DETECTION STRATEGIES FOR IMAGE CUBE TRAJECTORY ANALYSIS

Ingo Feldmann, Peter Kauff, Peter Eisert

Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institute, Berlin, Germany
Email:{feldmann, 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. For a moving camera each 3D point is represented by a trajectory in a so called image cube. In our previous work we have shown that it is possible to reconstruct the 3D scene from the parameters of these trajectories. A key component for this process is the trajectory detection within the cube. It is based on the image cube parameterization as well as the robust estimation of the trajectory color and trajectory color variation. In this paper we will focus on the second problem in more detail. We propose an algorithm which estimates the trajectory parameters in sub-pixel resolution with high accuracy. The corresponding 3D scene structure can be reconstructed with high level of detail even for complex scenes, multiple occlusions and very fine structures.

Index Terms— 3D scene reconstruction and modeling, multi view scene analysis, structure from motion

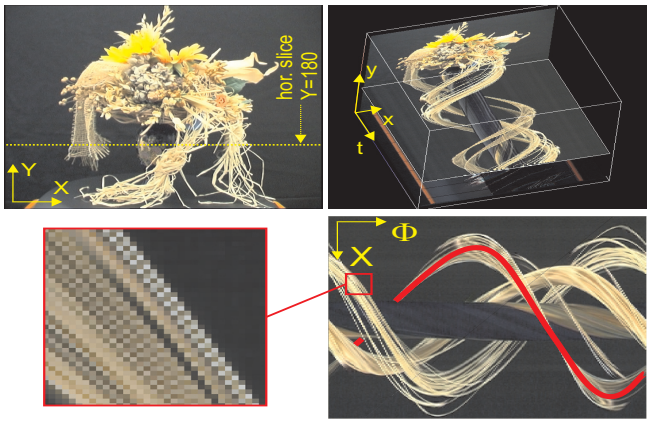

Fig. 1. 'Flower' sequence, circular camera path, **top left)** sample image, **top right)** image cube representation, **bottom)** trajectory structure in the image cube

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 Plane Image (EPI) analysis [3]. An Epipolar Plane Image can be thought of being a horizontal slice (or plane) in the *image cube* that can be constructed by collating all images of a sequence [1, 4]. It is defined for a linear equidistant camera movement only. In this case projections of 3D object points become straight point trajectories in the image cube which occur as lines on corresponding EPIs. The principle of EPI analysis is the detection of all point trajectories (the *EPI-lines*) in all available EPIs. The related 3D points are reconstructed from the parameters (shape, color) of the detected EPI-lines.

The advantage of this approach is the parallel analysis of all available views. Compared to other multi view approaches, such as for example the *voxel coloring* technique [5], a maximum of available information is exploited for the reconstruction of 3D scene structure. This gives a maximum of reconstruction accuracy for both, the geometric as well as the colorimetric properties of the reconstructed 3D scene points. Further, complex scene structures with multiple occlusions and high level of detail can be handled efficiently [4]. The disadvantage of EPI analysis is its restriction to linear equidistant camera movements.

For non-linear camera movements 3D points do not appear as lines on corresponding EPIs. Rather, they are represented by complex trajectories on arbitrary 3D surfaces within the image cube. The EPI-line approach cannot be applied for this case. Fig. 1 top right illustrates this at the example of a circular moving camera. One idea to solve this problem was presented in [6]. The authors suggest a piecewise linear approach where small segments of the object point trajectories are approximated by lines. Unfortunately, this reduces the amount of reference images and the robustness of the 3D reconstruction significantly.

In [7, 8] we have introduced a new concept called *Image Cube Trajectory (ICT) analysis* that overcomes the restrictions of EPI analysis and extends it to more general camera movements. The main idea of the proposed approach is the parameterization of the image cube based on the estimated camera parameters[1]. This parameterization is used to derive shape and position of 3D point trajectories (the so called ICTs) in the image cube as well as the trajectory occlusion ordering scheme. Based on this information an occlusion compatible search strategy is defined. It is based on the construction of a so called *search space* [10] which, again, depends on the estimated camera parameters and the given image cube parameterization. The search space represents the existence probability of all possible ICTs within the image cube. This probability is determined by statistical

[1]Robust self-calibration systems are well known in the literature [9]

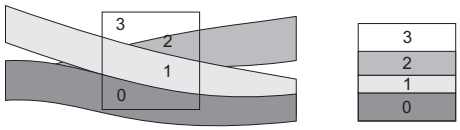

Fig. 2. Model for sub-pixel color accumulation in the image cube, **left)** trajectory model, **right)** simplified version

evaluation of the color variation along a given ICT. This method is based on the simplified model that a 3D point is projected into all camera positions with the same color.

