Model-based human shape reconstruction from multiple views

Jonathan Starck *, Adrian Hilton

Centre for Vision, Speech and Signal Processing, University of Surrey, UK

Abstract

Image-based modelling allows the reconstruction of highly realistic digital models from real-world objects. This paper presents a model-based approach to recover animated models of people from multiple view video images. Two contributions are made, a multiple resolution model-based framework is introduced that combines multiple visual cues in reconstruction. Second, a novel mesh parameterisation is presented to preserve the vertex parameterisation in the model for animation. A prior humanoid surface model is first decomposed into multiple levels of detail and represented as a hierarchical deformable model for image fitting. A novel mesh parameterisation is presented that allows propagation of deformation in the model hierarchy and regularisation of surface deformation to preserve vertex parameterisation and animation structure. The hierarchical model is then used to fuse multiple shape cues from silhouette, stereo and sparse feature data in a coarse-to-fine strategy to recover a model that reproduces the appearance in the images. The framework is compared to physics-based deformable surface fitting at a single resolution, demonstrating an improved reconstruction accuracy against ground-truth data with a reduced model distortion. Results demonstrate realistic modelling of real people with accurate shape and appearance while preserving model structure for use in animation.
© 2007 Elsevier Inc. All rights reserved.

Keywords: Image-based modelling; Deformable model; Multiple resolution mesh; Shape-from-silhouette; Stereo vision; Character animation

1. Introduction

In this paper, we address model-based reconstruction of people using a generic model of human shape and kinematic structure. A novel multiple resolution model-fitting technique is presented that combines multiple shape cues from camera images for coarse-to-fine model-based reconstruction with preservation of animation structure. Previous work concentrates on reconstructing dynamic shape to render the appearance of people from a novel viewpoint, termed *free-viewpoint video* [1,2]. These techniques have no prior scene structure to construct a consistent representation that can be instrumented for animation and synthesis of new content. *Model-based* scene recon-

struction is a well-established approach [3–5] in which a prior scene model is adapted to fit available shape data and so provides a consistent structure that can be used for synthesis. Previous model-fitting techniques have constrained the freedom in model deformation for robust reconstruction by restricting the space of feasible shape using a coarse input model or by restricting models to closed surfaces. Techniques also allow the vertex structure used in animation to be distorted in model deformation. In this work, we introduce constraints to explicitly preserve the mesh parameterisation in model deformation and present a hierarchical model that decomposes the degrees of freedom in model fitting for any prior high-resolution surface mesh with no restriction on the surface triangulation or topology. A structured representation is derived in reconstruction that supports control of dynamics for the synthesis of new animated content as illustrated in Fig. 1 as well as manipulation of viewpoint for free-viewpoint rendering.

* Corresponding author. Fax: +44 1483 686031.
 E-mail address: j.starck@surrey.ac.uk (J. Starck).

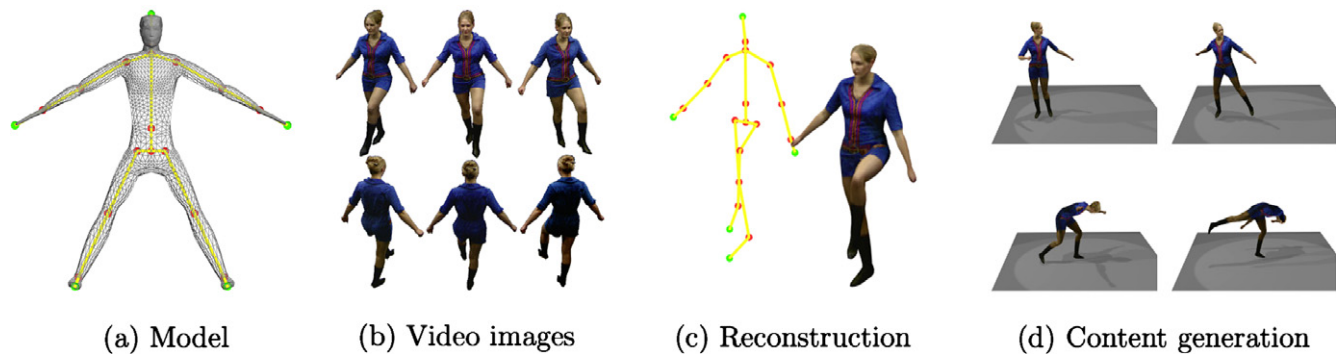(a) Model   (b) Video images   (c) Reconstruction   (d) Content generation

Fig. 1. A model-based approach to multiple view reconstruction matches a prior surface model to multiple view images providing a structured scene representation that allows novel content generation from the images.

The paper is organised as follows. In Section 2, we review related work. In Section 3, a multiple resolution deformable model is presented for hierarchical data fitting. In Section 4, a framework is presented for coarse-to-fine model fitting to multiple shape cues for image-based reconstruction. Multiple resolution coarse-to-fine model fitting is then evaluated in Section 5 and applied to human shape reconstruction from multiple views. The application of model-based reconstruction for human animation synthesis is finally demonstrated in Section 6 and the contributions made in this work are summarised in Section 7.

2. Related work

Over the past decade there has been extensive research in computer vision and graphics focused on modelling the appearance and movement of real people. Approaches to modelling include 3D scanning [6], reconstruction from static images [7,8], multiple view images [4,9] and reconstruction from multiple view video [1,10,11]. Kanade et al. [1] and Moezzi et al. [10] introduced reconstruction of sequences of people from multiple view video. Stereo reconstruction [1,12] and shape from silhouette [10,13] have been used to reconstruct 3D video sequences captured from multiple camera systems and video-rate reconstruction has been achieved using an image-based scene representation [2].

Previous work concentrates on reconstructing the dynamic shape and rendering the appearance of people captured in multiple view video for free-viewpoint visualisation or "Virtualized Reality" [1]. There is no assumption on the observed scene topology or geometry enabling reconstruction of complex scenes potentially with multiple people. There are however two limitations in this approach. First, reconstruction using silhouette or stereo data alone suffers from visual artefacts. Shape from silhouette requires a large number of views to accurately reconstruct a scene and does not necessarily reconstruct concavities. Stereo reconstruction fails in regions of uniform appearance and close to silhouette edges. Second, general surface reconstruction does not provide a scene structure suitable for surface animation to synthesise new dynamic content.

2.1. Object-centred scene reconstruction

An object-centred approach to scene reconstruction was introduced by Fua and Leclerc [14,15] to combine multiple visual cues in reconstruction using an initial surface mesh for a scene and then optimising surface shape to match both stereo and shading cues across multiple views. The work demonstrates a refinement of the initial surface subject to the restriction that the model is initially close to the solution for local optimisation. This surface-based approach allows inference of occlusions and discontinuities in camera images and the combination of different complementary visual cues in reconstruction. Object-centred reconstruction has been used to improve reconstructed scene models of people in *model-enhanced stereo* [13] where an initial surface reconstruction is used to restrict the search range to refine stereo correspondence. More recent techniques have been presented to refine a robust initial estimate of shape derived from silhouettes by incorporating stereo and photo-consistency from multiple views [11,16,17]. These techniques combine shape cues for robust reconstruction without the requirement for a prior surface model but provide no structure for animation synthesis.

2.2. Functional modelling

Terzopoulos [18] introduced the concept of 'functional modelling' in which a prior model is used to recover shape. A model-based approach has a number of important advantages: (i) the prior model can be designed for rendering and manipulation as a standard computer graphics model; (ii) the model can be instrumented with a kinematic structure for animation; (iii) a model provides prior shape information to regularise multiple view reconstruction; and (iv) reconstruction then provides a consistent structure for all surfaces. This approach has been used widely to derive animated models of the human face, see, for example, [5,19]. Whole-body human shape reconstruction has been presented previously to recover either static shape and articulated motion [4,20] or dynamic shape [21] from image silhouettes and from multiple shape cues [9].

Optimisation of a prior model to match image data was first introduced in computer vision and computer graphics by Terzopoulos et al. [3] who proposed a class of models that describe the shape and motion of physically based deformable surfaces. A significant body of literature exists on shape reconstruction from 2D and 3D data using deformable models. Montagnat et al. [22] review different representations, McInerney and Terzopoulos [23] review applications in medical image analysis and Nealen et al. [24] review techniques for simulation of deformable objects in computer graphics. Deformable contour models, known as *snakes* [25], have been widely used for 2D image segmentation. Deformable 3D surface models have been applied for segmentation and shape recovery from 3D range data [26] and 3D medical imaging data sets [27]. A variety of shape representations have been explored to achieve this, including surface meshes [28], particle systems [29], superquadrics [30] and implicit representations [31]. Surface evolution has been realised through discretisation of the deformation energy functions using finite differences [28] and finite elements [32] or through level-set evolution of an implicit surface [33].

Surface models used in animation have a highly structured mesh to give high-resolution representation in areas of deformation and efficient representation in other areas. Preserving this vertex parameterisation is important in reconstructing models that can be used for animation. Previous formulations for deformable triangulated surfaces make no explicit constraint on the vertex parameterisation for model animation [22]. Mesh deformation has been constrained in model fitting by parameterising shape deformation [34,4], using global transformations [27,35], free-form deformations (FFDs) [36] or by multiple resolution fitting using subdivision of a base model [37,38]. These techniques either restrict the feasible space of model deformations or require a coarse prior model. Parameterised multiple resolution models have also been constructed from high-resolution input models to approximate the original surface for geometric processing [39]. Delingette [40] proposes the use of a 2-simplex mesh that allows exact constraints to be formulated to parameterise vertex positions. A 2-simplex mesh is constructed as the dual of the polyhedral graph defined by the triangulated surface, however the constraint can only be defined for a closed surface and provides no inverse operation to recover the original triangulation exactly.

