Accurate Correspondences From Epipolar Plane Images

Martin Matoušek, Tomáš Werner, Václav Hlaváč

Czech Technical University in Prague
Center for Machine Perception
Karlovo nám. 13, 120 00 Prague, Czech Republic
http://cmp.felk.cvut.cz

Abstract. We present an algorithm for finding correspondences in epipolar plane images (EPIs). An EPI is a 2-dimensional spatio-temporal image obtained from a dense image sequence that is rectified so that each scene point is projected to the same row in all frames. Scenes with opaque Lambertian surfaces without occlusions are assumed. The approach is based on finding lines with similar intensities in an EPI for each image row separately, by dynamic programming. We focus on the correspondence accuracy. The high accuracy is enabled by a wide base line as in a stereo and by more data available. However, the matching is easier than in stereo as the displacement between neighboring frames is very small. No feature extraction is used, the algorithm is purely signal-based. We show the feasibility of the approach on synthetic data. A very good sub-pixel accuracy is demonstrated. The work in progress is reported aiming at using real data and sequences with occlusions.

1 Introduction

Finding image correspondences is fundamental for the 3-D reconstruction of real scenes. The approaches can be classified as follows. The first class is based on *stereo*, using a small number (often two) images far from each other. The second class is based on *tracking* features in image sequences [3] with large inter-frame displacements. In each frame, features (lines, corners) are detected and tracked, often using a priori knowledge about the camera motion in the tracking step. The third class uses *optical flow* computation. In a dense image sequence with typically sub-pixel inter-frame displacement the differential characteristics are computed on a local neighborhood of the current pixel, and the motion of the corresponding scene point is estimated from them.

Our approach combines advantage of wide baseline used in stereo and of using large amount of data as in optical flow computation.

Consider a dense image sequence, i.e., with small inter-frame displacements. It can be viewed as a 3-D *spatio-temporal* image. A special situation occurs if the sequence is rectified so that the projection of each scene point moves along a scene-independent curve in images. This allows to separate the 3-D spatio-temporal image to a set of 2-D *epipolar plane images* (EPIs) [7, 1, 6], of which each can be processed independently. The simplest such rectification causes the corresponding image points to be in a single scan-line in each image, allowing to process each row of the sequence separately. It is an analogy to epipolar rectification of a stereo pair. However, it is not always possible to rectify the whole sequence simultaneously.

The obvious approaches to finding correspondence in an EPI are feature tracking [1, 6] and optical flow. Features bring the well-known problem of feature invariance—a scene part can be detected as a feature in one image and remain undetected in another image. The optical flow finds the correspondence indirectly, by using 'image velocities'. The problem is in low accuracy and error accumulation caused by using differential characteristics computed on local neighborhoods of current pixels only.

Our approach avoids both these disadvantages—it is signal based (i.e., no features have to be detected). However, it uses the global information present in the whole EPI. It finds the optimal correspondence mapping by minimizing a total cost function using dynamic programming. The cost for a single match is computed from a large number of pixels rather than just from two as in the traditional stereo [3]. The optimal mapping is computed with sub-pixel precision naturally.

2 Epipolar Plane Images

An EPI is a spatio-temporal image obtained from a dense image sequence that is rectified so that each scene point is projected to the same row in all frames. Not every image sequence can be separated to EPIs. Examples of the situations when it *is* possible are as follows: *(i) an affine camera rotating around a fixed axis, (ii) a projective camera translating along a straight line.* To concentrate on the matching itself, we consider the EPIs as simple as possible, i.e. have chosen the (ii) setup. Moreover, the camera optical axis is orthogonal to the motion direction, so that the images are rectified without needing any other image-to--image homographies. Thus, images of a single scene point have the same vertical co-ordinate in all frames of the sequence.

Stacking the captured 2-D images in a 3-D block with dimensions x, y (columns, rows of images) and t ("time", a parametrization of a camera movement) forms a spatio-temporal 3-D image, Fig. 1. Each cut of this 3-D image is a separate EPI for a constant y. An important property of an EPI in our configuration is,that one spatial point makes a *straight line* in the EPI, formed by projections of this point in each image.

We assume that rows in the captured images are independent (i.e., no inter--line cohesivity constraints are considered) and we can work with each row separately. Thus, the search for the mapping can be decomposed to the search in each EPI separately.

