TOWARDS ARBITRARY CAMERA MOVEMENTS FOR IMAGE CUBE TRAJECTORY ANALYSIS

Ingo Feldmann, Peter Eisert, Peter Kauff

Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institute, Berlin, Germany
Email:{feldmann, eisert, kauff }@hhi.fhg.de

ABSTRACT

Image Cube Trajectory (ICT) Analysis is a new and robust method to estimate the 3D structure of a scene from a set of 2D images. In case of moving camera each 3D point is represented by a trajectory in an image cube. In previous publications we have shown that it is possible to reconstruct the 3D scene from the parameters of these trajectories. Nevertheless, the algorithm was restricted to simple parameterized camera movements. In this paper we will discuss the problem of arbitrary camera motion. We benefit from the fact that in many cases, even for hand-held cameras, the rotational variation of the estimated camera parameters is much higher than the translational variation. We will show that it is possible to compensate rotational and translational deviations by transformation and subsampling of the image cube. We obtain more uniform smooth trajectory structures which can be analyzed by standard ICT analysis algorithms.

1. INTRODUCTION

The estimation of depth information from 2D images has received much attention in the past decade. The basic problem of recovering the 3D structure of a scene from a set of images is the correspondence search [1]. Given a single point in one of the images its correspondences in the other images need to be detected. Depending on the algorithm two or more point correspondences as well as the camera geometry are used to estimate the depth of that point [2]. However, for complex real scenes the correspondence detection problem is still not fully solved. Especially, in the case of homogeneous regions, occlusions, or noise, it still faces many difficulties. It is now generally recognized that using more than two images can dramatically improve the quality of reconstruction.

One method for the simultaneous consideration of all available views is Epipolar Image (EPI) analysis [3]. An Epipolar Image can be thought of being a horizontal slice (or plane) in the so called *image cube* [1, 4] that can be constructed by collating all images of a sequence. It is defined for a linear equidistant camera movement parallel to the horizontal axis of the image plane only. In this case projections of 3D object points become straight lines called *EPI lines*. The principle of EPI analysis is the detection of all EPI-lines in all available EPIs. From the EPI-line parameters (slope and offset) either the related 3D point or the corresponding depth can be evaluated. The advantage is the joint detection of point correspondences for all available views. Occlusions as well as homogeneous regions can be handled efficiently [4]. The big disadvantage of the algorithm is its restriction to linear equidistant camera movements.

For non-linear parameterized camera movements 3D points are represented by more general trajectories in the 3D image cube. The EPI-line approach cannot be applied for this case. One idea to overcome the restrictions of EPI analysis is a piecewise linear analysis approach where small segments of the object point trajectory are approximated by straight lines [5]. Unfortunately, this reduces the amount of reference images and the robustness of the 3D reconstruction significantly.

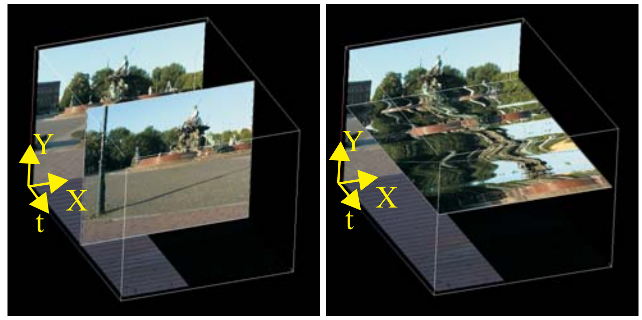

Fig. 1. Image cube representation of the 'Neptun'-sequence recorded with a hand-held camera, **left)** time slices (constant t) represent images, **right)**, horizontal slices (constant Y) illustrate the ICT structures

In [6, 7], we have introduced a new concept called *Image Cube Trajectory Analysis* that overcomes the restrictions of EPI analysis and is able to jointly exploit all available views for more general camera configurations. Without loss of generality we have demonstrated this technique at the example of a circularly moving camera.

The proposed ICT analysis method is a two step approach. In a first step the camera parameters are estimated such that all camera positions are known in advance. Robust camera self-calibration systems are well known in the literature [8]. In this way it is possible to parameterize the trajectory structures in the image cube. This information is used to derive an occlusion compatible ICT search order which is essential for robust ICT detection algorithms. Further, the camera parameters are used to define discrete ICT search spaces which can be optimized and adapted to the resolution of the captured images [9].

