Geometric Fusion for a Hand-held 3D Sensor

Adrian Hilton and John Illingworth

Centre for Vision, Speech and Signal Processing

University of Surrey, Guildford GU25XH, UK

a.hilton@surrey.ac.uk

Abstract

This article presents a geometric fusion algorithm developed for the reconstruction of 3D surface models from hand-held sensor data. Hand-held systems allow full 3D movement of the sensor to capture the shape of complex objects. Techniques previously developed for reconstruction from conventional 2.5D range image data cannot be applied to hand-held sensor data. A geometric fusion algorithm is introduced to integrate the measured 3D points from a hand-held sensor into a single continuous surface. The new geometric fusion algorithm is based on the *normal-volume* representation of a triangle which enables incremental transformation of an arbitrary mesh into an implicit volumetric field function. This system is demonstrated for reconstruction of surface models from both hand-held sensor data and conventional 2.5D range images.

Key-words: Object Modelling; 3D Reconstruction; Geometric Fusion; Range Image Integration

1 Capturing 3D models of real objects

Realistic object representation is a primary goal of computer graphics research [Curless and Levoy, 1996]. Model construction is a major bottleneck in the application of computer generated imagery. Applications demanding realistic models of real objects include film animation, virtual museums, tele-shopping, tele-conferencing, immersive virtual reality, military simulation and multimedia education/entertainment.

Currently manual techniques are used to build models of real objects and have been used to build commercial 3D model-banks [Infografica, 1996, Datalabs, 1996]. However, this is an expensive process requiring an expert modeller or animator and takes several months to obtain a single instance of a complex object. Reconstruction costs prevent realistic modelling of objects with a high level-of-detail. In addition, each new instance of an object must be reconstructed from scratch. This prohibits realistic modelling of individual instances of organic objects which exhibit large shape variations such as people or animals.

Image-based representations have been introduced [Gortler et al., 1996, Levoy and Hanrahan, 1996] which achieve photo-realistic image synthesis of objects. This approach encodes the projection of all light rays in a scene as a four dimensional function. This approach achieves highly realistic visualisation but is currently limited to static objects and scenes with fixed lighting.

1.1 3D Surface Measurement

Automatic reconstruction of models of 3D objects has received considerable interest in computer vision research using both active [Hoppe et al., 1992, Curless and Levoy, 1996, Hilton et al., 1998, Pito, 1996a, Soucy and Laurendeau, 1995b, Turk and Levoy, 1994] and passive [Niem and Wingebermhle, 1997, Fitzgibbon et al., 1998] surface measurement technologies. Active surface measurement techniques project a structured light pattern such as a laser stripe onto the object surface and by a process of optical triangulation measure the distance to points on the stripe. This approach enables high-accuracy dense 3D surface measurement. Conventional active ‘2.5D range’ image sensor acquire surface measurements on a grid corresponding to either a plane or cylinder. Capture of the full surface shape for 3D objects requires the acquisition of multiple 2.5D images. Automatic acquisition of full surface shape requires techniques for viewpoint planning. For complex

objects this may require the sensor to enter the object space and risk collisions with the unknown object surface. A general solution to this problem remains an open research issue, see [Pito, 1996b] for a review of this work.

Hand-held active measurement devices have recently been developed to enable full surface shape measurement for complex objects. The basic principle is to allow full six degree-of-freedom movement of an active surface measurement device in 3D space around the object. The freedom of movement enables an operator to select viewpoints to capture the full object surface. This utilises our expert knowledge in selecting sensor positions within the object space that avoid collisions. The user can also select and maintain the sensor viewpoint to be approximately normal to the object surface to achieve accurate and repeatable measurement. Providing a fast solution to acquiring surface shape for complex real objects. Several hand-held sensor systems have been developed [Fisher et al., 1996] and are commercially available including systems by 3D Scanners (www.3dscanners.com), Polhemus(www.polhemus.com) and EOIS (www.eois.com). The sensor position and orientation is measured using an articulated arm, electromagnetic sensor or optical device. The choice of position sensor determines the system cost, measurement volume, accuracy of the surface measurements and any restrictions on movement. An initial calibration of the system registers the sensor coordinate system with the position sensor coordinate system. This enables measurements of points on the object surface with respect to a single global coordinate system. The accuracy of the surface measurements for a hand-held sensor are dependent on both the accuracy of the position sensor and the accuracy of the range measurement. Throughout this work it is assumed that the position sensor is repeatable with a zero mean error over the entire object surface. In this paper we focus primarily on the 3D Scanners ModelMaker system (Figure 1(a)) for which the techniques presented were originally developed.