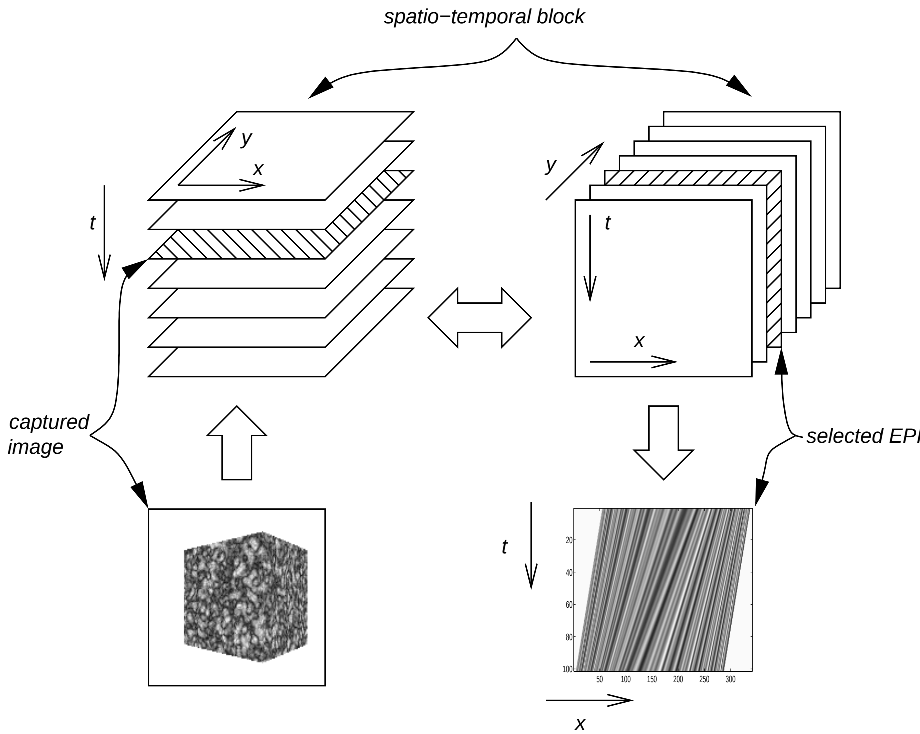


Fig. 1. Spatio-temporal block formed by captured images, and the EPI example.

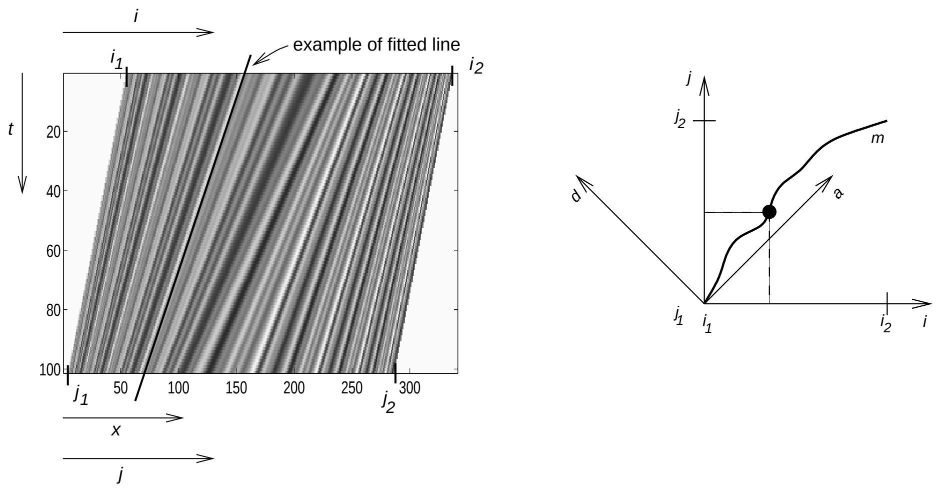
3 Matching

3.1 Definition, Search Space Parametrization

For a fixed y, i.e. one selected EPI, let i and j denote the x-axes in the first and last image of the sequence, respectively, Fig. 2a. A *single correspondence* is defined as a pair (i, j) in these coordinates. Such a pair corresponds with exactly one line in EPI, it means, that the position of this line unambiguously determinates the single correspondence, Fig. 2a,b.

Whole set of correspondences can be considered as a *mapping* between i and j. As no occlusion is assumed, ordering and uniqueness constraints apply [3]. Thus, the mapping is a strictly increasing continuous function. The mapping can be visualized as a plot in the i-j coordinate system, see Fig. 2b.

