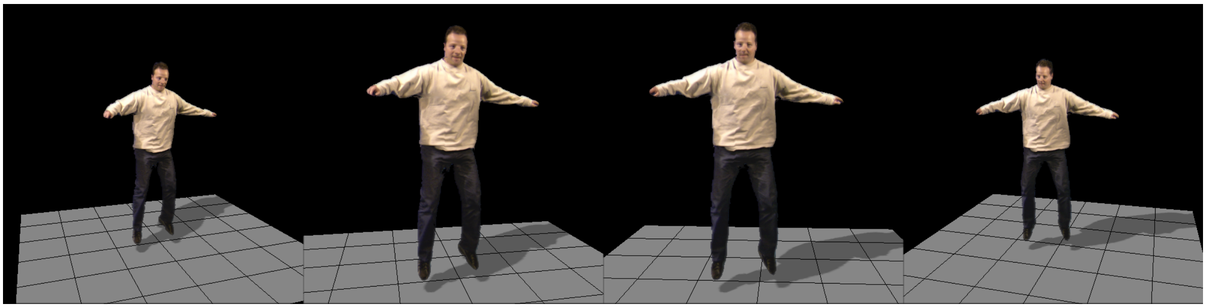
Vision, Video, and Graphics (2003)
P. Hall, P. Willis (Editors)

Towards a 3D Virtual Studio for Human Appearance Capture

J. Starck and A. Hilton

Centre for Vision, Speech and Signal Processing
University of Surrey
Guildford, GU2 7XH, UK



Abstract

This paper introduces the concept of a "3D Virtual Studio" for human appearance capture, akin to the motion capture studio for human motion tracking. Ultimately the 3D Virtual Studio should enable video-realistic reconstruction of a moving person from any viewpoint. A mesh-based stereo technique is presented to reconstruct a moving person from multiple camera views. This technique optimises a surface mesh to match stereo and silhouette data in a constrained coarse-to-fine framework, recovering sub-pixel image correspondence in the presence of inexact camera calibration. We compare this approach for scene reconstruction to conventional shape from silhouette and multiple view stereo. We then demonstrate view-dependent rendering and show improved resolution with the recovered image correspondence. We then demonstrate how this approach can be used to capture the dynamic shape and appearance of a computer graphics model of a person.

1. Introduction

The challenge of creating realistic computer generated scenes is leading to a convergence of computer graphics and computer vision technology. Where computer graphics deals with the complex modelling of objects and simulation of light interaction in a virtual scene to generate realistic images, computer vision offers the opportunity to capture and render such models directly from the real-world with the visual realism of conventional video images.

One of the key challenges is the creation of realistic human models, a central component of most visual media. In the film industry for example we have seen an explosion in the use of computer generated imagery with human models used for stunt doubles in "mixed reality" clips and even to form the entire cast of a major Hollywood film. However, the production of such computer graphics models is currently a high cost and labour intensive task, limiting the application to the big-budget film, advertising and game industries. In the broadcast industry and multi-media production use of computer graphics has instead concentrated on the virtual studio in which action is shot against a constant background and actors can be composited live with real and virtual footage. Recently the concept of using multiple cameras in a virtual studio has been introduced to capture actors in 3D [8]. Three-dimensional production or 3D video was first popularised by Kanade et al. [10] who coined the term "Virtualized Reality". Presenting an event in 3D allows visualisation in the same way as virtual reality, providing an immersive viewing experience.

In this paper we introduce the concept of the multiple camera "3D Virtual Studio" for 3D production of visually realistic computer graphics models of people. We liken this to the development of the motion capture studio that has found widespread use in generating believable movements for character animation. In the 3D Virtual Studio the dynamic shape and appearance of a computer graphics model can be captured from a person.

There are two strands of research in computer vision that will lead to the realisation of a 3D Virtual Studio for human appearance capture. Firstly marker-free visual motion capture from which the motion of person can be tracked from multiple cameras. Secondly accurate scene reconstruction to capture the dynamic shape and appearance of a person moving in the studio. In this paper we concentrate on scene reconstruction. We consider two conventional techniques for reconstruction, shape from silhouette and stereo vision, and introduce a surface optimisation technique to recover surface geometry of a scene integrating both stereo and silhouette data. This new technique for scene reconstruction is applied to optimise geometry and recover sub-pixel accurate image correspondence to render a virtual view of an actor captured in a studio. We demonstrate how this technique can be applied to capture the dynamic shape and appearance of a person as a computer graphics model.

2. Related Work

Acquisition of visually realistic models of real objects and scenes has been a long standing problem in both computer graphics and vision. Range scanning technologies, which actively project a structured light pattern onto the object surface, have been widely used as the basis for reconstruction of accurate models of static scenes [14]. Whole-body scanning systems capture the static shape of a person in a fixed pose. Starck et al. [23] used this technology to generate detailed animated human models from a single whole-body surface scan. Allen et al. [3] captured the upper body shape of a person in multiple static poses to characterise changes in body shape for animation.

