HANDLING OF HOMOGENEOUS REGIONS FOR IMAGE CUBE TRAJECTORY ANALYSIS

Ingo Feldmann, Matthias Heinrichs, Peter Kauff, Peter Eisert

Image Processing Department
Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institute
Einsteinufer 37, 10587 Berlin, Germany
Email:{feldmann, heinrich, kauff, eisert}@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. The motion of points in 3D space is represented by trajectories in an image cube. The advantage of this method is that the motion information of a single 3D point can be jointly represented and analyzed for all available views of an image sequence. In this paper we will discuss the problem of 3D structure estimation in the case of homogeneous image regions. We will present a new approach for detection and compensation of depth artifacts which is based on the construction of a particular search space. Homogeneous regions in the image cube cause special patterns in this space. They can be detected by standard image processing methods. At the simple example of a linear moving camera we demonstrate the efficiency of the proposed algorithm.

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 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 (see fig. 1 for an example). It is defined for a linear equidistant camera movement parallel to the horizontal axis of the image plane. In this case projections of 3D object points become a straight line called *EPI-line*. 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 corresponding depth can be evaluated. The advantage is the joint detection of point correspondences for all available views. Occlusions can be handled efficiently [4]. The big disadvantage of the algorithm is its restriction to linear equidistant camera movements.

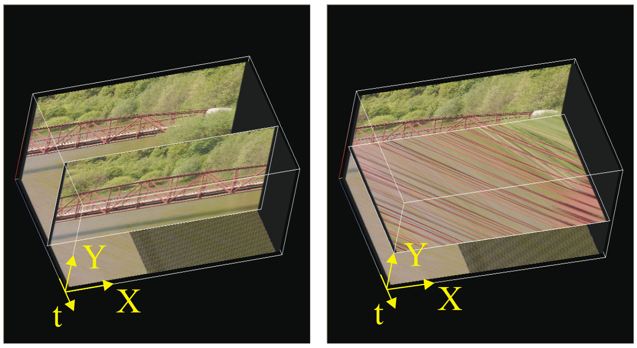

Fig. 1. Image cube representation of the 'Bridge' sequence, **left)** time slices (constant t) represent images, **right)**, horizontal slices (constant Y) represent Epipolar Images

2. IMAGE CUBE TRAJECTORY ANALYSIS

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. In [5, 6], we have proposed a new concept called *Image Cube Trajectory (ICT) analysis* that overcomes the restriction of EPI analysis and is able to jointly exploit all available views for more general camera configurations. The basic approach is inverse to the conventional way of EPI analysis where in a first step the EPI lines are detected by standard segmentation algorithms (edge detection filters etc., see [3, 7]) and structure information is computed afterwards from the line parameters. In contrast, we choose a hypothetical 3D point first and project it into the image cube space where it is represented by its trajectory. This hypothesis is checked in a second step by evaluating color constancy along the entire trajectory [5]. The best matching ICTs are considered to belong to the object surface. This inverse approach allows the consideration of arbitrary camera movements but requires the decision about a 3D search strategy. This problem was addressed in [8]. The authors derived a non-uniform sampling that minimizes the number of ICT evaluations for a given camera resolution while maintaining optimal image resolution.

At this point, it is important to consider the occlusion compatible ordering of the point trajectories. It is described in [4] for the case of a linearly moving camera . 3D points which are closer to the cameras 'move' faster (i.e. have larger disparities in the cam-

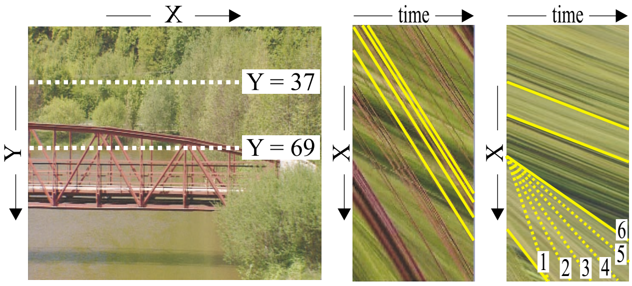

Fig. 2. 'Bridge' sequence, **left)** sample image, **middle)** EPI of line Y=69, trajectories with higher slope are detected first and masked out for subsequent analysis **right)** EPI of line Y=37, problem of homogeneous regions, due to occlusion compatible search order the trajectories will be detected in the illustrated order

era image planes) than those which are further away. Therefore, an ICT of a point with a small depth will have a larger slope than that of a more distant point. Because closer 3D points always hide those which are further away, the ICTs with larger slopes always hide those with smaller slopes. Fig.2 middle illustrates this at the example of the 'Bridge' sequence. Due to the occlusion ordering the ICTs with the largest slope need to be detected first. Successfully detected ICTs are masked out and excluded from subsequent analysis steps. In this way, the algorithm is able to detect 3D points in the scene scanning it from front to back. 3D points at the same depth but different horizontal and vertical positions do not occlude each other. They can be detected in arbitrary order.