Previous work on the topic of ICT analysis mainly deals with several aspects of image cube parameterization [7, 8] as well as the development of efficient search strategies [10] and the optimization of the considered search space [11]. Nevertheless, the quality of the reconstruction result highly depends on the robustness of the evaluation of ICT existence, i.e. the detection of the ICTs in the image cube and the reconstruction of the overall trajectory color. Especially for scene structures with high level of detail this task may become rather complex. Fig. 1 bottom illustrates this for the example of the 'flower' sequence. The width of the trajectories is smaller than the image pixel size. Therefore, the projected 3D points are blended with the background as well as the neighboring points on pixel level.

The problem of robust ICT detection and color reconstruction will be the main focus of this paper. We propose a new algorithm which works on sub-pixel level and extends the rather simple and straight forward approach proposed in [7]. In the following, we will firstly describe the general process of image cube trajectory detection. Afterwards, we discuss the problem of enhanced trajectory color reconstruction on sub-pixel level. We propose an accumulative color and mask buffering scheme. Further, we will discuss the influence of quantization error to the prediction result. We show, that this error can be estimated from the trajectory shape. Finally, the proposed algorithms are evaluated by experiments.

2. TRAJECTORY DETECTION IN THE IMAGE CUBE

The main two purposes of trajectory detection are to determine the probability of an ICT to exist in the image cube and to estimate the overall ICT color. The algorithm is based on the given camera setup, the derived image cube parameterization and the corresponding search space. For each search space positions a set of ICT parameters is generated which describes the ICT shape in the image cube. In order to evaluate the probability of ICT existence in the cube the statistical properties of the color elements along a given ICT shape are determined. One can think of this process as correlating a virtual ICT with the real image cube.

To solve this task, we have proposed a method which is based on the assumption that the color of the projected 3D points in the image cube is constant [7]. Therefore the probability of a given ICT to represent a 3D point on the object surface can be measured by the color variation along the trajectory. The straight forward way to solve this problem is to derive the ICT existence probability from the standard deviation of the colors of all ICT elements [7]. If the trajectory exists the standard deviation is low. In this case, the mean value of all ICT color elements represents the overall ICT color, i.e. the reconstructed color of the corresponding 3D point.

While this method works well for wide trajectory stripes, its robustness decreases drastically for ICTs which have a width equal or less than one pixel. In this case the original pixel color will be blended with scene background and neighboring trajectories. This problem is illustrated in fig.1 bottom left for a set of real trajectories and in fig. 3 top for a single synthetic ICT.

3. SUB-PIXEL TRAJECTORY COLOR ESTIMATION

In order to evaluate the statistical properties of a given ICT (i.e mean value and standard deviation) it is necessary to reconstruct the original ICT color for each trajectory element, i.e. for each corresponding pixel in the image cube. For sub-pixel trajectories the size of the ICT elements is by definition less than the pixel size. A correct color reconstruction is not possible. To overcome this problem the ICT analysis algorithm benefits from two facts. Firstly, the shape of the trajectory is known from the image cube parameterization. This information can be used to determine the overlapping areas for the pixels of the corresponding ICT. Secondly, the ordering of trajectories in the image cube is known. This is useful for the reconstruction of multiple overlapping trajectories. Note, that in general, the ICT analysis is based on a front-to-back search strategy [7]. Successfully detected ICTs are excluded from subsequent analysis steps by pixel wise masking in the image cube. For sub-pixel trajectories a sub-pixel masking scheme is required.

In order to simplify the task of color reconstruction for a single ICT element we have developed the accumulative color model illustrated in fig. 2 right. To differentiate between the color components of multiple ICTs at a single image cube pixel position we introduce index $i = 0 \ldots n-1$ with n as the total number of trajectories at this position. Each ICT is considered to have a certain width. The extension of an ICT within a pixel is simplified and modeled by the weighting factor $k_i = 0 \ldots 1$ which is normalized to the pixel size. The total pixel color $\mathbf{r}^{pix}$ is considered to be generated from the weighted sum of all ICT color parts $\mathbf{r}_i^{ICT}$ as

$$\mathbf{r}^{pix} = k_0\mathbf{r}_0^{ICT} + k_1\mathbf{r}_1^{ICT} + k_2\mathbf{r}_2^{ICT} + \ldots + k_{n-1}\mathbf{r}_{n-1}^{ICT} \quad (1)$$

with total weighting factor of $\sum_{i=0\ldots n-1} k_i = 1$.

