Sabry F. El-Hakim – Fabio Remondino – Lorenzo Gonzo – Francesca Voltolini

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

Abstract: Motivated by the need for a fast, accurate, and high-resolution approach to documenting heritage and archaeological objects 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 objects. The size of the object may be large and of any arbitrary shape which presents a challenge to all existing 3D techniques. Although range sensors can directly acquire high resolution 3D points, they can be costly and impractical to set up and move around archaeological sites. Alternatively, image-based techniques acquire data from inexpensive portable digital cameras. We present a sequential multi-stage procedure for 3D data capture from images designed to model fine geometric details. Test results demonstrate the utility and flexibility of the technique and prove that it creates highly detailed models in a reliable manner for many different types of surface detail.

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 from any view point and at a variety of lighting conditions. Creating such 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 short time typically allowed on the site to avoid disturbing work or visitors; flexible and scalable due to the wide variety of shapes and sizes of objects and sites; and highly accurate for documentation purposes. It is also essential that, in addition to rich texture maps, the model should contain dense 3D data on all surfaces to guarantee a realistic experience even close up. Models without fine geometric details including surface irregularities will exhibit overly smooth, flat-looking surfaces and polygonised silhouettes that are easily detected by the human eye.

Several sensing technologies exist. Range sensors like laser scanners can capture accurate geometric details, but they remain costly, difficult to use, bulky, require a stable platform, and the results are 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 and yet it is likely that points needed for reconstruction, like corners and edges, are not captured. Image-based modelling (IBM) techniques can produce accurate and realistic-looking models but they require a high level of user interaction which limits the amount of details a model can have. Fully automated methods are still unproven in real applications and require a large number of closely spaced images, which is impractical on most sites. In this paper, we propose an image-based approach that requires only a limited amount of human interaction and is capable of capturing fine geometric details with high accuracy. It can also cope with a wider image baseline than fully automated techniques. It relies on high resolution, well calibrated images in a strong geometric configuration to guarantee high geometric accuracy. The approach is sequential starting with a basic sparse model created with a small number of manually measured points. The model is segmented into surface patches and then acts as a guide for an automatic procedure to add the fine details. Three techniques are used, depending on the area in question: 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 to stereo matching we employ depth from shading (DFS).

Previous Work

Since a large body of work on 3D modelling exists, we only focus here on the most accepted and tested approaches to modelling heritage and archaeological objects and sites. 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 considered.

Debevec used a two step image-based approach; the first step is manual and produces a basic model then in the second step dense stereo matching is used to add details (DEBEVEC 2003). The first step makes assumptions about the surface shape to determine the camera positions, and thus is more suited for regular architecture structures than archaeological remains. The accuracy of determining the camera positions is only as good as the shape assumptions are valid, which affects stereo matching since it relies on the camera positions. Pollefeys et al. also used a two-step approach where in the first step the internal and external camera parameters of a short baseline image sequence, along with sparse 3D points on extracted features, are computed automatically (POLLEFEYS ET AL. 2003). This information is then used in the second step to constrain dense stereo matching 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 need for many closely-spaced images in an archaeological site to successfully employ the fully automated procedure may be a problem on large complex sites.

Window-based dense stereo techniques have been in development for several decades. They can be divided into two categories: (1) standard local approaches that select the disparity which gives the minimum difference between a window and a template (winner takes all), and (2) 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 / ZHANG 2006) and graph cuts (BOYKOV / VEKSLER / ZABIH 2001). Some performance evaluations exist (more recently SCHARSTEIN / SZELISKI 2002; SIETZ ET AL. 2006; MAYORAL / LERA / PEREZ-ILZARBE 2006), but the test objects used for the evaluations are usually small and in a laboratory environment, which may not transfer well to large, complex sites. Banks and Corke compared different similarity measures for dense stereo matching techniques (BANKS / CORKE 2001). A wide base-line matching technique, used for heritage applications, has been also developed (STRECHA / FRANSENS / VAN GOOL 2004). The technique requires a set of sparse initial depth estimates obtained from viewpoint-invariant features. However, more performance and accuracy evaluation is needed for large complex objects. Fassold et al. started from an initial model obtained by stereo then refined the model and add details with shape from shading (FASSOLD ET AL. 2004).

