

Digitizing Archaeological Excavations from Multiple Views

Xenophon Zabulis
Informatics and Telematics Institute
Thessaloniki, Greece
xenophon@iti.gr

Alexander Patterson
University of Pennsylvania
Philadelphia, USA
aiv@seas.upenn.edu

Kostas Daniilidis
University of Pennsylvania
Philadelphia, USA
kostas@cis.upenn.edu

Abstract

We present a novel approach on digitizing large scale unstructured environments like archaeological excavations using off-the-shelf digital still cameras. The cameras are calibrated with respect to few markers captured by a theodolite system. Having all cameras registered in the same coordinate system enables a volumetric approach. Our new algorithm has as input multiple calibrated images and outputs an occupancy voxel space where occupied pixels have a local orientation and a confidence value. Both, orientation and confidence facilitate an efficient rendering and texture mapping of the resulting point cloud. Our algorithm combines the following new features: Images are back-projected to hypothesized local patches in the world and correlated on these patches yielding the best orientation. Adjacent cameras build tuples which yield a product of pairwise correlations, called strength. Multiple camera tuples compete each other for the best strength for each voxel. A voxel is regarded as occupied if strength is maximum along the normal. Unlike other multi-camera algorithms using silhouettes, photoconsistency, or global correspondence, our algorithm makes no assumption on camera locations being outside the convex hull of the scene. We present compelling results of outdoors excavation areas.

1. Introduction

The goal of our approach has been to provide an inexpensive system for 3D-capture of large scale archaeological excavations. Classical fieldwork involved measure by taking point shots using theodolites and trying to combine them in 3D models. Recently, range digitizers like laser scanners or structured light systems have become available to purchase or even to rent. Such systems provide depth maps that have to be registered in a global excavation model and pos-

sibly also combined with color information from images. Though we are pursuing in parallel these active sensor techniques, we also follow an approach which can be accomplished with just one digital still camera and a laptop. The effective cost is much lower than even renting an active sensor system and with no need of any additional equipment except a theodolite for calibration, a device used anyway in any excavation site.

Our system consists of three components: a Euclidean camera calibration using markers shot by a theodolite device, an algorithm for volumetric reconstruction, and a walk-through visualization. Recovery of the 3rd dimension in the second step is based on a novel algorithm which is volumetric and produces as output an occupancy voxel grid where each voxel is assigned with a confidence value, called strength, and a normal of a local planar patch. Availability of normals facilitates the posterior step of finding surface connectivity. Our algorithm is based on three main steps: correlation and orientation selection, view selection, and occupancy decision. Images are back-projected to hypothesized local patches in the world and correlated on these patches yielding the best orientation. Multiple camera pairs compete each other for the best strength for each voxel. A voxel is regarded as occupied if its strength is maximum along the estimated normal. Sub-voxel precision in the localization of occupied voxels is achieved by a novel approach to RBF interpolation.

Our algorithm successfully addresses most of the challenges in stereo and multiple views: Correlation among multiple views is achieved by taking the winning normalized cross-correlation from the “best” camera pair. The application of modified normalized correlation is significantly less sensitive to brightness variation than usual photoconsistency metrics. Occlusion discontinuities with respect to one pair are eventually captured as non-occlusion by other camera pairs. Local deformations in wide-baseline stereo are

explicitly modeled by estimating a local orientation at every voxel. Misregistration of 2+1/2D depth maps is avoided by having a common voxel space and by eliminating induced occupancies through local maxima detection. Last but not least, unlike other multi-camera algorithms using silhouettes, photoconsistency, or global correspondence, our algorithm makes no assumption on camera locations being outside the convex hull of the scene.

1.1. Related Work

The first extensive application in digitizing archaeological sites from images was accomplished by Pollefeys et al. [3, 4, 5]. The difference to our work is that Pollefeys' system does not use any reference points but a denser image sequence which enables both calibration and registration of depth views. We use few calibrated cameras and one single volume where registration is implicit in the volumetric representation. We argue that our "strong" Euclidean calibration with respect to an earth based frame is more robust to multiple views with small overlap. Most of the archaeology digitization efforts focus on laser scanners and structured light systems and will not be reviewed here. A recent survey can be found in [6, 7] and [8].

