Mixtures of von Mises Distributions for People Trajectory Shape Analysis

Simone Calderara, Andrea Prati, *Senior Member, IEEE,* and Rita Cucchiara *Member, IEEE*

***Abstract*—People trajectory analysis is a recurrent task in many pattern recognition applications, such as surveillance, behavior analysis, video annotation, and many others. In this paper, we propose a new framework for analyzing trajectory *shape*, invariant to spatial shifts of the people motion in the scene. In order to cope with the noise and the uncertainty of the trajectory samples, we propose to describe the trajectories as a sequence of angles modeled by distributions of circular statistics, i.e., a mixture of von Mises (MovM) distributions. To deal with MovM, we define a new specific expectation-maximization (EM) algorithm for estimating the parameters and derive a closed form of the Bhattacharyya distance between single von Mises pdfs. Trajectories are then modeled with a sequence of symbols, corresponding to the most suitable distribution in the mixture, and compared each other after a global alignment procedure to cope with trajectories of different lengths. The trajectories in the training set are clustered according to their shape similarity in an off-line phase, and testing trajectories are then classified with a specific on-line EM, based on sufficient statistics. The approach is particularly suitable for classifying people trajectories in video surveillance, searching for abnormal (i.e., infrequent) paths. Tests on synthetic and real data are provided with also a complete comparison with other circular statistical and alignment methods.**

***Index Terms*—Circular statistics, trajectory shape analysis, von Mises distribution.**

I. Introduction

MANY SCENARIOS where video analysis is applied, such as video surveillance or sport video retrieval, require learning and classification of patterns of movement. In particular, the *trajectory* of moving people is one of the most common and discriminative pattern descriptors. Indeed, recent advances in object tracking algorithms have made it possible to obtain reliable and long people trajectories also across multiple cameras. These trajectories are often prone to errors, noise and uncertainty but their analysis can allow the inference of people's behavior; for instance, in video surveillance it could help to identify abnormal movements, by detecting patterns that deviate from normal cases. Without any *a priori* knowledge about the context, trajectory patterns can be classified as normal or abnormal only by considering their statistical occurrence. Many techniques which have been successfully applied in large databases and often adopted in time series data mining cannot be suitably employed in surveillance since they are too sensitive to noise. Instead, the adoption of statistical pattern recognition approaches, which learns the probability distribution of normal patterns, divides them in clusters of similarity, and then, given a new pattern, evaluates its fitness through the posterior probability, could be convenient also for trajectory analysis [1].

Basically, people trajectories can be represented in the reference ground plane as sequences of 2-D coordinates $\{(x_1, y_1), \cdots, (x_n, y_n)\}$ often associated with the motion status, e.g., the punctual velocity or acceleration [1], and similarity can be computed in the (x, y) domain. However, in many cases, instead of analyzing the spatial relationships of single trajectory points, the focus should be given to a more global descriptor of the trajectory, i.e., the *trajectory shape*. The shape is independent of the starting point and could constitute a very effective descriptor of the taken action. In surveillance of large public spaces, the trajectory shape could discriminate between different behaviors, such as people moving on a straight path or moving in a circle.

To copewith this, we propose a new method to model trajectory shapes by means of sequences of angles using suitable pattern recognition techniques in order to provide a comparison invariant to spatial shifts but taking into account the sequence of directions and their changes. Since angles are periodic variables, the classical assumption that measures are Gaussian-distributed is unsuitable. By exploiting the circular statistics, used until now only in few other pattern recognition contexts, e.g., to model wind directions [2] or audio signals [3], we propose a new representation based on a mixture of circular pdfs that naturally deals with the peculiarities of angular and periodic data.

