REAL-TIME RECONSTRUCTION OF INCOMPLETE HUMAN MODEL USING COMPUTER VISION

Ing. Vítězslav BERAN, Doctoral Degree Programme (1)
Dept. of Computer Graphics, FIT, BUT
E-mail: beranv@fit.vutbr.cz

Ing. Igor POTÚČEK, Doctoral Degree Programme (2)
Dept. of Computer Graphics, FIT, BUT
E-mail: potucek@fit.vutbr.cz

Supervised by: Dr. Pavel Zemčík

ABSTRACT

This paper covers improved real-time method for human model reconstruction based on features extracted from silhouette. The incomplete 2,5D model is constructed from the silhouette and its depth is finally evaluated by considering the prior knowledge of human body shape.

1 INTRODUCTION

Service of computers in communication between people constantly brings new demands on quality of interaction and also various novel interaction tasks. An opportunity also exists to find ways how to improve the current state of communication in point of view of realism.

Communication systems based on videoconference structure often lack spontaneity between remote conversing participants. However, some approaches exist that attempt to overcome the problems mentioned above. One of them is application of augmented reality techniques. Augmented reality systems mostly use special display equipment and allow augmenting an image of a real environment with virtual objects. Such objects can be models of the remote participants. Using specific user models in such systems is necessary due to severe distortion when the plain user model is rotated [1].

The need to preserve real-time execution requires simplification of the user model reconstruction process. Furthermore, videoconference applications mostly use only one camera for a scene monitoring, which is very limiting. The research is focused on enhancement of silhouette feature extraction techniques [2].

2 OVERVIEW OF THE STRUCTURAL MODEL RECONSTRUCTION

Due to the complex nature of human body, tracking of humans in video sequences is very difficult task. Several approaches exist how to reconstruct an articulated structural model that is specified up-front. The **Model-matching** approach creates a hierarchical structure of sticks connected by joints. According to physical attributes of the matched model, constraints of the sticks and joints are applied and parameterized. The task of this approach is to find such parameters for which the projection of the structural model best fits the image of the required object. The reconstructed model contains a complete 3D description but the computational cost is very high.

Another approach can be **Usage of the 2D features and human shape knowledge**. Several features are extracted from the silhouette of required object and are consequently used, together with prior human body shape features, for construction of human body model. Due to single view limitation, only incomplete (2,5D) model can be obtained.

3 OUTLINE OF THE ALGORITHM

The reconstruction process can be divided into two stages: *2D image analysis* and *3D feature evaluation and incorporation*. In the first stage, a body boundary is acquired and smoothed. The obtained silhouette then serves for a body mesh construction that is an interim data structure. Further analysis of the silhouette and the corresponding image region, the second stage, is used for gathering additional features. The results are treated as parameters of the final depth evaluation process. The depth evaluation algorithm applied to the plain body mesh combines the prior knowledge about human body shape and probability information about special parts of model (face, neck, hands, strange objects, etc.).

3.1 2D IMAGE ANALYSIS

General image processing methods can be exploited in the body silhouette evaluation algorithm to enhance the silhouette image. This enhancement, in combination with advanced techniques such as skin detection, can sufficiently increase the probability that the picked region is the right one from the human body.

The algorithm that exploits a temporal body region, its silhouette, and auxiliary skin regions has the following steps:

- background image is pre-acquired (without the human body),
- background subtraction is applied to each frame of an input image sequence, and differentiating regions are extracted,
- coordinates of the centroid (center of gravity), region boundary and area of the extracted regions are calculated,
- regions are scanned for skin-colored pixels [3], so the new temporal regions represent a face and hands (can increase a probability of proper localization of human body region candidate), the region with the largest area is also considered as appropriate candidate,
- from the calculated region parameters, the system searches for a contour pixel on the extracted region, the silhouette contour is traced, and the 2D coordinates of contour pixels are listed,
- suppression of noise damaging a contour smoothness is achieved by the dilation and erosion algorithms that are applied only on contour points, so that the speed of whole process is maintained,

- the contour tracking is executed again on the filtered region and finally, the contour is approximated into line segments of polygon.

Fig. 1: *The input image and constructed mesh.*

The body silhouette (polygon) is used for a body model mesh construction [1] (see figure 1). A depth of the model is evaluated in next process stage.

3.2 3D FEATURE EVALUATION AND INCORPORATION

Body parts separation serves for better evaluation of the model depth. Each body part has different property from a depth influence point of view. Cuts, breaking the body region, must lie on a negative minima curvature and satisfy certain criteria [5]. Further, if boundary points may be joint in more than one way to decompose a silhouette, the shortest one must be used. There are three factors affecting the salience of the part: the size of the part relative to the whole region, the degree to which the part protrudes and the strength of its boundaries. The algorithm is depicted on figure 2.

