Computer Vision Winter Workshop 2006, Ondřej Chum, Vojtěch Franc (eds.)
Telč, Czech Republic, February 6–8
Czech Pattern Recognition Society

[←]

Region matching for omnidirectional images using virtual camera planes

Thomas Mauthner, Friedrich Fraundorfer, and Horst Bischof

Institute for Computer Graphics and Vision, Graz University of Technology
mauthner@sbox.tugraz.at, fraunfri@icg.tu-graz.ac.at, bischof@icg.tu-graz.ac.at

Abstract *This paper proposes a new method for interest region matching in omnidirectional images, which uses virtual perspective camera planes. Perspective views are generated for each detected region depending on the region properties. This removes the distortions from the omnidirectional imaging device and enables the use of state-of-the-art wide-baseline algorithms designed for perspective cameras. We successfully applied our new method to mobile robot localization. Our approach was used to create landmark correspondences for motion estimation and map building. Experimental results for region matching, 3D reconstruction for map building and motion estimation will be shown.*

1 Introduction

One of the most important problems in computer vision is the search for correspondences in images from different viewpoints. This process of matching points or regions between different views of a scene is well researched for perspective images. Recently a respectable number of local image detectors have been developed (see [10, 20, 12, 9, 8, 11, 7]) to afford robust solutions for wide-baseline stereo, structure from motion, ego-motion estimation or robot navigation. The most advanced methods normalize each detected region using a local affine frame (LAF) to deal with the distortion introduced by perspective projection. Different methods for LAF constructions are possible and depend on the interest or region detector used. The first approach was introduced by Baumberg et al. [1] based on grayscale covariance matrices. Tuytelaars and van Gool proposed an ellipse-fitting method [19, 20] for intensity based and edge based interest regions. Odrzalek and Matas proposed several normalization methods e.g. [14] for their MSER detector. One of these used the covariance matrices of region borders and a direction found by an extremal point to normalize the regions shape

However, such methods cannot be directly applied to images of non-perspective cameras like omnidirectional imaging devices. Omnidirectional cameras introduce non-linear distortions because of their use of hyperbolic or other curved mirrors (see Figure 1 and Figure 5). Furthermore no normalization method has been established for omnidirectional cameras. On the other hand, omnidirectional cameras provide a lot of benefits. One advantage is the 360° field-of-view which results in the effect that corresponding points can only disappear in the case of an occlusion. One approach for matching in omnidirectional images is proposed in [18]. It uses adaptive windows around detected Harris points to generate normalized patches used for comparison, always supposing that the displacement of the omnidirectional system is significantly smaller than the depth of the surrounding scene, i.e. there is no significant scale change in the vicinity of the interest point.

As shown in [13] it is possible to generate a perspective view out of the information given by the omnidirectional image (if the mirror parameters are known). Several perspective views are generated using user-selected parameters like viewing direction, focal length and image size. In [6] a method is proposed to simulate the motion of virtual perspective cameras based on several omnidirectional images.

In this paper we present a method to apply standard wide-baseline region matching to images from a catadioptric camera system. In a first step interest regions are detected using a method which is unaffected by the non-linear distortions (like the MSER-Regions). Next, virtual perspective cameras are generated for each detected region and each region is re-sampled to obtain a perspective image. Finally standard methods can be employed for region matching. In our case we normalize the image with an LAF and extract SIFT descriptors [9] for feature matching.

Section 2 describes the camera model for the omnidirectional camera. In section 3 we describe the generation of perspective views using virtual camera planes. Section 4 gives the details about correspondence detection. Motion estimation and 3D reconstruction from omnidirectional images are outlined in section 5. Experimental results are given in section 6 and finally we draw conclusions in section 7.