This paper has two main objectives: to provide a mathematical description of the possible use of mixture of von Mises (MovM) pdfs in pattern analysis, and to define a complete approach for comparing people trajectories and detecting their abnormal shapes. The main contributions of the paper are the following:

1) the use of circular statistics for handling periodic feature vectors, by defining original expectation-maximization (EM) algorithms for the estimation of parameters of

Manuscript received April 27, 2010; revised August 27, 2010; accepted October 2, 2010. Date of publication March 10, 2011; date of current version April 1, 2011. This work was supported in part by the Project BESAFE funded by the NATO SfP Program. This paper was recommended by Associate Editor J. Cai.

S. Calderara and R. Cucchiara are with the Dipartimento di Ingegneria dell'Informazione, University of Modena and Reggio Emilia, Modena 41125, Italy (e-mail: simone.calderara@unimore.it; rita.cucchiara@unimore.it).

A. Prati is with the Dipartimento di Scienze e Metodi dell'Ingegneria, University of Modena and Reggio Emilia, Reggio Emilia 42122, Italy (e-mail: andrea.prati@unimore.it).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2011.2125550

mixtures of different circular statistical distributions, namely von Mises, Wrapped Gaussian, and approximate Wrapped Gaussian pdfs;
2) the formulation of a closed form of the Bhattacharyya distance for von Mises (vM) pdf comparison; it can be exploited to compare both sets and sequences of angles;
3) the definition of a complete approach for trajectory shape analysis in applications of people surveillance based on a sequential representation of angles with circular statistics;
4) the proposal of an on-line EM for providing real-time implementation of the approach with sufficient statistics.

Different circular statistics are tested on synthetic and real data of video surveillance. While synthetic data are used as a proof of concepts of many properties (such as robustness to noise), real data constitute a strong test of the effectiveness of the proposal. We provide a dataset of more than 1000 trajectories, taken in different days by an automatic surveillance system. Many clusters of similarity are taken initially and updated in time with a continuous learning paradigm, and anomalies are detected with a very high degree of accuracy. Further experiments demonstrated that the method is robust and easily applicable in other contexts of shape analysis.

II. Related Works

Trajectory analysis has been studied in depth in different contexts and, especially in the last years, also for people surveillance. People trajectories differ from other trajectories of moving objects (e.g., vehicles) for the unpredictability of the paths, for which simple predictive models would fail. Trajectories represented as time series are normally compared point-to-point, with many different measures, of which Euclidean distance is the simplest, but it does not take into account different lengths. Dynamic time warping (DTW) [4], [5], and longest common sub sequence (LCSS) [6] are often exploited with good results in querying for similarity. Many authors stated that despite their robustness in finding similarity among large sets of data, their efficiency decreases in the presence of noise and uncertainty [7]–[9]; for this reason, Chen *et al.* [10] proposed a distance measure called edit distance in real sequence exhibiting better performance dealing with trajectories affected by Gaussian noise. The work of Pelekis *et al.* [7] gave a very interesting overview of possible distances used to analyze databases of trajectories, by comparing plain Euclidean distance, DTW, and LCSS. Marteau defined a time warp edit distance which outperforms the other distances [11].

However, in typical video surveillance scenarios, the trajectories are detected by means of computer vision algorithms and thus are affected by errors, noise, and high uncertainty. They have very different lengths because of the different speeds of the objects and the inaccuracy of tracking algorithms. Moreover, surveilled scenarios often change dynamically so that long-lasting storage of people paths is not possible and many examples of similar paths are not available. For instance, the work of Zhang *et al.* [8] discussed several methods for video surveillance (Euclidean distance [12], an improved version of principal component analysis [13], DTW and LCSS), and showed that, in the case of noisy data, PCA+Euclidean measure is the most robust solution. Hausdorff distance was adopted by both Lou *et al.* [14] and Junejo *et al.* [15], although it is unable to distinguish trajectories in the same path but opposite directions [8]. In [15], trajectories are clustered with Hausdorff distance and DTW is only used within the clusters to find the mean trajectory.

