Correspondences From Epipolar Plane Images, Experimental Evaluation *

Martin Matoušek, Václav Hlaváč
Czech Technical University, Faculty of Electrical Engineering
Department of Cybernetics, Center for Machine Perception
121 35 Prague 2, Karlovo náměstí13, Czech Republic
phone +420 2 2435 7637, fax +420 2 2435 7385
{xmatousm,hlavac}@cmp.felk.cvut.cz

Abstract

The method seeking correspondences in Epipolar Plane Images (EPI) is presented. The principal idea is to employ dense sequence to get more information which could guide the correspondence algorithm. Theoretically a simple technique finding homogeneous straight lines in EPI suffices to establish correspondences.

This paper verifies the method in real experiments. The paper builds on results of [6]. Real data uncovered several optical and mechanical phenomena which introduce additional problems. They are revealed in the paper one by one and undone.

Keywords: epipolar plane image, correspondence, Lambertian surface

1 Introduction

We report about our attempts to improve correspondence between images, i.e. seeking coordinates of pixels which match the single point in the 3D scene. The underlined idea is to use a dense epipolarly aligned sequence of images to get more information than the wide baseline stereo has at hand. However, we do not follow the optical flow paradigm which uses local differential operators and cannot avoid error accumulation.

We have suggested the method seeking for correspondences and tested it on synthetic images [6]. This paper shows experimental results on real data, describes several pitfalls we trapped in, and comments on methods allowing us to extricate from traps. We believe that our relatively thorough study uncovered phenomena which are not so obvious.

*This research was supported by the Czech Ministry of Education under Research Programme MSM 212300013 and by the Grant Agency of the Czech Republic under Project GACR 102/00/1679.

2 Related Work

The first group of papers relates to computational binocular stereo. The state-of-the-art methods in stereo are of interest for us because we like to compare our results with them. The dynamic programming-based methods are commonly used [4]. We use the better correspondence matcher suggested by R. Šára [10].

The second group deals with contributions that use more than two images to establish correspondences. The paper [9] verifies correspondences by comparing intensity values in more than two images. The other possibility is to combine wide base and a larger data set. The moving wide base stereo rig provides two dense sequences [5] in which features are tracked. Short sections of tracked trajectories are obtained and ease finding correspondences.

The third group of papers uses the completely stacked sequence (spatio-temporal block) and aims at deriving scene structure from it. For instance, signal processing techniques (filtering) can be used to detect occlusions [3, 7].

3 Epipolar Plane Image

Let us assume the sequence of images $f(x, y, t)$ where x, y are coordinates of the pixel and t is an index in the sequence. Consider a particular case—a sequence of images captured by cameras with collinear locations of projection centers. Epipolar lines belonging to a selected epipolar plane (one epipolar line per image) can be stacked along the parameter t and forms an *epipolar plane image* (EPI).

Let us simplify the situation further and consider even more special case of an epipolarly rectified sequence which was sampled equidistantly in t. The situation is illustrated in Figure 1a.

Such simplified EPI has several properties that are useful.

- The space in which correspondence is sought is 1D only. It is assumed that epipolar lines do not depend on each other. This property allows to decompose the correspondence problem to simpler subproblems (which can be processed in parallel).
- The point in 3D scene maps to the *straight line* in EPI.

4 Simplifying Assumptions

Our aim is to analyze the simplest case. Therefore the following assumptions were considered.

- Surfaces in the scene are opaque and have Lambertian reflectance.
- The scene is occlusion-free, i.e., ordering and uniqueness constraint are not violated.
- The scene is covered by a ‘sufficiently good’ texture. The unambiguous solution of the correspondence problem can be obtained only if there are no extended regions with zero or undefined gradient in a light-field [2]. Let us note that EPI is a special case of a light--field.

