Epipolar Plane Images as a Tool to Seek Correspondences in a Dense Sequence*

Martin Matoušek, Václav Hlaváč

Czech Technical University, Faculty of Electrical Engineering
Department of Cybernetics, Center for Machine Perception
121 35 Prague 2, Karlovo náměstí 13, Czech Republic
phone +420 224 357 637, fax +420 224 357 385
{xmatousm,hlavac}@cmp.felk.cvut.cz

Abstract. We present a method seeking correspondences in a dense rectified image sequence, considered as a set of Epipolar Plane Images (EPI). The main idea is to employ dense sequence to get more information which could guide the correspondence algorithm. The method is intensity based, no features are detected. Our spatio-temporal volume analysis approach aims at accuracy and density of the correspondences. Two cost functions are used, quantifying the belief that given correspondence candidate is correct. The first one is based on projections of one scene point to the spatio-temporal data, while the second one uses two-dimensional neighborhood (in image plane) around such projections. The assumed opaque Lambertian surface without occlusions allows us to use a simple correspondence seeking algorithm based on minimization of a global criterion using dynamic programming.

1 Introduction

One of the fundamental problems in computer vision is correspondence seeking. Established correspondences (coordinates of pixels that are projections of a single point in the 3D scene) are precondition for reconstruction of 3D shape from intensity images.

The state-of-the-art approaches for correspondence seeking can be classified according to data processed as follows. The first class works with a set of two or more unorganized images. Computational binocular stereo is commonly used here. The well-known methods for occlussion-free scenes are based on dynamic programming [3]. For scenes with occlusions, pairing algorithms not imposing the monotonicity constraint are used [9]. Larger set of images can make search for correspondences easier. The verification of a correspondence candidate by comparing intensity values in multiple unorganized images is used in [8].

The methods belonging to the second class work on organized set of images, usually forming a sequence. Detecting and tracking features in sequence [10]

* This research was supported by the Czech Ministry of Education under the project MSM 212300013 and by the Grant Agency of the Czech Republic under the project GACR 102/03/0440 and by the European Union under the project IST-2001-32184.

is one possibility. Another possibility is to combine wide-baseline stereo and a sequence. The moving wide-baseline stereo rig provides two dense sequences [4] in which features are tracked. Short sections of tracked trajectories are obtained and simplify finding correspondences.

Images of the dense sequence (a sequence with small interframe displacements) stacked to a spatio-temporal block are the the input of the third class of methods. Signal processing techniques (filtering) can be used to detect occlusions [2, 7]. Another example is computation of differential characteristics in local neighborhood of the current pixel to estimate optical flow. Epipolar plane image analysis [1] also belongs to this class of approaches.

Each of the mentioned approaches has different advantages and drawbacks. Wide-baseline stereo has sufficient spatial accuracy, but often cannot resolve an ambiguity due to e.g. repeating-pattern. Local methods (tracking, optical flow) usually find acceptable correspondences between adjacent frames, but error accumulation occurs for longer sequences. Tracking may allow to detect features with high accuracy, but still only sparse set of correspondences is obtained.

Our approach to spatio-temporal block analysis utilizes large data set and is supposed to provide more scene information than stereo. Ambiguity known in stereo is reduced. We use global information present in the whole spatio--temporal block (or in a single EPI) without error accumulation (as in local methods) and no features have to be detected. This allows to compute a dense set of correspondences with sub-pixel accuracy, and large distance between end frames of the sequence leads to sufficient spatial accuracy. Unlike feature-based EPI analysis methods [1], our approach is purely signal based. The difference of another intensity based correspondence verification methods (e.g. [8]) is in employing the highly organized structure of the EPI, which significantly reduces the search space.

2 Problem Formulation

We are given a dense sequence of images in which each two images are rectified. We assume no occlusions in the captured sequence and rigid opaque Lambertian scene with enough texture. This allows to use uniqueness and monotonicity constraint, and to assume that a point on the surface is projected to each image with the same intensity. Our task is to establish a set of correspondences between end images of the sequence, focusing on density and accuracy of the correspondences.

