

Spherical Object Tracking using Von Mises-Fisher Distribution in Catadioptric System

¹Khalid Anisse, ²Radgui Amina and ¹Rziza Mohamed

¹LRIT, Rabat IT Center - Faculty of Sciences, Mohamed V University, B.P.1014 RP, Rabat, Morocco

²National Institute of Posts and Telecommunications, INPT, Rabat, Morocco

Article history

Received: 16-08-2020

Revised: 28-08-2020

Accepted: 19-09-2020

Corresponding Author:

Khalid Anisse

LRIT, Rabat IT Center -

Faculty of Sciences, Mohamed

V University, B.P.1014 RP,

Rabat, Morocco

Email: khalid.anisse@gmail.com

Abstract: Tracking objects on video sequences is a very challenging task in computer vision applications. However, there is few articles that deal with this topic in catadioptric vision. This paper describes a new approach of omnidirectional images (gray level) processing based on inverse stereographic projection in the image plane. Our work is based on minimizing the distance between two models. The model named von Mises-Fisher distribution have as input the gabor phase and the measure used is Kullback-Leibler Divergence (KLD). In one hand, this model matching respect the deformed geometry of omnidirectional images due to using the spherical neighbourhood. In the other hand, the simulation results show that our approach gives as better performance in terms of overlapping estimation.

Keywords: Object Tracking, Gabor Phase, Omnidirectional Image, Model Matching, VonMises-Fisher Distribution, Kullback-Leibler Divergence

Introduction

In the context of monitoring system, we describe a method for images processing and analysing. Indeed, Object Tracking is an important task in computer vision applications including video surveillance, Radar, robot navigation.

This work defines this process in a omnidirectional images sequence. This is a challenging task because these images contain significant distortions because of mirror geometry. This is why; we must take them into account during processing.

This work is focused on two main goals: The one including detection and matching features. After that, the adapted process according to spherical coordinates (θ, ϕ) over the image plane in order to make possible the non-rigid objects tracking.

The omnidirectional image provided by a camera with a Single View Point [(SVP) (Geyer and Daniilidis, 2000; Barreto and Araujo, 2001; Chiuso and Picci, 1998)] is modelled by a spherical image (Fig. 1). This model called unified projection model and was introduced in (Geyer and Daniilidis, 2000) and used in (Baker and Nayar, 1998). Also (Mei and Rives, 2007) make using this model for camera calibration.

Following the catadioptric mirror shape used (parabolic in our case), the projection process onto the sphere and the catadioptric plan is shown in Fig. 2.

The 3D point P (Equation 1) is projected onto the sphere in P_s .

This point then changed to new reference frame centred in P_m using ε (camera parameter). This value ranges between 0 (planar mirror) and 1 (parabolic mirror). Then, the point $P_m: (x_s; y_s; z_s-\varepsilon)$ is projected from the coordinate system to the point $m = (u; v; 1)$ onto the normalized plane. The image point P_i is obtained by $p_i = \kappa * m$. Where κ is a 3×3 matrix containing the camera intrinsic parameters. The matrix κ and ε values are obtained by the calibration procedure. For our case (parabolic mirror), $\varepsilon = 1$ and $\kappa = f(\alpha_u; \alpha_v; u_0; v_0)$. In this study, we assume that our Catadioptric camera is calibrated.

Spherical coordinates of P_s are defined in Equation (1):

$$\begin{cases} X_s = \cos(\varphi)\sin(\theta) \\ Y_s = \sin(\varphi)\sin(\theta) \\ Z_s = \cos(\theta) \end{cases} \quad (1)$$

The stereographic projection of P_s from the sphere to the catadioptric plane can be expressed on Cartesian coordinates in Equation (2):

$$\begin{cases} u = \frac{X_s}{1-Z_s} \\ v = \frac{Y_s}{1-Z_s} \end{cases} \quad (2)$$

discussion of the synthetic data experiments, while Section V concludes the paper.