2 Central Panoramic Camera

The catadioptric system used in this work is a combination of a perspective camera and a hyperbolic mirror. Geometric background of such a system and transformation between world coordinates and image coordinates are presented in [17, 15] or [4].
Figure 3 illustrates the geometric configuration. The camera is mounted in a way that its optical axis coincides with the z-axis of the mirror and the focal-point C lies in one of the two focal-points of the mirror F. This is achieved by a two step calibration process, in which first the perspective camera is calibrated to obtain the camera calibration matrix K and second the whole system is calibrated using an optical calibration method shown in [17]. Figure 3 also illustrates

the concept of the virtual camera plane (VCP). Instead of imaging rays projected onto the curved mirror surface we are imaging the intersections of the rays with a virtual plane. In general the position and rotation of the introduced camera plane can be arbitrary, however, it should be chosen in a way to minimize the perspective distortions.

3 Virtual Camera Plane

The use of virtual cameras for creating perspective views can be divided into following steps:

1. Detect regions in omnidirectional image
2. Compute mirror coordinates regions
3. Establish VCP for every region
4. Compute region border in virtual image
5. Resample mirror image to VCP

Since the geometry might be distorted the radiometry does not change too much between two views. Therefore the Maximally Stable Extremal Regions (MSER) provide the needed stability for distinguished regions under great viewpoint changes for omnidirectional images.

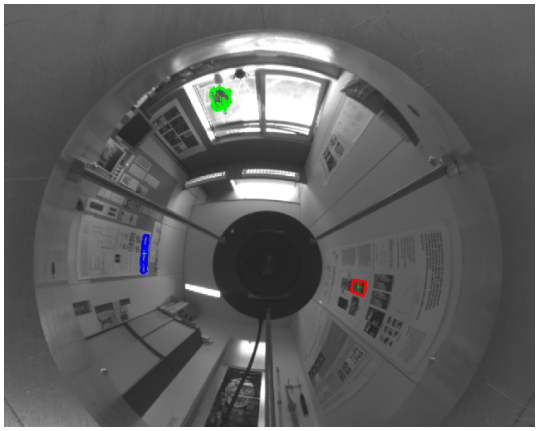


Figure 1: Detected MSER regions in an omnidirectional image captured in the room-sized environment used for the experiments.

We start by detecting regions using the MSER detector. For each detected region the center of gravity (CoG) in image coordinates is computed, using the pixels from the region boundary. The CoG of each region will be the origin of the coordinate system for the corresponding VCP. Next, the CoG and the regions border points are transformed into the coordinate system of the mirror, yielding the coordinates on the mirror surfaceX_{CG}. With the known 3D coordinates of the region a virtual camera plane (VCP) in 3d space can be defined for this region. The ray connecting the focal point of the mirror F' with X_{CG} defined by $v_{CG} = F' + \lambda X_{CG}$ is used as the normal vector of the VCP respectively the optical axis of the virtual camera. The origin of the VCP is given by X_{CG} but can also moved along the vector v_{CP}. The distance of the plane origin to F' defines the focal-length of the VCP. Two extra vectors $t_1 = [-v_{CG}(y), v_{CG}(x), v_{CG}(z)]$, normal to v_{CG}, and $t_2 = v_{CG} \times t_1$ as the vector-product between v_{CG} and t_1 are used as the basis-vectors to define the VCP in the mirror coordinate system. With the given origin coordinates of the plane given by X_{CG} and the two plane vectors t_1 and t_2 every point on the VCP can be reached by $X_{VCP} = aX_{CG} + bt_1 + ct_2$, where X_{VCP} is the 3D coordinate of a point on the VCP and b,c are the lengths of the vectors t_1 and t_2.

