Flatland: A Tool for Transforming Historical Sites into Archival Drawings

Rattasak Srisinroongruang and Eric Sinzinger and Glenn Hill

Texas Tech University

Abstract

Detailed documentation of historical sites is important for archaeological discovery and cultural preservation. The traditional method of documentation is to hand sketch 2D drawings of the region. Laser range finders can be used to produce highly accurate geometric representations of the historical sites, and high resolution images provide vital detail. However, archaeologists are both used to and prefer a flat, two dimensional archival drawing of the region. Flatland provides the missing link for archaeologists between three dimensional representations and archival drawings. There are two critical pieces of Flatland — texture mapping and geometric transformation.

The texture map acquisition phase aligns the world geometry with the high resolution images. Many historical sites contain rocky, uneven terrain and structures that do not contain distinguishing features that would allow for automated methods of correspondence selection. Instead, manually selected correspondences between the point cloud and the high resolution image are used to compute the texture map. If the selected region is nonplanar, then a camera projection matrix is computed to determine the texture map for the point cloud. However, when the selected region is planar, the camera projection matrix can not be computed, and instead a homography is used to determine the texture map.

The geometric transformation phase allows the archaeologists to essentially unfold both square and cylindrical surfaces to representations with a single dominant plane. A square room can be unfolded into four panels where geometric distance is preserved within each panel. An elliptical surface is unrolled about a flat rectangle with the height of the elliptical surface and the width equivalent to the circumference of the surface. The use of Flatland is demonstrated on scans from the Mesa Verde National Park.

1. Introduction

Traditionally archaeologists record historical sites by generating hand drawn 2D sketches. The sketches also require notation of interesting features or critical regions. This process is both time consuming and inaccurate. Three dimensional scanning technology could be used to generate accurate, detailed geometric representations of the region. However, the geometry does not project well onto an archaeologist's standard mode of data representation — flat paper.

A tool, Flatland, is presented that allows archaeologists to register high resolution photographs of an archaeological site with range data taken of the same area and then to transform the geometry into a representation better suited for 2D archival. This registration is done without a reference coordinate system defined through either the traditional methods of a marker system or surveying critical points that would require use of additional equipment by the archaeologists and whose accuracy is then still dependent on reliably determining the surveyed points in the captured data.

The need for Flatland is three-fold. First, Flatland allows the archaeologists to generate an archival documentation of the entire scanned region. Second, with a

textured high resolution mesh of the object, tracing a plan can be done in 3D, resulting in more interactive and efficient work flow. By tracing in 3D, sampling the data to a 2D space is not required, preserving the dimensionality of the captured data. Third, high fidelity representations of cultural sites in 3D provide the capability for expanding site documentation and archival compared to the current sketches generated by archaeologists.

There are three critical issues for using laser scanning in generating accurate historical representations. The first is that the resolution of current range scanners is insufficient to represent important details such as fine cracks. These details, however, are visible with a photograph taken by a high-resolution camera. Therefore, texture mapping must be incorporated to present the detail which induces the second critical issue. The second critical issue is the alignment of the range map to the image. Because of the nature of the data, automated methods to register the point cloud with the photographs are not feasible. The rocky nature of many archaeological sites inhibits the use of shape and geometry in producing accurate registration results. The third issue is the geometric transformation from complete three dimensional representation to flat paper.

The Vanishing Treasures program is an initiative designed to preserve historic sites at national parks in the arid climate of regions located in the western United States. The presented framework is part of a larger system being employed by archaeologists at Mesa Verde national park, an area containing over 600 sites of interest to staff archaeologists, of which only four have been documented at the desired level of detail. One site within Mesa Verde, Cliff Palace with over 200 rooms, is believed to the largest cliff dwelling in North America. The whole documentation system is also being incorporated at Canyonlands National Park, Hovenweep National Monument, Arches National Park, and Flagstaff area parks.

