Color 3D Digital Human Modeling and Its Applications to Animation and Anthropometry

Bao-zhen Ge[1], Qing-guo Tian[1], K. David Young[2], and Yu-chen Sun[1]

[1] College of Precision Instruments and Opto-electronics Engineering and
Key Ministry of Education Laboratory of
Opto-electronics Information and Technical Science
Tianjin University Tianjin 300072. China
gebz@tju.edu.cn
[2] Embedded Systems Institute and Department of Electronic and Computer Engineering
The Hong Kong University of Science and Technology
Clear Water Bay, Hong Kong
kkdyoung@ust.hk

Abstract. With the rapid advancement in laser technology, computer vision, and embedded computing, the application of laser scanning to the digitization of three dimensional physical realities has become increasingly widespread. In this paper, we focus on research results embodied in a 3D human body color digitization system developed at Tianjin University, and in collaboration with the Hong Kong University of Science and Technology. In digital human modeling, the first step involves the acquisition of the 3D human body data. We have over the years developed laser scanning technological know-how from first principles to support our research activities on building the first 3D digital human database for ethnic Chinese. The disadvantage of the conventional laser scanning is that surface color information is not contained in the point cloud data. By adding color imaging sensors to the developed multi-axis laser scanner, both the 3D human body coordinate data and the body surface color mapping are acquired. Our latest development is focused on skeleton extraction which is the key step towards human body animation, and applications to dynamic anthropometry. For dynamic anthropometric measurements, we first use an animation algorithm to adjust the 3D digital human to the required standard posture for measurement, and then fix the feature points and feature planes based on human body geometric characteristics. Utilizing the feature points, feature planes, and the extracted human body skeleton, we have measured 40 key sizes for the stand posture, and the squat posture. These experimental results will be given, and the factors that affect the measurement precision are analyzed through qualitative and quantitative analyses.

1 Introduction

Three dimensional (3D) digitization encompasses the applications of 3D information extraction methods to extract the 3D surface data of an object, real time 3D data

V.G. Duffy (Ed.): Digital Human Modeling, HCII 2007, LNCS 4561, pp. 82–91, 2007.

pre-processing and generation of visual 3D computer graphic models of the object. The 3D information extraction methods are subdivided into two classes: contact and non-contact. A typical representative of the contact type is the coordinate measurement machine (CMM) [1] whose precision can be of microns, however the measurement speed is rather slow and furthermore inapplicable to soft objects. Non-contact methods primarily apply optical techniques which can be further subdivided into passive and active types. Passive optical measurement methods include monocular imaging and binocular stereo imaging approaches [2]. Active optical measurement methods include: (1) phase measurement approaches such as phase profilometry [3], Moire contouring [4], color-coded projection grating method [5] whose main advantage is fast extraction speed, but suffer from lower resolution, and greater limitation of the measurement volume by the optical aperture; (2) laser ranging techniques which are based on the principle of direct measurement of the time of flight difference between a target optical beam and a reference beam, and its conversion into distance which yields resolution typically about 1 mm [6]; (3) laser scanning measurement techniques which employ structured laser light source to illuminate the target objects and detect the reflected light distorted by the object's surface with CCD arrays, triangulation principle is applied to determine the 3D point locations in a reference coordinate system. Different methods have been developed depending on the types of structured laser light sources: point source structured light method [7], line source structured light method [8], and double line source structured light method [9]. Optical methods are widely use in many areas due to their inherit advantages of allowing non-contact, high resolution, and high speed measurements. In particular, structured light triangulated measurement based 3D color digitalization has become widely adopted, The approach taken by Yu [10] and the work at HKUST [11] both assume the target object which is a human body is stationary, and the structured laser line and sensing CCD arrays move vertically to obtain the 3D point cloud data of the human; color information however was not collected. Xu [12] developed 3D color measurement modeling method from monochrome 3D measurement theory, and applied to obtain 3D color information for small size objects. Hu [13] utilized color CCD arrays and structured laser line source, and with a constant speed turning platform and high speed optoelectronic shutters, obtained the target object's 3D point cloud and the corresponding color information. Pulli [14] deployed four NTSC format color CCD cameras and a white light projector affixed to a turning platform to form a scanning system.

