An Efficient Dynamic Multi-Angular Feature Points Matcher for Catadioptric Views

Sio-hoï Ieng, Ryad Benosman, Jean Devars
Laboratory of Instruments and Systems of Ile de France
University Pierre and Marie Curie, 4 place Jussieu,
75252 Paris cedex 05
sio-hoi.ieng@ufr924.jussieu.fr, benosman@ccr.jussieu.fr, devars@ccr.jussieu.fr

abstract

A new efficient matching algorithm dedicated to catadioptric sensors is proposed in this paper. The presented approach is designed to overcome the varying resolution of the mirror. The aim of this work is to provide a matcher that gives reliable results similar to the ones obtained by classical operators on planar projection images. The matching is based on a dynamical size windows extraction, computed from the viewing angular aperture of the neighborhood around the points of interest. An angular scaling of this angular aperture provides a certain number of different neighborhood resolution around the same considered point. A combinatory cost method is introduced in order to determine the best match between the different angular neighborhood patches of two interest points. Results are presented on sparse matched corner points, that can be used to estimate the epipolar geometry of the scene in order to provide a dense 3D map of the observed environment.

1 Introduction

Catadioptric sensors are more and more intensively studied since the last decade. By mixing reflective surfaces and cameras one can take 360° views in one shot, without any mosaicing requirement. For a general overview of omnidirectional techniques, the reader can refer to [8]. Images obtained by such a sensor suffer from non linear projective geometric distorsion. We can expect the classical image operators to be modified in order to handle this issue. In this paper, we revisit feature point matching process because the blind use of applied matchers, used for non deformed images give poor results. There is few work related to catadioptric sensor matching issues [7, 10]. Matching two pixels follows basically a same scheme that consists in computing a similarity score between two small neighborhoods around the pixels. But such a sensor has non homogeneous spatial resolution, as studied in detail in [5, 6]. Hence we developed a multi dynamical window based solution that overcomes the resolution problem. The shape and the dimensions of the windows have to be functions of the position of the point in the image. The windows are built on the mirror surface instead of generating them directly on the image plane. Neighborhoods are obtained by a backprojection on the image plane. Due to the circular aspect of the image, the approach has to handle substantial rotations problems as will be detailed in the following sections. This paper is organized as follows. Section two starts by giving a general overview of the problem. In section 3, we explain the delicate computation of neighborhoods around a feature point and introduces the 1-patch matching method. The dynamic multi-angle aperture matching approach is introduced in section 4. Finally section 5 gives the experimental results of both 1-patch and N-patch approaches and provides a comparison of both methods.

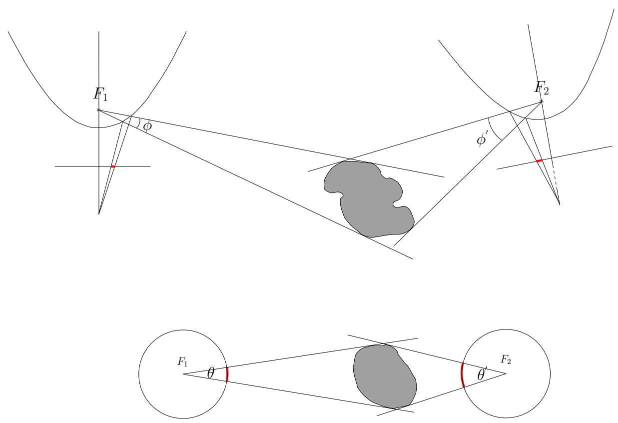

Figure 1: *Depending on the direction and the distance of the scene object, angle apertures will vary for each sensor.*

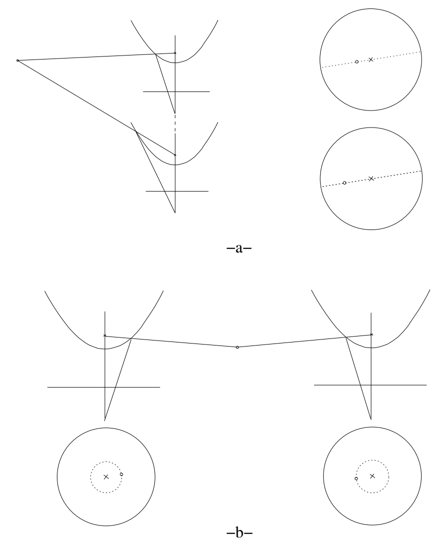

Figure 2: *Particular configurations where implicite epipolar geometry can be used to solve the search of correspondances.*

