

# Reconstruction of Animated Models from Images Using Constrained Deformable Surfaces

J. Starck, A. Hilton, and J. Illingworth

Centre for Vision, Speech and Signal Processing  
University of Surrey  
Guildford, GU2 7XH, UK  
`j.starck@eim.surrey.ac.uk`

**Abstract.** In this paper we present a technique to generate animated 3D models from a limited set of images. We adopt a model-based approach to shape recovery to obtain a specific model parameterisation for animation. A generic animatable 3D mesh model is first manually posed to match the pose in the images. Shape information from multiple 2D image silhouettes is integrated by reconstructing the occupied volume of an object. The mesh is then deformed to match the reconstructed volume using a robust coarse to fine matching technique. We formulate an irregularly parameterised generic mesh as a shape constrained deformable surface to preserve the prior shape information and specific parameterisation in the model. Our interest lies in the generation of animated human models in a multi-camera studio. We assess the technique using artificial test cases and present results for a dancer in a studio.

## 1 Introduction

We introduce a novel model-based technique to recover shape from a limited set of calibrated images in order to generate animated models of people in a multi-camera studio. Applications include advertising, computer games, video conferencing, user-interface agents, virtual environments and clothing retail.

Active sensing systems such as laser scanning or structured light projection are used to capture highly accurate body surface data through high-cost specialist scanning booths. In the studio setting, passive image based techniques such as stereo reconstruction [7] and shape from silhouette [3,10] have been adopted. These techniques produce static surface models and do not recover the necessary structure to animate the model. The reconstruction of animated human models has been addressed through model-based shape from silhouette [5] and model-based shape from stereo [4]. The model-based approach deforms a generic model to fit the available shape information, the reconstructed model can then be animated using the pre-defined animation structure of the generic model.

Our approach extends current methods of shape from silhouette [15] from 2D silhouette matching to 3D volume matching using the approximate reconstructed volume of an object from multiple image silhouettes, the “visual-hull”. Matching of articulated models in 2D can be problematic as only small differences in shape

and pose can lead to large variations in the apparent image silhouette. We make the assumption that where ambiguities occur in a single 2D view, shape information is available from additional views to resolve the ambiguity in matching. We therefore integrate the shape information in multiple images by matching to the 3D visual-hull. The visual-hull provides only a convex approximation to the shape of an object, we therefore formulate the generic model as a shape constrained deformable surface in matching, to preserve the prior shape information in the model.

Model-based shape from volume has been introduced previously for segmentation from volumetric medical image data using anatomical templates. Terzopoulos et. al. [16] first introduced deformable models as a method of shape recovery. This has proved effective in medical image analysis, exploiting prior shape knowledge in the shape of the model for segmentation while also allowing for significant variation in the final segmented shape [11]. Shape constraints have been included in models using global shape parameters as well as probability distributions on shape variables to capture expected variations in shape [14]. The shape of the human body can be parameterised [17], however we expect significant variability in shape with clothing and consider only a locally constrained shape. Our formulation for a shape constrained surface follows the work of Montagnat and Delingette [12], who made use of the specific parameterisation of 2-simplex meshes to formulate a constrained model. Here we extend the concept of shape constrained models to irregularly parameterised triangular meshes, where we attempt to preserve both the shape and the initial mesh parameterisation that may be optimised for model animation. We also combine the constrained deformable surface model with a coarse to fine matching technique to achieve robust matching of the model to the volume.

This paper describes a novel technique for model-based shape reconstruction of animated models from the approximate volume of an object, using a novel shape constrained deformable model and robust coarse to fine matching of the model to volume surface. Section 2 introduces volume reconstruction from multiple view image silhouettes. Section 3 formulates a shape constrained deformable surface model. In Section 4 we define the assignment of the model to the surface of the reconstructed volume and give the algorithm for coarse to fine matching of the model to the volume surface. Finally in Section 5 we present results for generating animated human models, first in artificial test cases demonstrating the reconstruction of scanned human data-sets and secondly in a real test case of a dancer in a studio.