3 Spatio-Temporal Volumes, Epipolar Plane Images

Let us consider dense image sequence, i.e., the sequence with small inter-frame pixel displacements. Such a sequence can be viewed as a 3D *spatio-temporal* volume $f(x, y, t)$; x, y denote the image coordinates and t is an image index. Particular scene point is projected as a curve $\mathcal{C}_X$ in (x, y, t) space. This curve is parametrized by t. Note that such a curve need not be defined for all t in case of occlusions.

Let the sequence $f(x, y, t)$ be captured by the pinhole cameras with projection centers lying on a line (the baseline), such that all adjacent centers have the same distance. Moreover, all cameras have the same intrinsic calibration parameters, all optical axes are perpendicular to the baseline and parallel to each other. Thus, the whole sequence is epipolarly rectified, it means that each pair of images is rectified. In this casem, each curve $\mathcal{C}_X$ is formed by (parts of) a line with constant y coordinate (Figure 1a).

The baseline generates a bundle of epipolar planes. Epipolar lines, belonging to a single epipolar plane, are aligned with image rows. They are stacked along the parameter t to form *epipolar plane image* (EPI, Figure 1b).

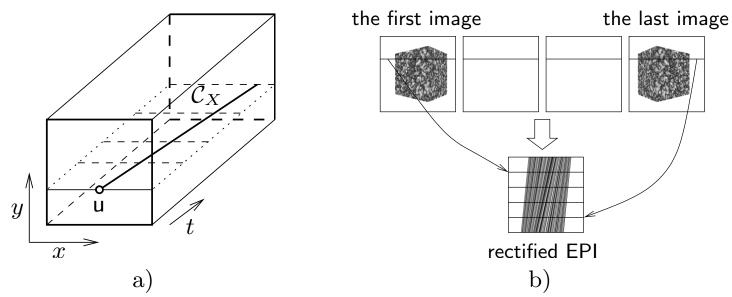

Fig. 1. (a) Rectified spatio temporal volume and (b) EPI formation scheme.

A sequence obtained under the above assumptions has the following useful properties:

- Search space is reduced to 1D only. Independence of epipolar lines allows to treat each EPI separately.
- A single scene point maps to a *straight line* in the EPI, which has in case of Lambertian reflectance a constant intensity profile.

The assumption of constant intensity of the line in the EPI corresponding to a particulare point is not entirely true in real world due to e.g. noise. Moreover, it is very difficult in practice to capture a sequence rectified accurately enough. In [5] we analyzed several phenomena violating the assumptions of the method.

4 Seeking Correspondence

A correspondence between a point in the first and in the last image is denoted by a pair (i, j) of column coordinates in the selected row. The (i, j) coordinates also determine end points of a line in the EPI (Figure 2a); such a line is a realization of a single point in the scene. Given a correspondence candidate, the cost $c(i, j)$ quantifies confidence that (i, j) is realization of some observed scene point. We used two following cost functions.

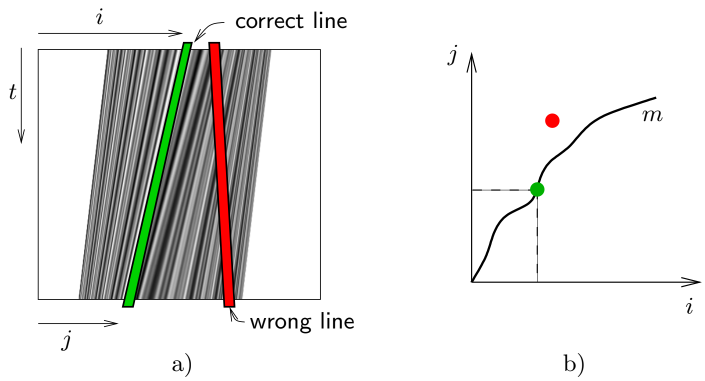

Fig. 2. (a) Selected EPI with two lines shown and (b) search space in which correspondences are sought.