2 On the difficulty of matching catadioptric images

Catadioptric images need a new family of matching methods adapted to their resolution. While the linear planar sensor case can be solved under certain assumptions, this becomes much harder with catadioptric images. An example of special configurations can point out the main difficulties encountered with such sensors.
As shown by figure.1, the neighborhood patch size for each sensor is subject to change not only according to the distance to the object but also with its orientation. In most situations, there is no trivial way to solve this matching process. However, some configurations can make things easier.

- Considering the case where the sensors' optical axis are vertically aligned (see figure.2.a), epipolar geometry can be implicitely recovered. Since the epipolar plane is defined by the focal points and the scene point, the epipolar curve is the mirror profile. The search for correspondances can be constrained to a diameter on the pictures.

- Under the constraint of equidistance and planar motion, one can limit the search area to a circle of same radius (see figure.2.b).

For each configuration, the matching is made easier because of the implicit use of epipolar geometry. However, these required assumptions limits the use and capabilities of such sensors. In general configurations, blind search is applied. Hence a solution is needed in order to deal with the non homogenous resolution.

3 Dynamical neighborhood shape determination

Matching feature points is based on similarity measures between neighborhoods. As any motion in the scene is reflected on the image plane by pixels' displacements combining rotations and translations, it is then obvious that we have to face two main difficulties at the same time:

- fixing the windows' size with regard to the distance of the pixel to the center. Since the resolution changes along a radial direction, windows' size will change accordingly.

- reorienting the window. This is required for a reliable correlation computation. This is partially

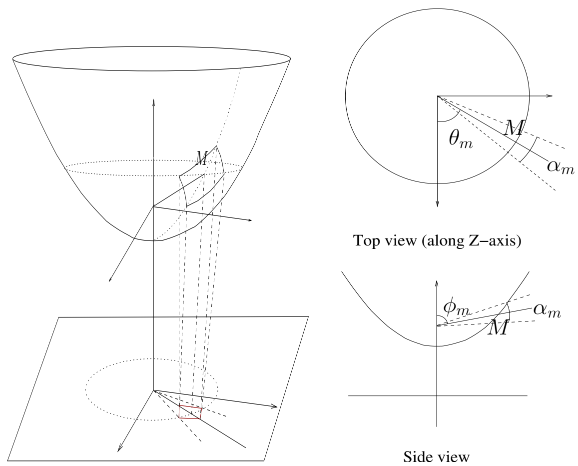

Figure 3: *Windows construction. A quadrilateral is defined on the mirror surface. Its vertices are obtained by setting elevation and azimuth ranges around the point O. The projection of this patch gives the matching window.*

treated in this paper as the matching orientation should also include a reorientation of the patches according to the determined epipolar geometry. This will be the topic of a future paper.

Hence, classical approaches [4] that use predefined size patches, often squares, exploiting or not the epipolar geometry, can not guarantee to retrieve the right surrounding of the pixel.
Appropriate windows are resizable patches, defined on the mirror surface and reprojected on the image plane [7]. Projections back and forth between the image plane and the mirror surface request a precise mapping between incident rays and their corresponding pixels in the image. As shown in [9], the single view point constraint of the sensor is seldomly fulfilled and can hardly be assumed valid if metrics is the issue. This is due to the mecanical fixing precision of the mirror and the discrete properties of the image. In order to provide a precise calibration of the sensor the method cited in [9] is applied and provides the intrinsics parameters of the sensor.

3.1 Appropriate windows

Square patches are an obvious shape used to define neighborhood windows because they are easy to handle. In the framework of our sensor, it is no longer a square but rather a diamond defined by its vertices. This construction is summarized in few steps.

