# Tracking of multiple moving objects on the unit sphere using a multiple-camera system on a mobile robot

Josip Ćesić, Ivan Marković, and Ivan Petrović

University of Zagreb, Faculty of Electrical Engineering and Computing
Departement of Control and Computer Engineering
Unska 3, Zagreb, Croatia

**Abstract.** Detection and tracking of moving objects with camera systems mounted on a mobile robot presents a formidable problem since the ego-motion of the robot and the moving objects jointly form a challengingly discernible motion in the image. In this paper we are concerned with multiple-camera systems, namely the Ladybug®2 camera, whose perspective images were used to detect motion and subsequently perform the tracking of multiple objects on the sphere. This enabled us to account for the continuity of the scene which is achieved by the sensor in an image stitching process on the sphere. The objects are tracked on the sphere with a Bayesian filter based on the von Mises-Fisher distribution and the data association is achieved by the global nearest neighbor method, for which the distance matrix is constructed by deriving the Rényi $\alpha$-divergence for the von Mises-Fisher distribution. The prospects of the proposed method are tested on a synthetic and real-world data experiments.

## 1 Introduction

Detection and tracking of moving objects (DATMO) in the surroundings of a mobile platform or a vehicle is a fundamental step in many different applications. Whether this information is being used to navigate in an environment populated by moving objects, or to discard regions which are designated as belonging to moving objects, the DATMO begins with the detection part which entails processing the information on the raw data level and is later followed by a tracking scheme which assumes an estimation process. In that sense the detection methods depend strongly on the nature of the sensor and the phenomena it senses, while the tracking part can depend on the space in which the measurements are taken. Namely, in the present paper we utilize a multiple-camera system which forms an omnidirectional image by stitching a series of perspective images on a unit sphere which is the space in which our measurements reside.

When a camera is placed on a mobile robot the task of detecting moving objects becomes increasingly more complicated since the total motion in the image is a combination of the mobile robot ego-motion and the motion of moving objects. In [1] moving object detection with a single perspective camera was achieved by calculating the optical flow and optimizing the bilinear transformation to warp the image between the consecutive frames, after which the images are differentiated and motion is detected. Then the particle filter was used to track the moving objects in the image and a laser range finder is used to infer about the location in 3D. In [2] the detection was based on monocular scene reconstruction and affine transformation of a triangle mesh in order to perform the image warping. The tracking of the moving object and the scene reconstruction was performed using the extended Kalman filter. Performing motion detection with the extension of the classical structure-from-motion (SfM) to dynamic scenes with multiple rigidly moving objects, called multibody SfM, can be found in [3, 4]. In [5] people were detected

by using histogram of oriented gradients in a stitched and unwrapped panoramic image using the Ladybug®2 multiple-camera system on a mobile robot.

When tracking multiple moving objects the problem of data association plays one of the crucial roles. To solve this problem the methods that can be used are the global nearest neighbor (GNN) which attempts to find the single most likely data association hypothesis at each scan [6], joint probabilistic data association filter (JPDA) filter where multiple hypotheses are formed after each scan and then these hypotheses are combined before proceeding further with the next scan [7], the multiple hypothesis filter (MHT) where multiple data association techniques are formed and propagated. Also, another method for tracking of multiple objects is the probability hypothesis density (PHD) filter [8] which does not solve the data association problem by itself but a solution has been presented in [9] for the Gaussian mixture PHD [10]. Particular implementation of the aforementioned tracking methods depends if the Kalman filter, Gaussian mixture or sequential Monte Carlo methods are used. However, to the best of the authors' knowledge, these methods have not been previously applied in a multiple target tracking scenario with a distribution on the unit sphere—the von Mises-Fisher distribution.

In this paper we propose a novel method for tracking multiple moving objects on the sphere using the GNN framework which is based on the Bayesian tracking with a von Mises-Fisher distribution. To solve the task, first we derive means for validation gating on the sphere based on the measurement likelihood and the predicted state in order to discard unlikely measurements from the association procedure. To calculate the most likely data association hypotheses we propose the Rényi $\alpha$-divergence as a distance measure between the von Mises-Fisher distributions and expressions thereof are derived in the paper. The Rényi $\alpha$-divergence is a class of generalized distances and includes some well known distances such as the Kullback-Leiber and Bhattacharyya distance. The experimental results of the tracking are first presented on a synthetic data example and subsequently on experiments obtained by the Ladybug®2 multiple-camera system. The system at hand is omnidirectional, but since it forms the omnidirectional image by stitching images of five perspective camera (the top one was not regarded) the detection of moving objects was performed by a method developed for perspective cameras [1]. Subsequent moving object clustering and measurement generation for the tracking process was then performed on the sphere.