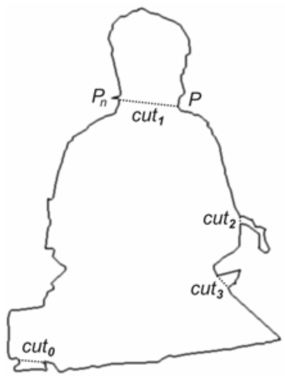


Fig. 2: *Body parts separation using cuts.*

$$P_n = \arg\min_{P'} |PP'| \qquad (1)$$

$$\frac{\langle PP' \rangle}{|PP'|} > T \qquad (2)$$

… where $\langle\ \rangle$ represents an arc length,

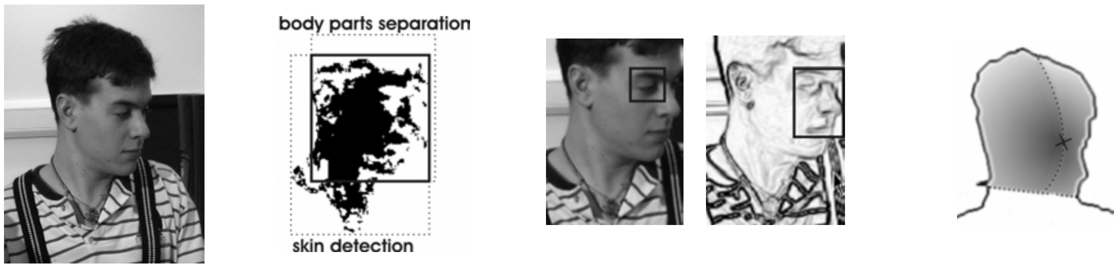
$(\ )$ is a length of the cut and

T is a ratio threshold.

Get the point P from silhouette with negative minima curvature. Find another point P_n that meets the conditions (1) and (2). The body can be broken by cut P,P_n only if one of the new body part does not contain any other cut. In combination with skin-colored region information, the probability of particular body part recognition is increased.

Head orientation refinement can eliminate distortion that comes up when the user's head is significantly rotated. Due to a similarity of left and right part of face, to find the head orientation correspond with task to find an axis of the face.

The face region is obtained from body parts separation and skin detection results (figure 3 (b)). Center of the face is then evaluated from dark patches around eyes and clutter of edges around eyes, nose and mouth (figure 3 (c)). The face centre position determines a required head rotation and serves for definition of final head shape refinement (figure 3 (d)).



(a) (b) (c) (d)

Fig. 3: *Stages of head orientation refinement process.*

Depth evaluation is the final process that combines all previously obtained features and creates a depth map. Each feature affects the map with different weight. The primary function defining an elementary model shape is a polynomial function of the distance *h(x,y)* to the silhouette border. The final *z* coordination function can have several forms. With respect to the speed of the method the *z(x,y)* main depth function has the following form (3).

$$z(x,y) = \text{height}.\left(1 - \frac{(\max - h(x,y))^2}{\max^2}\right) \tag{3}$$

where *height* is the maximum *z* value in the final model and *max* is the maximum distance found in the silhouette that corresponds to the value of this point.

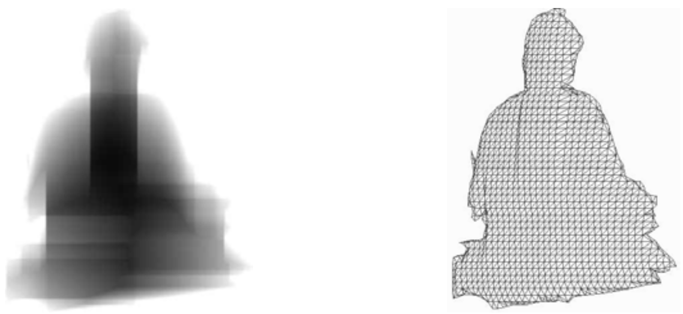


Fig. 4: *The main depth function applied on the plain mesh.*

A contribution of other features is not so significant and just refines the final model appearance.

4 RESULTS

The main purpose of the research is to improve the model appearance when the model is rendered in different viewpoint (e.g. model is rotated, see figure 5). Therefore, when the algorithm result is mentioned, it is meant in the realistic model appearance point of view. The proposed algorithm was tested in an office environment with invariable light conditions and with just one participant. This way, it was possible to study the behavior of proposed method for various human body poses and different participants.

Final model shape is refined and its appearance is enhanced without adversely affecting a method computation cost. A disadvantage of this refining approach is its low robustness. Worse results are obtained when the user is not properly clothed (no T-shirt and the like), for instance, is not sitting behind the table, etc. However, such nuisances are not so disturbing, because the participant is focused on model's head that is mostly processed properly.

Fig. 5: *Rotation of plain and incomplete 2,5D models.*

5 CONCLUSION

The method is usable particularly in systems providing a face-to-face communication in augmented environment. The lack of method robustness can be improved when the reconstructed model is in progress during the video sequence.

Further refinement of the model appearance could be achieved by incorporating some time variant features. Some edges in the user region keep their relative positions in contrast to hand edges, for example, when the hand occludes or blend with user's torso. Using edge detection might avoid not only the mentioned problem but also unsuitable model appearance when some strange object occludes the user's body.

The research state and applied algorithms give sufficiently good results to be used in face-to-face communication in augmented environment even without the improvements suggested above.

REFERENCES

[1] Beran, V.: Augmented reality in videoconference system, Brno, BUT 2003

[2] Potmesil, M.: Generation octree models of 3d objects from their silhouettes in a sequence of images, in Computer vision, Graphics and Image Processing, 40:1-20, 1987

[3] Potúček, I.: Person Tracking Using Omnidirectional View, Vědecký sborník, CZ, p. 603-607, 2003, ISBN 80-214-2379

[4] Huazhong Ning, Liang Wang, Weiming Hu, Tieniu Tan: Model-based Tracking of Human Walking in Monocular Video Sequences, IEEE Region 10 Technical Conference on Computers, Communication, Control and Power Engineering, 2002

[5] Anurag Mittal, Liang Zhao, Larry S. Davis: Human Body Pose Estimation Using Silhouette Shape Analysis, Conference on Advanced Video and Signal Based Surveillance, 2003, p. 263-270