2.3. Contribution

In this paper, an image-based reconstruction framework is presented that uses a prior surface model to provide a domain to combine multiple shape cues for robust reconstruction and to provide a model structure instrumented for animation. A functional modelling approach is adopted in which a prior triangulated surface is formulated as an elastically deformable model, solving for the shape deformation in data fitting by discretisation of the deformation energy functional. The novelty in this framework is first the introduction of a multiple resolution deformable model that explicitly preserves the vertex parameterisation in the triangulation without restriction on the shape or topology of the prior model. A coarse-to-fine model optimisation scheme is then presented to combine multiple cues from silhouette, stereo and feature data for multiple view human shape reconstruction. Generation of human models from camera images is evaluated and the application for model animation is demonstrated.

3. Multiple resolution model

The goal of model-based reconstruction is to take a prior triangulated surface, designed for animation and rendering in a standard computer graphics pipeline, and adapt the model to recover shape and image correspondence from multiple camera views while preserving the vertex parameterisation and animation structure of the model. In this section, a technique is presented to construct a multiple resolution deformable model representation for a prior high-resolution mesh. In Section 3.1, mesh simplification is used to construct a hierarchy of meshes M_l from an input mesh M. A novel mesh parameterisation technique is introduced in Section 3.2 to enable propagation of shape deformation between levels in this mesh hierarchy. A hierarchical deformable model is then proposed in 3.3 for multiple resolution data fitting that preserves the vertex parameterisation in the model and hence parameterisation-dependent model properties such as animation weights.

3.1. Hierarchical mesh representation

Mesh simplification is used to construct a hierarchical mesh representation. A triangulated surface M is defined by a pair (V, K) where V is a finite set of points, the vertex positions of M, and K a *simplicial complex* representing the vertex connectivity defining the topology of M. For a triangulated surface in $\mathbb{R}^3$ the simplicial complex is defined by three simplices, the vertices $\{i\} \in K$, edge connections $\{i,j\} \in K$, and face connections $\{i,j,k\} \in K$ in the mesh. The goal of simplification is to construct a hierarchical representation for M consisting of a set of progressively simplified meshes $M_l = (V, K_l)$ each using a subset of the original vertex positions V defining the geometry of M.

Simplification is performed by iterative edge contraction according to the process proposed by Hoppe et al. [41,42]. Each contraction operation deletes a single vertex and a single edge from the complex K. Subset placement is used to retain a single geometric representation V at all levels of detail. To delete a vertex v_1 and edge (v_1, v_2), all edge and face connections in the 1-neighbourhood of v_1 are updated as $v_1 \rightarrow v_2$. All degenerate edges and faces in the 1-neighbourhood are then deleted.

Mesh simplification produces a new representation for each edge contraction operation. A hierarchy of meshes

M_l, also termed a Dobkin–Kirkpatrick (DK) hierarchy [43], is constructed by successively removing a distinct set of vertices in the mesh. For each edge contraction the vertices in the 1-neighbourhood of the deleted vertex are marked as fixed and only non-fixed vertices are scheduled for removal. Once all non-adjacent distinct vertices are deleted, the resulting complex can be recorded as a level in the mesh hierarchy K_l. This process is repeated generating a hierarchy of complexes until either no more legal operations can be performed or the mesh reaches a predefined error tolerance with respect to the original surface. In this work, surface reconstruction is addressed using both an open mesh for the human face and a closed mesh for the whole body, the hierarchical representation constructed for these models is illustrated in Fig. 2.

3.2. Surface parameterisation in $\mathbb{R}^3$

The mesh hierarchy M_l provides a multiple resolution representation for a triangulated surface M. In coarse-to-fine model-fitting shape optimisation is scheduled at progressively finer resolutions. Optimisation of the vertex subset $i \in K_l$ for a coarse mesh will however leave the remaining vertices in the higher resolution representations unaffected. In this section, a surface parameterisation is introduced for an irregular triangulated mesh that enables local reconstruction of position from the vertex neighbourhood and propagation of model deformation from coarse-to-fine resolutions.

For an irregular triangular mesh, local parameterisation of vertex position is inherently ambiguous. A vertex can have two or more vertices in the 1-neighbourhood and it is not possible to consistently define a local reference frame in terms of the neighbourhood to express the vertex position. For a closed surface, every face has a constant 3-connected face neighbourhood as illustrated in Fig. 3. For a
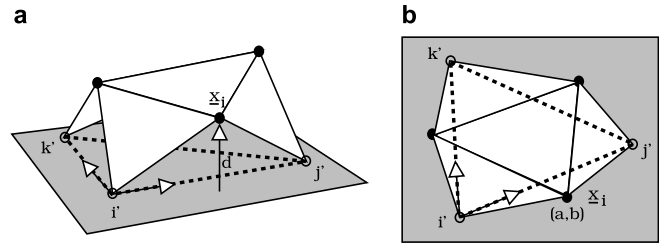

Fig. 3. Vertex parameterisation on an irregular triangular mesh. A local reference frame is constructed from the three vertices i', j', k' forming the 1-connected face neighbourhood for a triangle. Vertex position is parameterised as (a) the displacement along the frame normal and (b) barycentric coordinates in the triangle i', j', k'.

vertex $\underline{x}_i$ on a face, a consistent reference frame can be defined using the three vertices i', j', k' forming the 1-connected face neighbourhood. A translation and rotation invariant local parameterisation is defined for a vertex using a reference frame defined with respect to each face that contains the vertex. Vertex position is parameterised using barycentric coordinates (a, b) in the triangle formed by i', j', k' plus a displacement d along the normal $\hat{\underline{n}}$. With more than one reference frame a redundant parameterisation is obtained and vertex position can be averaged across frames, corresponding to a least squares fit to each of the corresponding parameterised positions

$$\hat{\underline{n}} = \frac{(\underline{x}_{j'} - \underline{x}_{i'}) \times (\underline{x}_{k'} - \underline{x}_{i'})}{\|\underline{x}_{j'} - \underline{x}_{i'}\|\|\underline{x}_{k'} - \underline{x}_{i'}\|} \tag{1}$$

$$\underline{x}_i(i', j', k') = (1 - a - b)\underline{x}_{j'} + a\underline{x}_{j'} + b\underline{x}_{k'} + d\hat{\underline{n}} \tag{2}$$

For an open mesh the face connectivity is not constant and a face can have one, two or three connected faces in the 1-neighbourhood. Boundary frames can be handled as a special case as shown in Fig. 4. With only one or two faces in the 1-neighbourhood a reference frame can still be constructed making use of the vertices on the face itself as well
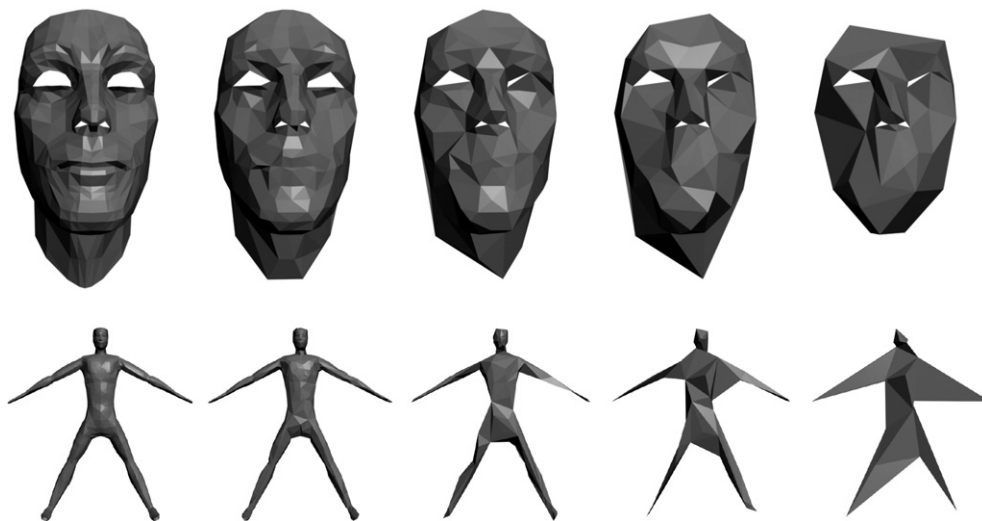

Fig. 2. Hierarchical mesh representation constructed for a generic head and whole-body model used for image-based reconstruction. The models are open and closed with irregular vertex connectivity.
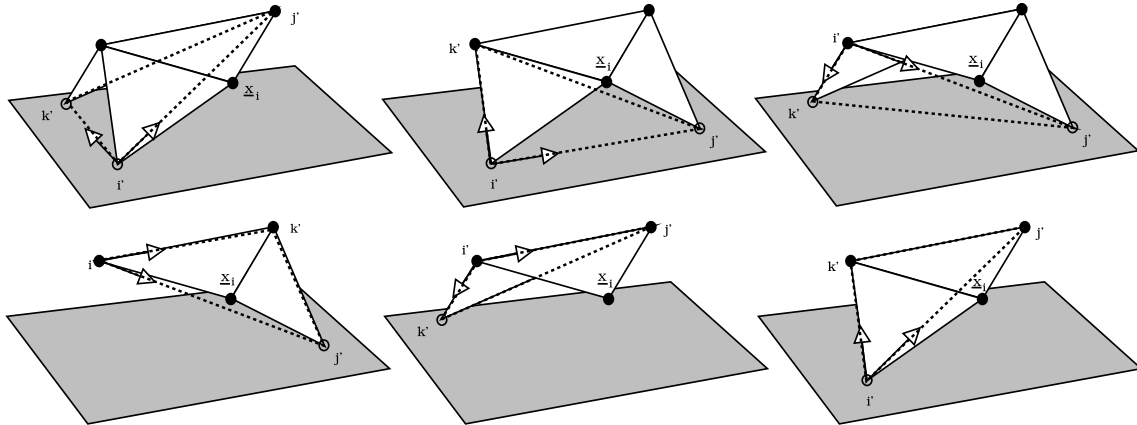
Fig. 4. Construction of a reference frame i', j', k' for a vertex i on a face in an open mesh with only (top) two or (bottom) one face in the 1-neighbourhood. Missing vertices in the face neighbourhood are replaced with vertices of the central triangle to provide a complete reference frame i', j', k'.

as those forming the face neighbourhood to form the local reference frame i', j', k'.