This paper is divided into two halves: the first half introduces a color 3D human digitalization system which is based on the use of structured laser line source and multi-axis synchronized scanning. Basic principle of operation of the system will be given, with more in depth discussions on the method used for the calibration of the 3D sensors and color sensors. The second half covers the human surface mesh modelling and processing aspects, which includes skeleton extraction algorithms based on radial basis functions and steepest descent, and an animation computation algorithm which is based an improved skinning method. Finally, the developed animation algorithm is applied to human body size measurements for anthropometry and garment design.

2 Color 3D Human Digitization System Design

2.1 System Structure

After a thorough examination of the advantages and disadvantages of different approaches in laser scanning based 3D color digitalization, we present in this paper an approach which exploits independent yet correlated 3D sensor and color sensor development in order to improve the system flexibility. The system is pictorially represented in Figure 1. The laser scanning support structure is consists of a rectangular base on which four vertical columns are attached; on each of the vertical columns, a 3D sensor is mounted. Interleaving with the 3D sensor columns are four additional vertical supports on which four color sensors are mounted. The maximum volume covered in one complete scan is cylindrical: 1000mm diameter × 2000 in height. On the surface of a cylinder whose center is at the center of the rectangular loading base and at a radius of 500mm, the typical resolution is 1mm in the horizontal direction, 2mm in the vertical direction. One complete scan takes 16.7 seconds at a vertical spacing of 2mm.

The 3D sensor applies the principle of triangulation and utilizes a laser line structured light source. In order to minimize the obstruction of light by the target object, a dual CCD camera structure is employed to receive the reflected light beam which is modulated by the target object being measured. The single axis 3D sensor structure is schematically represented as Figure 2. On the vertical column, a mechanical translation platform which is constructed from the combination of a servo motor drive and a ball screw and slide is responsible for moving the sensor in the up down direction. On the moving platform, two gray scale CCD cameras are mounted at identical inclined

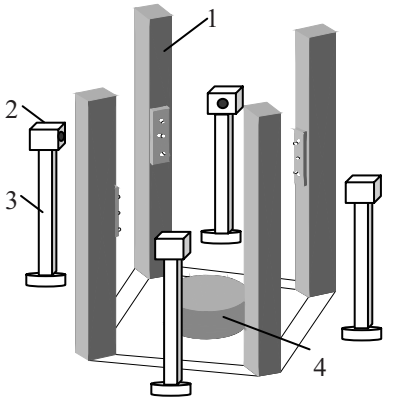

Fig. 1. 3D Laser Scanner based Color Digitalization System Schematic Diagram: 1. 3D Sensors 2. Color CCD Camera 3. 3D Camera Support 4. Load target platform.

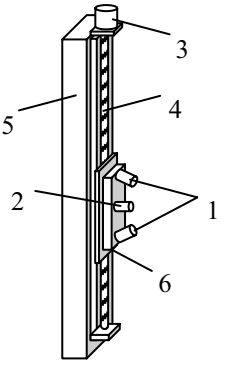

Fig. 2. Single axis 3D Sensor schematic diagram: 1 Gray Scale CCD Camera 2 Laser Line source 3 Servomotor 4 Ball Screw 5 Support Column 6 Sensor base

angles and equal distance from the centrally mounted laser line source. The key issue in this 3D color scanning system is calibration.

2.2 3D Sensor Calibration

The calibration of the 3D sensor is basically the calibration of the gray scale cameras. Yuan [15] applied a perspective transformation matrix method to calibrate the cameras, however he ignored nonlinear effects such as lens distortions. Although this method attains faster calibration speed, accuracy is lower. Faig [16] applied nonlinear optimization method, which took into account a variety of parameters to calibrate the camera system, and achieved higher accuracy, however the calibration outcome depends on the choice of initial values, making the calibration process rather unreliable. Reimar and Roger [17] combined the perspective transformation matrix approach with nonlinear optimization to perform calibration, although it can achieve high accuracy, this method requires solving for system parameters from nonlinear equations, it is also very time consuming. For comprise between speed and accuracy, we propose an approach which is based on linearized transformation matrix for partitioned regions to calibrate our 3D sensors.