In practice, for ICT analysis the total number n of all ICTs contributing to the final pixel color $\mathbf{r}^{pix}$ is not known in advance. Further, it is necessary to consider the influence of ICT self-occlusions within the pixel. Therefore, for a successful detection algorithm it is essential to use an occlusion compatible front-to-back search ordering to guarantee that in case of occlusions, $\mathbf{r}_i^{ICT}$ always occludes $\mathbf{r}_{i+1}^{ICT}$ (fully or partly). For an arbitrary analysis step i, all relevant ICTs can be labeled according to their depth hierarchy and one can define three groups of trajectories: The ICTs which are closer to the camera, the ICT at the current position, and the ICTs which are further away.

Assuming an error free front-to-back detection algorithm, the first category of ICTs is known from the detection history. The idea of this work is to introduce two pixel based accumulation buffers for these elements. The first buffer $\mathbf{b}_i^{acc} = \sum_{j=0\ldots i} k_j^{min}\mathbf{r}_j^{ICT}$ contains the weighted sum of all successfully detected ICT color components. The second buffer $k_i^{acc} = \sum_{j=0\ldots i} k_j^{min}$ contains the cumulative sum of the corresponding weighting factors. To guarantee a maximum of $k_{n-1}^{acc} \leq 1$ we introduce the corrected weighting factor $k_i^{min} = min(k_i, (1 - k_{i-1}^{acc}))$ with $k_0^{min} = k_0^{acc} = k_0$. This condition is required to limit the maximal pixel area for multiple occlusions to a normalized size of one. It ensures, that for an analysis step j with $j > i$ and $k_i^{acc} = 1$ all $k_j^{acc} = 0$. In other words, the color contributions of all subsequent ICT components will be discarded if the area of the pixel is already fully occupied from ICT

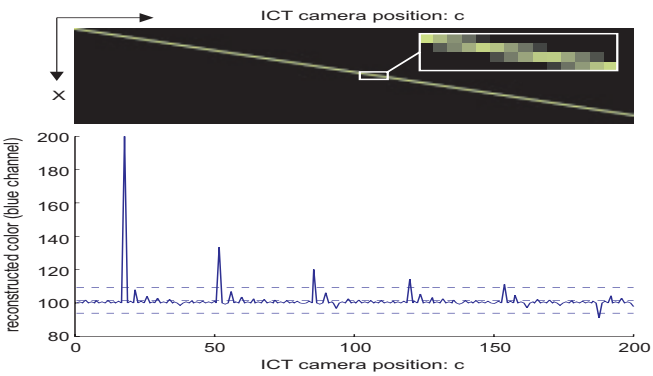

Fig. 3. Influence of quantization error to ICT color reconstruction, **top)** synthesized ICT **bottom)** reconstructed color (solid), mean value and standard deviation (dashed).

components which are closer to the camera. Note, that the accumulation of the weighting factors k_i^{min} can be interpreted as a sub-pixel masking operation with k_i^{acc} as being the sub-pixel masking buffer.

According to the three categories of trajectories mentioned above we define the weighted ICT color component at the current position (category two) as $\mathbf{b}_i^{ICT} = k_i^{min}\mathbf{r}_i^{ICT}$. Category three, the color components of all subsequent analysis steps, are considered as background. We define $\mathbf{b}_i^{BG}$ as the weighted background color. The final pixel color becomes

$$\mathbf{r}^{pix} = \mathbf{b}_{i-1}^{acc} + \mathbf{b}_i^{ICT} + \mathbf{b}_i^{BG} \tag{2}$$

The ICT color for step i can be predicted by

$$\mathbf{r}_i^{ICT} = \frac{1}{k_i^{min}}\mathbf{b}_i^{ICT} = \frac{1}{k_i^{min}}\left(\mathbf{r}^{pix} - \mathbf{b}_{i-1}^{acc} - \mathbf{b}_i^{BG}\right) \tag{3}$$

For the proposed front-to-back search algorithm, the weighted background color $\mathbf{b}_i^{BG}$ is not known in advance. One way to overcome this problem is to estimate $\mathbf{b}_i^{BG}$ from the neighboring pixel color $\mathbf{r}_i^{neigb}$. This solution is based on the assumption that the number of ICT components n contributing to the final pixel color $\mathbf{r}^{pix}$ is small, i.e. the depth structure is rather homogeneous on sub-pixel level. For this case exists a correlation between the current pixel and its neighbor which can be used for the color estimation. In eq. (3), the weighted background color component becomes

$$\mathbf{b}_i^{BG} = \left(1 - (k_i^{min} + k_{i-1}^{acc})\right)\mathbf{r}_i^{neigb} \tag{4}$$

Note, that this assumption cannot be guaranteed for all elements of a point trajectory. Nevertheless, experiments have shown that in practice it still holds for most of the ICT elements. In this way, the trajectory detection algorithm benefits from the fact that a large number of samples (i.e. all available images of the sequence) are used to reconstruct the color of a single 3D point. False color reconstruction results are handled as outliers in final statistical analysis of the ICT color reconstruction algorithm.