Compared to active range scanning, passive reconstruction from images or *image-based modelling*, enables greater flexibility in scene capture and provides the dynamic appearance inherent in video images. Debevec et al. [4] first demonstrated the visual realism that can be achieved in rendering novel views of a static scene from photographs. Kanade et al. [10] demonstrated the ability to recover 3D models of dynamic scenes from multiple video images. Techniques for shape estimation from multiple cameras include reconstruction of volume from image silhouettes, termed the *visual-hull* [13], volume from colour consistency between images, termed the *photo-hull* [22, 12], and surface recovery from stereo correspondence between pairs of camera images [10, 19]. Multiple camera systems have been developed to reconstruct dynamic sequences of people, Moezzi et al. [18] demonstrated the use of the visual-hull, Vedula et al. [25] made use of the photo-hull, and Kanade et al. [10] fused multiple stereo depth-maps into a surface model of a person.

The visual-hull provides a bounding approximation on the shape of a scene and cannot model concavities that are self-occluded in image silhouettes. Colour consistency techniques suffer from holes or false cavities in the volume of a scene where consistency cannot be correctly estimated between views, and the fattening of areas of the scene where there is insufficient colour information in the images to distinguish different surfaces. Finally stereo correspondence can fail in regions of poor image texture or occlusion boundaries and can produce noisy depth estimates with inexact matches between images. Techniques have therefore been considered based on models of the scene geometry and integrating multiple visual cues to improve scene reconstruction. Fua and Leclerc [6] introduced object-centred reconstruction in which an initial surface estimate is optimised to match stereo and shading cues between images. Vedula et al. [26] proposed a model-enhanced stereo system where an initial reconstructed scene model is used to refine the search range for stereo correspondence to improve stereo matches for reconstruction. These techniques make use of reconstructed geometry to improve the estimation of image correspondence. Model-based techniques have been proposed that make use of a prior model of scene geometry to constrain shape recovery in the presence of visual ambiguities such as lack of image texture that makes correspondence ambiguous. Hilton et al. [9] present model-based shape from silhouette to recover whole-body models of people. Plankers and Fua [20] adopt a model consisting of implicit volume primitives to recover the gross upper-body shape and pose from stereo and silhouette data.

In this paper we present a technique to integrate both stereo and silhouette data to optimise either estimated surface geometry or a prior surface model to match multiple camera images. Stereo correspondence is used to optimise surface shape to sub-pixel accuracy for recovery of colour texture. This provides improved resolution in rendering images in the presence of inexact surface geometry or inexact camera calibration compared to current approaches that make the assumption that a reconstructed surface is in correspondence between images [18, 25, 10]. The shape of the model is used to constrain the search for stereo correspondence in a coarse-to-fine framework that enables shape recovery from noisy stereo data. This provides a wider range of convergence compared to local optimisation techniques [6]. The framework incorporates multiple shape cues. This provides improved surface reconstruction in the presence of visual ambiguities compared to techniques that rely on a single shape cue [18, 10].

Techniques for shape from silhouette and stereo correspondence are first presented together with the new approach to optimise surface geometry. These techniques are then compared for scene reconstruction in a multiple camera studio. Rendering a virtual view is then compared for the different approaches using view-dependent rendering. View-dependent techniques are now increasingly used to produce greater visual realism in rendering by using the camera views closest to the novel viewpoint [4, 21, 16]. Finally we compare the object-centred and model-based approach to scene reconstruction and demonstrate the capture of the shape and appearance of animated computer graphics models of people in a multiple camera studio.

3. Scene Reconstruction

Our multiple camera "3D Virtual Studio" contains 8 Sony DXC-9100P 3-CCD colour cameras, providing PAL-resolution progressive scan images at 25Hz. A blue-screen backdrop is used for foreground silhouette segmentation and the set-up provides a capture volume of up to 2.5m x 2.5m x 2.5m. Intrinsic camera parameters are calibrated using the Camera Calibration Toolbox for Matlab from MRL-Intel [2]. Extrinsic camera calibration is performed using a wand-based technique.

3.1. Shape from silhouette

Various techniques have been developed for the reconstruction of the visual-hull from multiple camera images. Here we adopt a simple algorithm to generate the set of volume elements, called voxels, that reproject to the segmented image silhouettes. We divide the scene into a set of N×N×N voxels with (N+1)×(N+1)×(N+1) corners. All voxels are initially set as unoccupied. Each corner is tested for overlap with each image silhouette and each voxel is then set as occupied if at least one corner overlaps all image silhouettes. The discrete volumetric representation is then converted to a surface mesh by isosurface extraction using a variation on the marching cubes algorithm [15].

