AFRL-RH-WP-TR-2010-0023

Static and Dynamic Human Shape Modeling – A Review of the Literature and State of the Art

Zhiqing Cheng
Infoscitex Corp.
4027 Colonel Glenn Highway
Suite 210
Dayton OH 45431-1672

Kathleen Robinette
Biosciences and Protection Division
Biomechanics Branch

April 2009

Interim Report for October 2007 to December 2008

Air Force Research Laboratory
711th Human Performance Wing
Human Effectiveness Directorate
Biosciences and Protection Division
Biomechanics Branch
Wright-Patterson AFB OH 45433

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the 88th Air Base Wing Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RH-WP-TR-2010-0023 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//SIGNED//

Julia Parakkat, Work Unit Manager
Biomechanics Branch

//SIGNED//

Mark M. Hoffman, Deputy
Biosciences and Protection Division
Human Effectiveness Directorate
711th Human Performance Wing
Air Force Research Laboratory

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
27-04-2009	Interim	October 2007 – December 2008

4. TITLE AND SUBTITLE
Static and Dynamic Human Shape Modeling – A Review of the Literature and State of the Art

5a. CONTRACT NUMBER
FA8650-07-C-6854

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
63231F

6. AUTHOR(S)
Zhiqing Cheng
Kathleen Robinette

5d. PROJECT NUMBER
7184

5e. TASK NUMBER
02

5f. WORK UNIT NUMBER
71840226

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Infoscitex
4027 Colonel Glenn Highway, Ste 210
Dayton, OH 45431-1672

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Materiel Command
Air Force Research Laboratory
711th Human Performance Wing
Human Effectiveness Directorate
Biosciences and Protection Division
Biomechanics Branch
Wright-Patterson AFB OH 45433-7947

10. SPONSOR/MONITOR'S ACRONYM(S)
711 HPW/RHPA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-RH-WP-TR-2010-0023

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release, distribution is unlimited

13. SUPPLEMENTARY NOTES
88ABW/PA cleared on 19 Jun 09; 88ABW-2009-2663

14. ABSTRACT

This report provides a literature review of human shape modeling. The major topics related to both static and dynamic human shape modeling are addressed. Recent research on these topics is introduced and various theories, techniques, and methodologies are discussed. The report covers the state of the art of human shape modeling methods and highlights key technologies. The contents of the report are arranged as follows: 1. Introduction; 2. Static Shape Modeling; 3. Pose Change Modeling; 4. Dynamic Modeling; 5. Concluding Remarks; 6. References.

15. SUBJECT TERMS
Static Shape Human Modeling; Pose Change Modeling; 3-D Dynamic Human Shape Modeling

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Julia Parakkat
U	U	U	SAR	55	19b. TELEPHONE NUMBER *(include area code)* NA

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. 239.18

THIS PAGE IS INTENTIONAL LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

THIS PAGE IS INTENTIONALLY LEFT BLANK

SUMMARY

This report provides a literature review of human shape modeling. The major topics related to both static and dynamic human shape modeling are addressed. Recent research on these topics is introduced and various theories, techniques, and methodologies are discussed. The report covers the state of the art of human shape modeling methods and highlights key technologies. The contents of the report are arranged as follows: 1. Introduction; 2. Static Shape Modeling; 3. Pose Change Modeling; 4. Dynamic Modeling; 5. Concluding Remarks; 6. References.

THIS PAGE IS INTENTIONAL LEFT BLANK

1.0 INTRODUCTION

Creating a realistic, morphable, animatable, and vivid human shape model is one of the grand challenges for anthropometry and computer graphics. Since people are accustomed to seeing other humans with an acute ability to detect flaws, convincingly modeling human shape, motion, and appearance is difficult. A variety of human shape modeling methodologies is available, which can be classified as either creative or reconstructive depending on how the model is constructed (Thalmann et al. 2004). Some earlier anatomically based modelers, such as Scheepers et al. (1997), Shen and Thalmann (1995) and Wilhelms and Van Gelder (1997) fall into the creative approach. These models intend to mimic actual components of the body and consist of multi-layer structures for simulating individual muscles, bones and tissues. Lately, much work has been devoted to the reconstructive approach. This approach builds a three-dimensional (3-D) geometry of the human body automatically by capturing existing shape data from stereo imagery (Devernay and Faugeras 1994), 3-D scanners (Cordier and Thalmann 2002), and two-dimensional (2-D) images either from video sequences (Fua 1999) or from photos (Hilton et al. 1999, Lee at al. 2000). As 3-D whole body laser scanning becomes a viable technology, building a human shape model from the range scan data becomes a major approach (Allen et al. 2003, Azouz et al. 2005b).

From the perspective of the motion status of the subject to be modeled, human shape modeling can be classified as either static or dynamic. Static shape modeling creates a model to describe human shape at a particular pose, usually a standing pose. The major issues involved in the static shape modeling include shape description, registration, hole filling, shape variation characterization, and shape reconstruction. Dynamic shape modeling addresses the shape variations due to the pose changes or while the subject is in motion. The major issues involved in the shape modeling of pose changes include pose identification, skeleton modeling, and shape deformation. The shape modeling of motion involves issues of motion tracking, shape capturing, shape reconstruction, animation, and inverse kinematics.

Extensive investigations have been performed on human shape modeling. The following list is a small selection of articles that represent the state-of-the-art of human modeling technology:

1. Articulated Body Deformation from Range Scan Data.[2]
2. From 3D Shape Capture to Animated Models.[22]
3. The space of human body shapes: reconstruction and parameterization from range scans.[3]
4. "Synthesizing Animatable Body Models with Parameterized Shape Modifications.[37]
5. Continuous Capture of Skin Deformation. [35]
6. Analysis of Human Shape Variation using Volumetric Techniques.[11]
7. SCAPE: Shape Completion and Animation of People.[6]
8. Capturing and Animating Skin Deformation in Human Motion.[29]
9. A Framework for Natural Animation of Digitized Models.[1]
10. Detailed Human Shape and Pose from Images.[9]

The models presented in these papers exhibit various unique features that may be of great interest to the H-MASINT program. A more detailed description of these models is provided in a separate presentation. The major topics of human shape modeling will be addressed in the

following sections where various theories, techniques, and methodologies used in human shape modeling, including those presented in the above list, will be discussed.

2.0 STATIC SHAPE MODELING

2.1 Shape Description

In general, a shape descriptor is a set of numbers that are produced to describe a given shape. A shape may not be entirely reconstructable from its descriptor, but the descriptors for different shapes should be different enough that the shapes can be discriminated. Shape description is a fundamental problem for human shape modeling. Traditional anthropometry is based on a set of measurements corresponding to linear distances between anatomical landmarks and circumference values at predefined locations. These measurements provide limited information about the human body shape (Robinette et al. 1997). With the advances in surface digitization technology, a 3-D surface scan of the whole body can be acquired in a few seconds. While whole body 3-D surface scan provides very detailed description of the body shape, the verbose scan data cannot be used directly for shape analysis. Therefore, it is necessary to convert 3-D scans to some form of useful representation for the latter purpose. For searching and mining from a large 3-D scan database, Robinette (2003) investigated 3-D shape descriptors where the Paquet Shape Descriptor (PSD) (as shown in Figure 1.) developed by Paquet and Rioux (1998) was examined in detail. While PSD is able to discriminate or characterize different human shapes, it is not invertible. In other words, it is impossible to reconstruct a human shape from PSD.

An ideal human shape descriptor should be concise, unique, and complete for human shape description, efficient for shape indexing and searching, and invertible for shape reconstruction. Finding such a descriptor still remains a challenge. Alternatively, various graphic elements or graphic representation methods can be used to describe the human shape. For instance, Allen et al. (2003) and Anguelov et al. (2005b) basically dealt directly with the vertices or polygons of a scanned surface for shape description; and Allen et al. (2002) used subdivision of the surface in their pose modeling. Ben Azouz et al. (2004) utilized volumetric representation to convert vertices to voxels in their human shape modeling. While these methods guarantee reconstruction, they are not quite efficient for shape identification, discrimination, and searching.

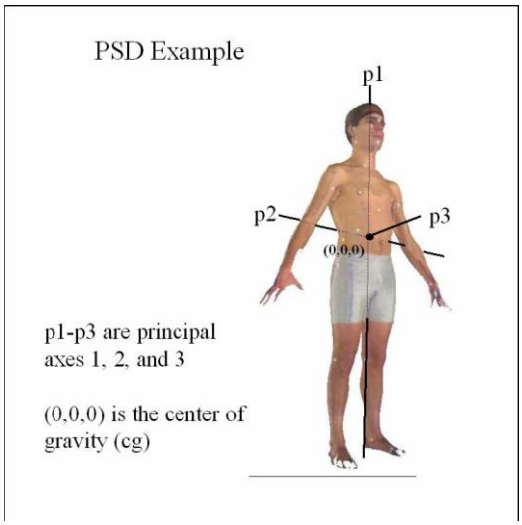


Figure 1. Illustration of principal axis system and PSD (Robinette 2003).

2.2 Registration or Point-to-Point Correspondence

Three-dimensional full body scanning technology allows efficient full scale digitization of the human body and provides a tremendous amount of shape information about the body. The 3-D scan data can be used to measure, compare, and conduct statistics and has applications in, for example, the ergonomic design of products such as automobiles, furniture, and clothes, and also in human identification. Processing the surface data collected by scanners, however, proves to be a challenge. The main difficulty is the surface registration or point-to-point correspondence among the scan data of different subjects. The digitized scan data may have different number of points since the human body comes in all shapes and sizes. This correspondence information is essential to many problems such as the study of human variability (Allen et al. 2003, Ben Azouz et al. 2005b) and pose modeling and animation (Allen et al. 2002, Anguelov et al. 2005b), as long as multiple subjects or multiple poses are involved. One methodology for establishing point-to-point correspondence among different scan data sets or models is usually called non-rigid registration. Given a set of markers between two meshes (two approximate representations of a surface), the task of non-rigid registration is to bring the meshes into close alignment while simultaneously aligning the markers.

Allen et al. (2002, 2003) solved the correspondence problem between subjects by deforming a template model which is a hole-free, artist-generated mesh to fit individual scans, as shown in Figure 2. The resulting individually fitted scans or individual "models" all have the same number of triangles and point-to-point correspondences. The fitting process relies on a set of anthropometric landmarks provided in the Civilian American and European Surface Anthropometry Resource (CAESAR) database (Robinette et al. 1999). These landmarks were marked on the measured subjects prior to scanning. They provide the seed correspondences that guide the deformation of the rest of the points.

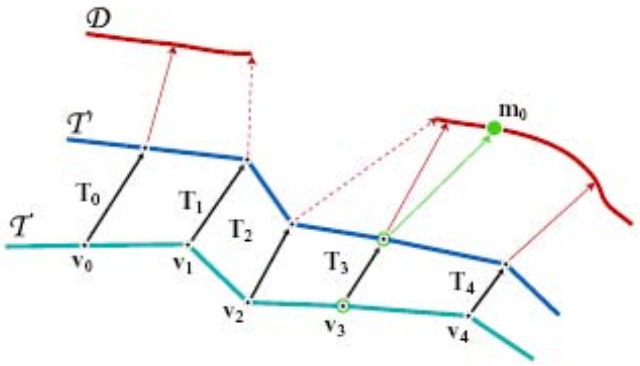
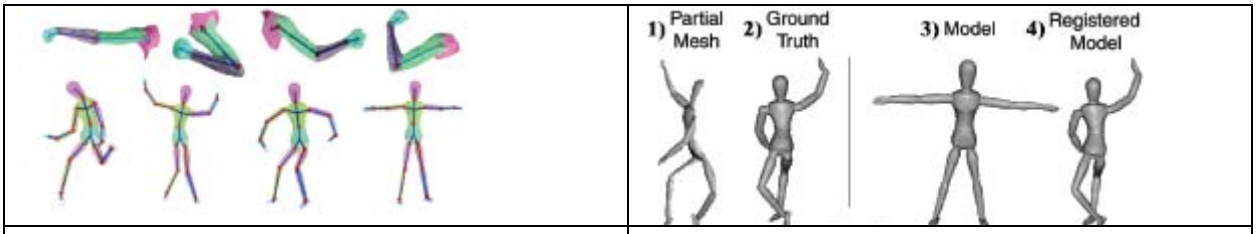


Figure 2. Summary of matching framework (Allen et al.. 2003).

To find a set of affine transformations Ti, that, when applied to the vertices vi of the template surface T, result in a new surface T0 that matches the target surface D. This diagram shows the match in progress; T0 is moving towards D, but has not yet reached it. The match proceeds by minimizing three error terms. The data error, indicated by the red arrows, is a weighted sum of the squared distances between the transformed template surface and D. Note that the dashed red arrows do not contribute to the data error because the nearest point on D is a hole boundary. The smoothness error penalizes differences between adjacent Ti transformations. The marker error penalizes distance between the marker points on the transformed surface and on D (here v3 is associated with m0).

Anguelov et al. (2005a) developed an unsupervised algorithm for registering 3-D surface scans of an object among different poses undergoing significant deformations. The algorithm called Correlated Correspondence (CC) does not use markers, nor does it assume prior knowledge about object shape, the dynamics of its deformation, or scan alignment. The algorithm registers two meshes with significant deformations by optimizing a joint probabilistic model over all point-to-point correspondences between them. This model enforces preservation of local mesh geometry, as well as more global constraints that capture the preservation of geodesic distance between corresponding point pairs. The algorithm applies even when one of the meshes is an incomplete range scan; thus, it can be used to automatically fill in the missing surface areas of a partial scan, even if those missing surfaces were previously only seen in a different configuration. The results produced by the CC algorithm are shown in Figure 3.

Anguelov et al. (2005b) used the CC algorithm to obtain the markers for the non-rigid registration, as shown in Figure 4. The CC algorithm computes the consistent embedding of each instance mesh (the mesh of a particular pose) into the template mesh (the mesh of a reference pose) that minimizes the deformation from the template mesh to the instance mesh and matches similar-looking surface regions.



(a) The results produced by the CC algorithm were used for unsupervised recovery of articulated models. 15 puppet parts and 4 arm parts, as well as the articulated object skeletons, were recovered.

(b) Partial view completion results. The missing parts of the surface were estimated by registering the partial view to a complete model of the object in a different configuration.

Figure 3. The CC algorithms for point-to-point registration (Anguelov et al. 2005a).

