

Tracking Moving Objects With a Catadioptric Sensor Using Particle Filter

Francois Rameau, Désiré Sidibé, Cedric Demonceaux, David Fofi

▶ **To cite this version:**

Francois Rameau, Désiré Sidibé, Cedric Demonceaux, David Fofi. Tracking Moving Objects With a Catadioptric Sensor Using Particle Filter. 11th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras, Nov 2011, Barcelone, Spain. pp.1-6. hal-00627905

HAL Id: hal-00627905

https://u-bourgogne.hal.science/hal-00627905

Submitted on 3 Nov 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Tracking Moving Objects With a Catadioptric Sensor Using Particle Filter

François Rameau Désiré Sidibé Cédric Demonceaux
David Fofi
Université de Bourgogne, Le2i UMR 5158 CNRS, 12 rue de la fonderie, 71200 Le Creusot, France

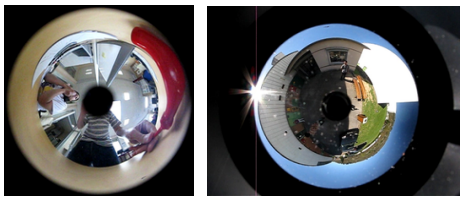
Abstract

Visual tracking in video sequences is a widely developed topic in computer vision applications. However, the emergence of panoramic vision using catadioptric sensors has created the need for new approaches in order to track an object in this type of images. Indeed the non-linear resolution and the geometric distortions due to the insertion of the mirror, make tracking in catadioptric images a very challenging task. This paper describes particle filter for tracking moving object over time using a catadioptric sensor. In this work different problems due to the specificities of the catadioptric systems such as geometry are considered. The obtained results demonstrate an important improvement of the tracking accuracy with our adapted method and a better robustness to clutter background and light changes.

1. Introduction

Tracking a moving target in complex environments is an important task for variety of applications including video surveillance, mobile robots navigation or human-computer interactions. Despite many efforts, tracking is still a challenging problem due to the presence of noise, changes of illumination and occlusions that introduce uncertainty in the estimation of the targets state.

Though many methods have been developed, a very few of them deal with the particular case of catadioptric sensors. Such a sensor provides a wide field of view of the scene. Thus the utilization of catadioptric vision systems allows to enlarge the possibilities offered by a conventional perspective camera by the creation of a panoramic view with a single camera. This kind of system can for instance allows the surveillance of the totality of a room with only one camera while many conventional cameras would be needed for the same task. This is also a really desirable system for robotic applications especially for navigation because the vision of the entire environment of the scene permits a good localization in space and the creation of a navigation map.



(a) (b)

Figure 1. Difficulties due to the geometry of catadioptric sensors. (a) Image distorsion; (b) Illumination changes.

Catadioptric sensors are also useful in various domains like 3D reconstruction or structure from motion.

However, because of the non-linear mapping from 3D world to the 2D image plane, using catadioptric images is not straightforward, as the case of perspective images. In fact, this kind of system leads several inconveniences, notably a strong distortion of the image relative to the shape of the mirror as can be seen in the figure 1(a) where the shape of the red screwdriver is totally modified. A high sensitivity to illumination changes due to the mirror (see figure 1(b)) is also a problem. Therefore, for employing such a sensor in practice, many of the algorithms developed for perspective image systems have to be adapted, and new methods have to be developed. Indeed, most of the conventional visual trackers are not able to properly follow a target through a video sequence taken with a catadioptric camera. Consequently the main purpose of this paper is the adaptation of conventional particle filter to the catadioptric geometry. This is done by adapting the window used to define the object appearance on the unitary sphere.

This paper is organized as follows. First, a brief review of existing tracking methods for catadioptric sensor is given in Section 2. Section 3 describes the neighbourhood adaptation method, and Section 4 is dedicated to the color histogram representation of the target object. The color particle filter adapted to catadioptric geometry is defined in detail in Section 5. Experimental results are shown in Section 6 and some conclusions are derived in last section.

