Augmented Multi-User Communication System

Vítězslav Beran

Brno University of Technology
Faculty of Information Technology
Božetěchova 2, 612 66 Brno, CZ
+420 777 286078

beranv@fit.vutbr.cz

ABSTRACT

This paper presents improvements carried out to enhance the visual interaction of computer users in existing communication systems. It includes the usage of augmented reality techniques and the modification of a method for user model reconstruction according to particular requirements of such applications. Promised achievement is to prepare the background for further development of multi-user interface, videoconference or collaborative workspace.

The aim of our research is replacing the standard computer interface components by equipment used in augmented reality and so immerse the user into augmented environment. Such approach allows to user positioning virtual objects in his workspace. One of possible techniques for precise virtual object pose evaluation widely used in augmented reality applications is to employ special tracking markers.

Traditionally, communication systems of videoconference type represent a remote user using his sprite (plain, billboard-like) model. The lack of realistic appearance, when the participant is displayed as a sprite model, can be eliminated by its artificial reconstruction. The method gain depth information from knowledge of human anatomy and hence it is able to create the artificial relief model of the remote user.

Categories and Subject Descriptors

I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism – *virtual reality, visible line/surface algorithms*.

General Terms

Experimentation.

Keywords

augmented reality, mixed reality, image analysis, object reconstruction, grid based triangulation

1. INTRODUCTION

New, continually developing, information technology trends are still bringing new ways of using computers; new types of users are arising; new types of data and interactions are coming up. Videoconference systems fix the focus of participants on a monitor or LCD display and do not allow them to be immersed in the virtual communication environment. These challenges are compound by attempts to support three-dimensional collaboration workspace [1]. One of the possibilities how to create some very natural environment for human communication with the computer is using augmented reality techniques. Augmented reality (AR) is a technology that attempts to restrain drawbacks of the virtual reality. Instead of the virtual reality techniques AR does not create any artificial environment around the user but only the user's augmented view. The computer-created virtual objects are inserted into image of the real world so the authenticity and naturalness are maintained [2][3].

The research attempts to utilize advantages of AR for an improvement of the current videoconference system standard. Using the equipment and methods of AR to build up such communication environment whose impression is as realistic as possible and that enables users to communicate in such natural way as during the real "face-to-face" communication. The flatness of the remote participant sprite model becomes readily apparent if the user is moving. The displacement between the remote participant and the camera in front of him does not correlate with the displacement between the user's eyes and the remote participant model (resp. marker). These displacement differences would cause the distorted appearance of the remote participant.

To avoid the distortion, the assumption, that the image of the user is an image of a human, allows reconstruct incomplete three-dimensional model (relief). Such final spatial model represents the particular user in stereometric and realistic environment. Unfortunately, it does not allow full freedom of viewpoint movement such as views from above or behind, but permits at least a wider range of viewpoint movement without heavy distortion.

2. SYSTEM STRUCTURE

The Augmented Multi-User Communication System (AMUCS) is designed to provide an extension of the environment surrounding the user, i.e. includes other participants of communication into the user's room. Two main tasks should be supplied by the system (see Figure 2). Firstly, is necessary to find spatial positions in the user's workspace where replaces the virtual objects and models to. The scene is captured in the user's head direction in space (HMC – head mounted camera) and markers are detected. To simplify the scene recognition and position evaluation, special contrasting markers are convenient to use. The AMUCS detects code on the markers and according to them replaces the marker with a model of another particular collaborator. The resulting video stream is displayed back to the user on a head mounted display (HMD). Secondly, instead of a sprite model of remote

AVI '04, May 25-28, 2004, Gallipoli (LE), Italy

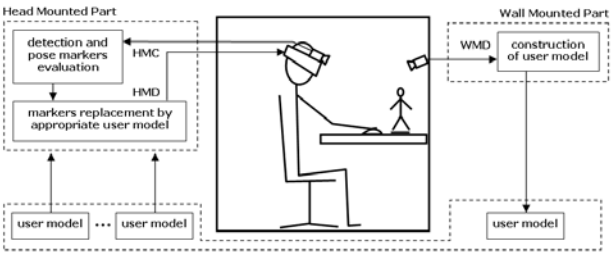

Figure 1. Structure of the user workplace in the AMUCS.

participant, the knowledge of human anatomy allows to artificially create and use his relief model. The user is scanned (WMC – wall mounted camera), separated from background and its model is reconstructed.

3. MODEL CONSTRUCTION

The necessity of model reconstruction arises from a requirement for an enhanced reality of the communication environment. The quality and dimensionality of reconstructed model deeply depends on the number of its images from different viewpoints. The realistic matter is not the only attribute that determinates the quality of the AMUCS. There is another feature with the same importance, the speed. Different classes of quality in shape representation exist such as sprite, relief or full 3D model. The AMUCS captures the user only from one viewpoint so only restricted two-and-a-half dimensionality can be achieved (relief).