A primary difference between hand-held sensor data and conventional 2.5D range image data is that the measurements are not structured on a grid. With a hand-held sensor such as the ModelMaker system the operator moves the projected light stripe back and forth across the surface in an action similar to paint spraying. This results in a series of stripes of 3D point measurements on the object surface which are not structured on a regular grid and do not have a common viewpoint. The raw point measurements are illustrated in Figure 1(b). Algorithms for surface reconstruction from conventional range image data cannot be applied to hand held sensor data. This paper introduces a new algorithm to enable reconstruction of surface models from hand-held sensor data.

1.2 Surface Reconstruction

Range image integration algorithms developed for conventional sensors can be classified into two categories: *mesh-based* which integrate directly overlapping mesh regions into a single mesh [Pito, 1996a, Soucy and Laurendeau, 1995b, Turk and Levoy, 1994, Rutishauser et al., 1994]; and *volume-based* which construct an intermediate volumetric implicit surface representation for fusion of overlapping surface measurements [Curless and Levoy, 1996, Hilton et al., 1996, Roth and Wibowo, 1995].

It has been shown [Curless and Levoy, 1996, Hilton et al., 1996] that volumetric approaches achieve improved reconstruction of complex geometry and greater computational efficiency. Algorithms for reconstruction of object models from conventional range image sensor data have assumed that the data are structured on a regular 2D grid [Curless and Levoy, 1996, Hilton et al., 1998, Pito, 1996a, Soucy and Laurendeau, 1995b, Turk and Levoy, 1994, Rutishauser et al., 1994]. This assumption allows reliable estimation of the local surface topology based on the distance between adjacent measurement on the grid [Soucy and Laurendeau, 1995b, Turk and Levoy, 1994]. In addition, many algorithms assume that the range image measurements are taken from a common view direction [Curless and Levoy, 1996, Soucy and Laurendeau, 1995b] in order to evaluate measurement overlap. Both of these assumptions prohibit the use of previous range image integration algorithms for fusion of hand-held sensor data.

Exceptions to this are algorithms, which address the more general problem of reconstruction from unstructured data [Boissonnat, 1984, Hoppe et al., 1992, Mencl, 1995]. However, algorithms for unstructured data do not achieve reliable reconstruction from measurement data of complex objects as the Euclidean distance between measurement points is used to estimate local surface topology. This distance metric fails where different parts of the object surface are in close proximity such as crease edges and thin sections [Hilton et al., 1998].

In this paper we introduce a new geometric fusion algorithm for reliable reconstruction of object models from hand-held or conventional 3D sensor data. The algorithm is based on the construction of an intermediate volumetric implicit surface representation for integration of overlapping measurements. The normal-volume representation of a triangle is used to incrementally transform an arbitrary mesh into an implicit volumetric representation. For hand-held sensor data we take advantage of the relationship between adjacent measurement stripes to reliably estimate local surface topology. This approach avoids prior assumptions of a grid measurement structure or single viewpoint enabling

reconstruction from hand-held sensor data. The reconstructed object surface for a typical object is illustrated in Figure 1(d).

A digital colour camera in the hand-held sensor also enables acquisition of set of multi-view colour from known locations. Texture mapping techniques [Niem and Wingebermhle, 1997] are used to project the images onto the reconstructed 3D surface and integrate overlapping image regions. This results in a highly realistic colour 3D object model suitable for computer generated imagery.

2 Geometric fusion to reconstruct a 3d model

The goal of geometric fusion is to obtain a surface model from the captured 3D point measurements. This section presents a general algorithm for reconstructing a surface by fusion of arbitrary sets of surface measurements into a single volumetric representation. A single triangulated mesh model of the surface can then be obtained from the fused volumetric representation.

Hoppe [Hoppe et al., 1992] introduced the use of an intermediate implicit volumetric representation for fusion of unstructured 3D point measurements. This approach was extended [Curless and Levoy, 1996, Hilton et al., 1996] to achieve reliable reconstruction of detailed surfaces from conventional range images by assuming that the measurements are on a regular grid in order to estimate the local surface topology.

The new geometric fusion algorithm presented here enables reliable and efficient fusion of surface measurements from both conventional range images and hand-held 3D sensor data. This is achieved by introducing an incremental procedure for transformation of an arbitrary mesh to a volumetric field function. This approach does not require the measurement data to be either structured on a grid or have a common view direction. The geometric fusion algorithm for generating a single integrated surface model proceeds in four stages:

1. **Surface Topology Estimation:** Estimate the local surface topology to triangulate the measurements.
2. **Volumetric Representation:** Transform the surface triangulation to an implicit volumetric representation.
3. **Geometric Fusion:** Integrate overlapping volumetric representations into a single implicit surface.