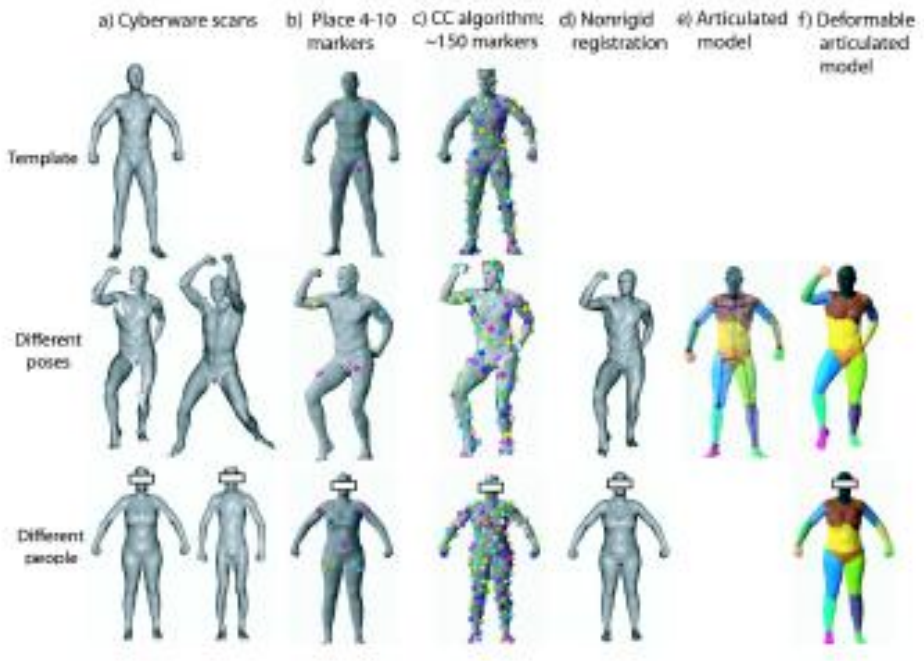


Figure 4. Mesh processing pipeline used to generate a training set (Anguelov et al. 2005b). *(a) Two data sets spanning the shape variability due to different human poses and different physiques were acquired. (b) A few markers were selected by hand to map the template mesh and each of the range scans. (c) Correlated Correspondence algorithm is used to compute numerous additional markers. (d) The markers are used as the input to a non-rigid registration algorithm, producing fully registered meshes. (e) A skeleton reconstruction algorithm is implemented to recover an articulated skeleton from the registered meshes. (f) The space of deformations due to pose and physique is learnt.*

Ben Azouz et al. (2005b) used a volumetric representation of human 3-D surface to establish the correspondences between the scan data of different subjects. By converting their polygonal mesh descriptions to a volumetric representation, the 3D scans of different subjects are aligned inside a volume of fixed dimensions, which is sampled to a set of voxels. A human 3-D shape is then characterized by an array of signed distances between the voxels and their nearest point on the

body surface. Correspondence is achieved by comparing for each voxel the signed distances attributed to different subjects. Surfaces are reconstructed from the volumetric description using the marching cubes algorithm. The main advantage of the volumetric representation is that correspondence between the surfaces of different subjects is achieved without using anatomical landmarks, thus avoiding a time-consuming process of placing markers at the landmark locations before scanning. This landmark-free approach allows the comparison and the extraction of certain modes of variation of the human body. However, it has the limitation that the volumetric representation provides only an approximation of the anthropometric correspondence between different human subjects (Ben Azouz et al. 2006).

2.3 Hole Filling

Surfaces acquired with scanners are typically incomplete and contain holes. Filling a hole is a challenging problem in its own right, as discussed by Davis et al. (2002). A common way to complete a hole is to fill it with a smooth surface patch that meets the boundary conditions of the hole (Curless and Levoy 1996, Davis et al. 2002, Liepa 2003). While these methods fill holes in a smooth manner, which is reasonable in some areas such as the top of the head and possibly in the underarm, other areas should not be filled smoothly. Therefore, Allen et al. (2003) developed a method that maps a surface from a template model (a hole-free, artist-generated mesh) to the hole region by minimizing the smoothness error, as shown in Figure 5.

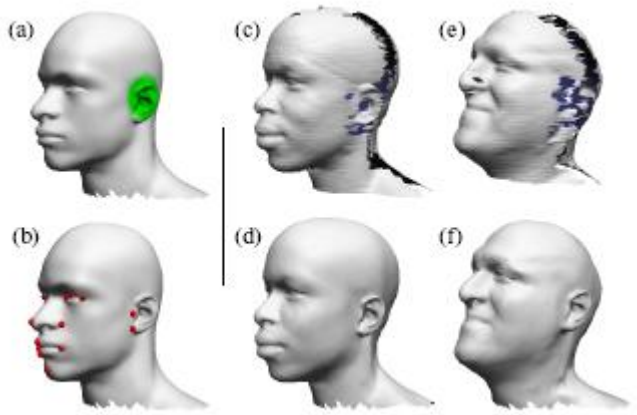


Figure 5. Using a template mesh to synthesize detail lost in the scan (Allen et al. 2003). *(a) The template mesh. Since the ear was not well scanned, the weight of ear vertices is set to zero (shown in green). (b) Since the template mesh does not have the CAESAR markers, a different set of markers based on visually-identifiable features is used to ensure good correspondence. (c) A head of one of the subjects. Interior surfaces are tinted blue. (d) The template head has been deformed to match the scanned head. Note that the ear has been filled in. (e) Another scanned head, with a substantially different pose and appearance from the template. (f) The template mapped to (e). The holes have been filled in, and the template ear has been plausibly rotated and scaled.*

Alternatively, hole-filling can be based on the contour lines of a scan surface (Ben Azouz et al. 2005b), as shown in Figure 6. These approaches work well when the holes are small compared to the geometric variation of the surface. In contrast to small hole-filling, the filling of large holes or partial view completion is often required. Anguelov et al. (2005b) developed a model-based

method that projects the template surface to the target region with a set of observed markers via optimization, as illustrated in Figure 7.

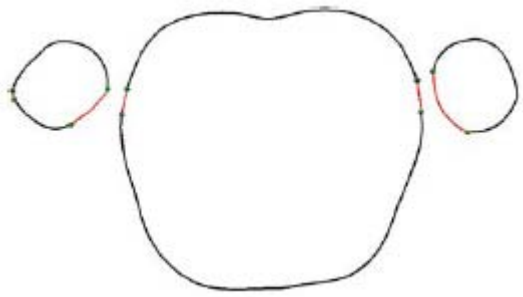


Figure 6. Hole-filling of a slice (Ben Azouz et al. 2005b). Black curves correspond to original data and red curves correspond to estimated data.

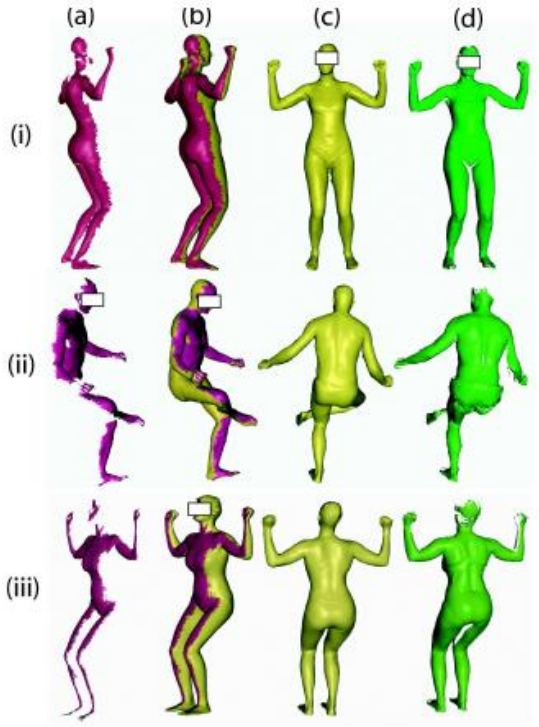


Figure 7. Examples of view completion (Angelov et al. 2005b).
Each row represents a different partial view scan. Subject (i) is in the shape data set but not in the pose data set; neither subjects (ii) and (iii) nor their poses are represented in the data set. (a) The original partial view. (b) The completed mesh from the same perspective as (a), with the completed portion in yellow. (c) The completed mesh from a view showing the completed portion. (d) A true scan of the same subject from the view in (c).

2.4 Shape Variation Characterization

The human body comes in all shapes and sizes. Characterizing the range of human body shape variation has applications ranging from better ergonomic design of human spaces (e.g., chairs, car compartments, and clothing) to easier modeling of realistic human characters for computer animation. Characterizing human shape variation is traditionally the subject of anthropometry—the study of human body measurement. The sparse measurements of traditional anthropometric shape characterization curtail its ability to capture the detailed shape variations needed for realism. While characterizing human shape variation based on a 3-D range scan could capture the details of shape variation, the method relies on three conditions: noise elimination, hole-filling and surface completion, and point-to-point correspondence.

Also, whole body scanners generate verbose data that cannot be used directly for shape variation analysis. Therefore, it is necessary to convert 3-D scans to a compact representation that retains information about the body shape. Principal components analysis (PCA) is a potential solution to this need. After de-noising and hole-filling, the central issue in applying PCA to 3-D anthropometric data is bringing the surfaces of all subjects in correspondence to each other. Allen et al. (2003) captured the variability of human shape by performing PCA over the displacements of the points from the template surface to an instance surface. Anguelov et al. (2005b) also used PCA to characterize the shape deformation and then used the principal components for shape completion. Ben Azouz et al. (2005b) applied PCA to the volumetric models where the vector is formed by the signed distance from a voxel to the surface of the model. While these methods retain the details of human shape and reveal the global variations of shape variations, the problem size of PCA is usually huge, as the vector usually contains a great number of elements.

Principal component analysis helps to characterize the space of human body variation, but it does not provide a direct way to explore the range of bodies with intuitive control parameters, such as height, weight, age, and sex. Allen et al. (2003) showed how to relate several variables simultaneously by learning a linear mapping between the control parameters and the PCA weights. Ben Azouz et al. (2004, 2005b) attempted to link the principal modes to some intuitive body shape variations by visualizing the first five modes of variation and gave interpretations of these modes, as shown in Table 1 and Table 2 as well as Figure 8 and Figure 9. While PCA is shown to be effective in characterizing global shape variations, it may smear local variations for which other methods (e.g., wavelets) may be more effective. Still, PCA is just one tool in the statistician's toolbox, a tool that treats the data as samples drawn from a single, multi-dimensional Gaussian distribution. Applying more sophisticated analyses (e.g., mixtures of Gaussians) to determine the true landscape of human shape variations remains an area for future work (Allen et al. 2003).

Table 1 Mode interpretation for non-normalized heights (Ben Azouz el al 2005b)

Mode	Interp.	Variability
1st	weight & height	35.0%
2nd	*weight/(height)*^{3}	15.0%
3rd	alignment artifact	9.53%
4th	leaning posture	4.02%
5th	muscularity	3.17%

Table 2 Mode interpretation for normalized heights (Ben Azouz el al 2005b)

Mode	Interp.	Variability
1st	weight	33.86%
2nd	leaning posture	15.11%
3rd	muscularity	8.93%
4th	arm-torso spacing	4.0%
5th	head position	3.64%

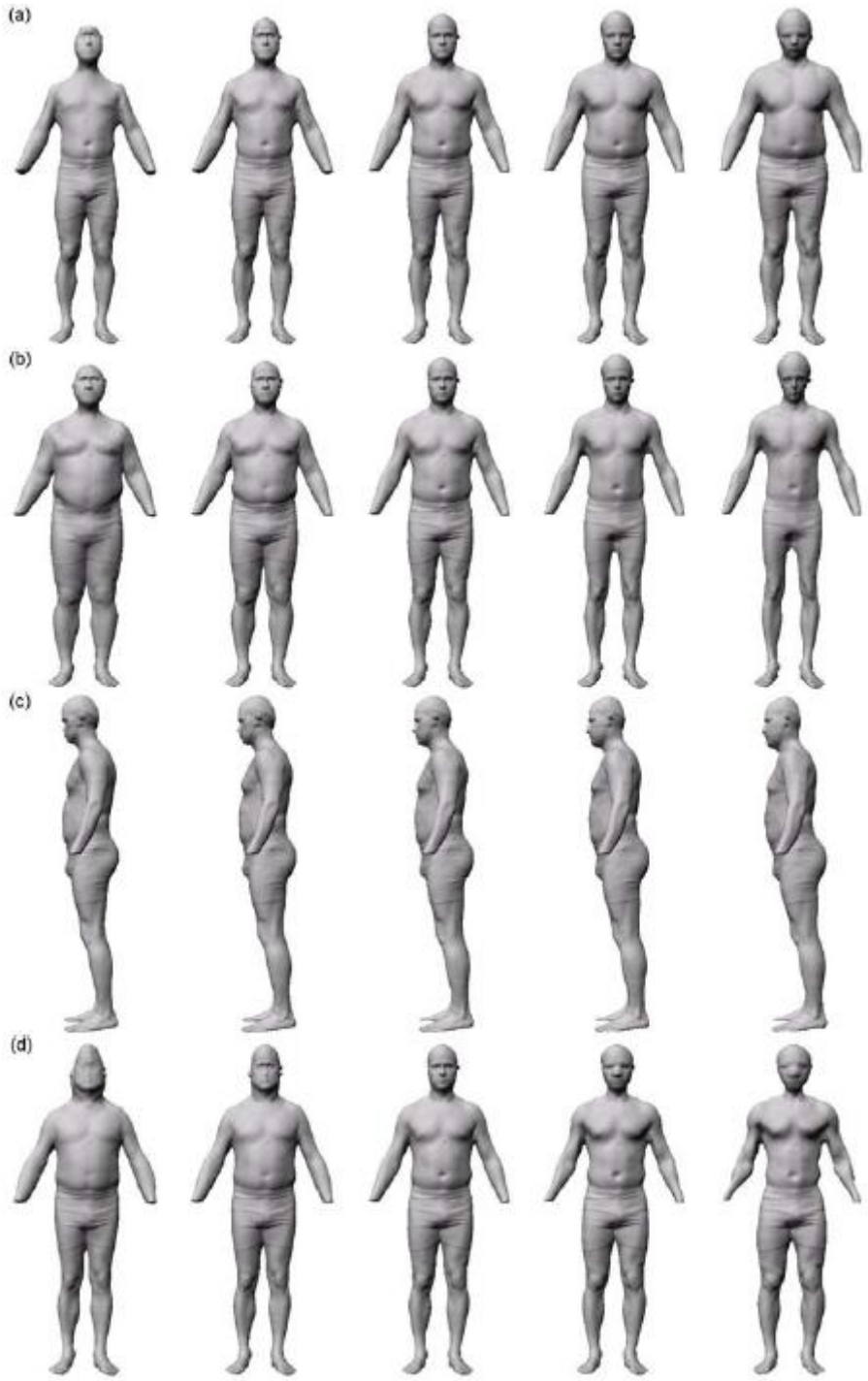


Figure 8. Shape variation induced by some of main components (Ben Azouz et al. 2004). *Components are sorted in decreasing order of their variances. (a) The first component is correlated to the height. (b) The second component is correlated to the weight and the height. (c) The third component reflects a posture variation. (d) The fourth component corresponds to a variation of muscularity.*

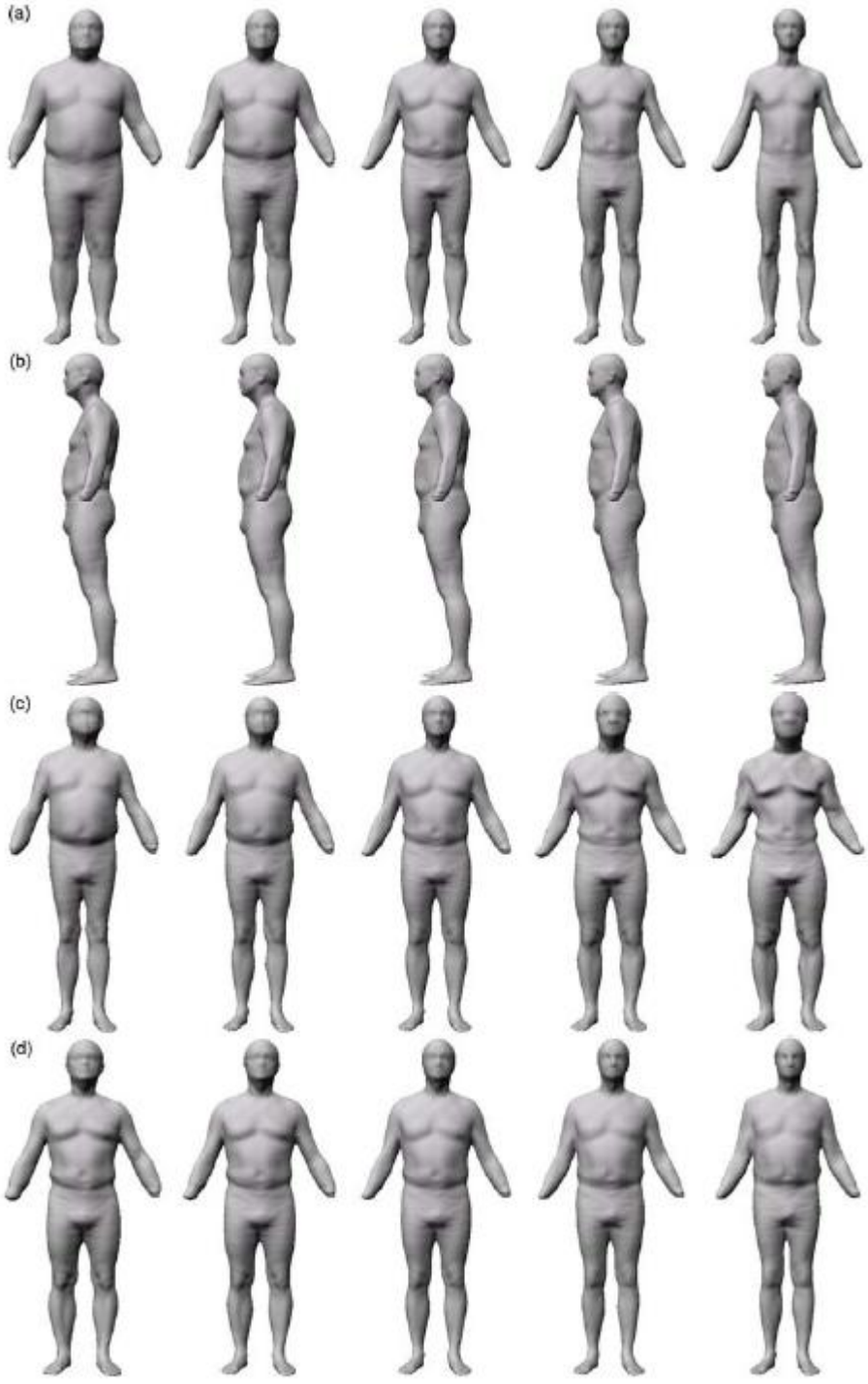


Figure 9. Shape variation induced by some of the main components after height normalization (Ben Azouz et al. 2004).

