

# Non-Contact Fiducial Based 3-Dimensional Patch Merging Methodology and Performance

Laurence G. Hassebrook, Charles J. Casey and Walter F. Lundby

Department of Electrical and Computer Engineering,  
Center for Visualization and Virtual Environments,  
University of Kentucky  
Lexington, Kentucky, USA.

---

## Abstract

Traditional 3-Dimensional patch merging is performed by first obtaining a skeleton using photogrammetry, followed by 3-Dimensional surface patch scanning followed by patch merging onto the skeleton geometry. This approach is effective but generally involves the use of physical fiducials to be mounted onto the object surface and requires costly scanning hardware and software. We present a more unified non-contact approach where the fiducials are projected light patterns and the patch scanner is software based, in that the cameras and digital projectors are off the shelf general purpose technologies. We combine the patch depth acquisition with digital color photography, each having different pixel resolutions. We evaluate the manual merging of the captured patch data and present a new point reduction method based on voxel grid mapping. We demonstrate the method's implementation and capability with an artifact from the CSS Alabama.

*Key words:* Structured Light Illumination, Registration, 3-D scanning

---

## 1 Introduction

We present a demonstration and discussion of 3-dimensional patch merging using light projected fiducials and structured light illumination. Traditional surround structured light scanning systems require physical fiducials to be placed at regular intervals around the surface of an object. These fiducials are photographed from different directions and processed into a 3-dimensional skeleton using photogrammetry.<sup>1</sup> A series of scanned surface patches are then mapped onto the skeleton and merged using some form of an

iterative closest point algorithm.<sup>2</sup> The result is a surround manifold representing the object's surface. In the special circumstance where the surface has a complicated structure, then other registration methods<sup>3</sup> may be used that do not require fiducial markers.

Our method uses light projected patterns instead of the physical fiducials. These fiducial patterns are

---

<sup>1</sup>Pascal Fua and Vincent Lepetit, "Vision Based 3D Tracking and Pose Estimation for Mixed Reality." *Emerging Technologies of Augmented Reality: Interfaces and Design*. Edited by Michael Haller, Mark Billinghurst, Bruce Thomas; University of Canterbury, Idea Group Inc. (2007) : 1-22.

---

<sup>2</sup> P. J. Besland and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 14, no. 2, pp.239-256, February 1992.

<sup>3</sup> Joaquim Salvi, Elisabet Batlle, Carles Matabosch and Xavier Llado, "Overview of surface registration techniques including loop minimization for three-dimensional modeling and inspection," *Journal of Electronic Imaging*, 17(3) (2008): 031103-1.

projected by an array of projectors surrounding the object. A Structured Light Illumination (SLI) scanner, consisting of a color digital camera, a digital projector and a video camera is used to capture a series of patches which include depth, a color texture photograph and a color photograph of the marker projection. The video camera and projector scan the depth using the well known Phase Measuring Profilometry (PMP).<sup>4</sup> A high resolution color digital camera is used to capture the marker pattern and the surface texture image. The fiducial projection, pattern projections and camera captures are sequenced by a computer. For our demonstration, the CSS Alabama artifact was small so we rotated the object along with an array of fiducial projectors and kept the scanner relatively in place but varied the height and elevation angle of the scanner to obtain the surround scan. The resulting patches are first manually merged based on alignment of the fiducial patterns. A finer manual alignment is then performed on correspondence of the local patch depth variation. A series of 4x4 coordinate transformation matrices are generated for parent/child pairing. Using graph theory, these are combined into a set of transformations to a common reference frame. The patches are transformed into a point cloud. Points are kept or eliminated based on predicted quality and redundancy using a new voxel grid mapping method.<sup>5</sup>

<sup>4</sup> Veera Ganesh Yalla and L.G. Hassebrook, "Very-High Resolution 3D Surface Scanning using Multi-Frequency Phase Measuring Profilometry," Edited by Peter Tchoryk, Jr. and Brian Holz, *SPIE Defense and Security, Spaceborne Sensors II*, Orlando, Florida, Vol. 5798-09, pp 44-53 (2005).

<sup>5</sup> Delicia Siaw-Chiing Woon, Laurence G. Hassebrook, Daniel L. Lau, and Zhenzhou Wang, "Implementation of Three Dimensional Linear Phase Coefficient Composite Filter For Head Pose Estimation," *Automatic Target Recognition XVI, SPIE Defense and Security Symposium*, edited by Firooz A. Sadjadi, Orlando, Florida. Vol. 6234, pp 62340I-1 to 62340I-12 (April 2006).

We present the acquisition methodology in section 2, the merging process in section 3, the texture mapping in section 4, point reduction algorithm in section 5, performance in section 6 and conclusions in section 7.

## 2 Acquisition Methodology

The artifact is from the CSS Alabama and is shown in Fig. 1. The artifact is primarily a metal fragment covered concretion. The artifact is held in a 5 point clamp shown in Figs. 1 and 2. The bottom point of the artifact is resting on top of a center point and surrounded by the 4 remaining clamp fingers. The clamp ends are Styrofoam covered with black felt. The Styrofoam distributes the pressure on the contact regions and the felt minimizes the reflected light. The artifact was mounted so that any 3 adjacent finger clamps would hold the artifact, so that the 4<sup>th</sup> finger clamp could be pulled away to reveal a unobstructed surface to be scanned.



Figure 1: The artifact is a metallic piece of the CSS Alabama.

As shown in Fig. 2, the artifact and 5 point clamp, along with 5 digital projectors are mounted on a turntable that is manually rotated. The digital projectors are used to project the alignment grids onto the artifact. There are 4 projectors that surround the artifact along its sides and then one projector is positioned above to project an alignment pattern onto the top of the artifact. The projectors are controlled by an Nvidia Tesla GPU system but any multi-display video graphics card may be used. An example of the alignment grid projection is shown in Fig. 4.