Two comprehensive review papers of stereo algorithms and their performance can be found in [1, 2]. The simplest multi-camera approach is silhouette-based [9] and exhibits the lowest computational complexity. It assumes that cameras lie outside the convex hull of the scene and that objects can be segmented from their background.

Space-carving [10] approaches provide reconstructions only in terms of determining if a voxel is occupied or not. They do not provide surface normal information which can boost the rendering quality. Regarding such approaches, we found the photoconsistency operator constraining in two respects. First, spectral radiometric/color calibration is difficult to achieve and retain. Second, correlation metrics are more powerful matching operators, because they account for local order of pixels. Due to this property of correlation, our method is able to estimate normals.

Unlike other multi-camera algorithms using silhouettes, photoconsistency, or global correspondence, our algorithm makes no assumption on camera locations being outside the convex hull of the scene. The global optimization algorithms in [11, 12] make assumptions on the camera locations and optimize only over one depth direction. The variational setting of the multiple view stereo in [13] can yield local minima of the cost function and, thus, the method is not guaranteed to converge. In [14], oriented particles were used to reconstruct surfaces from multiple views, but the original data were disparity maps subject to problems discussed in Sec. 1. In [15], a surfel was used to estimate orientation but occupancy was determined by space-carving and very accurate radiometric calibration was required.



Figure 1. A typical calibration image with scene markers selected and labeled to generate the transformation.

2. Calibration

Camera calibration is a calculation of the mapping from image pixels into rays in three dimensional space. This calculation can be done using complex camera models, but our method utilizes a simple pinhole model which does not account for radial distortion. We can use this model because of the high quality lenses on consumer grade cameras that give rise to small lense distortion.

Our approach to calibration has two important characteristics: it computes the camera calibration matrix with respect to a global archaeological site grid and the resources required are many times below that of laser scanning systems. A site grid is a predefined coordinate frame that is consistent over the entire archaeological dig site. Site grid calibration produces vertices that all have the same coordinate reference frame, and thus makes merging models a matter of file concatenation. A second is that models are compatible with CAD which archaeologists already use for vectorized maps and objects. Distances can be measured easily between single and multiple reconstructions making for an entire quantitative site model.

2.1. Calculation

The calibration method described in [16] was used to calibrate the cameras from the input correspondences. We construct a matrix \mathcal{P} such that $\mathcal{P}l = 0$ where $l_{[12 \times 1]}$ is a column vector of the rows of the projection matrix m such that $l = (m_1, m_2, m_3)^T$, where $m_i = (m_{i1}, m_{i2}, m_{i3}, m_{i4})$. P_n^T is the n^{th} correspondence's homogenous world coord-

ordinates and u_n, v_n its image coordinates. Then:

$$0 = \mathcal{P}l = \begin{pmatrix} P_1^T & 0 & -u_1 P_1^T \\ 0 & P_1^T & -v_1 P_1^T \\ \vdots & \vdots & \vdots \\ P_n^T & 0 & -u_n P_n^T \\ 0 & P_n^T & -v_n P_n^T \end{pmatrix} \begin{pmatrix} m_1^T \\ m_2^T \\ m_3^T \end{pmatrix} \quad (1)$$

should hold, but since Eq. 1 is an overconstrained system ($n > 6$), we must minimize the error ($|\mathcal{P}l|$) to get the best fit. This is done by taking the singular value decomposition (SVD) of matrix \mathcal{P} which has rank 11, if $n > 6$ and all the points are not co-planar. The lowest singular value vector from SVD yields the minimum error.

2.2. Process

A total station, which is standard surveying, is used by archaeologists to measure exact locations of artifacts, to generate CAD models of walls, and to create detailed maps of their sites. Using this total station, markers can be measured relative to each other. More importantly, they can be measured with respect to every other marker in a particular dig site because of the site grid.

Markers are laid out in a sparse framework in the scene to be reconstructed as shown (see Fig. 1). Our process uses colored rubber balls cut into hemispheres because they are large enough to be visible in the images and they are visible from a variety of camera orientations. The balls were set in the scene such that from any view there were not all coplanar markers visible, which resulted in enough constraints for the projection matrix. The survey was then performed with a total station to learn their site three dimensional coordinates.