(a) 3DScanners ModelMaker system

(b) Raw point stripes

(c) Triangulated patches (partial)

(d) Fused model triangulation

(e) Fused model surface

Figure 1: Model reconstruction from hand-held sensor data

4. **Triangulation:** Triangulate the implicit surface to generate a single mesh model.

2.1 Surface Topology Estimation

Conventional range sensor data is structured in a 2.5D grid enabling estimation of the local surface topology based on the distance between adjacent points. A *step discontinuity constrained* triangulation [Hilton et al., 1998] has previously been used [Curless and Levoy, 1996, Pito, 1996a, Rutishauser et al., 1994, Soucy and Laurendeau, 1995a, Turk and Levoy, 1994] as a initial step to generate a mesh representation which approximates the local topology of the measured surface.

For hand-held sensor data measurements are taken in stripes across the object surface as the sensor is moved. We perform a step discontinuity triangulation between points in adjacent stripes to estimate the local surface topology. As in previous work a threshold distance, $t_d = 3\Delta x$, is used to test if the surface is continuous between adjacent points on consecutive stripes: $|\vec{x}_i - \vec{x}_j| < t_d$ where Δx is the known sampling resolution of the sensor system for a given surface distance. If adjacent points satisfy this criteria then they are connected in a local surface triangulation. Discontinuities between patches occur due to either a step on the object surface, the sensor being moved rapidly or the sensor changing direction. For a single patch there may not be a common viewpoint as the sensor orientation is changing continuously during acquisition. This triangulation results in a series of overlapping surface patches which approximate the surface geometry and topology. The resulting triangulation for hand-held sensor data is illustrated in Figure 1(c) for a subset of the total patches generated from the raw stripe data. The objective is then to integrate these surface patches into a single surface representation. Due to the lack of a grid structure or single viewpoint for surface patches they can not be integrated using previous approaches for conventional 2.5D range image data.

2.2 Volumetric Representation

In this section we introduce a general algorithm for converting an arbitrary triangulated mesh M to a volumetric representation. An arbitrary topology closed manifold surface S can be represented in implicit form as an iso-surface of a spatial field function, $f(\vec{x})$, where $\vec{x} = [x, y, z]$ is any point in Euclidean space, R^3 . Thus we can represent a surface by defining the field function $f(\vec{x})$ as the signed distance from a point, $\vec{x}$, to the nearest point on the surface. This gives the iso-surface $f(\vec{x}) = 0$ for all points on S and $f(\vec{x}) \neq 0$

elsewhere. A discrete volumetric representation can be implemented by uniform spatial subdivision into voxels cells and local planar approximation for voxels near the implicit surface, $f(\vec{x}) = 0$. In practice, we require a discrete volumetric representation for computational efficiency of the fusion algorithm [Curless and Levoy, 1996].

Let us define a voxel grid based on a uniform spatial subdivision with voxel centres $\vec{c}_{\underline{i}} = [x_i, y_j, z_k]$ given by:

$$\vec{c}_{\underline{i}} = [\Delta x(i + \frac{1}{2}) + x_{min}, \Delta y(j + \frac{1}{2}) + y_{min}, \Delta z(k + \frac{1}{2}) + z_{min}] \quad (1)$$

The discrete representation is defined in the range $[\vec{c}_{min}, \vec{c}_{max}]$ with uniform voxel resolution in each spatial dimension $\Delta c = \Delta x = \Delta y = \Delta z$. The number of voxel cells in each dimension is given by $N_V = [\frac{(x_{max}-x_{min})}{\Delta c}, \frac{(y_{max}-y_{min})}{\Delta c}, \frac{(z_{max}-z_{min})}{\Delta c}]$. The volume occupied by the voxel cell corresponding to centre $\vec{c}_{\underline{i}}$ denoted $V(\vec{c}_{\underline{i}}) = [x_i \pm \frac{\Delta c}{2}, y_j \pm \frac{\Delta c}{2}, z_k \pm \frac{\Delta c}{2}]$. For a given point $\vec{x}$ we can evaluate the corresponding voxel index $\underline{i}$ by:

$$\underline{i} = [\frac{(x - x_{min})}{\Delta c}, \frac{(y - y_{min})}{\Delta c}, \frac{(z - z_{min})}{\Delta c}] \quad (2)$$