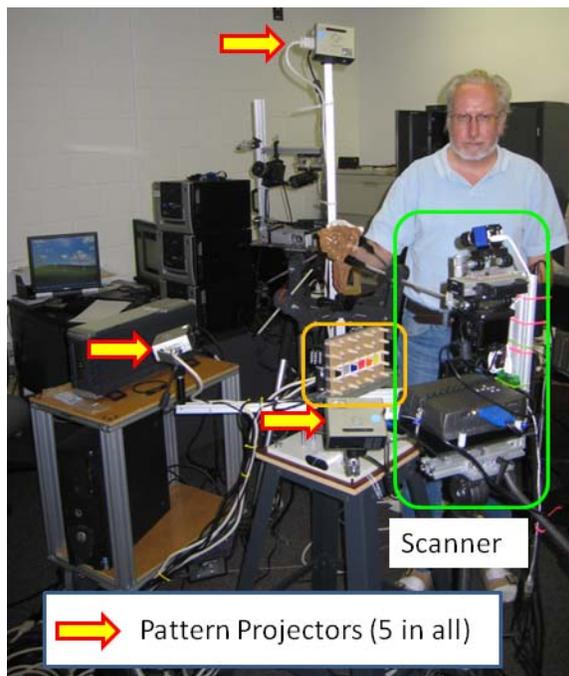


Figure 2: Scanning apparatus.

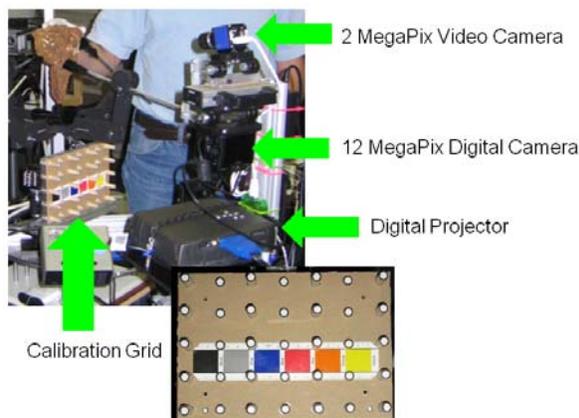


Figure 3: Scanner detail showing digital projector, video camera for depth and digital camera for color texture. A calibration grid consisting of pegs of different lengths is below the artifact.

To capture the surface depth, we use multi-frequency PMP SLI. The scanner is shown in Figs. 2 and 3. A 1600x1200 pixel MatrixVision BlueFox B&W CCD camera is used to capture the SLI information. An InFocus digital projector is used to project the PMP patterns. The color texture imagery was captured by a 4000x3000 pixel Canon G9 digital camera. Both cameras (USB) and the

projector (SVGA) were controlled by a MS XP based Dell laptop computer. The scanner was mounted on a standard tripod. The basic concept of SLI is triangulation. This concept is shown in Fig. 5 where given a single light ray from the illumination, a depth variation results in a lateral position change on the sensor. PMP uses a 2-D pattern sequence in place of the single ray of light. Thus, each pixel in the camera sensor will receive a single ray of light from a different spot on the 2-D pattern sequence. Since a sequence of patterns is presented, each camera pixel will receive a temporal signal whose signature refers back to a specific coordinate on the projector. A depth variation will laterally translate these temporal signals from one pixel location to another. However, if a high frequency pattern is used as shown in Figs. 6 and 7, then depth ambiguities will arise if the depth varies enough to shift the pattern by one or more wave lengths. The Base frequency in Fig. 6 is used to prevent depth ambiguity but because it is a low frequency, it will have a lower Signal-to-Noise Ratio (SNR) than a higher frequency. The multi-frequency PMP uses the base frequency to estimate a non-ambiguous depth which in turn is used to unwrap the “phase” of the higher frequencies, in succession.



Figure 4: Fiducial patterns remained fixed in on surface during rotation of artifact.

The scanner was calibrated using the calibration grid shown in Figs. 2 and 3. The grid was made by regularly spacing different length cylindrical pegs. The length of the pegs are accurately determined and the ends of the pegs are analyzed for the center point which is the calibration point location. The calibration process is only necessary one time. It

involves scanning the depth of the grid as well as photographing it with the digital camera. Once calibrated, the cameras and projector must remain locked in position with respect to one another but the entire unit may be re-orientated.

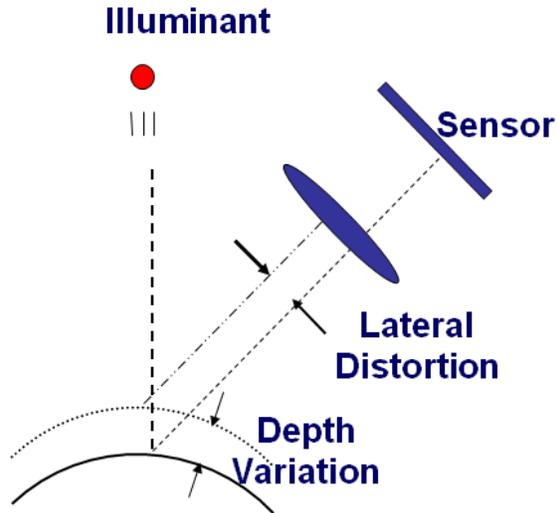


Figure 5: Structured Light Illumination triangulation.