For further formulation and implementation is more convinient not to work with the system i-j, by with the system a-d originated by rotating 45 degrees counter-clockwise, Fig. 2b. The a-coordinate means diagonal of the original system, while the d-coordinate has the meaning of disparity. The two co-ordinate systems are related by $a = (j + i)/\sqrt{2}$, $d = (j - i)/\sqrt{2}$. Now, a single correspondence can be considered as a pair (a, d) and the mapping can be defined as the



a) b)

Fig. 2. (a) The EPI example with one straight line selected. (b) Correspondence space. The relation between $i-j$ and $a-d$ coordinates and one single correspondence as an element of a mapping.

continuous function

$$d = m(a)\,, \tag{1}$$

which is defined on whole set of a. Ordering constraint induces the condition

$$|m'(a)| < 1\,. \tag{2}$$

The mapping m is constrained by fixing its end points. The end points (i_1, j_1) and (i_2, j_2) are obtained by segmenting the object from the background. The found end points are shown in Fig. 2b.

3.2 Cost of a Single Correspondence

The cost $c(a,d)$ of a single correspondence (a,d) is defined as a function of the corresponding line in the EPI. This function should quantify "non-uniformity" of the set of intensities on the line. One intensity value for each (integer) t step is used. We have chosen the *variance* of the intensities as the cost function. However, other choices can be considered too.

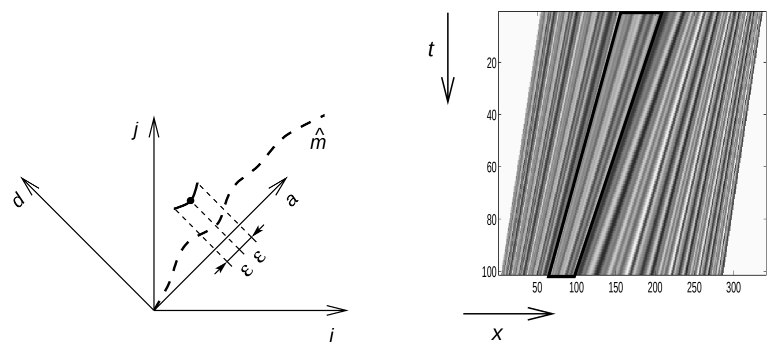
The domain of the function $c(a,d)$ is R^2. Since the lines in the EPI do not leed through points with integer coordinates, some (we use bilinear) interpolation of intensities must be used in any case and real domain of $c(s,d)$ does not care any additional complication.

Larger region of EPI than a single line can be used for evaluating a matching cost. We define the modified *single correspondence cost on the neighborhood* c_ε

as

$$c_\varepsilon(a,d) = \int\limits_{a-\varepsilon}^{a+\varepsilon} c(a', d + \widehat{m}(a') - \widehat{m}(a))\, da' \,, \tag{3}$$

where $\widehat{m}(a)$ is a rough estimate of the correct mapping m^*. It means, that the single correspondence cost c is integrated on a neighborhood ε (a-axis) on the estimate matching $\widehat{m}$ passed through a given co-ordinates (a, d), Fig. 3. Indeed, for $\varepsilon = 0$ it is $c_\varepsilon(a, d) = c(a, d)$.



a) b)

Fig. 3. (a) Evaluating the single correspondence cost on an ε-neighborhood, using an estimate matching $\widehat{m}$. (b) c_ε is integrated as a trapezoidal area in the EPI, although the area need not be sampled uniformly, which would correspond to linear $\widehat{m}$.

3.3 Matching Criterion

In an ideal situation, when images without any noise are captured with infinity resolution and the scene have a sufficient texture, the matching cost for and only for the correct match is zero. In a real situation, the matching must be find by minimizing some total cost function $J(m)$.

Considering, that mapping m is some *function* in the a-d system, a functional criterion can be defined. The total cost function is defined as

$$J(m) = \int\limits_{0}^{\sqrt{2}N} c(a, m(a))\, da \,, \tag{4}$$

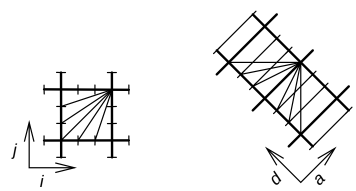
where N is the x-size of captured images and c is the cost. Both c or c_ε penalties can be used. This criterion was formulated such that for a constant matching cost c (a surface with constant observed intensity) no matching m from the set of all possible matchings is preferred, i.e., $J(m)$ is independent on m.

3.4 Dynamic Programming Solution

The optimal mapping m^* is obtained by a minimization of the criterion (4). This global minimum can be found by dynamic programming.