Vertex parameterisation is constructed as part of the mesh simplification process in creation of the hierarchical model representation. At each vertex deletion the vertex position is parameterised with respect to the local reference frames. The parameterisation is thus defined in terms of the remaining set of vertices in the mesh. In the final hierarchical representation the vertex deletion order can be traversed in reverse to progressively reconstruct vertex positions, ensuring that only those vertices in a lower level of detail are used in reconstruction.

3.3. Deformable model formulation

A hierarchical deformable model technique is now presented for multiple resolution model fitting. Each mesh in the hierarchy M_l is formulated as a discretized elastically deformable surface model [3] that preserves the mesh parameterisation, the relative position of the surface vertices as defined in Section 3.2, to ensure that the animation structure for the model remains valid. The energy function for surface optimisation provides a trade-off in data fitting using a data term E_{data} and regularisation of model distortion with respect to the original surface configuration using a shape term E_{shape}

$$E_{\text{model}} = (1 - k)E_{\text{data}} + kE_{\text{shape}} \qquad (3)$$

The energy functionals are defined by a set of local constraints C with the cost for each constraint defined using an error function $E(C)$. This formulation can be naturally extended to include robust error kernels in data fitting. In this work, we use the $L^2 - norm$, providing a least squares minimisation of the constraints

$$E(\underline{C}) = \frac{1}{2}\underline{C}^{\text{T}}\underline{C} \qquad (4)$$

Surface deformation is regularised using a shape constraint defined in each local reference frame for a vertex. Physically based deformable model techniques traditionally

make use of an elastic membrane and a thin-plate energy term in optimisation to regularise both surface stretch and bending deformations in a model. Minimisation of the membrane energy leads to a minimal area solution and minimisation of the thin-plate energy leads to a minimum curvature solution in data fitting. Here, the initial parameterization of the model is preserved explicitly as well as the surface shape. As illustrated in Fig. 5, constraining shape alone is no guarantee that the mesh parameterisation will be preserved.

The shape cost is designed to provide an elastic constraint during optimisation that restores the relative positions of the vertices i, i', j', k' within each local frame. The distortion in the vertex parameterisation is constrained according to the deviation $\underline{C}_{\text{p}}$ in the parameterised vertex position $x_i(i', j', k')$ in the frame and the scale of the frame is constrained by minimising the deviation C_{e} in the edge length for the frame $\{i', j', k'\}$. Here $\underline{x}^0$ denotes the original location for a vertex

$$C_{\text{e}}(i, j) = (\underline{x}_i - \underline{x}_j)^{\text{T}}(\underline{x}_i - \underline{x}_j) - (\underline{x}_i^0 - \underline{x}_j^0)^{\text{T}}(\underline{x}_i^0 - \underline{x}_j^0) \quad (5)$$

$$\underline{C}_{\text{p}}(i, i', j', k') = (\underline{x}_i - \underline{x}_i(i', j', k')) \qquad (6)$$

$$E_{\text{shape}} = \frac{1}{N_i}\sum_{i \in K_l} \frac{1}{N_{i', j', k'}}\sum_{i', j', k' \in P_{i,l}} E(C_{\text{e}}(i', j'))$$
$$+ E(C_{\text{e}}(j', k')) + E(C_{\text{e}}(i', k'))$$
$$+ E(\underline{C}_{\text{p}}(i, i', j', k')) \qquad (7)$$

A multiple resolution shape constraint is obtained by constructing a separate vertex parameterisation $\{i', j', k'\}$ for the vertices $i \in K_l$ at each level of detail in the mesh hierarchy M_l. Hierarchical deformable model optimisation pro-
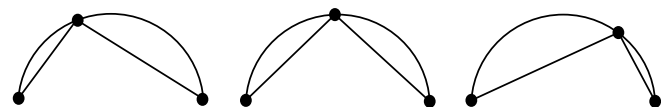

Fig. 5. Constant surface curvature can be obtained with different vertex parameterisations. Constrained model fitting preserves both parameterisation and curvature.

vides a multiple resolution constraint on shape distortion in data fitting.

4. Model-based shape reconstruction

In this section, a technique is presented for multiple resolution fitting of a prior mesh to multiple camera views. A hierarchical representation is first constructed for a prior model mesh M that decomposes the degrees of freedom (DOF) in the model into multiple levels of detail M_l for shape reconstruction. Shape fitting is then formulated as a coarse-to-fine hierarchical energy minimisation process in which the model shape is deformed to fit shape data in the camera images at progressively finer resolutions. The output is a deformed model that satisfies the multiple view appearance with a minimal distortion in the vertex parameterisation. In Sections 4.1–4.4, a data energy term is presented for coarse-to-fine model fitting to multiple shape cues. Finally in Section 4.5, the complete model optimisation framework is presented for multiple resolution model-based shape reconstruction.

4.1. Sparse data constraints

Feature constraints are required to specify exact correspondence in shape reconstruction, for example, where the animation structure in the model corresponds to a specific feature across multiple views. However, in surface optimisation a sparse set of constraints provide only a localised deformation of model shape. Previous approaches have therefore interpolated the influence of constraints either within the space containing the model [32] or across the model surface [9] and by relying on shape regularisation in optimisation to interpolate the data [44].

The hierarchical model representation is used to propagate the influence of sparse features in data fitting and to automatically define the region of influence on the surface of the model. A sparse set of exact constraints $\{\underline{d}_f\}_{f=1}^{N_f}$, $N_f \ll N_i$, are defined in $\mathbb{R}^3$, together with model correspondence $i(f)$. In construction of the model hierarchy the con-

strained vertices are retained at all levels. The exact constraints are therefore applied at multiple resolutions starting at the coarsest model with the influence interpolated across the surface between constraints while preserving the model surface configuration. The effect of satisfying a sparse set of feature constraints is demonstrated in Fig. 6 for a hierarchical deformable model in comparison with a single full-resolution deformable model.

4.2. Stereo data

Dense multiple view stereo is used to derive the surface shape that matches appearance across camera views. A model-based approach to correspondence estimation is introduced that performs stereo matching between views in a coarse-to-fine strategy in optimisation of the hierarchical model. The model provides an initial estimate of correspondence and visibility across multiple views and a direct search for correspondence is performed at multiple resolutions, providing a wider range of convergence compared to local optimisation techniques [14,15].

Stereo matching is performed with respect to a *key* camera with the closest viewpoint such that the surface can be assumed to be locally fronto-parallel. Matches are then located by a direct search in each visible adjacent camera, termed an *offset* camera, that forms a stereo pair with the *key* view. The search space is constrained by the predefined search range for the model along the epipolar in rectified images and according to the camera calibration error perpendicular to an epipolar line as illustrated in Fig. 7. The zero-mean normalised cross correlation (ZNCC) is used allowing for a linear variation in local intensity between views [45]. For each model vertex $\underline{x}_i$ a corresponding set of stereo data points are derived giving a complete stereo data set $\{\underline{d}_s\}_{s=1}^{N_s}$ and model correspondence $i(s)$.

4.3. Silhouette data

Stereo correspondence fails where there is insufficient appearance variation within an image for reliable matching and at depth discontinuities where the surface is perpendic-

(a) constraint　　　　(b) Single resolution　　　　(c) Multiple resolution
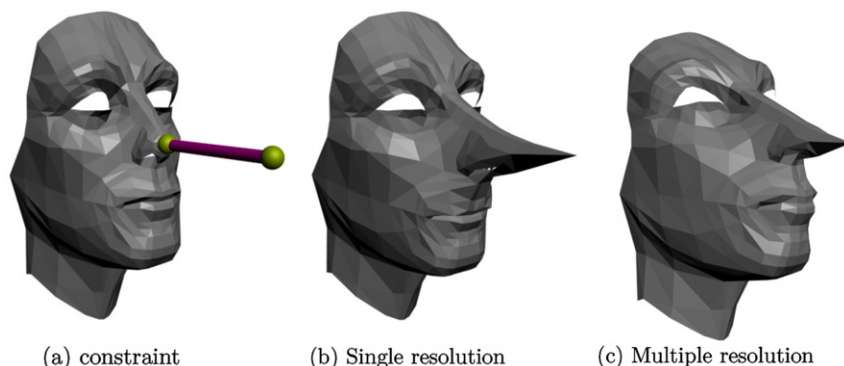
Fig. 6. Interpolating the influence of (a) a sparse feature constraint using (b) a deformable model at a single resolution and (c) a multiple resolution deformable model. The influence of the constraint is propagated across the surface in the multiple resolution model.
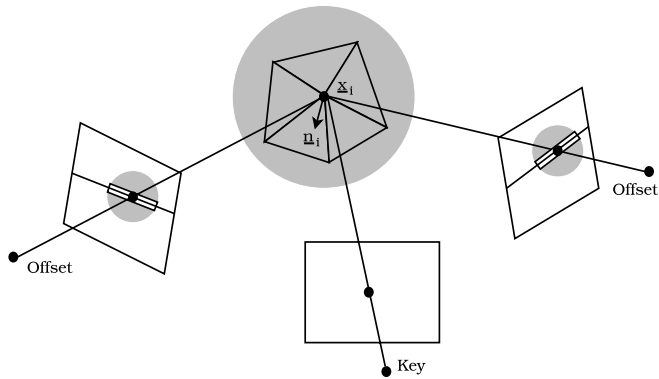
Fig. 7. Direct search for stereo correspondence in *offset* camera images with respect to a *key* camera view. The closest camera view is selected for the key view and a fronto-parallel surface is assumed for matching in rectified stereo camera pairs. The search range is defined by the matching scale along and the reprojection error perpendicular to the epipolar lines.

ular to the viewing plane. Shape from silhouette provides a bounding constraint on the surface shape that provides a complementary visual cue where stereo is ambiguous or fails. Previous model-based techniques make use of 2D image constraints for 3D reconstruction from image silhouettes [7,4]. This is based on the assumption that the projected shape of the model matches the silhouette contour. However for a general body pose, even small changes in model articulation can lead to large changes in the apparent 2D contour and an explicit 2D constraint is unreliable.