Volumetric surface representation is achieved by local planar approximation of the surface in the discrete voxel structure. Given an input triangulated mesh M composed of a set of vertices $P = [\vec{v}_0, \dots, \vec{v}_i, \dots, \vec{v}_{N_P-1}]$ and triangles $T = [t_0, \dots, t_j, \dots, t_{N_T-1}]$ where $t_j = [\vec{v}_r, \vec{v}_s, \vec{v}_t]$. We assume that M is a simply connected manifold triangulation with no self-intersections. Thus we can estimate the local surface normal for each triangle as $\vec{n}_{t_j} = \frac{(\vec{v}_s - \vec{v}_r) \times (\vec{v}_t - \vec{v}_r)}{\|(\vec{v}_s - \vec{v}_r) \times (\vec{v}_t - \vec{v}_r)\|}$. Vertex normals can be estimated by a weighted average of the adjacent triangle normals [Taubin, 1995]:

$$\vec{n}_{v_r} = \frac{\sum_{j \in r} w_j \vec{n}_{t_j}}{\sum_{j \in r} w_j} \quad (3)$$

A *volumetric envelope* is defined around the mesh M which enables us to convert to an implicit volumetric representation. An *offset surface* M' for mesh M is given by displacing each mesh vertex by a distance d_r in the vertex normal direction such that $\vec{v}'_r = \vec{v}_r + d_r \vec{n}_r$. If we let d_r be a constant offset distance d_{max} then the distance of all points on the offset surface M' is less than or equal to the offset distance d_{max} from the original mesh M . The offset surface is a continuous mesh but may not be a simple manifold due to self-intersection. Figure 2(b) illustrates the offset surface for a cross-section through a mesh. We can now define a volumetric envelope around M by two offset meshes M^+ and M^- such that each vertex is displaced by a distance d_{max} and

Figure 2: Volumetric surface representation

$-d_{max}$ in the normal direction respectively. The space enclosed by M^+ and M^- is a closed volumetric envelope such that every point inside this region is less than d_{max} from the mesh M . The volumetric envelope enclosed by the offset surfaces is illustrated for a cross section through a mesh in Figure 2(c).

For each triangle $t_j = [\vec{v}_r, \vec{v}_s, \vec{v}_t]$ in mesh M the offset meshes M^+ and M^- define a closed *normal volume*, $V_N(t_r)$, between $t_j^+ = [\vec{v}_r^+, \vec{v}_s^+, \vec{v}_t^+]$ and $t_j^- = [\vec{v}_r^-, \vec{v}_s^-, \vec{v}_t^-]$. The normal volume for a triangular element is illustrated in Figure 3(a). The concept of a *volumetric envelope* and *normal volume* or *fundamental prism* has previously been introduced [Cohen et al., 1996] to define a representation for mesh simplification with bounded approximation error.

Each side of the normal volume is a surface S_{rs} constrained by three vectors: the triangle edge $\vec{e}_{rs} = \vec{v}_s - \vec{v}_r$; and the corresponding vertex normals $\vec{n}_r$ and $\vec{n}_s$. If the vertex normals are equal $\vec{n}_r = \vec{n}_s$ then the surface S_{rs} is a plane. In general if the normals are not equal we can satisfy the constraints by defining the surface S_{rs} as a bilinear or Bezier patch [Cohen et al., 1996].

The volumetric envelope for mesh M between offset meshes M^+ and M^- is equivalent to the union of the normal volumes for all triangles $\bigcup_{r=0, N_T} V_N(t_r)$. Therefore, transforming the mesh M to an approximate volumetric representation can be reduced to an incremental process of transforming the normal volume for each triangle, t_r as follows:

1. Evaluate the normal volume $V_N(t_r)$.
2. Find the set of voxel centres inside the normal volume, $\{\dots, \vec{c}_1, \dots\} \in V_N(t_r)$.
3. For all voxel centres inside the normal volume $\{\dots, \vec{c}_1, \dots\}$ construct a local planar surface approximation from the nearest point on t_r and the corresponding normal, $[\vec{p}_1, \vec{n}_1]$.

To ensure that the normal volume, $V_N(t_r)$, encloses all points less than d_{max} from t_r we can evaluate the vertex offset distance for the j^{th} vertex, d_j , as: $d_j = \frac{d_{max}}{(\vec{n}_r \cdot \vec{n}_j)}$. It is

Figure 3: Normal-volume triangle representation

assumed that for any triangle in mesh M the angle between a triangle normal and adjacent vertex normal is $\leq 90^\circ$ to avoid the degenerate case where the normal volume reduces to zero. To obtain a closed volumetric envelope for a voxel size Δc it is necessary to set the offset distance $d_{max} \geq \sqrt{3}\Delta c$. This gives a discrete field function $f(\vec{x})$ representation for all points with offset distance less than $\frac{\sqrt{3}}{2}\Delta c$ of the implicit surface $f(\vec{x}) = 0$ or mesh M .

