Matching in Catadioptric Images with Appropriate Windows and Outliers Removal

Tomáš Svoboda[1] and Tomáš Pajdla

svoboda@vision.ee.ethz.ch, pajdla@cmp.felk.cvut.cz

[1] Computer Vision Group, Swiss Federal INstitute of Technology,
Gloriastrasse 35, 8092 Zürich, Switzerland.

T.Svododa and T.Pajdla. Matching in Catadioptric Images with Appropriate Windows and Outliers Removal. *Proc. of the 9th International Conference on Computer Analysis of Images and Patterns. Springer Verlag. Warsaw, Poland, 5–7 September, 2001.*

Center for Machine Perception, Department of Cybernetics
Faculty of Electrical Engineering, Czech Technical University
Technická 2, 166 27 Prague 6, Czech Republic
fax +420 2 2435 7385, phone +420 2 2435 7637, www: http://cmp.felk.cvut.cz

Matching in Catadioptric Images with Appropriate Windows, and Outliers Removal

Tomáš Svoboda[1] and Tomáš Pajdla[2]

[1] Computer Vision Group, Swiss Federal Institut of Technology, Gloriastrasse 35, 8092 Zürich, Switzerland, svoboda@vision.ee.ethz.ch

[2] Center for Machine Perception, Czech Technical University, Karlovo nám. 13, CZ 121-35 Praha 2, Czech Republic

Abstract. Active matching windows for matching in panoramic images taken by a catadioptric camera are proposed. The shape and the size of the windows vary depending on the position of an interest point. The windows size is then normalized and a standard correlation is used for measuring similarities of the points. A semi-iterative method based on sorting correspondences according to their similarity is suggested to remove possible outliers. It is experimentally shown that using this method the matching is successful for small and also big displacement of corresponding points.

Keywords Image matching, catadioptric panoramic cameras, active windows.

1 Introduction

The process of establishing correspondences between a pair of images is called image matching and it is a well established topic in computer vision, see [2] or [14] for further references. The image matching consists of three main steps. First, points of interest have to be detected in both images. The Harris corner detector [5] is a commonly used algorithm for solving this task. Second, the interest points are mutually compared and the most similar ones are paired into the corresponding pairs, shortly *correspondences*. Some points are often matched incorrectly. These false matches, called *outliers*, have to be removed, which is the final stage of an image matching algorithm.

In this work, we are interested in the second two steps of the matching algorithm. We show that fixed, rectangular windows, usually used for cameras with perspective projections, are no longer appropriate for catadioptric cameras with a nonlinear projection. Similar observation is stated in [1] and homogeneous filtering for viewsphere is proposed. We propose active windows for measuring similarities between points. Similarities between points are determined by computing by the standard 2D correlation. For removing outliers, we suggest a sequential method based on sorting correspondences.

The text is organized as follows. Section 2 introduces appropriate matching windows for panoramic cameras. A method for removing false matches is explained in Section 3. Experiments with real cameras are shown in Section 4.

2 Windows for panoramic images

2.1 Why do not standard windows work?

The crucial premise to a successful image matching is the fact that point similarities are measured on the projection of the same 3D scene planar patch. Rectangular windows commonly used for matching in perspective projections are no more adequate to the catadioptric cameras. The main reason may be a big camera rotation around its vertical axis, which however, is quite common case. The big camera rotation invoke large movement of corresponding points between consecutive frames. It is worse when the camera movement is combined with some significant translation. Let us consider a 3D point in the scene. Its projections in the cameras could not only change its azimuth but also the the distance from the image center. The standard rectangular window with fixed size and accordingly oriented w.r.t. the image axis clearly does not represent the same surroundings of the scene point. Hence, the main assumption of the successful matching is not fulfilled.