Using a perspective transformation model, a point in space $P(X_w,Y_w,Z_w)$ a mapped onto the corresponding point $P_f(X_f,Y_f)$ on the pixel map through rotation and resizing transformations. Since Z_w within the same light band of an object is identical, this relationship can be represented using the following matrix equation:

$$\begin{bmatrix} X_f \\ Y_f \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \bullet \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} = M \bullet \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} \tag{1}$$

The purpose of calibration is to determine the values of the elements in the matrix M so that the correspondence between the point (X_f,Y_f) and (X_w,Y_w) can be defined. If we let $m_{33} = 1$, then theoretically, taking four calibration point pairs will be sufficient to determine the remaining eight unknown matrix coefficients. However, in order to obtain better accuracies, we sample more calibration point pairs. After we compute the matrix M, the 3D coordinates of the corresponding points on the object can be determined from the pixel map.

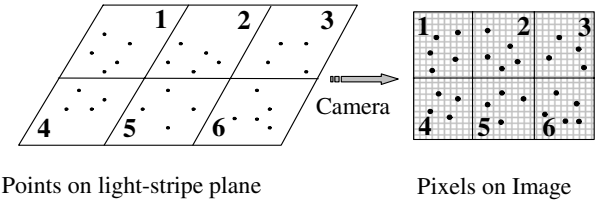

Fig. 3. Partitioned regions for calibration

In order to use linear transformation method while taking in account of lens distortion and other nonlineareffects, we adopt a partitioned region approachas indicated in

Figure 3 where points on the light stripe plane and the corresponding pixels on the imaging plane are subdivided into different zones. For each of the different zones, a matrix M as given by (1) is derived; yielding a number of different calibration matrix M_i, i=1, 2, …, N. In general, N=6 is sufficient.

2.3 Color Sensor Calibration

Color sensor captures 2D color images of the different side views of an object. Information of each pixel can be represented by a set of five variables (X_f,Y_f,R,G,B). Therefore the calibration of a color sensor is the determination of the projective relationship between a spatial point with color information (X_w,Y_w,Z_w,R,G,B) and a point (X_f,Y_f,R,G,B) on the color bitmap; and the imposition of the color value of the point on the bitmap on the corresponding point on the object. Although it is possible to use methods based on linearization with respect to partitioned regions to calibrate color sensors, however, due to the many-to-one mapping between the spatial point and the corresponding color bitmap pixel, a color sensor calibration method which is based on NP neural network is adopted.

A three layer BP neural network structure is utilized. There are three neurons at the input layer corresponding to the object's three spatial coordinates X_w, Y_w and Z_w . The output layer has two neurons which correspond to the color bitmap's pixel coordinates X_f and Y_f . According to the Kolmogorov theorem, only one hidden layer is needed. Thus there are a total of $2 \times 3 + 1 = 7$ neurons.

The calibration of color sensor is realized by a recursive learning process through which the weighting coefficients with respect to the neurons are determined. The learning process of a BP neural network is divided into one in the forward propagation path, and another one in the reverse propagation path. In the forward path, using the object's coordinates, the corresponding pixel coordinate estimates are computed; the coordinate deviations are also computed by comparing with the true pixel coordinates. In the reverse path, the coordinate deviations are propagated toward the input nodes, with the weighting parameters of the neurons updated as well. This process is repeated until the deviations are reduced below certain threshold value; and a set of BP neural network weighting parameters reaches their final values.

Figure 4(a) is an image showing the light strip bitmap of a mannequin's head sampled by a gray scale CCD camera. Figure 4(b) is the point cloud image of the same mannequin's head constructed by processing the light strip bitmap image with the 3D sensor calibration matrix M. There are a total of 35629 points in the point cloud data. Figure 4(c) depicts the color bitmap image of the mannequin's head captured by one of the color CCD camera; and Figure 4(d) is a 3D point cloud image embedded with strip color information which is derived from computing the estimated spatial coordinates using the converged BP neural network weighting parameters as gain values and the point cloud 3D coordinates as inputs. As shown, our system provides fairly accurate 3D color information.