Components are sorted in decreasing order of their variances. (a) The first component is correlated to the weight. (b) The second component reflects a posture variation. (c) The third component corresponds to a variation of muscularity. (d) The fourth component corresponds to a variation of mass distribution between the torso and the legs and a variation in the arms position.

2.5 Shape Reconstruction

Given a number of scan data sets of different subjects, a novel human shape can be created that will have resemblance to the samples but is not the exact copy of any existing shapes. This can be realized via morphing. There are four ways of morphing.

2.5.1 Direct Interpolation

We can morph between any two subjects' scans by taking linear combinations of their vertices. Figure 10 demonstrates this way of morphing (Allen et al. 2003). In order to create a faithful intermediate shape between two individuals, it is critical that all features are well-aligned; otherwise, features will cross-fade instead of moving.

Figure 10. Morphing between individuals (Allen et al. 2003).
Each of the key frame models (outlined) are generated from a Gaussian distribution in PCA space. These synthesized individuals have their own character, distinct from those of the original scanned individuals. The in-between models are created by linearly interpolating the vertices of the key frames.

2.5.2 Reconstruction from Eigen-space

After PCA analysis, the features of sample shapes are characterized by eigen-vectors or eigen-persons which form an eigne-space. Any new shape model can be generated from this space by combining a number of eigen-models with appropriate weighting factors. Figure 11 illustrates an example (Ben Azouz et al. 2005b).

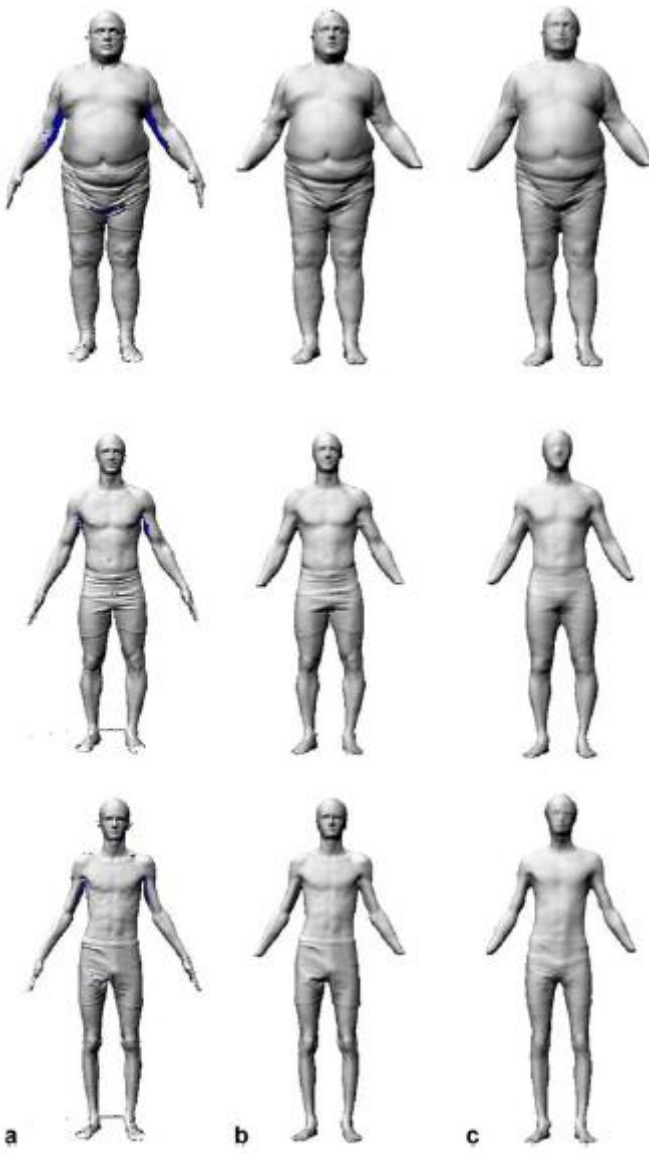


Figure 11. Reconstruction of human models (Ben Azouz et al. 2005b).
Use the first 64 eigenvectors extracted from the volumetric representation of 300 male subjects. (a) Original models; (b) repaired models; (c) reconstructed models.

2.5.3 Feature-based Synthesis

Once the relationship between human anthropometric features and eigen-vectors is established (as we discussed in the preceding section), a new shape model can be constructed from the eigen-space with desired features by editing multiple correlated attributes (such as height and weight), as shown in Figure 12.

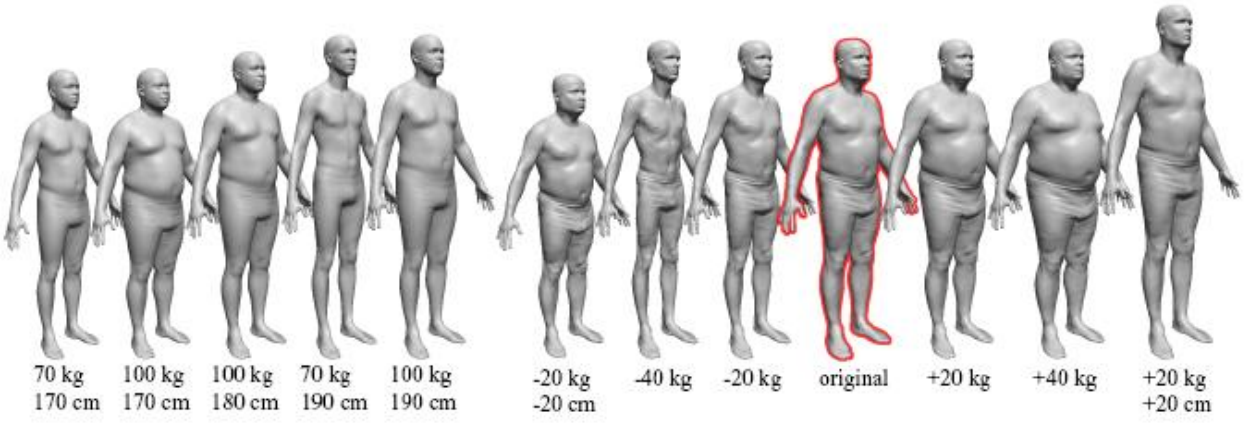


Figure 12. Feature-based synthesis (Allen et al. 2003).

The left part demonstrates feature-based synthesis, where an individual is created with the required height and weight. The right demonstrates feature-based editing. The outlined figure is one of the original subjects that has been parameterized. The gray figures demonstrate a change in height and/or weight. Notice the double-chin in the heaviest example, and the boniness of the thinnest example.

Seo et al. (2003) developed a synthesizer for obtaining higher level of manipulations of body models by using parameters such as fat percentage and hip-to-waist ratio. Figure 13 and Figure 14 are the illustrations of their models.

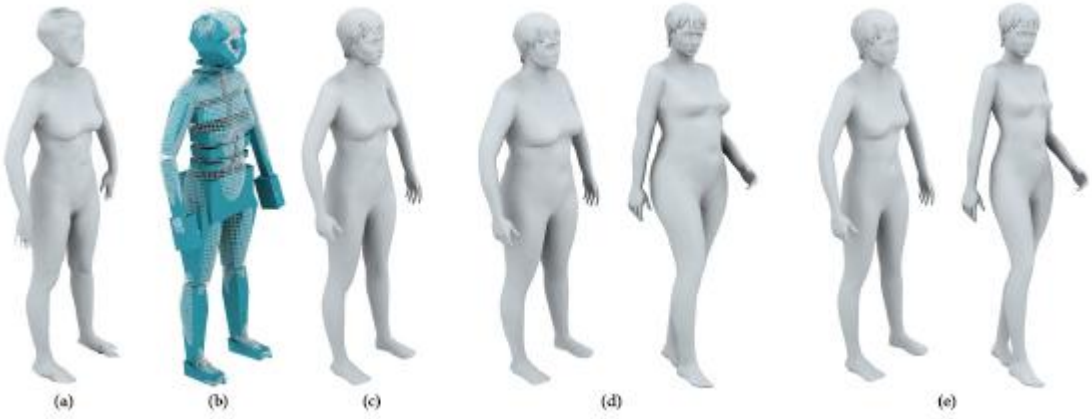


Figure 13. Synthesis of human bodies (Seo et al. 2003).

(a) scan data; (b) template model with animation structure is fitted to the scan data; (c) the fitted template mesh (d) modification of the physique (fat percent 38%) and modified posture; (e) modification of the physique (fat percent 22%) and modified posture.

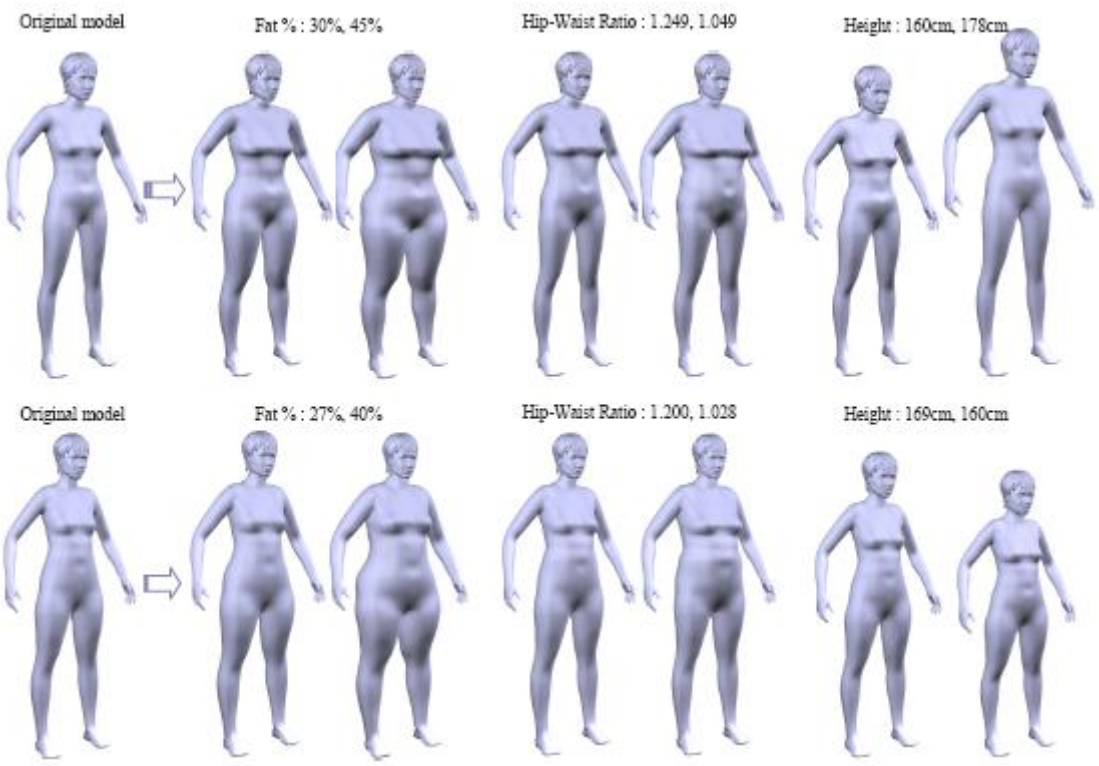


Figure 14. Modification of two individuals controlled by fat percent, hip-waist ratio and height (Seo et al. 2003).

2.5.4 Marker-only Matching

Marker-only matching can be considered as a way of reconstruction with provided markers. This is important for many applications such as deriving a model from video imagery, since marker data can be obtained using less expensive equipment than a laser range scanner (e.g., using a handful of calibrated photographs of a stationary subject). Figure 15 is an example from Allen et al. (2003).

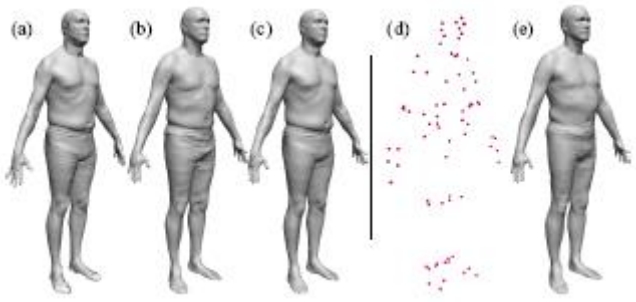


Figure 15. PCA-based fitting (Allen et al. 2003).

(a) A scanned mesh that was not included in the data set previously, and does not resemble any of the other scans. (b) A surface match using PCA weights and no marker data. (c) Using (b) as a template surface to get a good match to the surface without using markers. (d) Using very sparse data; in this case, only 74 marker points. (e) A surface match using PCA weights and no surface data.

3.0 POSE CHANGE MODELING

During pose changing or body movement, muscles, bones, and other anatomical structures continuously shift and change the shape of the body. For pose modeling, scanning the subject in every pose is impractical; instead, body shape can be scanned in a set of key poses, and then the body shape corresponding to intermediate poses can be determined by smoothly interpolating among these poses using scattered data interpolation techniques. The issues involved in pose modeling include pose definition and identification, skeleton model derivation, shape deformation (skinning), and pose mapping.

3.1 Pose Definition and Identification

The human body can assume various poses. In order to have a common basis for pose modeling, we need to have a distinct, unique description of difference poses. Since it is impossible to collect the data or create template models for all possible poses, it is necessary to define a set of standard, typical poses. This is pose definition. A convention for pose definition is yet to be established. One approach is to use joint angle changes as the measures to characterize human pose changing and gross motion. This means that poses can be defined by joint angles. By defining poses and motion in such a way, the body shape variation caused by pose changing and motion will consist of both rigid and non-rigid deformation. Rigid deformation is associated with the orientation and position of segments that connect joints. Non-rigid deformation is related to the changes in shape of soft tissues associated with segments in motion, which, however, excludes local deformation caused by muscle action alone.

In order to measure, define, or represent joint angles, one method (which may be the best method) is to use a skeleton model. In the model, the human body is divided into multiple segments according to major joints of the body, each segment is represented by a rigid linkage, and an appropriate joint is placed between the two corresponding linkages. The skeleton model derivation will be discussed in the following section.