## 2 Volume Reconstruction

A volumetric representation of an object can be obtained from multiple binary image silhouettes by a technique termed “Volume Intersection”, introduced by Martin and Aggarwal [9]. Each silhouette together with the camera view point forms a generalised cone in space that contains the object. The intersection of multiple silhouette cones provides a bounding volume, termed the “visual hull”,

on the object that approximates the object shape according to the position and number of cameras.

Volume intersection is generally performed by dividing the scene into a set of 3D voxels. The voxels are then labelled as either occupied or vacant according to the intersection with the silhouette cones. A voxel is occupied if its projection overlaps all silhouettes and conversely it is vacant if the projection does not overlap any single silhouette. Here we wish to extract the set of surface voxels to generate a surface model. The volume of interest is divided into  $N \times N \times N$  voxels with  $(N+1) \times (N+1) \times (N+1)$  corners. The projection of each corner is tested for overlap with each image silhouette, a corner being marked as non-overlapping if the projection lies outside any silhouette and over-lapping otherwise. A voxel is defined as occupied if it has at least one over-lapping corner. The surface set of voxels is extracted as the set of occupied voxels 6-connected to the non-occupied voxels. An octree representation of the surface voxels is constructed for efficient retrieval of closest surface voxels in model to voxel matching. The normal for each surface voxel is also calculated using a central difference operator on the occupied voxels.

Accurate camera calibration and silhouette extraction is critical for the reconstruction of detailed 3D shape. Calibration errors and silhouette noise typically lead to incorrect labelling of voxels as unoccupied. Our target environment is the virtual production studio and so we are free to use pre-calibrated fixed cameras and a blue-screen background for silhouette extraction.



**Fig. 1.** Volume reconstruction for 2, 4, 6 and 8 simulated cameras spaced at equal angles around the subject.

It is important to note that volume intersection does not uniquely determine the shape of an object, only the visual hull can be recovered. The visual hull is the

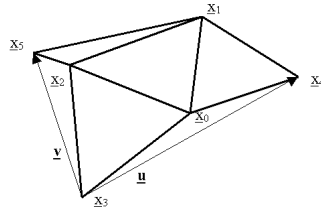
maximal volume that generates the original object silhouettes. Furthermore with only a limited number of viewpoints the apparent occupied volume increases, leading to protrusions around the object as shown in Figure 1. Here we make use of prior information on the human shape by seeking to match a generic human surface model to the surface of the occupied volume.

### 3 Constrained Surface Model

Standardised methods of representing 3D human models are provided by the VRML Humanoid Animation Working Group (H-Anim) [18]. We adopt a H-Anim 1.2 seamless model consisting of a single triangular mesh defining the body shape, attached to an underlying skeleton structure with 17 joints to synthesise the gross movements of the body and texture map(s) attached to change appearance. We make the assumption that the model is first posed to match the subject in the images either by manual positioning or selection of joint positions in the images [15]. The model mesh is then deformed to fit the extracted set of surface voxels.

Deformable surface models were introduced by Terzopoulos et. al. [16] as a method of recovering shape from images. The problem is formulated as an energy minimisation task across the model surface where the potential energy derived from data fitting is minimised along with the internal energy of the model, Equation (1). The deformation of the model in fitting is then regularised by the changes to the internal energy. The most prominent regularising functional treats the surface as a thin-plate material under tension, producing a smooth model that fits the data with a minimum area and minimum distortion.

$$E_{total} = E_{data} + \alpha E_{internal} \quad (1)$$



**Fig. 2.** Facet based local frame. Central vertex locations  $\underline{x}_0, \underline{x}_1, \underline{x}_2$  are defined by barycentric coordinates  $u, v, h$  in outer frame  $\underline{x}_3, \underline{x}_4, \underline{x}_5$