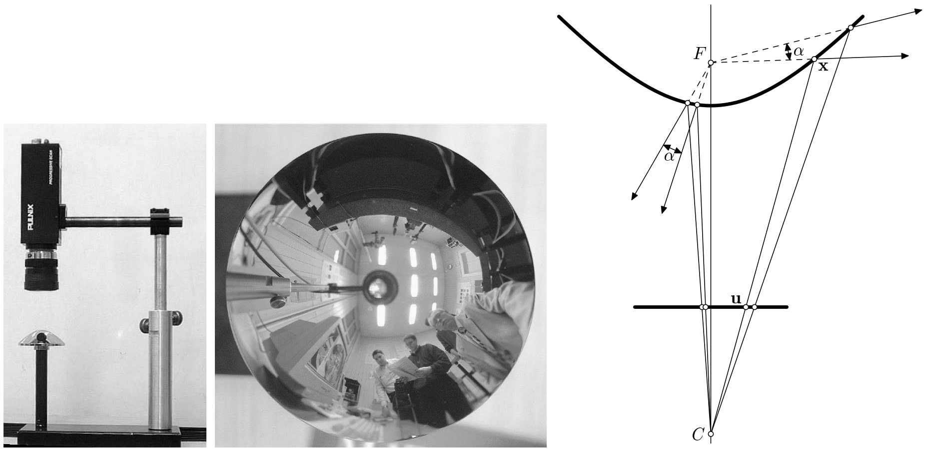

Fig. 1. Central catadioptric camera with a hyperbolic mirror (left) and a catadioptric image (middle). Size of the optimal window depends on the point's position (right). Parameters of the hyperbola are the same as for our real camera. For clarity, elevation range α is in the drawing enlarged to $12°$. F denotes the focal point of the mirror and C is the projection center of the camera.

2.2 Derivation of an appropriate window

Suppose that the displacement of the catadioptric camera center is significantly smaller than the depth of the scene. It means that the scale of the point's surroundings does not change as seen from these viewpoints, in other words, the surroundings is seen under the same spatial angle. This spatial angle intersects the mirror creating a small patch that, projected to the image plane, gives us the desired optimal window for matching. Note that the matching window will be bigger if the point is close to the mirror's periphery and smaller if the point is close to the image center, see Fig 1.

Statement 1 *Optimal (under assumptions stated above) window is bounded by four conics.*

Consider a surroundings of a space point bounded by lines. These lines project as conics in the image plane [3].

Practical considerations Debate spatial cones with the rectangular cross section. The dimensions of the cones are then defined by the *elevation* and *azimuth* range. These ranges play the role of the height and the width of the usual rectangular window and are to be preset, i.e. they will not be modified by the matching algorithm. The ranges has to be set carefully. Too small cones give us low descriptive information and more points will be mutually similar. However, too large cones highly probably breaks the assumption of locally planar neighborhood of the scene point. Moreover, they are computationally expensive during the matching process. One can afford relatively large cones, since our camera moves within a man made environment with a lot of planar structures, like walls, floors, ceilings or cabinets. The computation of conics would be impractical for the implementation. Let us put a constraint on the movements of cameras that is commonly satisfied in practice. We assume that the changes in the inclination of the cameras are neglectable. Therefore, we can assume that surroundings of the points are bounded by lines that are parallel and perpendicular to the common horizon of the cameras. Horizontal lines are projected to the conics with the common main axis, moreover, they can be approximated by the circles centered at the image center, since only small parts of the conics are used. Vertical lines project to the radial lines that intersect in the image center. Using this approximation, the points within the optimal window can be addressed by polar coordinates.

Computation of the window parameters Assume a point of interest, $\mathbf{u}$, found by, e.g., the Harris detector. This point is projected to the mirror point $\mathbf{x}$, see Fig 1. We omit the transformation formulas here in order to keep the derivation lucid. Interested readers are referred to [8, 10] for details. The Cartesian coordinates, $\mathbf{x} = [x, y, z]^\top$ are transformed to the spherical coordinates $\mathbf{x} = [\theta, \varphi, r]^\top$, where θ denotes the elevation, φ the azimuth, and r the distance of the point from the origin. Recall, that the origin is placed at the effective viewpoint, the focal point of the mirror. The corner points of the area made by the intersection of the spatial cone and the mirror surface are computed using the preset ranges of the azimuth and the elevation. Their radii r are computed from

$$r = \frac{-b^2}{e \sin\varphi - a}\,, \tag{1}$$

which is the equation in polar coordinates of the hyperbola centered at the focal point of the mirror, where a, b are the parameters of the hyperbolic mirror, and $e = \sqrt{a^2 + b^2}$. These corner points, after being projected back to the image plane, bound the window where the measure of similarity is to be computed. Note, that they do not bound a rectangular window but a part of an inter-circular area.