Shape from silhouette is satisfied by matching the surface of the *visual-hull*, the maximal volume that reproduces the 2D image silhouettes [46]. The visual-hull combines silhouette constraints across views and closest point assignment in $\mathbb{R}^3$ enforces a single silhouette constraint for each model vertex. Iterative closest point assignment in optimisation allows the correspondence to vary across the 2D contour and to switch between views removing the requirement for an explicit 2D constraint.

A volumetric reconstruction of the visual-hull is performed to derive the discrete set of voxels $\{\underline{d}_v\}_{v=1}^{N_v}$ on the visual-hull surface. Each model vertex, $\underline{x}_i$, is matched to the closest point on the visual-hull providing the correspondence $v(i)$ as follows:

$$v(i) = \operatorname{argmin}_{v \in \{1...N_v\}} \|\underline{x}_i - \underline{d}_v\|^2 \qquad (8)$$

4.4. Combining shape cues

Data constraints are defined in terms of three visual cues: shape from silhouette, stereo correspondence and sparse feature constraints. Silhouettes provide a robust constraint on shape, however the shape information available is limited giving only an approximate shape and image correspondence. Stereo matching allows recovery of more accurate geometry and image correspondence given sufficient variation in local appearance for reliable stereo

matching. Finally feature constraints allow exact constraints for user-defined image correspondence.

Shape cues are combined using a weight w_i^f, w_i^s, w_i^v that quantifies the confidence in fitting each data term at a vertex i. Matching is prioritised to satisfy exact constraints, then stereo data where matching is reliable and finally shape from silhouette. A binary weight is defined at a vertex $w_i^f = (0,1)$ enforcing feature matching. A soft weight is defined for stereo matching $0 < w_i^s < 1$. A constant silhouette term $w_i^v = 1$ is used to incorporate shape from silhouette where $w_i^s \to 0$. The data term for model fitting is then defined as follows:

$$\underline{C}_{\text{data}}(i) = w_i^f(\underline{x}_i - \underline{d}_{f(i)}) + (1 - w_i^f)w_i^s(\underline{x}_i - \underline{d}_{s(i)}) \\ + (1 - w_i^f)(1 - w_i^s)w_i^v(\underline{x}_i - \underline{d}_{v(i)}) \qquad (9)$$

$$E_{\text{data}} = \frac{1}{N_i} \sum_{i \in K_l} E(\underline{C}_{\text{data}}(i)) \qquad (10)$$

The confidence metric $0 < w_i^s < 1$ for stereo matching at a vertex *i* is derived using image entropy to quantify the ability to derive stereo correspondence. Different confidence metrics have been proposed to define the performance of stereo matching algorithms [47], including the matching score itself, the ratio between peak scores, the curvature of the quadratic fit in subpixel localisation, and the entropy of the image areas used in matching. With the ZNCC matching score, the variation in intensity in each image area is normalised and so both a high matching score and a high curvature in localisation can be obtained when matching regions of low intensity variation where correspondence is inherently ambiguous. Furthermore in coarse-to-fine matching the search region is successively reduced and peak scores cannot be compared within a small matching region. Instead image variation is used as a measure of image entropy to quantify regions of high intensity variation where good correspondence is likely to be obtained in stereo matching.

Image variation is precomputed in each camera view and a simple lookup operation is performed to define w_i^s in the *key* camera used for a vertex in stereo matching. Intensity variation is quantified using the standard deviation σ in pixel intensity for the image area used in stereo correlation and normalised in the range $0 < w_i^s < 1$ using a control parameter σ_0. This user-defined control parameter can be changed to restrict matching either to prominent features $\sigma_0 \to 127$ or all areas $\sigma_0 \to 0$ for an 8-bit image. Throughout this work a fixed value of $\sigma_0 = 10$ is adopted ensuring that stereo matching is restricted to regions where the intensity variation exceeds the maximum expected noise level for the camera system. Fig. 8 illustrates the confidence metric for a whole-body image of a person for different parameter settings

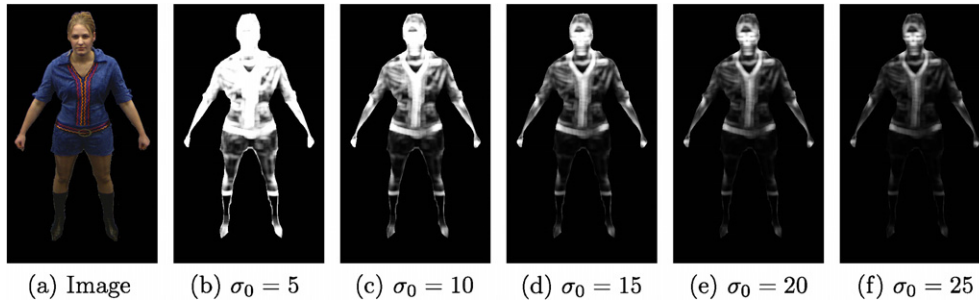$$w^s = 1 - \exp\left(-\frac{\sigma^2}{\sigma_0^2}\right) \qquad (11)$$

|(a) Image | (b) $\sigma_0 = 5$ | (c) $\sigma_0 = 10$ | (d) $\sigma_0 = 15$ | (e) $\sigma_0 = 20$ | (f) $\sigma_0 = 25$|

Fig. 8. The confidence metric w^s precomputed at all pixels in an image using a 13×13 image window. Brighter areas correspond to a high confidence value for stereo matching.

4.5. Hierarchical model optimisation

The complete framework is now described for image-based reconstruction from multiple camera views using a prior triangulated surface mesh. The input is a mesh M together with a set of calibrated camera images and optionally a set of exact feature constraints $\{\underline{d}_f\}_{f=1}^{N_f}$ with a predefined model correspondence $f(i)$. The set of surface voxels for silhouette data fitting, $\{\underline{d}_v\}_{v=1}^{N_v}$ is derived and an octree structure is constructed for fast closest point lookup. The set of stereo confidence images is calculated to lookup the stereo confidence score w_i^s in model optimisation. A hierarchical deformable model is then constructed from the mesh M for shape optimisation.

The reconstruction algorithm schedules model fitting first at the coarsest level M_0 in the mesh hierarchy and optimises the subset of vertices $\underline{x}_i \in K_0$. Model optimisation is guaranteed to converge monotonically to a local minima at each level as each iteration in optimisation reduces the total cost C. The vertex transformations are then propagated to the next level in the hierarchy $\underline{x}_i \in K_1$ and this level is then scheduled for optimisation. The process is repeated until the finest level of detail corresponding to the original surface mesh M is reached. At each level in the hierarchy M_l an iterative, alternating strategy of correspondence estimation followed by model optimisation is used for data fitting. At each iteration the set of stereo data points $\{\underline{d}_s\}_{s=1}^{N_s}$ for the model vertices and the silhouette assignment $v(i)$ are derived. The model M_l is then optimised to fit the feature, stereo and silhouette data constraints.

An annealing schedule is constructed for coarse-to-fine recovery of image correspondence in model optimisation. The search range for correspondence is specified by a surface error ϵ. The error is initialised as the worst-case error for the model and updated according to a geometric schedule $\epsilon = \lambda \times \epsilon$, $0 < \lambda < 1$ at each iteration of model optimisation. The convergence criteria for model optimisation is set where the maximum data error falls within the error tolerance at the next level of optimisation such that that the recovered correspondence remains valid at the next step of optimisation. Where a model is over-constrained by regularisation and data fitting is not possible a limit on the maximum number of iterations can be introduced. Transi-

tions are tied between levels of detail in the model hierarchy using a linear schedule starting at the coarsest model resolution. Optimisation terminates where the highest resolution mesh converges within the reconstruction accuracy of the camera system. The user-defined parameter λ is set to a constant value $\lambda = 0.7$ throughout this work, providing a conservative annealing schedule at the cost of a potentially greater computational complexity. In practise this parameter could be tuned to achieve a desired time-accuracy trade-off for a target application area.

Mesh optimisation at a level M_l and minimisation of the corresponding non-linear cost function E_{model} is performed using the conjugate gradient method. The step length in conjugate gradient descent is defined by performing a line search using the back-tracking algorithm. Techniques such as Gauss–Newton and Levenberg–Marquardt (LM) can also be applied for minimisation as the cost function is expressed in terms of squared error residuals. It is interesting to note that the LM algorithm is equivalent to an implicit integration approach to dynamic deformable model evolution using a backward Euler integration step. Implicit techniques have been proposed [48,49] to increase the step length and stability in dynamic deformations. In practise we found the increase in computational cost for the LM algorithm outways the advantage of large step lengths in optimisation and with large mesh deformations the algorithm spends the majority of iterations performing gradient descent steps.

5. Results

A framework has been presented to construct a hierarchical deformable model for a prior triangulated surface and then to optimise surface shape combining cues from silhouette, stereo and sparse feature data from multiple camera images. This technique provides the following key advantages in image-based modelling: *unrestricted model*, there is no restriction on the triangulated surface used in model fitting; *multiple resolution fitting*, hierarchical fitting avoids local minima in shape optimisation; and *vertex parameterisation*, the vertex parameterisation in the prior model is preserved such that the prior animation structure for the model can be reused to synthesise animations.