We seek to regularise data fitting by minimising the deformation of a triangular mesh from the default shape and default parameterisation. Local mesh shape has been approximated using quadratic surface patches either in vertex

based or facet based local frames [8]. For a vertex based scheme on an irregular mesh the default position of the central vertex, defining the default shape and parameterisation, is not well defined with respect to the neighbourhood of the vertex. We therefore consider a facet based local patch as introduced by Kobbelt et. al. [8] and shown in Figure (2). The default vertex positions of the central face  $\underline{x}_0, \underline{x}_1, \underline{x}_2$  can be defined by the barycentric coordinates in the frame of the surrounding vertices  $\underline{x}_3, \underline{x}_4, \underline{x}_5$  of the patch. We minimise the deviation of the mesh vertices from the default vertex positions reconstructed from the default barycentric coordinates in each facet based patch, preserving the local shape and relative parameterisation of the mesh. Equation (2) gives the internal energy of the shape constrained deformable surface where  $\underline{x}_i$  is a vertex in the mesh with valence  $N_i$ , and  $(u_{if}^0, v_{if}^0, h_{if}^0)$  are the default barycentric coordinates of the vertex in the  $f^{th}$  facet based frame.

$$E_{internal} = \sum_i \sum_f \frac{\|\underline{x}_i - \underline{x}(u_{if}^0, v_{if}^0, h_{if}^0)\|^2}{N_i} \quad (2)$$

Figure (3) demonstrates the effect of the internal energy on an irregular mesh with added noise and no data fitting term. The mesh recovers the original shape of the model under the shape constraint, even in the presence of severe distortions.



**Fig. 3.** Restoration of an irregular mesh (a) original shape (b) shape + noise (c) 50 iterations (d) 100 iterations (e) 150 iterations

## 4 Model to Volume Surface Matching

Our task is to match the shape constrained active surface model to the surface set of voxels from volume intersection, in order to deform the model to match the apparent shape of a subject in a studio. Kakadiaris et. al. [6] addressed a comparable problem in fitting deformable models to 2D images of body segments, making use of an “all-neighbours” assignment method for model to data matching. Here we use a similar assignment approach. A simple nearest neighbour type assignment would lead to distortions of the surface model according to the initial alignment of the model with the surface voxels. We therefore assign each vertex of the active surface model to all the surface voxels and use the robust point matching technique introduced by Rangarajan et. al. [13] to weight the assignment as the surface deforms.

The energy function we seek to minimise in data-fitting is the least-squared error between the model and the set of surface voxels. The energy function is given by Equation (3) where  $\underline{x}_i$  spans the set of  $I$  model vertices and  $\underline{y}_k$  spans the set of  $K$  surface voxels.

$$E_{data} = \sum_i \sum_k m_{ik} \|\underline{y}_k - \underline{x}_i\|^2 \quad (3)$$

The all-neighbour assignment is defined by a matrix  $M$  of continuous match variables  $m_{ik}$  subject to the following constraints (4): a voxel  $\underline{y}_k$  can match one model vertex  $m_{ik} = 1$ , partially match several vertices  $\sum_i m_{ik} \leq 1$  or may not match a vertex  $m_{ik} = 0$ , and a model vertex  $\underline{x}_i$  must be assigned to the voxels  $\sum_k m_{ik} = 1$ . The inequality constraint is transformed to an equality constraint by augmenting the match matrix by the slack variables  $m_{I+1,k}$ .

$$\begin{aligned} 0 &\leq m_{ik} \leq 1 \\ \sum_{i=1}^{I+1} m_{ik} &= 1 \\ \sum_{k=1}^K m_{ik} &= 1 \end{aligned} \quad (4)$$

We explicitly constrain the model to match the surface voxels. If the vertices were free to not match the voxels, then the model to voxel assignment could be lost where the restoring forces from the internal energy of the model become comparable to the assignment forces in matching. It is important to note that this approach is therefore only appropriate where the majority of the surface area is reconstructed by volume intersection. We rely on the shape constraint energy of the model to prevent excessive distortions in the presence of occlusions.

