3D Shape Capture for Archiving and Animation

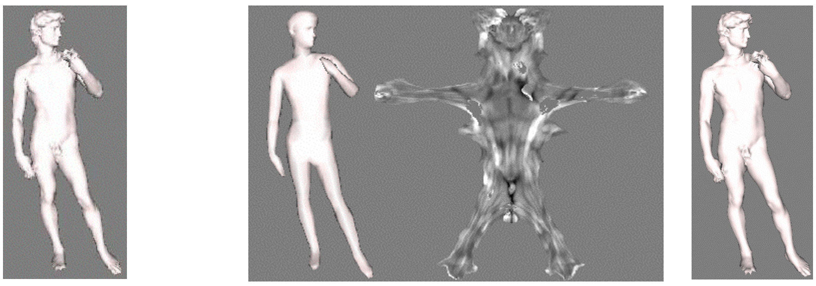
Adrian Hilton, Jonathan Starck, Gordon Collins, Michael Kalkavouras

Centre for Vision, Speech and Signal Processing
University of Surrey, Guildford GU25XH, UK
a.hilton@surrey.ac.uk

Abstract. This paper presents a novel model-based framework for rapid transformation of captured surface shape into a structured representation suitable for efficient storage, transmission, manipulated, visualization and animation. Captured surface shape can be derived from either active 3D sensors or passive reconstruction from multiple view images and video. Results are presented for archiving of historical artifacts and production of animated models of real people in a multiple camera studio.

1 Introduction

Over the past decade commercial systems for static 3D shape capture have become widely available. Computer vision research has also achieved automatic visual reconstruction of 3D surface shape from multiple view images using passive techniques such as shape-from-silhouette, stereo and structure from motion. Research has principally focused on the problem of reconstruction of surface models as polygonal meshes [1-4] and optimization for efficient shape representation [5-6]. However, only limited research has been conducted on the problem of transforming surface measurements into `functional models' [7] which are optimized with respect to the requirements of a particular application such as animation or visualization. In this paper we present model-based techniques developed to bridge-the-gap between 3D shape capture and the functional requirements of particular applications. Figure 1 shows our approach applied to Michelangelo's David[1] for efficient transmission.



(a) Original scan (b) Model-based representation (c) Reconstruction

Fig.1: Michelangelo's David reduced from 100Mb to <1Mb
(Original Courtesy Stanford Computer Graphics Lab. [1])

2 Model-based Representation of 3D Shape

In this section we present a model-based framework for reconstruction of structured *functional models* from capture data. The model-based approach utilises a priori knowledge of the functional requirements of a particular model such as the internal articulation structure for purposes of animation. Captured 3D surface measurements of natural organic objects using either passive or active sensors may have a wide variation in shape and pose and do not contain any information on the underlying non-rigid structure. Prior knowledge in the form of a generic model or user input can be utilized to explicitly identify the object pose and deviation from a generic class of objects. This leads to both efficient representation of highly detailed surface measurements and functional models suitable for computer animation. In this section we first give an overview of the model-based reconstruction framework then present each stage, further details can be found in [8,9].

2.1 Overview

The model-based framework for reconstruction of a functional representation given a generic model of the object comprises the following stages:

1. Manual Registration: The pose of the generic model is aligned with the captured 3D surface measurements.

2. Shape Fitting: Shape constrained least-squares minimization is used to conform the generic model to approximate the captured data.

3. Displacement Mapping: High-resolution surface detail is mapped onto the generic model to preserve the original surface detail.

Generic models for a wide variety of objects which have been optimized for efficient representation of shape and/or efficient animation are available from public model repositories on the web and companies such as ViewPoint DataLabs (www.viewpoint.com). For novel unknown objects a structured model can be derived directly from the the captured data subject to user specified constraints.

2.1 Manual Registration

Initially the generic model is manually aligned with the captured surface data by manually identifying key feature correspondences. The pose of the model is then optimized using least-square minimization of the distance between model and data features. For a static rigid model this recovers the six degrees-of-freedom defining translation and orientation of the model. In the case of articulated models (such as the human skeleton) or non-rigid surface models defined by key-features (such as the human face) optimization is also performed with respect to the additional parameters.

2.2 Shape Fitting

A novel shape-constrained fitting algorithm for arbitrary meshes was introduced in [9,10]. This algorithm is used to non-rigidly deform the generic model surface to approximate the shape of the captured surface data. A requirement for efficient representation and natural animation of the captured object is to preserve the generic model structure during fitting. Shape constraints in the fitting process ensure that the generic mesh parameterisation is preserved throughout the fitting process. The novelty of this approach lies in a unique parameterisation for arbitrary triangular meshes. Parameterisation of arbitrary triangular meshes [10] is achieved by defining the vertex location in terms of the faces surrounding a particular triangle.