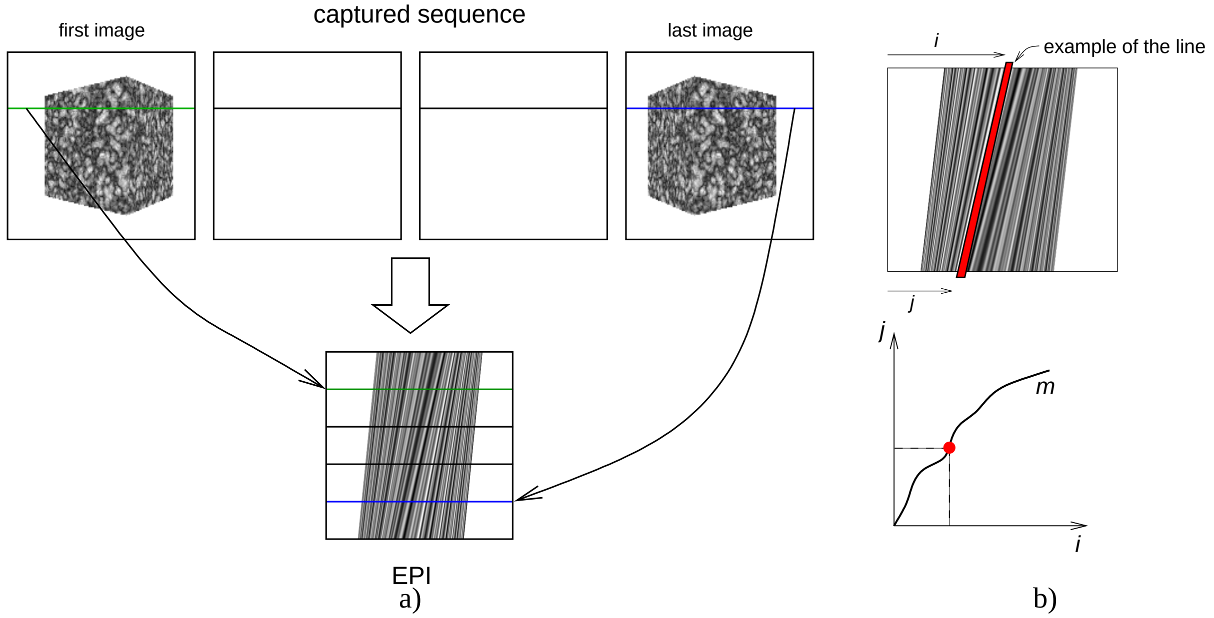


Figure 1: (a) EPI formation scheme. (b) (top) Selected EPI with one line shown and (bottom) search space in which correspondences are sought.

5 Seeking Correspondences

An elementary entity is the correspondence between a point in the first and in the last image. This correspondence of column coordinates in the first and the last image (denoted as a pair (i, j)) is realization of a single point in 3D scene. The *line in EPI* is another representation of the correspondence, see Figure 1b. Theoretically, the line corresponding to a single point in 3D scene should be of the same value. Practically, such line has only similar values due to various artifacts and noise. Analysis of related phenomena is actually the contribution of this paper, still to come.

The cost $c(i, j)$ quantifies our believe that the correspondence candidate agrees with the observed 3D point. The cost is calculated according to values (intensities) along the line found in EPI. *Variance* of the intensities was chosen similarly to [9]. The domain of the cost $c(i, j)$ is $\mathbb{R}^2$ because positions with subpixel accuracy can be obtained by interpolation neighbouring coordinates. Here we take advantage of the fact that there are many images at hand in the sequence.

Having defined cost function $c(i, j)$ it is possible to examine all possible correspondence pairs. The whole set of correspondences considered as *mapping* m between i and j is sought Figure 1b. The mapping is estimated by minimization of the global criterion $\mathcal{J}(m)$ which sums partial increments of $c(i, j)$ along the m.

The mapping m is a strictly increasing function because of ordering and uniqueness constraints. The estimate of the optimal mapping m^* can be computed by minimizing the criterion by dynamic programming which is computationally efficient. The dynamic programming is a discrete optimization method. In our case the raster is sampled in subpixel accuracy.

6 Experimental Evaluation on Real Data (Lessons Learned)

We have described the method so far. We can proceed now to actual experiments, problems that occurred, and our thoughts how to avoid them.

6.1 Experimental setup

The studied scene consisted of a single 15×15 cm bathroom tile. We captured reverse side of the tile. Our expectation was that unglazed surface of ceramics has almost Lambertian reflectance properties. The reverse side of the tile was sprayed by a random texture.

Images were captured by a stationary camera while the homogeneously illuminated scene (reverse side of the tile) was moved along the straight line. This configuration was chosen to minimize camera vibrations. The tile was moved on the translation table OPTEN by 15 cm. The images were captured in 1.5 mm steps, i.e, a hundred of images was taken.

A PULNIX TM-1001 digital CCD camera (1k×1k resolution, 9mm chip size) was used equipped with TAMRON 89698 25 mm lens mount (F-stop 12 was set). The camera was placed at the distance 90 cm from the scene.