Rangarajan et. al. [13] make use of deterministic annealing approach to the assignment by adding an entropy term to the energy function. The temperature  $T$  of the entropy term defines the degree of fuzziness in the assignment (5).

$$E_{data} = \sum_i \sum_k m_{ik} \|\underline{y}_k - \underline{x}_i\|^2 + T \sum_i \sum_k m_{ik} (\lg(m_{ik}) - 1) \quad (5)$$

The final energy equation for the shape constrained deformable model is then given by Equation (6).

$$E_{total} = \alpha \sum_i \sum_f \frac{\|\underline{x}_i - \underline{x}(u_{if}^0, v_{if}^0, h_{if}^0)\|^2}{N_i} + \sum_i \sum_k m_{ik} \|\underline{y}_k - \underline{x}_i\|^2 + T \sum_i \sum_k m_{ik} (\lg(m_{ik}) - 1) \quad (6)$$

For a fixed temperature  $T$  and fixed model configuration  $\underline{x}_i$  the match parameters  $m_{ik}$  can be derived by differentiating the energy function with respect to  $m_{ik}$  (7). The Sinkhorn balancing technique of alternating row and column normalisation of the match matrix  $M$  can then be performed to satisfy the match constraints (4) [1]. Equation (7) shows that a vertex assignment is therefore weighted according to the relative distance to a surface voxel  $\|\underline{y}_k - \underline{x}_i\|$  with the temperature  $T$  defining the effective region of matching in space. As the temperature tends to zero this method approaches nearest neighbour matching.

$$\begin{aligned} \frac{dE}{dm_{ik}} &= \|\underline{y}_k - \underline{x}_i\|^2 + T \lg(m_{ik}) = 0 \\ m_{ik} &= \exp\left(-\frac{\|\underline{y}_k - \underline{x}_i\|^2}{T}\right) \end{aligned} \quad (7)$$

Similarly for a fixed assignment we can derive the gradient of the energy function with respect to the mesh vertices  $\underline{x}_i$ , Equation (8), giving a gradient descent solution for the evolution of the mesh. We seek to fit the active surface model to the surface voxels by alternately updating the assignment to the surface voxels  $m_{ik}$  using Equation (7) and Sinkhorn balancing of the match matrix  $M$  to satisfy the constraints in Equation (4), then updating the vertex positions of the model  $\underline{x}_i$  using Equation (8). The temperature is reduced as the model is updated to refine the assignment, giving a coarse to fine approach to matching.

$$\frac{d\underline{x}^t}{dt} = -\frac{dE}{d\underline{x}^t} = -\alpha \sum_i \sum_f (\underline{x}_i - \underline{x}(u_{if}^0, v_{if}^0, h_{if}^0)) + \sum_i \sum_k m_{ik} (\underline{y}_k - \underline{x}_i) \quad (8)$$

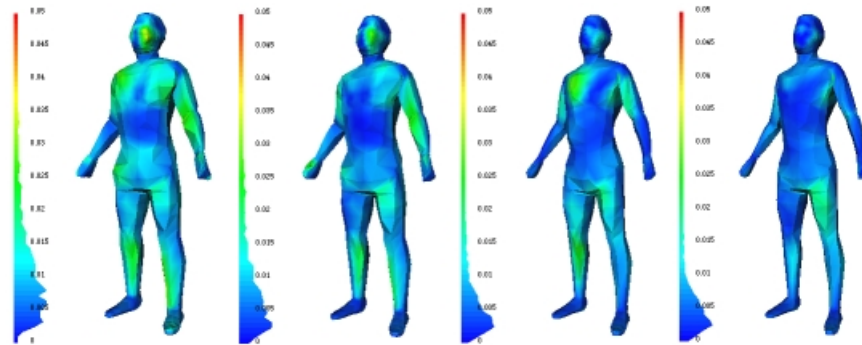
We start at an initial temperature  $T_{init}$  that gives a matching region larger than the greatest expected error, typically we set  $T_{init} = 0.4m^2$ . The match parameters are initialised according to Equation (7) and the slack matches  $m_{I+1,k}$  set to the minimum match  $m_{ik}$  for each voxel  $k$ . We then iteratively update the model and reduce the temperature until the final temperature  $T_{final}$  is reached.  $T_{final}$  is set in the order of the voxel size for a close fit to the voxel surface. At a set temperature and defined matches, the model is updated using explicit Euler integration steps until convergence. The convergence criteria is set as the maximum component of the gradient falling within the current region of matching, that is  $\max \{ \|\frac{dE}{d\underline{x}_i}\| \} \leq \sqrt{T}$ .