In the case of few and noisy data, a statistical model is still preferable. Chen *et al.* [16] presented a method to compare the trajectories after projecting them in a null-space to obtain a representation insensitive to projective transformation of the trajectories themselves. Although it increases significantly the robustness if the data are acquired from different viewpoints, it does not result discriminative enough to distinguish between trajectories that differ only in the main directions since the null-space representation is invariant to sheer, scale, and rotation transformations. In [17], an edit distance is proposed that gives a penalty to local temporal shifts, by determining the costs for insertion and deletion and accounting to the possible different speeds. Similarly, we propose to use a global alignment method over the trajectories and in particular the Needleman-Wunsch (N-W) algorithm [18] where the cost of gaps in the alignment is accounted. In this manner, two trajectories with similar shapes but different lengths are considered less similar to two trajectories with similar shape and same length [9], [19]. In [19], the use of this type of global alignment was initially provided on trajectories coded as a sequence of symbols of a codebook based on sampled angles and speeds, without any statistical model. Similarly, in [9] the authors proposed an edit distance with a similar global alignment approach on a sequence of symbols representing angles, speeds, and discontinuities. Only few papers have considered trajectories modeled by means of angles: for instance, in [20] a rotation invariant similarity measure based on LCSS computed on angles is provided, tested on very few synthetic curves.

In surveillance contexts, in order to search similarities and detect anomalies, the similarity measures have been typically defined in a statistical framework. In [21], the EM algorithm is adopted to learn the motion models' parameters given the spatial positions of the points of each trajectory as observations. Mecocci and Panozzo [22] suitably modified the iterative altruistic vector quantization algorithm to robustly cluster trajectories by pure spatial observations obtaining representative prototypes. The anomaly detection is based on fitting a spatial Gaussian on each prototype and statistically checking for fitness of new trajectory samples. Hu *et al.* [23] reported a complete system for learning statistical motion patterns with a two-stage fuzzy k-means. First, trajectories are clustered in the spatial domain, then each cluster is subclustered in the temporal domain. Porikli and Haga [24] proposed the use of a HMM-based similarity measure where each trajectory is modeled with a HMM and compared using the cross likelihood. The results are promising but, in general, a large amount of data is needed to avoid over-fitting in the HMM training phase. In [25], Basharat *et al.* proposed an

interesting framework to capture anomalous path transitions in tracking data. A mixture of Gaussians model is trained at each pixel location where a transition occurs and training data are constantly updated on-line for capturing new observations. Also our proposal substitutes the comparison on direct values with that on statistical pdfs. Since we aim at working on angular data, we propose a novel use of circular statistics and in particular of von Mises pdf.

A complete analysis of the vM distribution is reported in [26]. In general, the use of circular statistics for computer vision applications is very limited. Some periodic versions of the Gaussian distribution have been proposed for angular data [3], [27]. In [28], a Wrapped Gaussian is used to approximate vM and model the hue in the HSV color space, for modeling the background with the purpose of segmenting moving people. A very interesting approximated version of the Wrapped Gaussian has been presented by Bahlmann in [29] to deal with semi-periodic multivariate data in handwritten character recognition. An initial discussion on the possible use of MovM for trajectory analysis was presented in [30], here extended with the description of the complete approach, the EM for parameter estimation and the comparison with mixtures of other circular pdfs.

III. Circular Statistics for Angular Data Analysis

Many natural phenomena involve periodic measures and angular observations. Angular features are particularly suitable to model periodic events in images and videos, such as motion vectors, texture orientation, boundary directions, or trajectory angles. Dealing with sets of angles, periodic statistical models that preserve their periodic property and cope with noise and uncertainty are desirable.

Classical linear statistics suffers the severe flaw of not accounting for periodicity of the observations. For example, Gaussian pdfs are inappropriate since when applied on a periodic variable they will bring results that are strongly dependent on the arbitrary choice of the origin. Driven by these limitations, in the early 1960s circular statistics and pdfs that share a wrapping behavior have been introduced. Although the analytical treatment of circular statistics is well established, its numerical solutions and efficient algorithms for its exploitation in pattern recognition are not so frequent, due to the computational complexity of some calculations which made them too costly with past computational resources.

