Available online at www.sciencedirect.com
ScienceDirect
Computer Vision and Image Understanding 105 (2007) 30–41
Computer Vision and Image Understanding
www.elsevier.com/locate/cviu

A volumetric fusion technique for surface reconstruction from silhouettes and range data ☆

Y. Yemez *, C.J. Wetherilt

Multimedia, Vision and Graphics Laboratory, College of Engineering, Koç University, SarIyer, Istanbul 34450, Turkey

Received 21 April 2006; accepted 31 July 2006
Available online 15 September 2006

Abstract

Optical triangulation, an active reconstruction technique, is known to be an accurate method but has several shortcomings due to occlusion and laser reflectance properties of the object surface, that often lead to holes and inaccuracies on the recovered surface. Shape from silhouette, on the other hand, as a passive reconstruction technique, yields robust, hole-free reconstruction of the visual hull of the object. In this paper, a hybrid surface reconstruction method that fuses geometrical information acquired from silhouette images and optical triangulation is presented. Our motivation is to recover the geometry from silhouettes on those parts of the surface which the range data fail to capture. A volumetric octree representation is first obtained from the silhouette images and then carved by range points to amend the missing cavity information. An isolevel value on each surface cube of the carved octree structure is accumulated using local surface triangulations obtained separately from range data and silhouettes. The marching cubes algorithm is then applied for triangulation of the volumetric representation. The performance of the proposed technique is demonstrated on several real objects.

Keywords: Surface reconstruction; Shape from silhouette; Shape from optical triangulation; Volumetric fusion; Volume carving; Isosurface merging

1. Introduction

Today the art of inferring three-dimensional shapes of objects has become one of the major applications of computer graphics. There are many areas concerned with 3D reconstruction among which virtual reality applications, digital preservation of cultural heritage, machine vision, medical imaging are the most common. In general, 3D reconstruction methods can be collected under two groups: active and passive. Active methods make use of calibrated light sources such as lasers or coded light most typical example of which is the shape from optical triangulation method. Passive methods on the other hand, extract surface information by the use of 2D images of the scene. Among the most common that fall into this category are the techniques known as shape from silhouette, shape from stereo, and shape from shading. Many results are available concerning reliable reconstruction of objects using these methods. However, there is still a need for improved reconstructions since each specific method, active or passive, has its own drawbacks and deficiencies.

Shape from silhouette basically draws shape information by back-projecting multiple silhouettes into the world space as conical volumes and intersects these cones to generate a volumetric visual hull. Early examples of this technique were presented in [1] and later much improvement has been established concerning accuracy, and efficiency issues [2–5]. In general, the strength of the technique lies in its simplicity, robustness and accuracy especially when applied to convex shapes. The drawback of this method

☆ This work has been supported by the European Network of Excellence 3DTV (www.3dtv-research.org).
* Corresponding author. Fax: +90 212 3381548.
E-mail addresses: yyemez@ku.edu.tr (Y. Yemez), cwetherilt@ku.edu.tr (C.J. Wetherilt).

1077-3142/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2006.07.008

is that it fails to capture hidden concavities. However, the robust output of this method constitutes a solid initial foundation for further volume carving or surface deformation by incorporating other available cues with silhouette information.

Methods of shape from stereo seek to find correlations on separate images using texture or color information. This feature makes such techniques very sensitive to lighting conditions and renders them less effective as stand-alone methods. Several researchers as in [6,7] fuse shape from silhouette and shape from stereo in a volumetric fashion while others as in [8,9] adhere to deformation models for further enhancing the description of the object mesh initially obtained from silhouettes. Shape from shading methods, on the other hand, are based on the diffusing properties of Lambertian surfaces. They require controlled environments where the illumination of the object space and the object reflectance must be known. Textured objects pose severe problems and the proposed techniques are found to be mathematically unstable. In [10], shape from shading and shape from silhouette techniques are integrated by volumetric space carving, but the results show that the obtained reconstructions can not compete with the accuracy of active methods.

Shape from optical triangulation, as an active reconstruction technique, produces accurate point clouds. However, due to inherent camera and light occlusions, it poses problems in building complete and watertight reconstructions. In [11] several laser projectors are integrated to reduce light occlusions. In [12], next view planning techniques to optimize surface coverage are considered. In [13], large-scale enhanced acquisition systems have been used to overcome the occlusion problem presenting very accurate and successful results. In spite of all these enhancement efforts, and although some scanners perform better than others in generating more complete surfaces, final surface reconstructions for some objects always contain holes, no matter what kind of scanning configuration is implemented or how many scans are run. Depending on the material of the object, the projected light may get scattered or reflected from the object surface, which may further increase the proliferation of holes. The use of hole filling algorithms [14], multiple laser sources or cameras, and even mirrors [15] help improve the reconstructions but some portions of the object surface such as the inner walls of hollow parts are at best unreliably reconstructed. The problem of integrating aligned images while producing hole-free reconstructions has been addressed in [16] and [17]. The work in [16] is based on the design of a real-time acquisition system allowing to scan objects faster, with greater ease and with better coverage than conventional model acquisition pipelines. In [17], a volumetric technique for hole-free surface reconstruction from optical triangulation is presented. Although the proposed technique generally produces satisfactory results, it does not yield reliable reconstructions on parts of the object surface, that are visible by the sensor only at sharp angles due to occlusions. To overcome this problem, the authors of [17] point out the possibility of shaping such difficult parts of the object by carving via background extraction, which is in fact a simple variation of fusing shape from optical triangulation with shape from silhouette. Nevertheless, they do not further elaborate on this possibility and do not incorporate the silhouette information into their reconstruction scheme.

The problem of fusing shape from silhouette and shape from optical triangulation can be approached in several ways. Surface-based fusion with employment of model deformation frameworks is common for systems containing different kinds of information about the surface of the reconstructed object [8,18]. As an alternative, volumetric fusion can be favored to avoid topological problems that are likely to arise should a surface based fusion be applied. The silhouette model that assumes a solid and sound framework can be carved away volumetrically at the hidden hollow sections by the more precise cloud of range points to be followed by a triangulation process.

The work presented in [19] attempts to combine shape from silhouette and shape from optical triangulation for object volume reconstruction, but does not address the surface reconstruction problem and fails to produce clear results. The acquired range images are first converted to binary images so that the white pixels belong ''possibly'' to the object volume and the black pixels correspond to the background. The binary images are then directly fused with the silhouette information on an octree. This yields a fast algorithm for volumetric fusion, which is however liable to various topological problems. The paper does not address these problems and the presented method seems to be an approximate and ad-hoc technique that can not be used for volume reconstruction of complex objects. Another attempt to combine shape from silhouette and shape from optical triangulation is the one presented in [20] which proposes a volumetric technique for carving of the initial silhouette model with range data. However the voxels to be carved out can only approximately be determined with the proposed method and the authors do not describe how they handle topological problems that might arise during the volume carving process. Moreover, the fusion technique does not address the isosurface merging problem since it does not take into account the range surface normal information. As a result, the accuracy of the obtained surface reconstructions becomes strictly limited to the voxel resolution.