The paper is organized as follows. Section 2 presents utilized algorithm for motion detection based on image warping. Section 3 describes the proposed multiple object tracking method based on the von Mises-Fisher distribution, GNN, and Rény $\alpha$-divergence. Section 4 presents experimental results, while Section 5 concludes the paper.

## 2 Detection of Moving Objects

For the detection task we have used a spherical digital video camera system consisting of 6 monocular cameras (namely Ladybug®2) placed on a mobile platform. Since the system consists of six perpective cameras, they can be stitched together and thus form a spherical image. Many sensors, including monocular cameras, are in principle bearing-only sensing systems, whereas depth can be estimated only up to scale. Here arises the need for tracking of moving objects on the sphere. In the vein of previously presented problem, the position of an object detected in any of six images of the camera system is projected onto sphere, which served as an input for the tracking task. The detection of moving objects becomes quite a difficult problem ones the robot starts moving, since the ego-motion induces an inherent optical flow. Therefore, to ensure the efficiency of the algorithm while the platform moves, it is necessary to compensate ego-motion [1], [2].

The algorithm for ego-motion compensation provides a tranformation between two consequtive frames where the transformation may be estimated either directly or indirectly. Former relies on various localization systems (e.g. IMU, GPS, odometry), while latter method estimates the transformation through image processing, avoiding any additional sensors. Such indirect method is based on salient feature set tracked through consequtive images. For this purpose we have used Lucas-Kanade algorithm for sparse optical flow calculation presented in [11], [12], [13]. Once the optical flow is established, it is necessary to determine the transformation parameters for the compensation. Here we used a nonlinear model, which is able to handle both translational and rotational shifts, in particular a bilinear model given as follows

$$\begin{bmatrix} f_x^t \\ f_y^t \end{bmatrix} = \begin{bmatrix} a_0 \ f_x^{t-1} + a_1 \ f_y^{t-1} + a_2 + a_3 \ f_x^{t-1} \ f_y^{t-1} \\ a_4 \ f_x^{t-1} + a_5 \ f_y^{t-1} + a_6 + a_7 \ f_x^{t-1} \ f_y^{t-1} \end{bmatrix}, \tag{1}$$

where $a_i$, $i \in \{0, .., 7\}$ are transformation parameters, while $f_x^t$ and $f_y^t$ correspond to image coordinates of a feature at time $t$. In order to determine the most suitable transformation parameters, the least square error is minimized

$$J = \frac{1}{2} \sum_{i=1}^{N} ||f_i^t - \mathcal{W}(f_i^{t-1})||, \tag{2}$$

where $\mathcal{W}$ represents a transformation rule and $N$ corresponds to the size of the feature set. Since some features correspond to dynamic objects, the optical flow is caused by both ego-motion and moving objects. Therefore, in the vein of [1], the following procedure is used

- estimate the initial transformation $\mathcal{W}_0$,
- divide the feature set into two subsets by:

$$\begin{cases} f_i \in F_{in} & \text{if } ||f_i^t - \mathcal{W}(f_i^{t-1})|| < \epsilon \\ f_i \in F_{out} & \text{otherwise ,} \end{cases} \tag{3}$$

- re-estimate the transformation $\mathcal{W}$ using only $F_{in}$ subset.

This approach relies on the assumption that the ego-motion causes dominant optical flow. Once having the tranformation estimated, the previous image can be warped into the current moment in time

$$I_{t-1 \to t} = \mathcal{W}_{t-1 \to t}(I_{t-1}) \approx I_t. \tag{4}$$

Warped image $I_{t-1 \to t}$ corresponds approximatelly to image $I_{t-1}$, taken from the same position as $I_t$. Then the dynamic parts of image at time $t$ are detected after applying image differencing between images $I_{t-1 \to t}$ and $I_t$. Such image of differences is then divided into regions of size $16 \times 16$ pixels, whereas the mean level of difference is determined for each region. This mean can be considered as the amount of dynamics within the region. After applying a threshold, every region is declared as being either static or dynamic. Dynamic regions are then clustered, based on the Euclidean distances within the image.

Since every two contiguous cameras have an overlapping field of view, some objects might be recognized in two images. Therefore, at each time step, every detection is projected on the unit sphere, whence some detections are clustered together due to previous argumentation. An example of the detection process is shown in Fig. 1.