Range sensors, such as laser scanners, were used on their own in many projects (e.g. GODIN ET AL. 2002; ALLEN ET AL. 2004), while both range- and image-based techniques were combined in several others. Mueller et al. used an automated shape-from-video technique on large parts of the site while a structured light system was employed on small artefacts and statues in the 3D MURALE project (MUELLER ET AL. 2004). Alshawabkeh and Haala applied laser scanning to most parts of a site and completed the details on other parts not covered by the scans, due to occlusions and inaccessibility, with image-based techniques (ALSHAWABKEH / HAALA 2004). El-Hakim et al. took the opposite approach — they created the main models from image-based techniques and added some details from laser scanning (EL-HAKIM ET AL. 2004).

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 *(Fig. 1)*:

- Camera interior calibration at specific camera settings
- Image capture with the same camera settings as for calibration
- Image selection, ordering and pre-processing, if necessary
- Manual feature extraction for image position/orientation determination
- Scene or surface segmentation and initial model construction
- Dense stereo/multi-image matching or depth from shading on each segment of the initial model
- Creation of triangular mesh from the ensuing 3D point cloud

- Texture mapping together with texture geometric and radiometric correction
- Rendering and interactive visualization using techniques to handle very large data

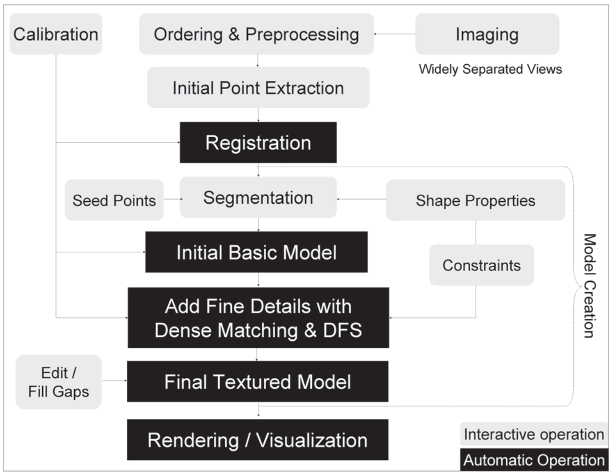

Fig. 1. The overall procedure – dark blocks indicate fully automatic operations.

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

- the camera is pre-calibrated for its internal parameters. This allows a more accurate estimation of all the parameters including lens distortion which is important to the success of all subsequent operations. It also allows for less restrictive image acquisition compared to on-site self-calibration. Pre-calibration is thus recommended for practical projects (Schindler / Grabner / Leberl 2003; Schouteden et al. 2001).
- the user manually extracts reference or seed points to be used for image orientation and to establish constraints for dense stereo matching. This means a much smaller number of images is needed since full automation of both image orientation and stereo matching requires a short baseline between images. It also gives more precise and reliable results without a significant increase in the overall project length.
- the user segments the scene to remove unwanted regions such as background, and divides the object or site into bounded regions to improve the matching and the modelling process. Scene segmentation has been also proposed by others (Medioni / Nivatia 1985; Hong / Chen 2004). Segmentation reduces processing time and helps during modelling regardless of the object size and complexity (Zeng et al. 2007). Segmentation in our case is used to:
  - restrict search regions to a single segment and allow effective constraints;
  - give smaller discontinuity within the segment, thus reducing matching problems at the boundaries;
  - handle occlusions, since they are mainly at the boundaries, while within a segment there is only rarely self-occlusion. Larger occlusions are handled by selecting the best images for a segment.