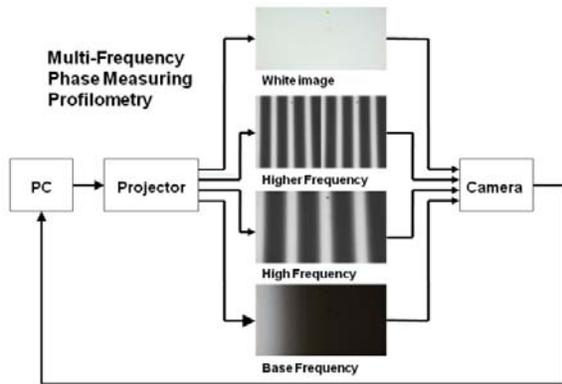


Figure 6: Multi-frequency phase measuring profilometry flow diagram.

There are four coordinate reference frames: (1) The world coordinates,  $\{x_w, y_w, z_w\}$ , in millimeters, defined by the calibration grid, (2) the video camera pixel coordinates  $\{x_c, y_c\}$ , (3) the projector pixel coordinates  $\{x_p, y_p\}$  and (4) the digital camera pixel coordinates  $\{x_d, y_d\}$ . The PMP patterns are aligned to remain constant along  $x_p$  so  $x_p$  is not used in the geometry. We use a pin hole

lens model for the cameras and projector perspective transformation from world coordinates.



Figure 7: A single PMP pattern projection.

The video camera perspective transformation equations are

$$x_c = \frac{m_1 x_w + m_2 y_w + m_3 z_w + m_4}{m_9 x_w + m_{10} y_w + m_{11} z_w + m_{12}} \quad (1)$$

and

$$y_c = \frac{m_5 x_w + m_6 y_w + m_7 z_w + m_8}{m_9 x_w + m_{10} y_w + m_{11} z_w + m_{12}} \quad (2)$$

where  $m_{12} = 1$ . The digital camera has the same transformation equations but with different coefficient values. The projector transformation is

$$y_p = \frac{m_{p5} x_w + m_{p6} y_w + m_{p7} z_w + m_{p8}}{m_{p9} x_w + m_{p10} y_w + m_{p11} z_w + m_{p12}} \quad (3)$$

and traditionally scaled to be in units of radian phase rather than projector pixel location.

The video camera is used to obtain a phase value for each camera coordinate. The inverse equations exist such that given  $\{x_c, y_c, y_p\}$ , a unique world coordinate,  $\{x_w, y_w, z_w\}$ , is determined for each valid camera pixel location.

The scanning procedure went as follows:

1. Position and fix scanner on tripod.
2. Rotate artifact into position. Fold back clamp fingers if necessary.
3. Project alignment patterns.
4. Capture alignment pattern with digital camera. Digital camera flash turned off.
5. Turn off alignment patterns.
6. Perform multi-frequency PMP scan with projector and video camera.
7. Capture texture photograph using the digital camera with flash on.
8. For rotation of object, repeat starting at step 2. After each rotation sequence raise or lower scanner and repeat starting at step 1. Else end process.

The scanning is performed in a dark room for maximum SNR. A total of 21 patches are captured at three different elevation angles.

### 3 Merging Methodology

A traditional approach to merging is to use an Original Equipment Manufacturer (OEM) software application to merge the patches captured by an OEM scanner. In contrast, our group has a track record of research and development of our own scanner technology.<sup>6,7</sup> Our system is software based, is independent of camera and projector technology and can be configured with multiple cameras and projectors. Our long term strategy<sup>8</sup> is to be able to acquisition a recording of an environment both spatially and temporally. Our motivation for investigating merging, rather than using OEM applications, is its eventual integration with the data acquisition process.

<sup>6</sup> Hassebrook et al, “3-Dimensional Gallery,” University of Kentucky, 5-D Studio, [www.engr.uky.edu/~lgh/data/data.htm](http://www.engr.uky.edu/~lgh/data/data.htm)

<sup>7</sup> FlashScan3D, LLC, “3-D fingerprint Data,” Home page, [www.flashscan3d.com/](http://www.flashscan3d.com/)

<sup>8</sup> Hassebrook et al, “3-D Data Acquisition,” Center for Visualization and Virtual Environments, [www.vis.uky.edu/3ddataacq.php](http://www.vis.uky.edu/3ddataacq.php)

Using the setup described in the previous section, we captured 21 patches numbered from 0 to 20. The patches are paired into a “parent-child” (PaCh) list. The PaCh list allows evaluation of forming the transformation matrices used to map all the patches into a single reference frame. The original PaCh relationships are shown in Fig. 8. The green arrows (ie., arrows connecting adjacent scans taken in sequence) have their arrow heads pointing to the parent and their tails connecting to the child patch.

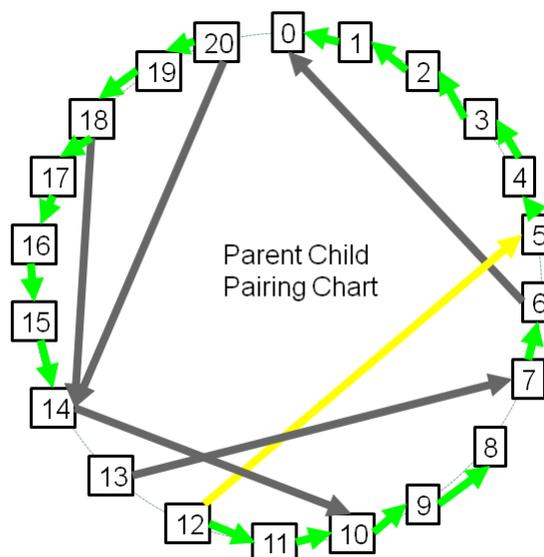


Figure 8: Parent/Child graph. Arrows indicate pairings of scans 0 through 20.