In this work we propose a novel, hybrid surface reconstruction scheme that combines the two techniques, shape from silhouette and shape from optical triangulation. The proposed scheme aims at eliminating the shortcomings of the combined methods and enabling the generation of high quality, robust, and complete 3D models. Without any prior assumptions about the object shape or orientation, the data acquired from the two techniques, are fused volumetrically and triangulated into a final mesh while preserving object topology. We are not aware of any published

methods for combining shape from silhouette with shape from optical triangulation that both underpin its feasibility as a successful method and yield satisfactory results. The main contribution of this work is (i) to show that it is possible to fuse silhouette-based and optical triangulation based methods, (ii) to provide examples of fusions conducted in this manner, and (iii) to demonstrate the improvements of the fused results over the results obtained from the individual methods.

The paper is organized as follows. We describe our hardware setup to collect range and silhouette data in Section 2 where we also address the preliminary steps before fusion, that are camera calibration and silhouette extraction. Sections 3 and 4 describe the shape from silhouette and shape from optical triangulation components of our fusion system, respectively. In Sections 5 and 6, we present the volume carving and the surface triangulation procedures that constitute our fusion scheme. Section 7 displays the experimental results and Section 8 provides concluding remarks.

2. Acquisition setup

Our setup consists of a CCD camera, a turntable, laser projectors, a backdrop for silhouette extraction, and an acquisition computer. The computer synchronizes the camera with the motion of the turntable and records the acquired data. The position of the turntable, which supports only rotational movement around a vertical axis, can be set to an angular precision of one degree. The object to be reconstructed is positioned on the turntable between the static camera and the backdrop. The laser projectors are pointed towards the object (see Fig. 1). A single laser projector yields a planar laser beam which appears as a stripe on the object surface. The laser projectors are positioned so that the projected beams are as vertical to the rotational axis of the turntable as possible. Two laser projectors are used to improve the coverage of the object surface. Also, the same laser projector may be used several times in different orientations during the same acquisition to obtain several scans for optical triangulation. The different laser projectors are not simultaneously used because of the difficulties involved in distinguishing the beams in discontinuous or occluded areas of the object. The laser projected images are acquired in the dark to discern the laser beams more easily. During the acquisition process, the laser projectors and the camera are maintained immobile. The only actuated component in the system is the turntable on which the object is placed.

The shape from silhouette method estimates the 3D model of the object from its silhouettes. Accurate extraction of the silhouettes is thus crucial for the quality of the reconstructed object. We employ the extraction method which was developed in [5]. This method depends on the use of a sharp contrast that must be maintained between the background and the object. The backdrop is saturated with light while the object is left in the dark creating a natural silhouette of the object. A simple thresholding then suffices to obtain the correct silhouette. This scheme works very well especially for objects with low reflectance properties. In certain circumstances, depending on the reflectance properties of the reconstructed object, the background and the object may become indistinguishable (see Fig. 2). Hence, to obtain a successful extraction, the strength of the light sources and the camera settings have to be fine-tuned. The background saturation method, when carefully tuned, produces very clean and accurate results circumventing the radiosity and color confusion problems.

Fusing the geometry information obtained from two different sources, i.e., silhouettes and laser data, in a coherent manner requires an accurate calibration of the turntable and the laser projectors with respect to the camera. For camera calibration we have used the toolbox which is publicly available in [21]. The camera calibration is

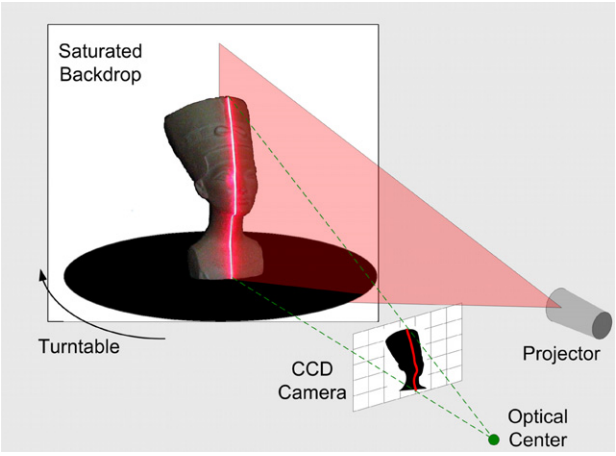

Fig. 1. Layout of the acquisition system.

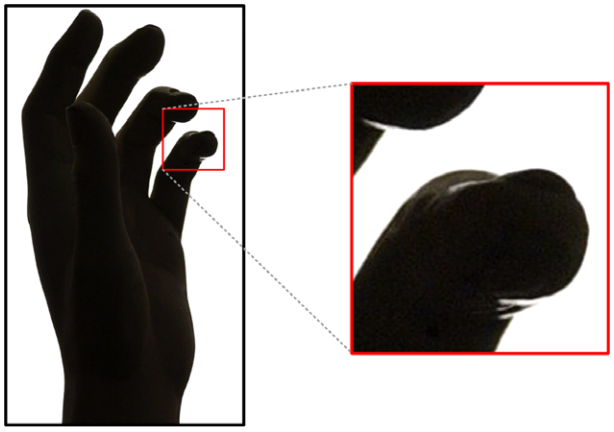

Fig. 2. Image of an object with saturated background. The zoomed section reveals the difficulty encountered sometimes in discerning the object from the background.

performed by using a special calibration object, which is basically a smooth checkerboard surface. The same checkerboard pattern is also used for calibration of the laser projector. The computation of the laser plane with respect to the camera frame is conducted similarly to the procedure presented in [22]. The camera observes the laser projected checkerboard calibration pattern in several poses and then makes use of the constraints imposed by the relation between the projection of the laser stripe and the plane of the calibration pattern. The turntable is calibrated using the technique proposed in [5].

3. Shape from optical triangulation

The structure of the scanner setup directly affects the acquisition process and the quality of the optical triangulation reconstruction. Therefore, it is important to distinguish the different types of available scanners [23]. The simplest and most common of all the structured light systems is the single-camera, single-stripe optical triangulation system. Since a single stripe covers only a small portion of the object, the scanner must be swept along the object to obtain a full range image. This is most commonly implemented with a circular sweep using a turntable on which the object is placed. Some systems also incorporate translational movement either to the scanner or to the object. A single sweep produces parallel stripes of sampled range points from which a patch of the object surface is triangulated. This system introduces the added computation of translational calibration and the problem of aligning overlapping parts of the multiple patches obtained from different views. Another alternative is the use of multi-stripe projectors that bear the requirement of distinguishing the stripes one from another [24,25,16].

