A Passive Full Body Scanner Using Shape from Silhouettes

Sebastian Weik
Institut für Theoretische Nachrichtentechnik
und Informationsverarbeitung
Hannover, Germany
weik@tnt.uni-hannover.de

Abstract

This contribution describes a camera-based approach to fully automatically create 3D models of persons. A setup of sixteen digital cameras is used to capture the images of the person to be scanned. Using a monochromatic background and the shape-from-silhouette approach a 3D model is created automatically. Using the original images the models are realistically textured. In future 3D tele presence applications or multi-player games these models have to be animated using an internal skeleton structure. A principal component analysis is utilised to extract the skeleton automatically from the models. Compared to other full body scanning approaches no mechanically moving components are used which makes the system especially suitable for low cost end user applications.

1. Introduction

Current video conferencing systems use 2D images to transmit and display visual information between different conference partners. As far as video is concerned such a system only enables the user to perform bilateral information exchange. The only way for more than two persons to take part in a 2D video conference is to broadcast the image information to all other participants and get the image streams from every other conferee. This requests an enormous technical effort, mainly bandwidth, and doesn't really give the impression of physically beeing close to each other. Another drawback is the fixed viewpoint of each participant's image stream which is determined by his local camera system and can not be changed arbitrarily at the receiver's site.

For future distributed video conferencing systems it is therefore desirable not only to transmit 2D video, audio and shared applications. Rather one would wish to create a 3D virtual copy of a real 3D scene for each user, containing highly realistic models of the participants. These models should look like, act and behave similar to the actual conferees.

The virtual environment would be provided and managed by a central session server. Each participant of such a virtual conference would run a client application and could distribute his own model and the respective movement via the central server [1]. The virtual environment and the models of the conferees are shared once in the beginning among the participants, the subsequent movements of the conferees are transmitted continously from the clients to the server. The server redistributes the actual motion to all clients providing the clients with the possibility to update the scene according to all motion parameters.

Main elements of such a system are the personal models which need to be constructed highly automatically when used in consumer market applications like 3D teleconferencing or multi-player games. The creation of those models is mainly divided in two parts: firstly the extraction of the shape and texture of the real person and secondly the automatic adaptation and fitting of an interior skeleton structure for animation purposes.

2. Mechanical Setup

As opposed to the modelling of rigid objects a turntable setup can not be used when modelling human beings; a turntable requires the person to remain rigid for the duration of the image acquisition process (e.g. several minutes). Therefore a different setup has been constructed here (Fig. 1). The person is situated in a monochromatic coated background which is used later on for silhouette extraction. To avoid the sequential acquisition the images are captured simultaneously using 16 digital cameras. The combination of background and camera positions need to fulfil mainly two important constraints: firstly, *all* cameras must see the complete person in front of the monochromatic background and secondly no camera should be visible from any other camera. Fig. 2 shows a photograph of the scanner during a presentation on CeBIT99 in Hanover/Germany.

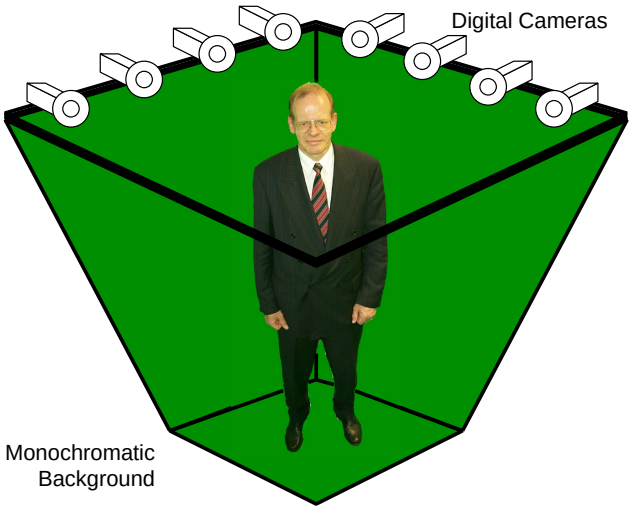

Figure 1. Principal measurement setup

Figure 2. Photograph of Body Scanner

3. Camera Calibration

Since a calibrated setup is required for the *shape-from-silhouettes* approach a special 3D calibration pattern has been designed. It is covered with calibration points of known 3D position and of approximately the size of a person. Prior to the actual scanning process the pattern is positioned in the middle of the scanner and the calibration images are taken. From these images the intrinsic and extrinsic camera parameters with respect to the calibration pattern are estimated automatically using the algorithm proposed by Tsai [2]. Fig. 3 shows a photograph of the pattern within the scanner.