Note that in Fig. 8, there are gaps in the green connections between {5 & 6}, {7 & 8}, {12 & 13}, {13 & 14}, and {20 & 0}. These gaps are usually associated with an elevation angle change in the scanner followed by a series of scans where the only rotation stage is moved between scans. Thus the patches overlap. If only the green arrow pairs were used, then we could not transform all of patches back to the same reference frame of patch 0. To resolve these gaps other non-adjacent PaCh pairs are formed. These pairs are indicated by the dark gray and light yellow arrows in Fig. 8. It is possible to produce scans where it is impossible to determine transformations for all patches that will bring them to the same reference frame. The question is how to verify that all patches can be transformed to the same reference frame? To solve this problem we formed a PaCh pairing graph as shown in Fig. 9. The graph is in the form of a matrix where the column indices correspond to the

parent indices and the row indices correspond to the child indices. A row, column location indicates a specific pairing. The {0,0} location is the first patch paired with itself. This results in a trivial transformation that does not change the “child” reference frame. The first green pair, located at {1,0}, will allow patch 1 to be transformed to the reference frame of patch 0. The reader can sequence through the pairings in Fig. 8 and see their locations in Fig. 9. There are a total of 23 pairing. Given the pairings indicated by the green locations in Fig. 9, we automated the determination of the transformations necessary. We can refer to these as the global transformations and they correspond to the first column in pairing matrix. Thus, the goal is to determine the transformations associated with the first column of pairing graph in Fig. 9. In order to reach that goal, many other pairings must be found.

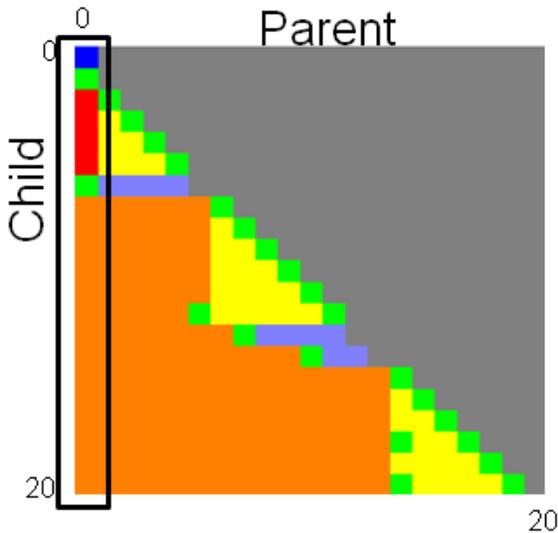


Figure 9: Pairing graph insures that all scans are transformed properly.

The coordinate transformation is based on well known affine transformations implemented by multiplication of 4x4 transformation matrices. For example, given one child world coordinate, an input vector is formed such that

$$\mathbf{p}_{child} = [x_{wc} \quad y_{wc} \quad z_{wc} \quad 1]^T \quad (4)$$

where the 4<sup>th</sup> element is unity and used to implement translation to the parent coordinate frame. The 4x4 transformation matrix is given by

$$\mathbf{A}_{n,m} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & T_{14} \\ a_{21} & a_{22} & a_{23} & T_{24} \\ a_{31} & a_{32} & a_{33} & T_{34} \\ d_{41} & d_{42} & d_{43} & d_{44} \end{bmatrix} \quad (5)$$

The “a” elements are used to scale and rotate the input coordinate. The “T” elements translate the input coordinate and “d” are used to carry perspective information. If the output coordinate is represented in vector form as

$$\mathbf{p}_{parent} = [x_{wp} \quad y_{wp} \quad z_{wp} \quad w]^T \quad (6)$$

Then a PaCh pair transformation of a single child coordinate point is written in matrix form as

$$\mathbf{P}_{parent} = \mathbf{A}_{p,c} \mathbf{p}_{child} \quad (7)$$

where “p” is the parent patch index and “c” is the child patch index. For example, given a series of adjacent PaCh pairs, the global transformation matrix from patch “n” to patch “0” may be given by the matrix multiplication as

$$\mathbf{A}_{0,n} = \mathbf{A}_{0,1} \mathbf{A}_{1,2} \cdots \mathbf{A}_{n-2,n-1} \mathbf{A}_{n-1,n} \quad (8)$$

So the global transformation would be

$$\mathbf{P}_0 = \mathbf{A}_{0,n} \mathbf{p}_n \quad (9)$$

Eq. (9) is performed on all the points of the n<sup>th</sup> patch thereby transforming it to the reference frame of patch 0. The objective is shown in Fig. 10.

The key concept to solving the pairing graph is to be able to find a path that yields Eq. (8) for n=0, 1, ... (N-1) where N is the total number of unique patches. where the indices of the transformation matrices are in order such that

$$\mathbf{A}_{0,n} = \mathbf{A}_{0,a} \mathbf{A}_{a,b} \cdots \mathbf{A}_{g,h} \mathbf{A}_{h,n} \quad (10)$$

For example, consider the green element at column 0, row 6 and the green element at column 6, row 7. These element locations satisfy Eq. (10) so that the

transformation from patch 7 to patch 0 can be determined by

$$\mathbf{A}_{0,7} = \mathbf{A}_{0,6} \mathbf{A}_{6,7} \quad (11)$$

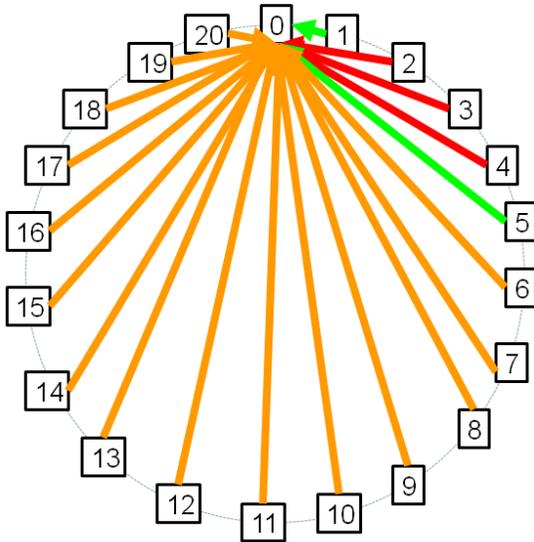


Figure 10: Scans are transformed to same reference frame as scan 0.