5.1. Evaluating model reconstruction

Previous model-based reconstruction approaches have required a specific triangulated surface structure, requiring either a closed genus-zero surface or a predefined low-resolution mesh. The technique presented in this paper is specifically designed for use with any high-resolution triangulated surface model. Two sets of experiments were performed in evaluation, first human face reconstruction using an open triangulated surface and second whole-body shape reconstruction with a closed genus-zero triangulated surface. The hierarchical representation for the models is shown in Fig. 2.

A quantitative evaluation is presented for the range of convergence in multiple resolution model fitting and the resulting distortion in the vertex parameterisation for the recovered shape models. A maximum range of convergence is required to overcome local minima in iterative surface optimisation. A minimal distortion is required in the vertex parameterisation for a model such that the animation structure for the model remains valid. A registration process is required to first align the models with target images for reconstruction and a manual registration step is performed in this work. The reconstruction accuracy is presented against a range of registration errors in model pose to evaluate the range of convergence in model fitting and the corresponding distortion in the final surface is quantified to evaluate distortion in the vertex parameterisation.

Reconstruction accuracy is defined using the root-mean square (RMS) error across the surface from the optimised model to ground-truth data. The error is computed by uniformly sampling each triangle of the model surface and for each point taking the distance to the closest point in the target ground-truth data set. Surface distortion is quantified using the Dirichlet energy for the mapping $f : \mathbb{R}^3 \to \mathbb{R}^3$ between the original and deformed vertex configuration. A tangent plane is defined, centred at each triangle and a local linear mapping A is calculated between triangles. The mapping $A = [f_u f_v]$ is a 2D affine map corresponding to the gradient of the piecewise linear mapping f defined in the local tangent space uv for a triangle. The total distortion in the triangulation is normalised as follows such that no distortion in the vertex parameterisation results in a unit distortion term

$$\text{distortion} = \frac{1}{N_{i,j,k}} \sum_{i,j,k \in K} \frac{1}{2} \left(\underline{f}_u^2 + \underline{f}_v^2 \right) \qquad (12)$$

5.2. Comparison to a standard deformable model

A reference implementation of a single resolution physics-based deformable model is provided for comparison of performance. A conventional node-spring model is adopted as widely used in model fitting [23] and in simulation of dynamic surface deformations, see, for example,

[50]. In this model the vertices of the mesh are treated as nodes and edges in the triangulation springs with a rest-length given by the initial configuration. The surface deforms as an elastic sheet and rigidity is introduced using diagonal springs connecting opposite vertices on triangles that share a common edge.

In this paper, a novel shape regularisation term is introduced to explicitly preserve mesh parameterisation in optimisation. For a fair comparison a fixed value $k = 0.5$ is used giving equal weight to data fitting and regularisation. The regularisation term k was then optimised for the reference model to obtain an equivalent performance in minimising mesh distortion. A random target data set was constructed for 10 trials and the full-resolution model optimised first using the proposed technique with $k = 0.5$, then as the reference model for a range of k values. The mean minimum value of k was then selected such that the reference model gave the same mesh distortion in data fitting as the proposed technique. For the face model $k = 0.89$ was obtained and for the whole-body model $k = 0.78$. This result in itself demonstrates that the proposed constraint on mesh parameterisation serves to reduce model distortion such that less regularisation is required in surface optimisation.

5.3. Sparse feature matching

The influence of sparse feature constraints is demonstrated in a series of experiments where the multiple resolution model is compared to the full-resolution reference implementation. Here the open-face model is optimised to fit seven feature constraints located at the corners of the eyes, mouth and on the tip of the nose as shown in Fig. 9(a). Optimisation is performed for 20 subjects (10 male, 10 female) and for a range of rotation errors in model pose applied about a random axis of rotation. The ground-truth surface for this test set, illustrated in Fig. 9(a), was reconstructed using three stereo pairs imaging an infra-red speckle pattern projected on the face [51].

The shape deformation induced by the feature correspondences in this problem can be solved relatively simply with a rigid-body or affine transformation. The evaluation serves to demonstrate the propagation of constraints across the mesh with the hierarchical representation avoiding local minima in optimisation. We expect a full-resolution mesh to have only a localised influence from sparse constraints. Fig. 9 shows the result of model optimisation for one subject with an initial pose error of 45°. For the reference model the constraints have only a local effect whereas the multiple resolution model recovers the global pose in fitting the constraints. The reconstruction error and the distortion in the mesh parameterisation are shown in Fig. 10. The multiple resolution model increases the range of convergence in fitting the ground-truth surface and reduces the distortion in the optimised mesh. The hierarchical representation for the model serves

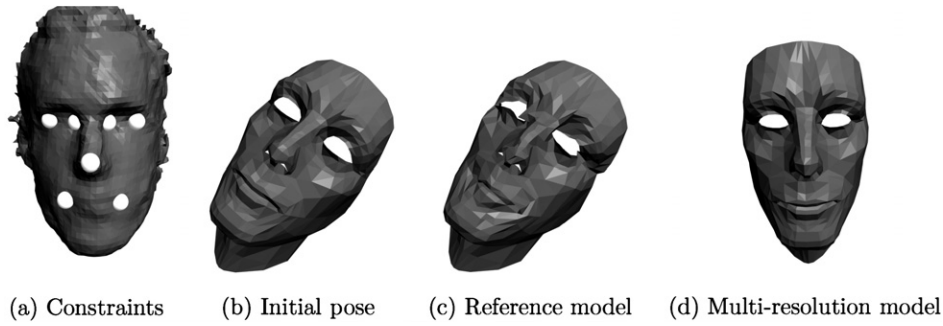(a) Constraints   (b) Initial pose   (c) Reference model   (d) Multi-resolution model

Fig. 9. Fitting sparse feature constraints for one subject with a 45° error in model pose, showing convergence to a local minima with the reference model and global fitting with the multiple resolution model.
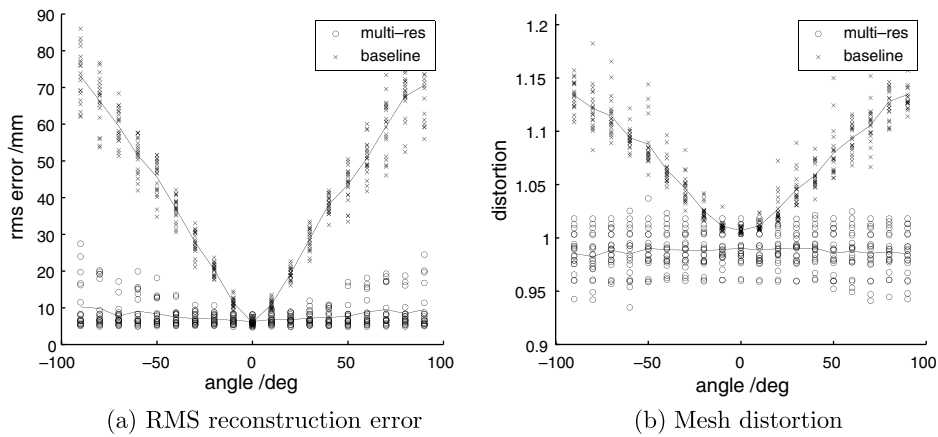

(a) RMS reconstruction error   (b) Mesh distortion

Fig. 10. (a) RMS surface error to ground-truth data and (b) mesh distortion, for the multiple resolution model compared to the reference model in fitting sparse feature points for a range of initial model rotations across 20 test subjects in face reconstruction.

to propagate the local effects of fitting sparse constraints at the lowest mesh resolution through to the full-resolution surface. The vertex parameterisation minimises the mesh distortion in optimisation and maintains the shape of the open surface mesh.

5.4. Combining sparse, stereo and silhouette shape cues

Surface optimisation combining silhouette, stereo and feature cues was evaluated for face reconstruction. Three camera views are used in model fitting as shown in
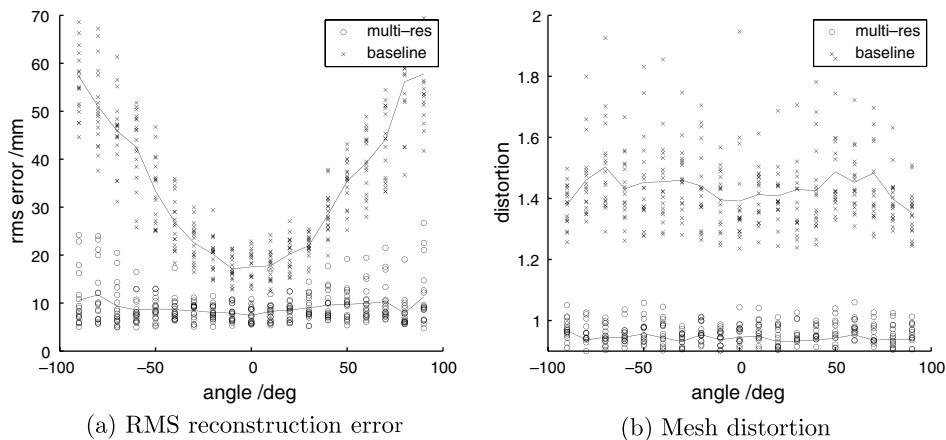

(a) RMS reconstruction error   (b) Mesh distortion

Fig. 11. (a) RMS surface error to ground-truth data and (b) mesh distortion, for the multiple resolution model compared to the reference model in fitting stereo, silhouette and feature data for a range of initial model rotations across 20 test subjects in face reconstruction.
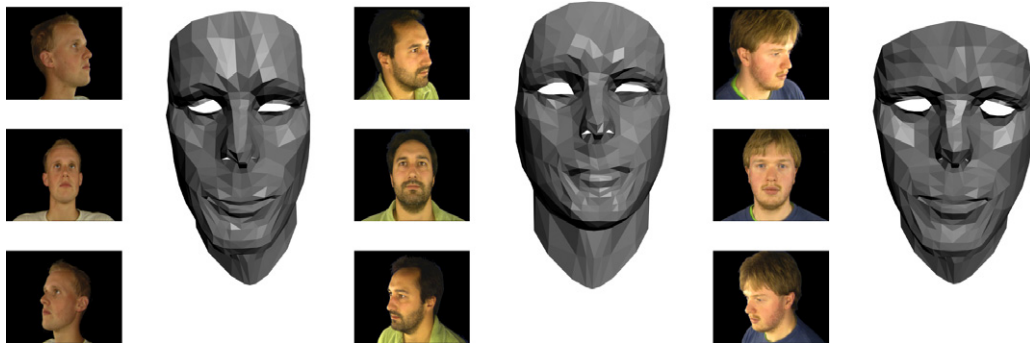
Fig. 12. Model-based shape reconstruction from three wide-baseline camera positions combining stereo, silhouette and feature data constraints.