2.3 Measure of similarity

Direct comparison of different windows is not possible because of their different size and non-integer coordinates of their boundaries. Differently sized windows would be normalized to some shape and size that enables direct measure of similarity, e.g. the correlation or sum of square differences. We re-map the irregular windows to rectangular shape, being the azimuth on the horizontal and the radius on the vertical axis. The resolution of the transformed window is determined by the largest window that surrounds a point close to the periphery of the mirror. It means that the smaller window are mapped oversampled. No interpolation is used in this backward mapping. The normalized rectangular windows are then resampled to a smaller resolution using bicubic interpolation. These windows can then be fed into a standard algorithm that measures the similarity of points. We use the standard 2D correlation

$$c = \frac{\sum_{ij}(I_{ij} - \overline{I})(J_{ij} - \overline{J})}{\sqrt{\sum_{ij}(I_{ij} - \overline{I})^2 \sum_{ij}(J_{ij} - \overline{J})^2}} , \tag{2}$$

where I and J are intensities in the image windows and $\overline{I}$ and $\overline{J}$ are average intensities in the windows.

2.4 Pairing

Mutual correlation coefficients of all points of interest in two images are computed and arranged in the matrix $C_{m,n}$, where m and n denote the number of interest points in the first and the second image respectively. Maximal correlation coefficients are found in each row and column of the matrix. Consider that a maximum has been found in row r. Its position denotes a column c. If the maximum of this column lies in the row r then the points, indexed by r in the first and by c in the second image, are considered as 1-1 corresponding pair.

3 Removing false matches

We exploit the properties of the *essential matrix*. Each pair of corresponding points has to satisfy the following equation

$$\mathbf{x}_2^\top E \mathbf{x}_1 = 0 , \tag{3}$$

where $\mathbf{x}$ are point coordinates on the mirror and E is the essential matrix. A derivation of (3) can be found in [6] for perspective cameras, and in [11] for catadioptric cameras. Equation (3) is a homogeneous linear equation in elements of E. It can be rearranged into the form

$$A\mathbf{e} = 0\,. \tag{4}$$

Using the Singular Value Decomposition (SVD) [4] of A, $\mathbf{e}$ is obtained as the right singular vector corresponding to the smallest singular value. To improve the conditionality of the estimation, all mirror points are normalized to unit length $\|\mathbf{x}\| = 1$, see [8] for details.

Many robust methods were proposed in the last two decades for solving (4), since its form is the same for both calibrated and uncalibrated perspective cameras. Robust, RANSAC based methods are suggested in [12, 13]. These methods are proved to be very robust but require to know the percentage of outliers a priori and they are computationally expensive. They are useful in the case that a lot of correspondences is provided, since their complexity does not depend on the total number of correspondences.

Here, we propose a semi-iterative method that requires no assumptions about outliers. We assume that the higher correlated points make better pair than the less correlated points. This assumption may fail when highly repetitive patterns are presented in the scene. Thus, more constraints are necessary. Ranked correspondences are used to generate sample sets of the correspondences, S_i. First set S_1 contains the 9 best correspondences[1]. Each S_{i+1} set is then constructed as a union of S_i with some (one or a few) best ranked correspondences not yet involved in S_i. Maximally $N-9$ sample sets can be constructed from N correspondences, yet practically, if N is large, 5 sets seem to be enough. Adding more correspondences, S_{i+1} should generate better results than the set S_i (only Gaussian noise is present) until the first outlier is included into the growing set.

For each sample set of correspondences, S_i, an essential matrix E_i is computed from (4). The quality of E_i has to be measured with respect to its expected properties. The matrix E has, in the noise-free case, rank 2 and its nonzero singular values are equal [7]. We propose the difference of two largest singular values of E_i as a quality measure

$$q_E = \sigma_1 - \sigma_2\,, \tag{5}$$

where σ_1 and σ_2 are the two biggest singular values of E_i. The correspondences that were used for computation the best E_i, i.e. E_i with the lowest q_E, are denoted as inliers and the rest as *outliers*. A more detailed description of this outlier removal can be, with alternative methods for sorting correspondences, found in [10].

[1] We do not need to compute E from the minimal set of correspondences. Instead of using 7 points, we use 9 that much is more robust in the presence of the Gaussian noise.

4 Experiments

An example of the optimal window is shown in Fig 2. The elevation and azimuth range are both equal to 5°. The point of interest was selected manually at the corner of one computer screen.