The current setup that we employ fits into the single-camera, single-stripe scanner category, supporting only a rotational movement of the object. The alignment of patches is not required, but in turn the surface coverage is limited. However it is possible to make multiple scans with the projectors positioned differently, to cover more of the object surface. Since the object is rotated with respect to the projector, the projected laser planes intersect in space, yielding irregular sampling of the surface and hence deficient range data on some parts of the surface for some objects.

The input to our shape from optical triangulation process is a series of laser images of the object in full rotation on the turntable. The routine starts with processing the laser images to draw 2D sample points from the laser stripes. A depth profile is computed from the sample points, creating a cloud of range points. Connecting nearest neighbors with triangles is a common strategy for systems that produce lattices of regular samples, such as in [17]. The triangulation method that we use undertakes a similar path by weaving a web of faces across range points of adjacent stripes. The fusion technique that will be presented in Sections 5 and 6 does not in fact explicitly use the connectivity information of a polygon mesh generated from the range points. The triangulation of range points rather serve us to estimate a surface normal for each range point. Note also that the oriented range points resulting from different scans will be integrated separately in the fusion process, thus they are not yet merged into a single surface representation.

The connectivity information of the triangulated range data is also used to obtain a denser set of range points. The faces of the obtained triangle mesh are subdivided to produce new points that can be incorporated into the fusion for volume carving. This enrichment in range data is carried out to support the high levels of the octree structure created by the shape from silhouette method. In practice, since the number of faces grows very fast with subdivision, a few steps of subdivision usually suffice for a satisfactory volume carving. Each subdivision step produces one new vertex and three new faces per old face. The new vertex is positioned at the centroid of the old face and its camera position is computed by averaging the camera positions associated with the old vertices. This scheme is simple to implement and adequately serves its purpose, which is to achieve a certain density of range points when carving the octree space.

4. Shape from silhouette

We use the shape from silhouette technique presented in [5], which suits well to our volumetric fusion framework. The visual hull of the object is computed by volume carving. An octree structure is used for this purpose, that represents the volumetric data in terms of cubes with varying sizes. These cubes (or nodes) may be in one of the three states: either totally inside the object (IN); totally outside the object (OUT); or intersecting the surface of the object (ON). Starting from a root node (representing the bounding box), the octree structure is constructed by recursively subdividing each parent cube. IN and OUT cubes need not be further subdivided since they do not comprise any surface information. The recursive subdivision continues only for ON cubes, until the unit cubes corresponding to leaf nodes of the highest octree level R are reached. When constructing the octree, the state of a traversed node is determined by projecting the corresponding cube onto each of the available camera planes and observing its location with respect to the silhouette boundaries. To extract a triangle mesh with higher precision from the volumetric model, the ON leaf cubes of the octree representation are triangulated using the marching cubes algorithm [26].

5. Volume carving

At the beginning of the volume carving process, the available data are collected under an octree representing

the visual hull and a set of oriented range points. The state of each node in the octree indicates if the voxel is IN, ON, or OUT. Each ON leaf node contains local triangulation information, which constitutes a surface patch enclosed by the dimensions of the cube. Each range point is linked with its position in the world space, the optical center of the camera from which it was observed, and a surface normal. The range data are accurate but incomplete mostly due to camera and light occlusions, whereas the resolution of the octree is chosen so as to match the initial density of the acquired range samples. The silhouette model, which ideally encompasses the optical triangulation model, can be carved where the hidden concavities lie and the missing range data can be filled in by the silhouette data to attain a complete and more accurate model.

The carving process changes the description of the octree and introduces new types of ON nodes that require different means of extracting surface information to achieve a final complete mesh representation. We categorize this differentiation under four types:

- TYPE 1 - Untouched ON nodes that contain only silhouette data.
- TYPE 2 - Nodes that comprise both silhouette and range data.
- TYPE 3 - Nodes that comprise only range data.
- TYPE 4 - Nodes that have neither silhouette nor range data.

5.1. Carving algorithm

Observing the fact that no part of the object volume intervenes the line between a visible range point and its projection on the camera screen is key to volume carving. A range point $\mathscr{R}_i$ and the position of the camera optical center at the time the range point is acquired, are tied to produce a line segment L_i, or a *scan line*, as it will be referred to from here on. This line segment does not intersect with the object surface. Therefore, removing those voxels from the initial volume that actually do intersect with scan lines will yield the volumetric definition that is sought. Recall from Section 3 that, prior to fusion, the set of range points is artificially made denser by subdividing the triangle mesh obtained from optical triangulation. This enrichment of the range data ensures that the volume carving process does not suffer from any rasterization artifacts and removes all the voxels that are visible from the sensor.

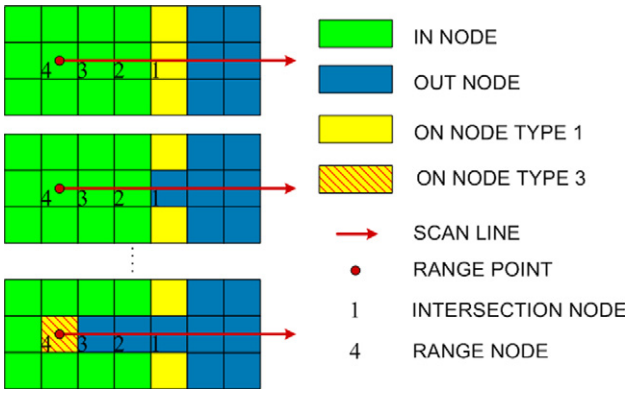

Fig. 3. Basic carving. The scan line successively carves the object grid starting from the intersection node, until the range node is encountered.

For each range point $\mathscr{R}_i$, the volumetric model is carved starting from the *intersection node*, S_{L_i}, which is the outermost node intersecting the scan line (see Fig. 3). Along the scan line, each IN or ON node is converted to OUT until the node that contains $\mathscr{R}_i$, that is the *range node* $S_{\mathscr{R}_i}$, is reached.