Tuples of pictures are taken of this scene from six to ten positions around it. The exact number depends on the amount of occlusions present in the scene and the size of the specific scene. A program was developed to aid in calibration of these images. It relied on the fact that once six points have been entered ($n \geq 6$) then the remaining points can be predicted by projecting their surveyed world coordinates using the intermediate projection matrix calculated from the subset of entered points. This prediction/confirmation allows for simple detection and correction of mislabeled or erroneously surveyed markers.

Calibration results are presented in Figure 2 by showing the position of the cameras with respect to the volume we are going to reconstruct.

3. Reconstruction from Multiple Views

We propose an operator that yields a measure of the confidence about the occupancy of a voxel in 3D space given a strongly calibrated image pair (I_1, I_2). It is applied at world

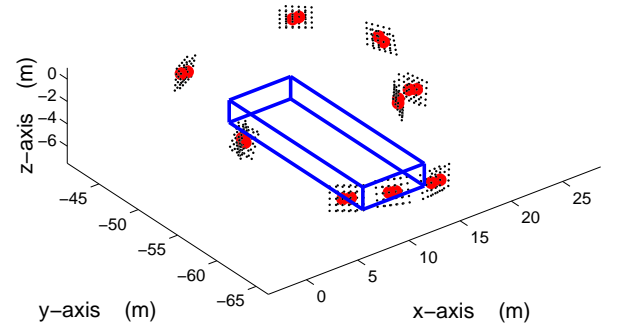


Figure 2. Camera positions obtained from calibration and the area (rectangle) to be reconstructed.

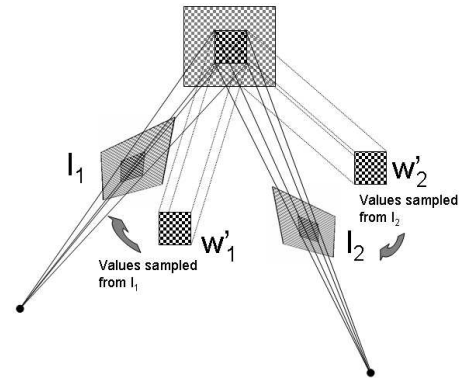


Figure 3. A surface is projectively distorted in images ($I_{1,2}$), but the collineations ($w_{1,2}$) from a planar patch tangent to this surface are not.

point $p \in R^3$ and outputs a confidence score $s(p)$ and a 3D unit normal $\vec{\kappa}(p)$.

Let a planar surface patch $\mathcal{S}_p^{\vec{n}}$, which size is $\alpha \times \alpha$ units of length (mm), centered at p , with unit normal \vec{n} . Back-projecting I_1 and I_2 onto $\mathcal{S}_p^{\vec{n}}$ yields two collineation images $w_1(p, \vec{n})$ and $w_2(p, \vec{n})$. Their formation rule is:

$$I_i (P_i \cdot (p + R(\vec{n}) \cdot [\Delta x \ \Delta y \ 0]^T)), \quad (2)$$

where P_i is the projection matrix of camera i , $R(\vec{n})$ is a 3×3 rotation matrix that maps $[0 \ 0 \ 1]^T$ to \vec{n} , and $\Delta x, \Delta y \in [-\frac{\alpha}{2}, \frac{\alpha}{2}]$ are local horizontal and vertical coordinates of a point on the patch (see Fig. 3).

To obtain $s(p)$ and $\vec{\kappa}(p)$, an $r \times r$ point lattice is assumed on $\mathcal{S}_p^{\vec{n}}$ and the correlation of $w_1(p, \vec{n})$ and $w_2(p, \vec{n})$, $Corr(w_1(p, \vec{n}), w_2(p, \vec{n}))$ is optimized as:

$$s(p) = \max_{\vec{n}} (Corr(w_1(p, \vec{n}), w_2(p, \vec{n}))), \quad (3)$$

$$\vec{\kappa}(p) = \operatorname{argmax}_{\vec{n}} (\operatorname{Corr}(w_1(p, \vec{n}), w_2(p, \vec{n}))). \quad (4)$$

Scalar $s(p)$ and vector $\vec{\kappa}(p)$ are then combined into vector $\vec{v}(p) = s(p)\vec{\kappa}(p)$. We note $w'_i(p) = w_i(p, \vec{\kappa}(p))$ and assume that I_1 and I_2 image Lambertian and textured surfaces as well as that world surfaces are locally planar.