2. Related work

Visual tracking with catadioptric images offers a wide field of possibilities, mainly for surveillance and navigation but this topic has received little attention and the literature is very limited. Among different approaches of visual tracking in catadioptric images, image unwrapping is certainly the most widely used technique. This method needs a Single View Point (SVP) camera in order to obtain a perspective image from the distorted one. The unwrapping process of the image is described in [15]. After the unwrapping, the geometrical distortion is reduced and a conventional tracking algorithm can be use. In [21] and [20] a particle filter is used on the unwrapped image while in [1] a KLT algorithm is applied. The limitation of this approach is the high computation time. Some strategies can be used to reduce the computation time, for example, in [21] a fisheye image is undistorted using a Support Vector Machine in order to accelerate the algorithm.

Tracking based on a priori is an efficient approach often used for robot navigation, visual servoing and 3D tracking. It needs a geometrical a priori knowledge about the object (feature) to track, for instance planes [12] or lines [14]. Plane tracking is mostly based on homography. For example, Mei *et al.* [13] use homography for planar template tracking. This method also requires certain conditions to work properly such as a small and smooth displacement of the plane between two successive images. Line tracking uses another strategy. First a line detection algorithm is apply on the catadioptric images to get lines features [3, 11, 14]. And tracking is performed using a conventional line tracking such as developed by Smith *et al.* [17].

Background subtraction can also be a good solution when the camera is static, because it is not necessary to take into account the distortion of the image. But background subtraction techniques are not robust to illumination changes and catadioptric sensors are very sensitive to such effects. This tracking method is often used for surveillance with catadioptric system for its simplicity and its reliability, for instance Boult *et al.* [6] give a good overview of this approach.

Another approach presented in literature is to adapt conventional visual tracking methods on the projected unitary sphere, so this adaptation can be used for every camera which respect the single view point constraint (fish eye, catadioptric, perspective). The adaptation of conventional techniques gives better results, because the neighbourhood is adapted to the geometrical particularity and to the non linear resolution of the image. The main concept is developed in [7], using the geodesic distance on the sphere. Nevertheless, only in [5] and [18] a color particle filter is adapted with this kind of technique and gives a better result than the conventional method.

The work presented in this paper is similar to the method developed by Bazin *et al.* [5], however we introduce many improvements such as the use of a kernel in order to have a better spatial localization of the object and the utilization of a multi-part histogram which gives a strong robustness to scale variation and clutter background. Furthermore a quantitative evaluation of the algorithm based on different metrics is performed.

3. Neighbourhood adaptation for catadioptric image

The adaptation of the neighbourhood is essential to improve the accuracy of visual tracking because everything in the image will be deformed under the same distortion model. Indeed using an adapted tracking window, the object will be selected and tracked with a highest precision especially when it is subject to a strong distortion. With a perspective image the neighbourhood is mostly defined according to the Euclidean distance and represented as a rectangle centered on the pixel of interest. Obviously, this "conventional" sampling is not appropriated for an omnidirectional image because it considers a constant resolution in the whole image and it does not take into account the distortion of the image.

Every image taken using a camera with a single view point (SVP) can be modelled by a spherical image. This unified projection model was introduced in [8] and [2]. The utilization of the unit sphere is the most suitable approach in order to adapt the neighbourhood for a SVP camera which include distortions. In fact, the projection onto the sphere takes into account the non linear resolution conforming to the shape of the catadioptric mirror (or the specific distortion generated by the fisheye lens) by using the geodesic distance.

Basically the spherical coordinates are defined as:

$$x = (\cos(\phi)\sin(\theta), \sin(\phi)\cos(\theta), \cos(\phi)), \qquad (1)$$

where x is a point lying on the sphere S^2, ϕ is the latitude varying between 0 and π, and θ is the longitude varying between 0 and 2π. The localization of a point with spherical coordinates is defined by two parameters (θ, ϕ) as illustrated in figure 2.

Firstly a center point, X_{sph}, of the neighbourhood has to be defined onto the sphere. X_{sph} has spherical coordinates

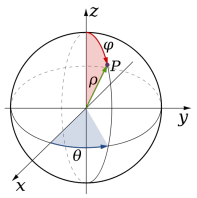
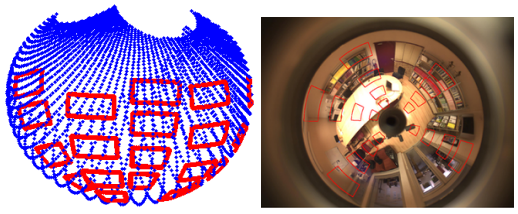


Figure 2. Spherical coordinates system



(a) (b)