3.2. Shape from stereo correspondence

Surface reconstruction from stereo is performed by extracting a 2.5D stereo depth-map for each camera pair in the studio. Here we use a two-stage dynamic programming technique proposed by Sun [24] to extract a surface that maximises the stereo correspondence between images and enforces continuity in the depth-map. We use a normalised cross-correlation metric to allow for linear changes in intensity between images with non-Lambertian surfaces or inexact intensity matched images. We also add the constraint that the disparity range for stereo correspondence lies within the visual-hull extracted from image silhouettes. This follows the model-enhanced stereo paradigm proposed by Vedula et al. [26] and removes outliers in stereo correspondence.

Multiple depth-maps are fused into a single surface representation using volumetric fusion as proposed by Narayanan et al. [19]. The volume of the scene is divided into a discrete set of voxels and a signed distance function is computed at each voxel that gives the distance to the surface estimated in the multiple 2.5D depth-maps. A surface mesh is then extracted from the volume using isosurface extraction. The signed distance at a voxel is derived by first projecting the centroid to each depth-map and deriving the distance to the closest surface-point. An average is then taken for all distances within a set tolerance of the closest distance across all depth-maps.

3.3. Surface optimisation

A surface optimisation technique is introduced to deform an initial model of the scene geometry to match stereo and silhouette data. The model deformation is formulated as an energy minimization task [17]. A cost-function is constructed consisting of a potential energy term derived from the fit of the model to the data, and an internal energy term that penalises the deviation from the desired model properties. The model is then deformed to minimize the total energy function, hence minimizing the error between the model and the data while the internal energy regularises the model deformation. In data fitting we use the cost of fitting to stereo data E_S and matching the shape from silhouette provided by the visual-hull E_V. The trade-off between these data terms is governed by a weighting β, and the influence of model regularisation, E_R, is governed by α.

$$E = \beta E_S + (1-\beta)E_V + \alpha E_R \tag{1}$$

We discretize the energy function at the vertices of our mesh $\underline{x}_i$ and use gradient descent for minimization. In terms of physics-based deformable models this is equivalent to a zero mass dynamic system. The deformation of the mesh vertices is then given by Equation 2.

$$\frac{d\underline{x}_i}{dt} = -\frac{dE}{d\underline{x}_i} = -\left(\beta\frac{dE_S}{d\underline{x}_i} + (1-\beta)\frac{dE_V}{d\underline{x}_i} + \alpha\frac{dE_R}{d\underline{x}_i}\right) \tag{2}$$

Stereo matching energy term

In stereo matching we use a direct search for stereo correspondence between images. For each mesh vertex we first determine the key view with the greatest surface visibility according to the camera with the closest viewpoint to the direction of the vertex normal. We then recover the disparity in each stereo pair that uses the key view. Here we make the simplifying assumption of a fronto-parallel surface at each vertex and use area-based normalized cross-correlation between rectified camera images [7]. For each offset image in each stereo pair we locate the sub-pixel match to the key image with the highest correlation score. We define the search

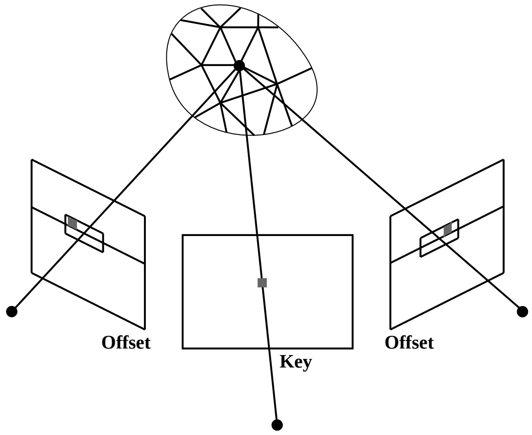


Figure 1: *Stereo matching between key and offset views, showing the search region along each epipolar line allowing for off-axis matches with inexact camera calibration.*

range along the epipolar line in each rectified offset image according to the expected error in the shape of the mesh. We also match up to a specified pixel error perpendicular to each epipolar line according to the expected accuracy of the camera calibration as illustrated in Figure 1.