How are the PaCh transformation matrices, Eq. (5) found? There are several methods for estimating the PaCh transformation matrices which include both automated and/or manual processes. We chose to investigate this problem by first manually determining control points. That is, for each parent-child pair, a set of control points is found by manual inspection. The first pass is performed at a lower resolution on the images with the alignment patterns. The second pass is in full resolution and the surface is stripped of color and textured with a metallic surface to reveal more depth detail.

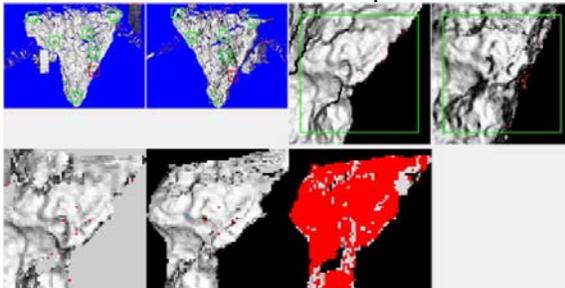


Figure 11: Manual control point matching interface.

A screen shot from our application is shown in Fig. 11. The upper left section shows the parent child pair. The rectangles within these images are the regions used to determine each control point. One of control point pairs is selected and shown in the upper right images. Using all the control points, a transformation matrix is determined and the resulting parent and transformed child control point region are shown in the lower left pair of images in Fig. 11. The lower right image is mostly red indicating the alignment of the parent and child region is very close in world coordinate space.

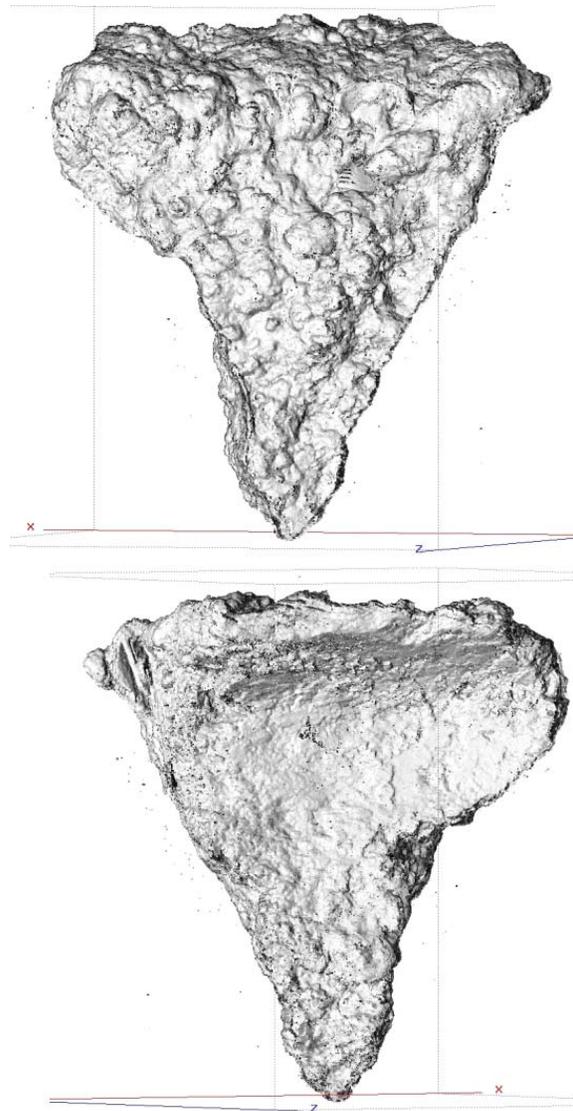


Figure 12: Merged patches. (top) Convex side of artifact. (bottom) Concave side of artifact.

Once all the global transformations are determined, the patches are all transformed to the patch 0 coordinate reference frame. Two views of the resulting point cloud are shown in Fig. 12. Top view is the convex side of the artifact. The concretion on the convex side is very coarse and contains organic deposits. The opposite side is concave and shown in Fig. 12 bottom. It is notably smoother and the concretion is primarily composed of oxidation of the metal.

## 4 Texture Mapping

The depth scan contains a non-color texture image. The color digital camera has significantly more resolution than the B&W video camera. Our objective to mix the resolutions via texture mapping, commonly called “skinning.” There are three approaches that we implemented: (1) Skin the depth data with the color data at the depth resolution, (2) skin the color data with the depth data at the color resolution and (3) up sample the depth data to approximately the color data resolution and then skin the upsampled depth data with the color data.

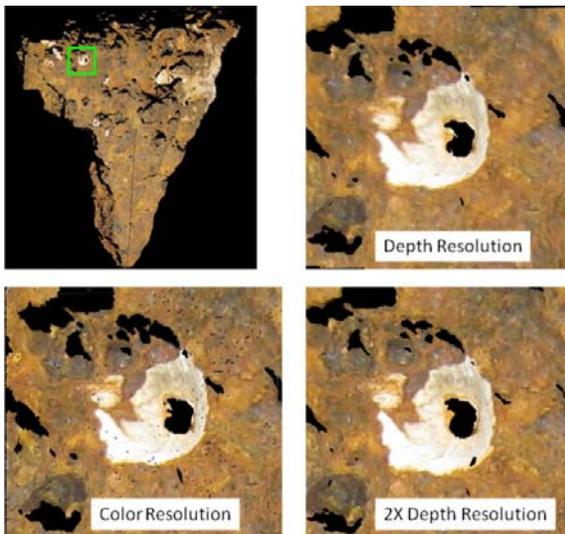


Figure 13: Skinning samples. (upper left) Cropped region. (upper right) Depth resolution. (lower left) Color texture resolution. (lower right) 2X depth resolution.