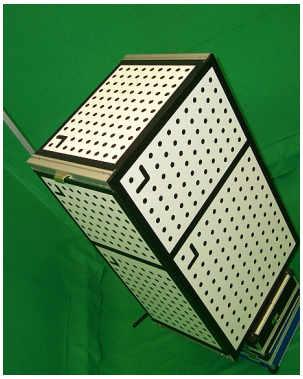

Figure 3. 3D calibration pattern

Figure 4. Input image and segmented foreground

4. Shape from Silhouette

The *shape from silhouettes* or ”method of occluding contours” approach is a well known technique for the automatic reconstruction of 3D objects from multiple camera views [3]. In this section the reconstruction technique is described in more detail.

The principle of the silhouette-based volumetric reconstruction can be divided into three steps. In the first step, the silhouette of the real object must be extracted from the input images as shown in Fig. 4. In the proposed environment the segmentation of the person against the background is facilitated by using the monochromatic background (”blue screen technique”).

In the second step, a volumetric cone is constructed using the focal point of the camera and the silhouette as shown in Fig. 5a. The convex hull of the cone is formed by the lines of sight from the camera focal point through all contour points of the object silhouette. For each view point

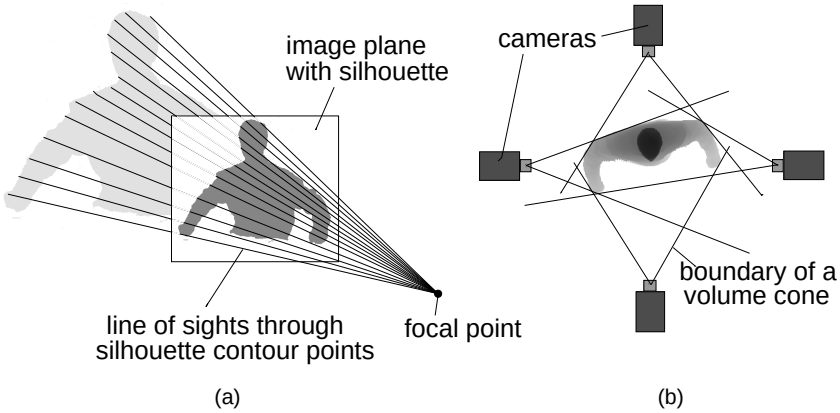

Figure 5. Volume reconstruction: (a) Construction of a volumetric cone, (b) Top-view of the cone intersection

Figure 6. Rendered views of a wireframe and a textured model in a virtual scene

such a volumetric cone is constructed, and each cone can be seen as a first approximation of the volume model.

In the last step, the volumetric cones from different view points are intersected in 3D and form the final approximation of the volume model. This is performed with the knowledge of the camera parameters, which give the information of the geometrical relation between the volumetric cones. In Fig. 5b a two dimensional top view of the intersection of the cones is shown. In Fig. 6 on the left a triangulated 3D point cloud representing the volume model surface is shown.

After the reconstruction of the geometry the model can automatically be textured using the original camera images giving a highly realistic impression [4]. To reduce the size of the resulting texture map the texture resolution is adapted to certain areas of the model, e.g. hands and facial areas are textured with a higher resolution. Fig. 6 right shows a rendered view of a textured model in a virtual scene.

5. Skeleton fitting

In order to widely use the 3D models in future 3D-telepresence or multi-player game applications they need to be animated. Therefore an internal skeleton structure is needed which controls the model movements. Normally this requires a tedious manual positioning of the joint positions within the model. In order to reduce the costs of model creation it is desirable to automate this process. Some research has been performed in this area and first results can be shown here.

As opposed to other algorithms that use the thinning of 3D data[5] finding the skeleton here is a multi-step process which is based on re-projected images of the voxel model of the person. As opposed to models created from a laser-scan device our models created from *shape-from-silhouettes* are inherently dense and do not contain any missing areas from measurement errors. This dramatically simplifies the automatic fitting of the skeleton to the model.