The task is different form ones often found in the papers on stereo by dynamic programming [2] in the fact that the search domain is real rather than integer. We could think of using the calculus of variations and solving a Lagrange PDE associated with (4) by established numerical methods, promising thus guaranteed accuracy, convergence, etc. However, for the sake of simplicity, we use a regularly discretized mesh, finer than for an integer-domain dynamic programming. Since $|m'(a)| < 1$, it is advantageous to have higher resolution in (a, d) search space in d-direction than in a-direction.

The optimal mapping curve m is approximated by a piecewise linear curve. A used search space discretization determines maximal length of a linear segment that approximates the corresponding part of the matching curve. It is appropriate to use denser discretization of an angle of the segment than only the horizontal, the vertical and the diagonal step. This leads to a quite unnatural discretization in case of the i-j co-ordinates (Fig. 4a). Solution in such parameterization is provided e.g in [4]. We use the a-d system discretization, whose implementation is straight-forward (Fig. 4b).



a) b)

Fig. 4. Discretization of the search space for the DP in the i-j (a) and the a-d (b) co-ordinate system.

4 Experiments on Synthetic Data

We tested the algorithm on synthetic data. Images are provided by the modified ray-tracer [5], together with the correct correspondences. Fig. 5 illustrates the experimental setup. A random granite-like texture was mapped onto the observed object. The experiments were carried out only for one selected EPI shown in Fig. 1b. Fig. 6 shows a zoomed-in part of function c in the i-j system (the part around the scene surface corner is selected) and the correct correspondence m^*, obtained from the ray tracer.

The computed mappings m found by minimization (4) are shown in Fig. 7—
—the same part of the surface is selected. The maximum and mean error is 0.5

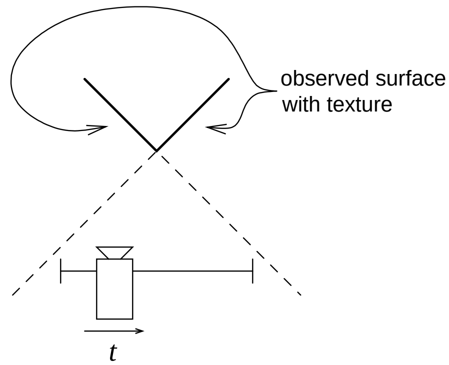


Fig. 5. Experimental setup, top-view.

and 0.1 pixels, respectively, for the single match cost, and 0.13 and 0.04 pixel for the the single match cost with a neighborhood $\varepsilon = 2$ pixels. The function c_ε (eq. (3)) was evaluated by approximating the integral by a finite sum, using 10 steps per pixel. Matching with the cost c_ε was computed by a simple iterative algorithm: At the begining, the initial matching $\widehat{m}^1$ was computed using a single-line cost function, c. Than $\widehat{m}^{k+1}$ was obtained from $\widehat{m}^k$ by solving (4) in which c_ε with $\widehat{m}^k$ was used. A few iterations suffice.

5 Conclusion and Outlook

We presented an algorithm for computing the correspondences between end images of a dense sequence, using all intermediate images between them. No feature detection is used, all pixel intensities directly contribute to the cost function. The optimal matching is found by minimizing a cost function using dynamic programming. The difference in optimization task formulation from the traditional dynamic programming stereo [2] is in the fact that the search space is real rather than integer valued. Moreover, the primitive cost is computed from a large number of pixels in all frames, rather than just two. This yields more accuracy.

Two matching penalties are introduced: the first using *single point* intensities and the second using a *small image neighborhood* rather than a single image point. In the latter case, the neighborhood shape is modified by the previously estimated matching. The matching cost uses all corresponding values from the sequence which enables to define it in sub-pixel accuracy.

Compared to the traditional stereo, the EPI based approach can obviously use the richer captured information for more reliable occlusion detection, for higher accuracy, and e.g. for resolving repeated-pattern-problem well-known in stereo [3]. As it is, the algorithm is suited for finding very accurate correspondence in an occlusion-free image sequence of a (nearly) Lambertian surface. 'Occlusion-free' exactly means that the same scene portion is visible in all frames of the sequence.

