Highly Specific Pose Estimation with a Catadioptric Omnidirectional Camera

Frédéric Comby, Olivier Strauss, Baptiste Magnier, Jean Triboulet, Cédric Demonceaux

Highly Specific Pose Estimation with a Catadioptric Omnidirectional Camera

Baptiste Magnier, Frédéric Comby, Olivier Strauss, Jean Triboulet
LIRMM Université Montpellier II
161 rue Ada, 34392 Montpellier cedex 5, France
Email: [magnier, comby, strauss, triboule]@lirmm.fr

Cédric Demonceaux
MIS, 7 Rue du Moulin Neuf,
80000 Amiens, France
Email: cedric.demonceaux@u-picardie.fr

***Abstract*—This article presents a new method for estimating the pose of para-catadioptric vision systems. It is based on the estimation of vanishing points associated with vertical edges of the environment. However, unlike classical approaches no feature (line, circle) extraction and/or identification is needed. A sampled domain of possible vanishing points is tested and histograms are build to characterize the soundness of these points. A specificity index allows to find the more relevant histogram and the pose of the sensor. This method has been tested on simulated and real images giving very promising results (maximum angular error of 0.18 degree).**

I. Introduction

Pose estimation is a classical issue in many robotics applications such as drone navigation and humanoid robotics. This topic is usually addressed using proprioceptive sensors such as inertial sensors: gyroscope and/or inclinometers, or exteroceptive sensors such as cameras. The main issue introduced by proprioceptive sensors is the bias drift of the measures it provides. This bias is due to the fact that pose estimation is achieved by integrating a noisy signal, which leads to an unbounded error growth.

Solutions have been proposed to partially cope with this bias drift. For example, in [1], the authors introduce a state estimation algorithm for rejecting noise and tracking bias of inertial measurement when the motion is intermittent. Another major drawback of inertial measurement is the precision of the pose measurement. Usually, classical performances of such devices are about one to five degree [2][3][4]. A more precise estimation may be useful, especially when the safety of the robot is engaged.

Using exteroceptive sensors, like perspective cameras, is a way to overcome this bias. In that case, the pose is estimated using 3D environment features at each frame and no integration process is needed. However, classical cameras often suffer from a limited field of view, which low their efficiency especially when features: lines, corners, have to be tracked all along a sequence. Solutions using non-conventional cameras (e.g. parabolic catadioptric cameras) have already been proposed to overcome this limited field of view matter [5]. These methods are usually based on tracking and identifying particular geometrical features such as lines in the omnidirectional image [6]. Such features may be difficult to track due to strong radial deformation brought by projections on the mirror and occlusions. The track can also be affected by the irregular resolution due to anamorphosis on the omnidirectional image.

This paper presents a new approach based on the pose estimation proposed in [7]. It consists in estimating the position of the two vanishing points associated to a set of vertical lines. In opposition to traditional approach, this method does not need any feature extraction. It is based on associating a particular histogram to each possible pose of the sensor. The property used here is the following: the more specific is the histogram, the more likely is the pose estimation. Thus the method consists in finding, in a set of possible orientations of the sensor, the one that is associated to the most specific histogram. Experiments on synthetic and real images are presented to illustrate the good performances we obtain with this method, i.e. a precision about 0.1 degree.

This article is organized as follows: the first section deals with the catadioptric sensor model we use and the relation between pose and vanishing points. The second section details the way histograms are build and the specificity indices we use to select the most probable orientation. The third section presents a synthetic and a real experiment.

II. Omnidirectional Catadioptric Vision

A. Central Catadioptric Model

Geyer and Daniilidis [8] have shown that an omnidirectional image is equivalent to a spherical image if the sensor satisfies the single view point constraint. The construction of this spherical image is depicted in Fig. 1. The projection of any point P of the 3D environment of the sensor can be decomposed into two projections: first P projects on the mirror (P_m) then this projection is projected on the image plane (P_i) via the optical device. Such a projection is equivalent to first projecting P on a unitary sphere centered on the mirror focus F (P_s is this projection), then projecting P_s on the image plane in P_i from the point O_s. Such an equivalent projection needs the catadioptric camera to be fully calibrated (i.e. know the intrinsic parameters of the projection and the parameters $\xi = |O_sF|$ and $\psi = |O_iF|$). Then, the image can be fully represented by the projection of all the points of the image plane on the equivalence sphere. This is what we call "the spherical equivalent image". This kind of model allows the use of the usual projective geometry defined for perspective cameras while using catadioptric cameras.