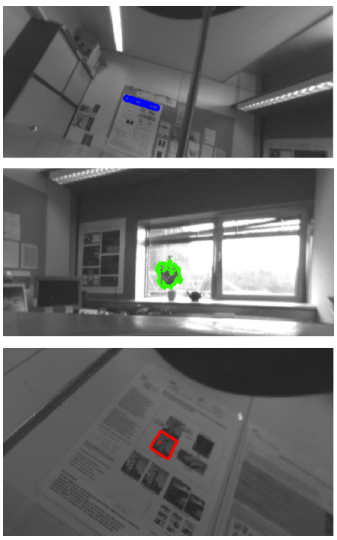


Figure 2: Virtual perspective images generated for the three selected regions shown in Figure 1. The shown images are bigger as used for region normalization to depict the effect of the removed mirror distortion more clearly.

The region border is transformed into image coordinates of the virtual image by directly computing the intersection of the rays going through the mirror points of the border and the virtual camera plane. They also define the size of the VCP. To generate the image of the virtual camera the points on the VCP are projected to the image plane and the intensity values are generated using bilinear interpolation. The discretization steps used for the basis-vectors define the geometric resolution of the VCP. Depending on the regions position on the mirror surface different resolutions are obtained for a constant discretization value. In this work the interpolation intervals on the VCP are set always constant and do not vary with the mirror coordinates.

4 Finding Correspondences

4.1 Region Normalization

Now the problem of finding corresponding regions in omnidirectional images is reduced to the correspondence problem in perspective images. This problem has been treated by many publications. The current solution is to resample the detected regions into a canonical coordinate system. Local affine frames presented for region normalization used by Tuytelaars in [19] or Obdrzalek and Matas [14] transform a given region into such a canonical frame, if possible. The region is assumed to be planar. For this case the affine transformation is a good approximation for perspective distortions of the detected regions. To establish such a local affine

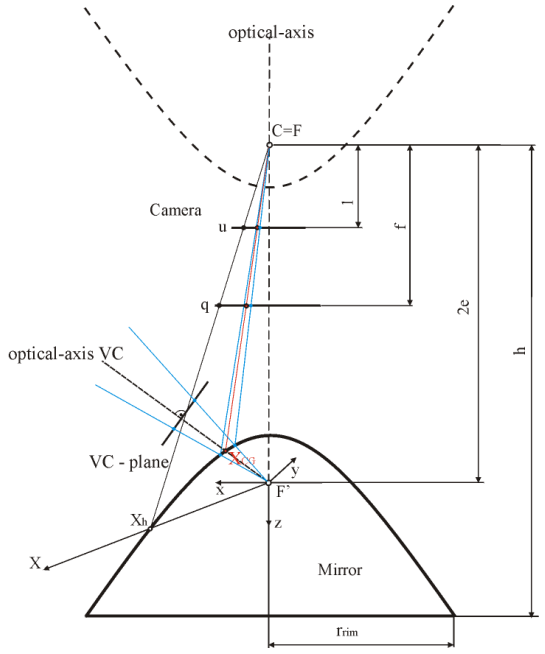


Figure 3: Geometric configuration of the used catadioptric system including the VC-plane concept.

frame the two-dimensional transformation from the actual region to the normalized form has to be found. To determine this transformation, 6 independent constrains have to be found, because of the six degrees of freedom of the 2D affine transformation.
Starting from the perspective view on the virtual camera plane, the center of gravity and the convex hull are calculated from the regions border points in coordinates of the VC-image.

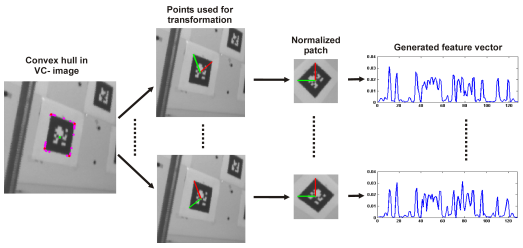


Figure 4: Normalization of a region. Centroid and convex-hull points are estimated from the region border in the VC-image. Multiple normalized frames are generated using different combinations of points. Finally feature vectors describing the normalized patches are computed using orientation histograms.