The number of surface voxels is potentially large, depending on the voxel size chosen in volume intersection. In practice we therefore restrict the number of voxels used in matching for each vertex to a closest sub-set. This also leads to

more efficient storage of the match matrix  $M$ . We use the octree representation of the surface voxels to efficiently retrieve the closest voxels. We also only match a vertex to those voxels whose surface normal lies in the same half-plane as the vertex normal in order to avoid incorrect surface matches.

## 5 Results

The technique was tested using virtual views generated for human 3D scanned data-sets, provided by Cyberware, in a simulated studio. A varying number of cameras were positioned in the studio at equal spacing and 3m radius around each data-set. An ideal pin-hole camera model was used with NTSC resolution and typical intrinsic parameters. Images of the data-sets were then rendered in OpenGL and used to test the shape reconstruction. Volume intersection was performed using a 1cm voxel size. Both the internal and data fitting energy terms represent geometric errors at the mesh vertices and we give equal weight to both terms in matching,  $\alpha = 1$ . We find this provides adequate data fitting without the need for fine tuning of the parameter  $\alpha$ , although we note that this may be case specific and that this requires further investigation.



**Fig. 4.** Error distribution (/m) in fitting to voxel surface from 2,4,6,8 simulated cameras.

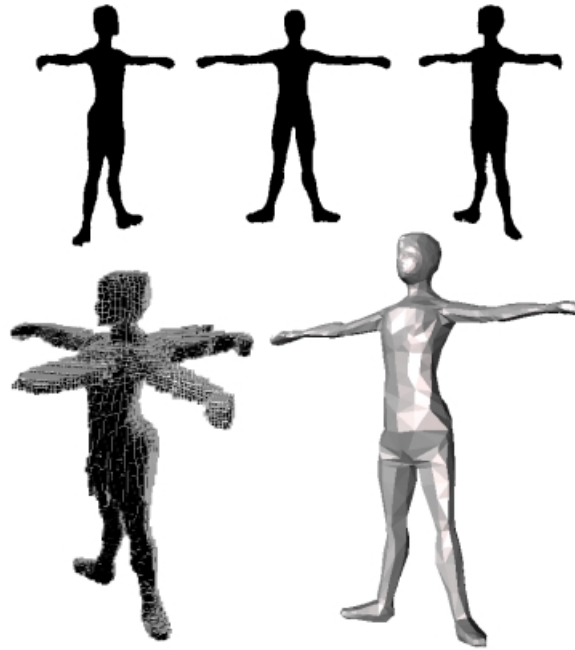
Figure (4) shows the distribution of errors between the deformed active surface model and the original scanned data-set across the surface of the model, for the surface voxels reconstructed from 2,4,6 and 8 cameras as shown in Figure 1. The Metro tool [2] has been used here to measure and display the errors. We can see that the shape constrained surface leads to feasible shape solutions even for a limited number of camera views in the presence of large deviations in the reconstructed volume from the true body shape. The shape constraint of the active surface model leads to a smooth approximation of the body shape. The model cannot recover the fine detail and concavities in the original data-set that are not reconstructed in the visual hull. The illusion of additional surface shape



information would be obtained by texture mapping the smooth model. Figure (5) shows the reconstructed shape for a dancer in a 6 camera studio.