The procedure described above is the basic carving algorithm. There are, however, further details that need to be incorporated into the algorithm to avoid possible topological problems and surface artifacts that might rather be visually disturbing. These problems and artifacts can be avoided by proper handling of some certain cases encountered during the carving process as briefly explained in the sequel (the complete carving procedure is given as a pseudo-code in Algorithm 1):

5.1.1. Breaking out from carving

When advancing along the scan line if a node with another range point is encountered, the carving process stops and resumes with the next range point. Converting every node in between the intersection node and the range node to OUT, regardless of whether any intermediate node contains any range point or not, would give way

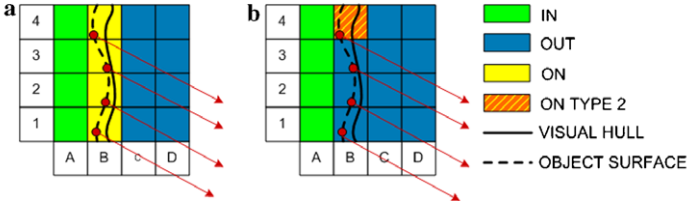

Fig. 4. Breaking out from carving. (a) The scan line of the range point in voxel B2 runs through B1. (b) This causes B1 to be converted to OUT and eliminated. Similarly, B3 scans through B2, and B4 scans through B3, ultimately, eliminating the whole layer of ON nodes. This is prevented by terminating the carving process for a scan line when a node that comprises another range point is encountered.

to possible degradation in the estimation of the object shape. Even a whole layer of ON nodes with range points representing an accurate shape could be eliminated as depicted in Fig. 4.

5.1.2. Testing the scan line for surface intersection

The scan line passing through an ON node does not necessarily cut the surface at that node as depicted in Fig. 5. Thus, the surface triangles inside the node generated by the marching cubes algorithm are tested for intersection of the scan line with the surface of the object defined by the silhouettes. If no intersection is encountered, then the node remains uncarved.

5.1.3. Incorporation of outlier range data

An ideal acquisition would produce all the scanned range points inside the boundaries of the silhouette-based visual hull. In practice, however, due to noise from input images and erroneous calibration parameters, some range points may happen to fall outside the visual hull. If such outlier range points are not taken into account at all, merging data from the two models may create jagged surfaces as depicted in Fig. 6. Thus when an OUT node with a range point is encountered, it is converted to ON TYPE 3, only if that OUT node is adjacent to an original ON TYPE 1 node from the initial visual hull. The ON nodes which no longer share a corner to an OUT node after inclusion of outliers are converted to IN to prevent multi-layering on the final triangulation. The range points falling far outside of the visual hull are discarded from fusion.

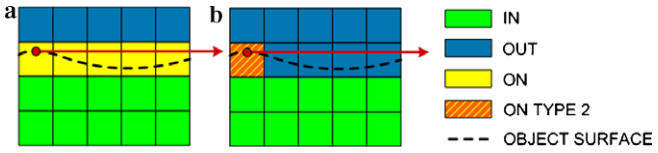

Fig. 5. Testing the scan line for surface intersection. (a) A range point on the surface and its scan line running through the depicted ON nodes. (b) Carving ON nodes without checking the scan line for intersection with the object surface may lead to their erroneous exclusion.

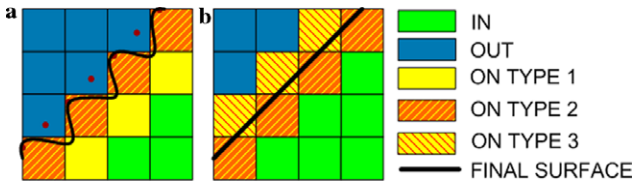

Fig. 6. Incorporation of outlier range data. (a) Triangulated patterns of dents appear where the range data are missed out. (b) When OUT nodes with range points located near the surface are converted to ON, the curving patterns disappear on the triangulation.

Algorithm 1. Carving the octree with range points.

Initially, each ON leaf node is marked as TYPE 1 or TYPE 2

For each range point $\mathscr{R}_i$
  Locate the range node $S_{\mathscr{R}_i}$
  If $S_{\mathscr{R}_i}$ is OUT
    If $S_{\mathscr{R}_i}$ is adjacent to an ON node
      Convert $S_{\mathscr{R}_i}$ to ON TYPE 3
    Else discard $\mathscr{R}_i$

For each ON node S_j
  If S_j is no more adjacent to any OUT node
    Convert S_j to IN

For each range point $\mathscr{R}_i$ not discarded before
  Find the intersection node S_{L_i} along the scan line L_i
  For each node S_j between S_{L_i} and $S_{\mathscr{R}_i}$ along L_i
    If S_j is associated with any range point
      If the state of S_j is TYPE 2
        Extract the triangulation of S_j using silhouettes
        If L_i intersects the patch of triangles derived from S_j
          Break carving with the current range point
      Else (the state is IN or TYPE 3)
        Change the state of S_j to TYPE 3
        Break carving with the current range point
    Else (the node is not associated with any range points)
      If the state of S_j is TYPE 1
        Extract the triangulation of S_j using silhouettes
        If L_i intersects the patch of triangles derived from S_j
          Change the state of S_j to OUT
      Else (the state is IN)
        Change the state of S_j to OUT

5.2. Restructuring the octree surface

At the end of this initial carving process, the states of the ON nodes in the octree vary from TYPE 1 to TYPE 3 as depicted in Fig. 7b. The original ON voxels from the visual hull that are left uncarved remain as TYPE 1 if no range point lies within the voxel. The original ON voxels from the visual hull that do comprise range data, however, are converted to TYPE 2, meaning that they are associated with surface information from both the silhouettes and range data. While some range points are located inside the ON voxels of the visual hull, others may be bounded by voxels that were completely inside the visual hull before volume carving is initiated. Since only the ON nodes contain silhouette based surface information, the IN voxels that are associated with range points after carving is initiated, make use of only the range data to extract surface

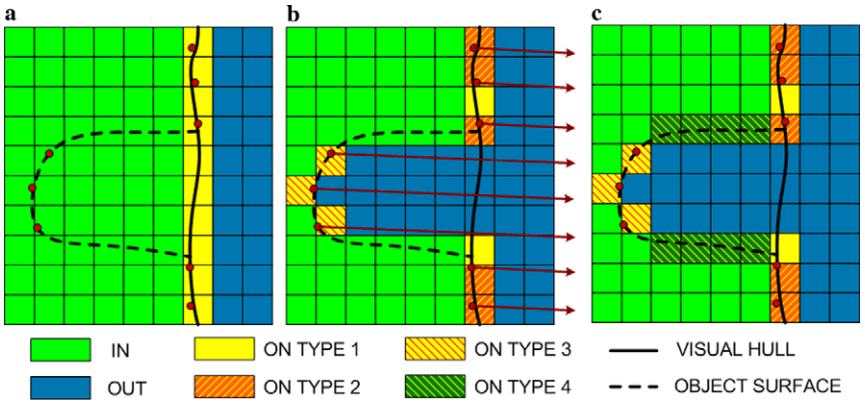

Fig. 7. Hole filling. (a) A sample grid before carving. (b) The same grid after carving; some IN nodes expose to OUT nodes along the edges of deep cuttings. (c) IN nodes sharing common sides with OUT nodes are converted to ON TYPE 4 to fill in holes.

information. The nodes representing such IN voxels in the object grid are converted to ON TYPE 3.