Efficient encoding of a triangle in a volumetric representation can be implemented by first evaluating the bounding box for the normal volume on the voxel grid $[\vec{i}_{min}, \vec{i}_{max}]$. Then testing each voxel inside the bounding box if the voxel center $\vec{c}_i$ is inside the normal volume. Testing if a voxel center is inside the normal volume can be implemented by evaluating if the point is inside each of the normal-volume sides. Not all the voxels in the bounding box need to be tested as each grid row of voxels can only enter and leave the volume once.

2.3 Geometric Fusion

Fusion of multiple overlapping meshes, $M_0, \dots, M_N$, can be achieved using the volumetric surface representation introduced in the previous section. For each mesh M_i we can define a closed offset envelope between M_i^+ and M_i^- with offset distance d_{max} . Thus for each

triangle t_r , we can define a normal volume $V_N(t_r)$ and incrementally transform the mesh M_i to a volumetric representation. In overlapping regions of surface measurements from different meshes the normal volumes will intersect provided the maximum measurement error $e_{max} \leq d_{max}$. If this condition is satisfied we can combine the field functions from different meshes into a single volumetric representation. In practice this condition can always be satisfied for a particular sensor by setting the voxel size Δc approximately equal to the maximum measurement error.

Fusion of multiple meshes requires an ‘overlap’ test to determine if surface measurements from different meshes in close spatial proximity correspond to the same or different regions of the measured object surface. Definition of a robust overlap test is critical for reliable surface reconstruction as discussed in previous work [Hilton et al., 1996]. As in previous work geometric constraints are used to estimate if overlapping measurements correspond to the same surface region based on the following criteria:

1. Spatial proximity: distance between overlapping measurements is less than the maximum distance d_{max} .
2. Surface orientation: overlapping surface normals with the same orientation $n_r \cdot n_t > 0$.
3. Measurement uncertainty: likelihood of measurement overlap based on estimates of measurement uncertainty.

Spatial proximity provides a coarse test of measurement overlap which has been used in previous work [Curless and Levoy, 1996]. However, this test is unreliable for sharp edges and for surfaces in close proximity. Surface orientation enables reliable reconstruction of crease edges and thin surface sections for continuous implicit surface representation [Hilton et al., 1996]. Some previous discrete volumetric representations for fusion of range image data [Curless and Levoy, 1996] stored only position information which does not allow a surface orientation test to be performed. The use of an orientation test increases robustness for reconstruction of complex surfaces. Surface orientation was previously used with a discrete volumetric representation by Roth et al. [Roth and Wibowo, 1995].

Measurement uncertainty for surface measurements can be estimated from the relative orientation of the surface and sensor viewpoint [Soucy and Laurendeau, 1995b, Hilton et al., 1996]. If overlapping measurements are determined to correspond to the same surface region they

may be combined according to a weighted average [Turk and Levoy, 1994] or maximum confidence [Pito, 1996a]. The weighted average for the $\underline{i}^{th}$ voxel is evaluated as:

$$\vec{p}_{\underline{i}} = \frac{\sum_{j=1}^{N_{\underline{i}}} w_{\underline{i}}(t_j) \vec{p}_{\underline{i}}(t_j)}{\sum_{j=1}^{N_{\underline{i}}} w_{\underline{i}}(t_j)} \quad \vec{n}_{\underline{i}} = \frac{\sum_{j=1}^{N_{\underline{i}}} w_{\underline{i}}(t_j) \vec{n}_{\underline{i}}(t_j)}{\sum_{j=1}^{N_{\underline{i}}} w_{\underline{i}}(t_j)} \quad (4)$$

where, $N_{\underline{i}}$ is the number of measurements and $w_{\underline{i}}(t_j)$ is the weight based on the measurement confidence for triangle t_j .

Fusion of overlapping surface measurements is illustrated in Figure 4. The cross-section through overlapping volumetric envelopes for two meshes are illustrated in Figure 4(a). The resulting volumetric envelope and integrated iso-surface after geometric fusion is illustrated in Figure 4(b). Overlapping regions of the volumetric representation are combined to obtain a single iso-surface according to the algorithm presented above.

The computational and memory cost of the volumetric representation depend on the number of voxels inside the volumetric envelope. For a free-form surface this is proportional to the surface area or the square of the voxel size. Explicit storage of all voxels has $O(N_v^3)$ memory cost which is prohibitively expensive. A run-length encoded voxel structure [Curless and Levoy, 1996] is used to achieve a storage cost which is proportional to the number of voxels inside the volumetric envelope. This is proportional to the surface area resulting in a relatively efficient storage cost of $O(N_v^2)$. The representation accuracy is inversely proportional to voxel size [Hilton et al., 1998].