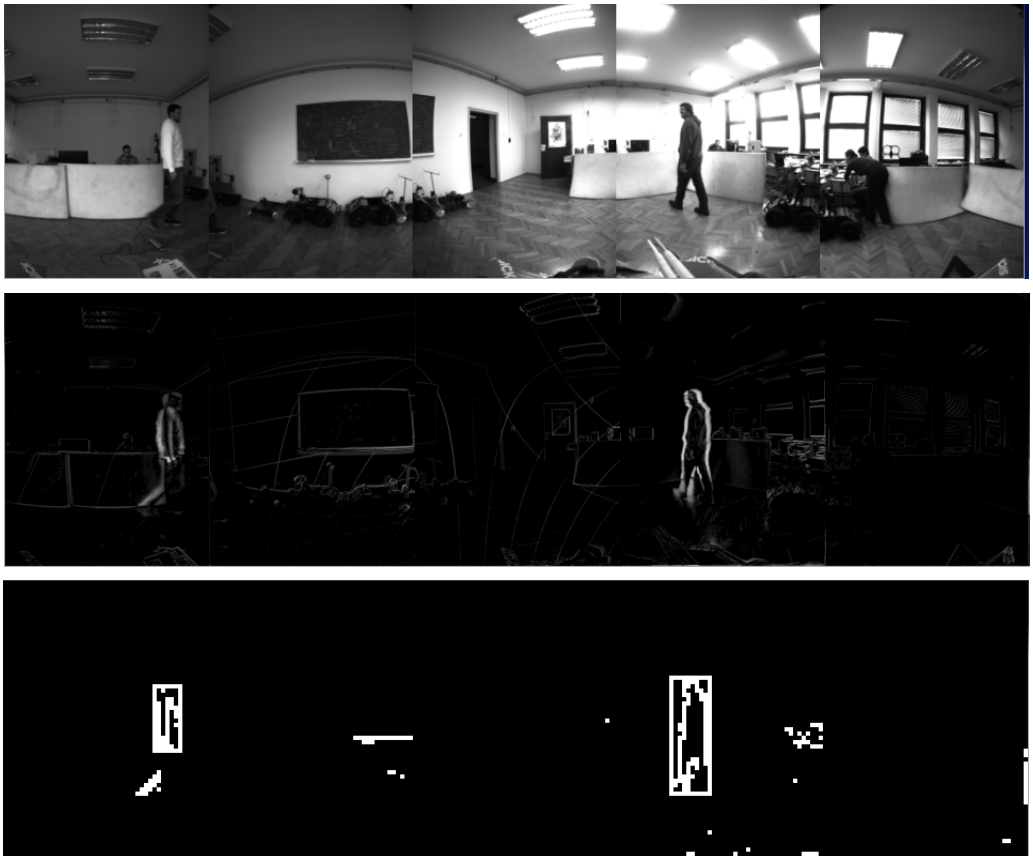


Fig. 1: A snapshot of the detection process. The uppper most image shows the concatenated images from the cameras. The middle image shows the result after subtracting the image $I_t$ captured at time $t$ from the warped image $I_{t-1\to t}$ captured at time $t-1$. The bottom image shows the detected motion inside of a white bounding box.

## 3 Tracking on the unit sphere

### 3.1 The von Mises-Fisher distribution

The von Mises-Fisher distribution serves as an all-purpose probability model for directions in space and directional measurement errors [14]. When considering directions in $p$ dimensions, i.e. unit vectors in $p$ dimensional Euclidean space $\mathbb{R}^p$, one can represent them as points on $S^{p-1}$, i.e. the $p-1$ dimensional sphere with unit radius and center at the origin. In other words, a $p$-sphere is defined as a set of points in $(p+1)$ dimensional Euclidean space, hence a 1-sphere is the circle and the 2-sphere is the surface of a ball in three-dimensional space. A three-dimensional unit random vector $\boldsymbol{x}$ is said to have a von Mises-Fisher distribution $\mathfrak{f}(\boldsymbol{\mu}, \kappa)$ if its probability density function (pdf) is of the following form [15]

$$p(\boldsymbol{x}; \boldsymbol{\mu}, \kappa) = \frac{\kappa}{4\pi \sinh \kappa} \exp\left(\kappa \boldsymbol{\mu}^{\mathrm{T}} \boldsymbol{x}\right), \quad \boldsymbol{x} \in S^2, \tag{5}$$

where $\boldsymbol{\mu}$, also a unit vector ($||\boldsymbol{\mu}|| = 1$), is the mean direction, $\kappa$ is the concentration parameter and $S^2$ is the unit 2-sphere. Because (5) is symmetrical about $\boldsymbol{\mu}$, the mean direction of $\boldsymbol{x}$ is

$\boldsymbol{\mu}$. For $\kappa > 0$, the distribution has a mode at the mean direction $\boldsymbol{\mu}$, whereas when $\kappa = 0$ the distribution is uniform. The larger the $\kappa$ the greater the clustering around the mean direction. Since (5) depends on $\boldsymbol{x}$ solely through $\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{x}$, the vMF is rotationally symmetric about $\boldsymbol{\mu}$.