The bathroom tile was chosen deliberately because the form of its reverse side suits our purpose. Even no ground truth about the 3D shape is available we can expect that the observed surface is bound by two parallel planes with known distance. The reverse side of the tile is covered by regular stripe-like hollows with depth approximately 0.2 mm. The approximate value of disparity between border images in the sequence was 427. This leads to expected difference of a disparity around 0.1 pixel.

Calibration markers (chessboard) were placed on the captured tile (Figure 6b) to ease epipolar rectification. The sequence $f(x, y, t)$ was separated to individual EPIs. Only one selected EPI was used in experiments.

6.2 Checking lines in EPI

This experiment attempts to verify the assumption that the intensities along the line in EPI corresponding to a single point in 3D scene are homogeneous and bear enough information about correspondences. First a single line in EPI was analysed. The intensity profile along the line mapping a correct correspondence should be constant in the ideal case. The observed profile is different, see top curve in Figure 2a. This would not matter if the 'correct' profile was significantly more homogeneous than 'erroneous' profile. This was tested by looking at profiles that are not correct, see two bottom curves in Figure 2a. We observed significant deformation of the profile.

Next test tries to identify what caused the problem. Fifty intensity profiles corresponding to 'correct' lines and covering the whole EPI were displayed in one figure. Each profile was normalized so that its maximal value is 1 and aligned according to the position of the pixel, see Figure 2b.

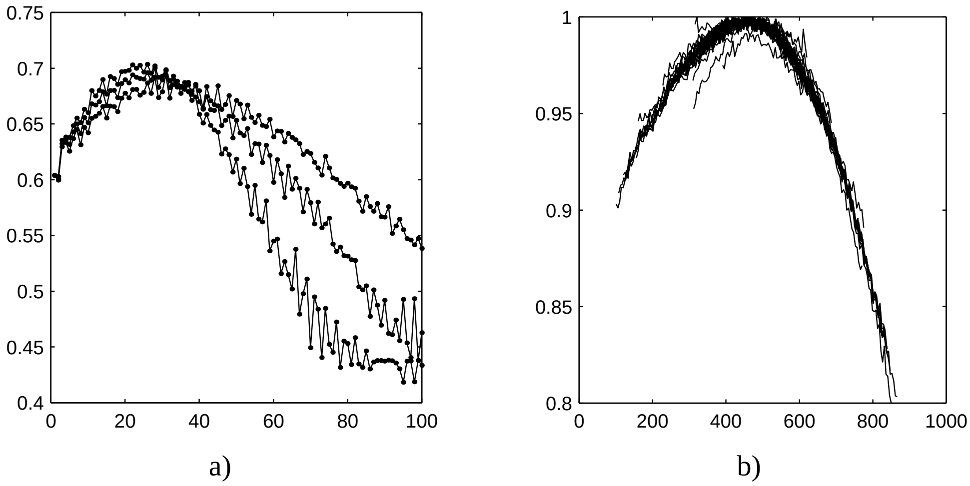


Figure 2: Intensity values related to the correct line in EPI. (a) Three intensity profiles. Horizontal axis gives t and vertical axis gives intensities. The top one corresponds to correct line (i, j). The other two profiles correspond to the modified lines $(i, j+1)$ and $(i, j-1)$. (b) Intensities along a several manually placed correct lines, normalized to maxima, and aligned according to position in image. Horizontal axis gives pixels and vertical axis gives intensities.

We noticed that the dominant shape of the profile depends only on the position in the image and does not depend on the position in the scene. We also observed that the intensity profiles are not symmetric. We concluded that the most significant phenomenon behind is photometric deficiency of the camera, namely field-darkening. Our decision was to perform proper photometric and radial camera calibration.

6.3 Photometric and Radial Calibration of the Camera

Field-darkening causes decrease of intensities for off-axis pixels. There are three main reasons for the phenomenon: (a) $cos^4\alpha$ effect, where α gives the angle between the ray and the optical axis, (b) vignetting effects which observed for open iris because not all iris is illuminated (likely absent in our case, F-stop = 12), and (c) pupil aberration (non-uniform light distribution across the aperture). The [1] concluded from experiments that the fall-off surface can be non-symmetric.