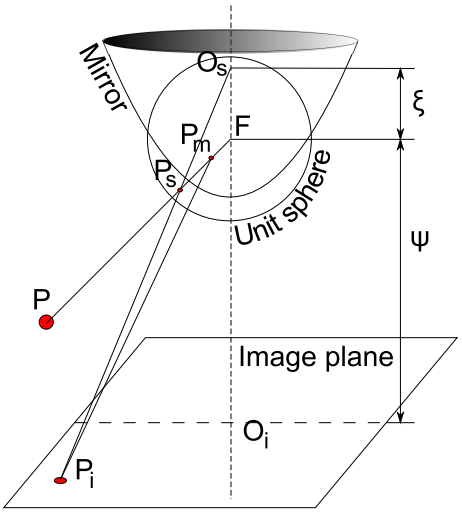


Fig. 1. Equivalence between catadioptric omnidirectional projection and two-step mapping via the sphere

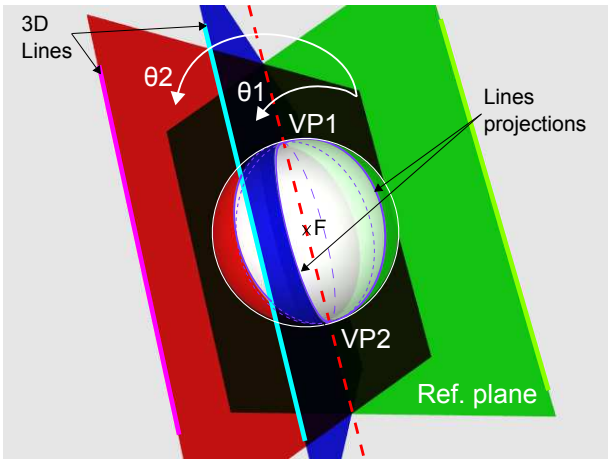


Fig. 2. Parallel lines projections on a spherical image. Each line produces a plane passing through the line, F and VP_1.

B. Relation between Pose and Omnidirectional Image

As proved in [8], a set of lines having the same direction $\overrightarrow{u}$ projects on the sphere into circles that intersect in two antipodal points VP_1 and VP_2 in the sphere, that corresponds to the vanishing directions. Demonceaux et al. [7] have proved that the direction of the vector $\overrightarrow{VP_1VP_2}$ is parallel to $\overrightarrow{u}$, and used this property to recover the 3D direction $\overrightarrow{u}$ of an unknown set of parallel lines (via a voting Hough-like process). Let us consider the set of all projections on the sphere of a set of points belonging to a 3D line. Then, each plane passing through one point of this set, the focus F and one of the antipodal points VP have exactly the same orientation θ (Fig. 2). They also contain the other antipodal point. Thus, if the environment is composed of a set of parallel lines, then a histogram of the orientations of the planes generated with every edge point of the image, F and VP_1 (or VP_2) is composed of peaks, each peak corresponding to a line direction. The reference plane is arbitrarily chosen and has no influence on the algorithm as we are only interested in its specificity, not on the location of peaks. If the environment is not composed only of parallel lines but also of a minority of other points, then these points will generate noise on the histogram.

III. A POLLING PROCESS FOR POSE ESTIMATION

A. Principle of the method

The idea that underlies our method is the following: as in indoors environment vertical edges are the most frequent, we will use their high density compared to other edge points to estimate the pose of the device. In fact, if the direction of the catadioptric device is known, so is known the position of the two antipodal points VP_1 and VP_2. Thus, the histogram of all the directions of planes passing through every edge point of the spherical image, F and VP_1 will be very specific. Indeed, the histogram will be composed of as many peaks as vertical edges of the environment and of a random noise due to other points not belonging to vertical edges (Fig. 3 (a)). A wrong orientation of the device will lead to mispositioning the two antipodal points VP_1 and VP_2. This mispositioning will alter the histogram and lower its specificity (Fig. 3 (b)). Thus, if the direction of the catadioptric device is unknown, then one can seek for this direction by constructing such a histogram for each possible direction. What rises from the geometrical properties of the spherical projection is that the more specific the histogram is, the more likely is its associated direction to be the true one (see Fig. 3).