A. Modeling Sets of Angles

Let us consider an unordered collection of independent and identically distributed (i.i.d.) angular features, i.e., a *set of angles*. We propose to model this set by means of the vM distribution because it is "circularly defined" and independent of the origin. The vM pdf is a special case of the von Mises-Fisher pdf [31]. When the variable is univariate, the vM pdf results to be

$$\mathcal{V}(\theta|\theta_0, m) = \frac{1}{2\pi I_0(m)} e^{m\cos(\theta-\theta_0)} \tag{1}$$

where I_0 in the normalization factor is the modified zero-order Bessel function of the first kind.

Parameter $m \geq 0$ denotes the *concentration* of the distribution around the mean θ_0 and for $m = 0$ the distribution collapses in a uniform distribution. The distribution is periodic so that $p(\theta + M2\pi) = p(\theta)$ for all θ and any integer M. The vM distribution yields many of the key properties for statistical inference that the normal distribution has for linear data.

If we adopt an unimodal vM pdf for describing a set of n i.i.d. samples of angular features $\boldsymbol{\theta} = \langle \theta_1, ..., \theta_n \rangle$, the pdf parameters can be inferred using the maximum likelihood (ML), with the following equations:

$$\theta_0^{ML} = \tan^{-1}\left\{ \frac{\sum_{i=1}^{n} \sin\theta_i}{\sum_{i=1}^{n} \cos\theta_i} \right\}$$

$$A\left(m^{ML}\right) \equiv \frac{I_1\left(m^{ML}\right)}{I_0\left(m^{ML}\right)} = \frac{1}{n}\sum_{i=1}^{n} \cos\left(\theta_i - \theta_0^{ML}\right) \tag{2}$$

where the solution for m^{ML} can be found by inverting (2) numerically.

Often a unimodal pdf is not significant enough to represent the manifold data. Accordingly, we propose to describe a set of angles with a MovM distribution. It can be defined as

$$MovM(\theta|\boldsymbol{\theta_0}, \boldsymbol{m}) = \sum_{k=1}^{K} \pi_k \mathcal{V}\left(\theta|\theta_{0,k}, m_k\right) \tag{3}$$

where K is the number of mixture's components and π_k the weight of the kth component.

The ML framework of (2) is not suitable for mixture of distributions since it cannot handle singularities. Since the mixture model depends on unobserved latent variables (defining the "responsibilities" of a given sample with respect to a given component of the mixture), a well known tool for finding maximum likelihood estimates of the mixture parameters is the EM algorithm. As far as we know, EM has never been analytically defined for a MovM distribution; therefore, in the following we derive the EM steps for a MovM distribution (details can be found in Appendix A).

1) *Expectation Step:* Given a $MovM(\theta|\boldsymbol{\theta_0}, \boldsymbol{m})$ composed of K components, the responsibility γ_{ik} of component k for sample θ_i can be estimated using the parameter values of the previous iteration (randomly initialized for the first iteration) as follows:

$$\gamma_{ik} = \frac{\pi_k \mathcal{V}\left(\theta_i|\theta_{0,k}, m_k\right)}{\sum_{s=1}^{K} \pi_s \mathcal{V}\left(\theta_i|\theta_{0,s}, m_s\right)}. \tag{4}$$

2) *Maximization Step:* The M step estimates the new values of the mixture's parameters $\boldsymbol{\pi} = \{\pi_1, \cdots, \pi_K\}$, $\boldsymbol{\theta_0} = \left\{\theta_{0,1}, \cdots, \theta_{0,K}\right\}$, and $\boldsymbol{m} = \{m_1, \cdots, m_K\}$. They can be estimated for each component

k as

$$\pi_k = \frac{1}{n}\sum_{i=1}^{n}\gamma_{ik} \tag{5}$$

$$\theta_{0,k} = \arctan\left(\frac{\sum_{i=1}^{n}\gamma_{ik}\sin\theta_i}{\sum_{i=1}^{n}\gamma_{ik}\cos\theta_i}\right) \tag{6}$$

$$A(m_k) = \frac{\sum_{i=1}^{n}\gamma_{ik}\cos\left(\theta_i - \theta_{0,k}\right)}{\sum_{i=1}^{n}\gamma_{ik}}. \tag{7}$$

Similarly to (2), the right side of (7) can be evaluated and then the value of m_k can be computed by inverting the function numerically.