Fig. 12 representing a challenging problem for both shape from silhouette where limited views are available and stereo reconstruction with wide-baseline camera positions.

Reconstruction accuracy and model distortion are shown in Fig. 11. The multiple resolution model again demonstrates a greater range of convergence compared to the reference model. Model distortion is now found to increase for the reference in fitting the unreliable stereo data across wide-baseline camera views. The multiple resolution technique serves to provide model convergence while maintaining the vertex parameterisation in fitting error prone data. The RMS reconstruction error demonstrates a systematic error compared to the ground-truth data arising from inexact correspondence between the wide-baseline images in regions of low image texture. Subjectively the recovered shape models shown in Fig. 12 provide a close approximation to underlying shape of the face.

5.5. Model regularisation

The proposed technique includes several tunable parameters and the parameter choice was evaluated for shape reconstruction against ground-truth for the face model. In all experiments the initial model error for stereo matching is set at 10 cm and reconstruction accuracy at 1 mm. Image entropy $\sigma_0 = 10$ is fixed according to the noise levels in the imaging system as described in Section 4.4. A conservative annealing schedule is fixed $\lambda = 0.7$ as described in Section 4.5. Model regularisation was then evaluated by changing the shape regularisation term k in Eq. 3 and number of levels of detail (LOD) in the model hierarchy.

Fig. 13 shows the RMS reconstruction error. The choice of k and LOD is expected to be coupled. The model hierarchy provides a multiple resolution constraint in data fitting, as the number of LOD increases, the region over which the shape is constrained is increased for a given k. Hierarchical optimisation both increases the range of convergence for the model and decreases the corresponding shape constraint k required in data fitting. The reconstruction error demonstrates that the range of convergence and accuracy increase up to a critical LOD and regularisation k. As model regularisation increases the model will become over-constrained. In a practical system the minimum number of LOD and regularisation k would be chosen to achieve the desired convergence and accuracy.
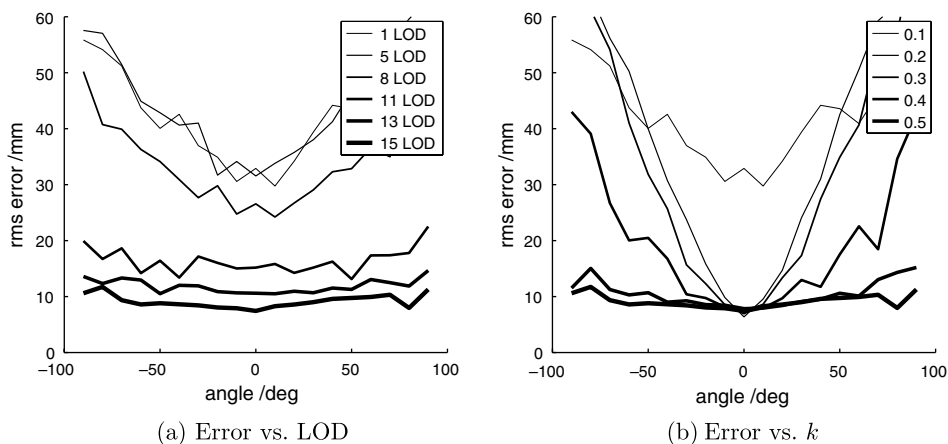

(a) Error vs. LOD

(b) Error vs. k

Fig. 13. RMS surface error to ground-truth data with different regularisation k and number of levels of detail (LOD) in the hierarchical model for 20 test subjects in face reconstruction. Individual data points are excluded for clarity and the mean error is shown.
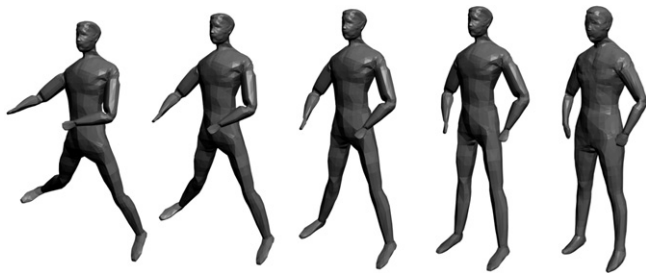
Fig. 14. Initial pose for the whole-body humanoid model showing (left to right) a 30°, 20°, 10° and zero registration error in skeletal pose in comparison with the reconstructed shape for zero pose error.

5.6. Whole-body shape reconstruction

Whole-body shape reconstruction was evaluated over a range of distinct body poses in a multiple view video sequence. Nine camera views were used forming four stereo pairs with one camera positioned overhead. The multiple resolution model was constructed from 15 LOD with a regularisation term $k = 0.5$ as presented for face reconstruction. Feature constraints were applied at the eyes, nose, mouth and limb tips. The body model has a skeleton structure with 17 articulated joints that are posed manually. The surface is animated using a standard vertex weighting scheme in which each bone in the skeleton is associated with a set of mesh vertices with a corresponding set of animation weights. The range of errors in pose used in reconstruction and the corresponding surface configuration is illustrated in Fig. 14.

The RMS reconstruction error and model distortion are shown in Fig. 15. Ground-truth shape is not available for this data set and a relative reconstruction error is derived with respect to the proposed technique with zero pose error. The pose error was introduced about a random axis of rotation in the articulated model structure up to a 30° limit, representing a challenging problem in reconstruction where the model can have self-intersections and up

to 0.5 m error at the limb tips. The proposed technique shows a reduced reconstruction error with an increased range of convergence along with a reduced model distortion.

Shape reconstruction and appearance recovery is shown in Fig. 16 for different body poses. The results demonstrate reconstruction in the presence of complex self occlusions in the camera images. The prior model provides the expected surface shape allowing the estimation of surface visibility in the images in surface optimisation. A disadvantage of model-based approaches to reconstruction should be noted here where the model lacks the detailed geometry of the hands. There is insufficient information in the camera images to reconstruct the fingers which are in the order of one pixel wide. The generic model does not contain hands and so does not provide the prior information necessary to provide a model complete with the structure of the hands. Model-based techniques are inherently limited by the prior model chosen for reconstruction. The model must be selected to represent the space of shapes that need to be recovered and model regularisation must be set to constrain model fitting where insufficient data is available to recover the desired shape.

6. Motion synthesis

The key advantage of a model-based approach to human shape reconstruction is now demonstrated. The whole-body model used in this work is instrumented with a skeletal structure to animate the model surface. The complex process of creating the model dynamics can be performed once and the reconstructed surface models can then be animated. This allows the synthesis of new dynamic content from the original camera images with the freedom to control viewpoint in visualisation.

The complete surface appearance for a person is derived as a single texture map for the model from the camera