2. Related Work

Cultural heritage itself provides an interesting application of computer graphics and vision techniques in a variety of real-world settings [BC00], each presenting unique and specific challenges. Cultural preservation using laser range finders has been performed varying in scale from statues [BRM*02, LPC*00, INHO03] to large buildings of period architecture styles [AST*03, HO02] and as well as artifacts in between [GBU00, GMR*05]. Techniques used for registration in urban or modern environments rely on the existence of dominant lines and planes in the original scan and photographs [LS05, LSY*06, SA00] as features for generating correspondences between two datasets. Debevec also relies on dominant lines in photographs using the framework presented in [DTM96].

A technique for generating planes to represent building facades is given in [MMP04]. Their technique, while useful for automating the process of scene reconstruction from point clouds, results in a loss of the precision in the original range scans that are beneficial for archaeologists. Another semi-automated method for generating 3D scenes whose primary concern is aesthetic quality rather than precision is described in [MMP03]. A general overview of issues related to 3D model and color acquisition is given in [BR02].

An automated method for registration of a 3D model with a photograph is given in [LHS01] that works well if the object of interest is easily extracted from its background in the photograph. This is not the case with the sites at Mesa Verde where the majority of sites and surrounding region are the same general shade of color. A method for automatic registration of natural sites is given in [TA04]. Their method requires known position of the sun for the photographs and uses the shadows cast a cue for optimization. Certain sites in Mesa Verde National Park are in caverns or beneath large overhangings that prevent any discerning shadows to be cast from the objects of interest.

A framework for representing 3D models with high-fidelity texture detail from range scans and photographs is given in [BMR01]. The authors use an image based approach from computed albedo and normal maps to improve the registration across multiple 3D scans. The authors primary concern is generating high quality textures to map onto relatively lower resolution 3D scans to produce an aesthetically pleasing result.

A framework for minimizing user selection of correspondences was presented in [FDG*05]. Their framework was concerned with minimizing redundant correspondences and inferring new correspondences when registering multiple photographs to a single mesh. This was done by representing the state of registration as a graph with nodes representing photographs and the mesh while arcs represent correspondences.

3. Methodology

Flatland takes an image of the site and a 3D representation of the site and outputs an "unrolled" textured mesh of the site. The archaeologist defines the regions of interests in the scans by selecting cutting planes and then defines correspondences between the photograph and the active cutting plane region. The input to the framework are photographs and meshes of the archaeological site. The meshes are obtained from

time-of-flight range scans that are then triangulated to provide facet information.

The cutting planes that define the region of interest in the mesh are determined from three points selected from the region of interest. The archaeologist can define multiple planes that are then tiled and presented for selection. The selected plane should be the region that most closely matches the current photographed being registered to. After selection of correspondences, the texture is mapped onto all of the points defined by the current plane selection. The type of mapping is selected from between the camera projection computation and the 2D homography computation.

Orthogonality between the range scans and associated digital pictures with the surface of interest cannot be assumed. This is necessary for the speed of documentation and by request of the staff archaeologists because of the working environment in which the scans are taken. Instead, multiple photographs across a region are to be registered with the range scan and blended to produce a detailed representation of the true site that can be traced by the archaeologists.

3.1. Plane-based Region Segmentation

A plane is defined by 3 selected points in the mesh that roughly describe the region that is to be mapped. Because of the nature of the data, the 3 selected points may not accurately fit the region that the archaeologist intended. To produce a more accurate plane that fits the data from the 3 selected points, a better fit plane is produced though optimization of least squares. The plane computed from the selected points is defined by the normal plane equation

$$Ax + By + Cz + D = 0. \tag{1}$$

A, B, C, and D are parameters of the model being optimized. Since the goal is the best fit plane for the data around the region specified by the archaeologist, the measure used to optimize the parameters is the distance from all the points in the region to the plane. Given the initial plane, the neighborhood of plane points is defined by

$$\Gamma = \{(x_i, y_i, z_i) \text{ s.t. } |Ax_i + By_i + Cz_i + D| < \epsilon\}. \tag{2}$$

where ϵ is an arbitrary parameter. Then, optimization is done after all points have been added to Γ, so that all points are equally weighted. When optimization was performed on the plane parameters after each point was added, an artificial weighting of the plane parameters towards the points that are added earlier in the iteration was observed. This artificial weighting resulted in some points being left out that would normally be added near the end of the iteration because the plane parameters would be weighted towards the earlier points.