Fig. 4. (a) Light strip bitmap

Fig. 4. (b) 3D point cloud

Fig. 4. (c) Color bitmap

Fig. 4. (d) Color point cloud

3 Digital Human Animation

Digital human animation is one of the most difficult and challenging topics in 3D animation. Research approaches can be divided into three categories: surface based method, volume based method and layer based method. Surface based method can be further subdivided into rigid model animation [18], deformation function representation [19], contour based deformation [20] skinning method [21] and sample based animation [22]. Volume based method mainly includes implicit surface method [23] and skeleton driven volumetric data method [24]. In layer based animation, the human body is decomposed into a skeleton layer, a muscle layer, a fat layer and a skin layer; independent model at each of the layer is constructed, and the ultimate skin deformation is realized by transferring the modeling characteristics from the most inner to the most outer layer [25]. Among the volume based methods, the implicit surface method can easily create digital human with nice shapes, while the volumetric data model is best suited to drive digital human model that is already represented by volumetric data. Since the volumetric data method generally utilizes a large number of data points, it is also the slowest among all the methods. Although layer based animation generally produces vivid and realistic animation, the model construction process is extremely complex. At present, the most commonly applied algorithms in digital human animation for skin deformation are those provided by the surface based method, and in particular, skinning method. The main drawback of this method is the need for user's manual interaction. Regardless of the approach to be adopted for digital human animation, the first task is to extract the 3D model's skeleton.

3.1 Extraction of Skelton of Digital Human

Interpolating an excessive number of mesh vertices to create an implicit surface representation of a 3D digital human is a very time consuming process; it is also unrealizable with ordinary computing machine. Although excessive mesh vertices and surface microstructures will create a large number of micro-segments in the resulting 3D skeleton which is beneficial to feature tracking and surface reconstruction, there is no intrinsic value to path planning and animation design. Therefore, before creating digital human's implicit representation, we apply the method proposed in [25] to simplify the mesh. In general, this approach can reduce the number of triangular mesh facets to 1% of the total mesh facets generated by the original mesh.

We apply the algorithm proposed by Ma [27] to extract the 3D skeleton from the simplified digital human mesh model. In the process of mesh simplification and skeleton extraction, each initial vertex's corresponding vertex in the simplified model, and each simplified model's vertex corresponding point in skeleton are recorded. Thus, to each initial vertex there is a corresponding skeleton point. This relationship is a many-to-one mapping. By taking this additional step, the process of binding the surface vertices to the corresponding skeleton points is simplified, and therefore the required computation time is reduced.

Fig. 5(a) shows the digital human's surface mesh model reconstructed from the contour data of a human body which is produced by the 3D laser scanning system introduced in Section 2. Fig. 5(b) depicts the resulting simplified model which contains about 1 % of the initial model's triangular mesh facets. Since the initial mesh model contains an excessive number of vertices and triangular mesh facets, simplification to only 1% of the original model leads to insignificant reduction in the digital human's fidelity. Fig. 5(c) and (d) are two different views of the resulting 3D skeleton extracted from the simplified model shown in Fig. 5(b).

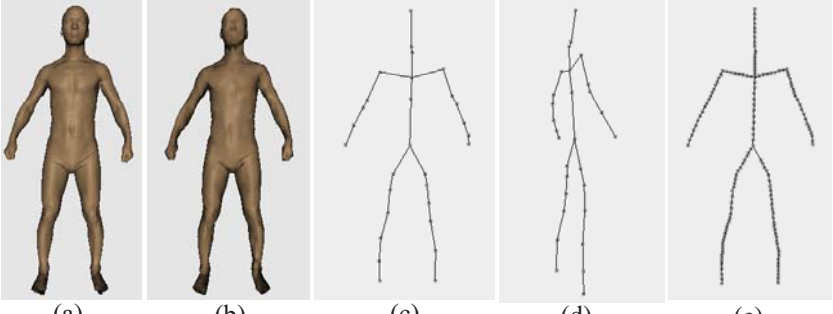

Fig. 5. Surface mesh model and the extracted 3D skeleton of a digital human

3.2 An Improved Skinning Method for Digital Human Animation

We introduce and apply an improved "skinning" method to realize digital human animation. In order to reduce the localized collapse and "candy wrapper" phenomena, the following three modifications are implemented:

(1) Extra joints on the 3D skeleton are added to minimize the distance between adjacent joints. Fig. 5(e) shows the resulting skeleton with extra equidistantly located joints added.

(2) A vertex on the surface is bound to multiple skeleton joints using the following equation:

$$T_i = \sum_{j=1}^{n} \frac{D - d_{j,i}}{(n-1) \times D} t_j \tag{2}$$

where t_j is the transformation matrix of the j[th] skeleton joint, T_i is the transformation matrix of the surface vertex i, $d_{j,i}$ is the distance between vertex i and joint j, D is total distance, and n is the number of joints that corresponds to vertex I; typically n=3. We first locate the joint j which has the minimal distance to surface vertex i; and then the surface vertex i is bound to the joint j and its adjacent joints j-1, j+1 by applying the transformation given by (2). Fig. 6 illustrates the idea where the black dot represents the surface vertex and the white circles are the skeleton joints.

