Online camera pose estimation in partially known and dynamic scenes

Gabriele Bleser, Harald Wuest, Mario Becker, Didier Stricker
{gbleser, hwest, beckerm, stricker}@igd.fhg.de
Department of Virtual and Augmented Reality
Fraunhofer IGD
Darmstadt, Germany

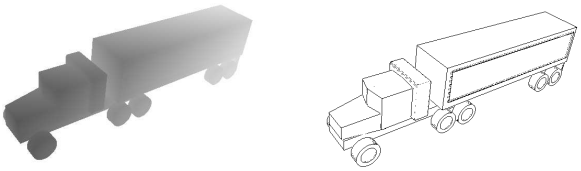

Figure 1: Z-buffer image (left) and polygonal model as result of the automatic 3D line extraction (right).

1 Demo description

Most of the markerless tracking systems nowadays require preprocessing steps like the calibration of reference images, the generation of keypoint databases or the construction of geometric models. Moreover, the camera movements are often restricted to the modeled parts of the environment. Since real world environments are in general not static, end users are not willing to spend much time on preparing their scenarios and existing geometric models are often not complete or not consistent, tracking systems relying mostly upon offline data are not always applicable.
We demonstrate a markerless camera tracking system that does not depend on any preprocessed data and needs only minimal 3D knowledge about the target scene [1]. A CAD model of one object in the scene is used to draw the initial connection between real and virtual world. The initialization is semi-automatic. A line model is generated automatically by rendering the given CAD model from a pose that puts the object in front of the camera and by extracting 3D contours from the z-buffer (cf. figure 1). This line model is then registered onto the image gradient for finding the initial pose. The camera therefore has to be moved in such a way that the projected lines are close to the appearance of the object in the video image (cf. figure 2).
If the initial pose has been found, point features are detected and tracked from frame to frame using a brightness invariant template matching algorithm. Features, which lie on the CAD model, are back-projected directly for obtaining 3D coordinates. Other features are triangulated after they have been seen in a sufficient number of views. As uncertainty is present in all computer vision steps,

Figure 2: Registration of the line model onto the image.

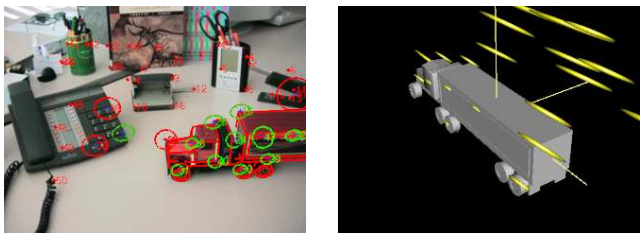

Figure 3: Uncertainties of the 3D feature locations projected into image space (left) and in 3D space from an external view (right).

the triangulated 3D locations of the features are first regarded as rough estimates of the real 3D locations, which are then refined everytime the feature is visible in the camera image. The evolution of a feature is usually as follows: A corner pixel is detected in a camera image and tracked over some frames. If a sufficient baseline has been reached, the feature is triangulated resulting in a 3D coordinate with a big uncertainty region. Every further observation of the feature is used to update its position and to shrink its uncertainty region. Figure 3 shows features in different reconstruction states.
Besides a convenient usage, there are some more key requirements, which are essential for the applicability of a tracking system in real world. This is basically precision, absence of drift and robustness against rapid camera movements, severe occlusion, changes of light and small modifications in the scene. In the system presented here, these issues are already handled on the 2D feature tracking layer and again in the subsequent geometry reconstruction processes. Therefore it is very robust, even if one removes the initializing object (cf. figure 4).

References

[1] Gabriele Bleser, Harald Wuest, and Didier Stricker. Online camera pose estimation in partially known and dynamic scenes. In *Proc. of International Symposium on Mixed and Augmented Reality (ISMAR), to appear*, 2006.

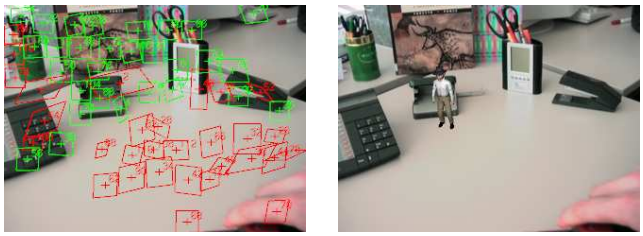

Figure 4: The removal of the toy truck is detected on the 2D feature tracking level. The patches are marked as invalid (red), while the augmentation remains stable.