3.2. Cylindrical Surface Unrolling

Although most man made structures are built of mainly rectangular surfaces, the second most common surface is cylindrical. Example surfaces include barn silos, castle turrets and ceremonial kivas. When archaeologists record sites like kivas they essentially unroll the cylindrical surface onto the flat printed page. Automating this process upon range data is nontrivial and requires the development of specialized algorithms.

The cylindrical data is first manually segmented from the other regions of the scan. Then the cylindrical data is centered and rotated so that the data is oriented about the y-axis. An eigen analysis of the surface normals of the data is used to determine the rotation. The least significant eigenvector corresponds to the current axis for the cylinder. A simple rotation orients the data about the y-axis.

Once oriented about the axis, the height, or y value of the 3D data is ignored, and the polar coordinates, (ρ, θ), is determined for all points. The points are divided into M segments based upon their angular rotation, θ. A spline is fit through the points in each segment using least squares. The complete spline, $s(\theta)$, determines the local "plane" of the surface. The unrolled point determined by (x, y, z) becomes

$$F(x, y, z) = (d(\theta), y, z(\theta, x, z)). \tag{3}$$

The term $d(\theta)$ represents the distance along the spline and is measured by

$$d(\theta) = \int_0^\theta s'(\theta) d\theta. \tag{4}$$

The term $z(\theta, x, y, z)$ represents the distance between the original point, (x, z) and the spline, $s(\theta)$. This is simply the Euclidean distance between the 2D points

$$z(\theta, x, z) = \sigma \left|\left|(x, z) - s(\theta)\right|\right|_2. \tag{5}$$

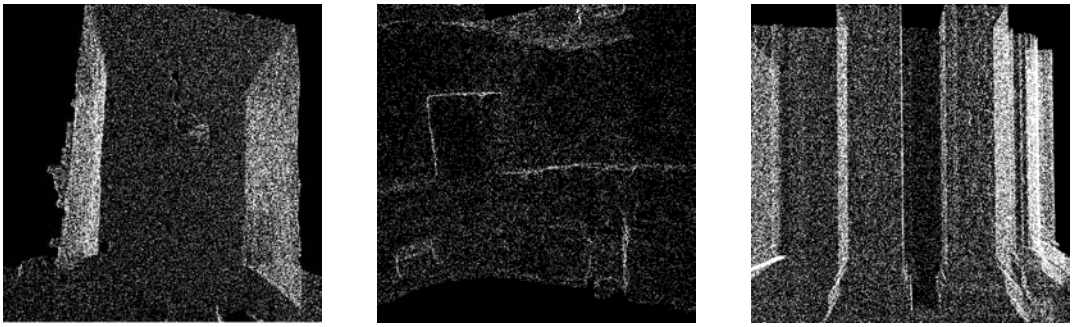
with a sign, σ, dependent upon whether or not the point is in front or behind the spline. This function allows the unrolled cylindrical surface to preserve the original depth characteristics.

3.3. Computation of Camera Projection Matrix

Computation of the projection matrix $\mathbf{C}$ requires the selection of six or more correspondences between the point cloud and the photograph. This is necessary because of the eleven degrees of freedom present (ignoring scale) in the projection matrix. Let $\mathbf{P}_i$

(a) (b) (c)

Figure 1: *Photographs used to register with the range scans. (a) Mesa Verde housing structure. (b) Mesa Verde kiva. (c) University building.*



(a) (b) (c)

Figure 2: *The corresponding scans of sites pictured in Figure 1.*

be the 3D homogeneous data points represented by (X_i, Y_i, Z_i, W_i) and $\mathbf{p}_i$ be the 2D image points represented by (x_i, y_i, w_i). For each correspondence between the 3D points and 2D points, $\mathbf{P}_i \leftrightarrow \mathbf{p}_i$, the following relationship is defined [HZ03]