3. HOMOGENEOUS REGIONS

One of the remaining problems is the detection of homogeneous regions. In the image cube such regions cause tubes of nearly constant color which occur as stripes in the EPI [4]. Fig. 2, right demonstrates this problem. The two marked regions belong to parts in the background of the scene with nearly constant color. The proposed conventional ICT search algorithm [5] cannot differentiate between the set of possible ICTs inside such stripes. Applying the above mentioned occlusion compatible search scheme the ICTs will be detected in the illustrated order of fig.2, right. The first detected ICT will be that with the highest slope inside the region. It is followed by the ICT with the next smaller slope etc. In this way the detected trajectories differ from the original scene geometry. This causes artifacts in the reconstructed depth or 3D structure.

A straight forward approach to handle this problem is shown in [3]. The authors consider the EPI as a set of regions which are analyzed by standard segmentation tools, such as edge detection filters etc. This introduces the problem of over or under segmentation which requires very precise parameter adjustments. In this way it is very difficult to obtain suitable results for more complex scenes, for example in natural environments containing multiple occlusions etc. Further, this technique is restricted to linear camera movement.

In this paper we will introduce a more generalized new method for handling of homogeneous regions. It is based on the construction of a special trajectory parameter space representation which we call *search space*. The advantages of the search space representation are its special properties. Such homogeneous regions

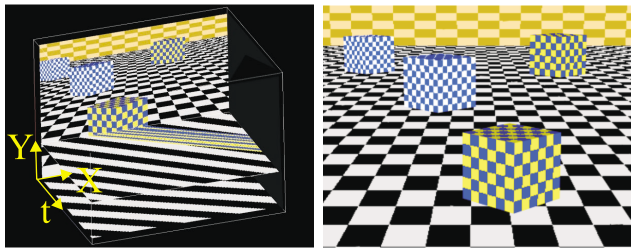

Fig. 3. Sample of the virtual scene 'Boxes' containing highly homogeneous regions, **left)** image cube representation, **right)** sample image

have well defined patterns. Further, the point trajectory representation has dual properties to the EPI. This introduces the possibility to develop and apply new and efficient algorithms. To introduce the basic principle of the proposed method we will restrict our discussion in this paper to the case of a linear moving camera. Note, that this is done without loss of generality. The basic principle can be applied to any other parameterized camera movement too.

4. ICT SEARCH SPACE REPRESENTATION

To analyze the problem of homogeneous regions for the ICT analysis method we use the special virtual scene shown in fig. 3, right. It contains many highly homogeneous regions at different depth levels and orientations. The EPIs are created from horizontal slices in the image cube (fig. 3,left). As shown in the figure the homogeneous regions of the scene occur as stripes in the EPI.

To generate the proposed search space we are looking at the parameters of all theoretically possible ICTs in the EPI. In order to measure the probability of the existence of a particular ICT we define the so called ICTs *normalized color constancy (NCC)*. It describes the normalized variation of the ICTs color components within the image cube. We have introduced this synonym due to the large number of existing approaches for the determination of the color variation for a set of samples. For our experiments we have used the inverse of the pixels variance in the RGB color space straight away. Other approaches are possible.

As outlined in [5] the ICT of an existing 3D point represents its projection into all available camera planes. Such, assuming diffuse surfaces the color of the ICT is constant. In practice, this theoretical assumption does not hold. Further, distortions are introduced due to camera noise, aliasing etc. For this kind of problems the NCC gives a confidence value about the probability of the existence of a 3D point for a given ICT in the EPI. The proposed search space can be thought of being a probability or confidence map for all theoretically possible ICTs in an EPI. For each of these ICTs it represents the NCC in dependency of the ICT parameters. In the case of a linearly moving camera the ICTs are straight lines which may be described by the parameters *slope* s and the *offset* b. For other camera movements different ICT parameterizations could be used. For example, in case of a circular camera movement the search space could be parameterized by the ICT parameters *amplitude* and *phase shift* (see [5] for more information). As mentioned before, within this paper we will restrict our discussion to the linear case.