The second step is the ICT detection itself. Based on the defined occlusion compatible ordering scheme the existence of ICTs in the image cube is determined by evaluating the color constancy of the parameterized point trajectories. Further, analyzing the properties of the considered search space, it is possible to detect and handle homogeneous regions efficiently [10]. The result of the ICT analysis are the parameters of the detected point trajectories. These parameters can be transformed in a subsequent step to

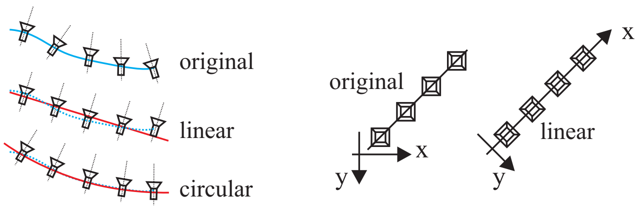

Fig. 2. Parameterized correction of camera rotation **left)** in the x-z-plane, **right)** in the x-y-plane

either image based depth maps or 3D voxel models.

One of the major differences of the ICT analysis method compared to other 3D multiview reconstruction approaches, such as, for example, the voxel coloring technique [11] is the analysis of the image-time domain straight away. For the evaluation of the structure of single 3D points (i.e. the ICTs) all available information of all existing images are analyzed simultaneously in a single processing step. In contrast to conventional methods the structure evaluation, occlusion handling as well as the masking of results is done in the image-time domain in the image cube.

In this paper we will discuss the problem of arbitrary non-parameterized camera movements. As an example, Fig. 1 left shows the image cube representation of the 'Neptun'-sequence which was recorded from a moving bicycle with a hand-held camera. The right-hand side of the figure illustrates the arbitrary non-parameterized ICT structure for this case. The goal of this paper is to preprocess the image cube in such a way that a more uniform parameterized trajectory structure can be derived which is necessary for the application of the proposed ICT analysis method.

The paper is structured as followed. Firstly, we discuss the problem of the parameterized correction of the camera rotation. We will show, that the rotational components of the cameras can be transformed to a virtually predefined camera setup without knowledge of depth information. In this way we extend conventional 2D image stabilization techniques to the case of a 3D camera path parameterization. Secondly, we will propose a method for a parameterized sub-sampling of the camera path to compensate variations of the camera translation. Finally, we will show some results to demonstrate the efficiency of the proposed algorithm.

2. PARAMETERIZED CORRECTION OF CAMERA ROTATION

The general idea of the parameterized correction of camera rotation is illustrated if Fig. 2. The left-hand side of the figure demonstrates the adaption of arbitrary camera rotations (top) to a predefined virtual setup (middle: linear, bottom: circular) in the x-z-domain. The same process can be applied to the other two axes of rotation as shown in the example on the right-hand side of the figure in the x-y-domain.

To solve the problem of parameterized correction of the camera rotation we benefit from two facts. Firstly, for many captured sequences the deviation of the rotational components of the camera motion is much higher than the deviation of the translational components. This holds even in the case of hand-held camera sequences as illustrated in Figs. 3 and 4. It can be seen that the translational components of the camera motion in the first figure are much more uniform than the rotational components in the second figure. Although this behavior cannot be generalized we have noticed that it still holds for many real life scenarios.

The second advantage of rotational deviations is that they can

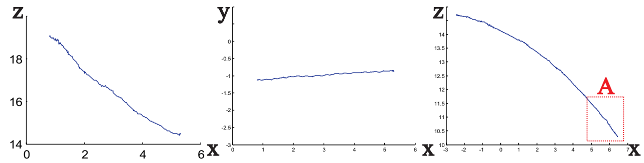

Fig. 3. Estimated camera translation, **left)** 'Park'-sequence, x-z-domain, **middle)** x-y-domain, **right)** 'Neptun'-sequence, x-z-domain