## 6 Conclusions

We have presented a method for 3D reconstruction of animated models from a limited number of camera images. We introduce a model-based approach to shape from silhouette that preserves the prior shape information in the model, regularising data fitting, and the parameterisation of the model for animation. A generic animated model is first manually posed. The model is then matched to multiple image silhouettes simultaneously by fitting to the reconstructed visual-hull of an object. We formulate an irregular triangular mesh model as a shape constrained deformable surface. The mesh is deformed to match the reconstructed volume using a robust coarse to fine matching technique. This has been applied to reconstruct smooth animated models of humans. Further work is required to test this approach for a variety of subjects wearing different clothing in a studio. The task also remains to apply colour texture from the images to generate photo-realistic models.



**Fig. 5.** Reconstructed model for a dancer in a studio showing 3 of the 6 captured image silhouettes and the reconstructed visual-hull

## References

1. H. Chui and A. Rangarajan. A New Algorithm for Non-Rigid Point Matching. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 44–51, 2000.
2. P. Cignoni, C. Rocchini, and R. Scopigno. Metro: measuring error on simplified surfaces. *Computer Graphics Forum*, 17(2):167–174, 1998.
3. L. Davis, E. Borovikov, R. Cutler, D. Harwood, and T. Horprasert. Multi-perspective analysis of human action. In *Proceedings of the 3rd International Workshop on Cooperative Distributed Vision*, Kyoto, Japan, 1999.
4. P. Fua, R. Plankers, C. Miccio, and D. Thalmann. *From Image Synthesis to Image Analysis: Using Human Animation Models to Guide Feature Extraction*. EPFL-LIG Computer Graphics Lab, Switzerland, 1998.
5. A. Hilton, D. Beresford, T. Gentils, R. Smith, and W. Sun. Virtual people: Capturing human models to populate virtual worlds. In *IEEE International Conference on Computer Animation*, May 1999.
6. I.A. Kakadiaris and D. Metaxas. Three-dimensional human body model acquisition from multiple views. *International Journal of Computer Vision*, 30(3):191–218, 1998.
7. T. Kanade, P. Rander, and P.J. Narayanan. Virtualized reality: Constructing virtual worlds from real scenes. *IEEE Multimedia*, 4(1):34–47, 1997.
8. L. Kobbelt, S. Campagna, J. Vorsatz, and H.P. Seidel. Interactive Multi-Resolution Modeling on Arbitrary Meshes. In *SIGGRAPH Conference Proceedings*, pages 105–114, August 1998.
9. W.N. Martin and J.K. Aggarwal. Volumetric descriptions of objects from multiple views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(2):150–158, 1983.
10. W. Matusik, C. Buehler, R. Raskar, S.J. Gortler, and L. McMillan. Image-based visual hulls. In *SIGGRAPH Conference Proceedings*. ACM SIGGRAPH, 2000.
11. T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: a survey. *Medical Image Analysis*, 1(2):91–108, 1996.
12. J. Montagnat and H. Delingette. Volumetric Medical Images Segmentation using Shape Constrained Deformable Models. In *First Joint Conference CVRMed-MRCAS*, volume Lecture Notes in Computer Science 1205, pages 13–22, March 1997.
13. A. Rangarajan, H. Chui, E. Mjolsness, S. Pappu, L. Davachi, P. Goldman-Rakic, and J. Duncan. A robust point matching algorithm for autoradiograph alignment. *Medical Image Analysis*, 4(1):379–398, 1997.
14. L.H. Staib and J.S. Duncan. Model-Based Deformable Surface Finding for Medical Images. *IEEE Transactions on Medical Imaging*, 15(5):720–731, 1996.
15. J. Starck, A. Hilton, and J. Illingworth. Human Shape Estimation in a Multi-Camera Studio. In *BMVC*, volume 2, pages 573–582, September 2001.
16. D. Terzopoulos, A. Witkin, and M. Kass. Symmetry-seeking models and 3D object reconstruction. *International Journal of Computer Vision*, 1:211–221, 1987.
17. D. Thalmann, J. Shen, and E. Chauvineau. Fast human body deformations for animation and VR applications. In *Proceedings of Computer Graphics International*, pages 166–174. IEEE Computer Society Press, June 1996.
18. VRML Humanoid Animation Working Group, <http://www.h-anim.org/>. *H-ANIM 1.1 Specification*, 1999.