For each vertex we derive a sub-pixel correspondence in each offset image and a reconstructed 3D position. The stereo energy term at each vertex, $E_S(\underline{x}_i)$, is defined as the squared error between the vertex position and the reconstructed 3D position $\underline{z}_{i,c}$ for each offset camera c. We therefore seek a least-square error fit to the matched vertex positions across the whole mesh, as given in Equation 4. The energy term is weighted according to the quality of the stereo matches as given by the correlation score $w_{i,c} \in [0,1]$, Equation 3. This enables a trade-off between fitting stereo data where good matches are obtained and fitting silhouette data where matching is poor.

$$\beta(\underline{x}_i) = \frac{1}{n_i^c} \sum_{c=0}^{n_i^c - 1} w_{i,c} \tag{3}$$

$$\beta E_S = \sum_i \frac{1}{n_i^c} \sum_{c=0}^{n_i^c - 1} w_{i,c} \|\underline{z}_{i,c} - \underline{x}_i\|^2 \tag{4}$$

In stereo matching it is important to account for self-occlusions to prevent incorrect matches between occluded and visible regions. We deal with self-occlusions by checking the visibility of each mesh vertex in each camera image and only matching between unoccluded views. Here we use the visibility algorithm introduced by Debevec et al. [5] that uses hardware accelerated OpenGL rendering. To test the visibility in a camera, the mesh is rendered to the camera viewpoint with a unique colour ID assigned to each polygon. For each front-facing vertex we can then retrieve the polygon at the projected location in the camera and check for occlusion against the polygon in screen space. In shape optimisation it is feasible to obtain incorrect visibility information as the mesh deforms. We therefore use a conservative visibility check, first by testing the visibility of the deformed mesh, then by checking the visibility of the mesh vertices against potential occluding regions in the visual-hull. The second visibility check is performed against the back-facing polygons of the visual-hull simply by using a back-face render of the visual-hull mesh in each camera view.

Silhouette matching energy term

Stereo matching can fail where texture is lacking in an image or where there is significant distortion in texture between views due to non-frontal surfaces or occlusion boundaries. Silhouette data is therefore incorporated by fitting the volumetric visual-hull as obtained in section 3.1. The visual-hull energy term, $E_V(\underline{x}_i)$, is defined as the squared error between the vertex position and the closest voxel on the visual-hull $\underline{y}_i$.

$$E_V = \sum_i (1 - \beta(\underline{x}_i)) \|\underline{y}_i - \underline{x}_i\|^2 \tag{5}$$

Shape regularisation energy term

The standard approach to shape regularisation is to treat a deformable model as a thin-plate material under tension. Here we use membrane tension for regularisation. The membrane functional for E_R across a parameterised surface $\underline{x}(u,v)$ is given in Equation 6 and the variational derivative is given by the Laplacian $\triangle(\underline{x})$. Under the simplifying assumption of a regular mesh parameterisation, the laplacian at a mesh vertex is given by the "umbrella-operator" in Equation 7 where the index v spans the 1-neighbourhood $\underline{x}_{i,v}$ of a vertex $\underline{x}_i$ [11]. The umbrella operator pulls vertices towards the centroid of the 1-neighbourhood. Intuitively the role of regularisation is to maintain a smooth, even parameterisation of the mesh surface during deformation.

$$E_R = \int\int \left(\|\underline{x}_u\|^2 + \|\underline{x}_v\|^2 \right) du dv \tag{6}$$

$$\frac{dE_R}{d\underline{x}_i} = -\frac{1}{n_i^v} \sum_{v=0}^{n_i^v - 1} (\underline{x}_{i,v} - \underline{x}_i) \tag{7}$$

Coarse-to-fine matching

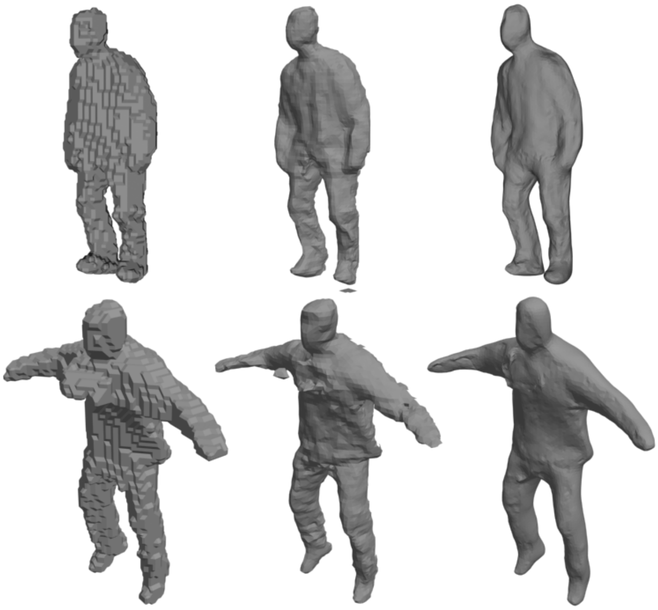
The shape optimisation process is performed in a coarse-to-fine framework in order to deal with noisy stereo matches. We start at an initial expected error for the surface mesh and locate the stereo matches up to the error estimate, together with the closest visual-hull point for each mesh vertex. We then update vertex locations to minimize the energy function. The expected error is successively reduced and the model again deformed to minimize the energy. The convergence criteria at each error level is set to the error estimate multiplied by the error reduction factor and the step length in steepest descent. Optimisation is stopped when the error level reaches the reconstruction accuracy of the camera set-up. The advantage of the coarse-to-fine matching and model deformation is that it allows the scene mesh to converge to a solution in the presence of noisy stereo data, increasing the range of convergence compared to local optimisation techniques [6].