Feature extraction for image orientation and the segmentation are executed manually with about ten mouse clicks per image. This creates a basic or approximate surface model. Image matching and DFS automatically add fine geometric details on each surface segment using three distinct approaches depending on the nature of each segment:

- A *fast stereo matching* approach that does not require manually measured points. As it uses only two images and a local matching approach, it may not handle large depth variations from the basic model.
- A *multi-image matching* approach, based on global optimization with least squares. It has no restriction on the object's shape but it requires some seed points at surface discontinuities.
- *Depth from shading* is applied on single images in areas where image matching does not work well, mainly untextured areas and those with repetitive textures. It is designed to compute the depth variation directly from the basic segment shape. DFS will particularly enhance the appearance of parts that do not require a very high geometric accuracy, such as rocks, bricks or petroglyphs.

Creating the Initial Model

We create basic models of surface elements such as planar walls, cylindrical shapes like columns, arches, doors, and windows using an approach initially developed by 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 the top and bottom circles of the column are then automatically added. For arches, first a plane is fitted to seed points on the wall. An edge detector is applied to the region to automatically sample points at constant intervals along the arch edge. The image coordinates of these points in one image, the known image parameters, and the plane parameters, are used to compute 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 inside surface point, the complete window or door is created.

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. As mentioned in section 3, image matching and DFS are used:

- For patches with a regular shape, an implicit function (e.g. plane, cylinder or quadric) is fitted using seed points and a relative stereo matching technique applied.
- For irregular patches with unknown approximate functions, an absolute multi-image matching technique is used.
- For patches unsuited for stereo matching (e.g. untextured), DFS is applied.

Dense Stereo/Multi-Image Matching

Occlusions, lack of texture, and light variations between images are persistent problems, especially with widely separated views. Dense stereo matching works best when sufficient texture variations or localised features are present on the surface. Therefore, we first analyse the intensity-level of the template window to select the areas where stereo matching will apply. This includes the mean, standard deviation, and second derivative of the grey-levels of the pixels in the window. If these 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 for a segment with known fitted function is as follows:

- 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. Only the corrections to the approximate 3D coordinates are computed.

The stereo matching is based on minimization of the normalised squares of the difference between a template and a 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 *Fig. 2*, point P1 in the template image has a corresponding point P2 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 limit 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. If the best-matched window differs from the template by more than a predetermined

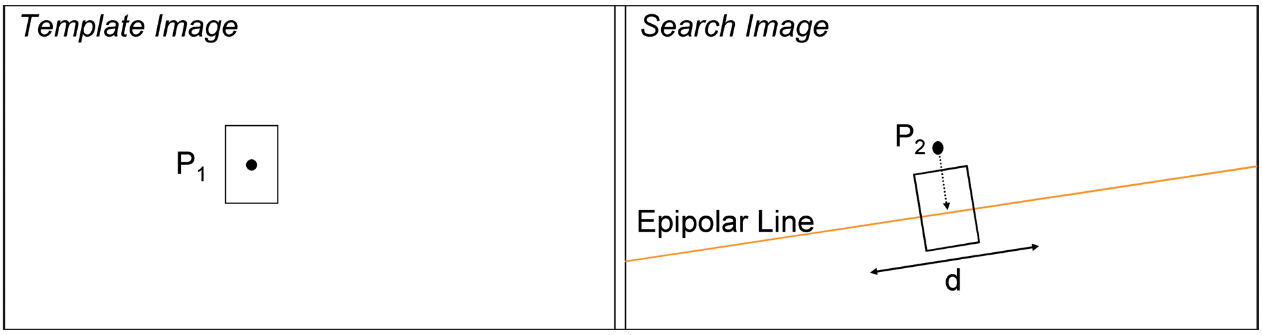

Fig. 2. Stereo matching with search constraints.

threshold (set based on light variation between the two images), the matching is considered unreliable (i.e. the region is occluded in the right image) and the system reverts to the basic model point, which is point P2.

