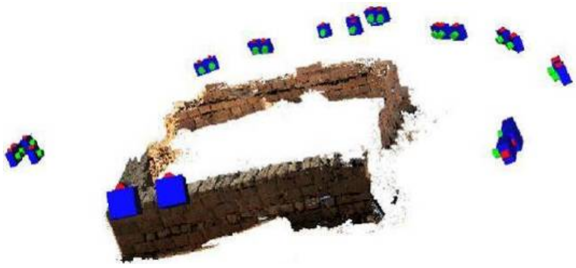

Fig. 15. 3D model of the building

The model has been converted in a 3D PDF file, and then merged with a Digital Elevation Model of the site. Because of the function of the buildings (settling tanks), we wanted to model the hydrometric network, at the site scale, and more precisely at the building scale. Since water resources remain the main problematic in desert areas, this works gives a lot of information about the organization of the monastery and its environment.
For smaller projects, we also used 3 stereopairs to model carvings spread on the ground, before taking them for longer analysis.

Fig. 16. 3D textured model of engraved stones scattered on the ground.

4. INTERESTS AND LIMITATIONS

To look at the situation in an objective way, we can consider two types of reflection. First, the answer to the initial needs (geometry, rendering) needs to be studied. Then, DStM should be compared to other techniques as Terrestrial Laser Scanning or traditional surveying in order to weigh the pros and cons of each one. It is then possible to define the best suited cases for each one of those techniques.

4.1 Criteria for compliance with initial requirements

The first obligation to validate DStM is the accuracy of the geometric model. We have to distinguish the several steps of the process in order to understand the combination of errors. Another distinction needs be done between localization and noise. Firstly, during the orientation, standard errors must be subpixellar in order to ensure good epipolar determination and orientation. To reach this precision, the use of coded targets is strongly recommended, increasing both precision and processing time. About the scaling, we can use control points (for multi stereopairs project, geolocalization) or just distances between targets (for single stereopair, no geolocalization is required). Finally, for the DStM, the noise is the main problem. Lots of parameters can lead to a noisy model: a global model that is not representative, leading to a wrong search interval along epipolar line, a bad radius for a matching region, a wrong epipolar determination due to an inaccurate calibration, a lack of texture, smooth images. If those aspects are fixed, the model can become a lot more precise. A denoise filter enables even better results. For non-noisy parts, the standard deviation of the position of the homologous point on the image does not exceed a couple of pixel. In this project, expected errors decrease as the size of interest zone reduces.

Non differential GPS : 10 m	*Prospection area : 25km radius*
Differential GPS : 0,2m	*Large site area: 1km radius*
Topographical survey, DEM : 7cm	*Site area: 500m radius*
Global Photomodel : 5 pixels # 3cm	*Building: 10m radius*
Stereo couple : 1pixel # 2-3mm	*Wall: 3m radius*

Fig. 17. Maximum expected errors and *approximate area* for each data type

For the analysis of carvings, a really good texture is necessary, and simple images only offer incomplete information. The input of the third dimension is an extraordinary add-on for visualization and simulation. With two strips of targets disposed around the epigraphs, it is almost instantaneous to get a stereo couple. Generally, the process does not exceed half an hour. Then, a 3D model (usually a 3D PDF) can be added to the database and linked to the GIS easily. For epigraphs, that generally does not exceed one meter, we can easily get a resolution of one millimeter with a single stereopair. The great advantage remains the high quality texture that allows degrading the model, without losing too much information on analysis and understanding.

4.2 Comparison to other techniques

Traditional sketches are still a common way to proceed in archaeology. Indeed, materially low cost, this technique allows extracting the information needed. However the use of this technique is really limited to architects. Realistic 3D models offer a lot of possibility for several purposes: simulation, historic rendering. With DStM, we really lose less information on materials, depth, and relief.

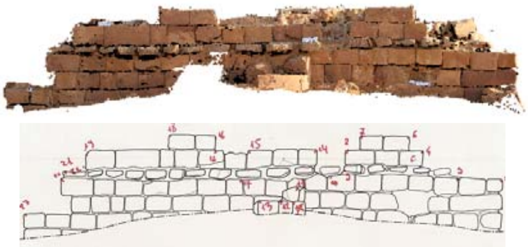

Fig. 18. Top: Meshed and textured point cloud.
Bottom: Architectural man-made sketch of the external face of the wall of the building.

But 3D models are never an end, they only consist the bases for work and surveys in many fields.
On the contrary, laser scanners return an accurate, dense 3D model. Texture information is basically not as great as a reflex camera, but can be done roughly. The drawbacks are mainly the cost, the complexity of transport, of use and post-processing, and the fragility. For those reasons many archaeological missions do not need and /or cannot afford a laser scanning survey. Only large and complex objects would require the use of laser scanners, but the results are often not fully exploited because of heaviness and complexity of data, and the texturing of the 3D model requires oriented high resolution images.

