An Intuitive Tool For Outlining Objects in Video Sequences: Applications To Augmented And Diminished Reality

Vincent Lepetit and Marie-Odile Berger

LORIA-INRIA Lorraine
E-mail: {lepetit,berger}@loria.fr

Abstract

In this paper, we present an efficient and easy to use method for outlining objects over a sequence. Applications of this algorithm are presented and special emphasis is made on diminished reality applications. Video sequences of our results can be seen at URL http://www.loria.fr/~lepetit/ismr2001.html
Key words: *Augmented reality, diminished reality, image-based rendering.*

1 Introduction

The objective of augmented reality (AR) is to add virtual objects to real video sequences in such a manner as to appear part of the 3D world. On the contrary, Diminished Reality (DR) aims at removing some objects from real video sequences. Applications include special effects for broadcast industry and also urban design when city planners want to visualize what the landscape will look like when a new construction is built, removed or replaced in the real scene. For AR, realistic image composition requires that the augmented pattern be correctly occluded by foreground objects. For DR, the part of the scene occluded by the removed object must be computed. For both cases, the occluding objects or the object to be removed must be outlined accurately over the sequence to allow realistic image composition. In this paper, we therefore propose an intuitive and efficient semi-interactive method for outlining objects over a sequence.

2 Outlining objects in video sequences

Theoretically, outlining objects in a sequence can be considered as a 2D process. However, especially in cluttered environments, tracking methods are not reliable and lack of accuracy. As computer vision methods now allow to compute the camera viewpoint over the sequence, 3D stereo-based reconstruction methods can be considered to segment the scene and to detect occluding objects. Unfortunately, uncertainties on the computed viewpoint or matching errors produce possibly large reconstruction errors which result in an inacurrate detection of the object over the frames. In order to overcome these problems, Ong [2] proposed a semi interactive approach to build the 3D model: the occluding objects are segmented by hand in selected views called key-frames. However, due to the uncertainty on the computed interframe motion, the recovered 3D shape does not project exactly onto the occluding objects in the key-frames nor in the intermediate frames.

In this paper, we also use the concept of key-frames but we do not attempt to build the 3D model of the occluding objects from all the key-frames. Figure 1 explains the way we compute a first estimation of the 2D object boundary in each frame of the sequence. First, the user points out a small number of key-frames which correspond to views where aspect changes of the occluding object occur. The user also outlines the occluding object on these key-frames. The starting point for our method is to build a good approximation of 3D occluding boundary which will be used for all the frames between two key-views. This 3D curve is built using stereo-triangulation from the two silhouettes outlined by the user. The projection of this approximated occluding boundary on the intermediate frames thus provides a fair estimation of the 2D occluding boundary (Fig. 1.a and b). A refinement stage is then needed to recover the actual 2D occluding boundary from the predicted one. To this aim, we use a deformable region-based matching scheme. Further details on this algorithms along with AR applications can be found in [1].

3 Diminished Reality

In this section, we especially consider diminished reality applications. Once the object to be removed has been outlined through the sequence, we have to fill in the mask with the background scene. To this aim, we use two strategies: if the removed object is thin, interpolation on image intensity or RGB components is used to fill in the mask; otherwise, the background scene is inferred from the whole sequence using 3D reconstruction.

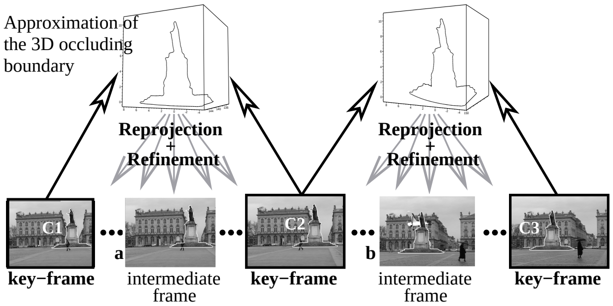

Figure 1 Overview.

