Effective High Resolution 3D Geometric Reconstruction of Heritage and Archaeological Sites from Images

S.F. El-Hakim[1], Fabio Remondino[2], Lorenzo Gonzo[3], Francesca Voltolini[3]

[1] National Research Council (NRC), Institute for Information Technology, Canada.
[2] Swiss Federal Institute of Technology (ETH), Institute of Geodesy and Photogrammetry, Switzerland.
[3] Centre for Scientific and Technological Research (ITC-IRST), Italy.

Abstract: Motivated by the need for fast, accurate, and high-resolution approach to document heritage and archaeological sites before they are removed or destroyed, the goal of this paper is to develop and demonstrate advanced image-based techniques to capture the fine 3D geometric details of such sites. The size of the site can be large and of any arbitrary shape which presents a challenge to all existing 3D techniques. Although laser scanners can acquire large number of 3D points, they can be impractical to set up and move around in archaeological sites. Alternatively, image-based techniques entail data from inexpensive portable digital cameras. We present a sequential multi-stage procedure for 3D data capture from images designed to model the fine geometric details. Test results demonstrated the utility and flexibility of the technique and proved that it creates highly detailed models in a reliable manner for many different types of surface detail.

Keywords: 3D Acquisition, 3D Modelling, Visualization, Performance evaluation, Heritage

1 Introduction

A textured 3D geometric model is a highly desirable form of object representation since it gives full geometric documentation and allows unrestricted interactive visualisation and manipulation at a variety of lighting conditions. Creating 3D models of heritage and archaeological objects and sites in their current state requires a technique that is portable due to accessibility problem, low cost due to limited budgets, fast due to the usually short allowed time on the site not to disturb work or visitor, flexible and scalable due to the wide variety and size of sites and objects, and highly accurate. It is also important that the model has dense 3D data on all surfaces to guarantee a realistic experience at close up. Even with rich texture maps, models without fine geometric details or surface irregularities will exhibit too smooth, flat-looking, surfaces and polygonised silhouettes that are easily detected by the human eye. Several sensing technologies are available. Range sensors like laser scanners can capture relatively accurate geometric details, but they remain costly, usually bulky, not easy to use, require stable platform, and influenced by surface properties. They also have limited scalability and flexibility since a range sensor is intended for a specific range and volume, therefore one designed for close range is not suitable for medium or long range. They may acquire millions of points, even on perfectly flat surfaces, often resulting in over-sampling. Nevertheless it is likely that the points really needed for reconstruction, like corners and edged, are not captured. Image-based modelling (IBM) techniques although can produce accurate and realistic-looking models remain highly interactive since current fully automated methods are still unproven in real applications, which limit the amount of details a model can have. A practical consideration is that the number of images taken is not too high particularly for large sites therefore wide baseline between images is expected. This causes a challenge to stereo matching techniques since it reduces the geometric and radiometric similarity between images and causes occlusions. In this paper we propose an image-based approach that requires only limited amount of human interactivity and is capable of capturing the fine geometric details with high accuracy. It can also cope with wide baseline using several advancements over standard stereo matching techniques. When high resolution well calibrated images are used at strong geometric configuration the accuracy will be very high. The approach is sequential starting with a basic model, without fine geometric details, that is created with a small number of interactively measured points. The model is segmented into surface regions or patches. Using this model as a guide, with the known camera parameters, an automatic procedure models the fine details with high-resolution meshes resulting in accurate documentation and photo-realistic visualisation. Three techniques are used, each where best suited: for regularly shaped patches such as planes, cylinders or quadrics, we apply a fast relative stereo matching technique, for more complex or irregular segments with unknown shape we use a global multi-image technique, and for segments unsuited for stereo matching we employ depth from shading (DFS). The remainder of the paper is organized as follows. A review of relevant previous work is given in the next section followed by an outline of the proposed image-based technique. Details of the stereo matching and DFS are given in sections 4 and 5. Examples and accuracy analysis with direct comparison with laser scanning are then presented followed by conclusions.

