VIRTUAL JEWEL RENDERING FOR AUGMENTED REALITY ENVIRONMENTS

Peter Eisert∗‡ *Christian Jacquemin*† *Anna Hilsmann*‡

∗: Humboldt Universität zu Berlin
†: LMSI-CNRS and University Paris 11
‡: Fraunhofer HHI, Berlin

ABSTRACT

Virtual Mirrors are augmented reality applications that capture a viewer's image and add 2D or 3D overlaid virtual elements, such as cloth patterns, or replace existing pieces of garments, such as shoes, by virtual ones. The purpose of our study is to focus on the addition of jewels that present complex rendering issues due to reflection and translucency, especially for real-time applications. We propose two complementary approaches which are both suited for particular objects. The image-based rendering technique relies on an image data-set of photos taken on a semi-circle of camera positions and a 3D reconstruction of the jewel. This approach is exploited to demonstrate real jewels with sophisticated refractions and highlights. The analytical 3D rendering technique can also visualize configurable virtual jewels and is based on the decomposition of the jewel into parts without non-local reflection or transparency. A graph-based rendering chain takes in consideration some of the self-reflection and lighting effects.

***Index Terms*—** Augmented reality, jewel rendering, image-based rendering

1. INTRODUCTION

Interactive virtual mirrors are applications that offer an augmented image of a user for the purpose of seeing her/himself with different garments and/or in a different environment [1, 2]. So far, most of the work in such projects has been devoted to virtual element registration through user tracking, and to a fluid and natural combination of virtual elements and input video, but little has been made about complex rendering issues. This research is focused on the rendering of virtual jewels that represent a good example of 3D additions with significant reflective and refractive properties.

In the following, we will propose two alternate and complementary methods for jewel rendering. The first one is an image-based technique which relies on the projection on a 3D reconstructed model of blended photos of a jewel under different angles of view. This technique is particularly suitable for realistic rendering of real existing jewelry with many highlights and refractions. However, later modifications of the objects is difficult since no semantic model is available. If interactive configuration of a virtual jewel is desired, the second technique presented is most suitable. It relies on the decomposition of the jewel into parts without non-local reflection or transparency and their combined rendering through a pipeline of environment-based or impostor-based lighting steps. This method allows to reproduce reflections of the environment and/or the user. The final image projected on the Virtual Mirror screen is the combination of the user's video capture with the jewel rendering, and additional light effects of the jewels on the user's image.

Blending the user's mirror image with jewel renderings relies on the registration of the 3D jewel models with the user's clothes through deformable surface tracking. Additional augmentations such as cloth color modification or texture overlay can also be combined with jewel augmentation resulting in a rich and flexible environment for virtual try-on of jewels and clothes as illustrated in Fig. 6.

2. IMAGE-BASED RENDERING THROUGH 3D IMAGE MAPPING

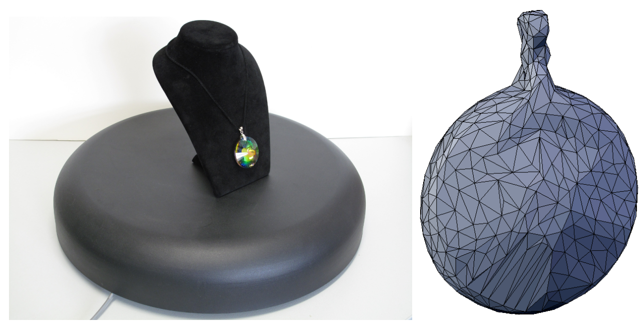

Fig. 1. Turntable setup for capturing and approximate geometry model reconstructed from the images.

Realistic rendering of real jewels with all their highlights, glitter, and refractions is difficult to achieve, especially if fine structures and sophisticated materials like crystals are present. One possibility to obtain realistic results without modeling all fine details is image-based rendering [3, 4]. From a large set of pictures taken of a real object from different viewing directions, new views can be interpolated by 2D image processing operations. Since the real images capture all complex lighting effects, very natural rendering is achieved. However, a large number of images might be required to cover all view variations, posing high demands on memory capacity and acquisition effort.