3) *Iterate Procedure:* Steps 1 and 2 are iterated until convergence (reached when the likelihood does not change too much between two consecutive iterations) or until a given number of iterations is reached.

If the considered mixture actually contains less than K main modalities, some estimated components will have the same parameters. Moreover, only the most significant components (in terms of weight) are retained until 90% of the samples are covered. Thus, the value of K can be kept arbitrarily high (although the problem known as "curse of dimensionality" should be considered) without significantly affecting the system's performance. However, more sophisticated approaches could be used to select the best K, as that proposed in [32].

B. Modeling Sequences of Angles

When some order of angular features is provided or imposed, speaking of "*sequences of angles*" is more appropriate and the model must take into account the ordering of data. Thus, we extend the model to incorporate both the statistical distribution (by means of a MovM) and the sequence ordering by transforming the sequence of angles into a sequence of symbols associated to the vM component of the mixture which maximizes the posterior probability.

Given a sequence $T = \{\theta_1, ..., \theta_n\}$ of angular values: 1) the EM algorithm is used to estimate the parameters for the components of the MovM describing the set of angles, and 2) each θ_i angle of the sequence is encoded with a symbol S_i representing the most similar pdf of the mixture, resulting in a sequence of symbols $\overline{T} = \{S_1, S_2, ..., S_n\}$. Assuming uniform priors, S_i is computed with a maximum-a-posteriori (MAP) approach as

$$S_i = \underset{k=1,...,K}{\arg\max}\ p\left(\theta_{0,k}, m_k|\theta_i\right) = \underset{k=1,...,K}{\arg\max}\ \mathcal{V}\left(\theta_i|\theta_{0,k}, m_k\right) \tag{8}$$

where $\theta_{0,k}$ and m_k are the parameters of the kth components of the MovM.

The representation of a trajectory as a sequence of symbols is a key contribution of our work because each single value, possibly affected by noise, is substituted with a pdf. In most frequent cases the sequences have different lengths and exact matching of either angular values or statistical distributions is not appropriate, especially in surveillance, since trajectories are also affected by temporal shifts, acquisition noise, or uncertainty. Hence, a suitable similarity measure $\Omega\left(\overline{T}_i, \overline{T}_j\right)$ must be defined. The similarity between the elements of the sequences must be ascertained after the alignment and warping of the trajectories to ensure the invariance to length differences.

Having to compare symbols, we can borrow from bioinformatics the method for comparing DNA sequences in order to find the best inexact matching between them, also accounting for gaps. We propose to adopt a *global alignment* technique, specifically the well-known N-W algorithm [18]. A global alignment (over the entire sequence) is preferable over a local one, because it preserves both global and local shape characteristics.

This algorithm is very onerous in terms of computational time if the sequences are long. For this reason, *dynamic programming* is used to reduce computational complexity to $O\left(n_i \cdot n_j\right)$, where n_i and n_j are the lengths of the two sequences. Dynamic programming overcomes the problem of the recursive solution to global alignment by not comparing the same subsequences for more than one time, and by exploiting tabular representation to efficiently compute the final similarity score.