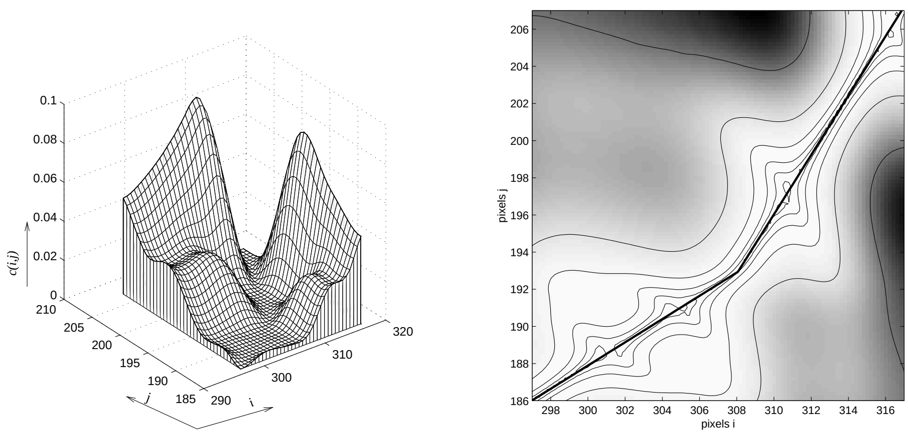
The situations where this condition is satisfied are not too frequent. However, the algorithm is meant as a *building block for a larger system*, allowing also for occlusions. The EPI based approach is especially suited for explicitly finding depth discontinuities in scenes—the evidence for a discontinuity is much richer in an EPI than in a sole pair of epipolar lines as in stereo. Once the regions in EPIs are found in which no discontinuities are present, the described algorithm can be used to find correspondence within these regions very accurately. Our contribution describes the part of a work in progress focused on the accuracy only.

Acknowledgements

This research was supported by the Czech Ministry of Education under Programme J04/98:212300013, and 4/11/AIP CR OCAMS, by the Grant Agency of the Czech Republic under Projects GACR 102/00/1679, and by the EU Fifth Framework Programme project 1999-29017 OMNIVIEWS.

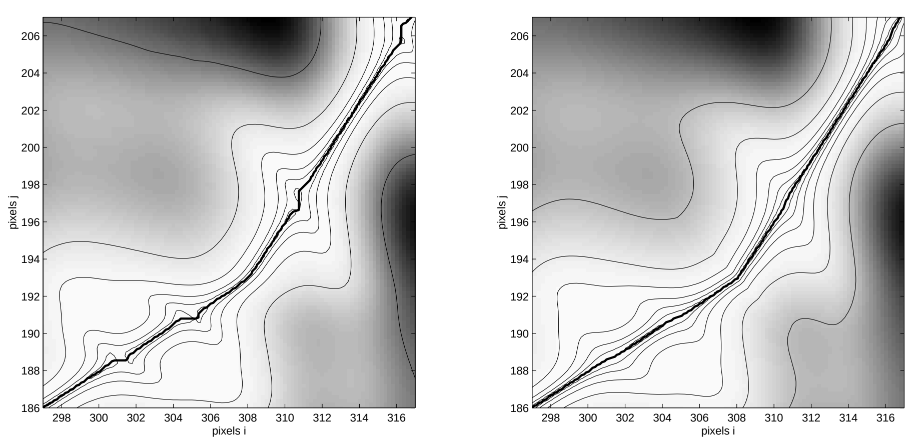
References

1. R.C. Bolles, H.H. Baker, and D.H. Marimont. Epipolar-plane image analysis: An approach to determining structure from motion. In *International Journal of Computer Vision*, pages 7–55, 1987.
2. Ingemar J. Cox, Sunita L. Hingorani, and Satish B. Rao. A maximum likelihood stereo algorithm. In *Computer Vision and Image Understanding*, pages 542–567, May 1996.
3. Olivier Faugeras. *Three-Dimensional Computer Vision*. MIT Press, Cambridge, Massachusetts, 1993.
4. Mario E. Munich and Pietro Perona. Continuous dynamic time warping for translation invariant curve alignment with applications to signature verification. In *ICCV'99: 7th IEEE International Conference on Computer Vision*, pages 108–115. IEEE Computer Society, September 1999.
5. Persistence Of Vision ray tracer. Free software, available from `http://www.povray.org/`.
6. 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.
7. GQ Zhou, J. Albertz, and K. Gwinner. Extracting 3D information using spatio--temporal analysis of aerial image sequences. In *Photogrammetric engineering and remote sensing*, pages 823–832, Bethesda, July 1999. American Society of Photogrammetry.



a) b)

Fig. 6. 3-D plot of selected part of the cost function c in the co-ordinate system i-j (a). True correspondence path (b), the underlying image showing the values of the cost function. White colour means lowest value, contours (with logarithmic scale) mark equal values.



a) b)

Fig. 7. Result of the DP matching using the single cost function c (a) and using the single cost function on a neighborhood c_ε (b). The cost functions are superimposed.