When $\mathcal{S}_p^{\vec{n}}$ is tangent on a world surface then $w_1(p, \vec{n})$ and $w_2(p, \vec{n})$ are images of the underlying surface texture devoid of projective distortions, since parameterization of Δx and Δy is regular. Thus $w_1(p, \vec{n}) = w_2(p, \vec{n})$, since they are sampled at the projections of the same world points. In contrast, if $\mathcal{S}_p^{\vec{n}}$ is not tangent on a surface then $w_1(p)$ and $w_2(p, \vec{n})$ collineate from different world points. Thus similarity of collineations is locally maximized at the loci and orientations of the imaged surfaces and their normals. Thus the loci of surface points and corresponding normals can be then estimated by detection of these local maxima.

The high level of computational complexity of the proposed operator is dealt with two optimizations. The first reduces the number of \vec{n} s considered in Eqs. 3 and 4. To find s and $\vec{\kappa}$, a coarse-to-fine iterative search in the (spherical) domain of normals is adopted. At each iteration, the search focuses into a cone around the highest scoring \vec{n} of the previous iteration. By logarithmically narrowing the opening angle of the cone a speedup of ≈ 100 is obtained.

The second is a scale-space approach to computing \vec{v} across space (see Fig. 4). At each scale, voxel volume is reduced by $1/2^3$ and correlation is computed only at the voxel neighborhoods for which a high correlation value was obtained at the previous iteration. At each scale, $\alpha_i = \alpha_0/2^i$ and $I_{1,2}$ are convolved with a Gaussian of $\sigma_i = \sigma_0/2^i$; r is kept constant. The purpose of these modulations is to efficiently compare w_1 and w_2 at coarse scales since, then, the density of the projections of the points on $\mathcal{S}_p^{\vec{n}}$ is less than image resolution. This way the comparison is also robust, since even a minute calibration error causes the miscorrespondence of sparse lattices on $\mathcal{S}_p^{\vec{n}}$, in the images.

To sufficiently reconstruct wide-area scenes, results are combined from multiple binocular pairs at each p by selecting the pair with the highest correlation: let T pairs, and $\beta(p) = \operatorname{argmax}_j (s_j(p))$, $\beta(p)$ and $j \in \{1, 2, \dots, T\}$; s_j , $\vec{\kappa}_j$, \vec{v}_j are the s , $\vec{\kappa}$, \vec{v} of pair j . Function $\beta(p)$ indicates the pair with the highest correlation at p . Then $\vec{V}(p) = \vec{v}_{\beta(p)}(p)$ yields a field of vectors of the same type as that obtained from a single pair. The principle is that the highest correlation value corresponds to the most matching backprojections which, in turn, correspond to the best estimation of the underlying surface pattern at p (if occupied). Treating then \vec{V} as gradient, a 3D version of Canny's edge detection [17] provides the local maxima up to voxel precision.

Often in outdoor scenes depth discontinuities are accompanied with cast shadows, which annihilate image texture and/or cause aperture phenomena. The result is spurious local maxima (correspondences) which, typically, violate vis-