Related Works

Object tracking in a complex environment needs a powerful algorithm. Consequently, many methods have been developed in literature and can be divided into tree groups: Tracking based on kernel, based on point and based on edges. These methods have been successfully employed in various application domains. They cannot be directly applied to images acquired by catadioptric cameras like the work's (Kumar, 2015). When the author attempts to estimate self-position of intelligent vehicles by fusion several techniques. Also in (Aggarwal, 2015), the same author use GPS information for estimation self-position of mobile robots. Mei *et al.* (2006), the author presents a homography-based approach for tracking multiple planar templates. First, the adaptation of conventional particle filter to the catadioptric geometry was purposed in (Lucas and Kanade, 1981). A few methods have been developed in the literature in this context. Consequently, the most adapted methods are based on statistic estimation. We find the adaptation of conventional particle filter to the catadioptric geometry that was purposed in (Ikoma, 2008). This is done by adapting the window used to define the object appearance on the unitary sphere. Also, authors in (Hurych *et al.*, 2011) propose a new method to display tracking result from weighted particles obtained from the estimation process by Sequential Monte Carlo (SMC) implementation of Probability Hypothesis Density (PHD) filter. Also Radgui *et al.* (2011) using spherical adaptation of the KLD method makes flow estimation in omnidirectional images.

In this study, we use methods based on kernel. Specially, based probability distributions whither give us information about area of block without knowing size or shape. In addition, in context of computer vision, there are many sets of probability distributions on the sphere in several fields.

In signal processing, spherical distributions have been used for multiple speaker tracking (Traa and Smaragdīs, 2014) and speaker clustering (Tang *et al.*, 2009). Most applications include quaternion-based orientation estimation (Kurz *et al.*, 2014; Glover and Kaelbling, 2014), protein structure modelling in biology (Kent and Hamelryck, 2005), machine learning (Bijral *et al.*, 2007; Banerjee *et al.*, 2005) and neuroscience (Leong and Carlile, 1998).

Early uses of spherical distributions was found in geology field (Bingham, 1974).

In recent years, these approaches have found their way into many domains like estimation and filtering not just to descriptive statistics. For example, the task

of tracking objects using omnidirectional cameras (Marković *et al.*, 2014a) can naturally be mapped to a spherical process problem by using the probability estimation. Thus, it even becomes possible to consider multiple objects tracking onto the unit sphere (Marković *et al.*, 2014b).

In the same context, we consider the Gabor phase information as estimator input. This is because it can take very different values even if it is sampled at image locations only a few pixels apart (Struc and Pavesic, 2010).

General Background

Wavelet Gabor phase

Gabor kernels or also called Gabor wavelets or filters have known to be a robust tool for feature detection. They represent an optimal localization in the spatial and frequency domains. They extract, multi-scale invariant, spatially local features of a suitable frequency band. In the spatial domain (Struc and Pavesic, 2010), the family of Gabor filters is defined as follows:

$$\Psi_{(u,v)}(x, y) = \frac{f_u^2}{\pi\alpha\beta} e^{(-f^2[\alpha^{-2}x^2 + \beta^{-2}y^2])} e^{(j2\pi f_u x')} \quad (4)$$

where, $x' = x\cos(\theta_v) + y\sin(\theta_v)$, and $y' = -\sin(\theta_v) + y\cos(\theta_v)$ denote axes that are rotated by $\theta_v = v * \frac{\pi}{8}$. f_u is the

frequency defined by $f_u = \frac{f_{\max}}{2^2}$. Each filter function

represents a Gaussian kernel function has centre frequency and orientation defined respectively by f_u and θ_v . α and β determine the ratio between center frequency and size of Gaussian envelope.

Parameters θ_v and f_u define the passband of the Gabor in frequency domain and the spectral characteristics of the Gabor kernel. The Gabor filter kernels are given by:

$$\begin{aligned} \Psi_+(x, y, f_u, \theta_v) &= g(x, y)\cos(2\pi f_u x'), \\ \Psi_-(x, y, f_u, \theta_v) &= g(x, y)\sin(2\pi f_u x'), \end{aligned} \quad (5)$$

$$g(x, y) = e^{\left(\frac{-\lambda^2 f_u (x^2 + y^2)}{4\pi}\right)},$$

where, λ is a constant.