2 Previous Work

Since a large body of work on 3D modelling exists, we only focus here on the most accepted and practically tested approaches for heritage and archaeological site modelling. We also report only on methods for creating detailed geometric models thus image-based rendering (IBR), which skips geometric modelling, and surveying and CAD techniques, which create sparse models, will not be included.

Debevec, 2003, used a two step image-based approach where the first step is interactive and produces a basic model then dense stereo matching is used in the second step to add details. The first step uses assumptions about the surface shape to find the camera positions, thus it is more suited for regular architecture structures than archaeological remains. Also the accuracy of determining the camera poses is as good as the shape assumptions are valid, thus upsetting the stereo matching which relies on camera locations. Pollefeys et al., 2003, also used a two steps approach where in the first step the internal and external camera parameters of short baseline image sequence, along with sparse 3D points on extracted features, are computed automatically. This information is then used to constrain dense stereo matching in the second step to add fine details. The drawback is that the first step may produce inaccurate data, either for the camera calibration and orientation, or for the 3D points, thus creating less accurate dense stereo matching in the next step. The requirement of large number of images on the archaeological site to achieve success in the fully automated procedure may be a problem on large complex sites.

Stereo matching techniques have been in development for several decades. Some performance evaluations exist [more recently; Scharstein, and Szeliski. 2002, Sietz et al, 2006, Mayoral et al, 2006] however the test objects used for the evaluations are usually small and in lab environment which may not scale well to large complex sites. Banks and Corke, 2001, compared different similarity measures for dense stereo matching techniques. Some wide base-line techniques with heritage application examples have been developed [Strecha et al, 2004]. The technique requires a set of sparse initial depth estimate obtained from viewpoint-invariant features. However more performance and accuracy evaluation is needed for large complex objects. In general stereo techniques can be divided into two categories: the standard local window-based approaches that select the disparity which gives minimum difference (winner take all), and the global approaches that actually solve for the disparity, along with other parameters such as shape and illumination differences, using minimization methods such as least squares [Remondino and Zhang, 2006] and graph cuts [Boykov et al, 2001]. Fassold et al., 2004, start from an initial model obtained by stereo then refine the model and add details with shape from shading.

Laser scanning alone was used in many projects [for example Godin et al 2002 and Allen et al, 2004], while laser scanning and image-based techniques were combined in several others. Mueller et al 2004 used automated shape from video on large parts of the site while a structured light system is employed on small artefacts and statues in the 3D MURALE project. Alshawabkeh and Haala, 2004 apply laser scanning to most of the site and complete the details on parts not covered by the scans, due to occlusions and inaccessibility, with image-based techniques. El-Hakim et al 2004 used an opposite approach - they created the main models from image-based techniques and added some details from laser scanning.

3 Synopsis of the Approach

First, the overall steps are outlined then we focus on the steps which need more clarification. Primarily it is a stepwise procedure where the main steps are:

1- Camera interior calibration at certain camera setting
2- Image capture with the same camera setting as for calibration
3- Image selection, ordering and pre-processing if necessary
4- Interactive feature extraction for image pose/orientation determination
5- Scene or surface segmentation and initial model construction
6- Dense stereo/multi-image matching or depth from shading on each segment of the initial model
7- Creation of triangular mesh from the ensuing 3D point cloud
8- Texture mapping together with texture geometric and radiometric correction
9- Rendering and interactive visualization using techniques to handle very large data

Some of the above steps are performed because in our experience best results are obtained when:

- Camera is pre-calibrated. This allows more accurate and complete calibration of all parameters including lens distortion which is important to the success of all subsequent operation. It also allows for less restrictive image acquisition than required for self-calibration.
- User segment the scene to remove unwanted regions, such as background, and divide the object or site into regions to improve the matching and the modelling process.
- Operator interactivity is used to extract reference or seed points to be used for image orientation and establish constraints for dense stereo matching. This again allows less restrictive image acquisition by elimination the need for short baseline and gives more precise and reliable results.

