OPTIMIZED SPACE SAMPLING FOR CIRCULAR 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. 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 represented and analyzed for all available images simultaneously. ICT analysis is based on the definition of an occlusion compatible search strategy in the image cube for known parametrized camera setups. In this paper we derive rules for an optimized sampling of the considered 3D space. They can also be applied to many other 3D reconstruction approaches such as voxel coloring [1]. We will restrict our discussion to a circular moving camera.

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 [2]. 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 [3]. 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 [4]. An Epipolar Image can be thought of being a horizontal slice (or plane) in the so called *image cube* [2, 5] 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 a straight line called *EPI line*. The principle of EPI analysis is the detection of all EPI-lines (and their slopes which correspond to depth) in all available EPIs. The advantage is the joint detection of point correspondences for all available views. Occlusions as well as homogeneous regions can be handled efficiently [5]. The big disadvantage of the algorithm is its restriction to linear equidistant camera movements.

In [6, 7], 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, such as the circular camera movement in fig. 2, left. For the special case of an inwards, tangential, or outwards facing circularly moving camera, we derived the analytical shape of the almost sinusoidal object point trajectories.

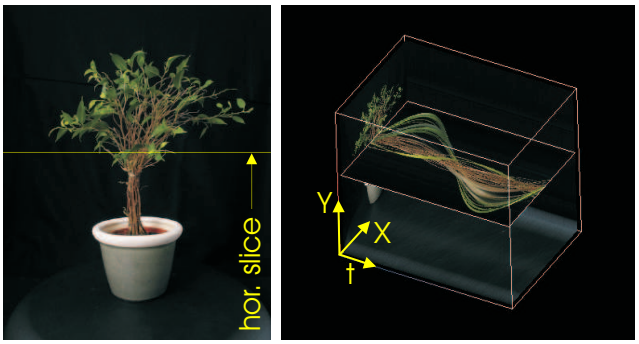

Fig. 1. Image cube representation for a sequence of a rotating object, **left)** original 'tree' sequence, **right)** image cube showing several ICTs on a horizontal slice in the $X-t$ domain.

(see fig. 1, right for an example) and proposed a new ICT matching method for robust depth estimation. In addition, we have analyzed the influence of perspective distortion to the exact ICT shape. Based on this we have introduced an efficient occlusion compatible search strategy in the image cube. We have suggested an inverse approach to the conventional way of EPI analysis where usually the EPI lines are detected [4, 8] and then depth is computed from the slopes of the lines. In contrast to that, we chose a 3D point, determine its trajectory through the image cube (image cube trajectory, ICT), and check the 3D point hypothesis by evaluating color constancy along the entire trajectory [6]. 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 space sampling. In this paper, we derive some non-uniform sampling that minimizes the number of ICT evaluations for a given camera resolution while maintaining cylindrical coordinates that help to specify an optimal occlusion compatible search order [6].

2. SAMPLING OF CYLINDER COORDINATES

The proposed ICT search strategy is based on the variation of ICT parameters in the image cube (i.e. the ICTs amplitudes, phases, and Y-coordinates). Therefore, we need to define an overall set of considered ICTs for our search. This corresponds to a finite set of discrete points in 3D space, i.e. a 3D sampling grid. Due to the defined occlusion compatible search order [6, 7], we focus on a circular sampling of the 3D space. A straight forward approach for this task is an equidistant sampled cylinder coordinate system as shown in fig. 2, right. Unfortunately, this solution is very inefficient, because it leads to a very high 3D resolution for small and a very low resolution for large radii.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

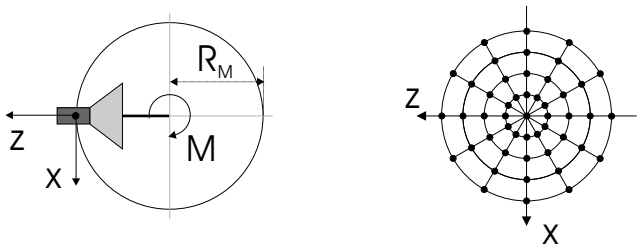

Fig. 2. **left)** Circular rotating camera (turntable setup), **right)** regular sampling of 3D space based on cylinder coordinates