In Equation (6), $G_{(u,v)}(x,y)$ represents the complex convolution which can be decomposed into real $R_{(u,v)}(x,y)$ and imaginary parts $I_{(u,v)}(x,y)$ as follows:

$$G_{(u,v)}(x, y) = I(x, y) * \Psi_{(u,v)}(x, y) \quad (6)$$

where, $I(x, y)$ is the input image. the local magnitude M and the phase Φ of the output are given by:

$$M(x, y, f_u, \theta_v) = \sqrt{R_{(u,v)}(x, y)^2 + I_{(u,v)}(x, y)^2},$$

$$\Phi(x, y, f_u, \theta_v) = \arctan\left(\frac{I_{(u,v)}(x, y)}{R_{(u,v)}(x, y)}\right). \quad (7)$$

The gabor phase is represented in Fig. 3.

We apply the Gabor filter over each block image. This block in perspective image corresponds to the different shape block in omnidirectional image. In our algorithm, we defined the neighborhood by size block given by $\delta\theta$ and $\delta\phi$ as follows:

$$N_s = \left(\left| \delta\theta \right| \leq \frac{1}{N}, 2\pi - \frac{1}{M} \leq \delta\phi \leq \frac{1}{M} \right) \quad (8)$$

N et M are the neighborhood orders. We extract the histogram phase of each block and it will be the input of simulator in order to create the model.

Von Mises-Fisher Distribution

The probability distribution $f(\mu, \kappa)$ onto the unit sphere defined by density function (pdf) firstly in (Mardia and Jupp, 1999) as Directional distribution:

$$f(x; \mu; \kappa) = C_d(\kappa) e^{\kappa \mu x}; x \in S^{d-1} \quad (9)$$

Is called von Mises-Fisher (vMF) distribution with parameters $\kappa \geq 0$ and $\mu \in S^{d-1}$ denote respectively the concentration and the mean direction:

$$C_d(\kappa) = \frac{\kappa^{\frac{d}{2}-1}}{2\pi^{\frac{d}{2}} I_{\frac{d}{2}-1}(\kappa)} \quad (10)$$

$C_d(\kappa)$ is the normalization constant, where I_p denotes the modified Bessel function (Wood, 1994) at order $p = \frac{d}{2} - 1$. The density in Equation (Ikoma, 2008) invariant by rotation around the mean direction. Also it is equivalent to the multivariate Gaussian distribution. This distribution is characterized by the maximum entropy Principe in order to maximize the Boltzmann–Shannon entropy according to the directional mean. Von Mises-Fisher distributions presents an exponential family (Rényi, 1961) defined by parameters $\theta = \kappa\mu \in R^d$ and the log-normalizing function given by:

$$F_d(\theta) = -\log C_d(\|\theta\|) \quad (11)$$

The vMF distribution is generally determined by the directional (angular) mean:

$$E[x] = \int_{S^{d-1}} x f(x; \mu; \kappa) dx = \Delta F_d(\theta) = A_d(\kappa)\mu, \quad (12)$$

where, A_d is the ratio of the following Bessel functions:

$$A_d = \frac{I_{d/2}(\kappa)}{I_{d/2-1}(\kappa)} \quad (13)$$

In order to simulate the VMF distribution, we obtain a random vector from the VMF knowing κ and μ . To get the desired direction, we only need to apply the suitable rotation.