For a Virtual Mirror application, we target at virtually adding a brooch, a necklace, or an earring to the mirrored view of a user captured by a camera. The lighting effects on the jewelry depend in principle on both the viewing and the lighting directions. Since the environmental illumination of the mirror setup can be assumed as relatively constant as well as the viewing position of the user which is constrained by the viewing angle of the camera, we assume that most variation is due to the motion in particular the rotation of the

∗e-mail: peter.eisert@hhi.fraunhofer.de
†e-mail: christian.jacquemin@limsi.fr
‡e-mail:anna.hilsmann@hhi.fraunhofer.de

object itself. The rotation in the image plane can be addressed by 2D image processing and the remaining two degrees of freedom of the rotation can be further reduced by the fact that the user can turn freely but is restricted to bend around a horizontal axis. Thus, we represent the object turning by a large array of images while the other degrees are handled by conventional rendering of an approximate geometry model. For that purpose, we reconstruct a 3D geometry model from the captured images as shown in Fig. 1. Since this model is only used to perform image warping in a very limited angular range, the accuracy of the model needs not be very precise as long as it covers the entire silhouette of the object. Additionally, it serves as depth estimate for disparity compensated image interpolation from the discrete set of captured images.

2.1. Acquisition of Images and Geometry

The jewel representation mentioned above consists of an approximate 3D shape model and a large number of texture images corresponding to different turning angles of the object. Therefore, for each new object, we have to capture the image array and create a geometry model. For the acquisition, we use a computer controlled turntable as shown in Fig. 1 that rotates a real jewel mounted on a bust. Illumination is fixed and adjusted in order to match approximately the situation in the final rendering. A digital camera automatically captures the object at different turning angles. In our experiments, we recorded one image every 2 degrees of rotation in a range of $\pm 90^o$ leading to 91 images. In this work, we capture images rotating around the vertical axis only, since the person in front of the mirror may turn freely, but horizontal rotation is extremely limited and will be covered by the 3D model. However, capturing with 2 degrees of freedom is straight forward and does not require any changes to the proposed method.

From this set of images, the geometry is reconstructed using a shape-from-silhouette framework. The jewel is segmented from the (known) background such that the visual hull of the object can be computed. This requires knowledge of the exact projection corresponding to each image. The intrinsic parameters of the camera are calibrated initially through a model-based calibration method, such that only the extrinsic parameters describing rotation and translation need to be determined for each captured view, which is robustly achieved by optimizing the mismatch between the silhouettes of the reconstructed object and the real ones [5].

2.2. Image-based Rendering

The interactive rendering of the jewel is made on the GPU as follows.

- The vertex shader computes the texture coordinates for the jewel according to the current viewing angle and the calibration parameters of the camera during pre-capturing. In order to support tri-linear interpolation, two pairs of texture coordinates are computed that correspond to the two closest capture angles α_f and α_c immediately below and above the current rotation of the jewel with respect to the observer.
- The fragment shader accesses the two texture parts corresponding to α_f and α_c based on the texture coordinates from the vertex shader, and blends the corresponding two color values. Two blending functions have been used for rendering: a uniform linear interpolation, and a cubic Bezier curve. The steepness of the Bezier curve is higher for pixels with important gray level differences to better highlight the flashing effects of specular light on the jewel facets.

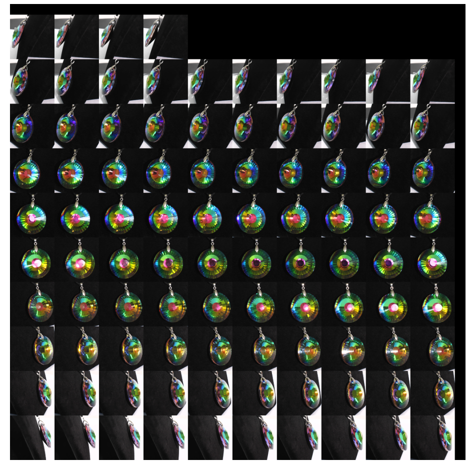

Fig. 2. Texture for image-based view interpolation.

3. ANALYTICAL RENDERING THROUGH JEWEL DECOMPOSITION

The image-based rendering approach is well suited for existing pieces of jewelry that do not require interactive redesign. If new configurable items should be visualized, an analytical 3D model-based approach is selected. Especially for reflective e.g. metallic surfaces, this approach can take environmental lighting conditions into account for convincing blending of real and virtual parts.