The von Mises-Fisher distribution, like many well known parametric distributions (Gaussian, Poisson, Gamma, Dirichlet etc.), is an exponential family [16]. A parametric set of probability distributions defined on a sample space $\mathcal{X}$ and parametrized by the natural parameter $\theta \in \Theta$ is called *exponential family* if their probability densities admit the following canonical representation

$$p(x;\theta) = \exp(T(x)\cdot\theta - F(\theta) + C(x)), \quad x \in \mathcal{X}. \tag{6}$$

$T(x)$ is called the minimal sufficient statistics, and functions $F$ and $C$ denote the log-normalizer and the carrier measure, respectively. It can be readily checked that the vMF distribution $\mathfrak{f}(\boldsymbol{\mu},\kappa)$ defined by (5) with standard parameters $\boldsymbol{\mu}$ and $\kappa$, is an exponential family parametrized by the natural parameter $\boldsymbol{\theta} = \kappa\boldsymbol{\mu}, \boldsymbol{\theta} \in \mathbb{R}^3$. The minimal sufficient statistics is $T(x) = \boldsymbol{x}$, the log-normalizer is given by $F(\boldsymbol{\theta}) = \log 4\pi \sinh(\kappa)/\kappa$, and the carrier measure is trivial $C = 0$.

### 3.2 Bayes tracker based on the vMF distribution

Since our goal is to track motion on the sphere, we proceed further by posing the problem as an estimation on a sphere thus devising a Bayesian state estimator (tracker) based solely on the vMF distribution [17,18]. A Bayesian estimation procedure of the a posteriori pdf consists of two steps: prediction and update [19], which in this case entails representing the state to be estimated $\boldsymbol{x}_t$ at time $t$ as the vMF distribution and successively predicting and updating this distribution. The prediction step involves calculating the pdf via the total probability theorem

$$p(\boldsymbol{x}_t \,|\, \boldsymbol{z}_{1:t-1}) = \int p(\boldsymbol{x}_t \,|\, \boldsymbol{x}_{t-1}) p(\boldsymbol{x}_{t-1} \,|\, \boldsymbol{z}_{1:t-1}) \mathrm{d}\boldsymbol{x}_{t-1}. \tag{7}$$

In this case we do not have a strict state evolution model, but we choose to add process noise governed by a centered vMF in the prediction stage which amounts to convolving our posterior at time $t-1$ with the vMF distribution representing the process noise. Let us assume that the prior state at $t-1$ is represented by a vMF distribution $\mathfrak{f}(\boldsymbol{\mu}_{t-1}, \kappa_{t-1})$ and that the motion model is represented by a vMF distribution centered at $\boldsymbol{\mu}_{t-1}$ but with concentration parameter $\kappa_{_Q}$. The result of (7) would not produce an another vMF distribution, but the result of this operation can be well approximated by a vMF with a suitably chosen value of the resulting $\kappa_{t|t-1}$ [15,17]

$$\kappa_{t|t-1} = A^{-1}(A(\kappa_{t-1})A(\kappa_{_Q})), \quad A(\kappa) = \frac{1}{\tanh\kappa} - \frac{1}{\kappa}. \tag{8}$$

Consequently, after the prediction step our state represented by a single vMF will have unchanged mean direction, $\boldsymbol{\mu}_{t|t-1} = \boldsymbol{\mu}_{t-1}$, but newly calculated concentration parameter via (8).

In the update step, the posterior at time $t$ is calculated via the Bayes theorem

$$p(\boldsymbol{x}_t \,|\, \boldsymbol{z}_{1:t}) = \frac{p(\boldsymbol{z}_t \,|\, \boldsymbol{x}_t) p(\boldsymbol{x}_t \,|\, \boldsymbol{z}_{1:t-1})}{p(\boldsymbol{z}_t \,|\, \boldsymbol{z}_{1:t-1})}. \tag{9}$$

The sensor model $p(\boldsymbol{z}_t \,|\, \boldsymbol{x}_t)$ is represented by a vMF distribution $\mathfrak{f}(\boldsymbol{z}_t, \kappa_{_R})$ centered at the detected moving objects in the image after being lifted to unit sphere, while the predicted state $p(\boldsymbol{x}_t \,|\, \boldsymbol{z}_{1:t-1})$ will be the result of the previously discussed (7). Given these two vMF distributions, $\mathfrak{f}(\boldsymbol{z}_t, \kappa_{_R})$ and $\mathfrak{f}(\boldsymbol{\mu}_{t|t-1}, \kappa_{t|t-1})$, the result of the update step is a vMF distribution $\mathfrak{f}(\boldsymbol{\mu}_{t|t}, \kappa_{t|t})$

with the following parameters [17]

$$\begin{aligned}\kappa_{t|t} &= \sqrt{\kappa_{t|t-1}^2 + \kappa_R^2 + 2\kappa_{t|t-1}\kappa_R(\boldsymbol{\mu}_{t|t-1} \cdot \boldsymbol{z}_t)} \\ \boldsymbol{\mu}_{t|t} &= \frac{\kappa_{t|t-1}\boldsymbol{\mu}_{t|t-1} + \kappa_R \boldsymbol{z}_t}{\kappa_{t|t}}.\end{aligned} \tag{10}$$

The two steps, governed by (8) and (10), will cyclically produce the estimate of the direction of the moving object. Methods for practical calculation of some of the aforementioned equations can be found in [20]. We have used this recursion in [18] in order to track a single moving object in the omnidirectional image acquired by a perspective camera with a fish-eye lens and to follow the object by visual servoing, but in the same paper the problems of multiple target tracking were not analyzed.