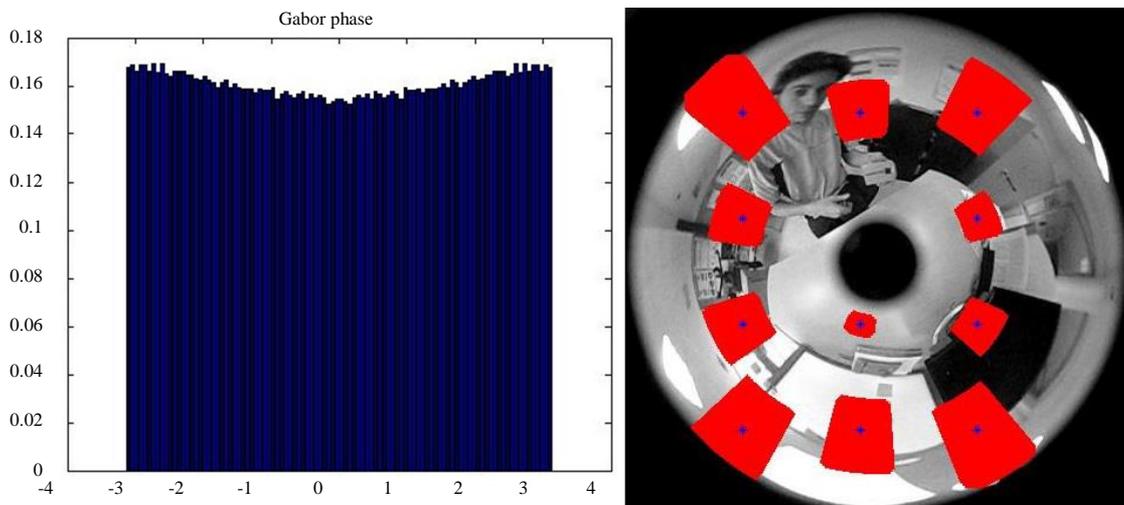


Fig. 3: Phase histogram at a particular wavelet for spherical block

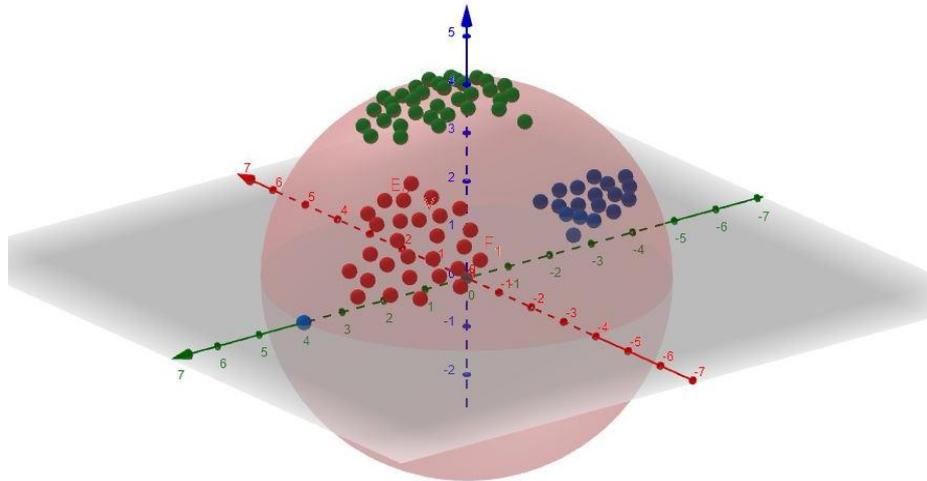


Fig. 4: Samples on the unit sphere of the von Mises-Fisher distribution with different mean directions and concentration parameters of 50 (red), 150 (green), 500 (blue)

Due to the present application of objects tracking, we are particularly interested in case of 3D domain (sphere) and vMF distributions on the unit sphere, where the above expressions simplify to $C_3(\kappa) = \frac{\kappa}{4\pi \sinh(\kappa)}$ and

$$A_3(\kappa) = \frac{1}{\tanh(\kappa)} - \frac{1}{\kappa}.$$

An example of a von Mises-Fisher distribution on the unit sphere with different mean directions and concentration parameters is shown in Fig. 4.

KL-Divergence for the VMF Distribution

In this study, we use a method that requires to compute the distance between the current model and others in their vicinity. We propose the following procedure based on statistical distances (Diethe, 2015).

Given that, we derive the Renyi α -divergence (Rényi, 1961) for the vMF distribution as a distance that will be used to determine the measurement.

We chose this divergence because is a statistical class of generalized distances including the Kullback-Leibler (KL) distance and the Bhattacharyya distance.

The Renyi α -divergence is given by the following expression:

$$D_R^{(\alpha)}(p, q) = \frac{1}{\alpha - 1} \log \int_{\mathcal{W}} p(x)^\alpha q(x)^{1-\alpha} dx. \quad (14)$$

And it is characterised by the real parameter α when in the limit $\alpha \rightarrow 1$. The KLD distance expression becomes as follows:

$$D_{KL}(p, q) = \int_{\mathcal{W}} \log \left(\frac{p(x)}{q(x)} \right) dx. \quad (15)$$

$$D_R^\alpha(p, q) = \frac{\alpha}{1 - \alpha} \log \left(\frac{4\pi \sinh \kappa_p}{\kappa_q} \right) + \log \left(\frac{4\pi \sinh \kappa_q}{\kappa_p} \right) - \left(\frac{1}{1 - \alpha} \right) \log \left(\frac{4\pi \sinh(\kappa_{pq})}{\kappa_{pq}} \right) \quad (16)$$

where, $\kappa_{pq} = \|\alpha \mu_p \kappa_p + (1 - \alpha) \mu_q \kappa_q\|$. For $\alpha = 1$, the KL distance will be:

$$D_{KL}(p, q) = \log \left(\frac{\kappa_p \sinh(\kappa_q)}{\kappa_q \sinh(\kappa_p)} \right) - \left(\frac{1}{\tanh(\kappa_p)} - \frac{1}{\kappa_p} \right) \mu_p^T (\kappa_q \mu_q - \kappa_p \mu_p) \quad (17)$$