Many [Schindler et al, 2003, and Schouteden et al, 2001] follow this process in practice rather than attempt full automation with un-calibrated camera. Scene segmentation has been also proposed by others [Medioni and

Nevatia, 1985, Hong and Chen, 2004]. Segmentation reduces processing time and helps in the modelling step [Zeng et al, 2007] since regardless of the object size and complexity, segmenting it into smaller and simpler segments makes it possible to model any site. The segmentation in our case achieves these objectives:

- Restricts the search region to the segment and allows effective constraints
- Smaller discontinuity within the segment thus reduces the matching problems at the boundaries.
- Handling of occlusions since they are mainly at the boundaries while within a segment there is only small self occlusion. Large occlusions are handled by selecting the best images for a segment.

The fourth and fifth steps are performed interactively (about ten mouse clicks per image) to accurately obtain the camera parameters and segment the scene into separate surfaces. This creates a basic approximate surface model, but the feature points measured are very accurate. We are working on increasing the level of automation in those steps. The sixth step, which in fact is combined with step 7 since we directly create a triangular mesh, is to automatically add fine details on each surface segment. Fine details are added with three different approaches, each where best suited:

1- A fast stereo matching approach that does not require manually measured points. But since it is uses only two images and a local matching approach, it may not handle large variations from the basic model.

2- The second matching approach is based on multi-images and global optimization with least squares and does not require approximate shape, but requires additional seed points at surface significant changes.

3- Depth from shading is applied on parts where stereo matching does not work well, mainly parts with none or repeated textures, and is applied on single image basis. It is designed to compute directly the depth variation from the basic segment shape. It does not provide high accuracy since it is an interpolative method but it will enhance the appearance of the parts that it applies to, such rocks, petroglyphs, or bricks.

4 Creating the Initial Model

We create basic models of surface elements such as planer walls, cylindrical shapes like columns, arches, doors, and windows using an approach initially developed in [El-Hakim, 2002]. For example, a column is automatically constructed from four seed points, two on the corner of the top crown and two on the corners of its base. From these points, the radius of the column and direction of its axis can be computed. The ratio between the upper and the lower circle is usually 0.85. 3D points on top and bottom circles of the column are automatically added. For arches, first a plane is fitted to seed points on the wall. An edge detector is applied to the region and points at constant interval along the arch are automatically sampled. The image coordinates of these points in one image, the known image parameters, and the plane parameters, are used to create the 3D coordinates. For windows and doors we need four outside corner points and one point on the inside surface. By fitting a plane to the corner points and a plane parallel to it at the surface point, the complete window or door is created.

5 Modelling the Geometric Details

Using the initial sparse model as a guide and knowing camera calibration and orientation parameters, we developed an automatic procedure to model fine details with high-resolution meshes to achieve accurate documentation and photo-realistic visualisation. Three techniques are used, each where best suited:

1. For patches with regular shape, fit an implicit function (e.g. plane, cylinder or quadric) using seed points and apply a relative stereo matching technique.
2. For irregular patches with unknown approximate function use an absolute multi-image matching technique.
3. For patches unsuited for stereo matching (e.g. un-textured), apply DFS.

5.1 Dense Stereo/Multi-Image Matching

Occlusions, lack of texture, and light variations between images are persistent problems for stereo matching especially with widely separated views. Template-based stereo matching works best when sufficient texture variations or localised features are present on the surface. Therefore, the first rule we use to select the areas where stereo matching will apply is intensity-level analysis of the template window. This includes mean, standard deviation, and second derivative of the grey-levels of the pixels in the window. If those are higher than preset thresholds, the stereo matching will proceed otherwise we consider the region to be too uniform for stereo matching and switch to DFS, which works best on smoothly shaded surfaces.