$$\mathbf{A}_i = \begin{pmatrix} \mathbf{0}^T & -w_i\mathbf{P}_i^T & y_i\mathbf{P}_i^T \\ w_i\mathbf{P}_i^T & \mathbf{0}^T & -x_i\mathbf{P}_i^T \\ -y_i\mathbf{P}_i^T & x_i\mathbf{P}_i^T & \mathbf{0}^T \end{pmatrix} \quad (6)$$

where $\mathbf{A}_i$ is a 3×12 matrix. The system with n correspondences to be solved is then defined as

$$\begin{pmatrix} \mathbf{A}_1 \\ \dots \\ \mathbf{A}_n \end{pmatrix} \mathbf{C} = \mathbf{0} \quad (7)$$

where $n \geq 6$ and each row of the projection matrix $\mathbf{C}$ is appended onto the previous row and represented in the system as a sixteen element column vector.

Data normalization is done before computation of the projection matrix. Normalization translates the centroid to the origin and scaled so that the RMS distance from the origin is $\sqrt{3}$. To work within the original dataset's space, the computed projection matrix is unnormalized afterwards. Let $\mathbf{N}_1$ be the transformation matrix for the 2D points and $\mathbf{N}_2$ be the transformation matrix for the 3D points, then the unnormalized matrix is computed as $\mathbf{N}_1^{-1}\mathbf{C}\mathbf{N}_2$.

The space and time complexity for computation of the projection matrix should be noted. Each correspondence provides a constraint on the system represented as a 3×12 matrix and, as such, the space complexity for the camera projection matrix computation with n correspondences is $\mathbf{O}(n)$. Singular value decomposition of the system of correspondences determines the camera projection matrix, leading to a time complexity of $\mathbf{O}(n)$.

3.4. 2D Homography

In the case that the mesh or the photograph contains solely data that lies on a plane or nearly on a plane,

Figure 3: *The result of texture mapping the adobe housing structure. Registration against the rocky, uneven surface of the building is done well. The floor region of the scan is not registered well against the photograph because of the lack of dominant correspondences to select from in the image.*

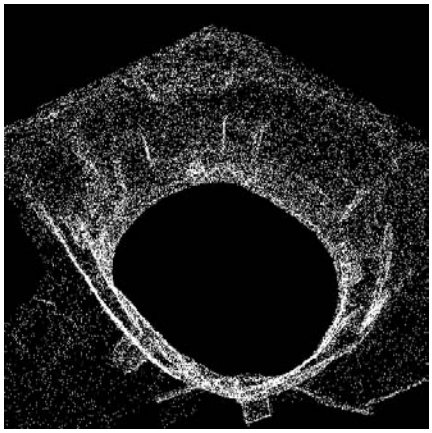


Figure 4: *A scan of the kiva site. The circular hole in the middle represents the floor of the kiva. There are 1,025,304 points in the scan.*

the computation of the camera projection matrix using the method above will not work. To solve this case, the problem is reduced to a homography between two 2D spaces. Reduction is done by projecting the current view of the mesh into the frame buffer and treating the buffer as another image. The user still selects correspondences in 3D space, and the selected points in the mesh are projected to 2D space before computation of the homography between the mesh and the photo.

Data normalization is performed in a similar manner to the camera projection matrix case before computation of the homography. Normalization allows the method to produce good results without requiring scaling of the frame buffer to match the dimensions of the photograph. Let $\mathbf{P}_i$ representing the 3D point projected into 2D space be denoted as (X_i, Y_i, W_i) and let $\mathbf{p}_i$ representing the 2D image points be denoted as (x_i, y_i, w_i).

Given at least four correspondences between the mesh and the photograph, $\mathbf{P}_i \leftrightarrow \mathbf{p}_i$, the relationship $\mathbf{A}_i\mathbf{C} = \mathbf{0}$ is the same as Equation 6 with 2D homogeneous coordinates. In this case, $\mathbf{A}_i$ is a 3×9 matrix and $\mathbf{C}$ is a vectorized version of the 2D homography matrix formed by appending the end of the previous row and represented as a nine element column vector. All constraint matrices are then arranged similarly to Equation 7 for $n \geq 4$.