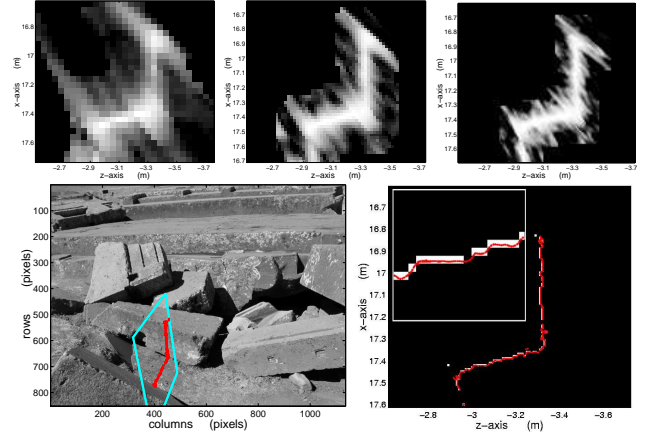


Figure 4. Top: $s(p)$ for 3 scales; left is coarsest. At 1st iteration: $\alpha_0 = 8\text{cm}$, voxel = $(4\text{cm})^3$, $r = 21$, $\sigma_0 = 5$. Bottom: image from the $\approx 40\text{cm}$ binocular pair and detected local maxima with interpolation (see forward). Local maxima are projected on the left image.

ibility rules. A countermeasure is to strengthen the stereo constraint and extend the operator to trinocular [18]. In this work, stereo is constrained to binocular and computation is reduced, by employing a visibility constraint: if a surface point is seen from a particular camera, there are no other surfaces between the camera and the point [10]. Local maxima that do not violate the constraint are detected by being unique along a ray of visibility, from the camera and through the voxel; they are henceforth called *solid* maxima. At each iteration, solid maxima are detected in each \vec{v}_i and the voxels from them to the corresponding cameras are set to be empty. The same voxels are then emptied in all \vec{v}_i , even if they are occupied, because this would mean that these occupancies are spurious. In Fig. 5 a flatland version of the process is illustrated. In the next scale-space iteration, correlation is not computed for the empty voxels, thus accelerating the correlation process.

Reconstruction of surfaces is performed by detecting the strong local maxima of $|\vec{V}(p)|$. The corresponding voxels are considered occupied, with normals: $\vec{\kappa}_{\beta(p)}(p)$. Sub-voxel precision is achieved by RBF ($|x|^3$) interpolation [19], but with the following novelty: the loci of on-surface points are only approximately known. Thus only off-surface pivot points are utilized, projected along the surface normal; one “inside” and one “outside” the surface at a distance of 80% of the voxel’s radius. The RBF’s values at pivot points are such that its 0-isosurface passes through the local maxima. These values are $c|\nabla\vec{V}(p)|$, since $|\nabla\vec{V}(p)|$ becomes 0 when p is a local maximum. Sign c is -1 for the closer of the two pivot points to the camera and 1 for the other.

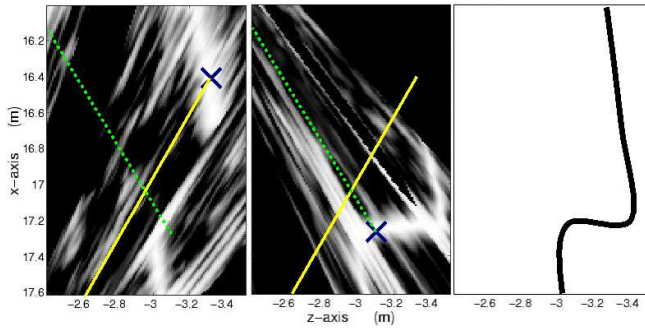


Figure 5. Left, middle: two correlation maps of the same XZ region, acquired from two opposite viewpoints ($\alpha = 2cm$, $\text{voxel} = (1cm)^3$, $r = 21$, $\sigma = 1.25$). A solid maximum is marked in each with 'x' and voxels from them to corresponding cameras are marked with dashed and solid lines. Each view uses the marked voxels from the opposite to suppress spurious maxima (not marked); e.g. middle at xz point $(17.1, -2.9)$. Right: real surface.

4. Results

Our goal is to reconstruct the excavation area in Tiwanaku on the Bolivian Altiplano. In this paper we present two parts of the Pumapunku temple which is a square temple with side length 300m. Scene 2 is an ancient retaining wall around the perimeter of the temple, and scene 1 is a section of stone strewn about by a treasure hunter with a couple of sticks of dynamite. Scene 1 will be an interesting puzzle for archeologists to piece together electronically. Fig. 6 shows two original image pairs from this scene.