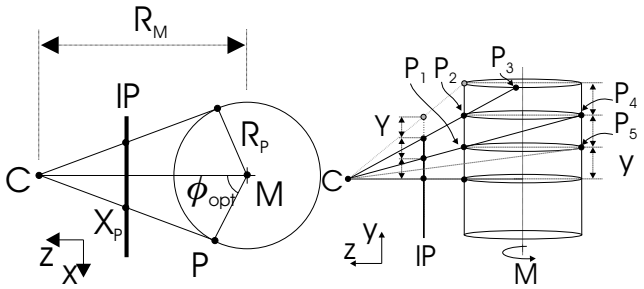

Fig. 3. **left)** Projection of a 3D point P to the image plane IP of a camera C, **right)** derivation of vertical 3D sampling grid for cylinder coordinates

The task of this paper is to derive a more efficient circular 3D sampling representation which is based on an optimized image cube representation of the ICTs. We will define a maximum set of discrete ICTs (i.e. ICT parameters) which still can be resolved and detected for a given image resolution. From this we will derive an optimized sampling of 3D space.

In the following chapter we will discuss the quantization of radii in 3D space first. Afterwards, we will consider the quantization of angles. Finally, we will define the vertical quantization (height). Note, that for matter of simplicity we will consider a fixed camera and a rotating scene in chapter 3.1, and vice versa, a rotating camera and a fixed scene in chapter 3.2.

3. OPTIMIZED HORIZONTAL SAMPLING

The task of this chapter is to derive a 3D space sampling rule for cylinder coordinates which is adapted to the properties of projective cameras. For linear, equidistant camera movement this problem can be solved applying the Plenoptic Sampling Theorem introduced in [9]. The authors present the dependency between the distance of the cameras, the resolution of the image planes and the 3D depth representation which is used to define a set of optimal depth layers parallel to the image plane of the cameras. The defined set of depth layers is the same for all considered cameras. In this chapter we will show that for circular camera movement an equivalent overall set of 3D layers does not exist. Rather, it highly depends from the considered pair of cameras. Nevertheless, we will see that it is still possible to define an overall maximum resolution bound for all cameras simultaneously which is adapted to the image cube resolution. Although this approach leads to an overestimated system it still provides an optimized solution for ICT analysis which is much more efficient than the equidistant approach illustrated in fig. 2, right.

3.1. Optimized Sampling of Radii

A rule for an optimized sampling of radii can be derived from the resolution of the image planes of the cameras. The projection of a 3D point P with the angle ϕ_P and radius R_P to the image plane IP of a camera C (see Fig. 3, left) can be defined as

$$X_p = f \frac{R_p \sin \phi_P}{R_M - R_P \cos \phi_P} \tag{1}$$

where R_M is the distance to the center of rotation M, f the focal length and X_P the resulting point on the image plane (see [6] for a detailed derivation). The 3D point P is related to a specific trajectory in the image cube (ICT) as the image cube represents the time dependent collection of all camera image plains (see [6, 10] for more details). If P is rotating around M [1] than the corresponding ICT is determined by the projected point X_P as a function of the rotation angle ϕ_P. The amplitude A of the considered ICT depends on the maximum deviation of $X_{P_{max}}$. The projection ray through $X_{P_{max}}$ is tangential to the circle defined by R_P. For this case the rotation angle ϕ_{opt} of point P can determined from the derivative

$$\left. \frac{\partial X_p(\phi_P)}{\partial \phi_P} \right|_{\phi_{opt}} = 0. \tag{2}$$

The combination of equations (2) and (1) leads to

$$\phi_{opt} = \arccos\left(-\frac{R_P}{R_M}\right). \tag{3}$$

The angle ϕ_{opt} specifies the maximum deviation $X_{P_{max}}$ in the image plane and such the maximal possible amplitude of a corresponding ICT. For the ICT analysis algorithm the number of different ICT amplitudes which can be resolved in the image cube is limited by the horizontal image resolution N_x. The projection rays through each of the discrete pixels (X_i, Y_i) are tangential to the optimized circles in 3D (see fig. 4, left). Using eq. (3) the radii of the optimized circles R_{opt_i} can be determined. We define angle α as $\tan\alpha = \frac{\Delta X}{f}$ and $\sin\alpha = \frac{R_{opt_i}}{R_M}$ where $\Delta X = \frac{w_h}{N_x}$ is the horizontal pixel size, $w_h = 2f\tan(FOV_h/2)$ the horizontal sensor width, f is the focal length, and $i = 0, 1, 2, .., N_x/2$ the positive horizontal pixel index. Such, the optimal radius is defined as

$$R_{opt_i} = 2\,\frac{R_m\,\tan(1/2\,FOV_h)i}{\sqrt{4\,(\tan(1/2\,FOV_h))^2\,i^2 + {N_x}^2}} \tag{4}$$