To compute the homography matrix $\mathbf{C}$, the SVD of $\mathbf{A}$ is computed. The homography matrix is the singular vector associated with the smallest singular value. Thus if the SVD computation of $\mathbf{A}$ is $\mathbf{UDV}^T$, then the homography matrix is the last column of $\mathbf{V}$. If the normalization transform for the photograph points is $\mathbf{N}_1$ and the normalization transform for the projected mesh points is $\mathbf{N}_2$, then the de-normalized homography matrix is computed as $\mathbf{N}_1^{-1}\mathbf{C}\mathbf{N}_2$. During computation of texture coordinates for the 2D homography, the mesh points are first projected onto the defined cutting plane and then the mapping is computed. This is necessary since the homography defines the relationship between 2D spaces.

Each additional correspondence incurs an additional 3×9 constraint matrix, resulting in a space complexity of $\mathbf{O}(n)$ for n correspondences. The determining factor in computation of the 2D homography in terms of time complexity is the SVD computation, leading to a complexity of approximately $\mathbf{O}(n)$.

4. Results

Results are presented on data collected from both Mesa Verde National Park and university buildings. Some results of the mapping are presented below. The first site is of an adobe housing structure at Mesa Verde within a neighborhood of similar structures. In Figure 1a, a photograph taken of the housing structure at Mesa Verde is to be mapped onto the scan shown in Figure 2a. In this case, there are sufficient multi planar features that can be selected to allow the use of the projection matrix computation in order to do the mapping. The results of this mapping are shown

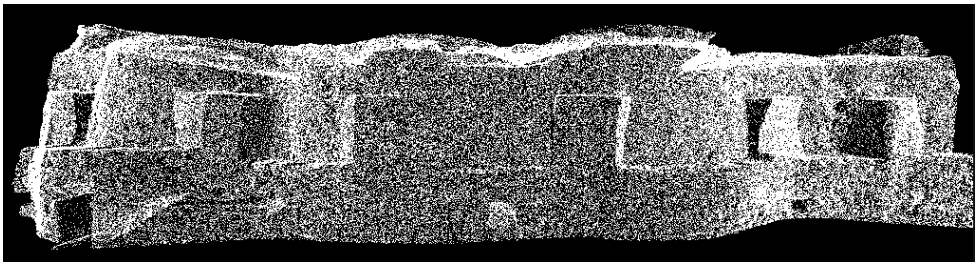


Figure 5: *The entire kiva scan after application of surface unrolling. The view is zoomed out to show the entire scan projected onto a planar region.*

in Figure 3. The mapping produced acceptable results on the main face of the building.

The next site was a kiva structure at Mesa Verde. The kiva is a subterranean room used for religious rituals by the Native American tribes in the region. Many are roughly elliptical or circular in nature. The original kiva scan is shown in Figure 4 and a photograph of a portion of the kiva is shown in Figure 1b. The irregular curvature on the surface of the kiva is visible in Figure 2b. There are, however, multi-planar layers in the kiva that make it suitable for use with the camera projection computation. A lower resolution kiva mesh used for visibility purposes is shown unrolled in Figure 5. The result of the mapping is shown in Figure 6. The extruded pillars are mapped appropriately from its representation in the photograph as is the rectangular hole in the bottom left of Figure 6.

Figure 6: *The result of texture mapping the kiva scan with the photograph in Figure 1b. The mapping aligned well with the extruded pillar like structures on the kiva wall as well as the border between the kiva wall and the floor.*

The side of a university building is used to show the error present when using the camera projection algorithm for points that lie on a single planer region. This test case is representative of large flat areas on the side of adobe structures that are prevalent at Mesa Verde. When mapping a high-resolution photograph without distinguishing multi-planar features, the camera projection computation fails. The scan from the university building test site is shown in Figure 2c and a photograph of the same region in Figure 1c. The results of the camera projection algorithm on this pair of scan and photograph is shown in Figure 7. The mapping shown in Figure 7 is from correspondences selected from the planar region encompassing the whole scan. Even if the correspondences were all selected from one planar region between two extruding pillars, the mapping would still produce an incorrect undulating pattern onto the scan. The scan lacks the resolution to select features on the carved designed portion of the building with the required precision. Instead, using the 2D homography computation results in the textured mesh shown in Figure 8. The mapping onto the planar background and the extruding pillars is correct. However, the mapping on the sides of the pillars are stretched because the photograph was taken orthogonal to the face of the building. The sides of the pillars are hence not represented in any region of the photograph.