4. INFLUENCE OF QUANTIZATION ERROR

The previous section highlighted the problem of spatial re-sampling. In the following, the task of color re-sampling will be discussed. Consider a continuous pixel color $\mathbf{r}^{pix}$ which will be sampled to its

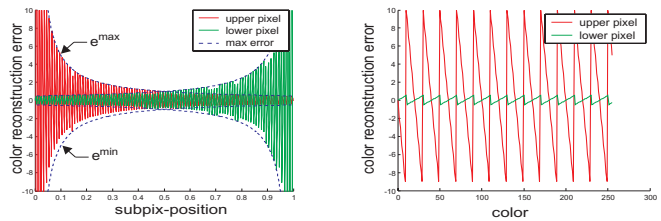

Fig. 4. Reconstruction error for sample ICT in fig. 3 in dependency **left)** of sub-pixel position, fixed color value (200), **right)** of ICT color, fixed sub-pixel position.

discrete version $\hat{\mathbf{r}}^{pix}$. The maximal quantization error for a single color unit is given by $\Delta q = \pm\frac{1}{2}$. The maximal reconstruction error $\mathbf{e}^{max}$ for a given ICT pixel color $\hat{\mathbf{r}}^{pix}$ can be derived from eq. (3).

$$\hat{\mathbf{r}}_i^{ICT} + \mathbf{e}^{max} = \frac{1}{k_i^{min}}\left(\left(\hat{\mathbf{r}}^{pix} + \Delta q\right) - \mathbf{b}_{i-1}^{acc} - \mathbf{b}_i^{BG}\right) \tag{5}$$

Where $\hat{\mathbf{r}}_i^{ICT}$ denotes the reconstructed ICT color which was estimated from the sampled pixel color. Note, that in order to reduce complexity the quantization errors of the accumulation buffers $\mathbf{b}_{i-1}^{acc}$ and $\mathbf{b}_i^{BG}$ are ignored. So, from eq.(5) the maximal ICT color reconstruction error can be derived as

$$\mathbf{e}^{max} = \frac{1}{k_i^{min}}\Delta q \tag{6}$$

In order to illustrate the influence of quantization error a linear ICT was synthesized (fig. 3 top). The color of this ICT was reconstructed for each of its elements based on the proposed algorithm. Fig. 3 bottom shows the result of the reconstruction. It can be seen that dependent on the ICT position the quantization error may cause large deviations from the original ICT color value of 100. Further, fig. 4 illustrates the dependency of quantization error from the pixel weighting factor k (left-hand side) and from the color value (right-hand side).

To overcome this problem, we introduce a threshold based approach. The idea is to estimate the maximal possible color reconstruction error for each ICT element. This is possible because the shape of the ICT is known. For each ICT element the corresponding weighting factor k_i^{min} can be determined. A threshold is used to mask ICT elements with large reconstruction error probabilities. These elements are excluded from overall ICT color variation estimation. In this way the robustness of the algorithm increases drastically.

5. EXPERIMENTAL RESULTS

To evaluate the efficiency of ICT color reconstruction the problem of ICT detection was simplified in order to minimize the influence of other distortions to the robustness of the reconstruction result (camera parameter estimation errors, camera noise etc.). Therefore, we have used a simple synthetic sequence with linear camera movement (see fig. 5 left). The trajectory search was restricted to a certain region in the 3D scene which is illustrated in fig. 5 left (region A). Note, that in the image cube this region is represented by the trajectories with a given offset range s as illustrated in region B on the right-hand side of the figure.