All steps except the local maxima detection were carried out in parallel, splitting the data into small cubes. If it would run on a single 2Ghz processor for a scene of $(9m)^3$, a voxelization of $(2cm)^3$, and a \vec{n} -granularity of 1° , the computation time would be 8 hours. However, this time can be significantly optimized if the coarse to fine search for local maxima is also paired with the required precision for \vec{n} and further warping optimizations. Evaluation of visibility constraints took $\approx 20mins$.

Figs. 7 and 8 show the reconstructions obtained for scenes 1 and 2, respectively. Both have a voxel resolution of $2cm^3$. Scene 1 was reconstructed from 12 images. We point out to the reader the minimal number of outliers given the large number of occlusions and orientation variation in this scene. For scene 2, we would like to mention the lack of any deformation in the wall due to misregistration of the 14 input views.

Fig 9 shows the improvement in reconstruction quality obtained by detecting the local maxima with sub-voxel precision, based on the RBF interpolation at $10\times$ the resolu-



Figure 6. Two original image pairs from Scene 1.

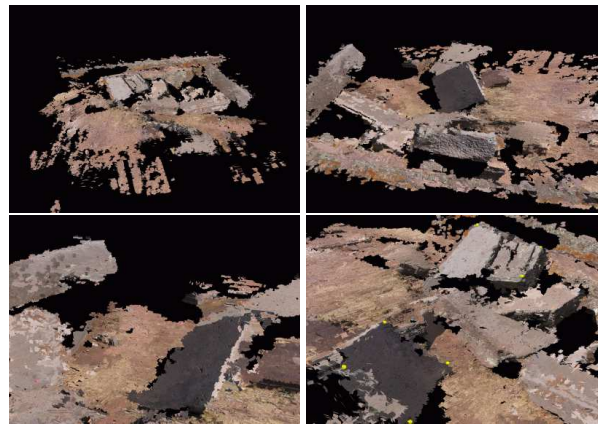


Figure 7. Reconstruction of scene 1 ($14m \times 11m \times 4.5m$, $\text{voxel}: (2cm)^3$).

tion. Due to the increased complexity of the interpolation operation and the large number of occupied voxels in our scenes, interpolation is performed taking into account only neighboring voxels (e.g. in a neighborhood of $(7\text{voxels})^3$). This is not a downside since distant voxels should not have any impact on the local structure of surfaces.

5. Visualization

A separate visualization system has been implemented to better see the reconstruction. It is closely related to the approach presented in [20], but has some major differences. These differences consist of certain simplifications that allow for a simpler implementation with our voxelized data.

Point clouds were used before in order to visualize this voxel data. A plane would be fitted to each voxel, and a

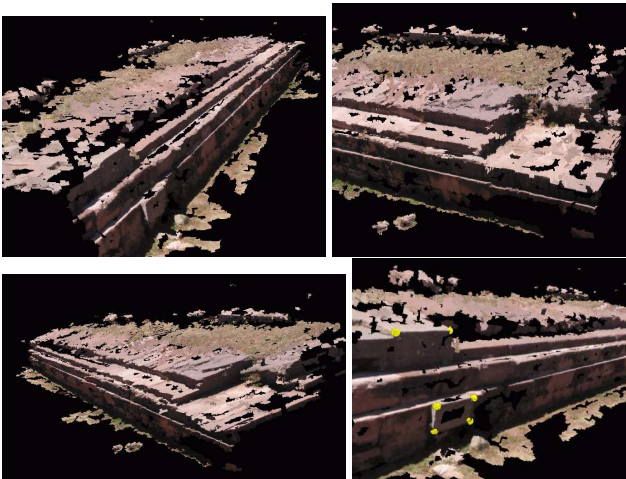


Figure 8. Reconstruction of scene 2 ($6\text{ m} \times 17\text{ m} \times 2\text{ m}$, voxel: $(2\text{ cm})^3$). Bottom right image contains marks, for volume/distance calculations.