The relative stereo matching approach, although fast and effective, requires an approximate surface shape. However, for irregular surfaces like many archaeological finds and sculptures, the approximate shape is unknown. Therefore, a more global, albeit slower 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 the photometric differences between images (Remondino / Zhang 2006). It uses more than two images simultaneously to increase its precision and reliability by matching the point in all the images in which it appears. 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 optimises the images for subsequent feature extraction and image matching. Image pyramids are also generated with several versions of the image in progressive spatial resolutions.
- *Multiple Primitive Multi-Image (MPM) matching:* we utilise 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 able 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 at each pyramid level: (i) features and edges are extracted and matched, (ii) matching primitives are integrated, and (iii) an initial mesh is generated. Within the pyramid levels, the matching is performed using an extension of the standard cross-correlation technique, while in the last (original) level only, a multi-photo geometrically constrained LSM is performed. The multi-image matching is guided by the object space and allows the 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 sufficiently accurate approximations for the two matching methods and increases the convergence rate.

Depth from Shading

DFS is applied where grey-level variations are not adequate for stereo matching and where sections appear only in a single image. Standard shape from shading techniques, which calculate the surface normal, were found to be inadequate in actual applications. Many unrealistic assumptions had to be made to come to a satisfactory conclusion, such as the camera looking orthogonally at a Lambertian surface or there only being one single light source located at infinity (Zhang et al. 1999). Our approach computes the depth directly, rather than the surface normal. It is applied to a work image: a greyscale version of the original with some pre-processing such as noise filtering and editing of unwanted shades. Using known depth and grey level at 8–10 points determined manually, we form a curve describing the relationship between the grey-levels and the depth variation from the basic model *(Fig. 3)*. The curve intersects the grey-level axis at the average intensity value of points actually falling on the basic model. We adjust the coordinates of a dense grid of points placed on the surface of the basic model segment according to shading using this curve. Our software is designed to instantly review the 3D shape details when the curve is created or manually modified. We now have a triangulated grid of points whose coordinates are altered from the initial basic model to account for the fine details.

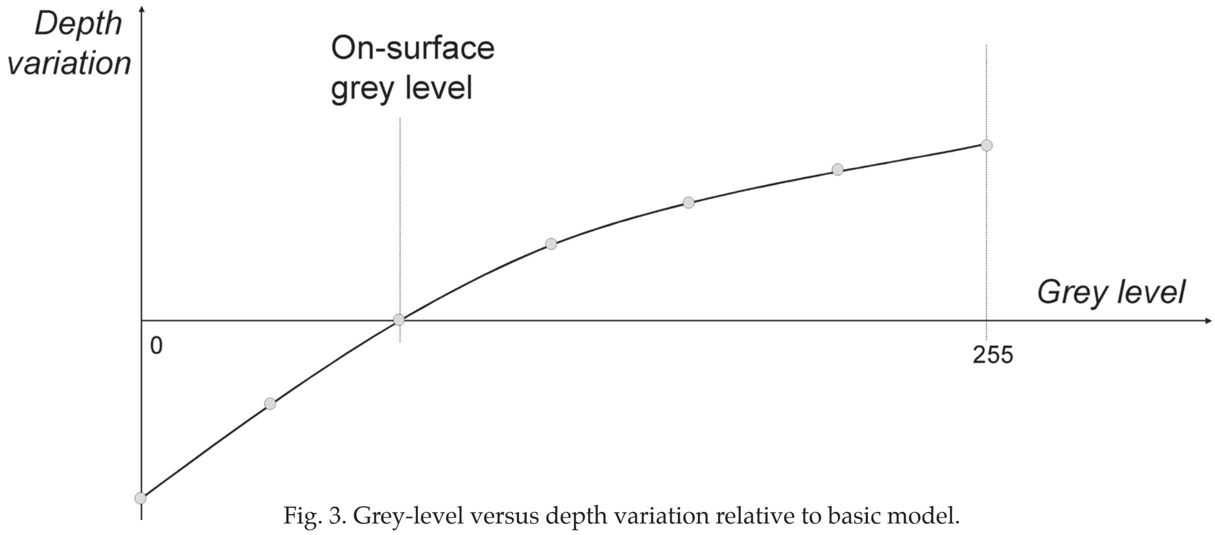

Fig. 3. Grey-level versus depth variation relative to basic model.