Another well known alignment algorithm is the DTW algorithm. The DTW algorithm aims to find the optimal warping path W between two sequences maximizing the warping score (i.e., minimizing the warping cost). The DTW technique has been successfully applied in heterogeneous contexts where the temporal shift between the considered sequences is nonlinear. It uses dynamic programming to compute the optimal alignment between two time series and is often adopted to obtain similarity measures or to match corresponding regions between two time-dependent sequences. Please note that DTW does not make use of gaps. Both DTW and global alignment (N-W) basically behave the same when the considered sequences do not present any time shifts. Conversely, when the sequences are characterized by different lengths, DTW tries to stretch the two sequences in order to find the optimal time warping path with the consequence of eventually adding additional matches. N-W, on the other hand, simply adds gaps to align the sequences leading to the advantage of being more susceptible to slight time series' changes by controlling the gap cost value. A comparison will be presented in the experimental results.

Whichever alignment technique is chosen, each element (a, b) of the table contains the alignment score of the symbol $S_{a,i}$ of sequence $\overline{T}_i$ with the symbol $S_{b,j}$ of sequence $\overline{T}_j$. However, we do not adopt it directly on the angular data since they can be affected by measurement noise, but on the symbols corresponding to vM pdfs. Thus, the one-to-one score between symbols is measured statistically as a function of the distance between the corresponding distributions. If the two distributions result sufficiently similar, the score should be high and positive, while if they differ significantly, the score (penalty) should be negative.

Unfortunately, the vM parameters θ_0 and m do not lie on a Euclidean space, thus the Euclidean distance in the

parameter space is not suitable. It is more effective to use the *Bhattacharyya* distance to compare two distributions p and q. We derived a closed-form expression of the Bhattacharyya distance in the case p and q are two von Mises distributions which can be written (with reference to symbols) as

$$c_B\left(S_{a,i}, S_{b,j}\right) = c_B\left(\mathcal{V}\left(\theta|\theta_{0,a}, m_a\right), \mathcal{V}\left(\theta|\theta_{0,b}, m_b\right)\right) = \sqrt{\frac{1}{I_0(m_a)I_0(m_b)}} I_0\left(\frac{\sqrt{m_a^2+m_b^2+2m_a m_b\cos\left(\theta_{0,a}-\theta_{0,b}\right)}}{2}\right) \tag{9}$$

where it holds that $0 \leq c_B\left(S_{a,i}, S_{b,j}\right) \leq 1$. The complete derivation is reported in Appendix B. Using a closed-form equation instead of an (approximated) computation of the integral has two main advantages: first, it provides an exact yet not approximated value, and second is much more efficient to compute.

The Bhattacharyya coefficient of (9) can be converted into a score to compute the global cost for aligning the two sequences. Adopting the same values proposed for DNA sequence alignment [33], we assume that two distributions are sufficiently similar (perfect match) if the coefficient is above 0.5, gaining a score of 2; conversely, the score (or penalty) for the perfect mismatch is -1. Moreover, we chose a gap penalty of -0.3. All these values have been selected empirically and left unchanged for all the experiments. We can write the general score as a linear function of c_B

$$d\left(S_{a,i}, S_{b,j}\right) = \begin{cases} 2\cdot(c_B), & \text{if} \quad c_B \geq 0.5 \\ 2\cdot(c_B - 0.5), & \text{if} \quad c_B < 0.5 \\ -0.3, & \text{if} \quad S_{a,i} \text{ or } S_{b,j} \text{ are gaps.} \end{cases} \tag{10}$$

Once the score of the best global alignment is computed (as sum of the scores in the best alignment path), it can be converted into a proper similarity measure $\Omega\left(\overline{T}_i, \overline{T}_j\right)$. Given the sequences of symbols $\overline{T}_i$ and $\overline{T}_j$, they are transformed into two aligned sequences of R symbols $\widetilde{T}_i$ and $\widetilde{T}_j$, whose similarity can be computed as

$$\Omega\left(\overline{T}_i, \overline{T}_j\right) \equiv \Omega\left(\widetilde{T}_i, \widetilde{T}_j\right) = \sum_{r=1}^{R} d\left(S_{r,i}, S_{r,j}\right). \tag{11}$$

IV. People Trajectory Shape Analysis

The previous section provides a way of modeling sets and sequences of angular features and a suitable similarity measure, potentially useful in many pattern recognition problems. In this section we apply them in trajectory shape analysis for people surveillance applications, in order to compare trajectories and detect abnormal paths.