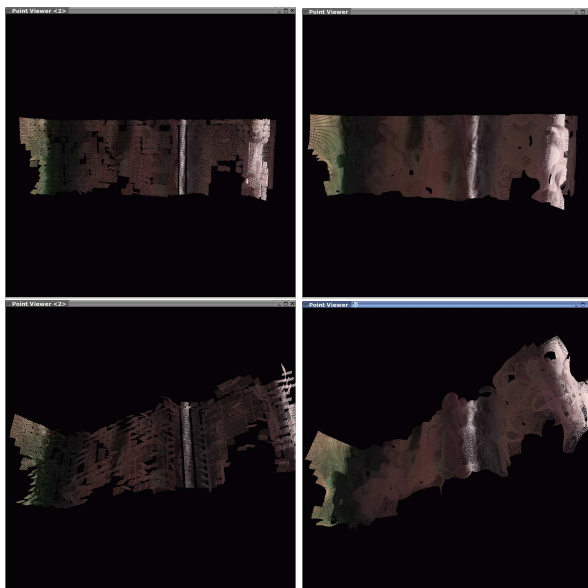


Figure 9. A vertical slice of the wall in scene 2. Left: point based rendering (voxel: $(2\text{ cm})^3$). Right: sub-voxel interpolation (interpolation step: $(0.2\text{ cm})^3$).

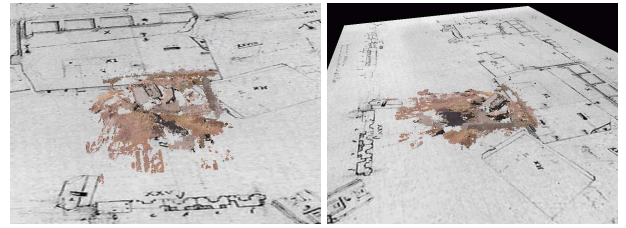


Figure 10. Image of the map of the Puma Punku temple with the reconstruction of a section of that temple overlaid on top of it.

small patch was parameterized to generate enough points to give a sense of the orientation of the voxel as well along with its location. This had the drawbacks that fifty or a hundred points were rendered for each voxel, and when data was examined at close range then the surface would evaporate into an ethereal set of dots. Our so-called facet approach renders each voxel as a 4 distinct vertex quad, or two triangles. This makes for simple input into the OpenGL pipeline where speed improvements to take advantage of current graphics hardware can be implemented.

Finally, since this is an archeological application, there are various features that are useful in the archeological context. One is the ability to measure stones both to determine volume and gain an estimate of weight and work required to shape and move the stone, and to measure distances from artifacts and other objects. This has been implemented by using a simple form of selection. Once a pixel has been chosen, then the depth value is read off from the z-buffer, and the ray in three space corresponding to that pixel is calculated from the OpenGL variable containing the projection matrix. In the bottom right of figure 7 four of these markers are visible allowing the stone to be measured from within the program.

A second improvement for archeological work is the incorporation of a site map into the visualization scheme. A map has been laid out under the model to give an idea of where the sections of the site that were not reconstructed would appear.

6. Conclusion

Archaeological excavation sites make the task of 3D reconstruction from images only very challenging. No assumption about the geometry of 3D structures can be made and the results have to be visually appealing for visualization but also sufficiently accurate for archaeological metrology. In this paper we present an integrated system of calibration, reconstruction, and visualization which requires only one digital still camera for acquisition and a theodo-

lite for calibration. Calibration is euclidean with respect to a georeferenced coordinate system. The reconstruction is a novel algorithm which outputs an occupancy volume where occupied voxels have orientation and a confidence value. The orientation is used in visualization which consists of rendering texture mapped local facets each with the estimated orientation. The reconstruction algorithm – as also shown in [18]– is very successful at slanted surfaces because it optimizes over the local orientation. It is immune to misregistration errors because it does not merge range views like traditional techniques. Instead, it decides for occupancy by selecting the best views as well as the local maximum in voxel space along the estimated orientation. Results have been obtained in two scenes of the Tiwanaku excavation and really show accurate reconstructions of slanted surfaces as well as global registration.