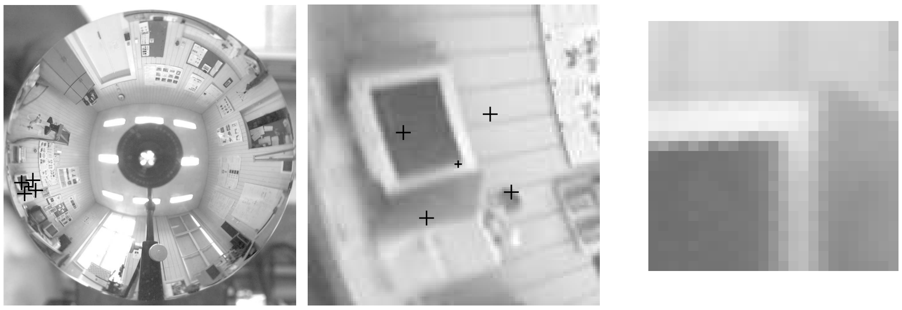

Fig. 2. The panoramic image is shown on the left. The point of interest is on the left, approximately in the middle of the image. Cropped surrounding around the point of interest. Four crosses denote boundary of the optimal window. The re-mapped and re-sampled normalized optimal window with a resolution of 25×25 pixels is shown on the right.

In the following experiments, all interest points were located by the Harris corner detector [5]. The smoothing parameter σ for image derivatives is set to 6.5, the detector parameter $k = 0.04$ is taken from [5] and the local maxima of the filter's response are searched in the 8-neighborhood of each point. The local maximum is considered as an interest point if it is greater then the threshold 10^4. All detector parameters above follow from the experiments presented in [9]. Around 200 interest points were found in images. Many points had no counterparts in the second image. Interest points lying outside the mirror and very close to its center (camera itself reflection) as well as points on the mirror holder were removed. The parameters of the active widows were $3° \times 3°$ spatial cone and 15×15 pixels normalized window size.

Fig 3 shows the reliability of the matching when the camera undergoes substantial rotation and the displacement of the corresponding points is very large. The amount of the rotation is best seen on the position of the window (very bright, saturated area). Points of interest are shown as small crosses. Possible 75 corresponding pairs were detected by the 1-1 pairing. The pairs were sorted according to their correlation coefficients and the essential matrix E was estimated with increasing number of pairs, see Figure 4. Together 17 points were denoted as correct pairs until the point No 46 was detected as the first outlier. The correctness of this decision can be visually verified on Figure 3.

The experiments with smaller displacement are omitted here because of space restrictions. The results are similar.

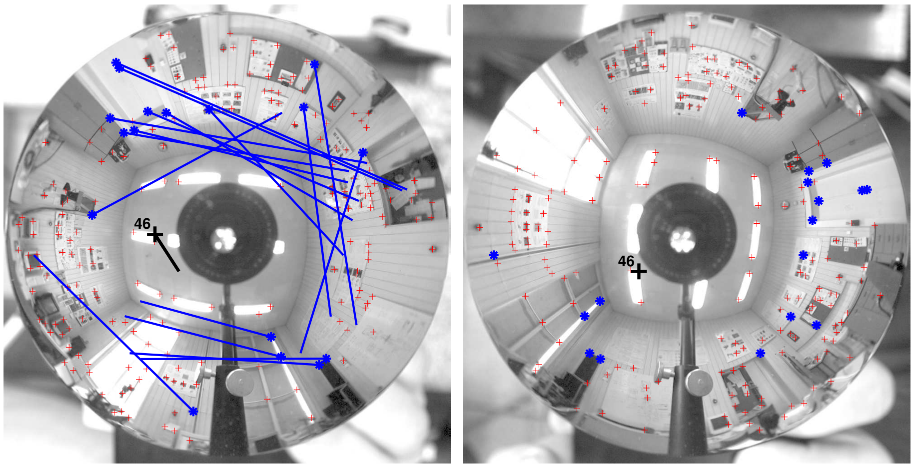

Fig. 3. The pair of the panoramic image. The movement between viewpoints includes a huge rotation around the vertical axis. The displacement of the corresponding points is larger than the half of the image. Positions of corrects pairs are denoted by the thick points. Lines show their displacement. The point No 46, denoted by a cross, is clearly an outlier because of the incoherent displacement.