The total number of radii samples N_R is determined by the horizontal image resolution in the image plain of the camera as $N_R = N_x/2$. Fig. 4, left illustrates the optimized radii at the example of a camera image resolution of $N_x = 14$. The right graph shows the corresponding ICTs in the $X - \Phi$ domain of the image cube where the rotation angle ϕ is represented by the ICTs phase Φ. It can be seen that the maximal amplitude of each ICT (which was obtained from the optimal radius R_{opt_i}) is sampled exactly in the image resolution. The defined radii sampling grid represents the maximal number of discrete ICT amplitudes which still can be resolved and detected in the image cube and such the optimal radii sampling for ICT analysis.

[1] Note, that for a static scene either a fixed scene and a rotating camera or a fixed camera and a rotating scene can be considered.

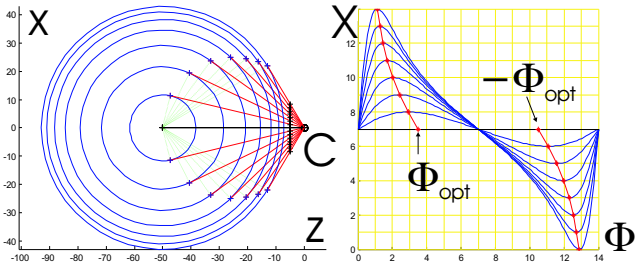

Fig. 4. **left)** Optimal radii grid in the 3D space, **right)** Optimal sampling of the ICT amplitudes in the $X - \Phi$ plane of the image cube

3.2. Optimized Sampling of Angles

In this chapter we will derive a rule for an optimized angular sampling. Firstly, following the discussion in [9] we will show that in case of a circular camera movement an optimal sampling does not exist for all available cameras simultaneously. Unlike the linear case, it is highly dependent on the considered cameras and the 3D image points. Nevertheless, for large sets of cameras we will derive an optimized rule that is based on the horizontal resolution of the cameras and leads to more practical results.

For the matter of simplicity, we will consider in the following a fixed 3D point P and a rotating camera C rather than a fixed camera and a rotating scene (fig. 5, right). The rotation angle between two subsequent cameras C_0 and C_1 is denoted as ϕ_{cam} with $\phi_{C_0} = 0$. It corresponds to the ICT phase Φ_{cam} in the image cube. The task of this chapter is to define the minimal discrete phase $\phi_{cam_{min}}$ which still can be resolved in the image cube. Such, we will obtain the maximal set of detectable ICTs in dependency of the image resolution and the camera setup. This corresponds to an optimal angular sampling in the 3D space.

Fig. 5, right illustrates the problem. The 3D point P is projected to the image planes of two subsequent cameras C_0 and C_1 with an angular distance ϕ_{cam}. The disparity d between the two projected points X_{C0} and X_{C1} can be expressed as $d = X'_{C0} - X_{C1}$. Substituting eq. (1) relative to ϕ_{cam} we obtain the dependency between the disparity d and the angles ϕ_P and ϕ_{cam} for a constant radius R_P

$$d = f R_p \left\{ \frac{\sin(\phi_P - \phi_{cam})}{R_M - R_p \cos(\phi_P - \phi_{cam})} - \frac{\sin \phi_P}{R_M - R_p \cos \phi_P} \right\}. \tag{5}$$

For the case of a linear moving camera it is shown in [9] that an optimal depth layering can be obtained by an equidistant sampling of the continuous disparity space d. The authors adapted the sampling interval Δd to the image resolution. In our case the circular camera movement leads to a non-linear solution which is illustrated in fig. 5, left. For the considered pair of cameras the corresponding optimal angular resolution occurs for the shown discreet angles $\phi_{P_{1,2,3,\dots}}$. It can be seen that it depends highly on ϕ_P, i.e. the position of point P. Further, it follows from eq. (5) that the sampled angles ϕ_{p_i} depend on the camera angles Φ_{cam_i} and $\Phi_{cam_{i+1}}$, i.e. the chosen cameras C_i and C_{i+1}. As in the ICT analysis approach we are dealing with a large set of cameras we obtain a different angular sampling for each considered subsequent camera pair.