The rendering of the virtual mirror application is based on a fixed screen and camera configuration, and the major changes of the environment are due to the user's motion in front of the mirror. In order to enhance the user's sense of presence in the environment, it is important to offer a good perceptive continuity between the real and virtual worlds. Seamless visual integration should augment the virtual world with real-world image and light information so that users can see their own reflections or shadows on the virtual objects [6, 7]. Since lighting is enclosed in the virtual mirror and faces the user, it is not likely that the user will cast shadows on the virtual jewels. But in the case of jewel with reflective metallic parts, the user is likely to observe her/his own reflection in the jewel. We wish to take into account these reflective effects through an environment-based and impostor-based rendering technique of 3D jewel models.

3.1. Environment Map

Image-based reflection and refraction rendering relies on environment maps—cubes or spheres textured with shots of the surrounding environment—so that they capture the light coming in all directions from the reflective or transparent object [8]. In case of changing environment the shots can be updated through live textures built from video images. For environment mapping or ray-tracing, video textures allow a proper blending of the real and virtual worlds by rendering live reflection or refraction [9].

Fig. 3 describes the construction of the environment map. We assume that the floor, ceiling, and side images are fixed and precomputed. Since the rendered scene represents the reflected image of the user in front of the mirror, the environment cube map is centered on the mirror. It is made of two symmetric halves: the real and the virtual scenes. Because the user is orientated towards the camera, only the lateral x+ and x- sides and y+ sides contribute to the rendering. The latter one is the composition of fixed shot 2 with the reflected live camera image (shot 4). The lower part of the figures presents

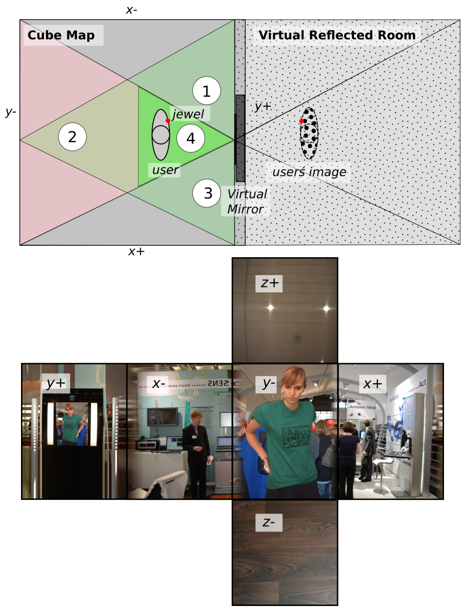

Fig. 3. Environment map of the *Virtual Mirror*

the global view of the environment map, in which the live image contributes to build the whole y- face and part of the y+ face. The y+, x+, and x- faces are used for computing jewel specular reflection. The y- video image is used for transparency, and for rendering the augmented reflection of the user through the overlay and composition with the rendered jewel images.

3.2. Rendering Architecture and Pipeline

To render reflections and transparency on complex objects at interactive rates, it is possible to use image-based rendering by replacing parts of the geometry by their rendering through billboards, depth maps [10], or layered distance maps [11]. In our application, we assume that the jewels are small and distant enough to avoid light interactions between them. For these reasons, each jewel rendering can be computed independently and depends mainly on the initial environment (the mirror and the room) and its changing part (the user captured by the camera).

The light-based rendering of the jewel and its combination with the video image is built on the combination of elementary passes. Each pass is the rendering of a virtual scene made of optional 3D elements and a background quad. This quad fits the view volume and its rendering is computed from bitmap textures or video images (such as the environment map faces), or frame buffer captures from previous intermediary steps. The output of each pass is the content of its frame buffer stored in a Frame Buffer Object (FBO), except for the last rendering step that is displayed on the screen. Thus the inputs to an elementary rendering pass are:

- static or animated textures (still images or video images), or FBOs (rendering of previous passes),
- zero, one, or more geometrical elements that possibly use the preceding image resources for their renderings,
- a background plane with a compositing shader that combines geometrical element renderings and other image sources.