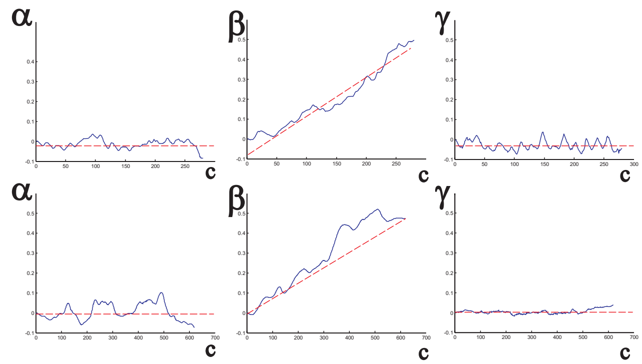

Fig. 4. Estimated camera rotation in terms of the rotation angles α, β, γ about the axes of the coordinate system x, y, z in dependency of the camera frame c. Parameterized correction of the rotational components to predefined virtual camera rotations (dashed line), **top row)** 'Park'-sequence, **bottom row)** 'Neptun'-sequence

be compensated without explicit knowledge about depth. Using homogeneous coordinates this can be derived easily from the projection of a 4D point $\mathbf{x}$ to the image planes of two cameras $\mathbf{X}$ and $\mathbf{X}'$

$$\mathbf{X} = \mathbf{P}\mathbf{x}$$
$$\mathbf{X}' = \mathbf{P}\mathbf{x}' \tag{1}$$

where $\mathbf{P}$ is the homogeneous projection matrix

$$\mathbf{P} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{2}$$

and f stands for the focal length of the cameras. The relation of the cameras to each other can be expressed by the the transformation

$$\mathbf{x}' = \mathbf{D}\mathbf{x} \tag{3}$$

where matrix $\mathbf{D}$

$$\mathbf{D} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0_3 & 1 \end{bmatrix} \tag{4}$$

is the homogeneous transformation between the cameras with the rotation matrix $\mathbf{R}$ and the translation vector $\mathbf{t}$.

From equations 1-4 follows for the components of the two images (X, Y) and (X', Y')

$$X' = \frac{r_{11}X + r_{12}Y + fr_{13} + t_x/z}{r_{31}X + r_{32}Y + fr_{33} + t_z/z}$$

$$Y' = \frac{r_{21}X + r_{22}Y + fr_{23} + t_y/z}{r_{31}X + r_{32}Y + fr_{33} + t_z/z} \tag{5}$$

Equation 5 shows that the displacement of an image point (X, Y) to the image plane (X', Y') is independent from the pixels depth z

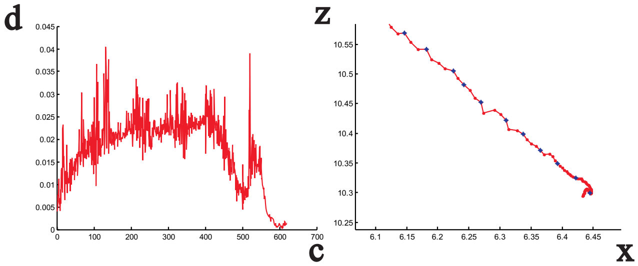

Fig. 5. 'Neptun'-sequence, **left)** camera sample density d evaluated by the Euclidic distance between subsequent translational camera samples, **right)** example for parameterized subsampling (dots) of the estimated camera sample path, reference: region A, Fig. 3, right

if there are no translational components **t** between the cameras. In other words, the rotational components of a camera can be adapted to arbitrary virtual parameterizations as long as they point roughly in the original direction (otherwise there is no common image information).

Based on this rule it is possible to adapt the components of camera rotation to pre-defined simple uniform parameterizations by transforming the original image planes of the camera. This is illustrated in Fig.4 for the example of the 'Park'- and the 'Neptun'-sequence. The rotational camera motion is expressed in terms of rotation angles α, β, γ around the axes of the coordinate system x, y, z in dependency of the camera sample c. The dashed lines in the figure show the chosen virtual parameterization reference. While the angles α and γ are kept at a constant level (i.e. no rotation) the angular rotational component β is transformed to a constantly increasing value. In practice the chosen parameterization corresponds to a constant (rotational) camera motion around the y axis.

3. PARAMETRIZED SUBSAMPLING OF CAMERA PATH

The second assumption of this paper is that many captured sequences contain much more images than finally needed for the ICT analysis. Even for short sequences it is not unusual to capture at least 500 or 1000 frames with a common video camera, in many cases even much more. From this assumption we have derived rules for a parameterized subsampling of the estimated camera path.