For the cameras we used, the depth data is captured at 1600 x 1200 and the color texture is captured at 4000 x 3000. In terms of total pixels depth and

color are ~2MegaPixel and 12 MegaPixel, respectively. However, this does not mean the pixel density is 1 to 6 because the Field of Views (FOVs) are not the same. The zoom control of the color camera is discrete so the actual pixel density ratio was 1 to 4 for total depth to color pixels. A pixel density 4 times higher is equivalent to upsampling the depth by 2X along the row and column direction. The data is stored in what we call the MAT5 format. There are 5 images all indexed in camera space, as defined in Eqs. (1) and (2), more commonly referred to as “UV” space. The first image is the texture or albedo image, the second is an 8 bit per pixel quality map and the last 3 images are X, Y and Z coordinate images. The significance of the UV space is that no matter how the X,Y,Z values are transformed, the points are not moved in UV space, so relative point position is preserved.

We present 3 texture mapping algorithms that are implemented in this research. The examples are shown in Fig. 13 for the cropped region of the upper left subimage.

### 4.1 Color to Depth Map

To map texture to the depth data, the depth data is indexed across the “UV” or camera coordinates. For each  $\{x_c, y_c\}$  camera coordinate, we have the  $\{x_w, y_w, z_w\}$  world coordinate stored in the X,Y,Z images located at the  $\{x_c, y_c\}$  element. Given the world coordinate we obtain the color texture UV coordinate using Eqs. (1) and (2) with the digital camera coefficients. The resulting digital camera coordinate,  $\{x_d, y_d\}$  will be fractional valued. For simplicity we rounded the coordinates to the nearest integer.

An alternative to nearest integer is to use a bilinear transformation of neighboring color pixels. We leave this approach for future implementation. A color mapped example of the result is shown in Fig. 13 upper right. The depth resolution can be observed by a close up of the point cloud as shown in Fig. 14 left.

### 4.2 Depth to Color Map

To map depth to texture data, the depth data is interpolated to the higher digital camera resolution.

However, we do not have a direct way to map from the color texture to either the depth UV space or world coordinates. Instead, the video camera's depth UV space is indexed in groups of 4 adjacent pixels forming a square region in the depth UV space. Each of the four corners the world coordinates are mapped to the associated points in the digital camera's color texture UV space. In the digital camera UV space, these point locations are no longer adjacent or in a square formation.

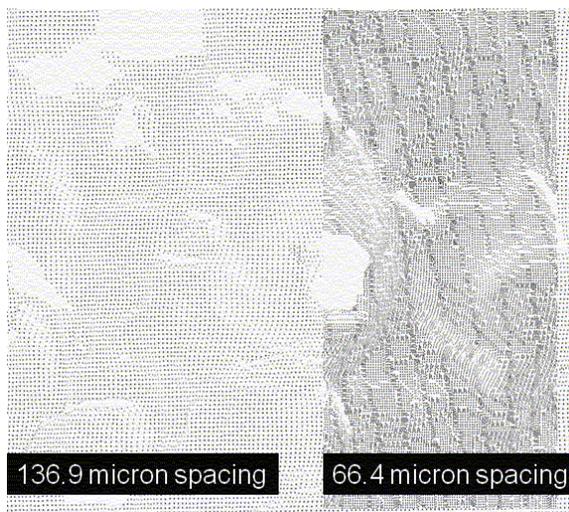


Figure 14: (left) Depth resolution. (right) Color resolution.

Because the digital camera UV space is higher resolution than the video cameras UV space, the four point locations form a convex quadrangle that probably contains more than one color texture points. The UV space of the color texture is bilinearly interpolated based on the quadrangle corner values. The results are shown in Fig. 13 lower left and Fig. 14 right. The color texture is significantly less blurred than the depth resolution method in the previous subsection. However, there are some holes and irregular spacing in the UV coordinates as shown in these images. We believe this is due to round off error and leave its correction to future research.

### 4.3 Color to Upsampled Depth

From Fig. 14 we see that the lateral spacing of the depth data is about twice that of the color. Given the complexity of the algorithm described in subsection 4.2, it becomes apparent that a much

simpler texture mapping approach is to upsample the depth by interpolating in between coordinates values and then applying the algorithm described in subsection 4.1 to the higher density depth samples. We upsampled by 2 in both the U and V direction. An example of the result is shown in Fig. 13 lower right. There are no pin holes as in the Depth to Color algorithm but is significantly less blurred than the algorithm described in subsection 4.1. The sample density is shown in Fig. 15.

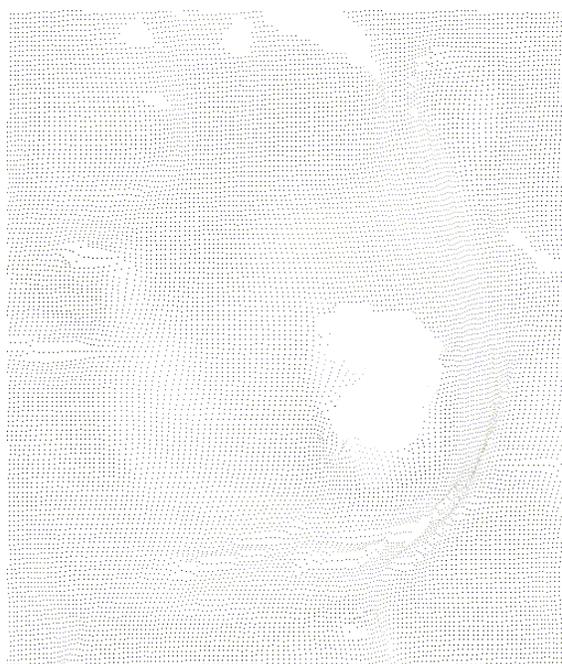


Figure 15: 2X upsampled, 68.5 micron spaced depth samples.