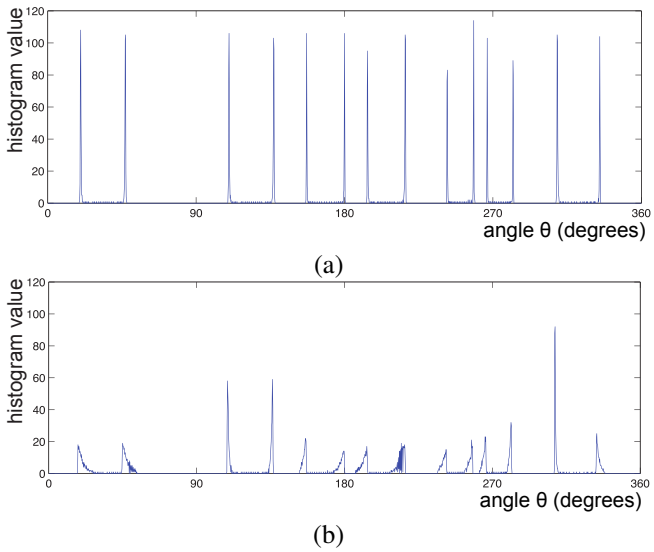


Fig. 3. Good and bad camera pose estimation leading respectively to a highly specific (a) and sparse (b) histogram of planes orientation.

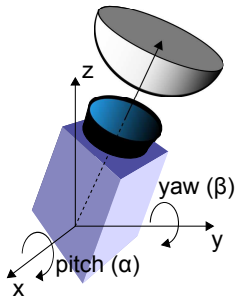


Fig. 4. Definition of the pitch and yaw angles for the camera orientation.

B. Construction of a histogram associated to each angle position

The overall method can be set as follow. First, the pose of the omnidirectional vision device is characterized by two angles α and β (e.g. pitch and yaw, see Fig. 4). We then define a sampled domain of all possible values of (α, β) and associate with each sampled ordered pair (α, β) a histogram. This histogram is built by computing the orientation of all the planes obtained by the triplet $(\overrightarrow{u}_{(\alpha,\beta)}, P_E, F)$, where $\overrightarrow{u}_{(\alpha,\beta)}$ is the orientation vector and P_E an edge point. After this construction, we chose for the pose, the value of (α, β) associated to the most specific histogram.

C. Specificity index of a histogram

In probability theory, it is instrumental to be able to order probability distributions from the point of view of their dispersion [9]. Numerous indices can be used to achieve this goal that are originally defined to make an information-based comparison between two distributions. The most popular of those indices are the Shannon entropy [10] and the Simpson index [11]. More recently, a specificity index has been rediscovered due to Birnbaum under the name of peakedness index [9] that is closely related to the Gini index. Due to the fact that a histogram can be seen as a density distribution estimator, such an ordering can also be used to compare two histograms from the point of view of their relative dispersion.
In this article, we consider the Shannon entropy and Simpson indices and the Birnbaum peakedness. Let $H = \{h_k\}_{k=1\ldots p}$ be the set of values of a histogram built on a partition made of p bins with n observations ($n = \sum_{k=1}^{p} h_k$). The Shannon entropy $\Phi(H)$, Simpson index $S(H)$ and Birnbaum peakedness index $B(H)$ are defined by:

$$\Phi(H) = log(n) - \frac{1}{n}\sum_{k=1}^{p} h_k.log(h_k), \quad (1)$$

$$S(H) = \frac{\sum_{k=1}^{p} h_k.(h_k - 1)}{n.(n-1)}, \quad (2)$$

$$B(H) = \frac{1}{n}\sum_{l=1}^{p}\sum_{k=l}^{p} h_{(k)}, \quad (3)$$

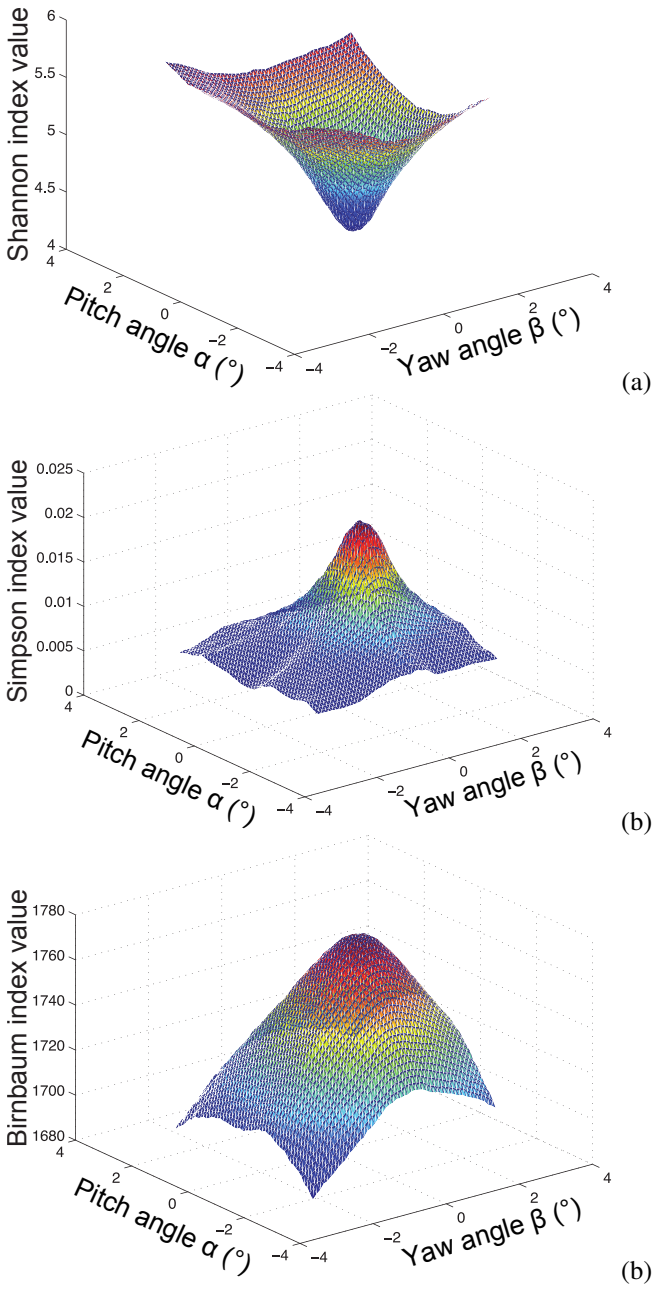
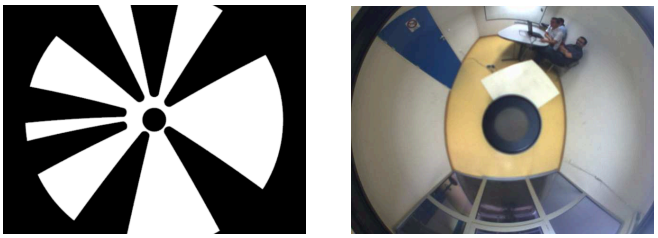


Fig. 5. Shannon (a), Simpson (b) and Birnbaum (c) indices for each ordered pair α and β of the [pitch, yaw] sampled space. The angles are in degrees

(.) being a permutation that sorts the $\{h_k\}_{k=1\ldots p}$ in ascending order, i.e. $h_{(1)} \leq \cdots \leq h_{(p)}$. Note that $S(H)$ and $B(H)$ are specificity indices while $\Phi(H)$ is a dispersion index. Therefore, finding the most peaked histogram will lead to maximize $S(H)$ and $B(H)$ and minimize $\Phi(H)$. Fig. 5 illustrates this property by plotting the three indices Shannon, Simpson and Birnbaum for all the possible pitch and yaw angles.



(a) (b)

Fig. 6. (a) Synthesized catadioptric image. (b) Exemple of real image tested.

IV. Experimentations

A. Experiment with Synthetic Images

The first experiment has been carried out with a synthetic image sequence. This sequence has been obtained by simulating, with the PovRay software, an environment made of seven black piles (Fig. 6(a)), and rotating the omnidirectional device around an axis perpendicular to its symmetry axis with a step of $\Delta\theta = 0.1^o$. In this experiment, since the omnidirectional vision sensor has been simulated, the projection and mirror parameters are perfectly known and thus the experimental results will not be altered by an ill calibration of the sensor. Table I presents the result obtained with smoothed and non smoothed specificity indices.

B. Experiment with Real Images

The second experiment has been carried out with a sequence of real images obtained by rotating a para-catadioptric sensor with a step motor with an accuracy of 0.001^o. As can be remarked on Fig. 6(b), the image is not made only of vertical lines. In this experiments, all edge points of the image have been used. Due to this fact, the obtained histograms are very noisy, so are the indices. Thus, it can be appropriate to smooth the indices before the sought for extrema: for example a 2D exponential smoothing filter (4) can be used. γ was set to 0.1 for the experiments, which is almost equivalent to a ± 10 pixels horizon filter.

$$H(x, y) = Ke^{\gamma(|x|+|y|)} \tag{4}$$

Table II presents the results obtained with smoothed and non smoothed specificity indices. The angular step of the rotation is also $\Delta\theta = 0.1^o$.