5.3. Hole filling

Carving the object space produces holes on the edges of the hollowed parts (neighboring IN and OUT voxels), which must be patched up in order to obtain a complete model. The definition of the octree structure asserts that no IN and OUT nodes can share a common corner, edge, or face on the space grid. As observed in Fig. 7b, the altered object space is liable to violate this requirement, especially around the edges of deeply scooped out portions of the object.

Our hole filling algorithm is very similar to the one proposed in [17], where, basically, a wall on the transition between the carved out voxels and the unseen voxels is established on the grid. In our case those IN nodes neighboring with any OUT nodes are converted to ON, specifically of TYPE 4. The only way to estimate the object shape for a TYPE 4 node is by examining the configuration of its neighboring nodes. Once each potential ON TYPE 4 node is identified in the octree, each one is associated with a point Q and a normal vector $\mathbf{N}$ to be used during the triangulation. The positioning of each IN and OUT voxel in the neighborhood (constituting 26 voxels) is used to interpolate Q and $\mathbf{N}$. Neighboring IN and OUT voxels simply act as opposing forces, as IN voxels are used to repel and OUT voxels are used to attract when determining the position of Q and the direction of $\mathbf{N}$ inside an ON TYPE 4 cube.

When picking potential ON TYPE 4 nodes, only those IN and OUT nodes that share a common face (rather than an edge or a corner) are taken into account. The reason behind this less aggressive approach is to avoid the creation of a multilayered shape during triangulation. The IN/OUT connectivity problem concerning IN and OUT nodes sharing a corner or an edge is resolved later in the triangulation process. Fig. 7c illustrates the same sample grid of Fig. 7b after hole filling is carried out.

6. Surface triangulation

The carved octree structure representing the final object volume can be triangulated using the marching cubes algorithm. To produce hole-free and manifold triangulations, this algorithm relies on the fact that common edges on neighboring voxels hold coherent isosurface information. On contrary to the shape from silhouette procedure, with the different kinds of voxels (TYPE 1–4) that are created during the volume carving in our fusion scheme, the isosurface information from adjoining voxels is not necessarily consistent. The problem of fusing isosurfaces was addressed before by Curless and Levoy [17] and Rocchini et al. [27] in a different context, for combining overlapping range maps resulting from separate optical scans. In our case, we resolve this problem by accumulating an isovalue at each grid point of the object surface since this strategy enables us to produce hole-free triangulations. Isovalues are computed using the local surface information obtained separately from range data and silhouettes.

6.1. Isosurface merging

An isovalue for a point on the grid determines the extent to which that the point is away from the surface. The closer to the object surface the closer the value approaches 0. IN and OUT points naturally have opposite signs. The intersection of the isosurface with the cube edges can be approximated by interpolating between corners of opposite signs. The isovalue at each grid point is estimated by averaging the *local* isovalues computed on the corresponding corners of the surface voxels sharing the grid point. A local isovalue on each corner of a voxel is given by the signed distance of the corner to the object surface inside that voxel. The strategy employed in computing local isovalues varies according to the type of the surface voxel, as explained in the sequel:

6.1.1. TYPE 1

The local isovalue of each corner is determined by its distance to the closest triangle inside that cube. The triangulation information is given by the shape from silhouette procedure. In most cases, the distance to a triangle can be well approximated by the distance to its plane. However in some cases, although the triangle itself is distant to a particular corner, the plane of the triangle may pass relatively closely to that corner. This observation leads us to the following strategy for computation of the distance to a triangle. If the corner is not incident to any cube edge that contains a vertex of the triangle, that is, if the distance from the corner to any vertex is more than the side length l of the maximum level cube, then we say that the corner is locally distant to the triangle. In this case, it is safer to approximate the distance to the triangle as the minimum of the distances from the corner to the three vertices. If the corner is not locally distant to the triangle, then its distance is approximated by the distance to the triangle plane. The local isovalue $\hat{f}$ for each corner C_i, $i = 1, \ldots, 8$, is thus computed as follows:

$$\hat{f}(C_i) = \min_j D_{ij}, \tag{1}$$

where D_{ij} denotes the distance from a corner C_i to a triangle T_j inside the cube:

$$D_{ij} = \begin{cases} d(C_i, P_{T_j}) & \text{if } \min_n d(C_i, V_{jn}) < l \\ \min_n d(C_i, V_{jn}) & \text{otherwise,} \end{cases} \tag{2}$$

where the function d gives the signed distance from a corner C_i of the cube, to either the plane P_{T_j} of a triangle T_j defined inside the cube or to a vertex V_{jn}, $n = 1, 2, 3$, of the triangle.

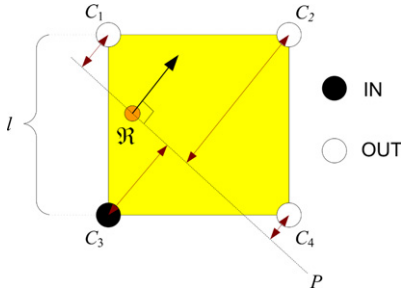

Fig. 8. Assignment of the local isovalues on the corners of an ON TYPE 3 voxel. A range point located in the voxel defines a plane P perpendicular to its normal.

6.1.2. TYPE 3

An ON node of this type may contain several range points. Each range point carries a normal vector from which a plane can be extrapolated that is perpendicular to the normal. The local isolevel value $\hat{f}(C_i)$ of each corner is computed as the signed distance to the plane obtained from the range point that is closest to the corner (see Fig. 8).

6.1.3. TYPE 2

The methods previously explained for TYPE 1 and TYPE 3 nodes are both applicable for computation of local isovalues of a TYPE 2 node since it comprises both silhouette and range data. In practice, since the range data are assumed to be more precise than the silhouettes, we favor the method employed for TYPE 3 nodes.

6.1.4. TYPE 4

Each node of this type is associated with a point and a surface normal (see Section 5.3). Thus its local isovalues are computed in the same way as a TYPE 3 node which comprises only range points.

An overall isovalue $f(C_k)$ for each grid point (or corner) C_k is finally computed by simply averaging all of its local isovalues $\hat{f}_m(C_k)$:

$$f(C_k) = \frac{1}{M_k} \sum_{m=1}^{M_k} \hat{f}_m(C_k), \tag{3}$$

where M_k is the number of the ON nodes that the grid point C_k is a corner to.