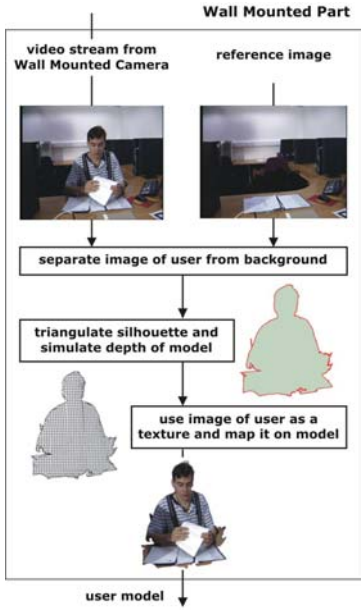

Figure 2. Stages of model construction process.

The first and also most sophisticated step is to find and separate a region with a participant from the background. A user's silhouette can be used to construct a mesh of its model. Contour polygon is triangulated and a depth algorithm simulates the 3D appearance. The model construction process is depicted on Figure 2.

3.1 Planar Mesh from Silhouette

A wide variety of possible constructs and techniques exist that offer several convenient approaches to tasks of the AMUCS in the model construction stage. The problem is that the system is supposed to provide a real-time tool and the lack of algorithm speed would be more disturbing in the final effect than a lower quality of reconstructed model.

The triangulation used is an extended version of the structured mesh method and achieves better fitting with the input polygon (see Figure 3(a)) and outperforms the method speed. The triangulation procedure is divided into two parts. The first stage of the triangulation process is the main grid triangulation. The additional grid vertexes from within extend the input polygon (Figure 3(b)). Algorithm browses over all of them, scans their neighbourhood, and creates appropriate triangles. In relation to existence of grid points in the actual point neighbourhood, the method creates the inner mesh structure. The partial result can be seen in Figure 3(c).

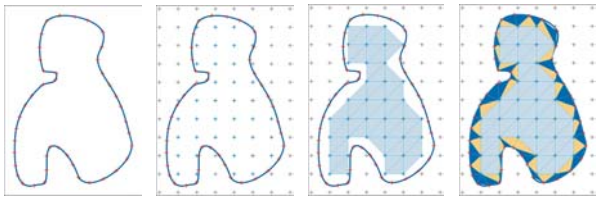

Figure 3. Mesh evaluation.

Such result is a mesh that only approximates the user's silhouette very roughly. Maximum decrease of error between the silhouette and the mesh is achieved by the more precise triangulation of the object boundary. Line segments of the silhouette are considered as edges of the final mesh. The triangulation is applied to the space delimited by the silhouette edges on one side and the edges of the rough mesh boundary on the other. The method processes the silhouette edges and searches for the closest most convenient grid vertexes. Such point must be on the right side of the vector $\overrightarrow{V_i V_{i-1}}$ (rule 1) and the distance between actual (V_i) and most convenient points must be minimal (rule 2). Figure 4 depicts the situation of detecting such point. When the grid vertex is found, the new triangle is added to the mesh and the next silhouette edge is processed. One special situation during this method can occur: if the new triangle is made up from a different grid vertex than was in the previous triangle, an extra passing triangle must be constructed (see odd-colored triangles on the Figure 3(d)). This triangle contains these two grid vertices and the silhouette vertex conjuncts actual and previous silhouette edges.

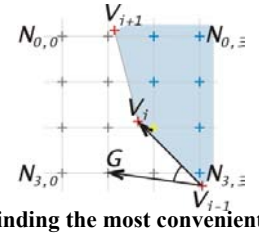

Figure 4. Finding the most convenient grid point.

The final mesh is in planar sprite representation. To evaluate the depth, the z coordination, of the mesh, the relief evaluation

algorithm must be executed with the presumption of the object represented by the mesh.

3.2 Relief Evaluation

In order to overcome the limitations of the planar sprite representation, depth information must be incorporated into the model. A fast, scalable, and robust method for reconstruction is "shape from silhouette" [4]. The planar mesh can be interpreted as some kind of shape from silhouette approach using two cameras; one main plus auxiliary camera for the depth. The application range of the planar sprite model can be extended even in the case when there is only one main camera available by using basic assumptions about the object shape. For human beings and some other objects, it is a good guess to assume that the volume expansion maximum occurs in the middle of the object. Further points in the silhouette edge have a big slope. This fact means that the surface normal is almost perpendicular to the optical axis of the camera here. A model that fits these conditions is an ellipsoid. This is suited for simple shaped objects and was successfully applied to head and shoulder scenes for modelling a person [5],[6].

The method is based on 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)* depth function can has form (1).