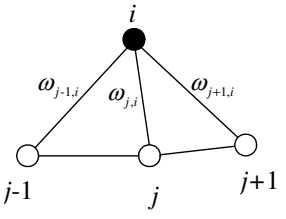

Fig. 6. Binding multiple skeleton joints to a surface vertex

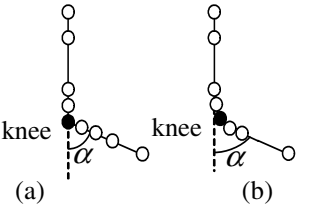

Fig. 7. An illustration of the compound joint technique

(3) Apply a compound joint technique to human body joints, such as the knee joint. Fig. 7 illustrates the kickback of the calf at an angle of α where the black dot denotes the knee joint location. Fig. 7(a) illustrates the case when all the joints below the knee joint are bent at an angle of α. In Fig. 7(b), two joints immediately above and below the knee joint are bent progressively with each of the joint bending at an angle of $\alpha/4$. This modification can effectively prevent the localized collapse and "candy wrapper" phenomena to occur. Since the distance between any two adjacent joints is very small, the compound joint technique produces minimal negative effects to overall fidelity.

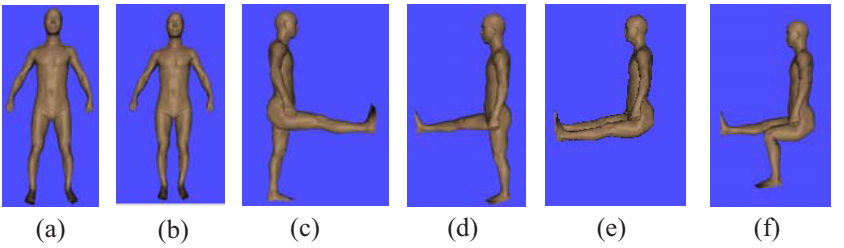

Fig. 8. A sequence of digital human re-posturing

Fig. 8 shows a sequence of digital human configurations where the legs are moved to present different postures. Fig. 8(a) is the digital human model in the initial stance; Fig. 8(b) indicates the posture after the feet move closer, showing a posture roughly at attention. Fig. 8(c) and (d) give two opposite views of a posture with the left leg raised to a level position. The posture showed in Fig. 8(e) result from raising both legs to level

positions, and finally Fig. 8(f) records the dropping of the right calf to a down right position from the posture given in (e).

4 Applications to Anthropometry

The development on digital human modeling as reported in the previous sections has been applied to obtain key human body measurements for anthropometry applications. Due to page limitation, details are omitted herein. We shall provide these results at the paper presentation in the conference.

5 Conclusions

In this paper we described the design of a color 3D human digitalization system, and elaborated on a 3D sensor calibration method based on linearized partitioned regions, and a color sensor calibration method based on BP neural network learning. This system has been demonstrated to acquire both 3D point cloud data and simultaneously the corresponding color information for large 3D objects such as the human body. The subsequently generated surface mesh model is further simplified to improve the speed of skeleton extraction. An improved "skinning" method with three new modifications has also been developed to improve the quality of digital human animation. These new developments have also been applied to obtain key human body measurements for anthropometry applications.

Acknowledgements. The work reported herein is partially supported by the China National Science Council through Project No. 60277009, and by Tianjin City Government Targeted Technology Program Support Project No. 06GFGPX05000.

References

[1] Guoxiong, Z.: The trend in development of coordinate measuring machines [J]. Chinese Mechanical Engineering(in Chinese) 11(2), 222–227 (2000)
[2] Narkhede, S., Golshani, F.: Stereoscopic imaging: a real-time, in depth look. IEEE Potentials 23(1), 28–42 (2004)
[3] Yujin, Z.: Image Engineering [M](in Chinese). Tsinghua University Press, Beijing (2000)
[4] Jin, H., Xiaobing, Z., et al.: 3D measuring techniques based on image processing[J]. Chinese Journal of Electron Devices(in Chinese) 25(4), 364–368 (2002)
[5] Demers, M.H., Hurley, J.D., Wulpern, R.C., et al.: Three dimensional surface capture for body measurement using projected sinusoidal patterns [A]. SPIE[C] 3023, 13–25 (1997)
[6] Frankchen., Shizhen, W.: Review for the measuring of 3-D object by optical method [J]. Cloud Light Technology(in Chinese) 34(2), 36–47 (2002)
[7] Tognola, G., Parazzini, M., et al.: A fast and reliable system for 3D surface acquisition and reconstruction. Image and Vision Computing 21, 295–305 (2003)
[8] Xu, Z.Q., Sun, C.K., et al.: 3D Color Reverse Engineering, Optoelectroincs (in Chinese). Laser 12(9), 937–939 (2001)