We obtained the fall-off surface by measuring the image irradiance depending on the position in the image for constant scene radiance. In [1] the homogeneous light source placed parallel to the camera was used. We could not use such a light source. We proposed alternative method. The small surface patch with almost Lambertian reflectance and almost homogenously illuminated was used as a light source. The camera was panned and tilted. Even if the light source is not exactly homogeneous the amount of light entering camera does not change because the camera center is at the same position. The measured surface is in Figure 3.

Having in mind that the method seeking correspondences reaches subpixel accuracy we also corrected the radial lens distortion. We used the RADIALD toolbox [8].

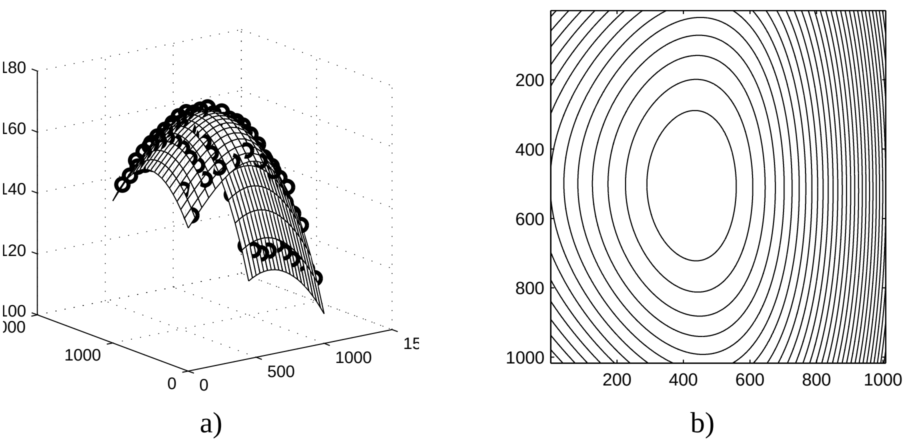


Figure 3: Measured fall-off phenomenon shown as the (a) 3D surface, and the (b) contour plot. The values (small circles) were measured in 10×10 regular grid and approximated by the second-order polynomial. Coordinates are given in pixels.

6.4 Do cameras lie on the straight line?

Theoretical model assumes many cameras lying on the straight line. Only one still camera is is used and the object moves in our case. This experiment validates the precision of the 'straight line' assumption.

Our experimental object, i.e. reverse side of the bathroom tile, contains calibration pattern (chessboard) which allows precise localization. We observed trajectories of selected location during movement, see Figure 4a. The y-differences of all targets from linear trajectory are depicted in Figure 4b. We conclude from uniformity of y-differences that our translation table is not precise enough and the 'straight line' assumption is significantly violated.

We attempted to suppress this effect. The mean value of y-differences for particular image in the sequence was used for shift the image. Figure 5 shows trajectories of the same selected calibration targets after shift correction. However, effects violating 'straight line' assumption cannot be undone in their entirety because the center of the camera changes position and the translation table moves in other directions too.

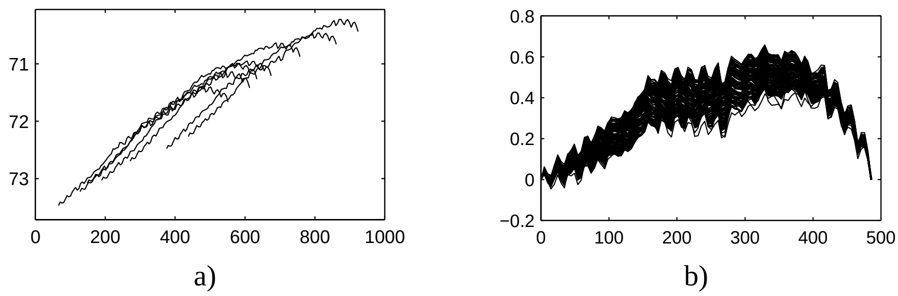


Figure 4: Plot of location of calibration targets during translation table movement. Coordinates are in pixels. (a) Trajectory of a few selected targets. (b) y-differences from linear trajectory for all targets.