The trajectory shape analysis approach is described in Fig. 1 and goes through two main phases. The off-line phase provides the clustering of the trajectory shapes into classes of similarity. This can be a stand-alone process (for a data mining application on surveillance data) or the training phase for a further on-line classification step of new samples. This process can be further specialized into a two-class classification problem to search for normal/abnormal trajectories in a statistical sense. A trajectory can thus be considered as "abnormal" if either it is similar to the those belonging to a class with few elements or, being different from all the others, it creates a new class.

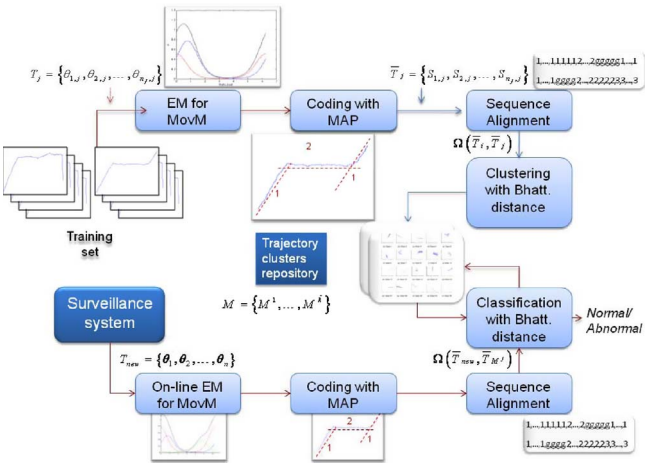


Fig. 1. Scheme of the system for trajectory shape analysis.

We represent each trajectory $T_j = \left\{\theta_{1,j}, \theta_{2,j}, \ldots, \theta_{n_j,j}\right\}$ as a sequence of n_j directions θ, defined in $[0, 2\pi)$. In the case of people surveillance the angles are computed as the tangent between two consecutive points in the ground plane. Each trajectory is modeled with a MovM distribution whose parameters are estimated through the EM. Then, the trajectory T_j is coded in a sequence of symbols $\overline{T}_j$, with each symbol $S_{a,j}$ corresponding to a single vM pdf. Eventually, trajectories are aligned and clustered with the similarity measure $\Omega\left(T_i, T_j\right) \equiv \Omega\left(\overline{T}_i, \overline{T}_j\right) \equiv \Omega\left(\widetilde{T}_i, \widetilde{T}_j\right)$.

The choice of the clustering algorithm is not critical: we only need a method capable of creating a not fixed number of clusters of different cardinality, eventually containing a single sample. This is required to achieve the maximum generality by allowing also "abnormal" trajectories to be included in the training set. Those trajectories will be clustered in small classes, due to their diversity. Our tests suggest that a satisfactory algorithm for clustering trajectory shapes is the *k-medoids* algorithm [34]. Moreover, in order to overcome the limitation of k-medoids clustering to have a fixed number k of clusters, we use its iterative version, which merges similar clusters until convergence, in order to find the "best" number $\widetilde{k}$ of clusters.

Let us suppose we have a training set $TS = \left\{T_1, \cdots, T_{N_t}\right\}$ composed of N_t trajectories and set $i = 0$ and $k(0) = N_t$. As initialization, each trajectory is chosen as prototype (medoid) of the corresponding cluster. The k-medoids algorithm iteratively assigns each trajectory T_j to the cluster C^m at the maximum similarity, i.e., given $k(i)$ clusters $C^1, \cdots, C^{k(i)}$ and the corresponding medoids $M^1, \cdots, M^{k(i)}$, $\widetilde{m} = \underset{m=1,\cdots,k(i)}{\arg\max}\, \Omega\left(T_j, T_{M^m}\right)$, where T_{M^m} is the trajectory corresponding to the medoid M^m. Once all the trajectories have been assigned to the correct cluster, the new medoid M^s for each cluster C^s is computed as that one which minimizes the intra-cluster distances, i.e., $T_{M^s} \equiv T_{\widetilde{p}} = \underset{\forall T_p \in C^s}{\arg\max} \sum_{\forall T_r \in C^s} \Omega\left(T_p, T_r\right)$. Given the initialization

step, we have as many clusters as the number of trajectories. Then, if there are two medoids with a similarity greater than a threshold Th, they are merged and the iteration starts again until all the medoids have a pairwise similarity lower than Th.