The proposed process is shown in Fig.5 at the example of the 'Neptun'-sequence. The left-hand side of the figure illustrates the camera sample density d evaluated by the Euclidic distance between subsequent translational camera samples. The variation of d results in a non-uniform ICT structure. To obtain a uniform parameterized ICT shape the cameras are subsampled, i.e. only those cameras are selected which provide a nearly uniform Euclidic distance as demonstrated on the right-hand side of the figure. In this way, it is possible to create almost uniform ICT structures even in case of highly non-uniform camera motion. Again, it is not possible to generalize this approach to all camera movements. Nevertheless, in practice it still holds for many different sequences. Note, that for complex camera paths the sequence might be split to smaller parts which are processed separately in the proposed way.

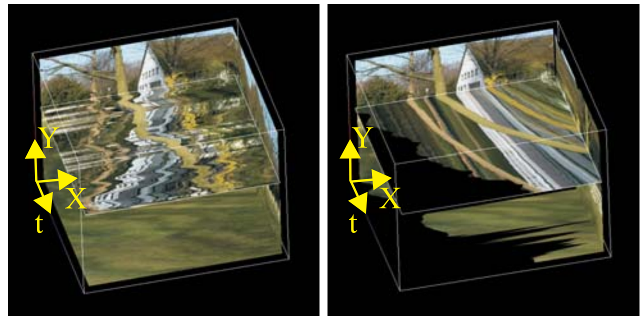

Fig. 6. Image cube representation of the 'Park'-sequence' **left)** before parameterized correction, **right)** after parameterized correction

4. RESULTS

In this section we will show several results of the proposed algorithms. We refer to three sequences. The 'Neptun'-sequence was captured from a moving bicycle with a hand-held camera. The bicycle was performing a circular motion while pointing with the camera to the center of the circle. The 'Park'-sequence was captured by a person walking through a park using a hand-held camera too. The translational motion was rather linear then circular whereas the rotational components occurred to be pointing to a common center too (see Fig. 4). Finally, the 'House'-sequence was captured from a moving car. It was aimed to keep the rotation as well as the translation components as uniform and linear as possible. But still deviations occurred due to the shaking and the velocity changes of the car.

A result for the parameterized correction of camera rotation and the parameterized subsampling of the camera path is illustrated in Fig.6 for the example of the 'Park'-sequence. To evaluate the structure of of the corrected ICTs horizontal image cube slices are shown in Fig.7. The top row of the figure demonstrates the corrected ICT structures for the 'Neptun'-sequence (see Fig. 1) and the bottom row for the 'Park'-sequence (see Fig. 6). Note, that due to the non-linear camera movement the point trajectories are still situated on non-planar surfaces within the image cube rather than on the illustrated horizontal planes (see [6]). But still, the figures illustrate the structure of the ICTs in a sufficient way. In both cases it can be seen that even in case of heavy distortions in the original image cubes an almost uniform ICT structure could be obtained.

Figures 8 and 9 demonstrate the result of a standard ICT 3D reconstruction algorithm applied to the parameterized and subsampled image sequences at the example of the 'Park'- and the 'House'-sequence. In both cases a 3D model of the scene as well as the corresponding depth maps could be reconstructed with good quality.

5. CONCLUSIONS

In this paper we have discussed the application of the Image Cube trajectory analysis technique to arbitrary camera movements. We have shown that it is possible correct deviations of the rotational components of the camera motion in order to adapt the camera movement to a uniform parameterized virtual reference path. Deviations of the translational components of the camera motion can be reduced by camera path subsampling. In order to apply standard ICT analysis techniques we have shown that the proposed methods