The relative stereo matching approach reduces the problems by using the basic model to narrow the search for matching. The procedure is as follows for each segment with known fitted function:

- A high-resolution approximate mesh of triangulated 3D points, which can be as dense as one vertex per pixel, is placed automatically on each segment according to its fitted function.
- The coordinates of the approximate mesh from the basic model are replaced with the final coordinates from the stereo matching. In fact the technique computes only the correction to the points.

Figure 1: Stereo matching with search constraints.

The stereo matching is based on minimizing the normalised squares of the difference between the template and the search window. The search is done along the epipolar line and we also limit the search to a disparity range computed from the basic model. For example in Figure 1, point P_1 in the template image has a corresponding point P_2 in the search image that is computed directly from the basic model. Based on maximum depth variation (roughly preset), we can easily compute the region on the epipolar line (distance d) where we restrict the search. The window in the search image is re-sampled to take into account the difference in orientation between the two images and surface orientation of the basic model. This accounts for the geometric variations between these two images and gives accurate and reliable results. We now apply another rule to accept or reject the matched points. If the best-matched window differs from the template by more than a threshold (pre-set based on light variation between the two images), the matching is considered unreliable or invalid (e.g. the region is occluded in the right image) and the system reverts to the basic model point, which is point P_2.

The relative stereo matching approach, although fast and effective, requires an approximate surface shape. However, for irregular surfaces like archaeological finds and sculptures, the approximate shape is unknown. Therefore, an extended, albeit slower, more global approach that does not require knowledge of an approximate surface has been developed. It is based on non-linear least-squares estimation that solves for several parameters including the matched pixel location and photometric images differences [Remondino and Zhang, 2006]. It uses more than two images to increase its precision and reliability by simultaneously matching the point in all the images it appears in. It is a coarse-to-fine hierarchical solution with automatic quality control. The approach performs three mutually connected steps:
Image pre-processing: the set of available images is processed with an adaptive smoothing filter in order to reduce the effects of the radiometric problems such as strong bright and dark regions and optimizes the images for subsequent feature extraction and image matching. Also image pyramids are generated to have several versions of the image with progressive spatial resolutions.
Multiple Primitive Multi-Image (MPM) matching: we utilize a coarse-to-fine hierarchical strategy for accurate and robust surface reconstruction. Starting from the low-density features in the lowest resolution level of the image pyramids, the MPM matching is performed with two or more images, incorporating multiple matching primitives (feature, edge, and grid points). Feature points are suitable to generate accurate surface models but they suffer from noise, occlusions, and discontinuities. Edges generate coarser but more stable models as they have higher semantic information and are more tolerant to noise. The MPM performs three operations in each pyramid level: (i) features and edges extraction and matching, (ii) integration of matching primitives and (iii) initial mesh generation. Within the pyramid levels, the matching is performed with an extension of the standard cross-correlation technique while only in the last (original) level a multi-photo geometrically constrained LSM is performed. The multi-image matching is guided from object space and allows reconstruction of 3D objects from all available images simultaneously. Moreover, at each pyramid level, a triangular mesh is reconstructed from the matched features. The mesh is used in the subsequent pyramid level for derivation of approximations and adaptive computation or self-tuning of the matching parameters.
Refined matching: Multi-photo geometrically constrained matching and least squares B-Spline snakes are used to achieve potentially sub-pixel accuracy matches and identify some inaccurate and possibly false matches. This is applied only at the original image resolution level. The surface derived from the previous MPM step provides well enough approximations for the two matching methods and increases the convergence rate.