We have set up two test scenarios. Firstly, the efficiency of color reconstruction was tested for the single trajectory t, illustrated in

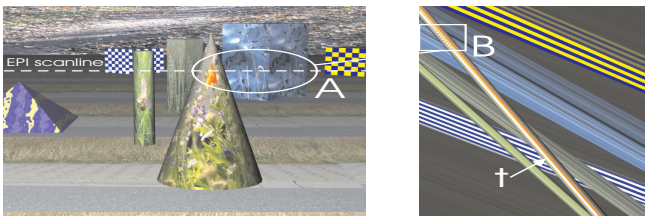

Fig. 5. Synthetic test sequence, **left)** sample image, **right)** trajectory structure in image cube

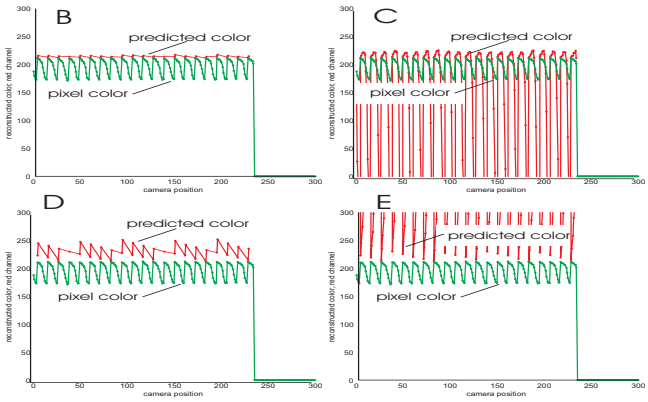

Fig. 6. Reconstructed color for trajectory t in fig. 5 right, **B)** proposed algorithm, **C)** no quantization error handling, **D)** no background color prediction, **E)** without both methods

fig. 5 right. The ICT color reconstruction results are compared with the original image colors. Fig. 6 shows the results for the proposed algorithm using: B) both: quantization error threshold and background color prediction, C) only background color prediction, D) only quantization error threshold, and E) just a simple color prediction. In case B) the ICT color was reconstructed with high quality. Both, the color prediction as well as the ICT color validation are very robust. In contrast, cases C, D, E show rather poor results. Especially, the neglected handling of quantization error enlarges the deviations from the original color drastically. The second test is illustrated in fig. 7. A full search was performed within the stated search region s (see fig. 5). Again, the four cases from the previous test setup were applied. It can be seen, that only for the proposed algorithm (case B), a reconstruction result with high accuracy can be obtained. In contrast, neglecting the quantization error handling (case C,E) as well as the background color prediction (case D,E) leads to poor reconstruction results.

6. CONCLUSIONS

ICT analysis is a powerful and new approach for 3D reconstruction. Previous work on this topic has shown that the estimation of overall trajectory color and the validation of its existence is still one of the bottlenecks of the algorithm. This paper proposes a new robust approach which overcomes this problem. It benefits from two aspects of the ICT analysis. Firstly, the trajectory shape and trajectory occlusion ordering can be derived from the image cube parameterization. Secondly, the parallel analysis on all available image data provides a large number of data samples which increases the robustness of the reconstruction result drastically. The proposed detection

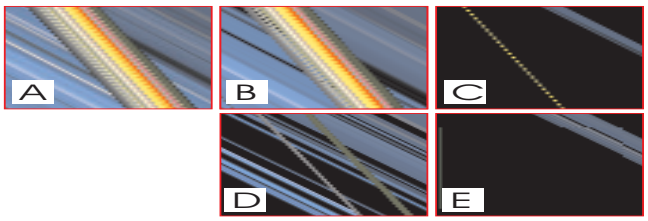

Fig. 7. Results for trajectory reconstruction, **A)** original image, **B)** proposed algorithm, **C)** without quantization error estimation, **D)** without background color prediction, **E)** without both methods

algorithm is based on an accumulative color prediction and masking scheme which works in sub-pixel resolution. It incorporates neighboring pixels as well as the reconstruction history in order to increase robustness. An efficient handling of distortions caused by the color quantization error further enhances the results. The robustness of the algorithm was proved by several experiments.

7. ACKNOWLEDGMENTS

The work presented in this paper has been developed with the support of the European Network of Excellence VISNET II (Contract IST-1-038398).

8. REFERENCES

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

[2] Paul A. Beardsley, Philip H. S. Torr, and Andrew Zisserman, "3d model acquisition from extended image sequences," in *ECCV (2)*, 1996, pp. 683–695.

[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] 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.

[6] 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.

[7] 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.

[8] I. Feldmann, P. Eisert, and P. Kauff, "Towards arbitrary camera movements for image cube trajectory analysis," in *Proc. International Conference on Image Processing (ICIP)*, Genova, Italy, Sep. 2005.

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

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

[11] 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, 2004, pp. 127–130.