Figure 3. Neighbourhood with fixed value $\delta\theta$=±0.2 and $\delta\phi$=±0.1. (a) Neighbourhood on the sphere; (b) Neighbourhood projected on the image plane

(θ,ϕ). Then, the neighbourhood around X_{sph} is defined using a range of variation ($\delta\theta$ and $\delta\phi$) as:

$$N_S(X_{Sph}) = \{X_S = (\theta', \phi')\epsilon S^2 | \, |\theta' - \theta| \leq \delta\theta \text{ and } |\phi' - \phi| \leq \delta\phi\}$$

Once the neighbourhood is obtained on the sphere it is back-projected onto the image plane to get the spatial neighbourhood on the image.

As we can see in figure 3, with constant parameters $\delta\theta$ and $\delta\phi$, the size of the window will vary according to its spatial position.

4. Color histogram representation

The tracking method used in this work is based on color features, more precisely on the color histogram of the object. Histograms are fast to compute and are a robust way to represent the appearance of an object.

4.1. Color distribution computation

Apperance of an object is represented by a color histogram which is computed using the following equation:

$$q_u = C \sum_{i=1}^{n_p} k(\| \, x_i \, \|^2) \delta[b(x_i) - u], \; u = 1 \ldots m, \quad (2)$$

where m is the histogram's size, δ is the Kronecker function, k a kernel profile, b(x_i) the bin number of a pixel x_i, $\| \, x_i \, \|$ the distance between the current point x_i and the center of the window, n_p the number of pixel in the window and C a normalization constant. The color distribution can either represented by a 3D histogram or by a 1D histogram obtained by concatenating the 3 histograms. In this work both methods have been tested, and, the full 3D histogram is selected because it generally gives better results.

4.2. Color space

Color histogram representation is very sensitive to changes in lighting conditions, and a catadioptric system is very subject to this inconvenience (because of reflection on the mirror for instance). Therefore, a suitable color representation must be adopted. In this work, we use the color space called "transformed rgb" since it is defined as robust to light and intensity changes and shifts [19]. The transformation is based on the normalization of each channel independently:

$$\begin{pmatrix} R' \\ G' \\ B' \end{pmatrix} = \begin{pmatrix} \frac{R-\mu_R}{\sigma_R} \\ \frac{G-\mu_G}{\sigma_G} \\ \frac{B-\mu_B}{\sigma_B} \end{pmatrix}, \quad (3)$$

where, σ is the standard deviation of the channel and μ its mean value.

4.3. Kernel Profile

In order to have a better localization of the target, the use of a kernel is highly recommended. In fact, the kernel is used to account for the spatial repartition of colors in the window. In our case, an Epanechnikov profile has been chosen (of course, it is possible to use other kernels with similar characteristics like Quartic kernel or Normal kernel) to compute a non-parametric density estimation of the color distribution. The kernel with Epanechnikov profile is defined by

$$k(x) = \begin{cases} \frac{1}{2} C_n^{-1}(n+2)(1-x) & if \; x \leq 1 \\ 0 & otherwise \end{cases} \quad (4)$$

where C_n is equal to the volume of the unit n-dimensional sphere.

An adaptation of the profile is necessary to deal with the distortion in catadioptric images. Since, we are working on the unitary sphere, the distance from the center of the window is computed using the geodesic distance. Considering two points lying on S^2 (x_{sph} and y_{sph}), the geodesic distance between them is equal to the inverse trigonometric function of the cosine of the scalar product of these two points according to [7]:

$$\forall x_{sph}, y_{sph} \in S^2 \; d(x_{sph}, y_{sph}) = \arccos(x_{sph}.y_{sph}) \quad (5)$$

4.4. Multi-part representation

In a catadioptric video sequence the scale of an object can change "suddenly" (for instance if the object is close to the exterior border of the mirror). Thus a robust method to adapt the size of the tracking windows is necessary. The multi-part target representation allows to drastically increase the accuracy and even permits to keep a good tracking result in difficult conditions (e.g. clutter background). In [10], the authors demonstrate the efficiency of this method for perspective images. Instead of computing only one histogram to represent the object model, we compute seven histograms according to the arrangement defined

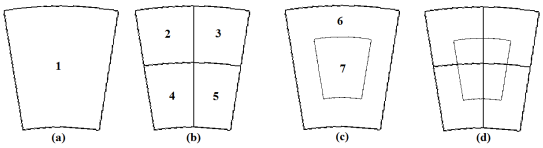


Figure 4. Multi-part representation. (a) whole patch; (b) division in 4 parts; (c) scale sensitive division; (d) final model

in figure 4. The first histogram is computed over the entire image window. Then, we compute 4 histograms by splitting the window. This layout originally permits to have an information about the rotation of the object, so it is basically used for tracking with a rotating tracking window. But different experimentals results show that it also allows to improve the performance of tracking even without a rotating window. The last division of the window into two parts (the inside and outside area of the window) allows a good scale adaptation.