Given a set of scan data, imagery, or photos, the determination or identification of the corresponding pose can be done by fitting a skeleton model to the data set. Alternatively, there are several methods for pose identification that are not based on skeleton models. Mittal et al. (2003) studied human body pose estimation using silhouette shape analysis. They developed an algorithm for human body pose estimation from multiple views that is fast and completely automatic. The algorithm works in the presence of multiple people by decoupling the problems of pose estimation of different people. The pose is estimated based on a likelihood function that integrates information from multiple views and thus obtains a globally optimal solution. Figure 16 shows the parts obtained via segmentation. Figure 17 illustrates the results of the algorithm for a person at a particular time instant from multiple perspectives.

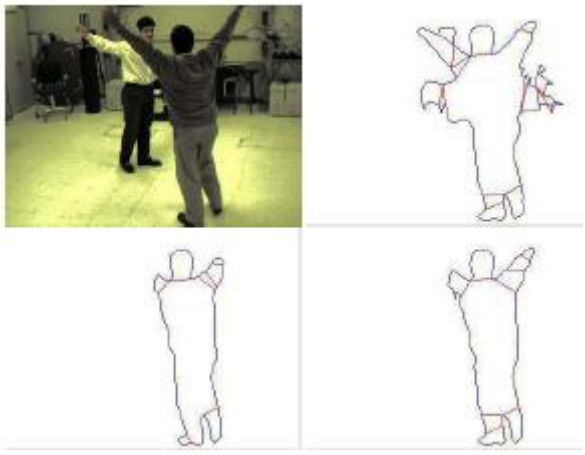


Figure 16. Multiple Body parts obtained via segmentation (Mittal et al. 2003).

Figure 17. A person at a particular time instant from multiple perspectives (Mittal et al. 2003).

Cohen and Li (2003) proposed an approach for inferring the body posture using a 3D visual-hull (a bounding geometry of the actual 3D object) constructed from a set of silhouettes. An appearance-based, view-independent, 3D shape description was introduced for classifying and identifying human posture using a support vector machine (a set of related supervised learning methods used for classification and regression). This shape representation is used to train a support vector machine that characterizes human body postures from the computed visual hull. Figure 18 is the overview of their proposed approach.

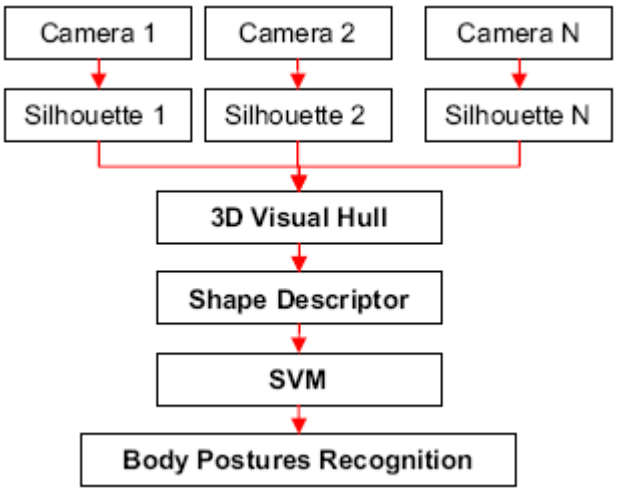


Figure 18. Overview of the proposed approach (Cohen and Li 2003).

3.2 Skeleton Model

Allen et al. (2002) constructed a kinematic skeleton model to identify the pose of a scan data set using markers captured during range scanning. The goal is to have a skeleton model that is a good representation of true human kinematics but not too complicated to construct and animate. Their upper body skeleton model is shown in Figure 19.

Anguelov et al. (2004) addressed the problem of unsupervised learning of complex articulated object models from 3-D range data and developed an algorithm that automatically recovers a decomposition of the object into approximately rigid parts, the location of the parts in the different poses, and the articulated object skeleton linking the parts. Figure 20 illustrates the procedure of this method, and Figure 21 shows a set of poses recovered.

Robertson and Trucco (2006) developed an evolutionary approach to estimating upper-body posture from multi-view markerless sequences. In their method, a skeleton model with 24 degree-of-freedom (DOF) was fitted to sparse 3-D stereo data set from an array of cameras. A particle swarm optimization algorithm was used which can incorporate constraints and does not require motion models. The high-dimensional search space is subdivided based on limb dynamics from application sequences and the hierarchical fitting from the least to the most uncertain body parts is performed. Figure 22 displays the joint positions and coordinate system of the skeleton model used. Figure 23 shows the phases of hierarchical fitting. Figure 24 illustrates two examples of fitted skeletons. Figure 25 is the example fittings.

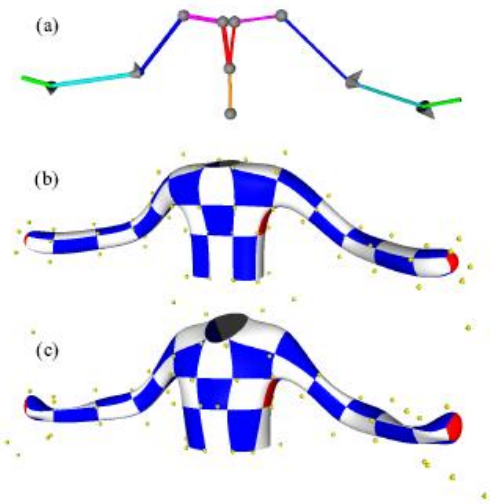


Figure 19. Upper body skeleton (Allen et al. 2002).
The large spheres are quaternion joints, and the cones are single-axis joints. (b) The control points for this skeleton, and the corresponding subdivision surface. The checkerboard pattern delineates the subdivision patches. (c) The control points and subdivision surface after refitting.

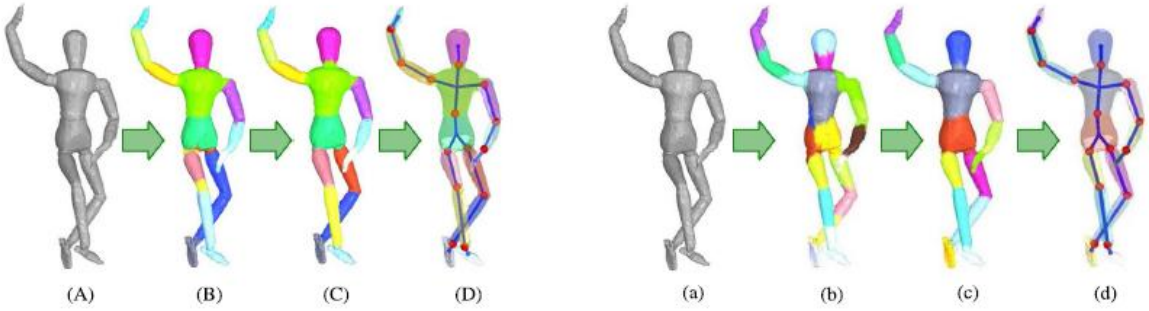


Figure 20. Illustration of the part-finding process (Anguelov et al. 2004).
(a) A template mesh is registered to all other meshes by CC algorithm. (B) The mesh is divided into parts by clustering the estimated local transformations for each template point, different parts are color-coded. (b) The mesh is randomly divided into small patches of approximately equal areas, different parts are color-coded. (C), (c) results in (B), (b) are used to initialize the EM algorithm which solves for the part assignments and the transformation for each part. (D), (d) the joints linking the rigid parts are estimated.

Figure 21. Four different poses from the puppet dataset (Anguelov et al. 2004). The articulated skeleton with 15 rigid parts is recovered automatically.

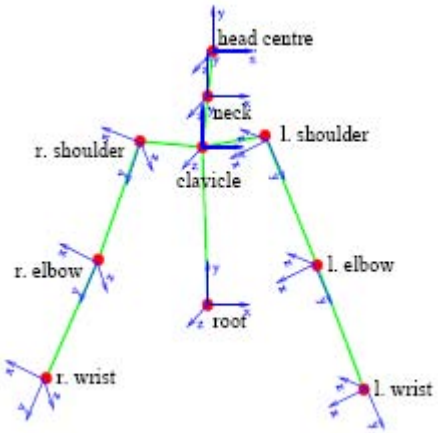


Figure 22. Parameterization of joint positions and coordinate systems (Robertson and Trucco 2006).

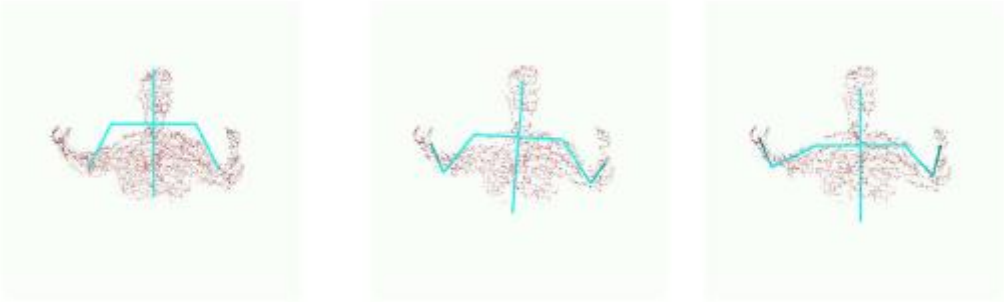


Figure 23. Some phases of the hierarchical fitting (Robertson and Trucco 2006).

Figure 24. Two examples of fitted skeletons (Robertson and Trucco 2006).

(a) Real teleconferencing scene with skeleton overlaid	(b) Real teleconferencing scene with skeleton overlaid (in plan)

Figure 25. Example fittings (Robertson and Trucco 2006).

Sundaresan et al. (2007) proposed a general approach using Laplacian Eigen-maps and a graphical model of the human body to segment 3-D voxel data of humans into different articulated chains. In the bottom-up stage, the voxels are transformed into a high dimensional (6-D or less) Laplacian Eigenspace (LE) of the voxel neighborhood graph. One-dimensional splines are fitted to voxels belonging to different articulated chains such as the limbs, head and trunk. A top-down probabilistic approach is then used to register the segmented chains, utilizing both their mutual connectivity and their individual properties such as length and thickness. The approach can deal with complex poses such as those where the limbs form loops. Figure 26 shows the model used and Figure 27 illustrates the procedure of the approach.

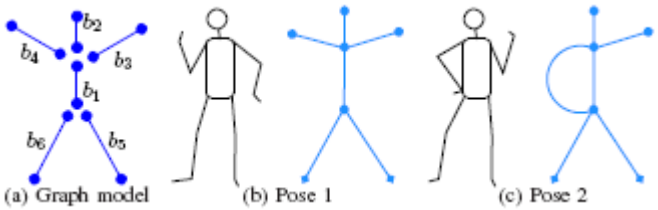


Figure 26. A model for the segmentation of 3-D voxel data (Sundaresan et al. 2007).
(a) Human body model with six articulated chains; (b) and (c) denote various poses.

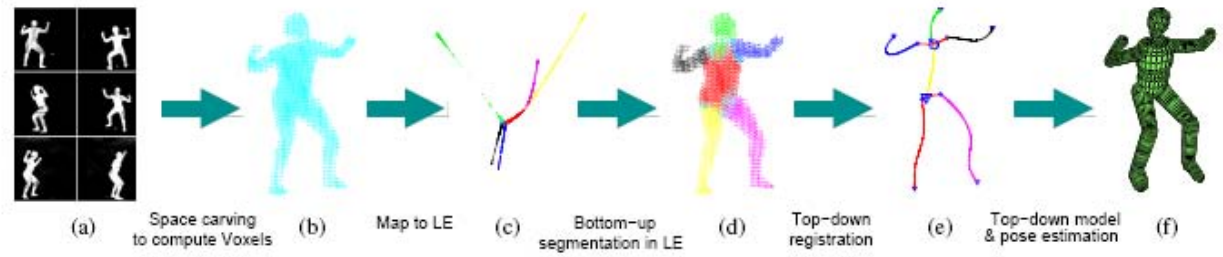


Figure 27. The steps in the segmentation in Laplacian Eigenspace (LE) to estimate human body model (Sundaresan et al. 2007).

3.3 Body Deformation Modeling

Body deformation modeling is also referred to as skinning in animation, since for 3D shape modeling only surface (skin) deformation is a concern. Two main approaches for modeling body deformations are anatomical modeling and example-based methods. The anatomical modeling is based on an accurate representation of the major bones, muscles, and other interior structures of the body. These structures are deformed as necessary when the body moves, and a skin simulation is wrapped around the underlying anatomy to obtain the final geometry of the body shape. The finite element method is the primary modeling technique used for anatomical modeling. There is a large body of work on anatomical modeling based approaches, including Wilhelms and Gelder (1997), Scheepers et al. (1997), and Aubel and Thalmann (2001).

An alternative paradigm is the example-based approach where an artist generates a model of some body part in several different poses with the same underlying mesh structure. These poses are correlated to various degrees of freedom, such as joint angles. An animator can then supply values for the degrees of freedom of a new pose and the generated body part for that new pose is interpolated appropriately. Lewis et al. (2000) and Sloan et al. (2001) developed similar techniques for applying example-based approaches to meshes. Both techniques use radial basis functions to supply the interpolation weights for each example, and, for shape interpolation, both require hand-sculpted meshes that ensure a one-to-one vertex correspondence exists between each pair of examples. Instead of using artist-generated models, recent work on the example-based modeling uses range-scan data.

Allen et al. (2002) presented an example-based method for calculating skeleton-driven body deformations. Their example data consists of range scans of a human body in a variety of poses. Using markers captured during range scanning, a kinematic skeleton is constructed first to identify the pose of each scan. Then a mutually consistent parameterization of all the scans is constructed using a posable subdivision surface template. The detail deformations are represented as displacements from this surface, and holes are filled smoothly within the displacement maps. Figure 28 illustrates how to deform the template surface to scanned surface. Figure 29 shows the body shapes for various poses made from skeletally driven subdivision surfaces. Figure 30 displays part-blending for surface generation.

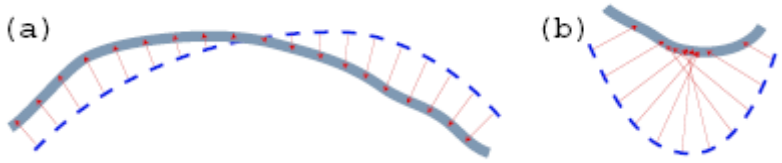


Figure 28. An example-based method for calculating body deformations (Allen et al. 2002).
(a) To construct a displaced subdivision surface, rays are cast (red arrows) perpendicular to the template subdivision surface (dashed blue line) to the nearest scanned surface (thick gray line). (b) If the template surface is too curved and the scanned surface is too far away, then the rays can cross, causing the parameterization to fold over on itself. This can be avoided by ensuring that the template surface is close to the scanned surface.

Figure 29. The body shapes for various poses made from skeletally driven subdivision surfaces (Allen et al. 2002).
Each of these 3D meshes is made from a skeletally driven subdivision surface. The displacements for the subdivision surface are interpolated from range-scan examples of the arm, shoulder, and torso in various poses. The joint angles for each pose are drawn from optical motion capture data.

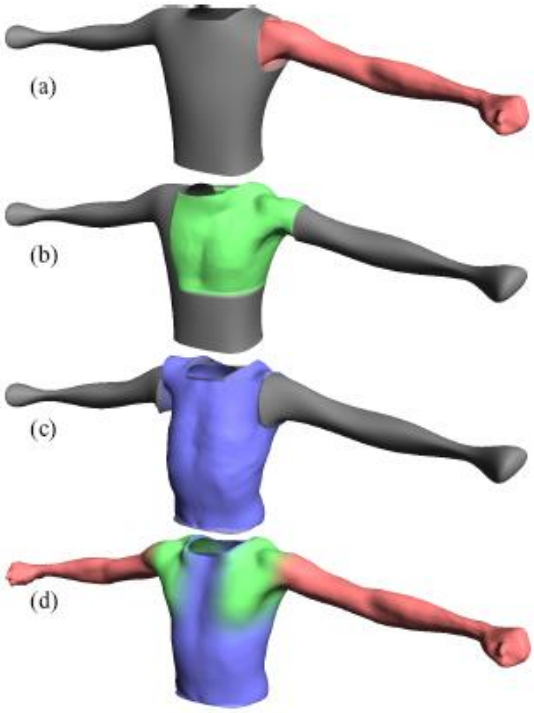


Figure 30. Blending three data sets (Allen et al. 2002).
(a) A sample arm pose. (b) A sample shoulder pose. (c) A sample torso pose. (d) Blend of arm, shoulder, and torso, with a mirrored right shoulder and right arm. Color indicates the blending weight.