C. Experimental results

For both real and simulated experiments, the angular histogram has been built on a high specific partition made of 1800 bins (i.e. the bin width equals 0.2^o). To characterize the accuracy of the method, three dispersion indices have been used: the standard deviation, $\sigma = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(\alpha_k^{measured} - \alpha_k^{true})^2}$, the mean absolute error, $|E| = \frac{1}{N}\sum_{k=1}^{N}|\alpha_k^{measured} - \alpha_k^{true}|$ and the maximal error $E_{max} = \max_{k=1}^{N}|\alpha_k^{measured} - \alpha_k^{true}|$. The simulated experiments shows that any specificity index can be used to achieve the sought for orientation since the obtained errors are comparable. This is illustrated in Fig. 7. As the yaw angle β remains constant during these experiments, and to ease the comparison, we only represent a cut of the three indices in the plane $\beta = 0$. All indices have been smoothed and normalized to be compared. Shannon index has also been inverted to ease the comparison. As we can see on Fig. 7, the Birnbaum index is a little more noisy than the two others. However, the extrema of each index are almost localized at the same place.

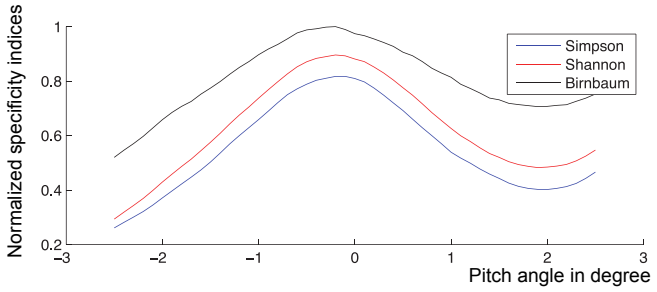


Fig. 7. Comparison of the three cuts of specificity indices in the plane $\beta = 0$.

Errors obtained on simulated sequences are mainly due to sampling noise introduced by image or histogram sampling. Within the real experiments, it appears that the Shannon and Simpson indices are less sensitive to the noise induced by minority edges than the Birnbaum index. When comparing the results obtained with and without a preliminary smoothing, it appears clearly that smoothing induces a gain in accuracy: it can go up to 0.18^o when using the Simpson index. This accuracy is also strongly related to the sampling step of the (α, β) space. The smaller the sampling step is, the better the accuracy is. For example the mean absolute error drops down from 0.11^o to 0.08^o with the Shannon index when the step on (α, β) goes from 0.1^o to 0.05^o. Nevertheless, the computing time increases with the size of the discretized (α, β) space. Under 0.05^o the accuracy seems to stabilize. These results are to compare with those presented in [6], where the mean error in orientation is about 0.1 and the maximum error less than 2. However, the scene is mainly composed of two calibrating patterns and presents a lot of parallel lines. This highly facilitates the pose estimation. In [7], the authors have no available ground truth to estimate errors. However, for a horizontal sensor they measured errors less than 2 degrees in both directions. So our algorithm seems to be more accurate. Obviously, comparisons of algorithms with similar conditions and sequences should be performed.

V. Conclusion

In this paper, we have proposed a new method for omnidirectional camera pose estimation. The main characteristic of this method is that it does not need any feature identification like lines or circles. It is based on estimating the orientation of the catadioptric device by exploring a sampled space of different possible orientations and associating a histogram to each orientation. The histogram construction is based on

TABLE I
ESTIMATION ERROR: EXPERIMENTATION WITH THE SIMULATED IMAGES WITH A (α, β) DOMAIN SAMPLING STEP OF 0.05^o

Sp. Index	smoothed	$\lvert E \rvert$	σ	E_{max}
Shannon	no	0.056^o	0.075^o	0.20^o
Simpson	no	0.063^o	0.082^o	0.20^o
Birnbaum	no	0.063^o	0.080^o	0.20^o
Shannon	yes	0.050^o	0.068^o	0.20^o
Simpson	yes	0.057^o	0.075^o	0.20^o
Birnbaum	yes	0.060^o	0.076^o	0.20^o

