Robust Estimation of Camera Parameters Pan, Tilt and Zoom for Integration of Virtual Objects into Video Sequences

Broszio, H., Grau, O.

Laboratorium fuer Informationstechnologie, University of Hannover
Schneiderberg 32, 30167 Hannover, Germany
Email: broszio@tnt.uni-hannover.de, grau@tnt.uni-hannover.de

ABSTRACT

The integration of virtual objects into an image sequence captured by a moving camera is of high interest for special effects in tv- and film-productions. For the composition, the camera parameters of the real camera are estimated and transferred to a virtual camera. The estimation accuracy is very critical, because even small estimation errors can be recognised as localisation jitter of the inserted virtual objects. This contribution presents a robust estimation technique for computation of camera parameters pan, tilt and zoom from an image sequence. Based on conventional feature detection and correspondence analysis, the robustness of the approach is achieved by using a highly efficient outlier detection of misaligned correspondences.

1. INTRODUCTION

The integration of virtual objects into an image sequence of a moving camera is of high interest for special effects in tv- and film- productions. A virtual camera generates a synthetic image of a virtual object which is mixed into the real camera image. The goal is the creation of a new scene with the illusion that the virtual object was already in the scene taken by the real camera. To make this illusion perfect, the parameters of the real camera are estimated and those of the virtual camera are adjusted accordingly.

The estimation of the camera parameters from the real image sequence is very critical, because even small estimation errors can be recognised as localisation jitter of the inserted virtual objects. To avoid this jitter a highly accurate estimation is needed. Further the approach must be robust against violations of the signal model that supposes a mainly static scene. The most important violations are due to noise in the camera images, moving objects and changing illumination in the scene.

The estimation should also be able to cope with long video sequences. For camera pans with large angles this also causes large displacements and completely new image content.

A number of the approaches for the estimation of camera parameters were developed for video coding purposes (like [3, 5]). Hoetter [3] proposed a differential approach that is directly based on the image luminance information and was also applied to a virtual studio application [7]. Due to the assumptions made and linearisations, the approach has a small convergence range and is not able to cope with large panning angles of the camera.

In contrast to the direct approaches for camera parameter estimation the indirect methods first compute a corresponding feature set or a displacement vector field. The camera parameters are then estimated in a second step based on this information (e.g. [5, 6]). A displacement vector field typically assigns a displacement to each image pixel, a corresponding feature set consists only of a number of features (most commonly corners) with a displacement according to the next image of the sequence. The difficulties in these approaches arise due to inaccurate or unreliable displacement information. Park [6] proposes a robust algorithm that detects erroneous displacements based on residual analysis in a least square approach that computes the camera parameters.

The new approach presented here uses an indirect computation of the camera parameters. In a first step corners are detected for each image. Then correspondences between features of an image to the next in the sequence are established. The final parameter estimation employs the random sampling procedure [1] that was successfully applied to a self-calibration for scene reconstruction [4]. The robustness of the presented approach is further increased using information about the image sequence.

The next section describes the proposed estimation algorithm in more detail. Section 3 gives some results and the paper ends with some conclusions.

2. ROBUST ESTIMATION OF CAMERA PARAMETERS PAN, TILT AND ZOOM

The estimation method consists of three processing steps: extraction of corners, determination of correspondences and robust estimation of the camera parameters.

2.1. Detection of Corners

The corners are detected with subpixel accuracy using the Harris corner detector [2]. For each image of the sequence a list of corner coordinates $L_n = \{\vec{p}_{n,1}, \ldots, \vec{p}_{n,i}, \ldots, \vec{p}_{n,N}\}$ is extracted, where $\vec{p}_{n,i} = (x, y)^T$ are the image coordinates of a corner i in the image n.

2.2. Correspondence Analysis

The corners of the list L_{n-1} of image $n-1$ are assigned to the corners L_n of the next image n by measuring cross-correlation between 5×5 windows surrounding the corners. The correspondences are established for those corners which have the highest cross-correlation. This results in a list of correspondences $C_n = \{c_{n,1}, \ldots, c_{n,j}, \ldots, c_{n,M}\}$ with correspondences $c_{n,j} = (\vec{p}_{n-1,j}, \vec{p}_{n,j})$.