### 3.3 Validation gate

In multiple target tracking it is often practical to devise a *validation gate* so as to reject highly unlikely measurements. This way the computational complexity of the association procedure can be significantly lowered [6]. In the case of tracking, the validity of a measurement $\boldsymbol{z}_t$ can be calculated by the following expression [21]

$$p(\boldsymbol{z}_t \,|\, \boldsymbol{z}_{1:t-1}) = \int p(\boldsymbol{z}_t \,|\, \boldsymbol{x}_t) p(\boldsymbol{x}_t \,|\, \boldsymbol{z}_{1:t-1}) \mathrm{d}\boldsymbol{x}_t. \tag{11}$$

If the underlying distribution is Gaussian then (11) produces a Gaussian distribution with the innovation vector and the innovation covariance matrix as the distribution parameters [21]. In this paper we propose to solve (11) for the case of the vMF distribution and utilize the result for validation gating. We can note that the two densities in (11) are exactly the measurement model and the prediction which yields

$$\begin{aligned} p(\boldsymbol{z}_t \,|\, \boldsymbol{z}_{1:t-1}) &= \int \frac{\kappa_{t|t-1}\kappa_R}{4\pi \sinh \kappa_{t|t-1} \sinh \kappa_R} \exp\{\kappa_{t|t-1}\mu_{t|t-1}^{\mathrm{T}} \boldsymbol{x}_t + \kappa_R \boldsymbol{z}_t^{\mathrm{T}} \boldsymbol{x}_t\} \mathrm{d}\boldsymbol{x}_t \\ &= \frac{\kappa_{t|t-1}\kappa_R}{4\pi \sinh \kappa_{t|t-1} \sinh \kappa_R} \int \exp\{\kappa_{t|t} \boldsymbol{\mu}_{t|t}^{\mathrm{T}} \boldsymbol{x}_t\} \mathrm{d}\boldsymbol{x}_t \\ &= \frac{\kappa_{t|t-1}\kappa_R}{4\pi \sinh \kappa_{t|t-1} \sinh \kappa_R} \frac{4\pi \sinh \kappa_{t|t}}{\kappa_{t|t}} \int \frac{\kappa_{t|t}}{4\pi \sinh \kappa_{t|t}} \exp\{\kappa_{t|t} \boldsymbol{\mu}_{t|t}^{\mathrm{T}} \boldsymbol{x}_t\} \mathrm{d}\boldsymbol{x}_t \\ &= \frac{\kappa_{t|t-1}\kappa_R \sinh \kappa_{t|t}}{\kappa_{t|t} \sinh \kappa_{t|t-1} \sinh \kappa_R}. \end{aligned} \tag{12}$$

The result of the derivation (12) is not a vMF distribution, but analogous to the case of the distribution on the unit circle, the von Mises distribution [15], we approximate this result with a vMF density $p(\boldsymbol{z}_t; \boldsymbol{\mu}_{t|t-1}, A^{-1}(A(\kappa_R)A(\kappa_{t|t-1})))$. For the vMF distribution the approximate $100(1-\alpha)\%$ validation region for $\boldsymbol{z}_t$ is [15]

$$\{\boldsymbol{z}_t \,:\, \boldsymbol{z}_t^{\mathrm{T}} \boldsymbol{\mu}_{\boldsymbol{t|t-1}} \geq \cos\delta\}, \tag{13}$$

which defines the intersection of the unit sphere with the cone having vertex at the origin, axis the mean direction $\boldsymbol{\mu}_{t|t-1}$ and semi-vertical angle $\delta$. Given the parameter $\alpha$, from numerical tables given in [15] we can define the angle $\delta$, which separates measurements which fall within the validation region from those that do not. In conclusion, if the scalar product $\boldsymbol{z}_t^{\mathrm{T}} \boldsymbol{\mu}_{t|t-1}$ is smaller than $\cos\delta$ then the measurement $\boldsymbol{z}_t$ is considered to be inside the validation region and is taken into account during the association process.