Examples and Performance Evaluation

We extensively tested our approach on hundreds of real artefacts of different types to assess its effectiveness under real application conditions (four examples are shown in *Fig. 4).*

We also performed a quantitative evaluation of the accuracy of our matching approach using several test objects in a controlled lab environment. One test object is shown in *Fig. 5a*. The lab allowed us to compare the results with ground truth under the same measurement conditions. For ground truth, the objects were scanned with two highly accurate close-range 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 the section 'Dense Stereo / Multi-Image Matching'. To compare these models with ground truth data, we used PolyWorks© Inspector software. The

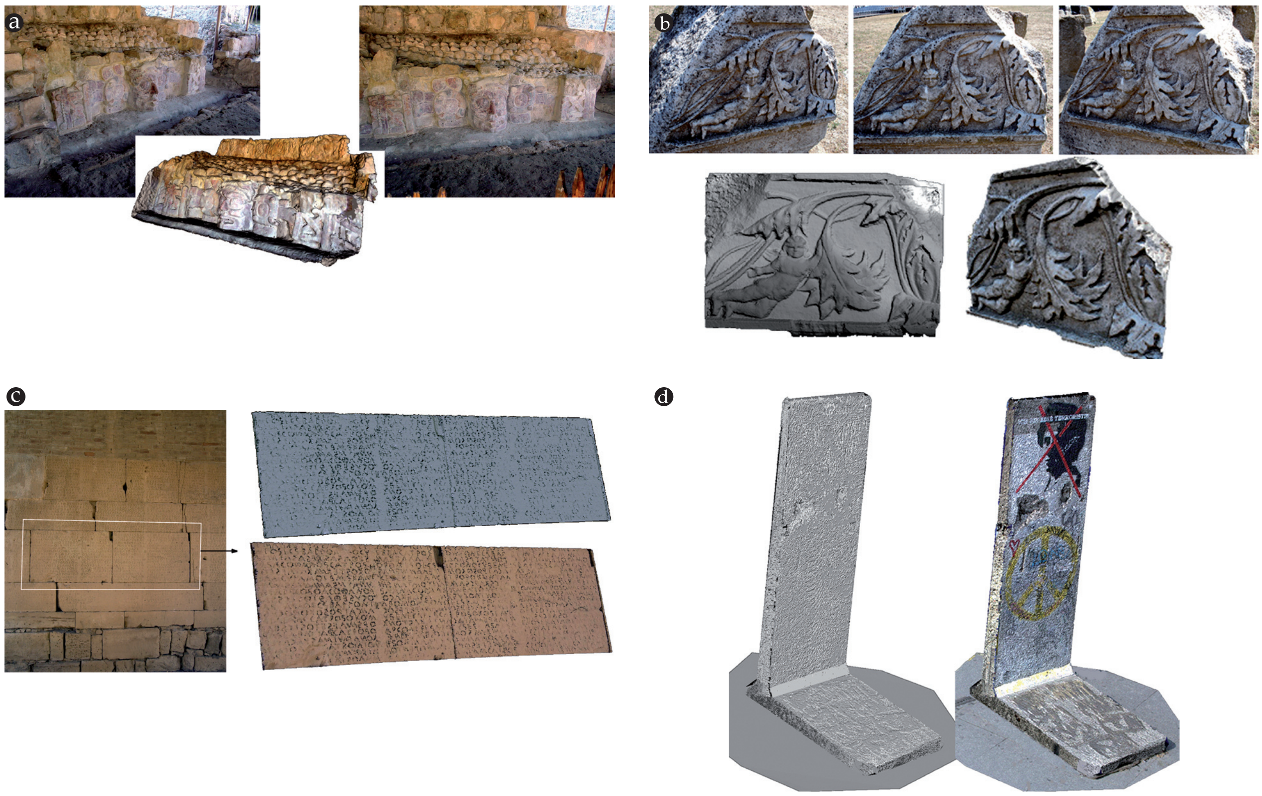

Fig. 4. Examples of stereo matching (a, b), DFS (c), and both stereo matching and DFS, each on a different patch (d).