Table 1: *Mesa Verde housing structure results*

Correspondences	Error(px)	Time(ms)
6	4.858	2.068
10	3.965	2.142
14	2.580	2.154

All photographs used for registration are 2048 × 2048. Computational results for the three sites are given in Tables 1, 2, 3. The computation time includes time for normalization of the correspondences and computation time to compute the 2D homogra-

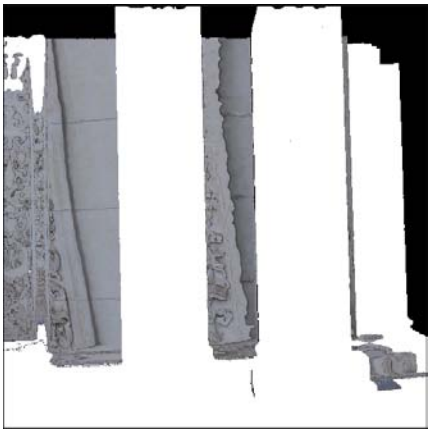


Figure 7: *The results of mapping the scan when selecting correspondences from a dominantly planar region and using the camera projection computation. The extruded pillars are not mapped because the computed texture coordinates are out of bounds of the original photograph.*

Figure 8: *An angled view of the building scan mapped with the 2D homography. The mapping against the extruded pillars is done well. The sides the pillars are stretched since the original photograph in Figure 1c did not have the sides of the pillars in view.*

Table 2: *Mesa Verde kiva results*

Correspondences	Error(px)	Time(ms)
10	34.333	2.116
14	30.473	2.146
18	10.767	2.198

phy in the case of the administrative building and the camera projection matrix in the two Mesa Verde scans. To provide a quantitative measure of the precision of the homography computation independent of any human error from selection of correspondences, a group of twenty correspondences is used as the ground truth. Then a subset of those correspondences is used to measure the accuracy of the homography matrices in computing the ground truth. The results are the average of three different subsets from the twenty correspondences for each site. In the case of the kiva, the highly irregular surface structure and difficulty in selecting correspondence points required a larger number of initial matches to achieve convergence.

Table 3: *University building results*

Correspondences	Error(px)	Time(ms)
6	9.267	1.857
10	5.520	1.936
14	3.273	1.998

5. Conclusion

A framework has been presented that allows an archaeologist to produce high fidelity 3D representations of culturally significant sites by combining high resolution photographs and 3D scans. By defining cutting planes in the scans, the archaeologist can transform the representation into flat 2D panels suitable for printing on paper and akin to their traditional methods of archival storage. However, using the Flatland system allows this work to be done accurately in a fraction of the time required to hand sketch the regions.

Beyond the development of an application for archaeologists, there are two main contributions of Flatland. First, Flatland is able to handle texture mapping on complex, natural geometry. This includes the capability to handle object such as cliff faces without clear correspondence points in the image or depth space, as well as nearly planar surfaces where there is insufficient information to compute the camera projection models. The second contribution of Flatland is its ability to segment and produced flattened panels including building faces and cylindrical rooms.

These flattened panels can be used by archaeologists to generate archival records of world heritage sites. The first client of Flatland will be archaeologists working at the Mesa Verde National Park. Based upon interactions with the Mesa Verde staff, future refinements to Flatland will be made to create a suitable

technology transfer tool to the archaeology community.

6. Acknowledgments

The Mesa Verde project is being supported by the KyArk fund of Kacyra Family Foundation.

References

[AST*03] Allen P. K., Stamos I., Troccoli A., Smith B., Leordeanu M., Hsu Y.: 3D modeling of historic sites using range and image data. In *Proceedings of the 2003 IEEE Conference on Robotics and Automation* (2003), pp. 145–150.

[BC00] Berndt E., Carlos J.: Cultural heritage in the mature era of computer graphics. *Computer Graphics and Applications, IEEE Transactions on 20*, 1 (2000), 36–37.