4. View Dependent Rendering

Surface optimisation provides the means to deform an estimated scene geometry to satisfy stereo matching between views in a multiple camera set-up. The framework incorporates silhouette data where stereo matches are poor or not possible due to occlusion and uses shape regularisation to control the model deformation. The result provides sub-pixel accurate image correspondence even with inexact camera calibration in highly textured regions where the alignment of images is important in rendering novel views. For shape from silhouette and stereo correspondence this image correspondence is found by projecting the vertices of the surface mesh into the images.

In view dependent rendering we make use of this image correspondence to render virtual views of the 3D scene. In order to smoothly blend between the camera images in rendering, we adopt a view-dependent vertex weighting as proposed by Pulli et al. [21]. The weight for vertex i with respect to camera c consists of two components, $\lambda_{i,c} = \lambda^1_{i,c}\lambda^2_{i,c}$. The first component defines the visibility of the vertex in each camera, $\lambda^1_{i,c} = \cos(\theta)$, where θ is the angle between the vertex normal and the vector from the vertex to the camera. The second component defines the proximity of the virtual view to each camera, $\lambda^2_{i,c} = \cos(\phi)$, where ϕ is the angle between the vector from the vertex to the camera and the vector from the vertex to the virtual camera. The mesh is textured from the camera views on a per-polygon basis and separate vertex weights for each polygon p are derived by only taking the weights for the views in which all polygon vertices are matched. In the presence of occlusions some polygons will not necessarily have vertices that are all matched in one camera view and so will contain no texture. We therefore also derive vertex colours in order to colour polygons that cannot be textured. Vertex colours are calculated as the weighted average of the matched camera image pixels according to the vertex weight $\lambda_{i,c}$.

The virtual view is generated using hardware accelerated OpenGL rendering. The mesh is first rendered with vertex colouring. Multi-pass texturing is then used to render the mesh from each camera image with the texture modulated by the blend weight at each polygon vertex. In the first instance of texturing a polygon, blending replaces the colour rendered mesh and subsequent passes add modulated texture.



(a) Visual-hull (b) Merged stereo (c) Optimisation

Figure 2: *Comparison of shape reconstruction.*

5. Results

5.1. Scene reconstruction

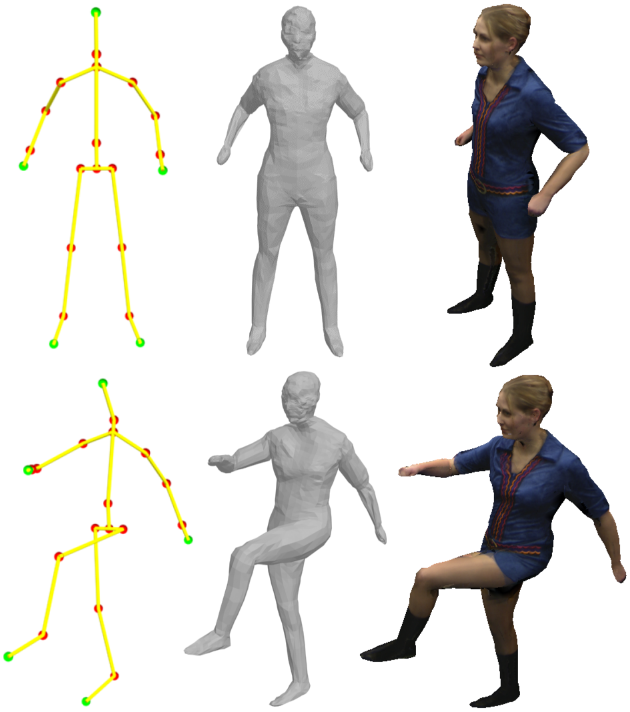
Surface optimisation is first compared with shape from silhouette and stereo vision for geometric reconstruction from multiple camera views. Here the visual-hull is used as an initial estimate of surface geometry and smoothed to obtain a smooth regularisation term from the surface shape. The surface mesh is then optimised starting at an initial expected error of 15cm and finishing at a 1cm error in optimisation. Figure 2 shows the reconstructed scene geometry in comparison to the visual-hull and the surface derived by merging multiple stereo depth-maps. Both stereo reconstruction and surface optimisation demonstrate a similar geometry and an improved shape in comparison with the visual-hull. With a lack of image texture stereo correspondence can fail, leading to noisy surface estimates and missing sections of geometry. Combining silhouette and stereo data demonstrates improved reconstruction where these visual ambiguities arise. All techniques fail to reconstruct the detailed geometry of the face due to a lack of resolution in the video images.