## 5 Grid Map Point Reduction

Grid mapping in 3-Dimensions is the process of spatially quantizing the world coordinates of points in a point cloud to a finite grid. The result is a set of 3-D rectilinear volume elements representing the points in the original point cloud. If the points are close enough, they are spatially quantized to the rectilinear volume element. These volume elements are called “voxels” which are the 3-D equivalent of 2-D pixels. Voxel grid maps can be used for a variety of applications from rendering to detection and pose estimation.<sup>5</sup> Most applications have a numerical advantage over the original point

cloud form, primarily due to the quantization of the space.

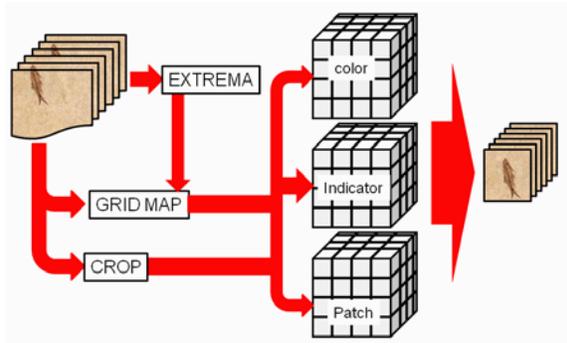


Figure 16: Point reduction using a voxel grid map. “Indicator” refers to quality value.

We used a voxel map to efficiently perform point reduction of the merged patches. We trade I/O operations and memory storage for numerical efficiency of the point reduction process. The algorithm is shown in Fig. 16.

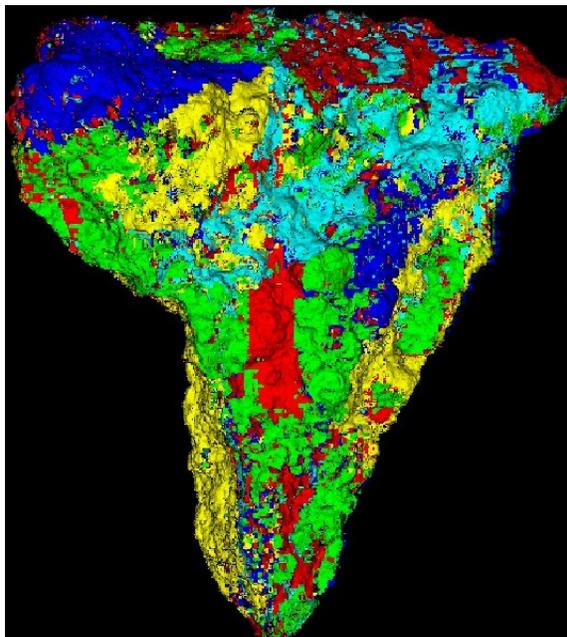


Figure 17: Colorized voxel mapping of artifact.

Given a set of patches stored on the hard drive, they are processed sequentially and their world coordinate extrema are determined. If all of these patches were read into memory simultaneously, the available memory may be exceeded. From the

extrema values, a grid map is allocated. On the second I/O pass, the grid map is treated as a 3-D Look Up Table (LUT). Each point in a patch is linearly mapped to a voxel. If the point has a higher quality value than previous points then its patch number and quality value will be assigned to a given voxel. Once all the voxels are tagged with the quality value and the dominant patch number, a third I/O pass is made of the data. Each valid point in the patch is mapped to the voxel grid and if its patch number matches that of the voxel, then that point is preserved. If its patch number does not match then it is removed from the patch by setting its quality to 0.



Figure 18: Voxel grid map with color texture.

There are several ways to weight the quality values. The primary way is during the scan process, its quality is determined in terms of SNR. Any points with a quality below a minimal threshold are discarded. Those remaining may be re-weighted to enhance the merging process. For example, we may want to preserve the continuity of patches so if one patch contains more adjacent pixels than another, then that patch is the preferred one. We processed our data with a  $128^3$  voxel cube and colored each pass differently. The result is shown in Fig. 17. Fig. 18 shows the same voxel grid mapping as Fig. 17 but with the original color texture.

In using a voxel map, if the voxels are too big, then patches will not fill in holes formed from shadowing. If the voxels are too small then redundant points will not be removed. We used cubical voxels and chose their size by trial and error. The result is shown in Fig. 18.

## 6 Performance

There are several performance aspects to this research. In many ways, this research is the system integration of the scanning process with the merging process. The scanning process took about 8 person hours to perform once the apparatus was constructed. The standard deviations (STDs) in millimeters are  $\sigma_x=0.295$ ,  $\sigma_y=0.247$ ,  $\sigma_z=0.571$  and are determined from the calibration grid shown in Fig. 3. The STD in the Z direction is about double the X and Y STD because the triangulation angle was about 30 degrees. These values are about 6 times higher than STD values for other scans that our group has performed. We believe the large STDs are due to errors in the construction of the prototype calibration grid.

The merging took about another 8 hours and the post data processing took about 20 minutes. Consider the manual merging in two stages, the course merge followed by the fine merge.

The first stage consisted of setting the control points of down-sampled scans with the fiducial markers as shown in Fig. 4. Because we were developing the software during this process, we were not able to get a statistical evaluation of this part of the process but our estimate is about 2 to 3 minutes per PaPh pair with 10 control points per pair. For 23 pairs, this part of the process is estimated to take about 1 hour. Although we can't say how much faster this would be with more experience, we believe that the process would speed up considerably, in part by better set up of the fiducial projection apparatus.