5.2 Depth from Shading

DFS is applied where grey-level variations are not adequate for stereo matching and sections appearing only in a single image. Standard shape from shading techniques, which compute surface normal, lacked success in actual applications due to its ill-posed formulation that requires unrealistic assumptions, such as the camera looks orthogonally at a Lambertian surface and there is only one single light source located at infinity [Zhang et al, 1999]. Our approach computes the depth directly, rather than surface normal. It is applied to a work image: a grey-level version of the original with some pre-processing such as noise removal filtering and editing of unwanted shades. Using known depth and grey level at 8-10 points determined interactively we form a curve describing the relation between grey-levels and depth variation from the basic model (Figure 2). The curve intersects the grey-level axis at the average intensity value of points actually falling on the basic model. By adjusting the curve the results can be instantly reviewed. We adjust the coordinates of the grid points on the surface of the basic model segment according to shading using this curve. We now have a triangulated grid of points whose coordinates are altered from the initial basic model to account for the fine details.

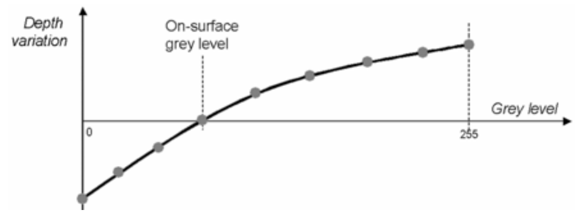
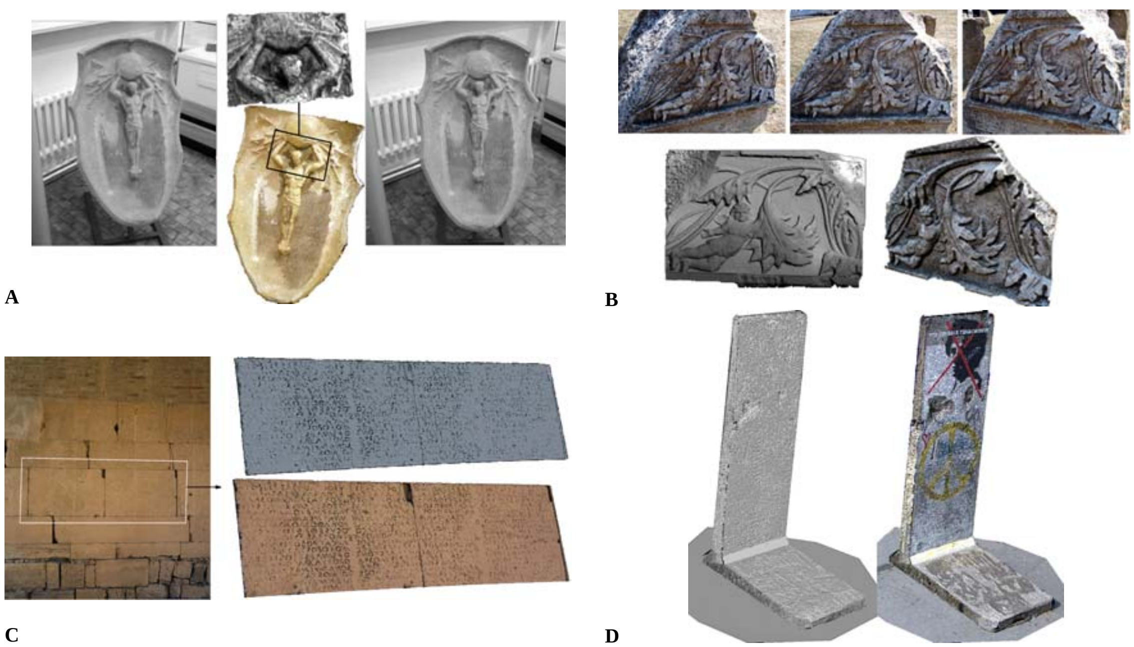


Figure 2: Grey-level versus depth variation relative to basic model.