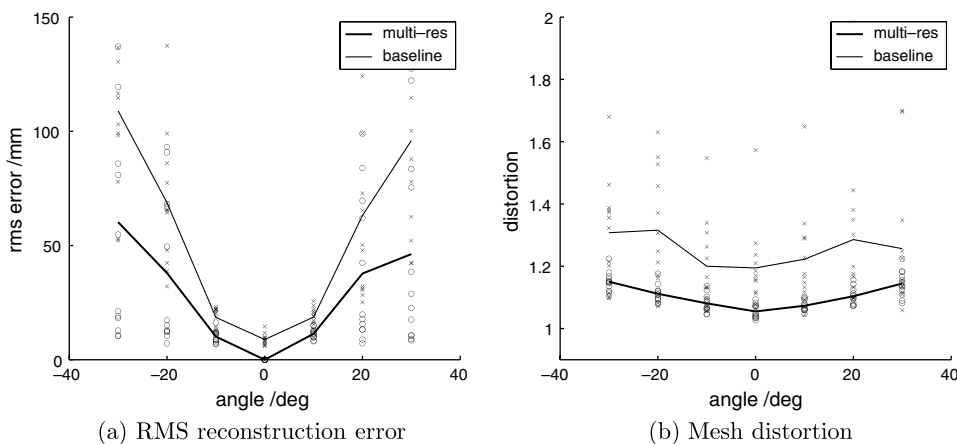(a) RMS reconstruction error

(b) Mesh distortion

Fig. 15. RMS surface error and model distortion for 12 different body poses for the multiple resolution model compared to the reference model. Reconstruction is evaluated for a 30° range of pose errors introduced in the articulated body model. Surface error is provided with respect to the multiple resolution model with zero pose error.
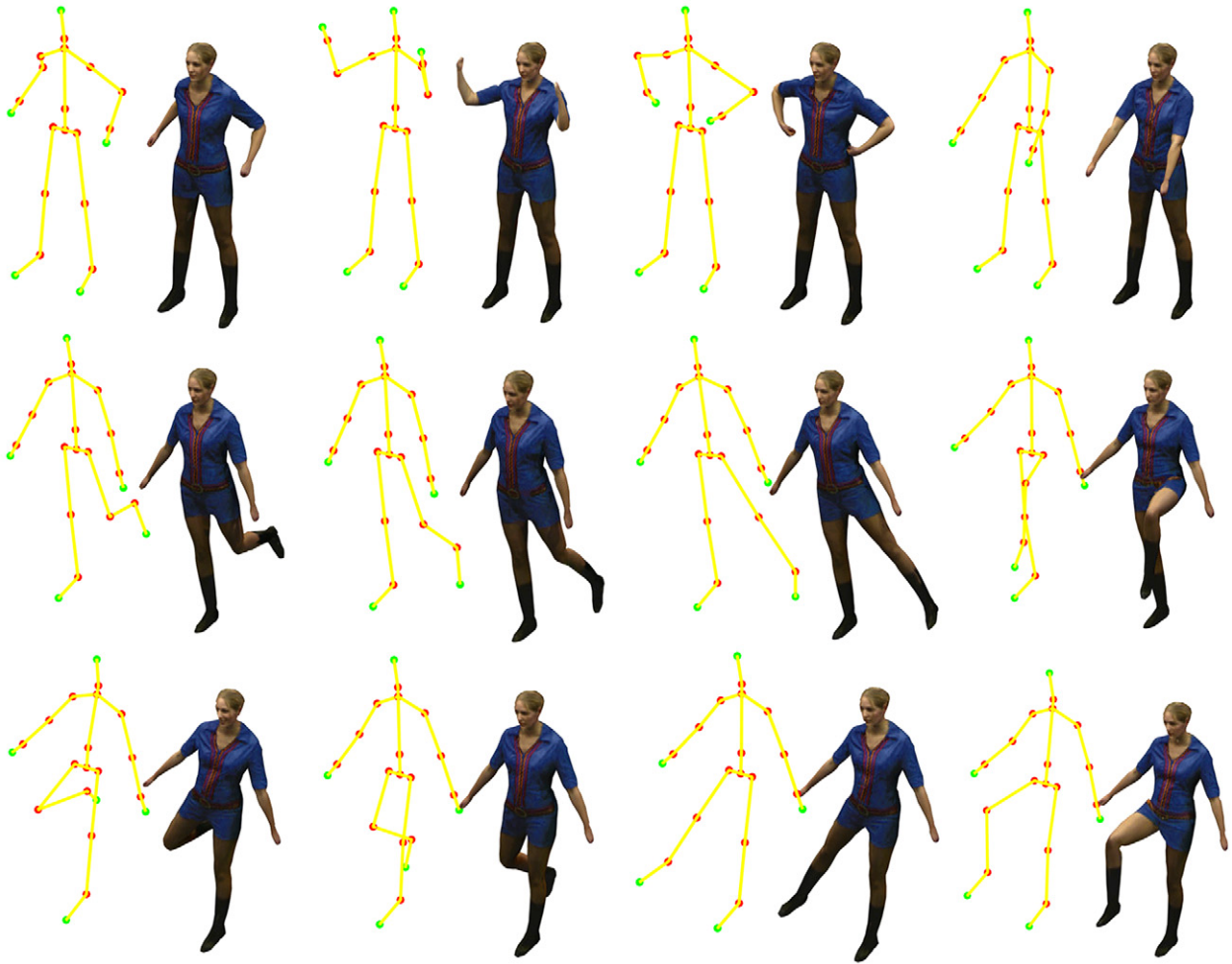
Fig. 16. Model-based shape reconstruction for a range of different body poses with self-occlusions.

images. The 2D texture domain is predefined for the humanoid model. Texture blending is then performed using a weighted average of the textures derived from different camera views. A smooth transition at the texture boundaries can be achieved with a smooth transformation of the relative weight given to each view. Techniques for

Fig. 17. Texture-mapped models for three subjects captured from nine camera views.
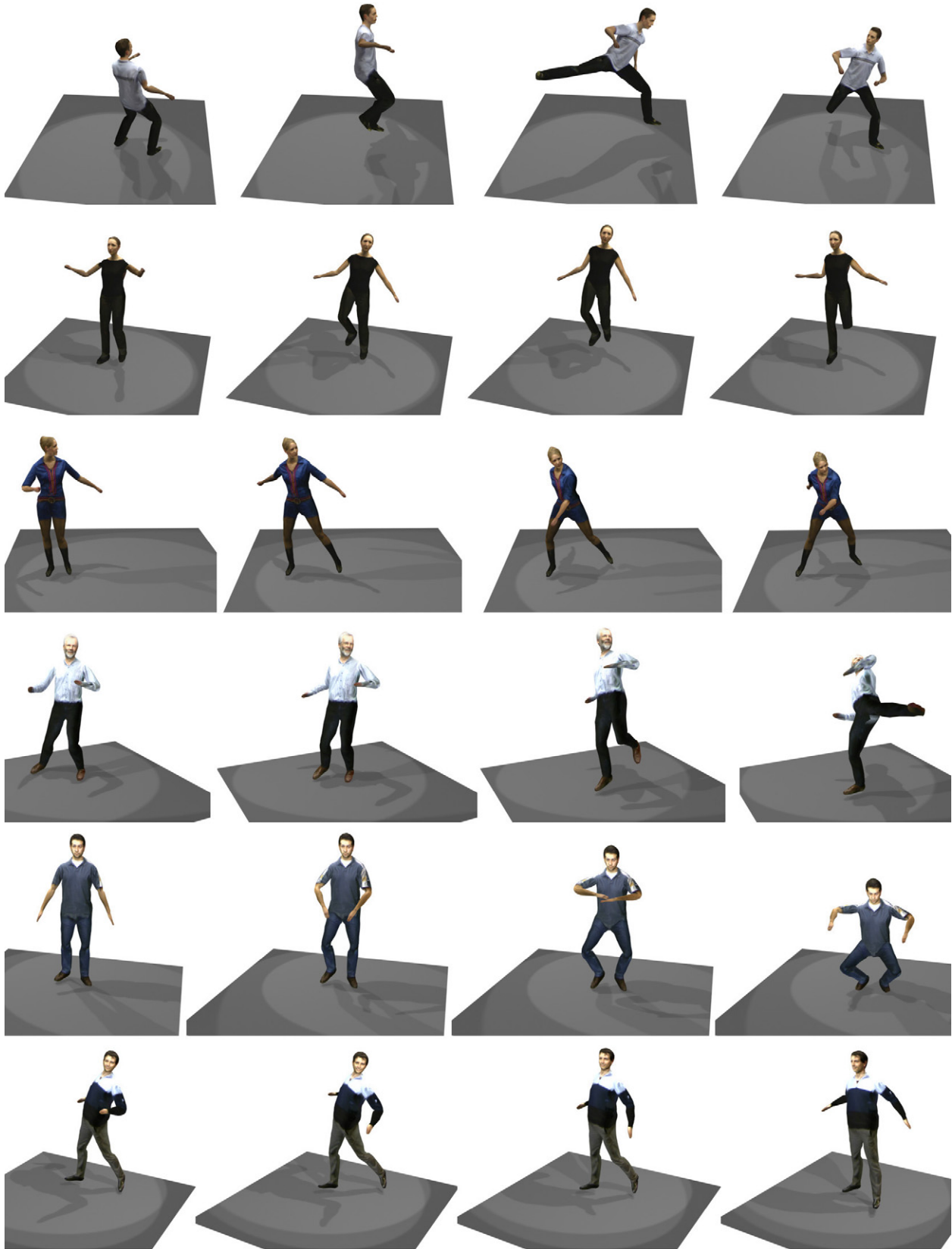
Fig. 18. Animated sequences for six reconstructed subjects created using the animation structure for the generic humanoid model.

blending make use of factors such as the relative orientation of the surface with respect to a camera view and the relative distance to the edge of the texture to define the weighting term [52]. Here a multiple resolution blending

technique is used as introduced by [53]. A multiple resolution approach ensures that the extent of texture blending corresponds to the spatial frequency of the features in the texture image, preserving the higher frequency detail in the resampled appearance. Fig. 17 shows the texture map recovered for models constructed for three different subjects. Given a texture mapped model we are free to generate new animated content as shown in Fig. 18.

7. Conclusion

In this paper, a framework is presented for model-based reconstruction from multiple view images. A general technique is presented for multiple resolution fitting of an open or closed irregular surface mesh of arbitrary topology designed for rendering in a conventional computer graphics pipeline. This allows fitting of a model designed for efficient rendering and animation without distortion of the mesh structure. A local shape parameterisation is introduced that allows reconstruction of vertex positions for surface fitting at multiple levels of detail in a hierarchical model. This local parameterisation is used to introduce an explicit constraint on the mesh vertices at different model resolutions minimising distortion in optimisation and preserving the validity of parameterisation-dependent properties such as animation weights.

A model-based approach to reconstruction is adopted introducing prior knowledge of the scene structure to infer visibility across views and to regularise shape reconstruction in the presence of ambiguous or noisy data. The reconstruction method combines silhouette, stereo and user-defined feature constraints with shape regularisation. Model-based stereo uses a direct search for stereo correspondence at multiple resolutions, avoiding convergence to local minima associated with local surface optimisation techniques. A trade-off is introduced between the complementary cues from silhouette and stereo data, and sparse feature constraints are interpolated across the model using the hierarchical representation.

The proposed technique was evaluated for face and whole-body shape reconstruction using open and closed surface models. The framework demonstrates an improved range of convergence and reconstruction accuracy with a reduced surface distortion in optimisation compared to a baseline single resolution deformable model. The key advantage of this approach is that a common structured representation is obtained for all reconstructed surfaces allowing animation synthesis from the original multiple view images. This structured representation is however based on the assumption that a single underlying model is sufficient to represent the space of observed shapes. Future research should address the representation of detailed surface geometry that is either not present in the underlying geometric model or cannot be represented by

the model surface such as long flowing hair or extensive clothing deformation.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.cviu.2007.10.001.