Therefore, it is not possible to define an overall unique angular quantization format which is adapted optimally for all available camera pairs. Nevertheless, it is still possible to find a compromise. Because we are interested in resolving the ICTs phase shift within the pixel grid we consider that ICT section which has the highest slope as a reference. For such a section the ICTs amplitude change in dependency of ϕ_{cam} will be maximal. This angle is derived from the projection rays through the borders of the pixel closest to the z-axis of the camera. We define the intersection of this ray with the circle having an optimized radius R_{opt_i} as the minimal angular sampling distance $\phi_{cam_{min}}$. In this way we obtain a different angular sampling for each considered radius. $\phi_{cam_{min}}$ can be derived by the application of trigonometric functions as:

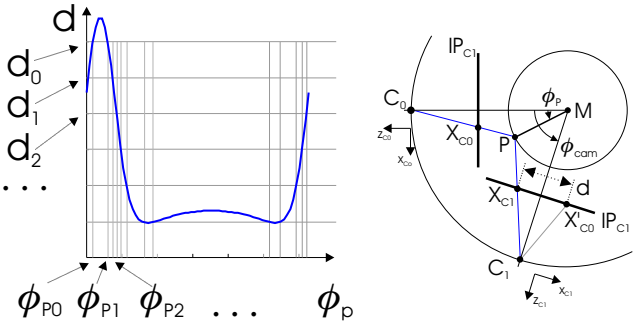

Fig. 5. Disparity d in dependency of the camera distance ϕ_{cam} and the point $P = (R_P, \phi_P)$ for a fixed radius R_P and a fixed camera angle ϕ_{cam}, **left)** d as a function of ϕ_P, a uniformly sampled disparity space leads to a non-uniformly angular resolution, **right)** setup in 3D space

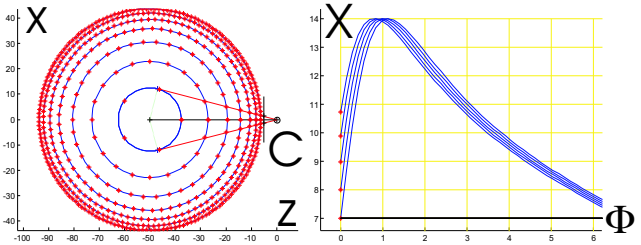

Fig. 6. **left)**, Optimized angle sampling in 3D space, **right)** Several subsequent ICTs shifted by the derived optimized distance

$$\phi_{cam_{min}} = \arcsin\left(R_M \frac{\sin(\alpha_{x_1})}{R_{opt_i}} \right) - \alpha_{x_1} \tag{6}$$

with $\alpha_{X_1} = \arctan \frac{\Delta X}{f}$ and $\Delta X = \frac{w_h}{N_X}$, where $w_h = 2f \tan FOV_h/2$. The equation represents the dependency between the optimized angular sampling $\Delta\Phi_{P_R}$ and the considered optimized radius R_{opt_i} for a fixed camera setup.

Fig. 6, left illustrates the optimized angular sampling grid in 3D space. The right hand figure shows several subsequent ICTs which are shifted by the defined optimized distance. It can be seen that the maximum resolution change in the X-axis between two subsequent ICTs is less or equal to the pixel size. Thus, an optimized 3D angular sampling grid for the ICT analysis approach can be defined.

4. OPTIMIZED VERTICAL SAMPLING

The optimal vertical sampling is straight forward. Again, the task is to define a vertical quantization grid which can be resolved in the image plane. Fig. 3, right illustrates the problem. The maximal

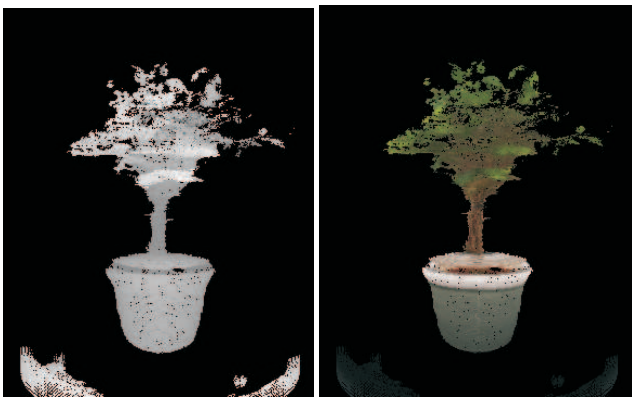

Fig. 7. Tree sequence, resolution optimized **left)** reconstructed depth map, **right)** reconstructed 3D model.

