VIRTUAL MIRROR: REAL-TIME TRACKING OF SHOES IN AUGMENTED REALITY ENVIRONMENTS

Peter Eisert, Jürgen Rurainsky, Philipp Fechteler

Fraunhofer Institute for Telecommunications, Heinrich-Hertz Institute
Einsteinufer 37, D-10587 Berlin, Germany
{eisert, rurainsky, philipp.fechteler}@hhi.fraunhofer.de
URL: http://iphome.hhi.de/eisert

ABSTRACT

In this paper, we present a system that enhances the visualization of customized sports shoes using augmented reality techniques. Instead of viewing yourself in a real mirror, sophisticated 3D image processing techniques are used to verify the appearance of new shoe models. A single camera captures the person and outputs the mirrored images onto a large display which replaces the real mirror. The 3D motion of both feet are tracked in real-time with a new motion tracking algorithm. Computer graphics models of the shoes are augmented into the video such that the person seems to wear the virtual shoes.

Index Terms— 3D modeling & synthesis, stereoscopic and 3D processing, motion detection and estimation, real-time systems

1. INTRODUCTION

The Virtual Mirror presented in this paper is a system for the visualization of customized shoes that a person can try and watch in front of a mirror although they do not exist in reality. This is achieved by using augmented reality techniques which combine real video with virtual objects represented by 3D computer graphics models. The use of a mirror environment enables the augmentation of the user with artificial objects without the user being forced to wear special glasses. Other approaches also use such virtual mirror techniques for applications like visual effects (face distortions) [1], mobile electronic mirrors [2], or for image communication purposes [3].

Our system as shown in Fig. 1 has been created for adidas and runs in their new store, opened in October 2006 at the Champs Elysées, Paris, France. In this innovation center, a customer cannot chose only shoes from the shelf but design personalized models. Besides particular fitting to the left and right foot, the client can change the design and colors of a shoe model at a special terminal and add individual embroideries and decorations. In order to give the customer an impression how the shoes will finally look like after being manufactured, the user can step in front of the Virtual Mirror.

Fig. 1. Installation of the Virtual Mirror in the adidas store, Champs Elysées, Paris.

There, a camera captures the customer wearing fitting boots with a standard design. A display replaces a real mirror and outputs the horizontally flipped camera image. The display is mounted such that the person appears at the same position, where the user would expect to see himself when looking into a real mirror. In order to enhance the virtual feeling of the framework, the background is segmented and replaced by a synthetic environment. A novel 3D motion tracker estimates the position and orientation for each foot using a model-based approach that is very robust and can easily be adapted to new shoe models. Once the exact feet positions in 3D space are known, the computer graphics models, that have been configured and colored according to the customer's wishes, are rendered and integrated into the video stream such that the real shoes are replaced by the virtual ones. Special care has to be taken for this augmentation, since the real scene in the 2D video can occlude parts of the virtual 3D scene. Therefore, visibility for all parts of the shoe has to be computed for a given position. Since all algorithms have been implemented with real-time constraints, the customer can move freely and

watch himself/herself with the new shoes that have been designed just some moments earlier.

2. SYSTEM DESCRIPTION

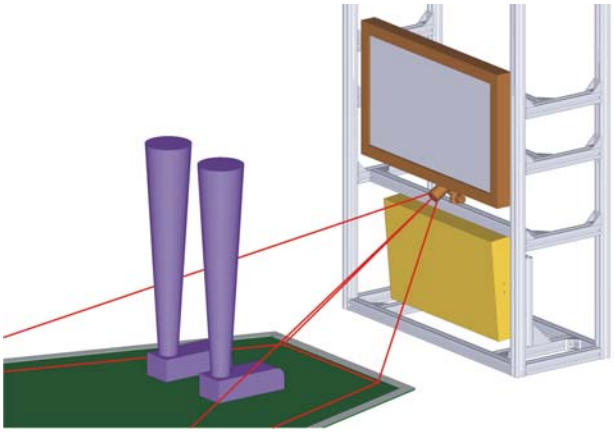

Fig. 2. Architecture of the Virtual Mirror.

The system mainly consists of a single camera and a display showing the output of the Virtual Mirror. The XGA firewire camera is mounted close to the display and looks down, capturing the feet of a person standing in front of the system. The legs of the user are segmented from the background and displayed on the screen after mirroring the video signal horizontally. The placement of the display and the viewing direction of the camera are chosen such that an average sized person sees about the same as he/she would expect when looking in a real mirror located at the same position as the display. In order to simplify the segmentation in a real environment with changing illumination and arbitrary colors of clothes, the floor in front of the camera is painted in green to allow the use of chroma keying techniques. Moreover, an additional light below the camera reduces the effect of shadows, since the illumination in the store is mainly from the top.