6.2. Final modification to the octree

By the octree definition, all the corners of OUT cubes must be OUT and likewise, all the corners of IN cubes must be IN. After merging the isovalues of the corresponding corners on adjoining nodes, the resultant states of the grid points may break this rule, i.e., the same point on the grid may be shared by both an OUT cube that defaults the shared point to be OUT and an ON cube in which the final isovalue assigned to the shared point is positive (defining the point as IN). If these contradictions are not resolved, there may appear holes on the surface reconstruction after marching cubes is applied. Thus, as depicted in Fig. 9, an OUT node with at least one corner declared as IN by a neighboring ON node, is converted to ON. The same applies for IN nodes that have corners declared to

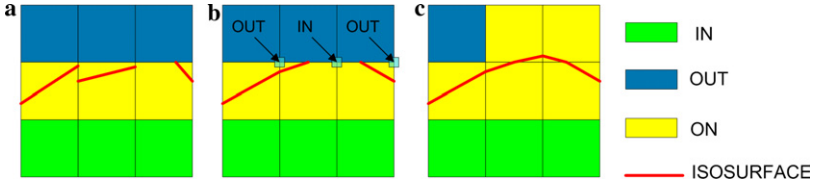

Fig. 9. (a) A sample grid before fusing the local isovalues. (b) After fusion, two OUT nodes end up sharing an IN corner with an ON node; a hole is exposed in the triangulated surface. (c) The same grid after covering up holes.

be OUT by neighboring ON nodes. Since IN and OUT cubes contain no isovalue information for their corners, the isovalues of the corners of the cubes that are converted must be copied from the corresponding corners on the neighboring ON nodes if available. In the case where no isovalue is available for a corner (when there is no neighboring ON cube), the isovalue defaults to -1 for OUT cubes and 1 for IN cubes. The default values are kept large to prevent over-peaking where the holes are sealed.

6.3. Mesh post-processing

The final triangular mesh obtained by applying the marching cubes algorithm is decimated via edge collapses so as to eliminate the faces with edges smaller than a threshold which is set as proportional to the grid size. Surface fairing is conducted to further smooth the object shape at sharp and jagged protrusions that look unnatural [28]. Fairing is exclusively applied to faces obtained from TYPE 4 nodes because these nodes produce especially rough surfaces on the transition from TYPE 4 nodes to other types of neighboring nodes.

7. Experimental results

The reconstructions of four models are presented to demonstrate the performance of the algorithm. As displayed in Fig. 10, the test objects are Greek1, Greek2 (two original ancient statuettes made of stone), Elephant and Hand (two ordinary statuettes made of wood and plastic, respectively). The size of the images acquired during the experiments is 2000×1310, which is sufficient to support the high level of detail demanded by the reconstructions. The numbers of silhouette and range images acquired vary depending on the shape and the size of the reconstructed object as listed in Table 1.

Table 1
Numbers of silhouette and range images acquired for each object

Object	Silhouette	Laser 1	Laser 2	Laser 3
Greek1	72	360	—	—
Greek2	72	360	—	—
Elephant	72	360	360	—
Hand	72	180	180	180

Figs. 11 and 12 illustrate the results obtained. The Greek1 sequence in the first row of Fig. 11 demonstrates the fusion process at octree resolution $R = 8$. The silhouette reconstruction typically lacks the detailed features while the optical triangulation exhibits many holes. It is important to note that the hollow parts on the handles of the Greek1 model are obtained from the silhouettes during the fusion process. The laser scanner of our acquisition system fails to detect these parts. Even more sophisticated optical triangulation scanners alone would not be able to reconstruct these parts of the object completely or reliably because of the high angle of incidence involved. The Greek2 sequence in Fig. 11 demonstrates the fusion process at resolution $R = 7$. The silhouette-based reconstruction shows the need for further carving. The fusion reveals a watertight reconstruction that includes cavity shape information deduced from range data. Note that the holes on the hair for example are not only successfully sealed with the silhouette information, but also shaped, to some extent, by the scan lines of the range points in the vicinity. The result of the fusion is, however, less sharp than that of the optical triangulation alone, which is caused by the relatively low resolution of the octree used. The Elephant sequence in Fig. 11 further shows that our fusion scheme can be used to generate very satisfactory results. The obstructed inner faces of the legs and the trunk and, though not visible at the depicted angle, the top sections of the model, are missing from the two separate scans of optical triangulation and are compensated with the silhouettes. Some minor artifacts at the belly of the Elephant are observed as a result of the deficiency of the range data around that area.

Fig. 12 presents the reconstructed models of the Hand object. The first row displays a succession of three optical triangulation reconstructions obtained from three separate scans along with two views from the silhouette model at $R = 7$. The second row first displays the three fusion results each obtained at $R = 7$ by using the corresponding 360-degree single scan data displayed as a triangulated surface in the first row. The last two images of the second row visualize two views from the overall fusion that incorporates all the three optical scans available. We observe that the range scanner misses some parts of the visible surface that are actually not occluded. This is mainly because our rotational range scanner produces laser planes that

Fig. 10. Original images from Greek1, Greek2, Elephant, and Hand objects.

Fig. 11. (Greek1, first row) Silhouette reconstruction, optical triangulation reconstruction, and two views from the model obtained by fusion at $R = 8$, respectively. (Greek2, second row) Silhouette reconstruction, two views from optical triangulation reconstruction, and two views from fusion at $R = 7$. (Elephant, third row) Silhouette reconstruction, optical triangulation reconstructions from two separate scans, and two views from fusion at $R = 8$.

converge towards certain locations of the object surface. However, incorporating three separate scans, the fusion process results in a very satisfactory carving and a watertight surface model at resolution $R = 7$.

In Figs. 11 and 12, we observe that the fusion algorithm captures some cavity information that is missing in the optical triangulation models on some parts of the object surface such as on the palm of the Hand and on the hair of the Greek2 statuette. This is thanks to the carving algorithm that makes use of not only the range points but also their scan lines, as demonstrated on the Hand object in Fig. 13. Although no range point is sampled on some parts of the palm of the Hand, the cavity information on these surface sections is partly recovered by the scan lines targeting at different locations on the object surface. These scan lines carve out all the voxels that lie between the corresponding range points and their projections on the camera screen. The surface information recovered by the fusion algorithm on such cavities is thus never exact, but still much more faithful to the real surface than the one obtained by the silhouettes.