To establish the local affine frame (LAF), the CoG and two points of the convex hull, each point providing two constraints, are used to define the two basis vectors of the LAF, as described in [14]. The constructed LAF defines an affine transformation into a canonical coordinate system with rectangular axes and unit length. Resampling the image region with this transformation yields the normalized image patch. Each triple of CoG and two border points yields a new LAF. To limit the number of constructed frames per region, only boundary corner points are used. Results of a virtual camera image and normalization of a region are shown in Figure 4.

4.2 Matching of detected Regions

After the normalization several patches are generated for every region. Matching is done using SIFT descriptors extracted from the normalized image patches [9]. For every LAF normalization a SIFT feature vector is extracted. Corresponding features are detected by nearest neighbor search in the feature space. To increase robustness backward matching is performed. Regions satisfying the matching conditions in both directions are stored as corresponding. Examples of matched regions are shown in Figure 5.

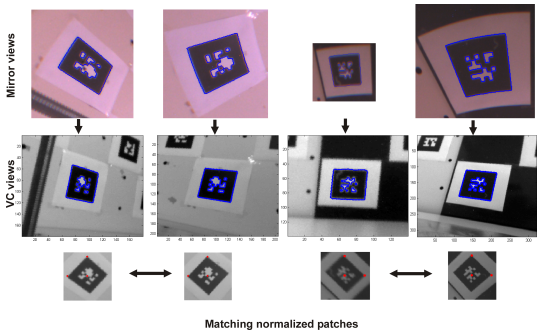


Figure 5: Matching results for view (first two columns) and scale changes. In the first row different views of two regions in the omnidirectional image are shown. The generated virtual camera images for each region are shown in the second row. As can be seen only perspective distortions are present in the images. Last row shows four corresponding normalized frames found by the matching process.

4.3 Point Correspondences

A drawback of using a region-based detector is that even for large regions only few points, in our case three, are known for a later 3D reconstruction. Also the accuracy of the used points is not necessary high. We therefore detect additional point correspondences within the matched image regions. For every pair of regions the two best matching frames are used. Harris points are detected in the first frame. Due to normalization the search area for a corresponding point in the second frame is known. There is no need to compute points also in the second frame, they can just be projected. For higher accuracy the correspondence is refined by area based correlation in the second frame. In addition sub-pixel accuracy is achieved by an interpolation method described in [16]. The process of finding corresponding points between matched normalized frames is shown in Figure 6.

5 Egomotion Estimation and 3D Reconstruction

The established correspondences from section 4 can be used to estimate the motion between the two given catadioptric cameras. The corresponding points of all regions are back-projected to their mirror surfaces and normalized to $\|X_H\| = 1$. The epipolar geometry for central panoramic cameras proposed in [17] allows to solve the egomotion, because we assume one camera has moved to another position, by solving a set linear equations. The corresponding points in mirror coordinates in the two images, known as X_{H1} and

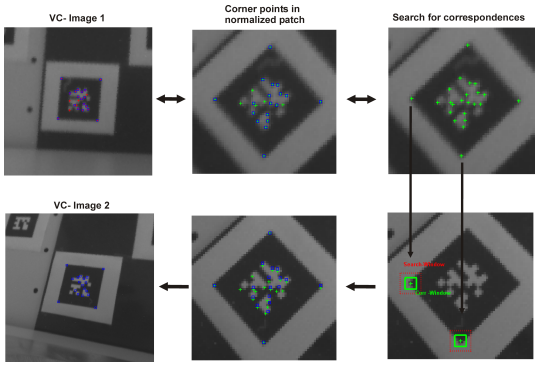


Figure 6: Corresponding points in normalized frames of two matched regions

X_{H2}, are used to estimate the essential matrix E by solving the equation $X_{H2}^T E X_{H1} = 0$. Rotation and translation can be extracted from the essential matrix (see [5]). To deal with possible outliers within the matched points a RANSAC algorithm [3] is used.