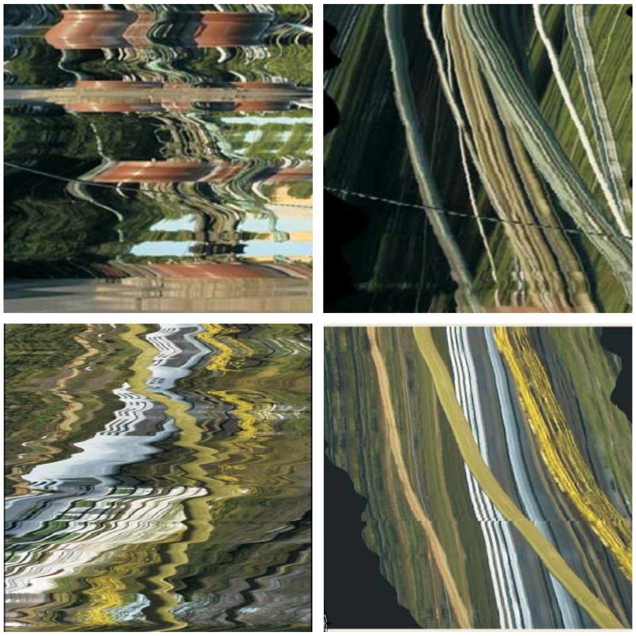

Fig. 7. Horizontal slices in the image cube illustrate the structure of the ICTs, **left column)** before parameterized correction, **right column)** afterwards. **top row)** 'Neptun'-sequence, **bottom row)** 'Park'-sequence.

can be applied to different kinds of camera movements. Even in case of hand-held cameras in many cases an almost uniform trajectory structure can be achieved which can be approximated by simple parameterization models, such as linear or circular camera movements. We have demonstrated the efficiency of the proposed algorithms for the 3D structure estimation on several examples.

6. ACKNOWLEDGMENT

The work presented in this paper has been developed with the support of the European Network of Excellence VISNET (IST Contract 506946).

7. REFERENCES

[1] J. P. Mellor, S. Teller, and T. Lozano-Perez, "Dense depth maps from epipolar images," Tech. Rep. AIM-1593, 1996.

[2] P. Beardsley, P. Torr, and A. Zisserman, "3D model acquisition from extended image sequences," in *Proc. European Conference on Computer Vision (ECCV)*, 1996.

[3] R. C. Bolles, H. H. Baker, and D. H. Marimont, "Epipolar image analysis: An approach to determine structure from motion," *International Journal of Computer Vision*, pp. 7–55, 1987.

[4] A. Criminisi, S. B. Kang, R. Swaminathan, R. Szeliski, and P. Anandan, "Extracting layers and analyzing their specular properties using epipolar-plane-image analysis," Tech. Rep. MSR-TR-2002-19, Microsoft Research, 2002.

[5] Y. Li, C.-K. Tang, and H.-Y. Shum, "Efficient dense depth estimation from dense multiperspective panoramas," in *Proc. International Conference on Computer Vision (ICCV)*, Vancouver, B.C., Canada, Jul. 2001, pp. 119–126.

[6] I. Feldmann, P. Eisert, and P. Kauff, "Extension of epipolar image analysis to circular camera movements," in *Proc. International Conference on Image Processing (ICIP)*, Barcelona, Spain, Sep. 2003, pp. 697–700.

Fig. 8. 3D reconstruction of the 'House'-sequence, **left)** original image of the sequence, **middle)** reconstructed 3D model, **right)** reconstructed depth map

Fig. 9. 3D reconstruction of the 'Park'-sequence, **left)** original image of the sequence, **middle)** reconstructed 3D model, **right)** reconstructed depth map

[7] I. Feldmann, P. Kauff, and P. Eisert, "Image cube trajectory analysis for 3D reconstruction of concentric mosaics," in *Proc. Workshop on Vision, Modeling, and Visualization*, Munich, Germany, Nov. 2003, pp. 569–576.

[8] C. Rothwell, G. Csurka, and O. D. Faugeras., "A comparison of projective reconstruction methods for pairs of views," Tech. Rep. 2538, INRIA, Jul. 1995.

[9] I. Feldmann, P. Kauff, and P. Eisert, "Optimized space sampling for circular image cube trajectory analysis," in *Proc. International Conference on Image Processing (ICIP)*, Singapore, Singapore, 2004, p. 127.

[10] I. Feldmann, M.Heinrichs, P. Kauff, and P. Eisert, "Handling of homogeneous regions for image cube trajectory analysis," in *Proc. International Workshop on Image Analysis*, Montreux, Switzerland, Apr. 2005.

[11] S. M. Seitz and C. R. Dyer, "Photorealistic scene reconstruction by voxel coloring," in *Proc. Computer Vision and Pattern Recognition*, Puerto Rico, 1997, pp. 1067–1073.