A standard deformable surface energy minimization framework is used which balances the external energy, *P*,associated with fitting the data with an internal energy, *S*, of the shape constraints.

$$E(\vec{x}) = P(\vec{x}) + S(\vec{x}) \quad \textbf{(1)}$$

The mesh parameterisation is used to define an internal energy constraint on the local surface shape and internal structure.

$$S(\vec{x}) = \sum_i \sum_f \frac{\left\| \vec{x}_i - \vec{x}_{if}(\alpha, \beta, h) \right\|^2}{N_i} \quad \textbf{(2)}$$

where (α,β) are the barycentric coordinates and h the height offset in the f^{th} face based frame for the i^{th} vertex. This local shape constraint ensures that the local mesh parameterisation is preserved. The external energy is defined from the distance between model vertices x and the data points y using all-neighbour assignment [9].

$$P(\vec{x}) = \sum_i \sum_j m_{ij} \left\| \vec{y}_i - \vec{x}_j \right\|^2 \quad \textbf{(3)}$$

$$\sum_i m_{ij} = 1 \qquad \sum_j m_{ij} = 1$$

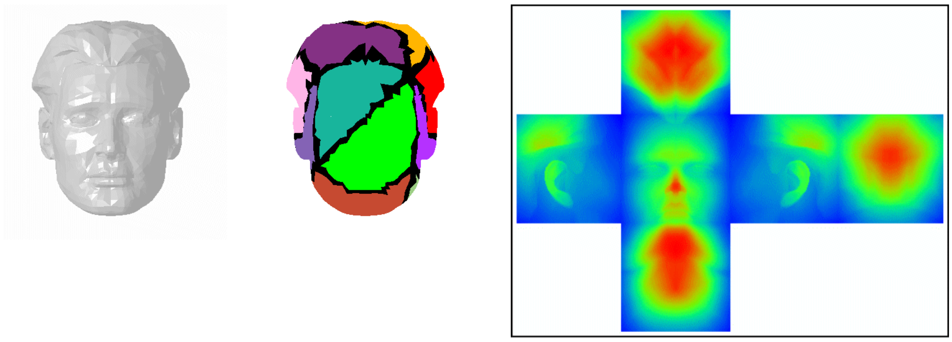
All-neighbour assignment ensures robust point matching where weights *m* are determined iteratively. This approach offers advantages over widely used nearest point matching metrics [9].

2.3 Displacement mapping of surface detail

Shape constrained fitting gives a good approximation to the object shape as illustrated in Figure 1(b). However, due to the relatively low-resolution nature of the generic model the resulting model only provides a smooth approximation. For natural organic objects such as people there is a large shape variation, therefore deviations will occur between the generic model either parametric or polygonal due to inability to directly represent surface detail. In previous work [8,9] displacement mapping techniques were introduced to represent the high-resolution surface detail as offsets from the low-resolution model.

The normal-volume [8] enables injective mapping of the captured data onto the generic model for a triangulated mesh. To form the normal volume each triangle vertex is offset along the vertex normal creating a continuous volumetric envelope surrounding the mesh. Any point from the triangle surface can then be mapped onto a corresponding control model point. Captured data points can then be represented in terms of their barycentric coordinates on a control model triangle and the offset from the surface.

A displacement map image is produced by computing the mapping for the verticies of each triangle and mapping to a 2D image space as in texture mapping. The distance to the surface is then resampled at image pixels. Figure 2 illustrates this process for a head model which is mapped to a simple cube. Figure 2(c) shows the resulting displacement map image where distance is psuedo-colour mapped. The original surface is reconstructed by re-sampling the image to offset point on the generic control model. Figure 1(c) shows the reconstructed model for Michelangelo's David which is visually indistinguishable from the original. The displacement map representation is highly efficient as the high-resolution mesh topology is represented implicitly in the image structure and the high-resolution geometry (offset distance) can be quantised to the desired accuracy.



a) Captured Shape b) Mapping to Cube c) Displacement map image
Fig. 2: Displacement map of head-to-cube using normal-volume mapping

3. Example Applications

The framework presented in section 2 has been used to derive model-based object representations from both active and passive 3D shape capture. Figure 1 demonstrates the representation of historical artifacts captured using high accuracy 3D surface measurement devices[1]. The application of a model-based framework enables an efficient structured representation to be reconstructed. The gross surface shape and topology are represented by the generic model surface, high-resolution surface detail is then efficiently represented by a displacement images as shown in Figure 1(b). Typically this reduces the storage required by two order of magnitude. The structured representation provides a straight forward mechanism for reconstruction at multiple level-of-detail by re-sampling the displacement map image at the desired resolution[9]. Adaptive sampling can be used to control level-of-detail non-uniformly across the surface. The displacement map image representation also facilitates editing of the surface by manipulation of the distance levels in the displacement map image in a process analogous to painting of texture maps. In addition, efficient progressive transmission can be achieved by transmitting the displacement map image as a series of mip-maps of increasing resolution. Standard image compression algorithm can also be applied to the displacement map image.