[9] Park, J., DeSouza, G.N., Kak, A.C.: Dual-beam structured-light scanning for 3-D object modeling. In: Proc. of Third International Conference on 3-D Digital Imaging and Modeling, pp. 65–72 (2001)
[10] Yuan, Y.C., Lo, Y.H., et al.: The 3D scanner for measuring body surface area: a simplified calculation in the Chinese adult. Applied Ergonomics 34, 273–278 (2003)
[11] Young, K.D.: 3D Mannequin Database - Hong Kong Innovation and Technology Commission funded Project Final Report, The Hong Kong University of Science and Technology (2002)
[12] Xu, Z.Q., Ye, S.H., Fan, G.Z.: Color 3D reverse engineering. Journal of Materials Processing Technology 129(1-3), 495–499 (2002)
[13] Hu, H.P., Li, D.H., et al.: 3D laser color scanning system. Advanced Technology Communications(in Chinese) 11(3), 25–30 (2001)
[14] Kari, P., Shapiro Linda, G.: Surface Reconstruction and Display from Range and Color Data. Graphical Models 62(3), 165–201 (2000)
[15] Ye, Y.: Camera Calibration method and boundary detection contour tracking algorithm research, Da Lian University doctoral dissertation, (in Chinese) (2002)
[16] Faig, W.: Calibration of close-range Photogrammetry systems: Mathematical formulation. Photogrammetric Eng. Remote Sensing 41, 1479–1486 (1975)
[17] Lenz, R.K., Tsai, R.Y.: Techniques for calibration of the scale factor and image center for high accuracy 3D machine vision metrology. Pattern Analysis and Machine Intelligence, IEEE Transaction 10(5), 713–720 (1988)
[18] Zordan, B.V., Majkowsk, A., Chiu, B., Fast, M.: Dynamic response for motion capture animation. ACM Transactions on Graphics 24(3), 697–701 (2005)
[19] Magnenat-Thalmann, N., Laperriere, R., Thalmann, D.: Joint-Dependent Local Deformations for Hand Animation and Object Grasping. In: Proceedings of Graphics Interface'88, pp. 26–33 (1988)
[20] Kalra, P., Thalmann, N.M., Moccozet, L., Sannier, G., Aubel, A., Thalmann, D.: Real-time animation of realistic virtual humans. Computer Graphics and Applications 18(5), 42–56 (1998)
[21] Mohr, A., Tokheim, L., Gleicher, M.: Direct manipulation of interactive character skins. In: Proceedings of the 2003 symposium on Interactive 3D graphics, pp. 27–30 (2003)
[22] Lewis, J.P., Cordner, M., Fong, N.: Pose space deformation: a unified approach to shape interpolation and skeleton-driven deformation. In: Computer Graphics, SIGGRAPH'2000 Proceedings pp. 165–172 (2000)
[23] Leclercq, A., Akkouche, S., Galin, E.: Mixing triangle meshes and implicit surfaces in character animation. In: Animation and Simulation'01, 12th Eurographics Workshop Proceedings, Manchester, England, pp. 37–47 (2001)
[24] Gagvani, N., Silver, D.: Animating Volumetric Models. Graphical Models 63(6), 443–458 (2001)
[25] Nedel, L.P., Thalmann, D.: Anatomically modeling of deformable human bodies. The Visual Computer 16(6), 306–321 (2000)
[26] Garland, M., Heckbert, P.S.: Surface simplification using quadric error metrics. In: Proceedings of SIGGRAPH, ACM Press ACM SIGGRAPH 2000, Computer Graphics Proceedings, Annual Conference Series, ACM, vol. 1997, pp. 209–216 (2000)
[27] Ma, W.-C., Wu, F.-C., Ouhyoung, M.: Skeleton Extraction of 3D Objects with Radial Basis Functions. In: IEEE Proceedings of the Shape Modeling International 2003 (SMI'03) (2003)