TABLE II
ESTIMATION ERROR: EXPERIMENTATION WITH THE REAL IMAGES WITH A (α, β) DOMAIN SAMPLING STEP OF 0.05^o

Sp. Index	smoothed	$\lvert E \rvert$	σ	E_{max}
Shannon	no	0.082^o	0.106^o	0.25^o
Simpson	no	0.080^o	0.111^o	0.40^o
Birnbaum	no	0.134^o	0.182^o	0.60^o
Shannon	yes	0.059^o	0.081^o	0.20^o
Simpson	yes	0.053^o	0.073^o	0.20^o
Birnbaum	yes	0.075^o	0.120^o	0.45^o

the particular geometrical properties of single viewpoint catadioptric cameras. Due to these properties, the specificity of the histogram reflects the soundness of the orientation it is associated with. Different specificity indices have been tested. From these tests, it appears that, in this context, the well known Shannon entropy provides the best results. Within both simulated and real image based experiment, this method can provide a pose estimation with a mean resolution of 0.06 degree. The resolution of the estimation is naturally highly linked to both sampling of the orientation space and noise in the images.

The test we achieved prove both feasibility and efficiency of this approach. However, a major drawback remains: the computation time can be quite long: about 19 seconds on a 2.13 GHz processor for a 640x480 image, 2600 sample of the orientation space and 1800 bins for each histogram. First of all optimizing the implementation can improve the overall performances. Second, we consider using histograms built upon fuzzy partition rather than binary ones [12] in order to reduce the number of histogram's bins without reducing the precision of the specificity index. Third, we propose to use the gradient orientation to remove, from the set of edge pixels, any point having an orientation that is too far away from a radial orientation (which should be the orientation of vertical edges). This third modification should drastically reduce the computation time and the noise in histograms. Finally, we propose to dynamically define the sampled orientation domain by using a Kalman filtering-based approach. It can allow a higher precision in the exploration domain and reduce also the computation time.

ACKNOWLEDGMENT

The authors would like to thank the ANR SHERPA project for their fundings.

REFERENCES

[1] E. A. Johnson, S. J. M. Bamberg, and M. A. Minor, "A state estimator for rejecting noise and tracking bias in inertial sensors," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, 19-23 May 2008, pp. 3256 – 3263.

[2] N. Metni, J.-M. Pflimlin, T. Hamel, and P. Soueres, "Attitude and gyro bias estimation for a vtol uav," *Control Engineering Practice*, vol. 14, no. 12, pp. 1511 – 1520, 2006.

[3] G. Muscato and G. Spampinato, "Kinematical model and control architecture for a human inspired five dof robotic leg," *Mechatronics*, vol. 17, no. 1, pp. 45 – 63, 2007.

[4] I.-W. Park, J.-Y. Kim, B.-K. Cho, and J.-H. Oh, "Control hardware integration of a biped humanoid robot with an android head," *Robotics and Autonomous Systems*, vol. 56, no. 1, pp. 95 – 103, 2008.

[5] J.-C. Bazin, C. Demonceaux, P. Vasseur, and I.-S. Kweon, "Motion estimation by decoupling rotation and translation in catadioptric vision," *Computer Vision and Image Understanding*, vol. In Press, Corrected Proof, 2009.

[6] B. Vandeportaele, M. Cattoen, and P. Marthon, "Realtime localization of a central catadioptric camera using vertical lines," in *International Conference on Computer Vision Theory and Applications (VISAPP)*, Setubal, Portugal, February 2006, pp. 416–421.

[7] C. Demonceaux, P. Vasseur, and C. Pégard, "Uav attitude computation by omnidirectional vision in urban environment," in *ICRA*, 2007, pp. 2017–2022.

[8] C. Geyer and K. Daniilidis, "Catadioptric projective geometry," *International Journal of Computer Vision*, vol. 45, no. 3, pp. 223–243, december 2001.

[9] D. Dubois and A. Hullermeier, "Comparing probability measures using possibility theory: A notion of relative peakedness," *International Journal of Approximate Reasoning*, vol. 45, no. 2, pp. 64–385, 2007.

[10] C. Shannon and W. Weaver, *The mathematical theory of communication*. University of Illinois Press, 1949.

[11] E. Simpson, "Measurement of diversity," *Nature*, vol. 163, p. 688, 1949.

[12] K. Loquin and O. Strauss, "Histogram density estimators based upon a fuzzy partition," *Statistics and Probability Letters*, vol. 78, no. 13, pp. 1863–1868, September 2008.