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

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

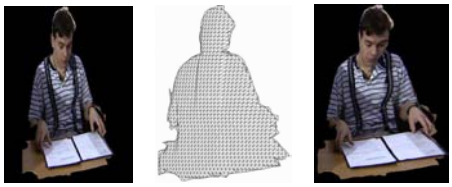

Figure 5. Different viewpoint: sprite, relief mesh and model.

The Figure 5 shows the result of a relief creation by using only one main camera and the assumption about the object geometry. The synthetic image still shows some artefacts due to the simplified three-dimensional shape. In contrast with the planar sprite model, the appearance of the model is much more realistic when moving the viewpoint away from the original camera position.

4. Marker Replacement

According to the main idea of AR, part of the input image should be overlapped by another image, generally an image of some 3D object. AR systems must know where to place this image of the virtual object. Precise information about position and orientation of the camera in the world coordinates is necessary not only because of an estimation of the area in which to match the image but also because of the rendering of a virtual object from the right viewpoint [7].

Several different techniques exist how to estimate extrinsic camera parameters and the speed and accuracy of methods correlate with the amount of known information. The AMUCS uses markers with special properties that enhance the speed and accuracy of the estimation. The Figure 6 depicts stages of whole process. The predefined parameters of markers and their patterns serve to assigning models to the right markers. The 3D pose of

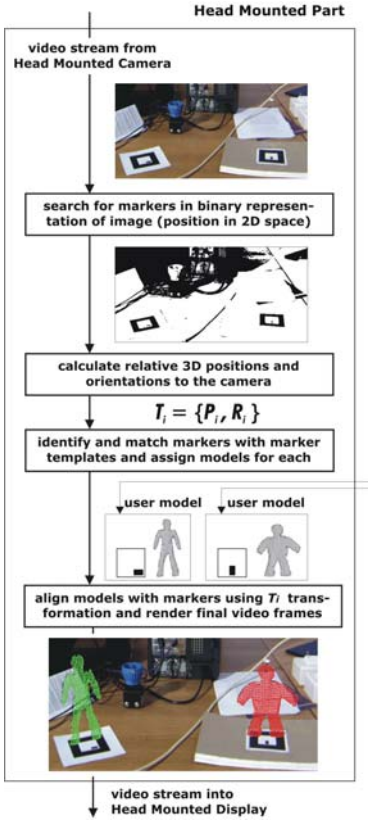

Figure 6. Stages of markers replacement process by models.

markers cannot be processed without intrinsic camera parameters. When the transformation $\boldsymbol{P_i}$, composed of transition $\boldsymbol{T_i}$ and rotation $\boldsymbol{R_i}$, for each marker is estimated, the models can be rendered from the right viewpoint and then matched with the input image.

Let us suppose that the tracking marker of predefined size and shape (size-known square, Figure 7). The crucial transformation ($\mathbf{T_m}$) from marker coordinates (X_m,Y_m,Z_m) to camera coordinates (X_c,Y_c,Z_c) is estimated by image analysis. The image processing techniques [8] are applied to obtain marker's parameters in form of vertexes of marker corners $m_i[m_{ix}, m_{iy}]$, $i=1..4$.

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{3,3} & \mathbf{T}_{3,1} \\ 0 \;\; 0 \;\; 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix} = \mathbf{T}_m \cdot \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix} \qquad (2)$$

Observing the marker projection on the image (2), parallel lines appear concurrent due to camera perspective projection. The image of line on the camera screen is actually projection of plane from 3D space. Using perspective projection matrix **P** obtained by camera calibration, the rotation matrix can be combined from vectors $\mathbf{v_1}$, $\mathbf{v_2}$, $\mathbf{v_3}$ as $\mathbf{R}_{3,3} = \begin{bmatrix} \mathbf{v}_1^T & \mathbf{v}_2^T & \mathbf{v}_3^T \end{bmatrix}$. The vector $\mathbf{v_1}$, resp. $\mathbf{v_2}$,

is the direction vector of two parallel sides of the marker given by the cross product of the plane's normal vectors $\mathbf{n_0}$ and $\mathbf{n_2}$, resp. $\mathbf{n_1}$ and $\mathbf{n_3}$. The vector $\mathbf{v_3}$ is perpendicular to $\mathbf{v_1}$ and $\mathbf{v_2}$.

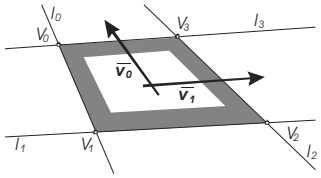

Figure 7. Marker rotation evaluation.