Anguelov et al. (2005b) developed a method that incorporates both articulated and non-rigid deformations. A pose deformation model was constructed from training range scan data that derives the non-rigid surface deformation as a function of the pose of the articulated skeleton. A separate model of shape variation was derived from the training data also. The two models were combined to produce a 3D surface model with realistic muscle deformation for different people in different poses, when neither appears in the training set. The method (model) is referred to as the Shape Completion and Animation for People (SCAPE), a data-driven method for building a human shape model that spans variation in both subject shape and pose. Figure 31 illustrates the process to generate training set. Figure 32 shows the examples of muscle deformations of the human body that can be captured in the SCAPE pose model.

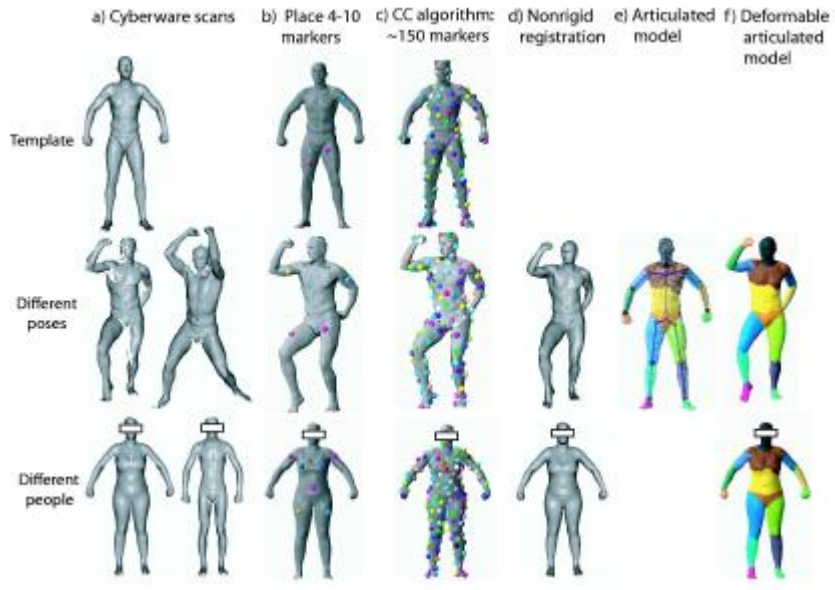


Figure 31. Mesh processing pipeline used to generate a training set (Anguelov et al. 2005b). *(a) Two data sets spanning the shape variability due to different human poses and different physiques; (b) A few markers by hand mapping the template mesh and each of the range scans; (c) Correlated Correspondence algorithm used to compute numerous additional markers; (d) The markers used as input to a non-rigid registration algorithm, producing fully registered meshes; (e) A skeleton reconstruction algorithm implemented to recover an articulated skeleton from the registered Meshes; (f) The space of deformations due to pose and physique.*

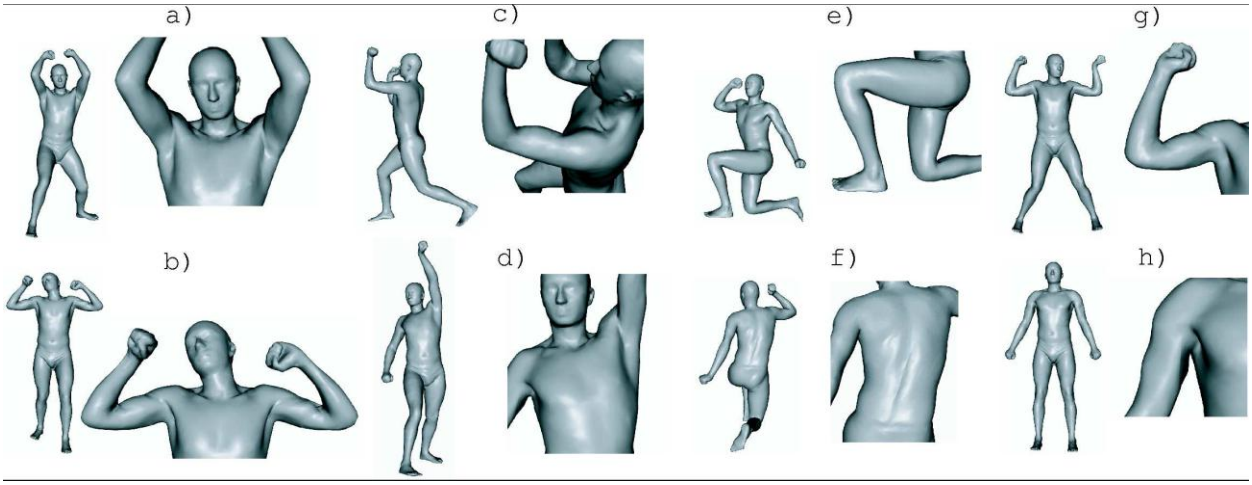


Figure 32. Examples of muscle deformations captured by the SCAPE pose model (Anguelov et al. 2005b).

Kry et al. (2002) developed a technique which allows subtle nonlinear quasi-static deformations of articulated characters to be compactly approximated by data-dependent eigen-bases which are optimized for real time rendering on commodity graphics hardware. The method extends the common Skeletal-Subspace Deformation (SSD) technique to provide efficient approximations of the complex deformation behaviors exhibited in simulated, measured, and artist-drawn

characters. Instead of using displacements for key poses (which may be numerous), principal components of the deformation influences for individual kinematic joints is pre-computed so that the optimal eigen-bases describing each joint's deformation subspace is constructed. Pose-dependent deformations are then expressed in terms of these reduced eigen-bases, allowing pre-computed coefficients of the eigen-basis to be interpolated at run time. Vertex program hardware can then efficiently render nonlinear skin deformations using a small number of eigen-displacements stored in graphics hardware. Figure 33 shows some results of this technique.

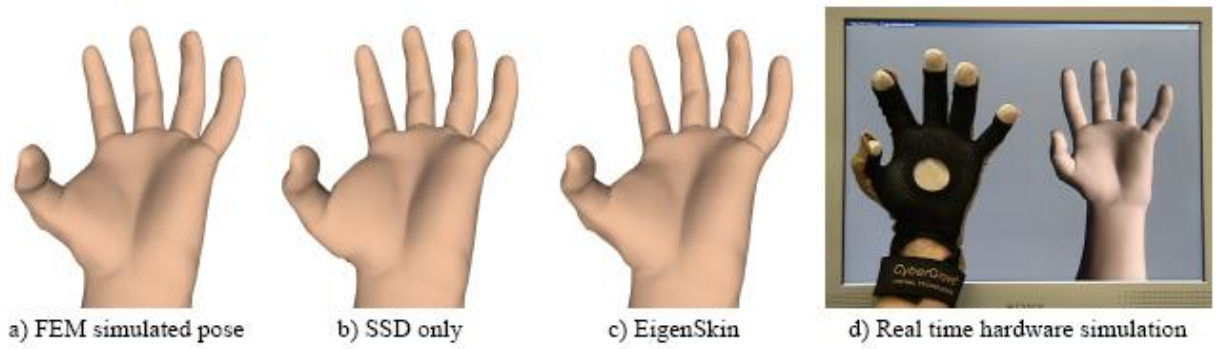


Figure 33. Comparison of EigenSkin and skeletal subspace deformation for an extreme pose not in the training data (Kry et al. 2002).

Note significant differences in the thumb between a) the new pose computed from our finite element hand model, b) skeletal-subspace deformation only, and c) EigenSkin with one eigen displacements and one normal correction per support. d) shows an EigenSkin hand example being animated using a CyberGlove. The hand model shown here consists of 55,904 triangles and is drawn using display lists with a GeForce3 vertex program.

3.4 Deformation Transfer—Pose Mapping

For pose modeling, it is impossible to acquire the pose deformation for each person at each pose. Instead, we need to transfer pose deformation from one person to another person for a given pose. Anguelov et al. (2005b) addressed this issue by integrating a pose model with a shape model reconstructed from eigen-space. As such, they were able to generate a mesh for any body shape in their PCA space in any pose. Figure 34 shows some examples of different synthesized scans, illustrating variation in both body shape and pose. The figure shows that realistic muscle deformation is achieved for very different subjects, and for a broad range of poses.

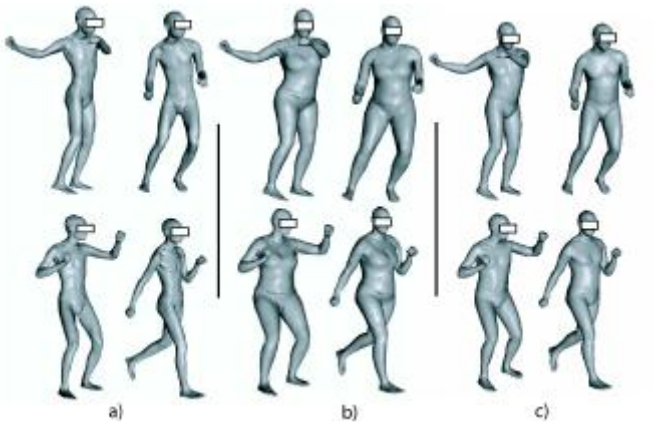


Figure 34. Deformation transfer by the SCAPE model (Anguelov et al. 2005b).
The figure shows three subjects, each in four different poses. Each subject was seen in a single reference pose only.

4.0 SHAPE MODELING OF HUMAN IN MOTION

4.1 Motion Tracking

Human motion tracking or capturing is an area that has attracted a lot of study and investigation. The acquisition of human motion data is of major importance for creating interactive virtual environments, intelligent user interfaces, and realistic computer animations. Today's performance of off-the-shelf computer hardware enables marker-free, non-intrusive optical tracking of the human body. In addition, recent research shows that it is possible to efficiently acquire and render volumetric scene representations in real-time. The following discussion presents some of the research particularly related to shape modeling.

The main challenge in articulated body motion tracking is the large number of degrees of freedom (around 30) to be recovered from motion data. Search algorithms, either deterministic or stochastic, search such a space without constraint and fall afoul of exponential computational complexity. Deutscher et al. (2000) studied general tracking without special preparation of subjects or restrictive assumptions and developed a modified particle filter for search in high dimensional configuration spaces. The new algorithm, termed annealed particle filtering, uses a continuation principle based on annealing and is shown to be capable of recovering full articulated body motion efficiently. Figure 35 and Figure 36 are the models used for the tracking, and Figure 37 is a comparison of the results.

Theobalt et al. (2004) developed a system to capture human motion at interactive frame rates without the use of markers or scene-introducing devices. Instead, 2-D computer vision and 3-D volumetric scene reconstruction algorithms are applied directly to the image data. A person is recorded by multiple synchronized cameras, and a multi-layer hierarchical kinematic skeleton is fitted into each frame in a two-stage process. The architecture of the system is shown in Figure 38.

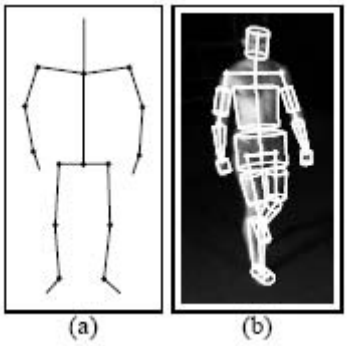


Figure 35. Model based on a kinematic chain with 17 segments (Deutscher et al. 2000).

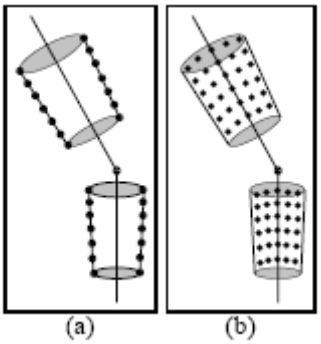


Figure 36. Configurations of the pixel map sampling points (Deutscher et al. 2000).

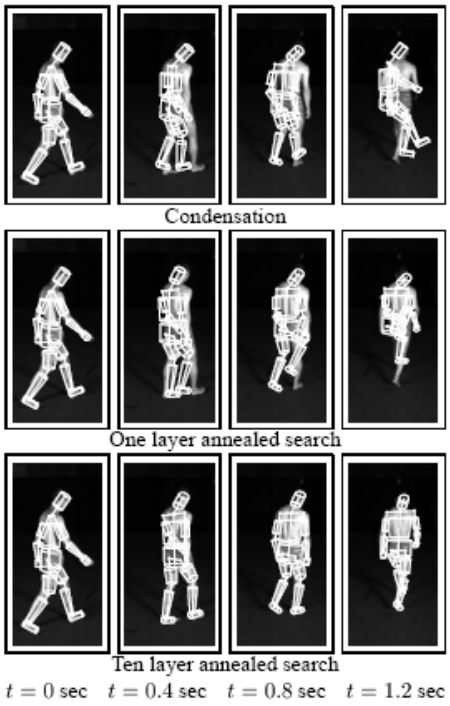


Figure 37. A comparison of Condensation with the annealed particle filter (Deutscher et al. 2000).

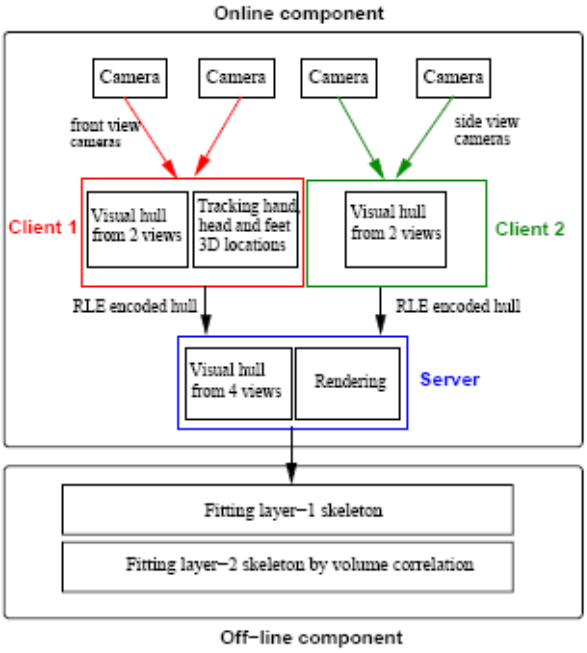


Figure 38. A system architecture for human motion capture at interactive frame rates (Theobalt et al. 2004).

4.2 Dynamic Shape Capture

During dynamic activities, the surface of the human body moves in many subtle but visually significant ways: bending, bulging, jiggling, and stretching. Park and Hodgins (2006) developed a technique for capturing and animating those motions using a commercial motion capture system and approximately 350 markers, as shown in Figure 39. Although the number of markers is significantly larger than that used in conventional motion capture, it is only a sparse representation of the true shape of the body. They supplemented this sparse sample with a detailed, actor specific surface model. The motion of the skin can then be computed by segmenting the markers into the motion of a set of rigid parts and a residual deformation (approximated first as a quadratic transformation and then with radial basis functions). The power of this approach was demonstrated by capturing existing muscles, high frequency motions, and abrupt decelerations on several actors. Figure 40 shows the capture setup.

Figure 39. Capture and animation of the dynamic motion of the human body (Park and Hodgins 2006).