Fig. 4, right shows an example for the search space of the 'Boxes' sequence. Note, that bright regions stand for a high NCC

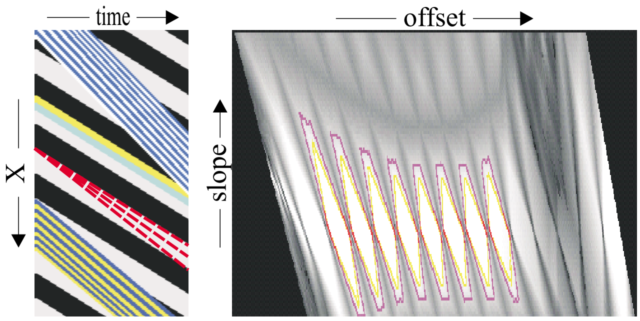

Fig. 4. 'Boxes' sequence, **left)** EPI, red lines illustrate possible wrong ICT matches, **right)** related search space, example for automatic pattern detection (yellow: NCC of 90%, pink: NCC of 70%). The overlapping border segments (red) correspond to points/lines B and D in fig. 6

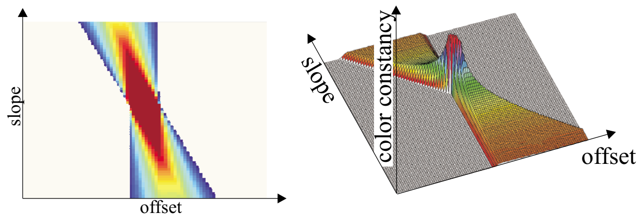

Fig. 5. Normalized color constancy for the case of homogeneous regions, **left)** 2D representation, **right)** 3D representation

value and vice versa. The search space is related to the EPI on the left-hand side of the figure. It can be seen that homogeneous stripes in the EPI cause particular patterns in the search space. These patterns can be detected by simple algorithms automatically. The figure shows the result of a threshold based border detection for those regions. The two colors yellow and pink stand for the different NCC values of 70% and 90%.

In order to benefit from the results of the pattern detection it is necessary to analyze the properties of the search space for the case of homogeneous regions. This problem is illustrated in fig. 6. The lines l_B and l_D in the left-hand side represent the borders of a homogeneous stripe in the EPI. They are represented as points P_B and P_D in the search space (right-hand figure) and define the minimum and maximum offset for any ICT within the region. Similar, the lines l_A and l_C mark the maximum and minimum ICT slope. They correspond to the points P_A and P_C in the search space.

At this point it should be mentioned that the geometrical properties of the EPI and the search space representation are dual to each other. Similar to other dual space approaches (see [9] for an example) lines in the EPI map to points in the search space and vice versa. Fig. 6 demonstrates this behavior. The intersection of the lines l_A and l_B defines a point p in the EPI (not shown in the figure). This point can be thought of being the intersection of all ICTs with the same constant offset. It is represented by the line $\overline{P_A P_B}$ in the search space. Vice versa, the point P_A in the search space is defined by the intersection of the lines $\overline{P_A P_B}$ and $\overline{P_A P_D}$ and corresponds to the line l_A in the EPI. In this way, all lines intersecting P_A in the search space correspond to points in the EPI lying on the line l_A. Note, that the duality of both spaces implies the possibility of dualizing the conventional ICT analysis

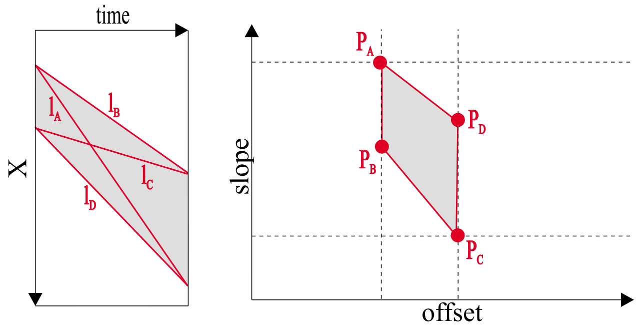

Fig. 6. Dual properties of **left)** the EPI and **right)** the search space, lines l_{A-D} in the EPI are represented as points P_{A-D} in the search space

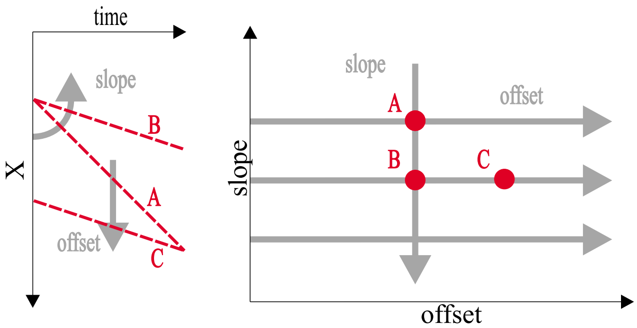

Fig. 7. Occlusion compatible search order (gray arrows) **left)** in the EPI, **right)** in the search space

algorithms to obtain new algorithms.