2.4 Triangulation

Having, constructed a fused volumetric implicit surface representation from the measurement data an explicit triangulated mesh representation can be extracted using an iso-surface polygonization algorithm. Marching Cubes is applied to obtain a fused mesh representation of the object surface [Bloomenthal, 1994, Lorensen and Cline, 1987].

Figure 5 shows reconstructed triangulated models of complex object using the geometric fusion algorithm. The top two objects were captured using the ModelMaker hand-held sensor system. Both data sets consist of approximately one million raw data points which are triangulated to form several hundred overlapping surface patches. Geometric fusion of the hand-held sensor data on a 200MHz Pentium PC platform is under two minutes for both objects. The resulting models consist of the order of one hundred thousand polygons. The bottom objects were acquired using conventional laser range image sensors

Figure 4: Fusion of field-functions for cross-sections through overlapping volumetric envelopes

from multiple views using a Cyberware scanner (bunny) and an NRCC scanner (knight) [Rioux, 1984]. The data sets consist of eight and ten range images respectively and take of the order of one minute on a SUN Sparc10 platform to reconstruct a fused model of several hundred thousand polygons.

These results demonstrate that the geometric fusion algorithm presented in this paper can be used for both conventional and hand-held range sensors. The data captured using the hand-held range sensor has approximately 0.2mm rms error resulting from the accuracy of the Faro Bronze articulated arm used. In addition there are offset errors for overlapping surface measurements taken on different sweeps of the sensor of the order of 0.2mm. The resulting models were reconstructed using a 0.2mm voxel resolution to integrate overlapping surface measurements. A smaller voxel resolution results in visible ridges on the model where overlapping measurements fall into different voxels. For the conventional range sensor data the rms error is $50\mu m$ or less due to the electro-optical noise of the sensor. The models were reconstructed with a 0.1mm voxel size resulting in a smooth surface model without visible artifacts due to the discrete voxel resolution. The volumetric approach introduced in this paper reconstructs correct surface models for both sources of data.

2.5 Geometric Fusion for Large Objects

Hand-held 3D sensors can be mounted on instrumented platforms or large articulated arms to enable capture of large objects [Levoy, 1999]. The problem is then to reconstruct a single surface model to the required accuracy.

Reconstruction of large objects at high-resolution is prohibitively expensive using a single resolution discrete volumetric representation. For example to reconstruct an object

Figure 5: Reconstructed models for a hand-held sensor(top) and conventional range image sensor (bottom)

the size of a car requires a volume of approximately $2.5m^3$. However, reconstruction accuracy must be of the order of 0.2mm to accurately represent fine surface details such as edges. The number of voxels would therefore be $N_v^3 = 12500^3$. With run-length encoding this gives approximately $N_v^2 = 156M$ occupied voxels with a memory requirement of 1.5Gb. This is prohibitively expensive, therefore we require techniques for reconstruction which do not require high-resolution reconstruction of objects independent of size.

The problem of reconstructing models of large objects has been addressed by subdividing the object space into a series of adjacent sub-volumes. Geometric fusion for measurements in each sub-volume is performed using the algorithm presented in the previous section to obtain an implicit surface representation. Sub-volumes are setup to overlap by exactly one voxel. The field-function $f(\vec{x})$ for the overlapping sub-voxels is exactly the same as identical input data are used to evaluate the field function. The Marching Cubes algorithm for implicit surface polygonization uses the field function values at the centre of each voxel to determine the intersection of edges between voxel centres with the iso-surface $f(\vec{x}) = 0$. Each intersection with the iso-surface results in a new vertex in the output fused mesh model. In the overlapping voxels of the sub-volumes the field-function values are the same resulting in the same iso-surface edge intersections. Marching Cubes generates a set of output mesh vertices on the boundary of the sub-volume which are identical for adjacent sub-volumes. Therefore, a single mesh representation can be extracted from multiple sub-volumes simply by merging output mesh vertices on the boundary of the sub-volume.

This approach has been used to obtain models of objects which require prohibitively large memory for a single volumetric representation. Sub-volumes were used to reconstruct the dwarf model shown in Figure 5 where the space was split into four sub-volumes. The reconstruction time using the sub-volume approach is the same as using a single volume. However, the peak memory requirement is one quarter of the memory requirement for reconstructing the model using a single volume. There is no difference in the resulting surface model using the sub-volume approach.

The computational cost of multiple sub-volumes is the same as for a single large-volume. Therefore, this process reduces the memory usage without requiring increased computational cost. Spatial subdivision with geometric fusion in multiple sub-volumes allows reconstruction of arbitrarily large objects without restrictions due to finite computer memory capacity.