Algorithm

The tracking algorithm of our approach is presented beside.

Evaluation Methodology

In this study, we use the Spatial Overlapping estimation. Overlapping estimation. This metric is the bounding box overlap as shown in Fig. 5. Given the tracked bounding box r_t and the ground truth box r_a ,

the overlap measure is defined as $S = \frac{r_t \cap r_a}{r_t \cup r_a}$ where \cap

and \cup represent the intersection and union of two regions, respectively and $|*|$ denotes the number of pixels in the region. To measure the performance on a sequence of frames, we count the number of successful frames whose overlap S is larger than the given threshold equal to 0.2.

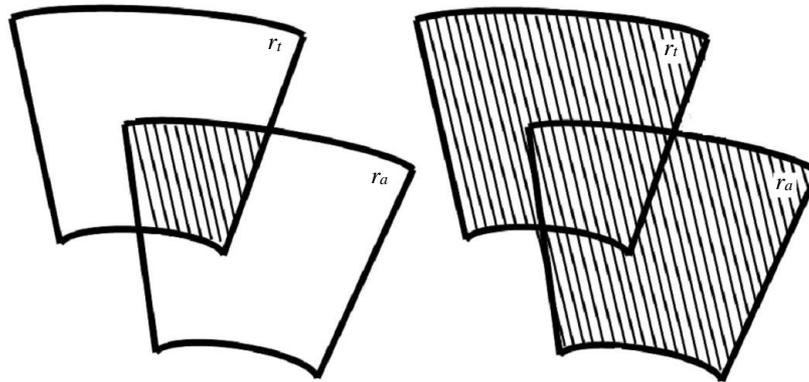


Fig. 5: Spatial overlap estimation. (left.) $r_i \cap r_a$; (right) $r_i \cup r_a$

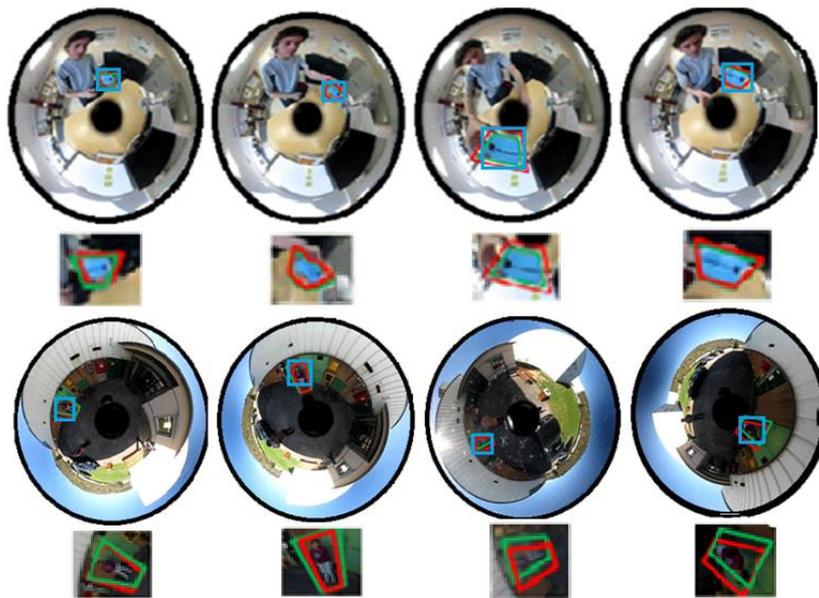


Fig. 6: Tracking results. Tracking results with the conventional Template matching are depicted in blue. Results with the proposed method are shown in red and the ground truth is in green

Table 1: Adapted Model Block Matching

Seq.	Average	SMM	Adapted MS (2)
Seq.1	Spatial	70,4%	65%
Seq.2	Ovelapping	65,3%	60%

In Table 1, we compared the adapted Mean Shift algorithm (Rameau *et al.*, 2011) with our approach based on adapted model matching in omnidirectional images in term of the Spatial Overlapping.