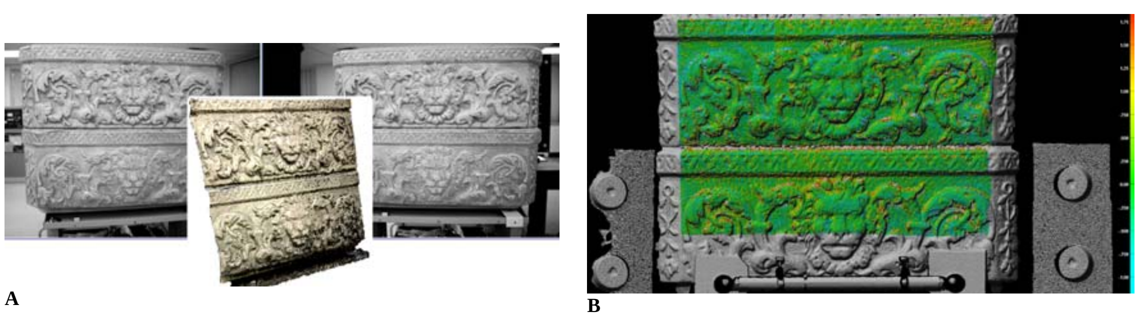
6 Examples and Performance Evaluation

We extensively tested our approach on hundreds of real data of different types to assess its effectiveness under real application conditions (examples are shown in figure 3). We also tested our matching approach on several test objects, one of which is in a lab environment shown in Figure 4a. For ground truth, the objects were scanned with two laser scanners: Surphaser© HS25X (0.48 mm accuracy) and ShapeGrabber© 502 (0.42 mm accuracy). The same objects were then modelled with both matching techniques described in 5.1. To compare these models with ground truth data, we used PolyWorks© Inspector software. Colour-coded results are illustrated in Figure 5b. The standard deviation of the differences between the scanned model and the image-based model was, on average, 0.54mm (Surphaser) and 0.52mm (ShapeGrabber) for all data sets and both matching techniques.



A B C D

Figure 4: Examples of stereo matching (A, B), DFS (C), and both on different segments (D)



A B

Figure 4: Lab test object (A) and colour-coded difference between scanned model and image-based model (B).

7 Conclusions

We presented a sequential segment-based approach that creates detailed models of any shape starting from an interactively-created basic model of the whole scene then automatically adds fine geometric details using two dense stereo matching techniques and depth from shading, each where best suited. It uses logical and easy to establish constraints to make it effective and well-posed. Extensive testing on various types of sites and objects proved it very effective. The resulting accuracy is comparable to highly precise close-range laser scanners but at a fraction of the cost and time.

8 References

Allen P., Feiner, S., Troccoli, A., Benko, H., Ishak, E., Smith, B., 2004, Seeing into the Past: Creating a 3D Modeling Pipeline for Archaeological Visualization, 3D Data Processing, Visualization and Transmission Symposium (3DPVT'04), Thessalonika, Greece.

Alshawabkeh Y. and Haala, N., 2004, Integration of Digital Photogrammetry and Laser Scanning for heritage Documentation, XXth ISPRS Congress, commission V proc., pp. 424-429.

Banks, J., Corke, P., Quantitative evaluation of matching methods and validity measures for stereo vision, Int. J. of Robotics Research, 20(7), 2001, pp. 512-532.

Boykov, Y., Veksler, O., Zabih, R., 2001, Fast approximate energy minimization via graph cuts, IEEE Trans. PAMI, 23(11), pp. 1222-1239.

Debevec, P.E., 2003, Image-based techniques for digitizing environments and artifacts, 4th Int. Conf. 3D Imaging and Modeling (3DIM'03), Banff, Canada, pp. 234-241.

El-Hakim, S.F., 2002. Semi-automatic 3d reconstruction of occluded and unmarked surfaces from widely separated views, Proc. ISPRS Commission V Symposium, Corfu, Greece, Sept, Vol. XXXIV, Part 5, pp. 143-148.