### 3.4 Data Association

In this paper we solve the data association on the sphere using the GNN and since the method requires a notion of distance to be calculated between the measurement likelihoods and the predicted states we propose the following procedure based on statistical distances. Given that, we derive the Rényi $\alpha$-divergence [22] for the vMF distribution as a distance which will be used to determine the measurement-to-track associations. The reason we chose this divergence is because it is a statistical and information theoretical class of generalized distances which includes cases such as the Kullback-Leibler (KL) distance and the Bhattacharyya distance. We have used this divergence in our previous work for reducing the number of the components in a mixture of von Mises distributions [23]. The Rényi $\alpha$-divergence is given by the following expression

$$D_R^{(\alpha)}(p,q) = \frac{1}{\alpha - 1} \log \int_{\mathcal{X}} p(x)^{\alpha} q(x)^{1-\alpha} \mathrm{d}x$$

and is parametrized by real parameter $\alpha$, which in the limit $\alpha \to 1$ yields the KL distance, which is defined as follows

$$D_{\mathrm{KL}}(p,q) = \int_{\mathcal{X}} p(x) \log \left( \frac{p(x)}{q(x)} \right) \mathrm{d}x.$$

In order to calculate the distance $D_F^{(\alpha)}$ between two vMF densities, $p(\boldsymbol{x}; \boldsymbol{\mu}_p, \kappa_p)$ and $q(\boldsymbol{x}; \boldsymbol{\mu}_q, \kappa_q)$ with their respective parameters $\theta_p$ and $\theta_q$, we will use the following closed form expressions for the exponential family of distributions

$$D_F^{(\alpha)}(\theta_p, \theta_q) = \begin{cases} \dfrac{1}{1-\alpha} J_F^{(\alpha)}(\theta_p, \theta_q), & \alpha \in (0,1), \\ B_F(\theta_q, \theta_p), & \alpha = 1, \end{cases} \tag{14}$$

where $B_F$ denotes the *Bregman divergence* [24] generated by the convex function $F$

$$B_F(\theta_q, \theta_p) = F(\theta_q) - F(\theta_p) - \nabla F(\theta_p) \cdot (\theta_q - \theta_p), \quad \theta_q, \theta_p \in \Theta \tag{15}$$

and where $J_F^{(\alpha)}$ denotes the *Jensen $\alpha$-divergence* [25] generated by the convex function $F$,

$$J_F^{(\alpha)}(\theta_p, \theta_q) = \alpha F(\theta_p) + (1-\alpha) F(\theta_q) - F(\alpha \theta_p + (1-\alpha)\theta_q), \quad \theta_p, \theta_q \in \Theta.$$

Bregman divergence is used for the case of $\alpha = 1$ since $B_F(\theta_q, \theta_p) = D_{\mathrm{KL}}(p,q)$—note that the parameters are in reverse order. For the case of the vMF distribution calculations reveal that for $\alpha \in (0,1)$

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

where $\kappa_{pq} = ||\alpha \boldsymbol{\mu}_p \kappa_p + (1-\alpha) \boldsymbol{\mu}_q \kappa_q||$ and that for $\alpha = 1$, i.e. the KL distance, we have

$$D_{\mathrm{KL}}(p,q) = \log \frac{\kappa_p \sinh \kappa_q}{\kappa_q \sinh \kappa_p} - \left( \frac{1}{\tanh \kappa_p} - \frac{1}{\kappa_p} \right) \boldsymbol{\mu}_p^{\mathrm{T}} \left( \kappa_q \boldsymbol{\mu}_q - \kappa_p \boldsymbol{\mu}_p \right). \tag{17}$$

With having defined the appropriate distance measure between the distributions we can form a distance matrix $\Omega = [d_{ij}]$ whose rows represent the objects currently being tracked and whose columns represent current measurements and $d_{ij}$ is the Rényi $\alpha$-divergence between the vMF representing the predicted state and the vMF representing the measurement likelihood. To solve the assignment problem, we used the Hungarian algorithm [26, 27] which yielded the association matrix having ones at the rows and columns designating the calculated measurement-to-track association.