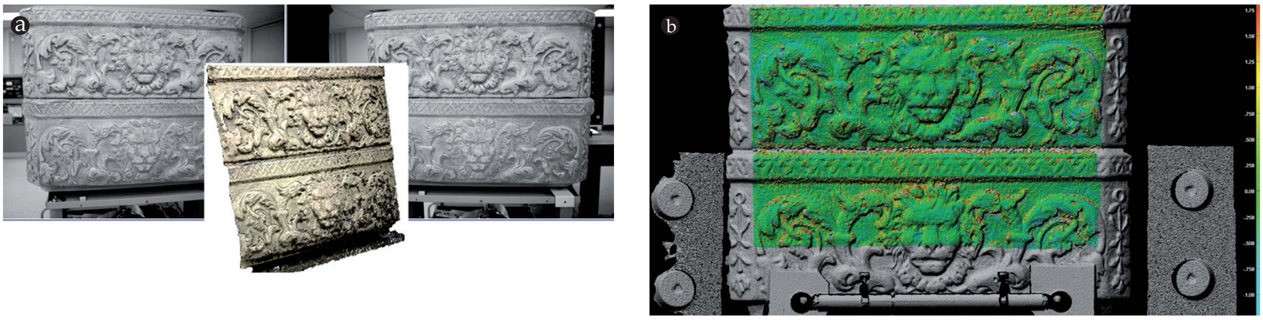

Fig. 5. Lab test object (a) and colour-coded difference between scanned model and image-based model (b).

colour-coded result of the comparison is illustrated in *Fig. 5b*. The standard deviation of the differences between the scanned model and the image-based model was 0.54 mm (Surphaser) and 0.52 mm (ShapeGrabber) averaged over all data sets.

Conclusions

We have presented a sequential segment-based approach that creates detailed models of any shape starting from an manually-created basic model of the whole scene then automatically adds fine geometric details using two dense stereo matching techniques and depth from shading as appropriate. It uses logical and easily established constraints to make it effective. Extensive testing on various types of sites and objects proved its effectiveness in capturing fine details with high accuracy. The accuracy is evaluated by comparing the results to those obtained with two precise close range laser scanners. The comparison showed that our technique is only about 0.5 mm different from the scanner data but requires only a fraction of the cost and time.

References

Allen et al. 2004

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

Alshawabkeh / Haala 2004

Y. Alshawabkeh / N. Haala, Integration of Digital Photogrammetry and Laser Scanning for Heritage Documentation. 20th ISPRS Congress, Commission V Proceedings 35:B5, 2004, 424–429.

Banks / Corke 2001

J. Banks / P. Corke, Quantitative Evaluation of Matching Methods and Validity Measures for Stereo Vision. Internat. Journal of Robotics Research, 20:7, 2001, 512–532.

Boykov / Veksler / Zabih 2001

Y. Boykov / O. Veksler / R. Zabih, Fast Approximate Energy Minimization via Graph Cuts. IEEE Trans. PAMI 23:11, 2001, 1222–1239.

Debevec 2003

P. E. Debevec, Image-Based Techniques for Digitizing Environments and Artifacts. Proceedings of 4th Int. Conf. 3D Imaging and Modeling (Bauff 2003), 234–241.

El-Hakim 2002

S. F. El-Hakim, Semi-Automatic 3D Reconstruction of Occluded and Unmarked Surfaces from Widely Separated Views. Proceedings of the ISPRS Commission V Symposium, Corfu, Greece, 34:5, 2002, 143–148.

El-Hakim et al. 2004

S. F. El-Hakim / J.-A. Beraldin / M. Picard / G. Godin, Detailed 3D Reconstruction of Large-Scale Heritage Sites with Integrated Techniques. IEEE Computer Graphics & Applications, 23:3, 2004, 21–29.