Finally, a similarity measure is computed between two color distributions using the following equation:

$$\rho(q,p) = \frac{\sum_{i=1}^{N} \rho[q^i, p^i]}{N}, \tag{6}$$

where N is the number of histograms (seven in our case) and q^i and p^i are respectively the histogram of the i^{th} subpartition of the window for the model object and the candidate object. And ρ is the Bhattacharyya coefficient calculated as:

$$\rho(q^i, p^i) = \sum_{u=1}^{m} \sqrt{q_u^i . p_u^i}, \tag{7}$$

with m the number of bins of the histogram.

5. Particle filter adapted to the unitary sphere

The conventional color particle filter tracker uses the (x,y) coordinates of pixels to propagate the particles directly in the image plane. In our approach the particles are diffused on the sphere to overcome the problem due to the distortion of the image, and to have a more "natural" behaviour adapted to catadioptric images. Then the state vector is defined by:

$$S = [\theta, \phi, \dot{\theta}, \dot{\phi}, \delta\theta, \delta\phi], \tag{8}$$

where the variables θ and ϕ are the coordinates of the center of the window, and $\delta\theta$ and $\delta\phi$ are the size of the window. The particles are propagated according to the dynamic model:

$$S_t = A.S_{t-1} + W_{t-1}. \tag{9}$$

The matrix A defines the deterministic component of the model, while W is a multivariate Gaussian random noise.

A similarity measure is computed between each particle and the target model one by one. In order to compute the weight necessary for the re-sampling operation, the following equation is used [9]:

$$w^{(i)} = exp((\rho^{(i)} - 1)\beta). \tag{10}$$

Here, $\rho^{(i)}$ is the similarity coefficient of the ith particle and β is a fixed parameter.

The estimated position of the object is obtained as the mean state of the system, given by:

$$E[S] = \sum_{i=1}^{N} w^{(i)} S^{(i)}, \tag{11}$$

where $S^{(i)}$ is the state of the ith particle and w$^{(i)}$ its weight.

6. Experiments and results

In this section, we evaluate the performance of the proposed tracking method using four catadioptric video sequences acquired in different conditions. The conditions include indoor and outdoor scenes, moving objects and persons, moving or fixed camera, difficult illumination changes and occlusion. Table 1 summarizes the characteristics of the sequences employed in the experiments.

The image size of the sequences is 640 x 480 and all algorithms are implemented using Matlab R2009b. We use Scaramuzza [16] 'OCamCalib' toolbox for calibration. We also generate manual ground-truths for each sequence and the adapted particle filter is compared with the conventional one.

6.1. Performances evaluation of a tracking algorithm

Many statistical methods are available to estimate the performance of a tracking system. In this paper three criteria are selected to evaluate the proposed method [4].

Firstly a ground-truth is needed for comparison with the tracking results. The ground-truths are created manually using a window adapted to the geometry of the catadioptric sensors. We use the following measures for evaluation: the overlapping between the ground-truth and the tracking window, the distance between the real center (from the ground-truth) and the center estimated by the tracking algorithm and the percentage of successfully tracked frames in a sequence.