All image processing algorithms, tracking, rendering and augmentation run on a single PC. It also hosts a web server that allows the control of the Virtual Mirror and interfacing with the adidas configuration and database system. In the algorithmic flow of the Virtual Mirror, several different components like

- segmentation and 2D image pre-processing
- gradient-based 3D tracking
- rendering and augmentation

have been developed, which are described in the following sections.

3. IMAGE PROCESSING

The calibrated camera [4] of the Virtual Mirror continuously captures the space in front of the system and transfers the images with a resolution of 1024 by 768 pixels to the image processing components. All automatic camera control is switched off to avoid unexpected behavior after changes in the lighting of the scene. In order to avoid interference with the illumination, the shutter time is synchronized with the flickering of the lights. The camera gain is re-computed to adjust for changing illumination each time when nobody is within the range of the camera.

This idle state is determined by a change detector, that exploits information about the spatio-temporal variations in the video signal. After the camera gain has been adjusted to the current lighting situation, a background image is computed by averaging 10 consecutive video frames. This background image is used for the segmentation of the mainly green background from the shoes and the legs.

In order to fulfill the real-time constraint, all image processing is applied in an image pyramid. The original image is filtered and down-sampled four times by a factor of two, until a final resolution of 64 by 48 pixels is reached. The segmentation algorithm starts on this lowest level by comparing all pixel colors with the corresponding ones in the background image. In the decision, whether the pixel belongs to the fore- or background, also a 3D RGB lookup table is used having 64^3 elements. This color cube is filled adaptively with the green background pixels. In order to handle also shadows and reflections on the floor, the resulting shape of background pixels in the RGB color cube is extended by cylinder- and cone-like models. After the pixels have been classified, small holes are filled and small regions are removed until only the two legs with the shoes remain. This segmentation mask is now propagated to higher levels, where only pixels are segmented (as described above) that originate from boundary pixels in the lower level. This process is repeated until the final resolution is reached, resulting in segmentation masks for each pyramid level.

From these segmentation masks, horizontal and vertical histograms are computed, which are used to determine, if a person has entered the virtual mirror. In case, both shoes are completely visible, the 3D tracking is initialized with the blob position, and 3D motion tracking is started.

4. 3D TRACKING

The 3D tracker estimates the two rigid body motion parameter sets corresponding to the shoes from a single frame of the camera. In total, 12 parameters are estimated using an analysis-by-synthesis technique similar to the face tracker described in [5, 6]. Instead of tracking a set of distinct feature points, the entire image is exploited for robust motion estimation. 3D computer graphics models specifying the shape of the shoes are rendered into a synthetic image approximating the camera frame. By reading out the z-buffer of the graphics card, information about the shoes' silhouettes and their dense depth information is obtained. The silhouette mask of the synthesized frame is now matched with the segmentation

mask from the foreground segmentation. All motion parameters (two sets of R_x, R_y, R_z, t_x, t_y, t_z) are optimized such that there is a perfect fit of real and synthetic silhouettes. The use of silhouette information as input for the tracker leads to robust results even for the highly specular materials of the sports shoes. However, texture and color information can be exploited in the analysis-by-synthesis loop exactly in the same way by additionally providing texture information to the computer graphics models.

The tracking process can be thought as finding the 3D parameter set that optimally matches the 2D silhouettes (and/or color information). However, a complete search in the 6- respectively 12-dimensional space would be very inefficient. Therefore, the parameters are directly computed using a gradient-based technique. For that purpose, the binary silhouette masks of synthetic and camera frames are first filtered using a separable 7 tap moving average filter. This operation transforms the binary object borders into linear ramps with constant gradient values. The closer a pixel is to the object, the higher the pixel values. By comparing different intensity values, information about mismatch of boundaries can be computed using the optical flow constraint equation [7]

$$\frac{\partial I(X,Y)}{\partial X}d_x + \frac{\partial I(X,Y)}{\partial Y}d_y = I(X,Y) - I'(X,Y), \quad (1)$$

where $\frac{\partial I}{\partial X}$ and $\frac{\partial I}{\partial Y}$ are the spatial derivatives at pixel position $[X\ Y]$. $I'-I$ denotes the intensity change between the filtered original and synthetic silhouette image. The 2D displacement vector can be related with the unknown motion parameters

$$[d_x\ \ d_y] = f(R_x, R_y, R_z, t_x, t_y, t_z) \quad (2)$$

using information about the rigid body motion model and the knowledge about the camera parameters [5]. Combination of (1) with (2) provides one additional equation for each pixel close to the object border. An over-determined linear set of equations is obtained that can be solved efficiently in a least-squares sense. Remaining errors in the motion parameter set caused by linearizations are resolved by applying the algorithm iteratively. Currently, four steps are used to converge to the desired resolution.