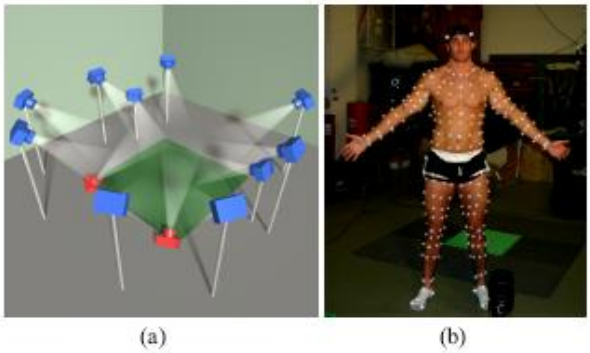


Figure 40. Capture setup (Park and Hodgins 2006).
(a) Twelve cameras surrounding a small capture region. Two cameras (shown in red) were aimed up rather than down to capture downward facing markers; (b) 350 small markers attached to the subject's body.

Sand et al. (2003) developed a method for the acquisition of deformable human geometry from silhouettes. Their technique uses a commercial tracking system to determine the motion of the skeleton and then estimates geometry for each bone using constraints provided by the silhouettes from one or more cameras. These silhouettes do not give a complete characterization of the geometry for a particular point in time, but when the subject moves, many observations of the same local geometries allow the construction of a complete model. Their reconstruction algorithm provides a simple mechanism for solving the problems of view aggregation, occlusion handling, hole filling, noise removal, and deformation modeling. The resulting model is parameterized to synthesize geometry for new poses of the skeleton. The procedure of the method is shown in Figure 41. The method uses a needle model (Figure 42). The input video included images of the subject in a wide variety of poses, as shown in Figure 43. The "goodness" of the model depends on training time and method, as shown in Figure 44.

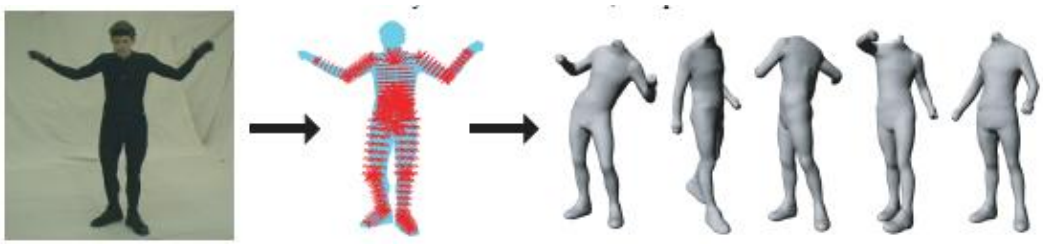


Figure 41. Extracting silhouettes from video sequences to build a deformable skin model that can be animated with new motion (Sand et al. 2003).

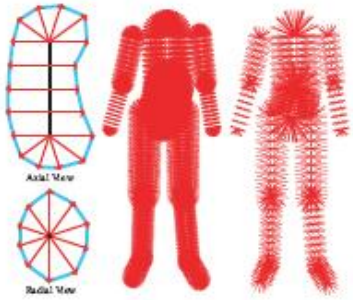


Figure 42. Deformable primitives used to describe the human body (Sand et al. 2003).

Figure 43. Input video including images of the subject in various poses (Sand et al. 2003).

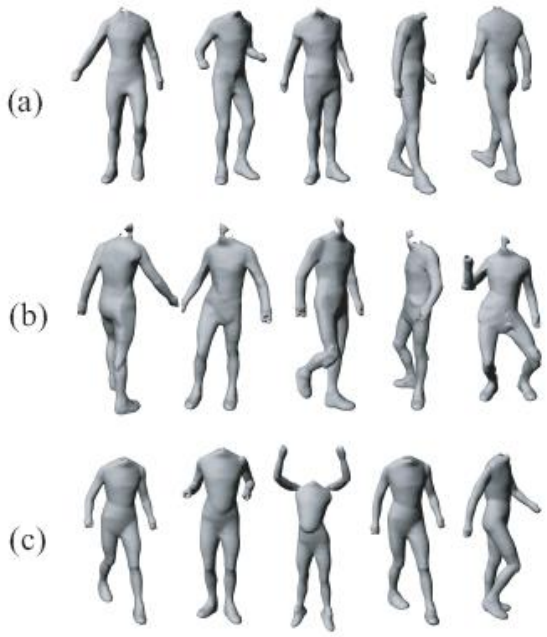


Figure 44. The models recovered from motions (Sand et al. 2003).
(a) With 3 minutes of motion observed with a single camera, a good model is obtained but its range of motion is limited. (b) With only 30 seconds of motion observed from a single camera, the model has a number of unpleasant artifacts. (c) When a model is trained without any deformation, the joints are poorly represented, illustrating that deformation is essential to an accurate human skin model.

4.3 Shape Reconstruction from Imagery Data

4.3.1 From Photos

Seo et al. (2006) presented a data-driven shape model for reconstructing human body models from one or more 2-D photos. Based on a data-driven, parameterized deformable model that is acquired from a collection of range scans of a real human body, the key idea is to complement the image-based reconstruction method by leveraging the quality, shape, and statistical information accumulated from multiple shapes of range-scanned people. In the presence of ambiguity either from the noise or missing views, the technique has a bias towards representing as much as possible the previously acquired knowledge on the shape geometry. Texture coordinates are then generated by projecting the modified deformable model acquired from range scan data onto the front view and back view images. The technique has shown to successfully reconstruct human body models from a minimum number of images, even from a single image input. Figure 45 is an overview of the approach. Figure 46 displays image projection and silhouette extraction. Figure 47 shows the distance between corresponding feature points and non-overlapping area error. Figure 48 illustrates a reconstructed model by using a single image input.

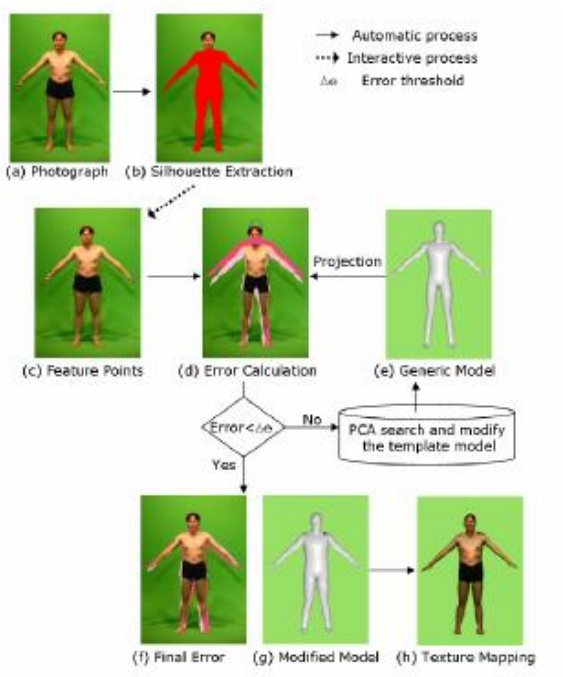


Figure 45. An approach for reconstructing human body models from 2-D photos (Seo et al. 2006).

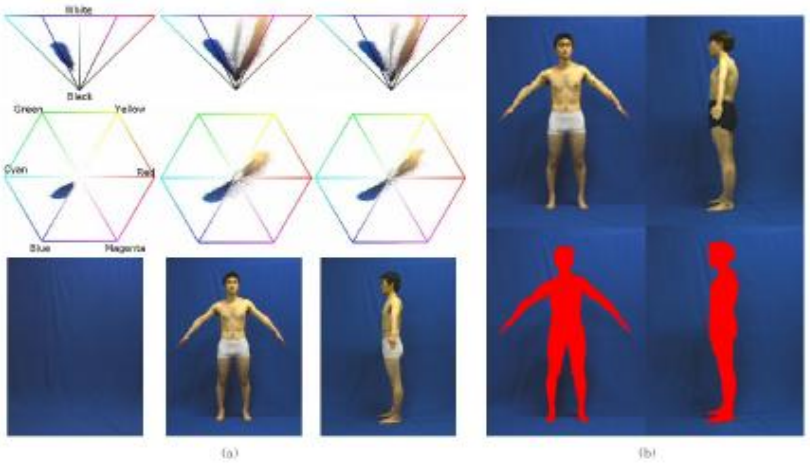


Figure 46. Image projection and silhouette extraction (Seo et al. 2006).
(a) Projection of images onto the HSV color space: Empty background (left), front (middle) and side (right) views. (b) Silhouette extraction results of two subjects with white (left) and black (right) underwears.

Figure 47. Distance between corresponding feature points (left) and non-overlapping area error calculated (right) (Seo et al. (2006).

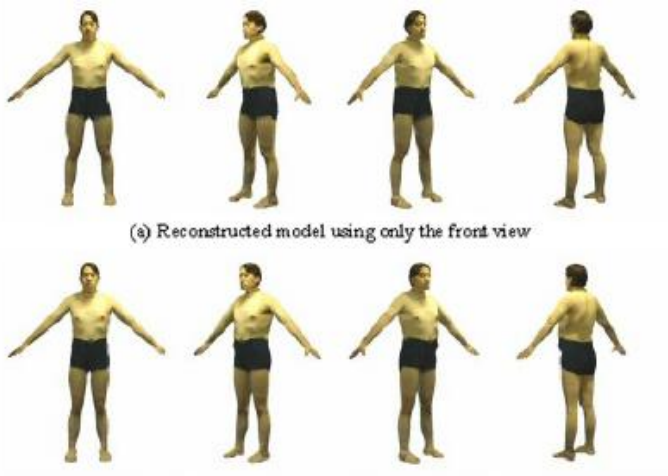


Figure 48. A reconstructed model by using a single image input (Seo et al. 2006).
(a) front image (b) side image.

4.3.2 From Video Sequences

One of the earliest studies of human motion capture from video was performed by D'Apuzzo et al. (1999), as shown in Figure 49. Given video sequences of a moving person acquired with a multi-camera system, they tracked joint locations during the movement and recovered shape information. The recovered shape and motion parameters can be used to either reconstruct the original sequence or to allow other animation models to mimic the subject's actions.

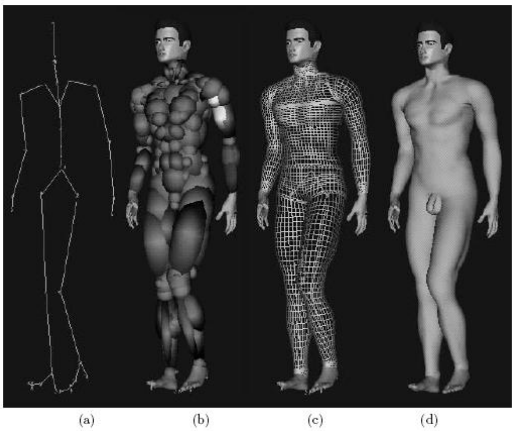


Figure 49. The layered human body model (D'Apuzzo et al. 1999).
(a) Skeleton. (b) Ellipsoidal metaballs used to simulate muscles and fat tissue. (c) Polygonal surface representation of the skin. (d) Shaded rendering.

One recent work is done by Balan (2007), as shown in Figure 50. While much of the research on video-based human motion capture assumes the body shape is known a priori and is represented coarsely (e.g. using cylinders or super quadrics to model limbs), in this paper, they proposed a method for recovering such models directly from images. Specifically, they represented the body using a recently proposed triangulated mesh model called SCAPE which employs a low-dimensional, but detailed, parametric model of shape and pose-dependent deformations that is learned from a database of range scans of human bodies. Previous work showed that the parameters of the SCAPE model could be estimated from marker-based motion capture data. Here they went further to estimate the parameters directly from image data, as shown in Figure 51 and Figure 52. They defined a cost function between image observations and a hypothesized mesh and formulated the problem as an optimization over the body shape and pose parameters using a stochastic search, as displayed in Figure 53. Their results (Figure 54 and Figure 55 showed that such rich, generative models enable the automatic recovery of detailed human shape and pose from images.

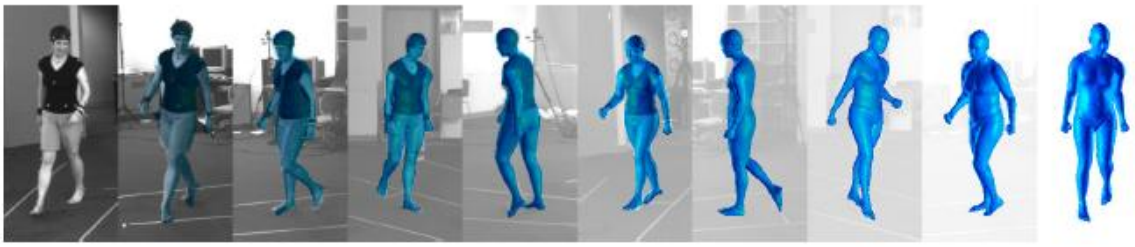


Figure 50. Detailed human shape and pose from images (Balan et al. 2007).

Figure 51. SCAPE from images (Balan et al. 2007).

Detailed 3D shape and pose of a human body is directly estimated from multi-camera image data. Several recovered poses from an image sequence of a walking subject are shown.

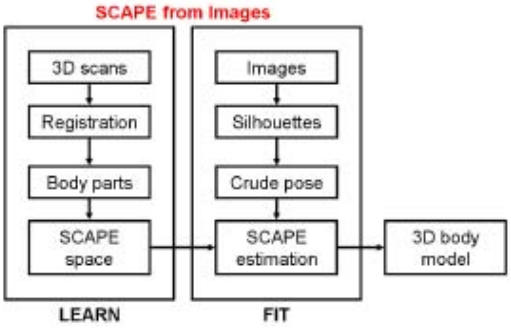


Figure 52. Algorithm Overview (Balan et al. 2007).

The learning phase is to build a 3D body model from range scans. The fitting phase is to fit the pose and shape parameters of the model to image data.

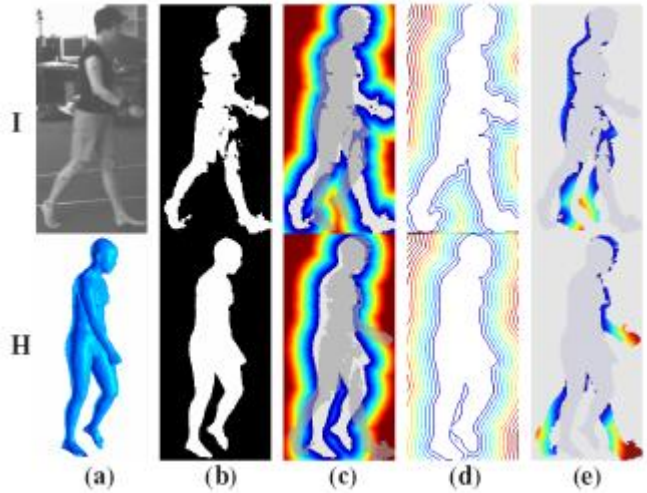


Figure 53. Cost function (Balan et al. 2007).

(a) original image I (top) and hypothesized mesh H (bottom); (b) image foreground silhouette FI and mesh silhouette FH, with 1 for foreground and 0 for background; (c) Chafer distance maps CI and CH, which are 0 inside the silhouette; the opposing silhouette is overlaid transparently; (d) contour maps for visualizing the distance maps; (e) per pixel silhouette distance from FH to FI.