The results in Fig. 6 show that the directional model matching in spherical domain gives better performances.

Algorithm 1 Spherical model matching

Require: N : Frame number; I_1, I_2 : current and next images; $\delta\theta = \frac{\pi}{40}$; $\delta\phi = \frac{\pi}{15}$

for $n = 1: N$ **do**

Ensure: Initialization: First image block

if $i = 1$ **then**

Select the object block $g_i(\delta\theta, \delta\phi)$;

Compute Gabor phase P_{g_i}

Estimate VMF parameters $[\kappa_{g_i}, \mu_{g_i}] = VMF(P_{g_i})$

else

Require: g_i, f_i : image blocks in I_1 and I_2

Ensure: research zone:

The Search zone SZ size is equal to $(M\delta\theta, W\delta\phi)$

$C_i = (p_i, q_i)$ = Center position of g_i

$C_s = (p, q)$ = Center position of Search zone

$$p \in \left[p_i - \frac{M\delta\theta}{2}, p_i + \frac{M\delta\theta}{2} \right]$$

$$q \in \left[q_i - \frac{M \delta \phi}{2}, p_i + \frac{M \delta \phi}{2} \right]$$

```
for  $i = 1: size(SZ, 1)$  do
  for  $j = 1: size(SZ, 2)$  do
    Compute Gabor phase  $P_{fi}$ 
     $[\kappa_{fi}, \mu_{fi}] = VMF(P_{fi})$ 
    Compute  $kld_{fg}(\kappa_{gi}, \mu_{gi}, \kappa_{fi}, \mu_{fi})$ 
    Where  $f_i(p_i, q_i)$  blocks in the vicinity of  $g_i$ .
    Storage all  $kld_{fg}$ 
  end for
end for
 $[I, j, kld] = argmin_{\omega} KLD$ 
Ensure: Block move:
 $w_i = i, t_i = j$ 
 $(w_i, t_i) = \text{position of } f_i$ 
 $M_i = [p_i + w_i, q_i + t_i]$ 
end if
end for
```

Conclusion

In this study, we present a tracking method in catadioptric system. It is based on an adapted Model Matching. In one hand, we the adapted process done by image projection onto the sphere. In the other hand, the model was defined by Von mises Fisher Distribution parameters. In tracking, we optimize the distance between models for finding the moving object block. Experimental results taking into account the change size and shape of block show the good performance of our approach. A direction of future work would be an extension for multiple objects tracking with real time processing.

Acknowledgment

Omnidirectional image sequence in this study was provided by Cedric Demonceaux, L2i Laboratory and Bourgogne Franche-Comte University. <http://le2i.cnrs.fr/-Cedric-Demonceaux,350->

Author's Contributions

Anisse Khalid: Participated in all experiments and tests and data-analysis. Contributed to the writing of the manuscript.

Amina Radgui: Conceptualization, Supervision and Investigation.

Mohamed Rziza: Review, editing and Formal Analysis.

Ethics

This article is original and contains unpublished work and results. The corresponding author confirms that all of the other authors have read and approved the manuscript and has no ethical issue involved.