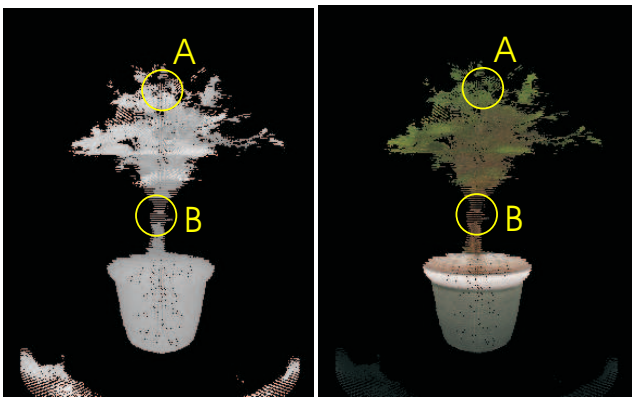

Fig. 8. Tree sequence, non optimized, **left)** reconstructed depth map, **right)** reconstructed 3D model.

vertical deviation occurs clearly for the points P_1, P_2 which are closest to the camera. The deviation for any other point, such as $P_{4,5}$ in the example is smaller. The considered radius R_{max} is the maximal used radius within the whole analysis and can be derived from eq. (4) for $i = N_x/2$. The vertical sampling interval Δy is determined by the vertical pixel spacing ΔY. It can be defined as

$$\Delta y = \frac{\Delta Y (R_M - R_{max})}{f}. \tag{7}$$

Note, that at the image border the sampling interval Δy could be slightly increased due to the limited sensor width (see point P_3). For the matter of simplicity we do not consider this effect which leads to a slight overestimation.

5. EXPERIMENTAL RESULTS

In order to prove the theoretical insights we have analyzed the natural tree sequence shown in fig. 1. We have used a turntable setup with 360 different positions corresponding to 360 cameras. The chosen image resolution was 280×360 pixel. Using our ICT analysis approach we have reconstructed a depth map and a 3D model for the case of an optimized space sampling first. 35 million 3D points were evaluated. Compared to the uniform quantization, the proposed sampling leads to a reduction of almost 50%. Fig. 7 shows the result. To show the effects of sample errors, the 3D reconstruction was performed without interpolation. In contrast, fig. 8 illustrates the reconstruction result for a non-optimized sampling grid based on half of the required quantization grid sizes for angular and hight quantization. For the second case, artifacts occur for vertical resolution in point A and for horizontal resolution in point B. Note, that an oversampling of the 3D space leads to the same result as presented in fig. 7 but increases the computational time.

6. CONCLUSIONS AND FUTURE WORK

In this paper we have presented an approach for the optimized sampling of the 3D space for the ICT analysis of a circular moving camera. We have derived dependencies between the image resolution, the camera setup parameters and the position of discrete 3D points in cylinder coordinates. The angular sampling is adapted to the requirements of the ICT analysis approach. For circular camera motion, only about half the trajectories need to be analyzed with the proposed optimized sampling positions. Our results were proved on theoretical simulations as well as on real image sequences. For our future work we plan to consider more general parameterized camera setups. Further, we plan to exploit the sampling accuracy within the ICT matching process.

7. ACKNOWLEDGMENT

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

8. REFERENCES

[1] 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.

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

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

[4] 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.

[5] 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.

[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.

[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] 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.

[9] J.-X. Chai, X. Tong, S.-C. Chan, and H.-Y. Shum, "Plenoptic sampling," in *Proc. Computer Graphics (SIGGRAPH)*, Los Angeles, USA, Jul. 2000, pp. 307–318.

[10] I. Feldmann, P. Kauff, and P. Eisert, "Image cube trajectory analysis for concentric mosaics," in *Proc. Int'l Workshop on Very Low Bit-rate Video Coding*, Madrid, Spain, Sep. 2003, pp. 341–350.