Although it is not possible to evaluate the accuracy obtained on the cavities that are missing in the optical triangulation models (unless a ground truth is made available), it is still possible to quantify the accuracy of the fusion algorithm and the improvement obtained with respect to the silhouette model. Let $d(\mathscr{R}_i, \mathscr{M})$ denote the Euclidean distance of the range point $\mathscr{R}_i$ to the nearest point on the reconstructed surface $\mathscr{M}$. We then define the average distance ϵ (per range point) of the range dataset to the reconstructed surface as follows:

$$\epsilon = \frac{1}{N} \sum_{i=1}^{N} d(\mathscr{R}_i, \mathscr{M}), \tag{4}$$

where N is the size of the range dataset. Note that ϵ is a directed distance metric that measures the faithfulness of the reconstructed surface to the accurate but incomplete range data. In Table 2, we display for each of the 4 different objects, the distance ϵ of the models obtained by shape from silhouette and fusion. The distance is calculat-

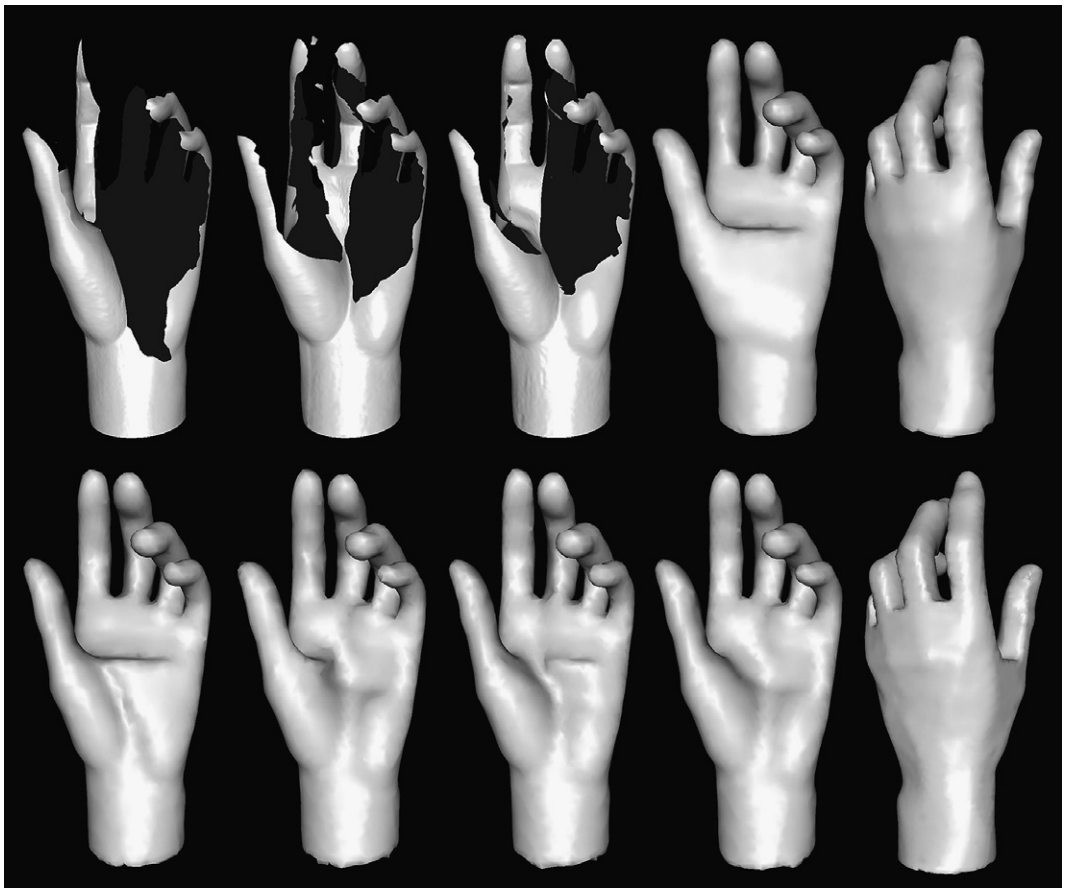

Fig. 12. Reconstructed models of the Hand at $R = 7$. (First row) Three optical triangulation reconstructions obtained from three separate scans and two views from the silhouette model. (Second row) Three fusion results each obtained by using the corresponding single optical scan displayed in the first row and two views from the overall fusion that incorporates all the three scans available.

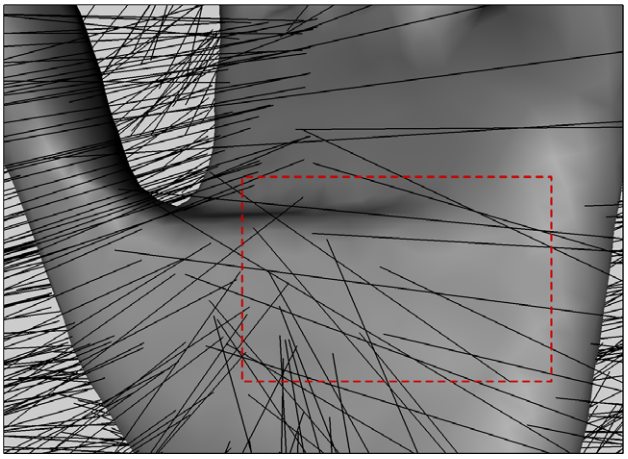

Fig. 13. Scan lines of the carving algorithm, demonstrated on the zoomed silhouette model of the Hand. Although no range point is sampled on some parts of the palm, the cavity information is partly recovered by the scan lines targeting at different locations on the object surface. Note that only a reduced set of scan lines resulting from three separate scans are displayed for illustration.

ed by assuming that each object is circumscribed by a bounding sphere with radius 100 units. For instance, for the Greek2 object which is in reality bounded by a sphere

Table 2
Average distance ϵ of the acquired range data to the models reconstructed by shape from silhouette (SFS) and fusion for each object

Model	ϵ-SFS	ϵ-Fusion
Greek1 ($R = 8$)	1.722	0.070
Greek2 ($R = 7$)	1.046	0.061
Elephant ($R = 8$)	1.708	0.087
Hand ($R = 7$)	0.961	0.268

of radius approximately 15 cm, the average distance of the range data to the fused model is about 0.1 mm. The results given in Table 2 indicate that the fusion algorithm improves the accuracy of the silhouette model while it recovers the cavity information to the extent that is possible with the available scan lines. The improvement obtained by fusion with respect to silhouette-based reconstruction is the least (about 4 times) for the Hand object as compared to the others since in this case the object surface does not contain severe cavities except at the palm which is however also missing in the optical triangulation models.

The resolution of the octree used for volume carving in our experiments was always chosen to be slightly less than the initial resolution of the acquired range samples. This is due to two reasons. First, our rotational scanner performs

irregular sampling and may thus yield deficient range data on some parts of the object surface. Second, the acquisition noise may occasionally cause the loss of range samples. The restriction on the use of higher octree resolutions results in slightly smoothed reconstructions, but in turn prevents possible artifacts and uncarved protrusions from appearing on the reconstructed surface.