References

- Aggarwal, A. K. (2015). Machine Vision Based Self Position Estimation of Mobile Robots. *International Journal of Electronics and Communication Engineering & Technology*, 6(10).
- Baker, S., & Nayar, S. K. (1998, January). A theory of catadioptric image formation. In *Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271)* (pp. 35-42). IEEE.
- Banerjee, A., Dhillon, I. S., Ghosh, J., & Sra, S. (2005). Clustering on the unit hypersphere using von Mises-Fisher distributions. *Journal of Machine Learning Research*, 6(Sep), 1345-1382.
- Barreto, J. P., & Araujo, H. (2001, December). Issues on the geometry of central catadioptric image formation. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001 (Vol. 2, pp. II-II)*. IEEE.
- Bijral, A. S., Breitenbach, M., & Grudic, G. (2007, March). Mixture of Watson distributions: a generative model for hyperspherical embeddings. In *Artificial Intelligence and Statistics* (pp. 35-42).
- Bingham, C. (1974). An antipodally symmetric distribution on the sphere. *The Annals of Statistics*, 1201-1225.
- Chiuso, A., & Picci, G. (1998). Visual tracking of points as estimation on the unit sphere. In *The confluence of vision and control* (pp. 90-105). Springer, London.
- Diethe, T. (2015). A Note on the Kullback-Leibler Divergence for the von Mises-Fisher distribution. *arXiv preprint arXiv:1502.07104*.
- Geyer, C., & Daniilidis, K. (2000, June). A unifying theory for central panoramic systems and practical implications. In *European conference on computer vision* (pp. 445-461). Springer, Berlin, Heidelberg.
- Glover, J., & Kaelbling, L. P. (2014, May). Tracking the spin on a ping pong ball with the quaternion Bingham filter. In *2014 IEEE international conference on robotics and automation (ICRA)* (pp. 4133-4140). IEEE.
- Hurych, D., Zimmermann, K., & Svoboda, T. (2011, March). Fast Learnable Object Tracking and Detection in High-resolution Omnidirectional Images. In *VISAPP* (pp. 521-530).
- Ikoma, N. (2008). Tracking of multiple moving objects in dynamic image of OMNI-directional camera using PHD filter. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 12(1), 16-25.
- Kent, J. T., & Hamelryck, T. (2005). Using the Fisher-Bingham distribution in stochastic models for protein structure. *Quantitative Biology, Shape Analysis and Wavelets*, 24(1), 57-60.

- Kumar, A. (2015). Image based methods for navigation of intelligent vehicles. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, 4, 8210-8215.
- Kurz, G., Gilitschenski, I., Julier, S., & Hanebeck, U. D. (2014). Recursive Bingham filter for directional estimation involving 180 degree symmetry. *Journal of Advances in Information Fusion*, 9(2), 90-105.
- Leong, P., & Carlile, S. (1998). Methods for spherical data analysis and visualization. *Journal of neuroscience methods*, 80(2), 191-200.
- Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision.
- Mardia, K. V., & Jupp, P. E. (1999). *Directional statistics*. John Wiley & Sons.
- Marković, I., Chaumette, F., & Petrović, I. (2014a, May). Moving object detection, tracking and following using an omnidirectional camera on a mobile robot. In *2014 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 5630-5635). IEEE.
- Marković, I., Bukal, M., Česić, J., & Petrović, I. (2014b, July). Direction-only tracking of moving objects on the unit sphere via probabilistic data association. In *17th International Conference on Information Fusion (FUSION)* (pp. 1-7). IEEE.
- Mei, C., & Rives, P. (2007, April). Single view point omnidirectional camera calibration from planar grids. In *Proceedings 2007 IEEE International Conference on Robotics and Automation* (pp. 3945-3950). IEEE.
- Mei, C., Benhimane, S., Malis, E., & Rives, P. (2006, September). Constrained Multiple Planar Template Tracking for Central Catadioptric Cameras. In *BMVC* (pp. 619-628).
- Radgui, A., Demonceaux, C., Mouaddib, E., Rziza, M., & Aboutajdine, D. (2011). Optical flow estimation from multichannel spherical image decomposition. *Computer Vision and Image Understanding*, 115(9), 1263-1272.
- Rameau, F., Sidibé, D. D., Demonceaux, C., & Fofi, D. (2011). Visual tracking with omnidirectional cameras: an efficient approach. *Electronics letters*, 47(21), 1183-1184.
- Rényi, A. (1961). On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. The Regents of the University of California.
- Struc, V., & Pavesic, N. (2010). From Gabor magnitude to Gabor phase features: tackling the problem of face recognition under severe illumination changes. INTECH Open Access Publisher.
- Tang, H., Chu, S. M., & Huang, T. S. (2009, April). Generative model-based speaker clustering via mixture of von Mises-Fisher distributions. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 4101-4104). IEEE.
- Traa, J., & Smaragdis, P. (2014, September). Multiple speaker tracking with the factorial von Mises-Fisher filter. In *2014 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)* (pp. 1-6). IEEE.
- Wood, A. T. (1994). Simulation of the von Mises Fisher distribution. *Communications in statistics-simulation and computation*, 23(1), 157-164.