El-Hakim, S.F. Beraldin, Picard, M., and Godin, G., 2004 Detailed 3D reconstruction of large-scale heritage sites with integrated techniques, IEEE Computer Graphics & Applications, 23(3), pp. 21-29.

El-Hakim, S.F., 2006, A sequential approach to capture fine geometric details from images, ISPRS Commission V Symposium, Image Engineering and Vision Metrology, Dresden, Germany, IAPRS&SIS Vol. XXXVI, Part 5, pp. 97-102.

Fassold, H., Danzl, R., Schindler, K., Bischof, H., 2004, reconstruction of archaeological finds using shape from stereo and shape from shading, 9th Computer Vision Winter Workshop (CVWW), Piran, Feb., Slovenia, pp. 21-30.

Godin, G., Beraldin, J.-A., Taylor, J., Cournoyer, L., Rioux, M., El-Hakim, S., Baribeau, R., Blais, F., Boulanger, P., Picard, M., Domey, J., 2002, Active Optical 3-D Imaging for Heritage Applications. IEEE Computer Graphics & Applications, 22(5), pp. 24-36.

Hong, L., Chen, G., 2004, Segment-based stereo matching using graph cuts, CVPR 2004, Vol. 1, pp. 74-81.

Pollefeys, M., Van Gool, L., Vergauwen, M., Cornelis, K., Verbiest, F., Tops, J., 2003, Image-based 3D Recording for Archaeological Field Work, IEEE Computer Graphics & Applications, 23(3), pp. 20-27.

Mayoral, R., Lera, G., Perez-Ilzarbe, M.J., 2006, Evaluation of correspondence errors for stereo, Image and Vision Computing, 24, pp. 1288-1300.

Medioni, G., Nivatia, R., 1985, Segment-based stereo matching, Computer Vision, Graphics and Image Processing, 31(1), pp. 2-18.

Mueller, P., Vereenooghe, T., Vergauwen, M., Van Gool, L., Waelkens, M., 2004, Photo-realistic and detailed 3D modelling: the Antonine nymphaeum at Sagalassos (Turkey), Computer Applications and Quantitative Methods in Archaeology (CAA2004), pp. 221-230.

Ogale, A., Aloimonos, Y., 2005, Shape and the stereo correspondence problem, Int. J. Computer Vision, 65(3), pp. 147-162.

Remondino, F, Zhang, L., 2006, Surface Reconstruction Algorithms for Detailed Close-Range Object Modelling, ISPRS Commission III Symposium, Bonn, 20-22 Sept, 2006, IAPRS&SIS, Vol. XXXVI, Part 3, pp. 117-123.

Scharstein, D. and Szeliski R. 2002, A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. Int. Journal of Computer Vision, 47(1-3), pp. 7-42.

Schindler, K., Grabner, F., Leberl, F., 2003, Fast on-site reconstruction and visualization of archaeological finds, CIPA Symposium, Antalya, Turkey.

Schouteden, J., Pollefeys, M., Vergauwen, M., Van Gool, L., 2001, Image-based 3D acquisition tool for architectural conservation, CIPA 2001, Potsdam, Germany, pp. 381-388.

Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R., 2006, A Comparison and evaluation of multi-view stereo reconstruction algorithms. IEEE Computer Vision and Pattern Recognition (CVPR 2006), N.Y., Vol. 1, pp. 519-528.

Strecha, C., Fransens, R., Van Gool, L., Wide-baseline stereo from multiple views: a probabilistic account, 2004, IEEE Computer Vision and Pattern Recognition (CVPR 2004), Washington, D.C., Vol. 1, pp. 552-559.

Zhang, R., Tsai, P.-S., Cryer, J.E., Shah, M, 1999, Shape from shading: A survey. IEEE Trans. PAMI, 21(8), 690–706.

Zeng, G., Paris, S., Quan, L., Sillion, F., 2007, Accurate and scalable surface representation and reconstruction from images, IEEE Trans. PAMI, 29(1), pp.141-158.