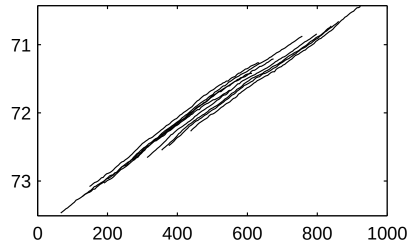
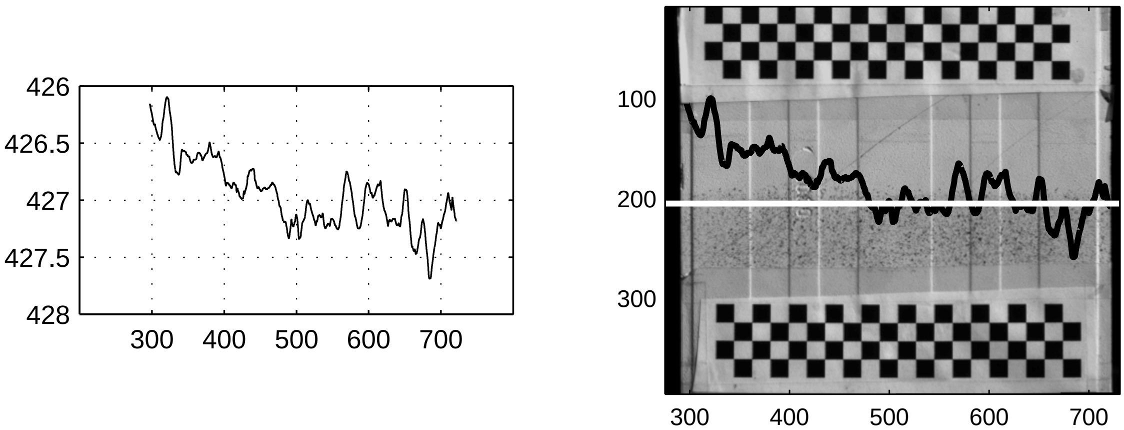


Figure 5: Plot of locations of a few selected calibration targets after shift correction. Coordinates are in pixels.

7 Final Results and Conclusions

All phenomena described above were understood, modelled, and suppressed. EPI made from corrected images was an input to the correspondence algorithm. The disparity plot is shown in Figure 6. The underlying depth variations can be detected but it is not so easy to distinguish them from noise.



a) b)

Figure 6: Computed disparity values. Both axes are in pixels. (a) Disparity only. (b) Disparity (scaled and shifted) overlayed on the original captured image. White line shows selected y-coordinate.

For comparison we calculated disparity using Šára's stereo matcher [10]. First and last images from the sequence were used. The method calculates correspondences with precision up to pixel correctly. Next the Šára's sub-pixel disparity correction (not published yet) was applied. In the resulting disparity map the 3D structure of the reverse side of the bathroom tile was not revealed at all.

Our method reveals depth variation where Šára's method does not but oscillates. It means that the solution is sometimes worse than Šára's stereo matcher at whole pixels.

To conclude, we admit that the phenomena involved are complicated. Even the idea aiming at finding distinct homogeneous lines in EPI is simple and appealing, it is difficult to meet our assumption in practical cases.

We like thing the problem over again and to do some more experiments in near future.

We acknowledge fruitful discussions with T. Werner, T. Pajdla, R. Šára, O. Drbohlav that oriented our work. V. Smutný helped us in setting experiments.

References

[1] Manoj Aggarwal, Hong Hua, and Narendra Ahuja. On cosine-fourth and vignetting effects in real lenses. In Jim Little and David Lowe, editors, *ICCV'01: Proc. 8th IEEE Intl. Conf. on Computer Vision*, volume 1, pages 472–479, Vancouver, British Columbia, Canada, July 2001. IEEE Computer Society Press, Los Alamitos, CA, USA.

[2] Simon Baker, Terence Sim, and Takeo Kanade. A characterization of inherent stereo ambiguities. In Jim Little and David Lowe, editors, *ICCV'01: Proc. 8th IEEE Intl. Conf. on Computer Vision*, volume 1, pages 428–435, Vancouver, British Columbia, Canada, July 2001. IEEE Computer Society Press, Los Alamitos, CA, USA.

[3] 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, Los Alamitos, CA, USA.

[4] 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.

[5] 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.

[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, Los Alamitos, CA, USA.

[8] Tomáš Pajdla, Tomáš Werner, and Václav Hlaváč. Correcting radial lens distortion without knowledge of 3-D structure. Research Report CTU-CMP-1999-4, Center for Machine Perception, Czech Technical University, Prague, Czech Republic, June 1997.

[9] 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.

[10] Radim Šára. Sigma-delta stable matching for computational stereopsis. Research Report CTU–CMP–2001–25, Center for Machine Perception, Czech Technical University, Prague, Czech Republic, September 2001.