The first cost function is defined considering only one scene point. The cost is calculated according to intensities along the line (i, j) in the EPI. *Variance* of the intensities was chosen as this cost, similarly as in [8]. The second cost function uses a small 2D neighbourhood around the point in each image of the sequence. We assume, that such a small neighbourhood is an observation of a local surface texture model, and we measure the consistency of all observations with this model. The sum of normalized correlations between each observation and the model serves as the consistency measure. The neighbourhood in the first image in the sequence was chosen as the model, however, other choices (e.g., mean of all observations) are possible.

The domain of the cost $c(i, j)$ is $\mathbb{R}^2$ in both cases because positions with sub-pixel accuracy can be obtained by interpolating neighbouring coordinates. Here we take advantage of the fact that there are many images at hand in the sequence. The defined cost function $c(i, j)$ allows to examine all possible correspondence pairs. We are looking for the whole set of correspondences considered as a *mapping* m between i and j (Figure 2b). The global criterion $\mathcal{J}(m)$ sums up $c(i, j)$ along m. The mapping m is a strictly increasing function because of ordering and uniqueness constraints. The estimate m^* of the optimal mapping is computed by minimizing $\mathcal{J}(m)$ by dynamic programming. During dynamic programming, discrete raster is subsampled to achieve sub-pixel accuracy.

5 Experimental Results

We conducted experiments having two aims. The first aim was to test the idea on synthetic data, which perfectly fits the assumptions of the algorithm. The second aim was to uncover the limits of practical usage on the real scenes.

5.1 Synthetic Data

Only the first cost function was tested on synthetic data, which were provided by the ray-tracer. The scene was an edge of a cube with random granite-like texture. The maximum and mean error was 0.13 and 0.04 pixel, respectively. The results on synthetic images are shown in Figure 3 and presented in more detail in [6].

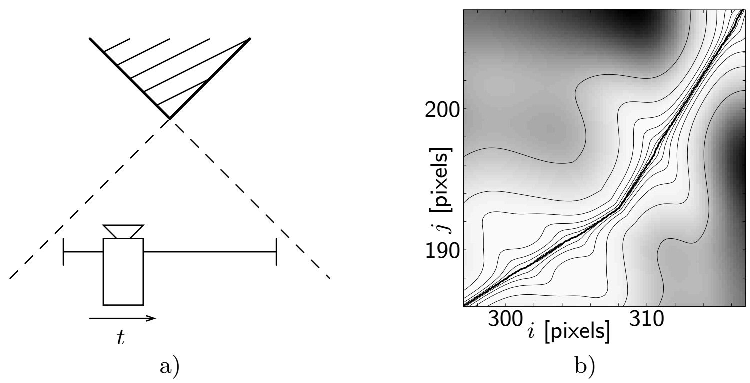

Fig. 3. (a) The setup for the synthetic experiment. (b) Result of the dynamic programming on synthetic data, the path corresponding to surface near the corner is zoomed in. The underlying image shows the values of the cost function; white means lowest value, contours (with logarithmic scale) mark equal values.

5.2 Real Data

The studied real scene consisted of a reverse side of a single 15×15 cm bathroom tile, sprayed by a random texture. The unglazed ceramic surface is expected to have almost Lambertian reflectance properties. Images were captured by a stationary CCD camera (1K×1K resolution, 9mm chip size, 25 mm lens) placed at the distance 90 cm from the scene. The homogeneously illuminated scene was moved along the straight line by 15 cm in 1.5 mm steps; a hundred of images was taken (Figure 4). Note that we are moving the scene instead of the camera because mechanical inaccuracy during movement has smaller influence in that case. To fulfill the assumption that the surface is observed with the same intensity from an arbitrary direction, we illuminated the scene by the light source placed sufficiently far from the scene.