3 Conclusions

In this paper we have introduced a new geometric fusion algorithm for reconstruction of 3D object surface models from hand-held sensor data. The principal advance of this approach is that it avoids the assumptions made in previous fusion algorithms developed for conventional range image data based on a grid structure with a common viewpoint. Results are presented for reconstruction of surface models from both hand-held sensor data and conventional range images.

The new geometric fusion algorithm is based on the use of the *normal-volume* of a triangle. This normal-volume is used to incrementally transform an arbitrary triangulated mesh into a discrete volumetric implicit field-function representation. This approach avoids any prior assumptions of a grid structure or common viewpoint. The field-functions for overlapping surface measurements can be integrated to define a single representation. Implicit surface polygonization is performed using the Marching Cubes algorithm to generate a fused mesh model of the object surface.

Large objects can be reconstructed by sub-dividing the space into overlapping sub-volumes. Geometric fusion is performed independently for each sub-volume. The resulting surface models for each sub-volume are then merged into a single model. This approach allows fusion of data for arbitrarily large object with finite memory resources. The computational cost is equivalent to fusion for a single volume enclosing the entire space, whereas the storage cost is equivalent to a single sub-volume.

Acknowledgement

This research was supported by the EPSRC, UK Funding Council on Advanced Fellowship AF/95/2531 held by Dr. Adrian Hilton and Grant GR/89518 ‘Functional Models: Building Realistic Models for Virtual Reality and Animation’.

References

- [Bloomenthal, 1994] Bloomenthal, J. (1994). An implicit surface polygonizer. *Graphics Gems ed. Heckbert, P.S.*, 4:324—350.
- [Boissonnat, 1984] Boissonnat, J. (1984). Geometric structures for three-dimensional shape representation. *ACM Transactions on Graphics*, 3(4):266—286.
- [Cohen et al., 1996] Cohen, J., Varshney, A., Manocha, D., Turk, G., Weber, H., Agarwal, P., Brooks, F., and Wright, W. (1996). Simplification Envelopes. In *ACM Computer Graphics Proc., SIGGRAPH, New Orleans, USA*, pages 119—128.

- [Curless and Levoy, 1996] Curless, B. and Levoy, M. (1996). A Volumetric Method for Building Complex Models from Range Images. In *ACM Computer Graphics Proceedings, SIGGRAPH, NewOrleans, USA*, pages 303–312.
- [Datalabs, 1996] Datalabs, V. (1996). *Viewpoint Catalog*. (<http://www.viewpoint.com>).
- [Fisher et al., 1996] Fisher, R., Fitzgibbon, A., Gionis, A., Wright, M., and Eggert, D. (1996). A hand-held optical surface scanner for enviromental modeling and virtual reality. In *Proc. Virtual Reality World, Stuttgart, Germany*, pages 13–15.
- [Fitzgibbon et al., 1998] Fitzgibbon, A. W., Cross, G., and Zisserman, A. (1998). Automatic 3D model construction for turn-table sequences. In Koch, R. and Van Gool, L., editors, *3D Structure from Multiple Images of Large-Scale Environments, LNCS 1506*, pages 155–170. Springer-Verlag.
- [Gortler et al., 1996] Gortler, S., Grzeszczuk, R., Szeliski, R., and Cohen, M. (1996). The lumigraph. In *ACM Computer Graphics Proc., SIGGRAPH, NewOrleans, USA*, pages 43–54. NewOrleans, USA.
- [Hilton et al., 1996] Hilton, A., Stoddart, A., Illingworth, J., and Windeatt, T. (1996). Reliable surface reconstruction from multiple range images. In *4th European Conference on Computer Vision*, pages 117–126. Springer.
- [Hilton et al., 1998] Hilton, A., Stoddart, A., Illingworth, J., and Windeatt, T. (March 1998). Implicit surface based geometric fusion. *International Journal of Computer Vision and Image Understanding, Special Issue on CAD Based Vision*, 69(3):273–291.
- [Hoppe et al., 1992] Hoppe, H., DeRose, T., Duchamp, T., McDonald, J., and Stuetzle, W. (1992). Surface reconstruction from unorganised points. *Computer Graphics*, 26(2):71–77.
- [Infografica, 1996] Infografica, R. (1996). *REM 3D Models Bank Catalog*. (<http://www.infografica.com>).
- [Levoy, 1999] Levoy, M. (1999). The Digital Michelangelo Project. In *Second IEEE International Conference on 3D Digital Imaging and Modeling, Ottawa, Canada*.
- [Levoy and Hanrahan, 1996] Levoy, M. and Hanrahan, P. (1996). Light Field Rendering. In *ACM Computer Graphics Proc., SIGGRAPH, NewOrleans, USA*, pages 31–42.
- [Lorensen and Cline, 1987] Lorensen, W. and Cline, H. (1987). Marching cubes: A high resolution 3d surface construction algorithm. *Computer Graphics*, 21(4):163–169.
- [Menc1, 1995] Menc1, R. (1995). A graph-based approach to surface reconstruction. In *EUROGRAPHICS 14(3)*, pages 446–456.
- [Niem and Wingebermhle, 1997] Niem, W. and Wingebermhle, J. (1997). Automatic reconstruction of 3d objects using a mobile monoscopic camera. In *Proc. IEEE International Conferences on Recent Advances in 3D Digital Imaging and Modeling, Ottawa, Canada*, pages 173–180.
- [Pito, 1996a] Pito, R. (1996a). Mesh integration based on co-measurements. In *International Conf. on Image Processing*, pages 397–400. Lusanne, Switzerland.
- [Pito, 1996b] Pito, R. (1996b). A sensor based solution to the next best view problem. In *International Conf. on Pattern Recognition*, pages 941–945. Vienna, Austria.
- [Rioux, 1984] Rioux, M. (1984). Laser range finder based on synchronized scanners. *Applied Optics*, 23(21):3837–3844.
- [Roth and Wibowo, 1995] Roth, G. and Wibowo, E. (1995). A fast algorithm for making mesh-models from multiple-view range data. In *Proceedings of the 1995 Robotics and Knowledge Based Systems Workshop*, pages 349–355.
- [Rutishauser et al., 1994] Rutishauser, M., Stricker, M., and Trobina, M. (1994). Merging range images of arbitrarily shaped objects. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 573–580.