4.3 Kind of works

For many archaeological missions, laser scanner does not respond to the needs: these are too expensive, too fragile, and too complex. Even if Dense Stereo Matching is still complicated for complex cases like large buildings, it offers more than an alternative for many applications.
In this paper, we showed that single stereopairs are really easy to process for a lot of small and medium size objects. Flintstones, pieces of wall, carvings or an excavation can be modeled in less than an hour. Another great advantage compared to other techniques is the flexibility: the size of the object only influences the base over height ratio, so that many objects, from tiny to much larger ones can be modeled in exactly the same way.
For Kilwa's epigraphists, this kind of results was really appreciated because of its flexibility, easiness of working with and analysis. There is no doubt that the diffusion format (*.PDF with 3D content) help them to familiarize with this kind of new data.

5. CONCLUSION

Even if Dense Stereo Matching commercial solutions only appeared a few years ago, they have offered surprisingly good results, from many points of view. Firstly, it is possible to reach high precision on both geometric and rendering sides. Secondly, photogrammetry conserves in DStM all its flexibility and cost advantages, with dense point clouds as output, as only laser scanner techniques offered until now. Finally, the data acquisition does not require a high level of knowledge in photogrammetry, so that with few recommendations, many people would be able to take photographs in the right way. Nevertheless, due to the texture properties of the object, this passive stereo solution cannot yet be applied to some particularly uniform or reflective materials. Passive stereo techniques (pattern or grid projection on the object) make this kind of objects measurable but are difficult to apply outdoor.
Compared to other techniques, it is impossible to say that DStM will fully replace today the laser scanners or the other traditional measurement and representation methods. But all the people interested in the same time by the geometrical and visual properties of an object will find in DStM an easy, cheap, fast and accurate way to document it.
All the advantages and evolutions will certainly help the Cultural Heritage community to integrate this technology, and to take two parallel images instead of one to easily get the 3D dimension.

6. REFERENCES

References from Books:
Hartley, R., Zisserman, A., 2004, Multiple View Geometry in Computer Vision, Cambridge University Press, 2nd edition, pp. 237-360.

Kasser, M., Egels, Y., 2001, Digital Photogrammetry, Taylor & Francis, pp. 168-221.

References from Journals:
Grussenmeyer, P., Al Khalil, O., 2002, Solutions for exterior orientation in photogrammetry, a review, The photogrammetric record, an international journal of photogrammetry, Vol. 17, N°100, October 2002, pp. 615-634.

Meyer, E., Parisel, C., Grussenmeyer, P., Revez, J., Tidafi, T. 2006, A computerized solution for the epigraphic survey in Egyptian Temples, In, Journal of Archaeological Science, Volume 33, Issue 11, November 2006, pp.1605-1616.

References from Other Literature:
Fassi, F., 2007, 3D modeling of complex architecture integrating different techniques – a critical overview. *The Int. Arch. of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Zurich, Vol. XXXVI-5/W47, 11p.

Furukawa, Y., Ponce, J., 2007. Accurate, dense, and robust multi view stereopsis. *In Proc. Int'1 Conf. on Computer Vision and Pattern Recognition, IEEE, 2007,* pp. 1-8.

Remondino, F., Fabio Menna, F., 2008, Image-based surface measurement for close-range heritage documentation, *The Int. Arch.of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Beijing, Vol. XXXVII-B Part B5, pp. 199-206.

Reznicek J. and Pavelka K., 2008, New low-cost 3D scanning techniques for cultural heritage documentation. *The Int. Arch. of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Beijing, Vol. XXXVII-B Part B5, pp. 237-240.

Seitz, S., Curless, B., Diebel, J., Scharstein, D., Szeliski, R., 2006, A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms, *In Proc. Int'l Conf. on Computer Vision and Pattern Recognition*. - IEEE, 2006. Vol. 1, pp. 519-528.

Strecha, C., von Hausen, W., Van Gool, J. L., Fua, P., Thoennessen, U., 2008, On benchmarking camera calibration and multi-view stereo for high resolution imagery. *In Proc. Int'1 Conf. on Computer Vision and Pattern Recognition, IEEE, 2008,* pp. 1-8.

Tikhonov A.N., 1963, Solution of incorrectly formulated problems and the regularization method, *Soviet Math Dokl* 4, pp. 1035-1038.

Zhang Z. ,1996, On the epipolar geometry between two images with lens distortion, *Proc. Int'l Conf. Pattern Recognition*, Vol. 1, pp. 407-411.