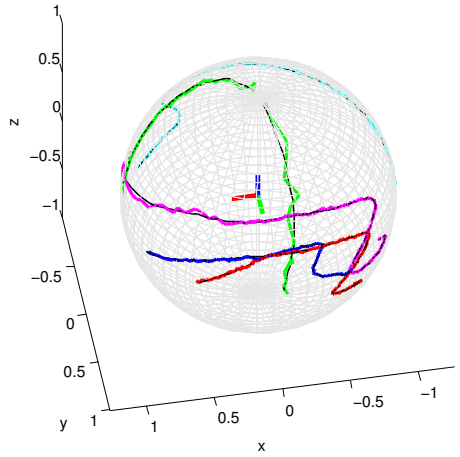
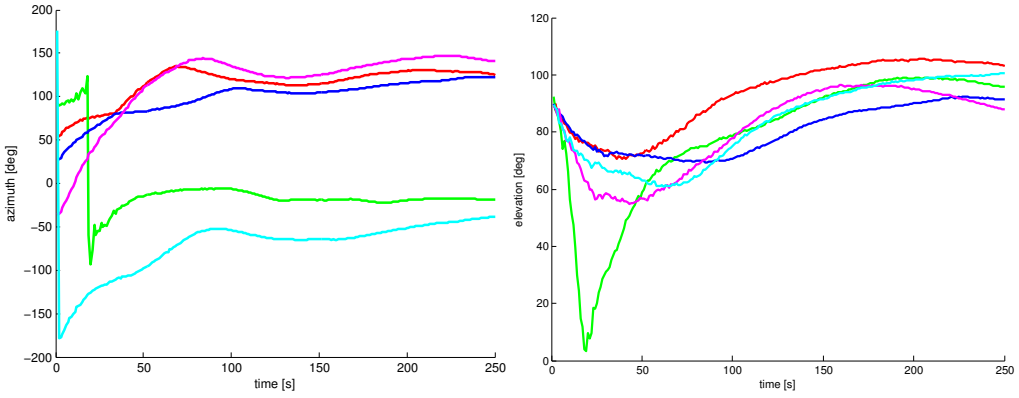



Fig. 2: Estimated trajectories on the unit sphere




(a) azimuth (b) elevation

Fig. 3: Estimated azimuth and elevation of the tracked objects on the unit sphere

## 4 Experiments

### 4.1 Synthetic data

The synthetic data included multiple moving objects in 3D following a jump-state Markov model. Basically, each trajectory had a constant velocity but experienced random changes in the azimuth and elevation angles—the angles either stayed the same or changed $\pm 5°$. Furthermore, the direction measurements of the trajectories were obtained by adding Gaussian noise to the Cartesian 3D coordinates and then normalizing the noisy vectors to the unit sphere. For the data association distance metric we have used the Rényi $\frac{1}{2}$-divergence, i.e. the Bhattacharyya distance, since it is symmetrical $D_R^{(\frac{1}{2})}(p,q) = D_R^{(\frac{1}{2})}(q,p)$. It is also positive definite, but like the other Rényi $\alpha$-divergence it does not satisfy the triangle inequality.

To test the algorithm a scene with five moving objects was generated. Estimated directions on the unit sphere are shown in Fig. 2, while the azimuth and elevations of the trajectories

are shown in Figs. 3a and 3b, respectively. The results show that the algorithm is capable of associating measurements to tracks so as to track all the moving objects in the scene. However, with the approach presented in this paper we are estimating only the direction of the objects without a motion model that predicts the future position. Given that, if two trajectories cross each other at approximately the same time, the algorithm will associate the measurement that is probabilistically closest in terms of the previous position. In that case, without taking into account the history of the motion, it is possible that the identities of two tracks might get confused. In the given example, this has happened at 34 s and can be more clearly seen in the azimuth angle in Fig. 3a where the red and blue trajectories were switched after the crossing happened. This effect is also visible in Fig. 2 where the blue and red trajectories crossed for the first time—the red trajectory should have continued in the north-east direction instead of switching to the straight motion to the east of the blue trajectory. Such situations could be alleviated by probabilistic data association techniques in the vein of [7], multiple hypothesis tracking in the vein of [28], or data association and track management in the vein of [9] for the Gaussian mixture PHD filter [10], but this is a subject of future research.

### 4.2 Real-world data

The real-world data experiment was conducted with the Ladybug®2 multiple-camera system mounted on a Pioneer 3DX mobile robot. The scene included two humans walking around the mobile robot randomly changing the direction while the mobile robot executed rotational and translational motion during the experiment. Each camera reported an image of 1024x768 pixels at 7fps which were collected and processed using the Robot operating system (ROS) [29] and OpenCV [30].

## 5 Conclusion

In this paper we have proposed a novel method for tracking multiple moving objects on the unit sphere. The method was tested in a scenario where a multiple-camera system was placed on a mobile robot and detection of motion was performed by warping images between consequtive frames and subtracting the same in order to detect motion in each camera image. After the subtraction each pixel with sufficient intensity representing the motion was lifted to the sphere where clustering of objects was performed and whose center of gravity yielded measurements in the form of the vMF distribution. Subsequently, the tracking was performed by a Bayesian tracker based on the vMF distribution and where the data association was solved by GNN where the Rényi $\alpha$-divergence was proposed as a distance metric. Furthermore, in order to alleviate the complexity of the association process, a validation gate was derived for the vMF filter. The unlikely measurements were disregarded during assoction but they did participate in the possible track initialization process which was handled by the M/N logic algorithm. In the end, the proposed method was tested on a synthetic and real-world data experiments which confirmed the prospects of the proposed method.

## Acknowledgements

This work has been supported by European Community's Seventh Framework Programme under grant agreement no. 285939 (ACROSS) and research project VISTA (EuropeAid/131920/M /ACT/HR).