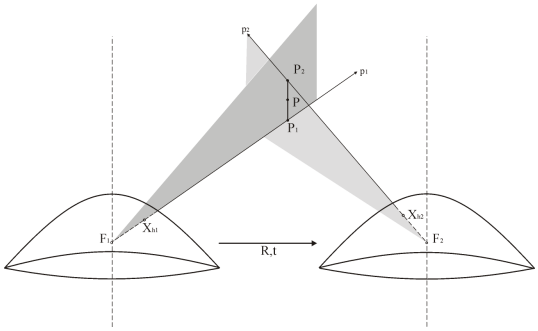


Figure 7: Mid-point reconstruction by using two corresponding points on the mirror surface.

To estimate the 3D coordinates from two corresponding points the mid-point method is used. The shortest transversal of the two rays given by $p1 = X_{H1}F_1$ and $p2 = X_{H2}F_2$ is computed (see Figure 7).

6 Experiments

In the experiments the proposed method is used on a mobile robot (Peoplebot), finding corresponding regions and points between two images in a real-world environment. One of our offices at the institute with size of about $15m^2$ is selected to be the test environment, including furniture, posters and other natural stuff as possible regions. The needed ground-truth for motion and reconstruction was created using the Peoplebots odometry and the laser range finder, refined in an off-line process using the commercial software ScanStudio from ActivMedia [1]. The virtual camera plane method for omnidirectional images is able to estimate motions within this environment without the need for any artificial landmarks resulting in a 3D reconstruction of detected regions.

First, we need to generate features for every image taken in the environment. To obtain useful results for motion estimation, detected regions should be spread as good as possible over the whole visible space. In an off-line task several matchings between positions are computed and motion estimation was done. In Figure 8 the results of the matching between two positions is shown. As expected we obtain good results on posters and parts of the furniture. Also homogeneous parts of the walls, reflections or areas around illuminations are matched. This results in a set of about 140 motions (results with a bad conditioned essential matrix are discarded). The translations between two images were among 300mm and 1100mm. The results of the egomotion estimation between images are transferred into the world coordinate system of the robot using the scale information of the odometry and the measured position in the robot path. In the world coordinate system the estimated positions can be compared with the given ground-truth and errors of the motions are computed. Estimated motions are shown in Figure 9. The solid black line defines the motion of the robot given by its odometry. Positions, where images are captured, are marked with green diamonds and numbered. Estimated motions are shown with dashed red lines and only motions with a motion error smaller 30mm are plotted for better visualization, which are about 30% of all motions. The egomotion estimation performs well in areas where corresponding regions can be detected with an adequate resolution. Especially the motions made between the first 9 positions near the wall and also positions from 25 to 30 perform very well. But also for long distances enough correspondences are matched, so good estimations are achieved e.g. $22 \rightarrow 25$ or $1 \rightarrow 5$ where we have baselines of 1061mm and 1092mm. Over all tested motions a median motion error of 45mm was achieved. The errors for all tested motions are shown in Figure 10. Most of the high error values in motion estimation are caused by degenerated alignment of points, e.g. a majority of the correspondences is located on a single plane or point correspondences are badly distributed.

From the estimated motion between two images and the corresponding points the reconstruction of the regions was done for all image pairs. The accuracy of the reconstruction depends on the accuracy of the motion estimation. The shown reconstruction in Figure 1 is made with a baseline of 725 mm and a rotation of 34° between the two mirror positions. The green plotted laser-map points gives us a visual ground-truth for the reconstruction, which shows that a really good reconstruction result can be achieved by our method. Points, lying on two opposite walls, are reconstructed with a good accuracy and even the flower pot at the window, also visible and matched in Figure 8, has a good position. As emphasized in our experiments also shown in [2], the reliable space for reconstruction and reconstruction accuracy increases with the baseline between two images, if matching accuracy is equal. We have to deal with the fact that we loss resolution and accuracy by moving over greater distances, but need this motion to get good reconstructions. As shown in the experiments, we fulfill the need for a matching method between two omnidirectional images with a wide baseline.