The movement of the tile is approximately, but not exactly, parallel to image x axis, so the sequence is rectified before the correspondence seeking. To enable the computation of rectifying homography, the tile is equipped with easy-to--locate calibration marks (chess-board). However, we need not know the full camera calibration for the rectification. Equidistant movement steps together with constant relative orientation of the movement direction and optical axis are ensured. Then the rectifying homography consists only of two elementary transformations, rotation around the optical axis and pan. These transformations remain constant for each image and thay can be computed from a set of known correspondences.

Fig. 4. Experimental setup. The camera and the moving table mounted on granite slab.

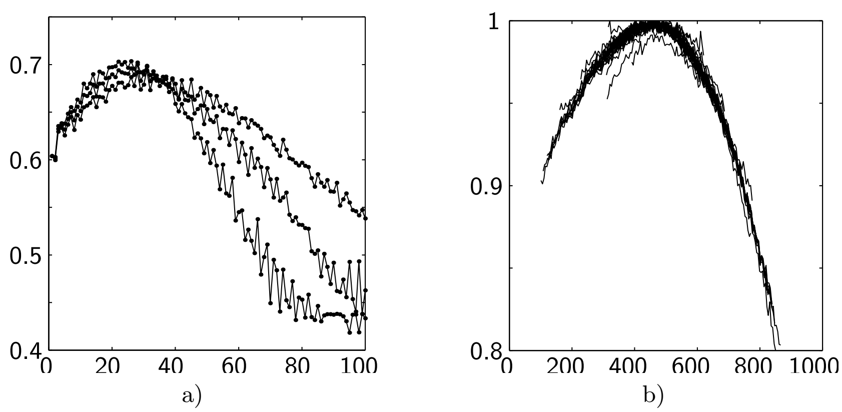

Fig. 5. (a) Three intensity profiles related to lines in an EPI. Horizontal axis corresponds to t and vertical axis to intensities. The top profile corresponds to correct line (i, j). The other two profiles correspond to the modified lines $(i, j+1)$ and $(i, j-1)$. (b) Intensities along a several manually placed correct lines, normalized to maxima and aligned according to position in image. Horizontal axis corresponds to pixels and vertical axis to intensities.

The back side of our bathroom tile is covered by regular stripe-like hollows which are approximately 0.2 mm deep—the surface is bounded by two parallel 'planes' with known distance. The approximate value of disparity between border images in the sequence was 427, which leads to expected difference of a disparity around 0.1 pixel between the two ‘planes’.

First, we use the variance-based cost function. Real data suffer from many phenomena, missing in ideal case, some of which are described in [5]. Figure 5 shows several intensity profiles along correct line in an EPI. It is apparent that the photometric field darkening strongly influences the profile. We measured the camera response to the constant irradiance in a few image locations (regular 10×10 grid). The third-order two-dimensional polynomial was fitted to the measured intensities and used for the correction of the field darkening.

The presence of noise and violation of Lambertian surface assumption motivated us to use the cost function based on correlation window (Figure 6).

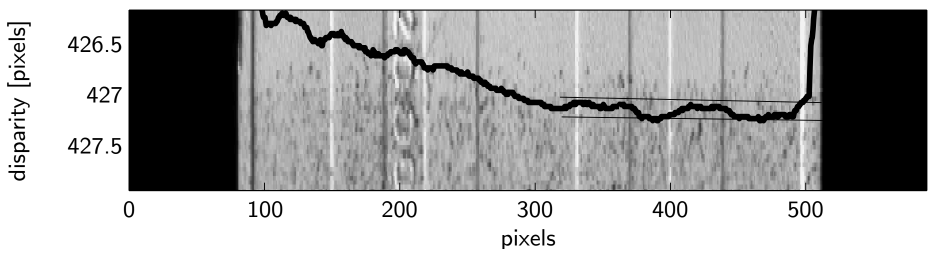

Fig. 6. Computed disparity values for a selected y-coordinate, overlaid on the first image of the sequence. Expected disparity for the two planes is shown as the two parallel lines. Let us note that the tile is non-planar in reality.