Since we do not use ray-tracing, each jewel is decomposed into elementary parts without non-local reflection or transparency, and multipass processing is used to render these non-local effects. Previous pass outputs are used as input images for subsequent steps, thus producing a cumulative scene rendering. Because of the dependency between inputs and outputs, the whole rendering process can be defined as a graph in which the nodes are the elementary rendering steps and the links are the dependencies between FBO inputs and outputs. In order to illustrate how the rendering graph and the composition of its elements are deduced from the structure of a jewel, we detail the case of a jewel made of a reflective metallic part on top of which is a refractive and reflective crystal (Fig. 4).

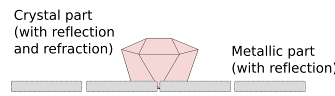

Fig. 4. Sample 2-part jewel.

The complete chain is illustrated by Fig. 5, and is made of 8 steps. Two parallel sequences 2-3-6 and 4-5-7 render the two parts of the jewel and produce a blurred mask to avoid aliasing and ensure a proper blending between the 3D rendering and the video image. A last step, implemented in a fragment shader, combines these outputs in a back to front order and adds them to the video image. Blurred masks are scaled for a texture-based shadowing of the jewels.

The graph rendering is implemented with *Virtual Choreographer*[1] XML language and associated 3D engine. The use of XML data encoding would make automatic scene generation quite simple. In addition, such XML encoded scenes could easily be translated to other XML-based 3D scene and behavior languages.

4. TRACKING 3D OBJECTS AND DEFORMABLE SURFACES

The translation and orientation of the jewels are controlled by the location of the vertices of a mesh-based model that is used to track different parts of the human body, depending on the kind of jewels. For example, for a necklace or a bracelet, the 3D tracking methods presented in [1] can be used to track the rigid body motion of the human neck or wrist. Here, we use the 3D model of the tracked part and exploit mainly silhouette information. In these cases the 3D position and orientation can be directly deduced from the 3D model in the current frame. For a brooch, which is attached to the clothes that can be stretched or bent, we use a 2D deformable mesh to track the elastic deformations of the cloth in the image plane, because the 3D reconstruction of elastically deformable surfaces is difficult in real-time [2]. The translation of an attached jewel is obtained by anchoring it to one of the vertices of the mesh and the orientation is deduced from the inverse homography transformation of 4 or more points of the mesh from their initial to their current position. In all our tracking methods we exploit an analysis-by-synthesis approach where the estimation is performed between the actual camera frame and a synthetic version of the previous frame, thereby avoiding error accumulation. This is essential for real-time applications like the Virtual Mirror to achieve an accurate augmented reality experience and to give the user the impression of truly wearing the virtual jewels.

5. CONCLUSION

This paper presented two rendering modes for virtual jewel augmentation in a Virtual Mirror application. Both methods have their

[1] http://virchor.sf.net/

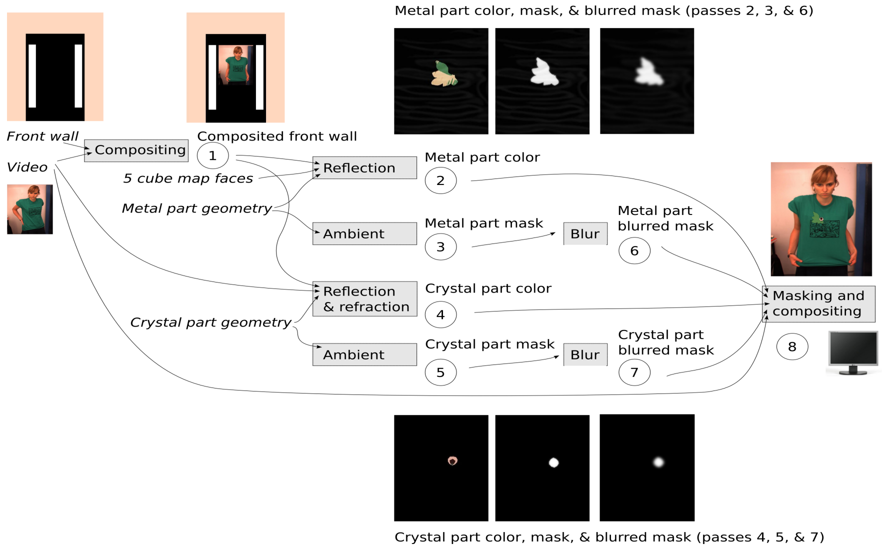

Fig. 5. Sample processing chain for the rendering of the jewel shown in Fig. 4.

advantages. The image-based technique (center pendant in Fig. 6) achieves very realistic results for real jewelry with many highlights, refractions and details without explicitly modeling all these fine details in the geometry since the real images capture all lighting effects (Fig. 7) . However, reflection of the user cannot be achieved as the set of images is captured a-priori. The multipass 3D light-based rendering (left brooch in Fig. 6) takes these reflective effects into account by using an environment- and impostor-based approach. However, in order to achieve realistic results, every fine detail has to be modeled in the 3D geometry in contrast to the image-based technique.