[BMR01] Bernardini F., Martin I., Rushmeier H.: High-quality texture reconstruction from multiple scans. *Visualization and Computer Graphics, IEEE Transactions on 7*, 4 (2001), 318–332.

[BR02] Bernardini F., Rushmeier H.: The 3d model acquisition pipeline. *The Computer Graphics Forum 21*, 2 (June 2002), 149–172.

[BRM*02] Bernardini F., Rushmeier H., Martin I., Mittleman J., Taubin G.: Building a digital model of Michelangelo's Florentine Pieta. *Computer Graphics and Applications, IEEE 22* (2002), 59 – 67.

[DTM96] Debevec P. E., Taylor C. J., Malik J.: Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach. In *Proceedings of the 23rd annual ACM SIGGRAPH conference on computer graphics and interactive techniques* (1996), pp. 11–20.

[FDG*05] Franken T., Dellepiane M., Ganovelli F., Cignoni P., Montani C., Scopigno R.: Minimizing user intervention in registering 2D images to 3D models. *The Visual Computer 21*, 8-10 (September 2005), 619–628.

[GBU00] Gaiani M., Balzani M., Uccelli F.: Reshaping the Coliseum in Rome: An integrated data capture and modeling method at heritage sites. *The Computer Graphics Forum 19*, 3 (September 2000), 369–378.

[GMR*05] Guidi G., Micoli L., Russo M., Frischer B., De Simone M., Spinetti A., Carosso L.: 3D digitization of a large model of imperial Rome. In *Proceedings of IEEE Fifth International Conference on 3-D Digital Imaging and Modeling* (2005), pp. 565–572.

[HO02] Hanke K., Oberschneider M.: The medieval fortress Kufstein, Austria - an example for the restitution and visualization of cultural heritage. In *Proceedings of International Society of Photogrammetry and Remote Sensing Comm V.* (2002), pp. 530–533.

[HZ03] Hartley R., Zisserman A.: *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.

[INHO03] Ikeuchi K., Nakazawa A., Hasegawa K., Ohishi T.: The Great Buddha project: Modeling cultural heritage for vR systems through observation. In *The Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality* (2003), pp. 7–16.

[LHS01] Lensch H. P., Heidrich W., Seidel H.-P.: A silhouette-based algorithm for texture registration and stitching. *Graphics Models 63*, 4 (2001), 245–262.

[LPC*00] Levoy M., Pulli K., Curless B., Rusinkiewicz S., Koller D., Pereira L., Ginzton M., Anderson S., Davis J., Ginsberg J., Shade J., Fulk D.: The digital Michelangelo project: 3D scanning of large statues. In *Proceedings of ACM SIGGRAPH 2000* (2000), pp. 131–144.

[LS05] Liu L., Stamos I.: Automatic 3D to 2D registration for the photorealistic rendering of urban scenes. In *Proceedings of the 2005 IEEE Conference on Computer Vision and Pattern Recognition - Volume 2* (2005), pp. 137–143.

[LSY*06] Liu L., Stamos I., Yu G., Wolberg G., Zokai S.: Multiview geometry for texture mapping 2D images onto 3D range data. In *Proceedings of the 2006 IEEE Conference on Computer Vision and Pattern Recognition* (2006), pp. 2293–2300.

[MMP03] Mayer H., Mosch M., Piepe J.: Modelling the walled city of Nicosia. In *Proceedings of IEEE Symposium on Visual Analytics Science and Technology* (2003), pp. 57–65.

[MMP04] Mayer H., Mosch M., Piepe J.: 3D model generation and visualization of Wartburg Castle. In *Proceedings of ISPRS International Workshop on Processing and Visualization Using High-Resolution Imagery* (2004).

[SA00] Stamos I., Allen P.: 3D model construction using range and image data. In *Proceedings of the 2000 IEEE Conference on Computer Vision and Pattern Recognition* (2000), pp. 1531–1536.

[TA04] Troccoli A. J., Allen P. K.: A shadow based method for image to model registration. In *Proceedings of 2nd IEEE Workshop on Image and Video Registration* (2004), pp. 169–178.