5.2. View-dependent rendering

Current techniques for view generation rely on rendering a novel view using reconstructed scene geometry under the assumption that the scene model is in correspondence between views. Errors in correspondence can arise either due to inaccuracies in reconstruction or inexact camera calibration. This becomes apparent as a misalignment and blurring of texture in rendering. The surface optimisation technique provides sub-pixel accurate correspondence for view-dependent rendering.

Figure 3 shows a novel viewpoint, mid-way between two cameras in the studio, Figure 3(a),(b), demonstrating view-interpolation between the pair of cameras. The visual-hull, Figure 3(c) shows the blurring effect with incorrect geometry. The merged stereo, Figure 3(d), shows improved resolution but suffers from missing and incorrect sections of geometry. Figure 3(e) shows the optimised surface and demonstrates the highest resolution with the recovered sub-pixel correspondence.

Figure 4 shows a sequence of rendered views from a multiple view video sequence. The virtual viewpoint moves into and pans around the dynamic scene. This demonstrates the flexibility in viewpoint control that is given by the 3D description of the scene. The virtual-views approach the resolution of the original camera images and the dynamic appearance of the clothing wrinkles produces a video-realistic result. Movie sequences can be viewed at [1].



(a) Skeleton (b) Model (c) Rendered view

Figure 5: *Model based reconstruction and rendering.*

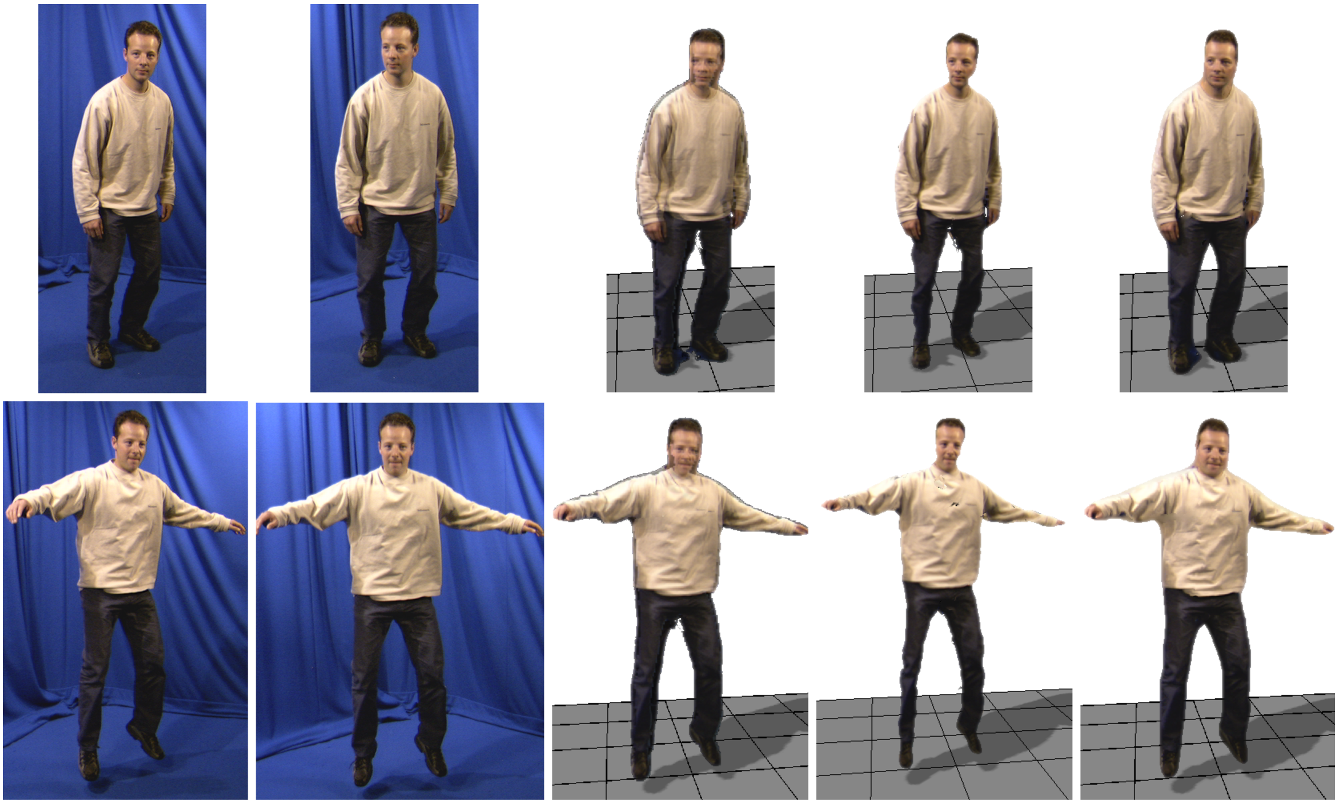
5.3. Model-based reconstruction

So far we have followed the object-centred approach to scene reconstruction, using the visual-hull as a robust initial estimate of the scene geometry and optimising the geometry to match both silhouette and stereo data. This framework can also be applied to update a prior model of the scene geometry. Figure 5 shows a humanoid computer graphics model that has been posed and the surface optimised to match an actor in a studio. Previous research [23] has introduced a manual technique to pose a model and a shape regularisation energy term to preserve prior model shape.