[1] ScanStudio is available from http://www.activrobots.com/

Figure 11: Reconstruction result for two images with a baseline of 725mm and a rotation of 34° given mm-coordinates. The green plot shows the map obtained by the laser range finder. Crosses mark the reconstructed matched regions and the circles mark the used camera positions. The matched regions are reconstructed at their correct position, most of them are posters at the walls (the laser map does not show the complete outline of the walls because of obstructions). Points lying in the motion direction of the robot are not plotted.

7 Conclusion

In this work we presented a new method for region matching in omnidirectional images. By introducing a virtual camera plane for each detected interest region it is possible to do matching as in the case of perspective cameras. Thus standard methods from wide-baseline stereo can be applied. In the experiments MSER regions were detected in the omnidirectional images. After re-sampling to perspective images, normalization by a LAF was performed and a SIFT descriptor was calculated. The experiments show that we obtain reliable and accurate region matches. The detected correspondences were successfully used for 3D reconstruction and motion estimation.

It should be stressed that the proposed method is not restricted to the catadioptric camera system we used. The method of the virtual camera planes can be applied to other configurations too.

Acknowledgement

Thanks to Hynek Bakstein from CMP for his valuable help in calibrating the omnidirectional camera system.

References

[1] A. Baumberg. Reliable feature matching across widely separated views. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head, South Carolina*, pages 774–781, 2000.

[2] P. Doubek and T. Svoboda. Reliable 3d reconstruction from a few catadioptric images. In R. Benosman and E.M. Mouaddib, editors, *Proceedings of the IEEE Workshop on Omnidirectional Vision 2002*, pages 71–78. IEEE Computer Society, 2002.

[3] M. A. Fischler and R. C. Bolles. RANSAC random sampling concensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of ACM*, 26:381–395, 1981.

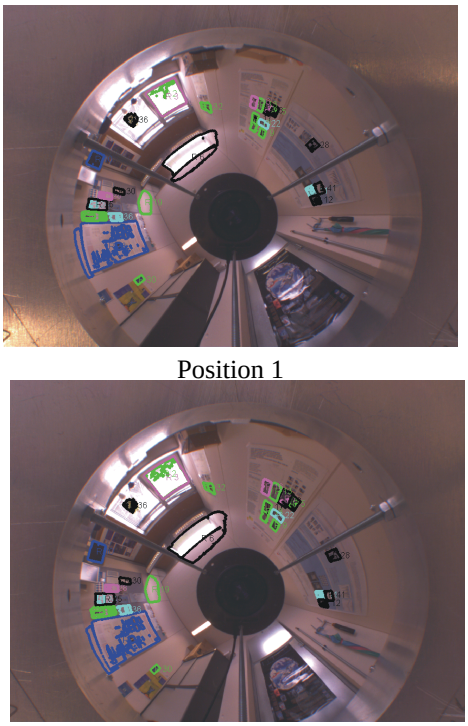
[4] J. Gluckman, S. Nayar, and K. Thorek. Real-time omnidirectional and panoramic stereo. In *DARPA Image Understanding Workshop*, November 1998.

[5] R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge, 2000.

[6] H. Ishiguro, K.C. Ng, R. Capella, and M.M. Trivedi. Omnidirectional image-based modeling: three approaches to approximated plenoptic representations. *Machine Vision and Applications*, 14(2):94–102, June 2003.

[7] T. Kadir, A. Zisserman, and M. Brady. An affine invariant salient region detector. In *Proc. 7th European Conference on Computer Vision, Prague, Czech Republic*, pages Vol I: 228–241, 2004.

[8] T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):79–116, 1998.



Position 1

Position 2

Figure 8: Matchings detected between position 1 and 2. Matching regions are plotted in equal colors and numbered identical.