For the moment we restrict our analysis to the detection of homogeneous regions in the dual search space. Fig. 5 shows the simulation of a typical search space pattern caused by a homogeneous region in dependency of the NCC value. It can be seen that for a NCC value of 100% (red color) the 2D shape of the region is the same as shown in figures 4 and 6. To reconstruct the width of the stripe in the EPI the points P_B and P_D need to be detected. In a first approach we are using a simple threshold based pattern detection algorithm at different constant NCC levels. Assuming a 4-connectivity we determine the borders of the detected patterns as a closed region. The points P_B and P_D can be determined by the mean value of the overlapping parts of the region borders for different NCC levels. Fig. 4 illustrates the result of this operation (red points). Having the parameters of the points P_B, P_D the borders of the homogeneous region are defined in the EPI by the corresponding ICT lines l_B, l_D (see fig. 6). Finally, the missing ICTs inside the homogeneous region can be interpolated from the two border ICTs.

As a last step, an occlusion compatible search scheme needs to be defined for the search space analysis. As shown in fig. 7, it can be derived from the EPI straight away. As outlined in [5] the ICT search in the EPI starts with the maximum slope varying the ICTs offset (gray arrows in fig. 7, left). This process is repeated for decreasing slopes. Such, an occlusion compatible pattern detection in the search space needs to be performed on horizontal scan lines starting with the maximum slope and decreasing this value step by

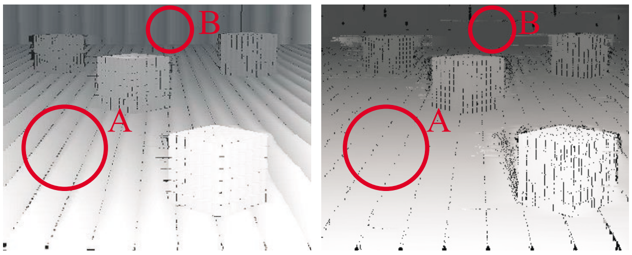

Fig. 8. Reconstructed depth map of the 'Boxes' sequence, **left)** conventional ICT search algorithm, **right)**, the proposed method

Fig. 9. 'Bridge' sequence, **left)** sample frame, **right)** reconstructed depth map using the proposed method

step (gray arrows in fig. 7, right).

5. EXPERIMENTAL RESULTS

For the final ICT analysis algorithm we have implemented an iterative method. In a first step only highly accurate detected ICTs (high NCC values) are accepted as valid results. For subsequent steps they are masked out from the source images. The process will be repeated by decreasing the overall reconstruction accuracy (i.e. the NCC value). For each new step the relating search space is recalculated. The iteration stops if the content of the whole EPI is detected successfully. The drawback of this method is that it increases the computational costs drastically. The main advantage is that the accuracy of the results is adapted automatically to the scene content. No careful interactive threshold adaption needs to be performed before the algorithm starts. The quality of the final result does not depend on the experience of the user. The algorithm runs fully automatically adapting itself to the complexity of the scene content. Practical tests have shown that for many cases 8 or 10 iterations are required. If the complexity of the analyzed scene becomes higher this number can increase up to 30 or 40 iterations.

One of the drawbacks of the proposed straight forward approach for the region detection in the search space is its restriction to a minimum region size of the homogeneous patterns. In practical tests we have evaluated that for stable results the region should contain at least 8-10 pixels to avoid artifacts. Otherwise we switch to the conventional ICT analysis algorithm.

The efficiency of the proposed algorithm for the detection of homogeneous regions is demonstrated in fig. 8. For the 'Boxes' sequence it shows the reconstructed depth map of a single frame for the conventional ICT analysis method (left-hand side) and the proposed algorithm (right-hand side). From the figure it can be seen that the artifacts inside the homogeneous regions in the left-hand image could be removed completely using the proposed method .

The resulting reconstructed depth map on the right-hand side has a uniform and smooth structure (see labeled parts A,B). Note, that in order to demonstrate the output of the algorithm no post processing was applied to fill the holes. Finally, fig. 9 shows the result of the reconstructed depth map for the natural 'Bridge' sequence. The demonstrated results illustrate that the problem of homogeneous region detection was solved successfully.

6. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a new approach for the detection and handling of homogeneous regions for the ICT analysis method. We have introduced a new search space representation of the EPI which has dual properties to the original EPI. The advantage of this method is that homogeneous regions have well defined patterns within this search space which can be detected by standard image processing tools. One of the most promising features of the proposed search space representation format is that there are no restrictions to the camera movement.

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] 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.

[6] 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.

[7] 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.

[8] 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, Oct. 2004, p. 127.

[9] S. Carlsson, "Duality of reconstruction and positioning from projective views," In IEEE Workshop on Representation of Visual Scenes, pages 8592., 1995.