In a first step a principal axis analysis is performed to transform the model into a defined position and orientation. Using a virtual camera – not to be confound with one of the real cameras – a synthetic silhouette from a frontal viewpoint is calculated. The outer contour of this image is used to extract certain feature points like the bounding box, the position of the neck, the hands and so on as can be seen in Fig. 7 on the left. In the last step the 2D joint positions of the desired skeleton are derived directly from the detected

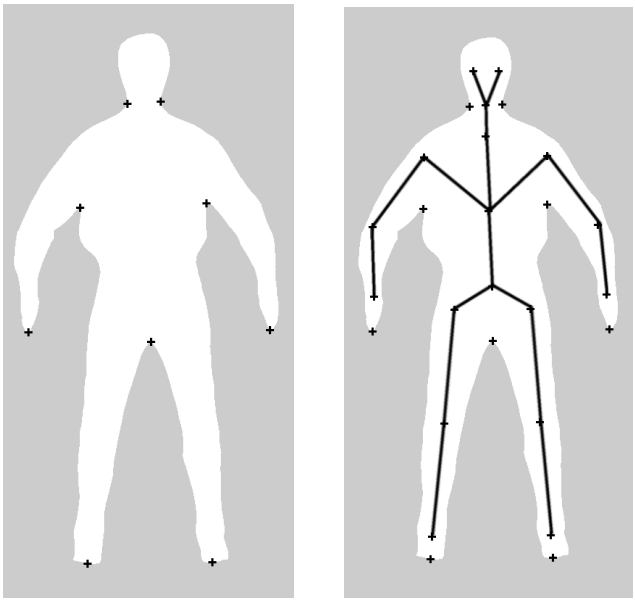

Figure 7. Extracted features (left) and calculated skeleton

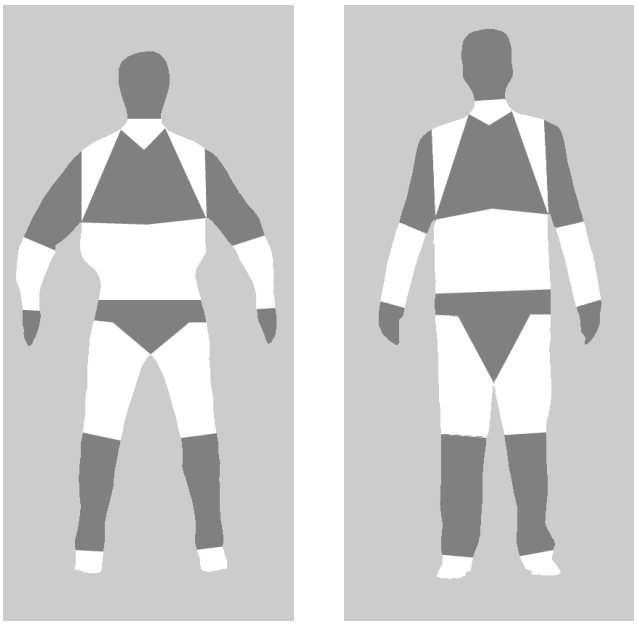

Figure 8. Examples of automatic segmentation in body parts

feature points using certain ratios (Fig. 7, right). Using the real model and the virtual camera these 2D joint positions are extended to their real 3D positions. In Fig. 8 the skeletons have been used to segment the model into different body parts. The found 3D joint positions and the mesh segmentation can directly be used to export the models into 3D animation tools.

6. Conclusions

An efficient and promising approach has been presented for anthropomorphic modelling. The 3D data of the real persons to be modelled is obtained using a *shape-from-silhouette* approach. A special setup with monochromatic background and 16 digital cameras capture the images simultaneously. Texturing the 3D models from the original images leads to a highly realistic appearance.

The presented approach guides the way to an automatic kind of creating anthropomorphic models as needed in 3D telecommunication applications or multi-player-games where a time consuming manual creation of a large number of different models is inappropriate.

7. Acknowledgements

The author would like to thank *Dimension 3D Systems* (www.dimension-3d.com) without whose funding part of this work would have been impossible.

References

[1] J. Wingbermuehle, S. Weik, “Towards Automatic Creation of Realistic Anthropomorphic Models for Real-time 3D Telecommunication”, Journal of VLSI Signal Processing Systems, Special Issue on Multimedia Signal Processing, Vol. 20, Oct. 1998, pp.81-96

[2] R. Y. Tsai, “A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses”, IEEE Journal of Robotics and Automation, Vol. RA-3. No.4, August 1987, pp. 323-344.

[3] R. Szeliski, “Rapid Octree Construction from Image Sequences”, Graphical Models and Image Processing: Image Understanding, Vol. 58, No 1, July, pp. 23-32, 1993.

[4] W. Niem, H. Broszio, “Mapping Texture from Multiple Camera Views onto 3D-Object Models for Computer Animation,” in Proceedings of the International Workshop on Stereoscopic and Three Dimensional Imaging, Santorini, Greece, 1995.

[5] C. Pedney, “Distance-ordered homotopic thinning: a skeletonization algorithm for 3D digital images”, Comput. Vis. Image Underst., vol.72, no.3, p404-13, 1998