Figure 3 illustrates the application of the framework to a Cyberware whole-body 3D scan data to create an animated model. A generic articulated human model consisting of approximately 2K polygons is initially manually aligned with the data. Shape constrained fitting followed by displacement mapping is then applied to reconstruct a representation of the high-resolution surface detail which can be animated efficiently.

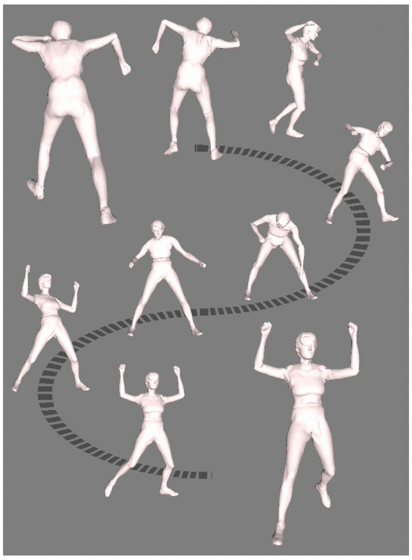


Fig.3: Animation of a whole-body scan

Fig.4: Multi-view reconstruction from images

Figure 4 shows the model-based reconstruction of a dancer model from multiple view images. The visual hull is initially reconstructed by intersecting the silhouette images from multiple calibrated cameras resulting in a discrete volumetric set of occupied voxels. A generic humanoid model is then manually aligned with the voxel set as shown. The resulting shape constrained fitting of the generic model to the voxel set is shown in the middle row. This illustrates that the model-based approach using shape constrained fitting is able to reconstruct a reasonable approximation in the area of the chest despite the significant error in the voxel set due to visual ambiguity. Finally the reconstructed model can be texture mapped and animated in a virtual scene.

4. Conclusions

This paper has presented a model-based framework for reconstruction of structured models from captured 3D surface shape. Results of applying the approach have been demonstrated for both active and passive 3D surface measurement. Algorithms have been developed to enable robust fitting of a generic structured model to the captured 3D shape. Displacement mapping is then applied for accurate representation of high-resolution surface detail. This approach results in a structured model which achieves efficient representation of captured data for visualization and transmission. The approach can also be used to reconstruct animated models given a known articulation or non-rigid surface structure for the object. Further research is developing the application of this framework to dynamic 3D sequences of surface shape.

References

[1] Levoy, M. et al. 'The Digital Michelangelo Project' ACM SIGGRAPH pp.131-144, 2000

[2] Curless,B. and Levoy,M. 'A Volumetric Method for Building Complex Models from Range Images' ACM SIGGRAPH, pp.115-123,1996

[3] Hoppe,H. et al. 'Surface Reconstruction from Unorganised Points', ACM SIGGRAPH, 1992

[4]A. Hilton, A.J. Stoddart, J. Illingworth, and T. Windeatt. `'Implicit surface based geometric fusion. *International Journal of Computer Vision and Image Understanding,* , 69(3):273--291, March 1998

[5] Hoppe, H. DeRose,T. Duchamp,T. McDonald,J. and Stutzle,W. 'Mesh Optimisation' ACM SIGGRAPH pp.19-26,1993

[6] Isenbug,M. and Snoeyink,J. 'Face Fixer: Compressing polygonal meshes with properties'ACM SIGGRAPH, pp.263-270, 2000

[7] Terzopoulos,D. 'From physics-based representation to functional modeling of highly complex objects' NSF-ARPA Workshop on Object Representation, pp. 347-359, 1994

[8] A. Hilton, J. Starck, and G. Collins. 'From 3d shape capture to animated models' *IEEE Conference on 3D Data Processing, Visualisation Transmission*, June 2002.

[9] J. Starck, G. Collins, R. Smith, A. Hilton, and J. Illingworth. Animated statues. *To Appear Journal of Machine Vision Applications*, 2002.

[10] J. Starck and A. Hilton. Reconstruction of animated models from images using constrained deformable surfaces. In *10th Conf. on Discrete Geometry for Computer Imagery. Vol.2301, Lecture Notes in Computer Science, Spriger-Verlag*, pages 382-391, April 2002