Overlapping estimation. The spatial overlapping is relative to the percentage of common area $A(St, Gt)$ between two bounding boxes (the window resulted from the tracker St and the window from the ground-truth Gt), as shown in figure 5. The spatial overlapping is computed as:

$$A(St, Gt) = \frac{Area(Gt \cap St)}{Area(Gt \cup St)} \tag{12}$$

Table 1. summary of particularities of used sequences

	Localization	Type of camera	Illumination changes	Occlusion	Target	Frames/sec	Number of images
Sequence 1	Outdoor	Non central	No	No	Person	30	780
Sequence 2	Indoor	Non central	reflective surface	No	Object	30	649
Sequence 3	Outdoor	Non central	dazzlement	Partial	Person	30	602
Sequence 4	Indoor	Central	No	Total	Object	15	481



(a) (b)

Figure 5. Spatial overlap estimation. (a) Area(Gt∩St); (b) Area(Gt∪St)

In order to determine the temporal overlapping (the number of frames where the tracking result is considered as good), we define a threshold as the minimum percentage of overlapping:

$$TO(St, Gt) = \begin{cases} 1 \; if \; A(St, Gt) > T_{Ov} \\ 0 \; if \; A(St, Gt) < T_{Ov} \end{cases} \quad (13)$$

So the temporal overlapping $TO(St, Gt)$ is the percentage of frames with spatial overlapping greater than or equal to T_{Ov} (20% in our experiments).

The distance between centers. The distance between the center of the ground-truth and the output of the tracker is a significant information about the accuracy of the tracking algorithm. This distance can be computed with the Euclidean distance on the image plane. In this case the distance must be as low as possible if we wish a accurate result. This metric is less representative of the performance of a tracker than the overlapping measure, but provides a complementary information. For instance a tracking algorithm can have a bad spatial overlapping score but can be properly centered on the target.

6.2. Results

Some tracking results are shown in figure 6. In the images, the blue bounding box corresponds to the ground truth object location, while the red and green bounding boxes correspond to the estimated positions by the proposed method and the conventional particle filter respectively.

The tracking performances are summarized in Table 2. As can be seen, our adapted method outperforms the conventional color based particle filter for all four sequences. For instance, the non-adapted method loses the target object after 44% of sequence 2 which comprises 649 frames, while our method do not lose the object in the entire sequence. This sequence is particularly interesting because it shows a reflective object moving far and close to the mirror, which generates strong scale variations and important distortion of the object. The robustness of the proposed method is also demonstrated with sequence 1, 3 and 4. In fact, even with strong scale variation, occlusion and a strong change in illumination, the algorithm can accurately follow the target thanks to the multi-part histogram representation and a suitable color space.

Note that for each sequence the temporal overlapping is computed based on Equation (13), and when the target is lost (spatial overlapping less than 20%), the mean spatial overlapping and the mean center distance given in Table 2 are computed based only on the correctly tracked frames of the sequence.

In all four sequences, our adapted method gives a better localization of the tracked object as shown by the mean distance between the true center of the object and the center given by the algorithm. This good accuracy is due to the utilization of the multi-part histogram representation. Figure 7 shows the tracking results with a sequence with clutter background. Using the method described by Bazin *et al.* [5], the tracked dwarf is totally lost at the frame 372. However with the multi-part histogram, the target is accurately followed until the end of the sequence (743 frames) with our method.

7. Conclusion

In this paper, a tracking method for catadioptric image sequences is presented. The method is based on adapting a particle filter to account for the specific geometry of catadioptric sensors. These adaptations are managed by the projection of the image onto the unitary sphere. A robust color space is also introduced to deal with the high sensitivity of catadioptric sensors to illumination changes. Experimental results with different sequences in different conditions show the good performance of the approach compared with a conventional color based particle filter for tracking. A direction of future work would be an extension for multiple object tracking, adding shape and texture features.

References

[1] H. Bakstein and A. Leonardis. Catadioptric image-based rendering for mobile robot localization. In *Proceedings of the 6th International Conference on Computer Vision*, pages 1–6, 2007. 2

Table 2. Mean results obtained with PF and our adapted PF for each sequence

Sequence	Method	spatial overlapping	center distance	temporal overlapping
Sequence 1	Conventional method	45%	5.7	100%
	Our method	71%	2.6	100%
Sequence 2	Conventional method	30%	5.5	44%
	Our method	60%	3.8	99.7%
Sequence 3	Conventional method	33%	7.7	88%
	Our method	65%	4.8	99.5%
Sequence 4	Conventional method	45%	6	96.8%
	Our method	66%	4.1	97.3%