References

[1] T. Kanade, P. Rander, P. Narayanan, Virtualized reality: constructing virtual worlds from real scenes, IEEE Multimedia 4 (1) (1997) 34–47.

[2] W. Matusik, C. Buehler, R. Raskar, S. Gortler, L. McMillan, Image-based visual hulls, ACM Transactions on Graphics (ACM SIGGRAPH 2000) (2000) 369–374.

[3] D. Terzopoulos, J. Platt, A. Barr, K. Fleischer, Elastically deformable models, Computer Graphics Annual Conference Series 21 (4) (1987) 205–214.

[4] J. Carranza, C.M. Theobalt, M. Magnor, H. Seidel, Free-viewpoint video of human actors, ACM Transactions on Graphics (ACM SIGGRAPH 2003) 22 (3) (2003) 569–577.

[5] Z. Zhang, Z. Liu, D. Adler, M.F. Cohen, E. Hanson, Y. Shan, Robust and rapid generation of animated faces from video images: a model-based modeling approach, International Journal of Computer Vision 58 (2) (2004) 93–119.

[6] B. Allen, B. Curless, Z. Popović, Articulated body deformation from range scan data, ACM Transactions on Graphics (ACM SIGGRAPH 2002) 21 (3) (2002) 612–619.

[7] A. Hilton, D. Beresford, T. Gentils, R. Smith, W. Sun, J. Illingworth, Whole-body modelling of people from multiview images to populate virtual worlds, The Visual Computer 16 (7) (2000) 411–436.

[8] I. Kakadiaris, D. Metaxas, Three-dimensional human body model acquisition from multiple views, International Journal of Computer Vision 30 (3) (1998) 191–218.

[9] J. Starck, A. Hilton, Model-based multiple view reconstruction of people, in: 9th IEEE International Conference on Computer Vision (ICCV 2003), 2003, pp. 915–922.

[10] S. Moezzi, L. Tai, P. Gerard, Virtual view generation for 3D digital video, IEEE Multimedia 4 (1) (1997) 18–25.

[11] J. Starck, A. Hilton, Virtual view synthesis of people from multiple view video sequences, Graphical Models 67 (6) (2005) 600–620.

[12] P. Narayanan, P. Rander, T. Kanade, Constructing virtual worlds using dense stereo, IEEE International Conference on Computer Vision (1998) 3–10.

[13] S. Vedula, P. Rander, H. Saito, T. Kanade, Modeling, combining, and rendering dynamic real-world events from image sequences, Proceedings of Virtual Systems and Multimedia (1998) 323–344.

[14] P. Fua, Y. Leclerc, Object-centered surface reconstruction: combining multi-image stereo and shading, International Journal of Computer Vision 16 (1995) 35–56.

[15] D. Samaras, D. Metaxas, P. Fua, Y. Leclerc, Variable albedo surface reconstruction from stereo and shape from shading, in: CVPR, IEEE Computer Society, 2000, pp. 1480–1487.

[16] J. Isidoro, S. Sclaroff, Stochastic refinement of the visual hull to satisfy photometric and silhouette consistency constraints, in: 9th IEEE International Conference on Computer Vision (ICCV 2003), 2003, pp. 1335–1342.

[17] C. Esteban, F. Schmitt, Silhouette and stereo fusion for 3D object modeling, Computer Vision and Image Understanding 96 (3) (2004) 367–392.

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

[19] L. Zhang, N. Snavely, B. Curless, S. Seitz, Spacetime faces: high resolution capture for modeling and animation, ACM Transactions on Graphics (ACM SIGGRAPH 2004) 23 (3) (2004) 548–558.

[20] K. Cheung, S. Baker, T. Kanade, Shape-from-silhouette of articulated objects and its use for human body kinematics estimation and motion capture, IEEE Conference on Computer Vision and Pattern Recognition (1) (2003) 77–84.

[21] P. Sand, L. McMillan, J. Popovic, Continuous capture of skin deformation, ACM Transactions on Graphics (ACM SIGGRAPH 2003) 22 (3) (2003) 578–586.

[22] J. Montagnat, H. Delingette, N. Ayache, A review of deformable surfaces: topology, geometry and deformation, Image and Vision Computing 19 (2001) 1023–1040.

[23] T. McInerney, D. Terzopoulos, Deformable models in medical image analysis: a survey, Medical Image Analysis 1 (2) (1996) 91–108.

[24] A. Nealen, M. Mller, R. Keiser, E. Boxerman, M. Carlson, Physically based deformable models in computer graphics, Eurographics 2005 state of the art report (STAR) (2005) STAR4.

[25] M. Kass, A. Witkin, D. Terzopoulos, Snakes: active contour models, International Journal of Computer Vision 1 (1988) 321–331.

[26] H. Delingette, M. Hebert, K. Ikeuchi, Shape representation and image segmentation using deformable surfaces, Image and Vision Computing 10 (3) (1992) 132–144.

[27] J. Montagnat, O. Faugeras, Spatial and temporal shape constrained deformable surfaces for 3D and 4D medical image segmentation, Technical Report 4078, INRIA, 2000.

[28] C. Lurig, L. Kobbelt, T. Ertl, Deformable surfaces for feature based indirect volume rendering, Computer Graphics International (1998) 752–760.

[29] R. Szeliski, D. Tonnesen, Surface modeling with orientated particle systems, Computer Graphics 26 (2) (1992) 185–194.

[30] D. Terzopoulos, D. Metaxas, Dynamic 3D models with local and global deformations: deformable superquadrics, IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (1991) 703–714.

[31] R. Whitaker, Volumetric deformable models: active blobs, Visualization in Biomedical Computing (1994) 122–134.

[32] L. Cohen, I. Cohen, Finite element methods for active contour models and balloons for 2D and 3D images, IEEE Transactions on Pattern Analysis and Machine Intelligence 15 (11) (1993) 1131–1147.

[33] R. Malladi, J. Sethian, B. Vemuri, Shape modelling with front propagation: a level set approach, IEEE Transactions on Pattern Analysis and Machine Intelligence 17 (2) (1995) 158–174.

[34] R. Plaenkers, P. Fua, Articulated soft objects for multiview shape and motion capture, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (9) (2003) 1182–1187.

[35] S. Sclaroff, L. Liu, Deformable shape detection and description via model-based region grouping, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (5) (2001) 475–489.

[36] E. Bardinet, L. Cohen, N. Ayache, A parametric deformable model to fit unstructured 3D data, Computer Vision and Image Understanding 71 (1) (1998) 39–54.

[37] J.-O. Lachaud, A. Montanvert, Deformable meshes with automated topology changes for coarse-to-fine three-dimensional surface extraction, Medical Image Analysis 3 (2) (1999) 187–207.

[38] Y. Duan, H. Qin, A subdivision-based deformable model for surface reconstruction of unknown topology, Graphical Models 66 (4) (2004) 181–202.

[39] I. Guskov, W. Sweldens, P. Schroder, Multiresolution signal processing for meshes, ACM Transactions on Graphics (ACM SIGGRAPH 1999) (1999) 325–334.

[40] H. Delingette, General object reconstruction based on simplex meshes, International Journal of Computer Vision 32 (2) (1999) 111–146.

[41] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, W. Stuetzle, Mesh optimization, ACM Transactions on Graphics (ACM SIGGRAPH 1993) 27 (1993) 19–26.

[42] H. Hoppe, Progressive meshes, ACM Transactions on Graphics (ACM SIGGRAPH 1996) (1996) 99–108.

[43] A.W.F. Lee, W. Sweldens, P. Schröder, L. Cowsar, D. Dobkin, Maps: multiresolution adaptive parameterization of surfaces, ACM Transactions on Graphics (ACM SIGGRAPH 1998) (1998) 95–104.

[44] R. Szeliski, D. Tonnesen, D. Terzopoulos, Modeling surfaces of arbitrary topology with dynamic particles, in: CVPR, IEEE Computer Society, 1993, pp. 82–87.

[45] O. Faugeras, B. Hotz, H. Mathieu, T. Vieville, Z. Zhang, P. Fua, E. Theron, L. Moll, G. Berry, J. Vuillemin, P. Bertin, C. Proy, Real-time correlation-based stereo: algorithm, implementation and applications, Technical Report 2013, INRIA, 1993.

[46] A. Laurentini, The visual hull concept for silhouette based image understanding, IEEE Transactions on Pattern Analysis and Machine Intelligence 16 (2) (1994) 150–162.

[47] G. Egnal, M. Mintz, R.P. Wildes, A stereo confidence metric using single view imagery with comparison to five alternative approaches, Image and Vision Computing 22 (12) (2004) 943–957.

[48] D. Baraff, A. Witkin, Large steps in cloth simulation, ACM Transactions on Graphics (ACM SIGGRAPH 1998) (2003) 43–54.

[49] M. Desbrun, M. Meyer, P. Schröder, A.H. Barr, Implicit fairing of irregular meshes using diffusion and curvature flow, ACM Transactions on Graphics (ACM SIGGRAPH 1999) (1999) 317–324.

[50] X. Provot, Deformation constraints in a mass-spring model to describe rigid cloth behavior, in: Graphics Interface, 1995, pp. 147–154.

[51] I. Ypsilos, A. Hilton, S. Rowe, Video-rate capture of dynamic face shape and appearance, 2004, pp. 117–122.

[52] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, D. Salesin, Synthesizing realistic facial expressions from photographs, ACM Transactions on Graphics (ACM SIGGRAPH 1998) (1998) 75–84.

[53] P. Burt, E. Adelson, A multiresolution spline with application to image mosaics, ACM Transactions on Graphics 2 (4) (1983) 217–236.