3.1 Fill in the mask using 3D reconstruction

For large objects, the interpolation approach is untractable. Instead, a piecewise planar reconstruction of the scene is achieved from interest points. Using standard computer vision algorithms, interest points can be extracted to sub-pixel accuracy and tracked through the sequence. For DR, we only consider the interest points which do not belong to the object we want to remove (this can be done automatically since the object is outlined in all the frames). This way, the 3D reconstruction of the scene without the object to be removed can be achieved using stereo algorithms with the computed viewpoints.

Let us now explain how the mask is filled in. In each image, we consider the projection of the set of 3D points. This set of 2D points is then triangulated using Delaunay algorithm. To fill in the mask, we only consider triangles which have a non empty intersection with the mask. Let t_1, t_2, t_3 be such a triangle and $T_1T_2T_3$ the corresponding triangle in the 3D set. The triangle t_1, t_2, t_3 is textured by selecting automatically the nearest image from the sequence in which $T_1T_2T_3$ is visible. Mapping texture is realized via an homographic transformation.

Sometimes, the triangulation obtained from the interest points does not completely overlap the mask. Besides the obvious case where the considered sequence does not allow to build the background of the object, this can also arise when interest points cannot be extracted in some parts of the images. Consider for instance Fig. 2. Here, interest points cannot be detected in the sky. This results in a triangulation that does not overlap the top of the head. Such problems generally concern a small area of the object. So interpolation is used to fill in these triangles.

3.2 Results

Our algorithm is demonstrated on the Stanislas square sequence (150 frames). Here, we aim at removing the Stanislas statue. Two key-views are used and the user-defined silhouettes are presented in Fig. 2.a. The statue is then automatically outlined in the in-between frames (Fig. 2.b-c). The mask is then filled in using the triangulation method (Fig. 2.d). Finally, the diminished image is shown in Fig. 2.f.

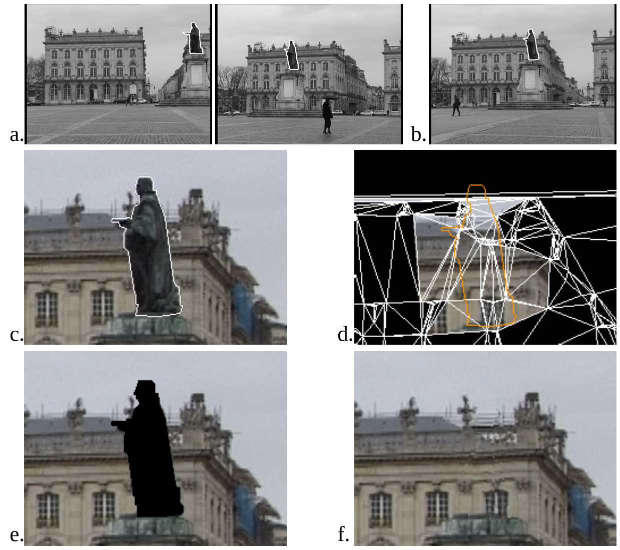

Figure 2 The Stanislas square sequence (150 frames): a. The two key-frames; b. an intermediate frame; c. the automatically outlined object in the intermediate frame; d. triangulation; e. mask of the removed object; f. final result.

Note that other results and full videos of diminished sequences are available at the URL http://www.loria.fr/~lepetit/ismr2001.html.

4 Conclusion

We have presented a general method for outlining objects over a sequence. The key concept is that accurate detection of the object can be obtained with moderate user interaction. This method can be of great interest for various applications in mixed reality. Besides augmented and diminished reality, this method can be used to perform special effects on the detected object (as colorisation) and can also be used for video composition (an object can be extracted in a video sequence and added on another one).

References

[1] V. Lepetit and M.-O. Berger. A semi automatic method for resolving occlusions in augmented reality. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina (USA)*, June 2000.

[2] K. C. Ong, H. C. Teh, and T. S. Tan. Resolving Occlusion in Image Sequence Made Easy. *The Visual Computer*, 14:153–165, 1998.