There are two advantages to the model-based approach. Firstly a model can provide prior shape information to constrain reconstruction in the presence of visual ambiguities such as self occlusion or lack of image texture. Secondly a model provides a consistent structure to capture a dynamic sequence. This structure can be instrumented for animation as shown in Figure 5(a) so that the model can be controlled to synthesise new content. It also fits in with current computer graphics production methods and opens up the possibility for the production of video-realistic computer graphics models. The current drawback of this approach to scene reconstruction is the requirement for the manual posing of a model to match an actor in multiple camera views. This would become an impossible task for a lengthy sequence as show in Figure 4.

6. Summary and Conclusions

In this paper we have presented a technique for mesh-based multiple view stereo. Estimated surface geometry is updated to match available stereo and silhouette data as a deformable mesh model. Optimisation of the mesh is performed in a coarse-to-fine framework in which the search range for stereo matches is gradually reduced to the calibration accuracy of the camera system, enabling convergence in the presence of noisy stereo data. Results demonstrate improved reconstruction compared to shape from silhouette and comparable reconstruction to shape from stereo correspondence. Improved reconstruction is obtained by combining silhouette and stereo data in the presence of visual ambiguities such as lack of image texture or occlusion boundaries. The technique also demonstrates improved resolution in rendering virtual views through the derivation of sub-pixel image correspondence even with inexact camera calibration. This technique for reconstruction enables the synthesis of virtual views of a person moving in a multiple camera studio. This can be applied for an object-centred approach to reconstruction allowing for arbitrary dynamic content in the scene or a model-based approach for the production of a video-realistic computer graphics model of a person.



(a) Camera image (b) Camera image (c)Visual-hull (d) Merged stereo (e) Surface optimisation

Figure 3: *Rendering a virtual view mid-way between two cameras.*

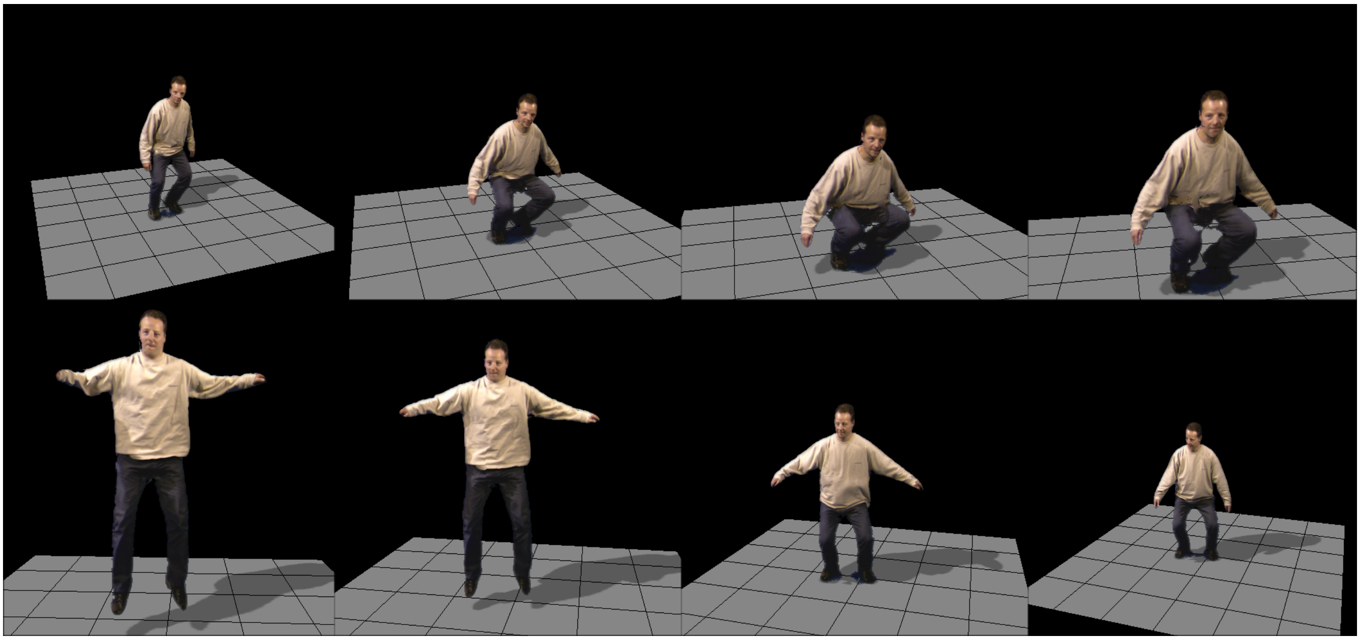


Figure 4: *Sequences of virtual views for multiple frames showing a view that moves into and pans around a dynamic scene.*