The second stage consisted of refining the control point positions at full video camera depth resolution. In doing this fine merge, it was advantageous to interchange the view the surface between a fiducial marker, metallic or color texture. The second stage was psychologically tedious and took about 6 hours. In Table 1, the

alignment of the PaCh pairs is shown in root mean squared (RMS) error with units of millimeters. As shown in the table, there were 21 patches combined in 23 pairs. The  $RMS_1$  measure was determined after the second stage was completed.

Table 1: Manual control point merge accuracy.

Pair	Parent	Child	$RMS_1$ (mm)	$RMS_2$ (mm)
0	0	1	0.98	0.81
1	1	2	0.31	0.28
2	2	3	0.21	0.18
3	3	4	1.43	0.33
4	4	5	0.69	0.49
5	5	12	1.71	0.31
6	0	6	0.72	0.55
7	7	8	0.53	0.42
8	8	9	0.87	N/A (N=4)
9	9	10	0.61	0.53
10	10	11	0.68	0.50
11	11	12	0.41	0.37
12	7	13	2.27	0.12
13	14	15	0.77	0.65
14	15	16	0.61	0.51
15	16	17	0.36	0.33
16	17	18	0.73	0.61
17	18	19	0.57	0.43
18	19	20	0.45	0.41
19	14	20	1.13	0.77
20	14	18	0.54	0.44
21	6	7	0.39	0.36
22	10	14	1.22	0.98

Note that pairs 3, 5, 12, 19 and 22 have error values greater than 1 mm. We hypothesized that these pairs had at least one misaligned control points. So we implemented a "backward elimination" algorithm to determine which control points added the most error for a given set. For  $N$  control points, 1 point would be dropped and the RMS error would be calculated. In this way, the point that resulted in the least decrease in RMS is identified. We then removed that control point and recalculated the RMS error. Pair 8 only had 4 control points so it was not possible to perform this algorithm because there had to be at least 5 points. The resulting RMS improved all other points as indicated by  $RMS_2$  in Table 1. This indicates that

human error is significant in the manual control point alignment process that we used.

The post-processing was fully automated and took a total of about 20 minutes for the coordinate transformation and point reduction of the 21 patches.

## 7 Conclusions

We presented a non-contact fiducial based patch alignment where the fiducials were projected onto the target object with light based pattern imaging. Integrated with the fiducial projection, is a SLI based 3-D surface scanning process. In the past, our group has implemented or developed various automated merging techniques which were very surface dependent. To reduce the surface dependencies, the merging in this research was limited to manual operations only. We also introduced a voxel grid mapping technique for point reduction of the merged patches.

For future research more experimentation is needed to establish objective statistical performance measurements. However, our subjective evaluation is that the fiducial pattern projection is very promising. We found it very time consuming to manually merge the patches without the use of the pattern markers.

## Acknowledgements

We would like to thank the University of Kentucky Center for Visualization and Virtual Environments, Department of Electrical and Computer Engineering and Blazie Professorship fund for support of this work. We thank Benjamin Rennison of the Clemson Conservation Center for providing the access to the artifact. We would like to thank Christopher Hassebrook for his assistance in the scanning process.

## Bibliography

1. Pascal Fua and Vincent Lepetit, "Vision Based 3D Tracking and Pose Estimation for Mixed Reality." *Emerging Technologies of Augmented Reality: Interfaces and Design*. Edited by Michael Haller, Mark Billinghurst, Bruce Thomas; University of Canterbury, Idea Group Inc. (2007) : 1-22.
2. P. J. Besland and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 14, no. 2, pp.239–256, February 1992.
3. Joaquim Salvi, Elisabet Batlle, Carles Matabosch and Xavier Llado, "Overview of surface registration techniques including loop minimization for three-dimensional modeling and inspection," *Journal of Electronic Imaging*, 17(3) (2008): 031103-1.

We also found that the fine merging was not performed accurately enough for this application. We believe that we would need about 6 times more accuracy to achieve a merging accuracy comparable to the point spacing of the individual patches. In order to do this, application dependent automated algorithms should be incorporated into future efforts.

The texture mapping algorithm performed quite well. Not shown in this research, were measurements of depth features locations versus color texture features. We could not determine measurable errors from the artifact data. The accuracy is high enough to require the use of a test grid designed for evaluating the texture mapping in future research.

The voxel grid mapping algorithm was introduced. Unfortunately, its performance is dependent on the merging accuracy so it was not evaluated. Numerically, the voxel grid map is equivalent to a LUT and has the same high efficiency since it does not require calculations. The cycle time is dominated by I/O speed related to sequencing through the patches.

- 
4. Veera Ganesh Yalla and L.G. Hassebrook, "Very-High Resolution 3D Surface Scanning using Multi-Frequency Phase Measuring Profilometry," Edited by Peter Tchoryk, Jr. and Brian Holz, *SPIE Defense and Security, Spaceborne Sensors II*, Orlando, Florida, Vol. 5798-09, pp 44-53 (2005).
  5. Delicia Siaw-Chiing Woon, Laurence G. Hassebrook, Daniel L. Lau, and Zhenzhou Wang, "Implementation of Three Dimensional Linear Phase Coefficient Composite Filter For Head Pose Estimation," *Automatic Target Recognition XVI, SPIE Defense and Security Symposium*, edited by Firooz A. Sadjadi, Orlando, Florida. Vol. 6234, pp 62340I-1 to 62340I-12 (April 2006).
  6. Hassebrook et al, "3-Dimensional Gallery," University of Kentucky, 5-D Studio, [www.engr.uky.edu/~lgh/data/data.htm](http://www.engr.uky.edu/~lgh/data/data.htm)
  7. FlashScan3D, LLC, "3-D fingerprint Data," Home page, [www.flashscan3d.com/](http://www.flashscan3d.com/)
  8. Hassebrook et al, "3-D Data Acquisition," Center for Visualization and Virtual Environments, [www.vis.uky.edu/3ddataacq.php](http://www.vis.uky.edu/3ddataacq.php)
-