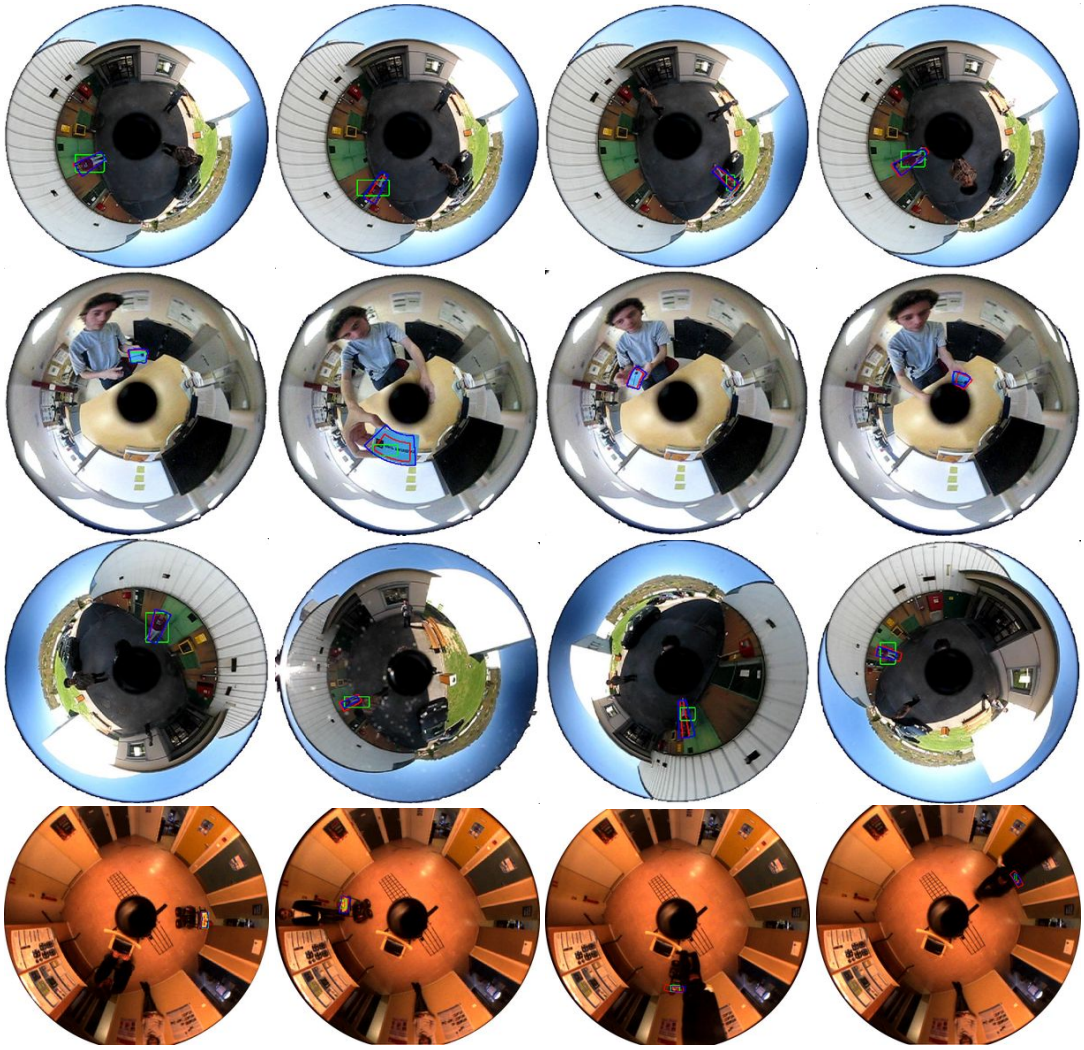


Figure 6. Tracking results. From top to bottom: sequence 1, 2, 3 and 4. Tracking results with the conventional particle filter are depicted in green. Results with the proposed method are shown in red, and the ground truth is in blue.

[2] J.P. Barreto and Helder A. Geometric properties of central catadioptric line images and their application in calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:1327–1333, August 2005. 2

[3] J.P. Barreto and A. Helder. Direct least square fitting of paracatadioptric line images. In *OMNIVIS*, page 78, 2003. 2

[4] F. Bashir and F. Porikli. Performance evaluation of object detection and tracking systems. In *IEEE In-*