- Feature points selection. This is usually solved by using corners detector (e.g. Harris detector).
- Back projecting these points to the mirror surface (see point M in figure.3)
- Fixed size quadrilateral definition, centered on mirror feature points.
- Vertices projection on the image plane

Corners extraction is provided by a classical Harris detector [1], and it is used “as it is", i.e. applied to the image disregarding its catadioptric propoperties. Since Harris is a high curvature detection and due to the fact that the hyperboloidal mirror introduces geometric distorsions, a more suited detector should be used. Works are being carried out, more detailed results will be provided in a future work. Back projection on the mirror can be achieved, since given an image pixel, we can compute the incident and the reflected ray associated to it (because of the previously calibrated sensor). The corresponding mirror point is the intersection of the camera ray and the mirror. The quadrilateral patch definition differs from classical approaches. We fix a characterized surface directly on the mirror rather than on the image plane. A point M lying on the reflective surface is determined by angular coordinates i.e. azimuth θ_m and elevation ϕ_m. Angular aperture α_m is fixed equally for both directions (see figure.3).
Projection of these windows on the image gives circles centered on the axis and lines along radial direction. For small angular ranges, such windows can be approximated by squares. According to the goal aimed for by setting the window on the mirror instead on the image, we are able to sustain the spatial resolution requirement. Dimensions of the window increase with the distance to the center, under the assumption of fixed angular parameters.

3.2 Matching and outliers removal

An optimal matching process requires good reconditioning of extracted neighborhoods. Each window is resized

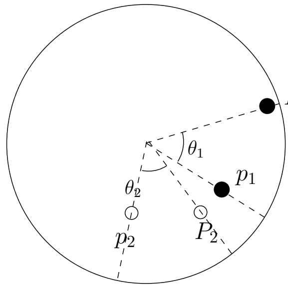

Figure 4: *Considering P_1 and P_2 matched respectively with p_1 and p_2. Projected in the same image plane, we can measure rotation angle θ_i associated to the displacement for eache couple. Assuming homogeneous displacement, θ_i will not differ much from a mean value $\overline{\theta}$.*

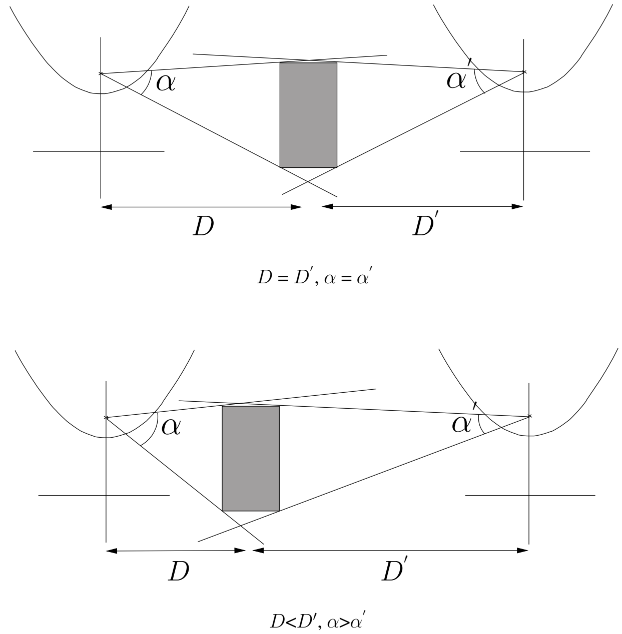

Figure 5: *Particular configurations of the position of objects in the scene performing translations. One can notice that the angular aperture α_m increases as the distance D increases.*

using a classical bilinear interpolation. In addition to this, existing methods imply the use of locally rotation invariant matching score (e.g. Zernike's moments [2]) for better results, but this is too computationally expensive, we will rather look for a fast and neat issue.
The defined mirror patches are constructed along a radial direction, we can interpolate the windows directly along these directions. Hence patches are correctly oriented and formatted at the same time before the matching process.
The correlation score computation should ensure one and only one best match for each point, we then use a centered and normalized cross correlation. Points are paired if they mutually give the best similarity scores.
This is still not enough to get reliable points. In order to remove outliers, a threshold is set to determine the measured angular disparity provided by matched points (see figure.4). Under the assumption that the error rate is moderate, point i is removed if the computed value $e^{-|\frac{\theta_i - \overline{\theta}}{2\pi}|}$ is not above a certain threshold. Setting the threshold on the mean angular value is only relevant in the case of homogeneous displacements. We can expect a defection of such methods if the scene contains more than one motion.