5. RENDERING AND AUGMENTATION

Once the 3D motion parameters for both shoes are determined, the 3D computer graphics models of the personalized shoes can be rendered at the correct position, such that the real shoes of the person in front of the Virtual Mirror are replaced. These 3D models can be individually configured by a customer by selecting a base model and choosing between different sole types, materials, and colors. Additionally, individual embroideries like flags and text can be attached. From this configuration data, an individual 3D model is composed.

For that purpose, geometry, texture, and colors of the 3D models have to be modified to represent the chosen design. Each shoe model consists of different sub-objects composed of triangle meshes which can be replaced to create different geometries. For the modeling of different surface materials, individual texture maps are chosen from a database. Additionally, colors can be assigned to the textures in order to customize the individual parts of the shoe. This way, the customer can chose among hundreds of thousands of models and design a shoe according to personal preferences.

Fig. 3. View dependent rendering for correct augmentation into video. Parts that should be later occluded by the legs (2D video) are not visible.

The left and right shoe are finally rendered using OpenGL at the position and orientation determined by the 3D tracker. In the rendering and augmentation process, first the background is rendered at a far distance. Then, the original video is rendered using the segmentation mask as alpha channel of the RGBA texture map. Finally, the shoes objects are overlayed, occluding the original shoes in the segmented video. However, the legs in the original 2D video should also occlude some parts of the shoes. By adding also a transparent non visible leg model to the scene, the z-buffer is manipulated such that all occlusions are recovered correctly and the 3D model can be augmented into the 2D video. Fig. 3 shows two examples of the shoe rendering with some parts, later occluded by the legs, removed.

6. EXPERIMENTAL RESULTS

In this section, some results of the tracking and rendering are presented. Four different shoe models are configured and the Virtual Mirror is started. A single camera captures the scene at a resolution of 1024 by 768 pixels. A user enters the green space in front of the system. In all cases, the shoes are cor-

Fig. 4. **Upper row:** Scene captured with the Virtual Mirror camera. **Lower row:** Virtual Mirror output augmented with customized shoes.

rectly detected, segmented, and tracked. Fig. 4 shows examples of the Virtual Mirror output. The upper row illustrates some frames of the original scene captured with the camera. For these frames, the corresponding results displayed on the Virtual Mirror are depicted in the lower row. It can be seen, that the 3D computer models follow the original 3D shoe motion correctly, even for rather extreme feet positions.

Since the entire framework should behave like a real mirror, real-time processing is required. All algorithms are therefore optimized for speed. Image processing algorithms are applied in an image pyramid and the tracking is also computed on a lower resolution level. In order to exploit multiple processor cores, a multi-threaded framework is setup with four threads for capturing/debayering, image processing, rendering, and system control. On a 2.3 GHz dual processor Pentium 4 based system, the Virtual Mirror takes about 48 ms for all computations with 4 iterations of the tracker, leading to a frame rate of more than 20 Hz.

7. CONCLUSIONS

We have presented a system for the real-time 3D tracking of shoes in a Virtual Mirror environment. From a single camera the rigid body motion of left and right foot are estimated using linear and low-complexity optimization methods. The tracking algorithm is hereby not restricted to shoe models, but can also be applied to other objects if a 3D geometry description is available. The motion information is then used to render new personalized sports shoes into the real scene such that the person can watch himself/herself with the new shoes. The system is accessible to the public in a store in Paris and new installations with updated features are also planned for New York and Beijing.

Acknowledgements

The work presented in this paper has been developed with the support of the European Network of Excellence VISNET II (Contract IST-1-038398).

8. REFERENCES

[1] T. Darrell, G. Gordon, J. Woodfill, and M. Harville, "A virtual mirror interface using real-time robust face tracking," in *Proc. Third International Conference on Face and Gesture Recognition*, Nara, Japan, Apr. 1998.

[2] A. Francois and E. Kang, "A handheld mirror simulation," in *Proc. International Conference on Multimedia and Expo (ICME)*, Los Angeles, USA, Jul. 2003, vol. 2, pp. 745–748.

[3] C. Cullinan and S. Agamanolis, "Reflexion: A responsive virtual mirror for interpersonal communication," in *Proc. ECSCW 2003 8th European Conference on Computer Supported Cooperative Work*, Helsinki, Finland, Sep. 2003.

[4] P. Eisert, "Model-based camera calibration using analysis by synthesis techniques," in *Proc. International Workshop on Vision, Modeling, and Visualization*, Erlangen, Germany, Nov. 2002, pp. 307–314.

[5] P. Eisert and B. Girod, "Analyzing facial expressions for virtual conferencing," *IEEE Computer Graphics and Applications*, vol. 18, no. 5, pp. 70–78, Sep. 1998.

[6] P. Eisert, "MPEG-4 facial animation in video analysis and synthesis," *International Journal of Imaging Systems and Technology*, vol. 13, no. 5, pp. 245–256, Mar. 2003.

[7] B. K. P. Horn, *Robot Vision*, MIT Press, Cambridge, 1986.