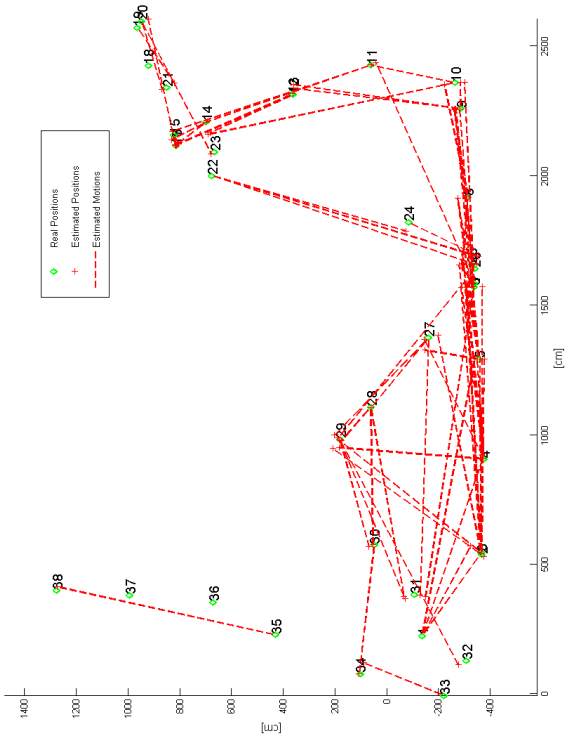


Figure 9: Motion estimation experiment: Motions between two camera positions (marked with green diamonds) with an error smaller 30 mm are plotted as red dashed lines. The right wall provides better features for matching, therefore more motion estimates achieved an error below 30 mm.

[9] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, November 2004.

[10] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *Proc. 13th British Machine Vision Conference, Cardiff, UK*, pages 384–393, 2002.

[11] K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. In *Proc. 7th European Conference on Computer Vision, Copenhagen, Denmark*, page I: 128 ff., 2002.

[12] Krystian Mikolajczyk and Cordelia Schmid. Indexing based on scale invariant interest points. In *Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada*, pages 525–531, 2001.

[13] Shree K. Nayar. Catadioptric omnidirectional camera. In *CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, page 482, Washington, DC, USA, 1997. IEEE Computer Society.

[14] Stepan Obdrzalek and Jiri Matas. Object recognition using local affine frames on distinguished regions. In *Proc. 13th British Machine Vision Conference, Cardiff, UK*, volume 1, pages 113–122, 2002.

[15] T. Pajdla, T. Svoboda, and V. Hlavac. Epipolar geometry for central panoramic catadioptric cameras. In *Panoramic Vision: Sensors, Theory, Applications*, pages 73–102, 2001.

[16] Roland Perko. *Computer Vision For Large Format Digital Aerial Cameras*. PhD thesis, Graz University of Technology, 2004.

[17] T. Svoboda. *Central Panoramic Cameras Design, Geometry, Egomotion*. PhD thesis, Czech Technical University, 1999.

[18] T. Svoboda and T. Pajdla. Matching in catadioptric images with appropriate windows, and outliers removal. In *CAIP '01: Proceedings of the 9th International Conference on Computer Analysis of Images and Patterns*, pages 733–740, London, UK, 2001. Springer-Verlag.

[19] T. Tuytelaars and L. Van Gool. Wide baseline stereo matching based on local, affinely invariant regions. In *British Machine Vision Conference BMVC'2000*, September 2000.

[20] T. Tuytelaars and L. Van Gool. Matching widely separated views based on affine invariant regions. *International Journal of Computer Vision*, 1(59):61–85, 2004.

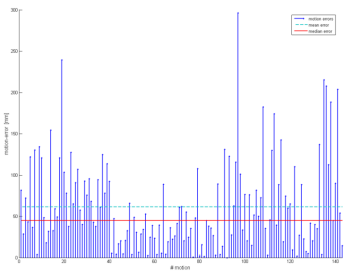


Figure 10: Errors of the estimated motions.