4 Multi angle-aperture approach

Let us consider the previous case where we set a same angular range for both views. One may wonder that such a choice could be too geometricaly restrictive and not be suited to all possible geometric configurations appearing in the scenes. This limitation covers the case when objects move but stay at a same distance to both sensors. This was used in a similar way as described in [7], and holds as it is used in the context of a mobile robot acquiring images at a high rate generating a small displacement between the two grabbed images. In a more general configuration, it becomes more difficult to determine the right couple of angular aperture to be applied on both images to perfectly suit the objects pose. As shown in figure.5, the more the scene object is far from a sensor, the more the angular aperture is small. Since it is a blind match, a judicious way to proceed is to compute a number of sampled angles values and find which combination of pairs provides the highest similarity score.

Using different patch size for matching, will logically improve the robustness. We then set a sample of angular values. The sampling unit value is α_1 it is set in our experiment to 0.05 rad. Each aperture is defined as $\alpha_i = i.\alpha_1$ as expressed by figure.6.

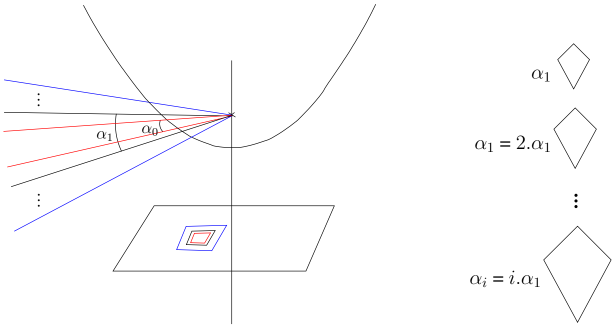

Figure 6: *Building matching windows by changing angular aperture. The α_i are ranges set for both azimuth and elevation. $\alpha_i = i.\alpha_1$ where α_1 is a step value used as an angular sampling unit.*

Matching several windows is robust, and provides a higher number of good match, this will be detailed in the following section. The computation increases according to the number of considered patches for each pixel. As shown by figure.7, if N angular ranges are used, we have to compute $2N$ patches. Each patch of the first image is compaired to each patch of the second one. A total of N^2 comparison is needed. N sets of matched points are obtained. We can then count through these sets, the number of times where two pixels are paired, this will implicitely give an index of confidence being the sum of their different matching scores. Different combinations of angular apertures will give differents pairs. Ambiguities are first solved using the index of confidence, then by looking which pairs provides the highest similarity score.

5 Experimental results

5.1 1-patch matching

Tests are performed on four couples of stereoscopic images. Extracted corners are used as matching features. Images include displacements as translations along X and Y axis and a huge rotation around Z. Without the knowledge of the extrinsics parameters, rotations around X and Y axis have to be moderated in order to provide potential matches. Tests are performed four times for differents angular apertures varying from 0,05 to 0,4 rad. The results expressed by Table.1 give a confirmation of the necessity of considering multiple angular aperture, as we can clearly see that the match is strictly related to the aperture.

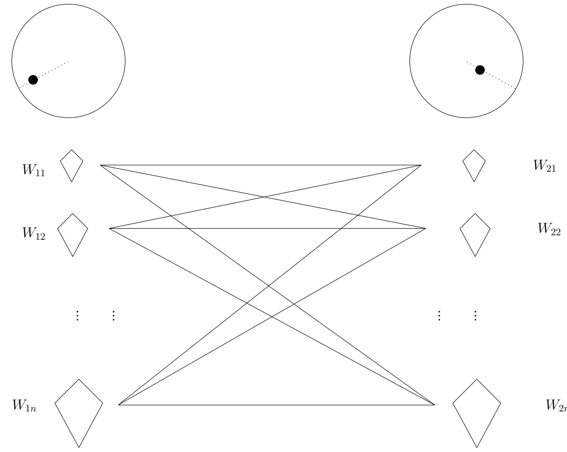

Figure 7: *N patches matching needs to perform N^2 times basic matching. Paired pixels are charaterized by their index of confidence and similarity score.*