Fassold et al. 2004

H. Fassold / R. Danzl / K. Schindler / H. Bischof, Reconstruction of Archaeological Finds Using Shape from Stereo and Shape from Shading. In: 9th Computer Vision Winter Workshop (Piran 2004) 21–30.

Godin et al. 2002

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

Hong / Chen 2004

L. Hong / G. Chen, Segment-Based Stereo Matching Using Graph Cuts. CVPR 2004 1, 74–81.

Mayoral / Lera / Perez-Ilzarbe 2006

R. Mayoral / G. Lera / M. J. Perez-Ilzarbe, Evaluation of Correspondence Errors for Stereo. Image and Vision Computing 24, 2006, 1288–1300.

MEDIONI / NIVATIA 1985
G. MEDIONI / R. NIVATIA, Segment-Based Stereo Matching. Computer Vision, Graphics and Image Processing 31:1, 1985, 2–18.
MUELLER ET AL. 2004
P. MUELLER / T. VEREENOOGHE / M. VERGAUWEN / L. VAN GOOL / M. WAELKENS, Photo-Realistic and Detailed 3D Modelling: The Antonine Nymphaeum at Sagalassos (Turkey). Computer Applications and Quantitative Methods in Archaeology, 2004, 221–230.
POLLEFEYS ET AL. 2003
M. POLLEFEYS / L. VAN GOOL / M. VERGAUWEN / K. CORNELIS / F. VERBIEST / J. TOPS, Image-Based 3D Recording for Archaeological Field Work. IEEE Computer Graphics & Applications 23:3, 2003, 20–27.
REMONDINO / ZHANG 2006
F. REMONDINO / L. ZHANG, Surface Reconstruction Algorithms for Detailed Close-Range Object Modelling. ISPRS Commission III Symposium, Bonn, 20–22 Sept, 2006. IAPRS&SIS 36:3, 117–123.
SCHARSTEIN / SZELISKI 2002
D. SCHARSTEIN / R. SZELISKI, A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. Int. Journal of Computer Vision 47:1–3, 2002, 7–42.
SCHINDLER / GRABNER / LEBERL 2003
K. SCHINDLER / F. GRABNER / F. LEBERL, Fast On-Site Reconstruction and Visualization of Archaeological Finds. CIPA Symposium (Antalya 2003) 463–468.
SCHOUTEDEN ET AL. 2001
J. SCHOUTEDEN / M. POLLEFEYS / M. VERGAUWEN / L. VAN GOOL, Image-Based 3D Acquisition Tool for Architectural Conservation. CIPA Symposium (Potsdam 2001) 381–388.
STRECHA / FRANSENS / VAN GOOL 2004
C. STRECHA / R. FRANSENS / L. VAN GOOL, Wide-Baseline Stereo From Multiple Views: A Probabilistic Account. IEEE Computer Vision and Pattern Recognition 2004 1, 552–559.
ZHANG ET AL. 1999
R. ZHANG / P.-S. TSAI / J. E. CRYER / M. SHAH, Shape from Shading: A Survey. IEEE Trans. PAMI 21:8, 1999, 690–706.
ZENG ET AL. 2007
G. ZENG / S. PARIS / L. QUAN / F. SILLION, Accurate and Scalable Surface Representation and Reconstruction from Images. IEEE Trans. PAMI 29:1, 2007, 141–158.

Sabry F. El-Hakim

National Research Council (NRC)
Institute for Information Technology
M-50, 1200 Montreal Road, Ottawa
Ontario K1A 0R6, Canada
sabry.el-hakim@nrc-cnrc.gc.ca

Fabio Remondino

Institute of Geodesy and Photogrammetry
ETH Zurich
Wolfgang-Pauli-Str. 15
8093 Zurich, Switzerland

Lorenzo Gonzo
Francesca Voltolini

Centre for Scientific and Technological Research
Fondazione Bruno Kessler
Via Santa Croce 77
38100 Trento, Italy