Acknowledgements

This work has started upon an initiative of Professor Clark Erickson, Department of Anthropology, University of Pennsylvania. A first version with no overlap with the system proposed here was developed by Dr. Geoff Egnal and enhanced by Dr. Dianna Xu, both graduated PhD students of the University of Pennsylvania. Jonathan Goulet made the first implementation of the calibration procedure. Dr. Alexei Vranich from the Museum of Archaeology and Anthropology, University of Pennsylvania, guided the fieldwork and the image acquisition. The authors are grateful for support through the grants: NSF-IIS-0083209, NSF-IIS-0121293, NSF-EIA-0324977, NSF-CNS-0423891, NSF-IIS-0431070, ARO/MURI DAAD19-02-1-0383, as well as the 3DTV European Network of Excellence, 6th Framework IST Programme.

References

- [1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *IJCV*, vol. 47, no. 1-2-3, pp. 7–42, 2002.
- [2] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in computational stereo," *PAMI*, vol. 25, no. 8, pp. 993–1008, 2003.
- [3] M. Pollefeys, R. Koch, M. Vergauwen, and L. V. Gool, "Virtualizing archaeological sites," in *Proceedings Virtual Systems and MultiMedia*, 1998.
- [4] M. Pollefeys, L. V. Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch, "Visual modeling with a hand-held camera," *Int. Journal of Computer Vision*, vol. 59, pp. 207–232, 2004.
- [5] M. Pollefeys, L. V. Gool, M. Vergauwen, F. Verbiest, K. Cornelis, and J. Tops, "Image-based 3d recording for archaeological field work," *IEEE Computer Graphics and Applications*, vol. 23, pp. 20–27, 2003.
- [6] S.F.El-Hakim, J.-A. Beraldin, and M. Picard, "Effective 3d modeling of heritage sites," in *3-D Digital Imaging and Modeling*, 2003.
- [7] P. Debevec, "Image-based techniques for digitizing environments and artifacts," in *3-D Digital Imaging and Modeling*, 2003.
- [8] A. Chalmers and H. Rushmeier, "Special issue on computer graphics in art history and archaeology," *IEEE Computer Graphics and Applications*, vol. 22, no. 5, pp. 22–23, 2002.
- [9] A. Laurentini, "The visual hull concept for silhouette-based image understanding," *PAMI*, vol. 16, no. 2, pp. 150–162, 1994.
- [10] K. N. Kutulakos and S. M. Seitz, "A theory of shape by space carving," *IJCV*, vol. 38, no. 3, pp. 197–216, 2000.
- [11] V. Kolmogorov and R. Zabih, "Multi-camera scene reconstruction via graph cuts," in *Proc. ECCV 02*, vol. 1, pp. 379–393, 1998.
- [12] K. Junhwan, V. Kolmogorov, and R. Zabih, "Visual correspondence using energy minimization and mutual information," in *Proc. ICCV 03*, vol. 2, pp. 1033–1040, 2003.
- [13] O. Faugeras and R. Keriven, "Complete dense stereovision using level set methods," in *Proc. ECCV 98*, vol. 1, pp. 379–393, 1998.
- [14] P. Fua, "Reconstructing complex surfaces from multiple stereo views," in *Proc. ICCV*, pp. 1078–1085, 1995.
- [15] R. Carceroni and K. Kutulakos, "Multi-View scene capture by surfel sampling: From video streams to Non-Rigid 3D motion, shape & reflectance," *IJCV*, vol. 49, no. 2-3, pp. 175–214, 2002.
- [16] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*. Prentice Hall Professional Technical Reference, 2002.
- [17] J. F. Canny, "A computational approach to edge detection," *PAMI*, vol. 8, no. 6, pp. 679–698, 1986.
- [18] X. Zabulis and K. Daniilidis, "Multi-camera reconstruction based on surface normal estimation and best viewpoint selection," in *Proc. of IEEE 3DPVT*, pp. 733–40, 2004.
- [19] G. Turk and J. F. O'Brien, "Modelling with implicit surfaces that interpolate," *ACM Transactions on Graphics*, vol. 21, no. 4, pp. 855–873, 2002.
- [20] L. Ren, H. Pfister, and M. Zwicker, "Object space ewa surface splatting: A hardware accelerated approach to high quality point rendering," *Computer Graphics Forum*, vol. 21, no. 3, pp. 461–470, 2002.