## References


1. B. Jung and G. Sukhatme, "Real-time motion tracking from a mobile robot," *International Journal of Social Robotics*, vol. 2, no. 1, pp. 63–78, 2009.
2. E. Einhorn, M. Filzhuth, C. Schröter, and H.-M. Gross, "Monocular detection and estimation of moving obstacles for robot navigation," in *European Conference on Mobile Robots (ECMR)*, 2011, pp. 121–126.
3. A. Kundu, K. M. Krishna, and C. V. Jawahar, "Realtime motion segmentation based multibody visual SLAM," in *Indian Conference on Computer Vision, Graphics and Image Processing -ICVGIP '10*, 2010, pp. 251–258.
4. K. E. Ozden, K. Schindler, and L. Van Gool, "Multibody structure-from-motion in practice." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 6, pp. 1134–41, Jun. 2010.
5. A. A. Mekonnen, C. Briand, F. Lerasle, and A. Herbulot, "Fast HOG based person detection devoted to a mobile robot with a spherical camera," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Nov. 2013, pp. 631–637.
6. S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems.* Artech House Publishers, 1999.
7. Y. Bar-Shalom and E. Tse, "Sonar tracking of multiple targets using joint probabilistic data association filter," *Automatica*, vol. 11, pp. 451–460, 1975.
8. R. P. S. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1152–1178, Oct. 2003.
9. K. Panta, D. E. Clark, and B.-N. Vo, "Data association and track management for the Gaussian mixture probability hypothesis density filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 3, pp. 1003–1016, 2009.
10. B.-N. Vo and W.-K. Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4091–4104, Nov. 2006.
11. B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision (darpa)," in *Proceedings of the 1981 DARPA Image Understanding Workshop*, April 1981, pp. 121–130.
12. C. Tomasi and T. Kanade, "Detection and tracking of point features," International Journal of Computer Vision, Tech. Rep., 1991.
13. F. Lu and E. Milios, "Robot pose estimation in unknown environments by matching 2d range scans," *J. Intell. Robotics Syst.*, vol. 18, no. 3, pp. 249–275, Mar. 1997.
14. N. I. Fisher, T. Lewis, and B. J. J. Embleton, *Statistical Analysis of Spherical Data.* Cambridge University Press, 1993.
15. K. V. Mardia and P. E. Jupp, *Directional Statistics.* New York: Wiley, 1999.
16. S. R. Jammalamadaka and A. Sengupta, *Topics in Circular Statistics.* World Scientific, 2001.
17. A. Chiuso and G. Picci, "Visual tracking of points as estimation on the unit sphere," *The Confluence of Vision and Control, Lecture Notes in Control and Information Sciences*, vol. 237, pp. 90–105, 1998.
18. I. Marković, F. Chaumette, and I. Petrović, "Moving object detection, tracking and following using an omnidirectional camera on a mobile robot," in *International Conference on Robotics and Automation (ICRA)*, 2014.
19. S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics.* The MIT Press, 2006.
20. W. Jakob, "Numerically stable sampling of the von Mises-Fisher distribution on S^2 (and other tricks)," Interactive Geometry Lab, ETH Zürich, Tech. Rep., 2012.
21. T. Bailey, B. Upcroft, and H. Durrant-Whyte, "Validation Gating for Non-Linear Validation gating for non-linear non-Gaussian target tracking," in *International Conference on Information Fusion*, 2006, pp. 1–6.
22. A. Rényi, "On Measures of Entropy and Information," in *Proc. Fourth Berkeley Symp. Math. Stat. and Probability*, ser. Vol. 1. University of California Press, 1961, pp. 547–561.
23. M. Bukal, I. Marković, and I. Petrović, "Composite distance based approach to von Mises mixture reduction," *Accepted for publication in Information Fusion*, 2014.

24. L. M. Bregman, "The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming," *USSR Computational Mathematics and Mathematical Physics*, vol. 7, no. 3, pp. 200—-217, 1967.
25. F. Nielsen, "Chernoff information of exponential families," *arXiv:*, vol. 1102.2684, 2011.
26. H. W. Kuhn, "The Hungarian method for the assignement problem," *Naval Research Logistics Quarterly*, vol. 2, pp. 83–97, 1955.
27. J. Munkres, "Algorithms for the assignment and transportation problems," *Journal of the Society for Industrial and Applied Mathematics*, vol. 5, no. 1, pp. 32–38, 1957.
28. D. Reid, "An algorithm for tracking multiple targets," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 843–854, 1979.
29. M. Quigley, B. Gerkey, K. Conley, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler, and A. Ng, "ROS : an open-source Robot Operating System," *IEEE International Conference on Robotics and Automation (ICRA), Workshop on Open Source Software*, 2009.
30. G. Bradski, "The OpenCV library," *Dr. Dobb's Journal of Software Tools*, 2000.