The typical execution time for the fusion algorithm, though not completely optimized, (excluding shape from silhouette and optical triangulation) is less than 10 min on a 1.7 GHz Pentium PC. The fusion for the Elephant model consisting about 200 K triangles at $R = 8$ for example takes about 7 min.

8. Conclusion

This work has described a novel surface reconstruction scheme that fuses the geometry information obtained separately from shape from silhouette and shape from optical triangulation techniques. The aim was to compensate for the problems associated with each method by the benefits of the other. The fusion method is based on volume carving followed by isosurface merging. The experiments show that it is possible to produce robust and satisfactory surface reconstructions. The most prominent property of the presented method is the ability to build cavity-sensitive and hole-free models of complicated objects containing severe occlusions and sharp hollows on their surfaces. The test objects of our experiments, especially the Hand and the Greek1, are examples to such complicated objects which are very difficult to reconstruct using only shape from optical triangulation, even with range scanners that are much more sophisticated than ours.

The foremost restraining factor in the overall system performance was found to be hardware related. The fixed single-stripe range scanner employed in our experiments, supporting only rotational object movement, may deliver deficient and irregularly distributed range samples for some objects. This defect restricts the use of higher resolutions than the ones conducted in the experiments. However, using range scanners that produce full range images (that support translational movement for instance) as discussed in Section 3, can significantly enhance the distribution of the range data and permit carving at higher octree resolutions, which would produce even sharper fusion results and better preserve the details of the object shape.

References

[1] C.H. Chien, J.K. Aggarwal, Identification of 3d objects from multiple silhouettes using quadtrees/octrees, Comput. Vis. Graph. Image Process 36 (2-3) (1986) 256–273.

[2] R. Szeliski, Rapid octree construction from range sequences, Comput. Vis. Graph. Image Process 58 (1) (1993) 23–32.

[3] W. Niem, J. Wingbermuhle, Automatic reconstruction of 3d objects using a mobile camera, Image Vis. Comput. 17 (2) (1999) 125–134.

[4] M. Tarini, M. Callieri, C. Montani, C. Rocchini, Marching intersections: an efficient approach to shape-from-silhouette, Vis. Model. Visual. VMV'2002 (2002) 283–290.

[5] Y. Yemez, F. Schmitt, 3d reconstruction of real objects with high resolution shape and texture, Image Vis. Comput. 22 (2004) 1137–1153.

[6] Y. Matsumoto, K. Fujimura, T. Kitamura, Shape-from-silhouette/stereo and its application to 3-d digitizer, Int. Conf. on Discrete Geometry for Computing Imagery (1999) 177–190.

[7] K.N. Kutulakos, S.M. Seitz, A theory of shape by space carving, Int. J. Comput. Vis. 38 (3) (2000) 199–218.

[8] C.H. Esteban, F. Schmitt, Silhouette and stereo fusion. for 3d object modeling, Comput. Vis. Image Understand 96 (3) (2004) 367–392.

[9] N. Brusco, L. Ballan, G.M. Cortelazzo, Passive reconstruction of high quality textured 3d models of works of art, in: 6th Int. Symp. on Virtual Reality, Archeology and Cultural Heritage VAST'05 (2005) 21–28.

[10] S. Savarese, H. Rushmeier, F. Bernardini, P. Perona, Implementation of a shadow carving for shape capture, in: Proc. 3D Data Processing Visualization and Transmission (2002) 107–114.

[11] J. Park, G.N. DeSouza, A.C. Kak, Dual-beam structured-light scanning for 3-d object modeling, in: 3rd Int. Conf. on 3D Digital Imaging and Modeling (3DIM'01) (2001) 65–72.

[12] C. Liska, R. Sablatnig, Adaptive 3d acquisition using laser light, in: Proc. Czech Pattern Recognition Workshop (2000) 111–116.

[13] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, D. Fulk, The digital michelangelo project: 3d scanning of large statues, in: Proc. SIGGRAPH'00 (2000) 131–144.

[14] J. Davis, S.R. Marschner, M. Garr, M. Levoy, Filling holes in complex surfaces using volumetric diffusion, in: Int. Symposium on 3D Data Processing, Visualization, and Transmission (2002) 428–438.

[15] A. Fasano, M. Callieri, P. Cignoni, R. Scopigno, Exploiting mirrors for laser stripe 3d scanning, in: Int. Conf. on 3D Digital Imaging and Modeling (3DIM'03) (2003) 243–250.

[16] S. Rusinkiewicz, O. Hall-Holt, M. Levoy, Real-time 3d model acquisition, in: Proc. SIGGRAPH'02 21(3) (2002) 438–446.

[17] B. Curless, M. Levoy, A volumetric method for building complex models from range images, in: Proc. SIGGRAPH'96 30 (1996)303–312.

[18] P. Fua, Y.G. Leclerc, Object-centered surface reconstruction: combining multi-image stereo and shading, Int. J. Comput. Vis. 16 (1) (1995) 35–56.

[19] S. Tosovic, R. Sablatnig, M. Kampel, On combining shape from silhouette and shape from structured light, in: Proc. 7th Computer Vision Winter Workshop (2002) 108–118.

[20] T. Terauchi, Y. Oue, K. Fujimura, A flexible 3d modeling system based on combining shape-from-silhouette with light-sectioning algorithm, in: Int. Conf. on 3D Digital Imaging and Modeling (3DIM'05) (2005) 196–203.

[21] J. Bouguet, Camera calibration toolbox for matlab, http://www.vision.caltech.edu/bouguetj/calib_doc/.

[22] Q. Zhang, R. Pless, Extrinsic calibration of a camera and laser range finder, in: IEEE Int. Conf. on Intelligent Robots and Systems (IROS) (2004) 2301–2306.

[23] F. Bernardini, H.E. Rushmeier, 3d model acquisition, Eurographics 2000, State of the Art Reports Proceedings (2000) 41–62.

[24] K.L. Boyer, A.C. Kak, Color-encoded structured light for rapid active ranging, IEEE Trans. PAMI 9 (1) (1987) 14–28.

[25] C.M.P.P.C. Rocchini, P. Cignoni, R. Scopigno, A low cost 3d scanner based on structured light, in: Proc. Eurographics'01 20 (3) (2001) 299–308.

[26] W. Lorensen, H. Cline, Marching cubes: a high resolution 3d surface construction algorithm, in: Proc. SIGGRAPH'87 21 (4) (1987) 163–169.

[27] C. Rocchini, P. Cignoni, F. Ganovelli, C. Montani, P. Pingi, R. Scopigno, The marching intersections algorithm for merging range images, Vis. Comput. 20 (2-3) (2004) 149–164.

[28] G. Taubin, A signal processing approach to fair surface design, in: Proc. SIGGRAPH'95 (1995) 315–358.