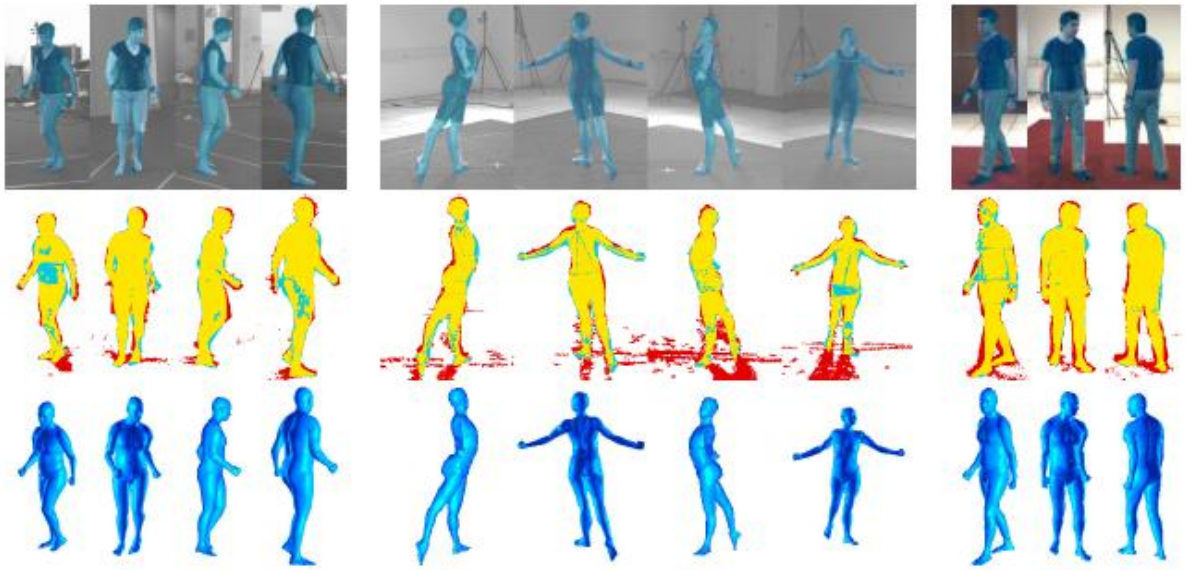


Figure 54. SCAPE-from-image results (Balan et al. 2007).

Reconstruction results based on the views shown for one male and two female subjects, in walking and ballet poses, wearing tight fitting as well as baggy clothes. Top: input images overlaid with estimated body model. Middle: overlap (yellow) between silhouette (red) and estimated model (blue). Bottom: Recovered model from each camera view.

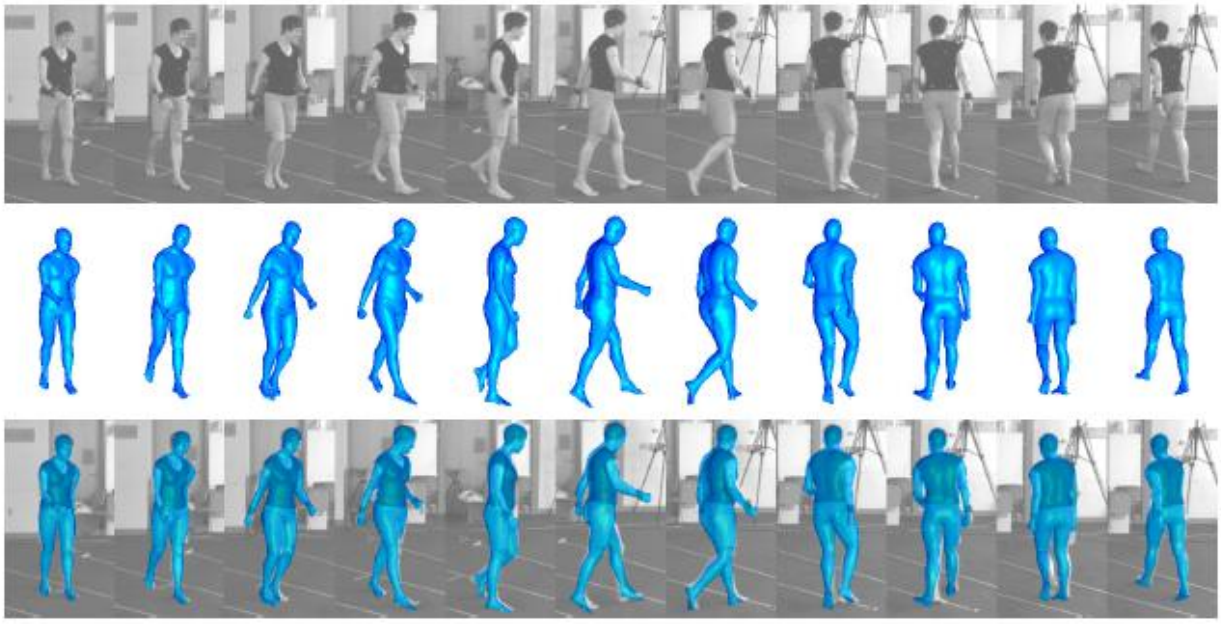


Figure 55. Automatic recovery of detailed human shape and pose from images (Balan et al. 2007).

First row: Input images. Second row: Estimated mesh models. Third row: Meshes overlaid over input images. By applying the shape parameters recovered from 33 frames to the template mesh placed in a canonical pose, we obtained a shape deviation per vertex of 8.8 ± 5.3mm, computed as the mean deviation from the average location of each surface vertex.

4.4 Animation

Realistic human representation and animation remains a primary goal of computer graphics research. The animation of the subject can be realized by displaying a series of human shape models for a prescribed sequence of poses. Hilton et al. (2002) built a framework for construction of animated models from the captured surface shape of real objects. Algorithms were developed to transform the captured surface shape into a layered model. The layered model is composed of an articulation structure, a generic control model and a displacement map to represent the high-resolution surface detail. Novel methods were presented for automatic control model generation, shape constrained fitting and displacement mapping of the captured data. The framework enables rapid transformation of captured data into a structured representation suitable for realistic animation. An example animated model from captured 3-D surface measurements is shown in Figure 56. The pipeline for reconstructing animated models from captured 3-D surface data is illustrated in Figure 57. Figure 58 displays an animation of a Cyberware whole body scan.

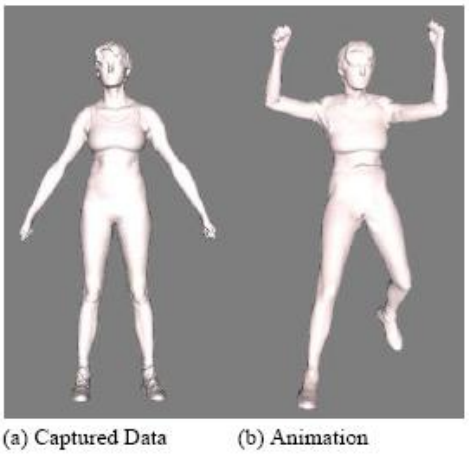


Figure 56. Example animated model from captured 3D surface measurements (Hilton et al. 2002).

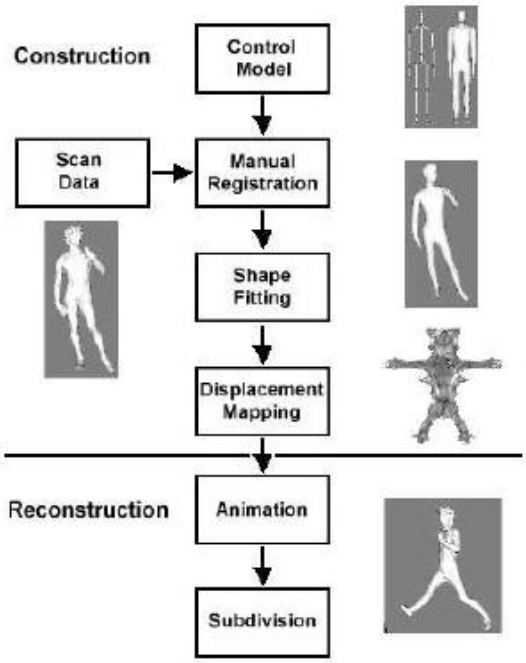


Figure 57. Functional Models Pipeline for Animating Michelangelo's David from captured data (Hilton et al. 2002).

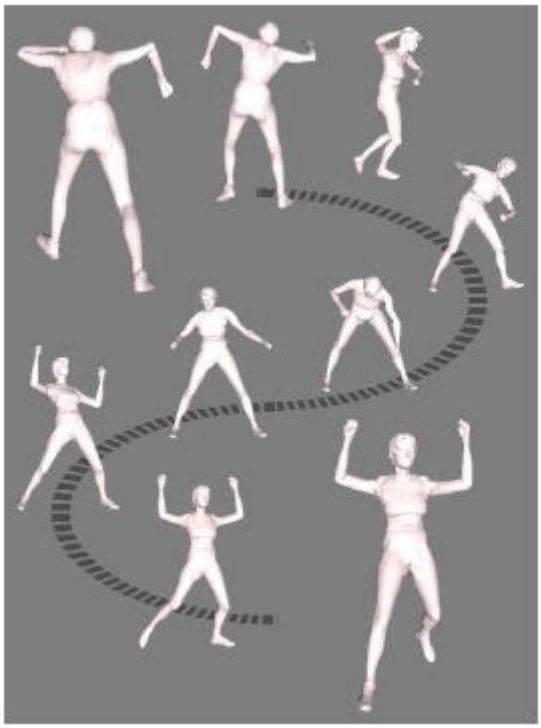


Figure 58. Animation of a Cyberware whole body scan (Hilton et al. 2002).
The model was reconstructed using a generic humanoid model of approximately 2500 polygons.

Seo et al. (2003) developed a synthesizer for manipulations of body models by using parameters such as fat percentage and hip-to-waist ratio. On any synthesized model, the underlying bone and skin structure is properly adjusted, so that the model remains completely animatable using the underlying skeleton. Figure 59 illustrates the motion captured animation applied to their models.

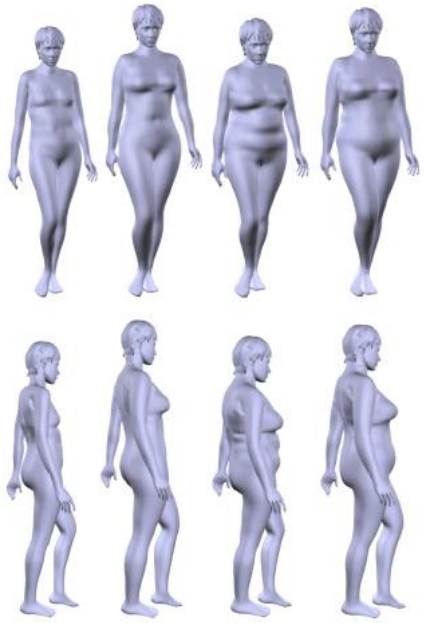


Figure 59. Motion captured animation applied to four of our models (Seo et al. 2003).

Aguiar et al. (2006) developed a novel versatile, fast and simple framework to generate high quality animations of scanned human characters from input motion data, as shown in Figure 60. The method is purely mesh-based and, in contrast to skeleton-based animation, requires only a minimum of manual interaction. The only manual step that is required to create moving virtual people is the placement of a sparse set of correspondences between triangles of an input mesh and triangles of the mesh to be animated. The proposed algorithm, as illustrated in Figure 61, implicitly generates realistic body deformations, and can easily transfer motions between human subjects of completely different shape and proportions. The approach handles many different types of input data, e.g. other animated meshes and motion capture files, in just the same way. Finally, and most importantly, it creates animations at interactive frame rates. Figure 62 displays the pose deformation transfer. Figure 63 shows how the acquired motion is realistically transferred to the final human body scan.

Figure 60. Confluent marker-based animation (Aguiar et al. 2006).
Subsequent frames showing the female scan authentically performing a soccer kick. Motion data have been acquired by means of a marker-based motion capture system. Note the realistic protrusion of the chest when she blocks the ball, as well as the original head motion.

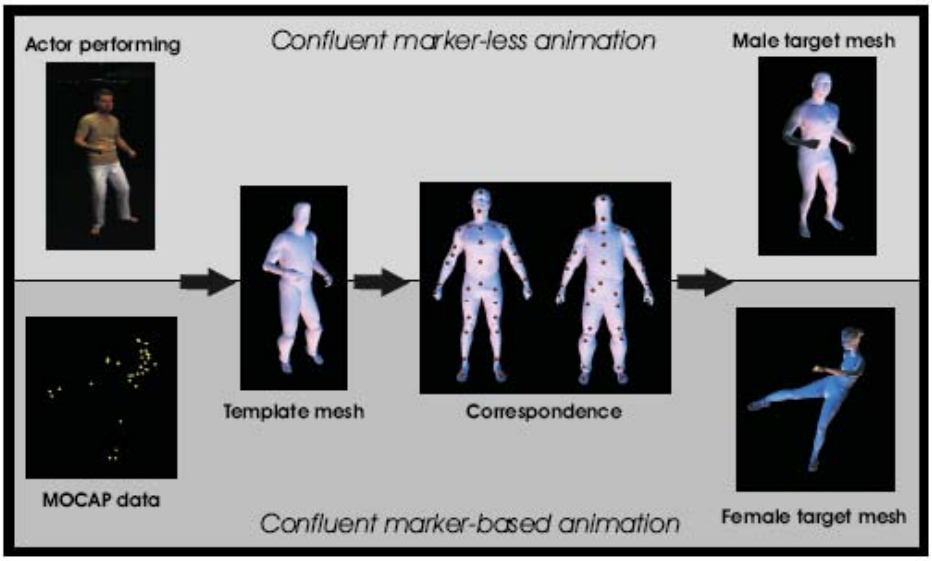


Figure 61. Illustration of the Conuent Motion pipeline (Aguiar et al. 2006).

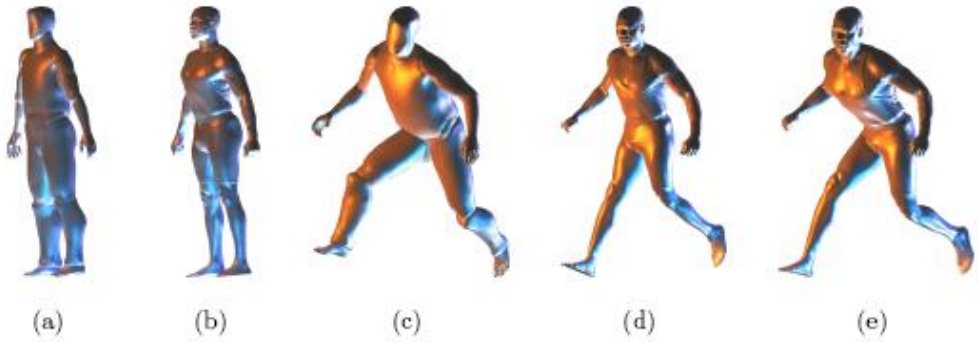


Figure 62. A template model (Aguiar et al. 2006).
(a) A high-resolution body scan (b) In their respective reference poses. Inuence of the number of markers on the quality of the deformation: the template in a pose obtained via motion capture (c). While 22 triangle correspondences already suffice to transfer this pose in good quality to the scan (d), 180 triangle correspondences reproduce even subtle details (e).

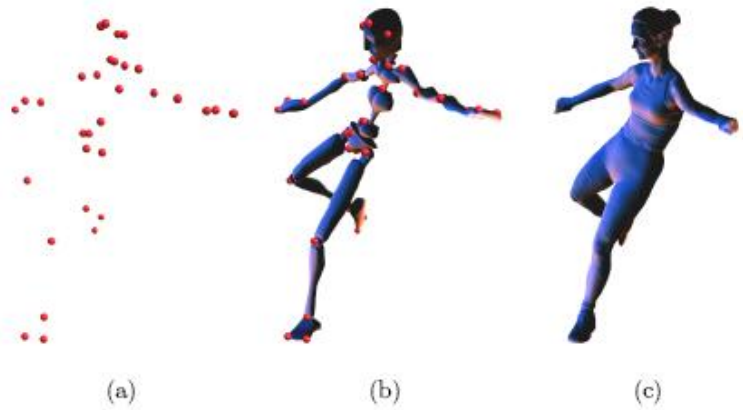


Figure 63. Input markers (Aguiar et al. 2006).
(a) are used to generate an intermediate biped model (b). By applying our deformation technique the acquired motion is realistically transfered to the final human body scan (c).

4.5 Video-driven Animation

Video-driven animation animates a subject using the motion of the same or a different subject captured from video. In Aguiar et al. (2006) (as shown in Figure 64), the motion parameters were extracted from raw video footage of human performances (top row). By this means, body poses of a video-taped individual can easily be mapped to body scans of other human subjects (second and third row). Note that skin deformations are naturally modeled (middle column). Scans are faithfully animated regardless of the differences in body shape and skeletal dimensions.