Since the rotation component $\mathbf{R_{3,3}}$ in the transformation matrix is now given, the translation components $\mathbf{T_{3,1}}$ can be estimated. Let's consider points $W_i[W_{iX}, W_{iY}, W_{iZ}]$, $i=1..4$, of the marker in initial position. First, the partly-unknown transformation $\mathbf{T_m}$ is applied to those points. The equation (3) is based on the assumption that the points $w_i[w_{ix},w_{iy}]$, the projection of transformed points W_i by projective matrix $\mathbf{P}$, should have the same coordinates as the points $m_i[m_{ix}, m_{iy}]$, $i=1..4$ (marker corners).

$$\begin{bmatrix} h_m \cdot m_{ix} \\ h_m \cdot m_{iy} \\ h_m \\ 1 \end{bmatrix} = \begin{bmatrix} h_w \cdot w_{ix} \\ h_w \cdot w_{iy} \\ h_w \\ 1 \end{bmatrix} = \mathbf{P} \cdot \begin{bmatrix} \mathbf{R}_{3,3} & \mathbf{T}_{3,1} \\ 0 \;\; 0 \;\; 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} W_{iX} \\ W_{iY} \\ W_{iZ} \\ 1 \end{bmatrix} \qquad (3)$$

Using four known points, the equation can be modified and parameters T_x, T_y and T_z can be obtained from the system of eight linear equations. The drawback of such method is that it may include error. Slight re-optimisation of rotation coefficients can minimize such error.

Models of other AMUCS participants are rendered in right viewpoint, merged with the input image and projected into HMD. Although the method used for the user model reconstruction does not give a full three-dimensional model, even the artificially evaluated model creates an improved illusion of real interaction. The background of a remote participant is flat and consequently distorted when the viewpoint is changed. A user model placed on its background is distorted too but due to its faked 2,5D model the distortion is not so serious.

5. Results

The model reconstruction was implemented and tested, especially the part of the silhouette extraction and its triangulation. The significant advantage of our method is its simplicity and robustness. Using the grid points, with their known position, brings the method fastness. When the method parameters, such as a grid point distance or a silhouette fine coefficient, are suitably set a computational complexity and memory requirement are $\mathrm{O}(n)$, because it is not necessary to search for point neighbours in all other points. Compare the results of techniques that need to execute searching are then $\mathrm{O}(n \cdot \log n)$.

Unfortunately, the method does not properly cope with the noise of the silhouette. It can be solved by additional image processing techniques that get rid of such a noise and smooth the user's contour and consequently its silhouette. A discussed image pre-processing significantly slows down the whole process. Tests with the median filter and dilation and erosion methods to reduce the contour noise were accomplished. The difference in final model appearance was not considerably significant in contrast with the slowing down of the performance.

6. Conclusion

The main goal of the research is focused on using augmented reality in the videoconference systems and partly on the implementation of such system. The motivation to use augmented reality techniques came from previous researches coping with a scene feature detection, etc. The AMUCS uses results of augmented reality development and is able to find the pose where place the model of remote participant to. Finally, the AMUCS execute our algorithm for user model reconstruction, obtains the model and uses it to create an augmented communication environment. The appearance of workspace is now more realistic due to the usage of the user's relief model instead of its sprite representation.

7. ACKNOWLEDGMENTS

The paper has been partly supported by EU IST Program project Augmented Multi-party Interaction, EU-HLT, 506811-AMI, 2004-2006.

References

[1] Billinghurst, M., Kato, H., *Collaborative Mixed Reality*, In Proceedings of the First International Symposium on Mixed Reality (ISMR '99), 1999, pp. 261-284, Berlin: Springer Verlag.

[2] Milgram, P., H. Takemura, et al. (1994). *Augmented Reality: a Class of Displays on the Reality-Virutality Continuum*. SPIE Proceedings: Telemanipulator and Telepresence Technologies . H. Das, SPIE. 2351 : pp. 282-292.

[3] Azuma, Ronald T., *A Survey of Augmented Reality*, Teleoperators and Virtual Environments 6, 4 August 1997, pp. 355-385.

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

[5] Ostermann, J., *Modelling of 3D moving objects for an analysis-synthesis coder*, Proc. of SPIE/SPSE Symposium on Sensing and Reconstruction of 3D Object and Scenes, B. Girod Ed., Proc. SPIE 1260, Santa Clara, California, USA, February 1990.

[6] Grau, O., Price, M., Thomas, G.A., *Use of 3-D Techniques for Virtual Production*, SPIE conference on Videometrics and Optical Methods for 3D Shape Measurement, San Jose, USA, 22-23 Jan.2001.

[7] Kato, H., Billinghurst, M., Weghorst, S., Furness, T., *A Mixed Reality 3D Conferencing Application*, HITL, University of Washington, http://www.hitl.washington.edu.

[8] Gonzalez, R., Woods, R., *Digital Image Processing*, ISBN 0201180758, Prentice Hall, Hardcover, 2nd edition, November 2001.