- [Soucy and Laurendeau, 1995a] Soucy, M. and Laurendeau, D. (1995a). A dynamic integration algorithm to model surfaces from multiple range images. *Machine Vision and Applications*, 8:53–62.
- [Soucy and Laurendeau, 1995b] Soucy, M. and Laurendeau, D. (1995b). A general surface approach to the integration of a set of range views. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14(4):344–358.
- [Taubin, 1995] Taubin, G. (1995). Estimating the tensor of curvature of a surface from a polyhedral approximation. In *Proceedings IEEE International Conference on Computer Vision, Boston, USA*, pages 902—907.
- [Turk and Levoy, 1994] Turk, G. and Levoy, M. (1994). Zippered Polygon Meshes from Range Images. In *ACM Computer Graphics Proceedings, SIGGRAPH, Orlando, Florida*, pages 311–318.

Hilton

Illingworth

Author Bibliography

Dr. Adrian Hilton received a B.Sc.(Hons.) in Mechanical Engineering and a D.Phil. degree from the University of Sussex in 1988 and 1992 respectively. In 1992 he joined the CVSSP at the University of Surrey working on 3D shape capture for realistic computer graphics and accurate inspection. This work received Best Paper awards from the journals Pattern Recognition in 1996 and IEE Electronics and Communications in 1999. Collaboration with 3D Scanners Ltd. resulted in the first hand-held system for capturing 3D object models which was awarded a 1996 EU IT 'Grand Prize' for innovation. In 1997 he was awarded a five year EPSRC Advanced Fellowship and in 1999 became lecturer in 3D Digital Media and Broadcast. Research into 3D human modelling from digital photographs has been exploited by AvatarMe Ltd. to develop the first 3D booth for generating animated models of people. His active research interests include model-based 3D vision for capturing human movement and photo-realistic 3D models for computer animation.

Professor John Illingworth is Professor of Machine Vision at Surrey University. He has a doctorate in Physics but has been active in research in computer vision, image processing and pattern recognition for the last 18 years. He has published over 130 papers in refereed journals and conferences. He is a Fellow of the Institution of Electrical Engineers and edits IEE Proceedings on Vision Image and Signal Processing.

Figure Captions

Figure 1: Model reconstruction from hand-held sensor data

- (a) 3DScanners ModelMaker system
- (b) Raw point stripes
- (c) Triangulated patches (partial)
- (d) Fused model triangulation
- (e) Fused model surface

Figure 2: Volumetric surface representation

- (a) Mesh and vertex normals
- (b) Offset surface
- (c) Volumetric envelope

Figure 3: Normal-volume triangle representation

- (a) Triangle normal-volume
- (b) Bilinear side of normal-volume

Figure 4: Fusion of field-functions for cross-sections through overlapping volumetric envelopes

- (a) Overlapping field-functions
- (b) Integrated field-function

Figure 5: Reconstructed models for a hand-held sensor(top) and conventional range image sensor (bottom)

Corresponding Author

Adrian Hilton

Centre for Vision, Speech and Signal Processing

University of Surrey

Guildford

GU25XH

UK

Tel: +44-1483-873956

Fax: +44-1483-876031

a.hilton@surrey.ac.uk