References

1. http://www.ee.surrey.ac.uk/research/vssp/vmrg/vcphuman.html.
2. www.vision.caltech.edu/bouguetj/calib-doc. Technical report, MRL-INTEL, 2003.
3. B. Allen, B. Curless, and z. Popović. Articulated body deformation from range scan data. pages 612–619, 2002.
4. P. Debevec, C. Taylor, and J. Malik. Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach. *SIGGRAPH Conference Proceedings*, pages 11–20, 1996.
5. P. Debevec, Y. Yu, and G. Borshukov. Efficient view-dependent image-based rendering with projective texture-mapping. *9th Eurographics Rendering Workshop*, 1998.
6. P. Fua and Y. Leclerc. Object-centred surface reconstruction: Combining multi-image stereo and shading. *International Journal of Computer Vision*, 16:35–56, 1995.
7. A. Fusiello, E. Trucco, and A. Verri. Rectification with unconstrained stereo geometry. *8th British Machine Vision Conference*, pages 400–409, 1997.
8. O. Grau. G. Thomas. 3d image sequence acquisition for tv & film production. *1st International Symposium on 3D Data Processing Visualization and Transmission*, pages 320–326, June 2002.
9. A. Hilton, D. Beresford, T. Gentils, R. Smith, W. Sun, and J. Illingworth. Whole-body modelling of people from multiview images to populate virtual worlds. *The Visual Computer*, 16(7):411–436, 2000.
10. T. Kanade, P.W. Rander, and P.J. Narayanan. Virtualized reality: Constructing virtual worlds from real scenes. *IEEE Multimedia*, 4(1):34–47, 1997.
11. L. Kobbelt, S. Campagna, J. Vorsatz, and H.P. Seidel. Interactive multi-resolution modeling on arbitrary meshes. *SIGGRAPH Conference Proceedings*, pages 105–114, August 1998.
12. K. Kutulakos and S. Seitz. A theory of shape by space carving. *International Journal of Computer Vision*, 38(3):199–218, July 2000.
13. A. Laurentini. The visual hull concept for silhouette based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2):150–162, 1994.
14. M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, and D. Fulk. The digital michelangelo project. *SIGGRAPH Conference Proceedings*, pages 131–144, 2000.
15. W.E. Lorensen and H.E. Cline. Marching cubes: a high resolution 3d surface reconstruction algorithm. *SIGGRAPH Conference Proceedings*, 21(4):163–169, 1987.
16. W. Matusik, C. Buehler, and L. McMillan. Polyhedral visual hulls for real-time rendering. *Eurographics Workshop on Rendering*, pages 115–125, 2001.
17. T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: a survey. *Medical Image Analysis*, 1(2):91–108, 1996.
18. S. Moezzi, L.C. Tai, and P. Gerard. Virtual view generation for 3d digital video. *IEEE Multimedia*, 4(1):18–25, 1997.
19. P.J. Narayanan, P. Rander, and T. Kanade. Constructing virtual worlds using dense stereo. *6th IEEE International Conference on Computer Vision*, pages 3–10, 1998.
20. R Plankers and P. Fua. Articulated soft objects for video-based body modeling. *IEEE International Conference on Computer Vision*, pages 394–401, 2001.
21. K. Pulli, M. Cohen, T. Duchamp, H. Hoppe, L. Shapiro, and W. Stuetzle. View-based rendering: Visualizing real objects from scanned range and color data. *8th Eurographics workshop on Rendering*, pages 23–34, 1997.
22. C.M. Seitz and C.R. Dyer. Photorealistic scene reconstruction by voxel coloring. *International Journal of Computer Vision*, 35(2):1–23, 1999.
23. J. Starck, G. Collins, R. Smith, A. Hilton, and J. Illingworth. Animated statues. *Machine Vision and Applications, Special Issue on Human Modeling, Analysis, and Synthesis*, 2002.
24. C. Sun. Fast stereo matching using rectangular subregioning and 3d maximum-surface techniques. *International Journal of Computer Vision*, 47(1/2/3):99–117, 2002.
25. S. Vedula, S. Baker, and T. Kanade. Spatio-temporal view interpolation. *Eurographics Workshop on Rendering*, pages 1–11, 2002.
26. S. Vedula, P. Rander, H. Saito, and T. Kanade. Modeling, combining, and rendering dynamic real-world events from image sequences. *Proceedings of Virtual Systems and Multimedia*, pages 323–344, 1998.