2.3. Motion Model

Assuming a camera with fixed position, the proposed algorithm to estimate pan φ, tilt θ and zoom s is based on the following motion model. The camera projections with focal length f_{n-1} for image $n-1$ and f_n for image n after a camera operation are illustrated for pan φ in Fig. 1. For the projected point $P(X, Y, Z)$ the relationship between a corner located at (x_n, y_n) in image n and (x_{n-1}, y_{n-1}) in image $n-1$ is pictured. The image coordinates can be described by the angles (α_n, β_n) and $(\alpha_{n-1}, \beta_{n-1})$. Using angles, the camera motion can be expressed as:

$$\alpha_{n-1} = \alpha'_n + \varphi \tag{1}$$

$$\beta_{n-1} = \beta'_n + \theta \tag{2}$$

$$\frac{f_{n-1}}{f_n} = \frac{\tan \alpha'_n}{\tan \alpha_n} \tag{3}$$

$$\frac{f_{n-1}}{f_n} = \frac{\tan \beta'_n}{\tan \beta_n} \tag{4}$$

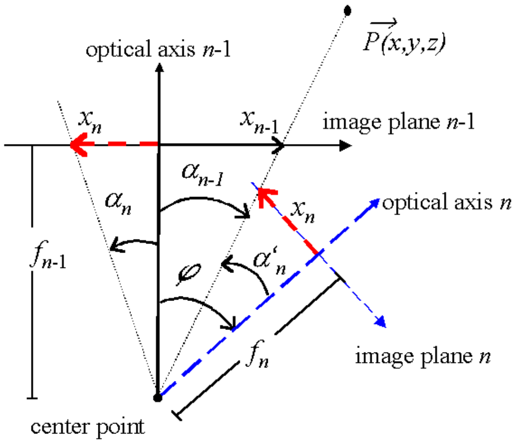

Figure 1: Rotation of camera by pan φ and variation of the focal length f cause the movement of corner x_{n-1} to x_n.

Assuming small angles $\alpha'_n, \alpha_n, \beta'_n$ and β_n, which is valid for corner coordinates near the optical axis, (3) and (4) can be approximated:

$$\frac{f_{n-1}}{f_n} = \frac{\alpha'_n}{\alpha_n} \tag{5}$$

$$\frac{f_{n-1}}{f_n} = \frac{\beta'_n}{\beta_n} \tag{6}$$

With $s = f_{n-1}/f_n$ the parameter can be calculated from two correspondences:

$$\alpha_{n-1} = \alpha_n s + \varphi \tag{7}$$

$$\beta_{n-1} = \beta_n s + \theta \tag{8}$$

2.4. Robust Estimation

Due to misalignments, usually some of the correspondences are incorrect. To achieve a robust matching, a random sampling algorithm [1] for outlier detection is employed. Two correspondences are used to estimate the camera parameters. Then the support of this parameter set is measured by evaluating how good the displacements of the other correspondences can be expressed by this parameter set. If it doesn't match the majority of the correspondences, a new parameter set is choosen randomly until a satisfying solution is found. Finally the parameters are estimated using a least squares technique applied to the correspondences found.

The estimation of camera parameters for long video sequences is critical concerning the accumulation of estimation errors. In difference to

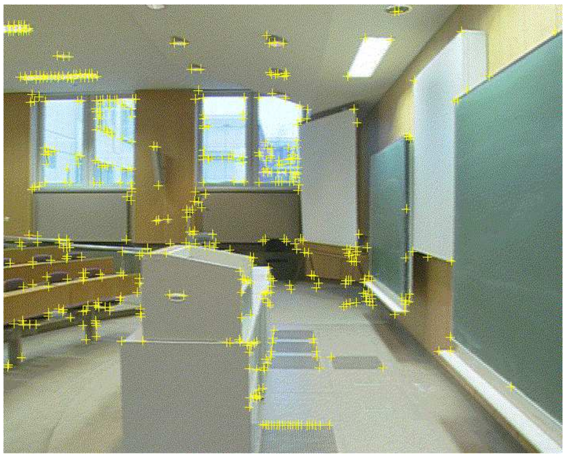

Figure 2: Extracted corners from an image of the sequence "lecture room".

other algorithms where the parameters are estimated from image to image, the suggested approach estimates the parameters of each image of the sequence in relation to a global coordinate system. Therefore all measured corners of a valid correspondence are registered into the global coordinate system. Thus, the estimation of a new camera view is relative to the global coordinate system.

This ensures that for images captured under large camera rotation angles the integrated virtual object will always appear at the correct place related to the real scene.

3. RESULTS

Fig. 2 shows the result of corner detection with 466 detected corners in the 1st image of the sequence "lecture room". For this corner set 326 correspondences were calculated (Fig. 3). From these, the applied random sampling algorithm selects 226 supported correspondences (Fig. 4), that were used to estimate camera parameters. With the estimated parameters a virtual object is composed into the real scene (Fig. 5).

Fig. 6 shows the difference between an image $n-1$ of the sequence and the compensated image n. The compensation is calculated by applying the estimated motion to each pixel of the image $n-1$. The image difference between the transformed image and the nth image shows the compensation error for a rotation $\varphi = 1.06$ degree, $\theta = 0.005$ degree and a zoom $s = 1.002$.