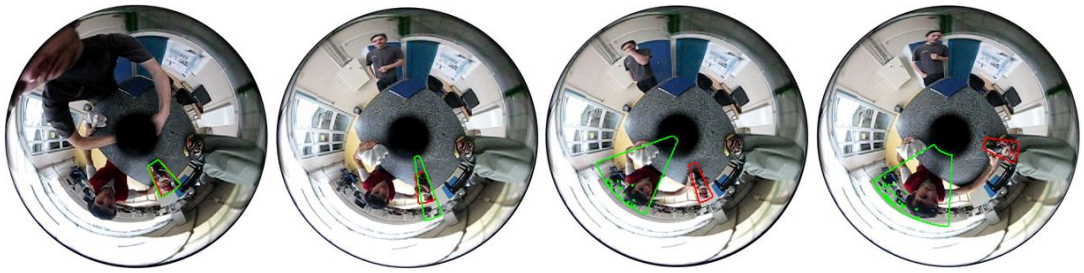


Figure 7. Comparison between a single histogram (green) and a multi-part histogram representation (red)

ternational Workshop on Performance Evaluation of Tracking and Surveillance (PETS), pages 7–14. Citeseer, 2006. 4

[5] J.C. Bazin, K.J. Yoon, , I.S. Kweon, C. Demonceaux, and P. Vasseur. Particle filter approach adapted to catadioptric images for target tracking application. In *20th British Machine Vision Conference, (BMVC'09)*, London, UK, September 2009. 2, 5

[6] T.E. Boult, R. Micheals, X. Gao, P. Lewis, C. Power, W. Yin, and A. Erkan. Frame-rate omnidirectional surveillance and tracking of camouflaged and occluded targets. In *in Second IEEE International Workshop on Visual Surveillance*, pages 48–55, 1999. 2

[7] C. Demonceaux and P. Vasseur. Omnidirectional image processing using geodesic metric. In *Proceedings of the 16th IEEE International Conference on Image processing*, ICIP'09, pages 221–224, Piscataway, NJ, USA, 2009. IEEE Press. 2, 3

[8] C. Geyer and K. Daniilidis. Catadioptric projective geometry. *International Journal of Computer Vision*, 45:223–243, 2002. 2

[9] Aude Jacquot. *Suivi dobjets en imagerie aérienne*. PhD thesis, Institut National Polytechnique de Grenoble, nov 2006. 4

[10] E. Maggio and A. Cavallaro. Multi-part target representation for color tracking. In *Proc. Int. Conf. Image Processing, Italy*, pages 729–732, 2005. 3

[11] E. Marchand and F. Chaumette. Fitting 3d models on central catadioptric images. In *IEEE Int. Conf. on Robotics and Automation, ICRA'07*, pages 52–58, Roma, Italy, 2007. 2

[12] C. Mei, S. Benhimane, E. Malis, and P. Rives. Constrained multiple planar template tracking for central catadioptric cameras. In *British Machine Vision Conference*, pages 4–7, 2006. 2

[13] C. Mei, S. Benhimane, E. Malis, and P. Rives. Homography-based tracking for central catadioptric cameras. In *IN IROS*, 2006. 2

[14] C. Mei and E. Malis. Fast central catadioptric line extraction, estimation, tracking and structure from motion. In *IROS*, pages 4774–4779, 2006. 2

[15] J. Pedro and A. Barreto. General central projection systems, modeling, calibration and visual servoing. Technical report, University of Coimbra,, 2003. 2

[16] D. Scaramuzza. Omnidirectional camera and calibration toolbox for matlab. `http://robotics.ethz.ch/~scaramuzza/Davide_Scaramuzza_files/Research/OcamCalib_Tutorial.htm`, May 2009. 4

[17] P. Smith, T. Drummond, and R. Cipolla. Layered motion segmentation and depth ordering by tracking edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:479–494, April 2004. 2

[18] M. Taiana, J. Gaspar, J. Nascimento, R. Bernardino, and Pedro Lima. 3d tracking by catadioptric vision based on particle filters. In *RoboCup International Symposium*, 2007. 2

[19] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1582–1596, 2010. 3

[20] F. Wallhoff, M. Zobl, G. Rigoll, and I. Potucek. Face tracking in meeting room scenarios using omnidirectional views. In *Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 4 - Volume 04*, ICPR '04, pages 933–936, Washington, DC, USA, 2004. 2

[21] X. Meng Z. Cao and S. Liu. Dynamic omnidirectional vision localization using a beacon tracker based on particle filter. In *Computer Vision, Xiong Zhihui (Ed.), ISBN: 978-953-7619-21-3*, pages 538–554, (2008). 2