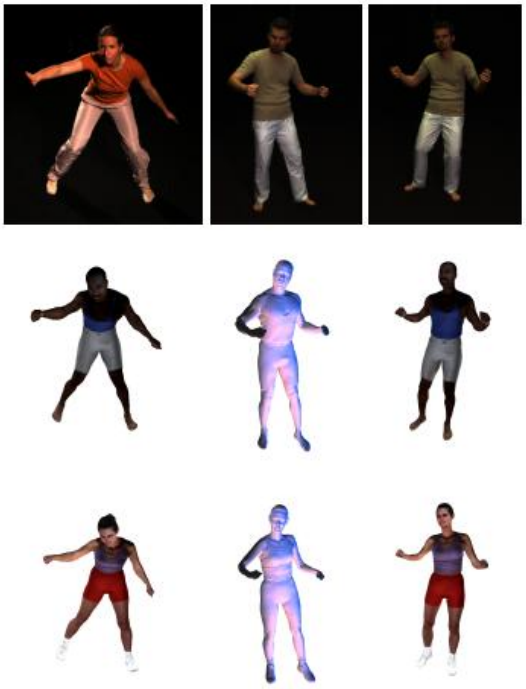


Figure 64. Video-driven animation (Aguiar et al. 2006).

4.6 Inverse Kinematics

Inverse kinematics (IK), the process of computing the pose of a human body from a set of constraints, is widely used in computer animation. However, the problem is inherently underdetermined. For example, for given positions of the hands and feet of a character or an animated figure, there are many possible character poses that satisfy the constraints. Even though many poses are possible, some poses are more likely than others — an actor asked to reach forward with his arm will most likely reach with his whole body, rather than keeping the rest of the body limp. In general, the likelihood of poses depends on the body shape and style of the individual person, and designing this likelihood function by hand for every person would be a difficult or impossible task. Grochow et al. (2004) developed an inverse kinematics system (as shown in Figure 65) based on a learned model of human poses that can produce the most likely pose satisfying those constraints in real time. Training the model on different input data leads to different styles of IK. The model is represented as a probability distribution over the space of all possible poses. This means that the model can generate any pose, but prefers poses that are most similar to the space of poses in the training data. Figure below illustrates the process of modeling learning. Figure 66 shows an example of creating a motion by key framing using three key framed markers.

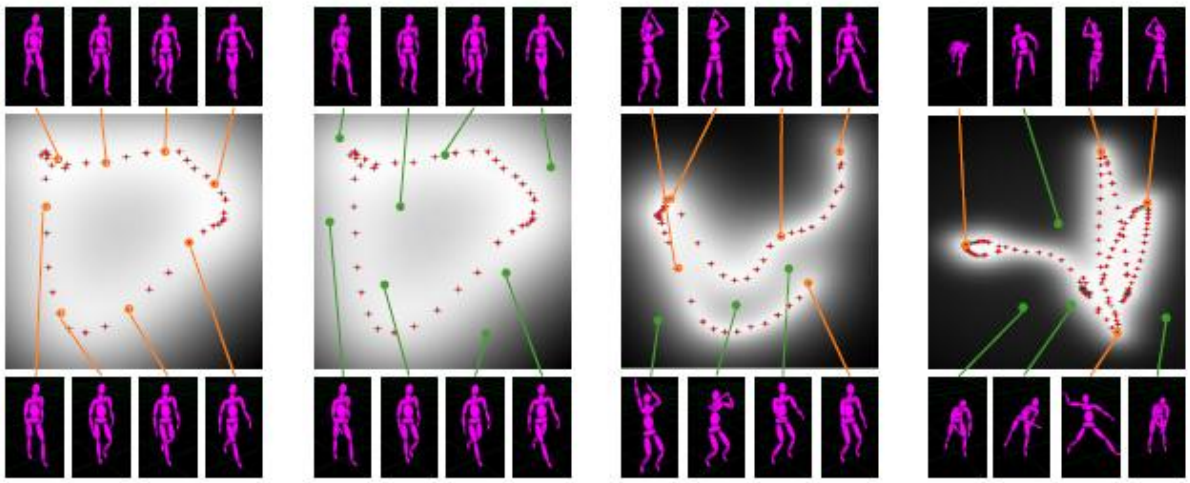


Figure 65. Latent spaces learned from different motion capture sequences (Grochow et al. 2004).

The motion sequences include a walk cycle, a jump shot, and a baseball pitch. Points: The learning process estimates a 2D position associated with every training pose. Plus signs (+) indicate positions of the original training points in the 2D space. Red points indicate training poses included in the training set. Poses: Some of the original poses are shown along with the plots, connected to their 2D positions by orange lines. Additionally, some novel poses are shown, connected by green lines to their positions in the 2D plot. Note that the new poses extrapolate from the original poses in a sensible way, and that the original poses have been arranged so that similar poses are nearby in the 2D space.

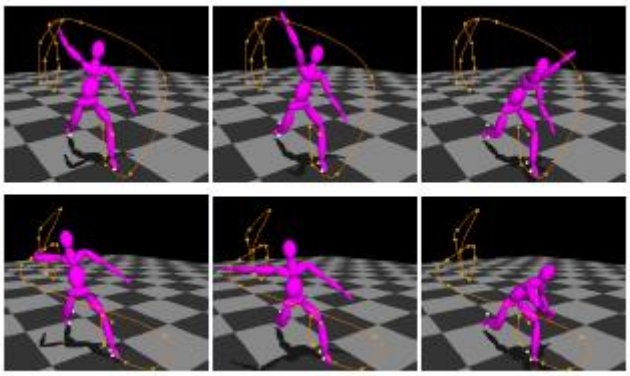


Figure 66. Trajectory key framing using a style learned from the baseball pitch data (Grochow et al. 2004).

Top row: A baseball pitch. Bottom row: A side-arm pitch. In each case, the feet and one arm were key framed; no other constraints were used. The side-arm contains poses very different from those in the original data.

5.0 CONCLUDING REMARKS

Human shape modeling spans various research areas from anthropometry, computer graphics and computer vision to machine intelligence and optimization. In addition to traditional uses, human modeling is finding many applications in some new areas, such as virtual environment, human identification, and human-borne threat detection. These new applications pose great challenges to human modeling technology. With the advancement of human modeling technology, creating a realistic, dynamic human model that can be rendered in real-time or nearly real-time is achievable**.**

REFERENCES

1. Aguiar, E., Zayer, R., Theobalt, C., Magnor, M. ,and Seidel, H.P. (2006). "A Framework for Natural Animation of Digitized Models," in MPI-I-2006-4-003 July 2006.
2. Allen, B., Curless, B., and Popovic, Z. (2002). "Articulated Body Deformation from Range Scan Data," in ACM SIGGRAPH 2002, 21-26 July 2002, San Antonio, TX, USA.
3. Allen, B., Curless, B., and Popvic, Z. (2003). "The space of human body shapes: reconstruction and parameterization from range scans," in ACM SIGGRAPH 2003, 27-31 July 2003, San Diego, CA, USA.
4. Anguelov, D., Koller, D., Pang, H.C., Srinivasan, P., and Thrun, S. (2004). "Recovering Articulated Object Models from 3D Range Data," in Proceedings of the 20th conference on Uncertainty in artificial intelligence, Pages: 18 – 26, 2004.
5. Anguelov, D., Srinivasan, P., Pang, H.C., Koller, D., Thrun, S., and Davis, J. (2005a). "The correlated correspondence algorithm for unsupervised registration of nonrigid surfaces," Advances in Neural Information Processing Systems 17, 33-40.
6. Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J. and Davis, J. (2005b). "SCAPE: Shape Completion and Animation of People." ACM Transactions on Graphics (SIGGRAPH) 24(3).
7. Apuzzo, N. D', Plankers, R., Fua, P., Gruen, A., and Thalmann, D. (1999). "Modeling Human Bodies from Video Sequences," in SPIE proceedings series Videometrics VI , San Jose CA, 28-29 January 1999.
8. Aubel, A., and Thalmann, D. (2001). "Interactive modeling of the human musculature," in Proc. of Computer Animation 2001.
9. Balan, A., Sigal, L., Black, M. , Davis, J., and Haussecker, H. (2007) . "Detailed Human Shape and Pose from Images," IEEE Conf. on Comp. Vision and Pattern Recognition (CVPR), 2007.
10. Ben Azouz, Z. B., Shu, C., and Mantel, A. (2006). "Automatic Locating of Anthropometrics Landmarks on 3D Human models," in Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06).
11. Ben Azouz, B.Z., Rioux, M., Shu, C., and Lepage, R. (2004). "Analysis of Human Shape Variation using Volumetric Techniques," in Proc. of 17th Annual Conference on Computer Animation and Social Agents Geneva, Switzerland. July 7-9, 2004.
12. Ben Azouz, B.Z., Rioux, M., Shu, C., and Lepage, R. (2005a). "Characterizing Human Shape Variation Using 3-D Anthropometric Data," International Journal of Computer Graphics, volume 22, number 5, pp. 302-314, 2005.
13. Ben Azouz, Z. B., Shu, C., Lepage, R., and Rioux, M. (2005b). "Extracting Main Modes of Human Body Shape Variation from 3-D Anthropometric Data," in Proceedings of the Fifth International Conference on 3-D Digital Imaging and Modeling (3DIM'05).
14. Cohen, I. and Li, H.X. (2003). "Inference of Human Postures by Classification of 3D Human Body Shape," in Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG'03).
15. Curless, B., AND Levoy, M. (1996). "A volumetric method of building complex models from range images," in Proceedings of SIGGRAPH 1996, 303–312.

16. Davatzikos, C., Tao, X.D., and Shen, D.G. (2003). “Hierarchical Active Shape Models Using the Wavelet Transform,” IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 22, NO.3, March 2003.
17. Davis, J., Marschner, S., Garr, M., and Levoy, M.(2002). “Filling holes in complex surfaces using volumetric diffusion,” in Proceedings of the First International Symposium on 3D Data Processing, Visualization and Transmission. Padua, Italy (2002).
18. Deutscher, J., Blake, A., and Reid, I. (2000). “Articulated Body Motion Captured by Annealed Particle Filtering,” in Proceedings of 2000 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'00), Volume 2, 2000.
19. Godil, A., Grother, P., and Ressler, S. (2003). “Human Identification from Body Shape,” in Proceedings of the Fourth International Conference on 3-D Digital Imaging and Modeling (3DIM’03).
20. Grochow, K., Martin, S.L., Hertzmann, A., and Popovic, Z. (2004). “Style-Based Inverse Kinematics,” ACM Trans. on Graphics (Proc. SIGGRAPH’04).
21. Guti´errez, M., Thalmann, D.,Vexo, F., Moccozet, L., Thalmann, N. M., Mortara, M., and Spagnuolo, M. (2005). “An Ontology of Virtual Humans: incorporating semantics into human shapes,” in Proc. Workshop towards Semantic Virtual Environments (SVE05), 2005, p. 56-57.
22. Hilton, A., Starck, J., and Collins, G. (2002). “ From 3D Shape Capture to Animated Models,” in Proceedings of First International Symposium on 3D Data Processing Visualization and Transmission, 2002 pp 246- 255.
23. Kry, P.G., James, D.L., and Pai, D.K. (2002). “EigenSkin: Real Time Large Deformation Character Skinning in Hardware,” In ACM SIGGRAPH Symposium on Computer Animation.
24. Lewis, J.P., Cordner, M., and Fong, N. (2000). “Pose space deformations: A unified approach to shape interpolation and skeleton-driven deformation,” in Proceedings of ACM SIGGRAPH 2000, ACM Press / ACM SIGGRAPH / Addison Wesley Longman, K. Akeley, Ed., Annual Conference Series, ACM, 165–172.
25. Liepa, P. (2003). “Filling holes in meshes,” in Proc. of the Eurographics/ACM SIGGRAPH symposium on Geometry processing, 200–205.
26. Maad, S. and Bouakaz, S. (2004). “Geometric Modelling of Moving Virtual Humans: A Survey of Various Practices,” Http://www710.univ-lyon1.fr/~smaad/CYBERII/.
27. Mittal, A., Zhao, L., and Davis, L. S. (2003). “Human Body Pose Estimation Using Silhouette Shape Analysis,” in Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS’03).
28. Paquet, E. and Rioux, M. (1998). “Content-based access of VRML libraries,” IAPR-International Workshop on Multimedia Information Analysis and Retrieval, in Lecture Notes in Computer.
29. Park, S.I. and Hodgins, J.K. (2006). “Capturing and Animating Skin Deformation in Human Motion,” ACM Transaction on Graphics (SIGGRAPH 2006), 25(3), pp 881-889,July 2006.
30. Rhee, T., Lewis, J.P., and Neumann, U. (2006). “Real-TimeWeighted Pose-Space Deformation on the GPU,” EUROGRAPHICS 2006, Vol. 25 No. 3.
31. Robertson, C. and Trucco, E. (2006). “Human body posture via hierarchical evolutionary optimization,” in BMVC06, 2006.

32. Robinette, K., Daanen, H., and Paquet, E. (1999). "The Caesar Project: A 3-D Surface Anthropometry Survey," in Second International Conference on 3-D Digital Imaging and Modeling (3DIM'99), pages 380–386, Ottawa, Canada, October 1999.
33. Robinette, K.M. (2003). "An Investigation of 3-D Anthropometric Shape Descriptors for Database Mining," Ph.D. Thesis, University of Cincinnati, 2003.
34. Robinette, K.M., Vannier, M.W., Rioux, M. and Jones, P. (1997). "3-D surface anthropometry: Review of technologies," in Neuilly sur-Seine: North Atlantic Treaty Organization Advisory Group for Aerospace Rearch & Development, Aerospace Medical Panel, 1997.
35. Sand, P., McMillan, L., and Popovic, J. (2003). "Continuous Capture of Skin Deformation," ACM Transactions on Graphics (TOG) 22, 3, 578-586.
36. Scheepers, F., Parent, R.E., Carlson, W.E., and May, S F. (1997). "Anatomy based modeling of the human musculature," in Proceedings of ACM SIGGRAPH 97, T. Whitted, Ed., Annual Conference Series, ACM, 163–172. Sciences-Springer, Vol. 1464, pp. 20-32, Springer, Hong Kong, China.
37. Seo, H., Cordier, F., Thalmann, N.M. (2003), "Synthesizing Animatable Body Models with Parameterized Shape Modifications," Eurographics/SIGGRAPH Symposium on Computer Animation (2003).
38. Seo, H., Yeo, Y.I., and Wohn, K. (2006). "3D Body Reconstruction from Photos Based on Range Scan," LECTURE NOTES IN COMPUTER SCIENCE, 2006, No. 3942, pages 849-860, SPRINGER-VERLAG.
39. Sloan, P.P., Rose, C., and Cohen, M. F. (2001). "Shape by example," in Proceedings of 2001 Symposium on Interactive 3D Graphics.
40. Sundaresan, A. and Chellappa, R. (2007). "Model driven segmentation of articulating humans in Laplacian Eigenspace," IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 2007.
41. Thalmann, N. M., Seo, H., and Cordier, F. (2004). "Automatic Modeling of Virtual Humans and Body Clothing," Journal of Computer Science and Technology, Volume 19 , Issue 5 , September 2004.
42. Theobalt, C., Magnor, M., Sch¨uler, P., and Seidel, H.P. (2004). "Combining 2D Feature Tracking and Volume Reconstructions for Online Video-Based Human Motion Capture," International Journal of Image and Graphics, Vol. 4 No. 4, 563-583, 2004.
43. Wilhelms, J., and Gelder, A.V. (1997). "Anatomically based modeling," in Proceedings of ACM SIGGRAPH 97, T. Whitted, Ed., Annual Conference Series, ACM, 173–180.
44. Yang, X.S. and Zhang, J.J. (2005). "Realistic Skeleton Driven Skin Deformation," in Proc. ICCSA 2005, LNCS 3482, pp. 1109–1118, 2005.