6 Conclusion

We presented the idea of how to use the set of epipolar plane images for correspondence seeking. The structure of EPI allows all image data from the sequence to be employed directly in computation. The main contribution lies in using purely intensity based method on the set of EPIs. The idea works well when the data processed are in accordance with the strict assumptions of Lambertian surface and accurately rectified sequence (synthetic case).

However, it has proven to be difficult to fulfill assumptions of the method in the real case, considering the accuracy level we aim to work at. The useful information extracted from the rectified spatio-temporal volume is disturbed by imperfection of the real measurements. As shown in [5], the inaccurate positioning of the camera has a strong influence. However, even in the real case, the results demonstrate the feasibility of the idea. The application of the stereo using the pairing algorithm [9] suffers from ambiguous matches.

Our outlook is to extend the concept on general (not-rectified) sequence. The structure of a general spatio-temporal volume, relating the image data of the dense sequence, should simplify the correspondence search in such cases. Also occurrence of occlusions can be handled more easily in this framework.

References

1. Robert C. Bolles, Harlyn H. Baker, and David H. Marimont. Epipolar plane image analysis: An approach to determining structure from motion. In *International Journal of Computer Vision*, number 1 in 1, pages 7–55, 1987.
2. George T. Chou. A model of figure-ground segregation from kinetic occlusion. In Eric Grimson, editor, *ICCV'95: Proc. 5th IEEE Intl. Conf. on Computer Vision*, pages 1050–1057, Cambridge, Massachusetts, June 1995. IEEE Computer Society Press.
3. Ingemar J. Cox, Sunita L. Higorani, Satish B. Rao, and Bruce M. Maggs. A maximum likelihood stereo algorithm. *Computer Vision and Image Understanding*, 63(3):542–567, May 1996.
4. Pui-Kuen Ho and Ronald Chung. Stereo-motion with stereo and motion in complement. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(2):215–220, February 2000.
5. Martin Matoušek and Václav Hlaváč. Correspondences from epipolar plane images, experimental evaluation. In Boštjan Likar, editor, *Proc. Computer Vision Winter Workshop*, pages 11–18, Bad Aussee, Austria, February 2002. PRIP.
6. Martin Matoušek, Tomáš. Werner, and Václav Hlaváč. Accurate correspondences from epipolar plane images. In Boštjan Likar, editor, *Proc. Computer Vision Winter Workshop*, pages 181–189, Bled, Slovenia, February 2001. Slovenian Pattern Recorgnition Society.
7. Sourabh A. Niyogi. Detecting kinetic occlusion. In Eric Grimson, editor, *ICCV'95: Proc. 5th IEEE Intl. Conf. on Computer Vision*, pages 1044–1049, Cambridge, Massachusetts, June 1995. IEEE Computer Society Press.
8. Sébastien Roy and Ingemar J. Cox. A maximum-flow formulation of the N-camera stereo correspondence problem. In Sharat Chandran and Uday Desai, editors, *ICCV'98: Proc. 6th IEEE Intl. Conf. on Computer Vision*, pages 492–499, Bombai, India, January 1998. IEEE, Narosa Publishing House.
9. Radim Šára. Finding the largest unambiguous component of stereo matching. In Anders Heyden, Gunnar Sparr, Mads Nielsen, and Peter Johansen, editors, *ECCV'02: Proc. 7th European Conf. on Computer Vision*, volume 3, pages 900–914, Copenhagen, Denmark, May 2002. Springer-Verlag.
10. Tomáš Werner, Tomáš Pajdla, and Václav Hlaváč. Correspondence by tracking edges in a dense sequence for image-based scene representation. In Tomáš Pajdla, editor, *Proceedings of the Czech Pattern Recognition Workshop '97*, pages 64–68, Prague, Czech Republic, February 1997. Czech Pattern Recognition Society.