Fig. 6. Two augmented jewel rendering modes combined with cloth recoloring and texture overlay. The center jewel is created by image-based rendering while the left brooch is explicitly modeled.

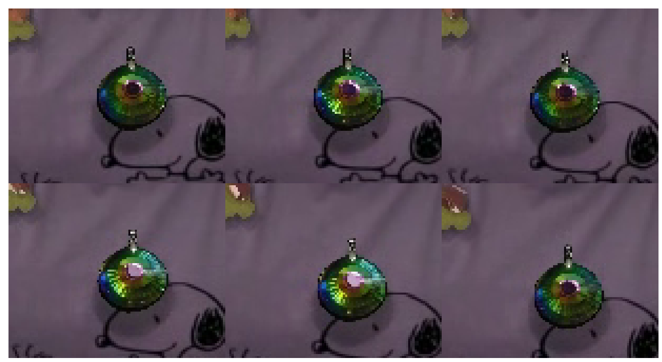

Fig. 7. Closeups of the sequence in Fig. 6 showing view dependent reflections of the jewelry.

6. REFERENCES

[1] P. Eisert, P. Fechteler, and J. Rurainsky, "3-D tracking of shoes for virtual mirror applications," in *Proc. of Int. Conf. on Computer Vision and Pattern Recognition (CVPR 2008)*, Anchorage, Alaska, June 2008.

[2] A. Hilsmann and P. Eisert, "Tracking deformable surfaces with optical flow in the presence of self occlusions in monocular image sequences," in *Proc. CVPR Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment*, Anchorage, USA, June 2008.

[3] Y. Mukaigawa, H. Miyaki, S. Mihashi, and T. Shakunaga, "Photometric image-based rendering for image generation in arbitrary illumination," in *Proc. ICCV*, July 2001, pp. 652–659.

[4] W. Matusik, H. Pfister, R. Ziegler, A. Ngan, and L. McMillan, "Acquisition and rendering of transparent and refractive objects," in *Proc. 13th Eurographics Workshop on Rendering*, Pisa, June 2002, pp. 267–277.

[5] P. Eisert, "3-D geometry enhancement by contour optimization in turntable sequences," in *Proc. Int. Conf. on Image Processing (ICIP)*, Singapore, Oct. 2004, pp. 3021–3024.

[6] F. Steinicke, K. Hinrichs, and T. Ropinski, "Virtual reflections and virtual shadows in mixed reality environments.," 2005, vol. 3585 of *Lecture Notes in Computer Science*, pp. 1018–1021, Springer.

[7] C. Jacquemin, B. Planes, and R. Ajaj, "Shadow casting for soft and engaging immersion in augmented virtuality artworks," in *Proc. of the 15th Int. Conf. on Multimedia*, New York, USA, 2007, pp. 477–480.

[8] J. Blinn and M. E. Newell, "Texture and reflection in computer generated images," *Commun. ACM*, vol. 19, no. 10, pp. 542–547, 1976.

[9] A. Pomi, G. Marmitt, I. Wald, and P. Slusallek, "Streaming Video Textures for Mixed Reality Applications in Interactive Ray Tracing Environments," in *Vision, Modelling and Visualization 2003 (VMV)*, 2003.

[10] V. Popescu, C. Mei, J. Dauble, and E. Sacks, "Reflected-scene impostors for realistic reflections at interactive rates.," *Comput. Graph. Forum*, vol. 25, no. 3, pp. 313–322, 2006.

[11] T. Umenhoffer, G. Patow, and L. Szirmay-Kalos, "Robust multiple specular reflections and refractions," in *GPU-Gems 3*, pp. 387–407. Addison-Wesley, 2007.