5 Conclusions

Active windows for matching in panoramic images were proposed. Active shape of the window yields, even with simple correlation, good matching results also in the case substantial displacement of the corresponding points. The performances of the active shape overcomes ones of the fixed, rectangular window mainly in the case of substantial rotation and of the changes in the distance of the point from the image center changes. A semi-iterative method based on sorting correspondences was suggested to remove outliers.

Highly repetitive patterns in the scene may spoil the matching. Many pairs exhibit good similarity however, they can be mutually commuted. Alternative method for sorting correspondences were proposed in [10].

Acknowledgment This research was supported by the Czech Ministry of Education under the grant VS96049 and 4/11/AIP CR, by the Grant Agency of the Czech Republic under the grant GACR 102/01/0971, by the Research Program J04/98:212300013 Decision and control for industry, and by EU Fifth Framework Program project Omniviews No. 1999-29017.

References

1. J. Chahl and M. Srinivasan. A complete panoramic vision system, incorporating imaging, ranging, and three dimensional navigation. In *IEEE Workshop on Omnidirectional Vision*, pages 104–111. IEEE Computer Society Press, June 2000.
2. O. Faugeras. *Three-Dimensional Computer Vision*. MIT Press, 1993.

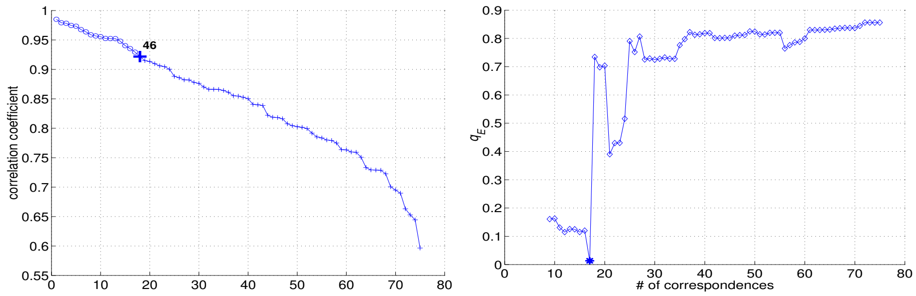

Fig. 4. Sorted correspondences on the left. Quality measure of the essential matrix on the right.

3. C. Geyer and K. Daniilidis. A unifying theory for central panoramic systems and practical implications. In D. Vernon, editor, *Proceedings of the 6th European Conference on Computer Vision*, number 1843 in Lecture Notes in Computer Science, pages 445–461, Berlin, Germany, June/July 2000. Springer.
4. G. H. Golub and C. F. v. Loan. *Matrix Computation*. The Johns Hopkins University Press, second edition, 1989.
5. C. Harris and M. Stephen. A combined corner and edge detection. In M. M. Matthews, editor, *Proceedings of the 4th ALVEY vision conference*, pages 147–151, University of Manchaster, England, September 1988.
6. R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK, 2000.
7. T. Huang and O. Faugeras. Some properties of the E matrix in two-view motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(12):1310–1312, December 1989.
8. T. Pajdla, T. Svoboda, and V. Hlaváč. Epipolar geometry of central panoramic cameras. In R. Benosman and S. B. Kang, editors, *Panoramic Vision: Sensors, Theory and Applications*, pages 85–114. Springer, 2001.
9. Z. Pohl. Omnidirectional vision – searching for correspondences. Master's thesis, Czech Technical University in Prague, Prague, Czech Republic, February 2001. In Czech.
10. T. Svoboda. *Central Panoramic Cameras Design, Geometry, Egomotion*. PhD Thesis, Center for Machine Perception, Czech Technical University, Prague, Czech Republic, April 2000.
11. T. Svoboda, T. Pajdla, and V. Hlaváč. Epipolar geometry for panoramic cameras. In H. Burkhardt and N. Bernd, editors, *the fifth European Conference on Computer Vision, Freiburg, Germany*, number 1406 in Lecture Notes in Computer Science, pages 218–232, Berlin, Germany, June 1998. Springer.
12. P. Torr and D. Murray. The development and comparison of robust methods for estimating the fundamental matrix. *IJCV*, 24(3):271–300, September 1997.
13. Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. *International Journal of Computer Vision*, 27(2):161–195, 1998.
14. Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luog. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence*, 78:87–119, October 1995. Also Research Report No.2273, INRIA Sophia-Antipolis.