4. CONCLUSIONS

A robust estimation technique for computation of camera parameters pan, tilt and zoom from

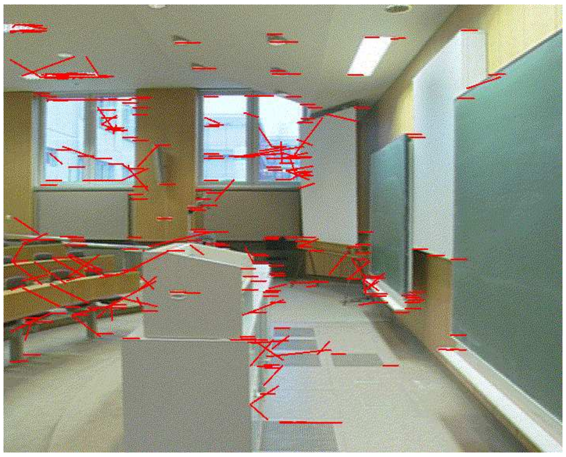

Figure 3: Displacement vectors of corresponding corners between two images.

Figure 4: Displacement vectors of the remaining correspondences after outlier elimination.

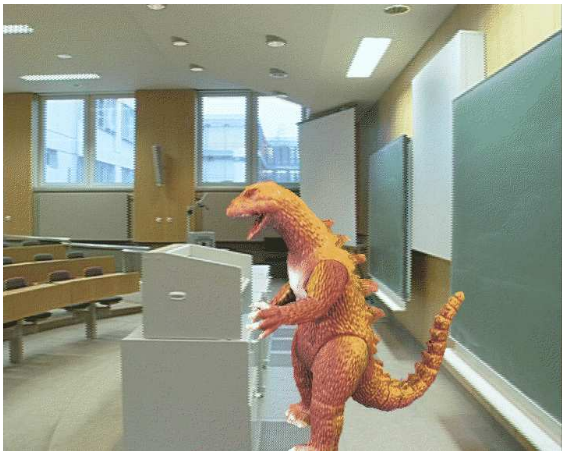

Figure 5: A virtual object is composed into the image sequence "lecture room".

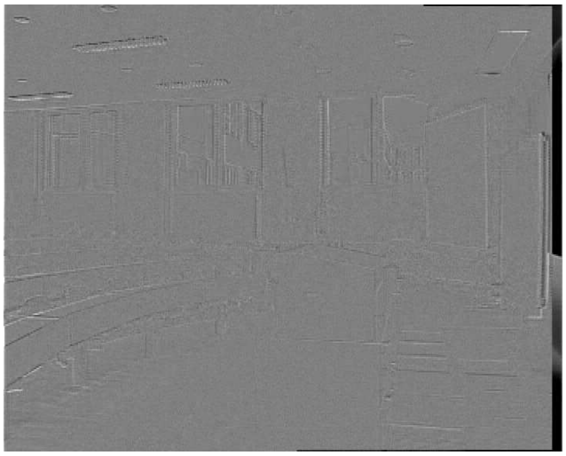

Figure 6: Difference of image n and the compensated image $n-1$.

an image sequence was presented. Based on conventional feature detection and correspondence analysis, the robustness of the approach is archieved by the use of a highly efficient outlier detection of misaligned correspondences.

The parameter estimator is based on a least square technique that computes very accurate parameters for pan, tilt and zoom of the camera. A virtual object occurs with minimal jitter in a composed image sequence. The difference of a compensated image pair shows that the errors are quite small (about 0.5 to 2 pel). The remaining errors are systematicly due to violation of the simple motion model. The main influence was found in experiments with different tv-cameras: the optical center of the camera is not located in the axis of rotation of the camera tripod. The second important influence are the radial distortions of the camera lens.

The influence of the different rotation axes can be minimised with professional tripods by adjusting the camera postition mechanicaly. The influence of the radial distortion can be avoided only by the use of high class optics. In the case of high quality equipment as found in tv- productions, the proposed approach gives promissing results. In the considered test sequences the errors due to model violation did not affect the realistic impression of the composed image sequence.

5. REFERENCES

[1] Fischler, M. A., Bolles, R. C., "Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography" *Commun. Assoc. Comp. Mach*, Vol. 24, p. 381-395, 1981.

[2] Harris C., Stephens, M., "A combined corner and edge detector" *In Proc. 4th Alvey Vision Conf.*, p. 147-151, 1988.

[3] Hoetter, M., "Differential esitmation of the global motion parameters zoom and pan" *Signal Processing*, Vol. 16, No. 3, March 1989.

[4] Koch, R., "Automatic 3D Model Acquisition from Uncalibrated Image Sequences" *Proc. of Computer Graphics International 1998 (CGI'98)*, June 22nd - 26th, 1998, Hannover, Germany.

[5] Park, J.-I., et al., "Estimation of Camera Parameters from Image Sequence for Model-Based Video Coding" *IEEE Tr. on CAS-VT*, Vol. 4, No. 3, June 1994.

[6] Park, J.-I., Lee, C. W., "Robust estimation of camera parameters from image sequence for video composition" *Signal Processing: Image Communication*, 9(1996).

[7] Thomas, G. A., "Global Motion Estimation for Registering Real and Synthetic Images" *European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Production*, 23.-24. Nov. 1994 Hamburg.