5.2 N-patches matching

In this experiment N is fixed to 3, nine 1-patch matching are computed with $\alpha_i \in \{0.05; 0.10; 0.15\}$. A scene combining a huge rotation and a small translation is tested. A database of 24 corresponding points is generated and presented to both methods to ensure a comparison. In order to check the robustness, a wrong point is added. Table.2 gives the results of each method. The N-patches provides more correct matched pixels that the 1-patch approach, which in itself is an expected result. Unless a false match is detected many times despite the use of several aperture, results with a high index of confidence are most of the case strongly reliable. Since points are sorted by this score, potential outliers will be grouped at the end of the list of paired points. Finally Table.3 shows the matched pairs according to the best combination of angle aperture.

6 Conclusion

So far the presented method gives robust and reliable results. We have shown that the presented approach is able to provide an effective sparse robust matching algorithm. The results are comparable with the classical planar ones. The geometry of the sensor ensures the method to be nearly rotationately insensitive. Future work will combine this approach on catadioptric epipolar curves to produce dense 3D reconstruction of scenes.

Number of pts1	Number of pts2	angular aperture	Number of match	% of outliers	Observations
224	224	0.4	70	5.70	huge rotation, moderate translation
		0.2	58	3.45	
		0.1	54	18.50	
		0.05	33	18.87	
151	248	0.4	46	21.74	huge rotation, translation
		0.2	41	21.95	
		0.1	29	27.59	
		0.05	21	47.62	
224	263	0.4	78	12.82	huge rotation, small translation
		0.2	67	4.48	
		0.1	50	14.00	
		0.05	29	31.03	
299	372	0.4	120	3.33	small rotation, translation, scene with textures
		0.2	119	5.88	
		0.1	94	15.96	
		0.05	55	23.64	

Table 1: 1-Patch matching results according to the retained aperture and corresponding number of matched and outliers points.

	Number of pts1	Number of pts2	angular aperture	Number of pairs	number of outliers
Constant angular apertures	25	24	0.05	10	5
			0.10	13	0
			0.15	16	1
N-patches	25	24		17	0

Table 2: Comparison between N-patch and 1-patch matching methods on a database of 24 corresponding points. Results are presented with their corresponding apertures values and outliers.

angular apertures	0.05	0.10	0.15
0.05	1;12		3
0.10		9;11;4	5;7;14
0.15		4;8;15;10	2;6;8;13;16

Table 3: Results of matched pair associated to the best combination of angular apertures giving the highest correlation score.

References

[1] C. Harris and M. Stevens - *A combined corner and edge detector* - In Proceedings of the fourth Alvey Visino Conference, pages 153-158, 1988

[2] A.Khotanzad and Y.H.Hong. *Invariant image recognitionby Zernike moments* - IEEE Transaction on Pattern Analysis and Machine Intelligence, 12(5), May 1990

[3] O. Faugeras and S. Maybank - *Motion from point matches: multiplicity of solutions* - Technical report RR-1157, Inria, February 1990.

[4] Z. Zhang, R. Deriche, O. Faugeras, Q.-T. Luong- *A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry* - Technical Report RR-2273, Inria, octo-ber 1995.

[5] S. Baker and S. K. Nayar - *A theory of catadioptric image formation* - ICCV, pages 35-42,1998.

[6] S. Baker and S. K. Nayar - *A theory of single-viewpoint catadioptric image formation* - IJVC, 35(2):1-22, November 1999.

[7] T. Svoboda and T. Pajdla - *Matching in catadioptrique images with appropriate windows and outliers removal* - CAIP, pages 733-740, 2001.

[8] R. Benosman and S. kang - *Panoramic imaging: Sensor theory and application* - Spinger Verlag, 2001

[9] J. Fabrizzio, P. Tarel et R. Benosman - *Calibration of panoramic catadioptric that made easier* - In proceedings of the IEEE workshop on omnidirectional vision, june 2002.

[10] S.-K. Wei and M. Urban and T. Pajdla - *Stereo matching of catadioptric panoramic images* - Technical report, CMP, August 2000