During the on-line phase, whenever a new trajectory T_{new} is collected (see Fig. 1), its statistical model is computed and compared to the cluster's medoids. Thus, the trajectory can be classified either as belonging to an existing cluster (of either normal or abnormal trajectories), or as a new trajectory which creates a new cluster (i.e., a class of trajectories never seen before).

In order to compute the statistical model for T_{new}, the EM algorithm described in Section III-A can result unsuitable for real-time classification being very time consuming. For this reason, similarly to what proposed for a mixture of Gaussians [35], we have derived an *on-line EM* algorithm for MovM.

On-line EM updating is based on the concept of *sufficient statistics*. A statistic $\psi(\theta)$ is sufficient for underlying parameter η if the conditional probability distribution of the data θ, given the statistic $\psi(\eta)$, is independent of the parameter η. Thanks to the Fisher-Neyman's factorization theorem [36], the likelihood function $L_\eta(\theta)$ of θ can be factorized in two components, one independent of the parameters η and the other dependent on them only through the sufficient statistics $\psi(\theta)$: $L_\eta(\theta) = h(\theta) g_\eta(\psi(\theta))$. It has been shown [37] that in the case of distributions of the exponential family (such as Gaussian and von Mises) the factorization theorem can be written as $p(\theta|\eta) = h(\theta)\, g(\eta) \exp\left\{\eta^T \psi(\theta)\right\}$. Considering a von Mises distribution and a set $\boldsymbol{\theta}$ of i.i.d. angles, we can decompose the expression of the distribution $p(\boldsymbol{\theta}|\theta_0, m)$ as follows:

$$\begin{aligned}
&\prod_{i=1}^{n} \frac{1}{2\pi I_0(m)} \exp\{m\cos(\theta_i - \theta_0)\} \\
&= \frac{1}{(2\pi I_0(m))^n} \exp\left\{ m \sum_{i=1}^{n} \cos(\theta_i - \theta_0) \right\} = \\
&= \frac{1}{(2\pi I_0(m))^n} \exp\left\{ \begin{bmatrix} m\cos\theta_0 \\ m\sin\theta_0 \end{bmatrix}^T \cdot \begin{bmatrix} \sum_{i=1}^{n} \cos\theta_i \\ \sum_{i=1}^{n} \sin\theta_i \end{bmatrix} \right\}. \qquad (12)
\end{aligned}$$

Thus, the sufficient statistics for a single von Mises distribution are

$$\psi(\boldsymbol{\theta}) = \begin{bmatrix} \sum_{i=1}^{n} \cos\theta_i \\ \sum_{i=1}^{n} \sin\theta_i \end{bmatrix}.$$

In the case of a mixture of distributions belonging to the exponential family, the on-line updating of the mixture's parameters can be obtained by updating the sufficient statistics (s.s.) of the mixture, computed as $\psi_M(\boldsymbol{\theta}) = \sum_{k=1}^{K} \gamma_k \psi_k(\boldsymbol{\theta})$, where $\psi_k(\boldsymbol{\theta})$ are the s.s. for the kth single distribution of the mixture. The updating process (having observed up to the sample $(i-1)$th) can be obtained as

$$\psi_k^i(\boldsymbol{\theta}) = \alpha(i)\gamma_k \psi_k(\theta_i) + (1 - \alpha(i))\, \psi_k^{i-1}(\boldsymbol{\theta}) \qquad (13)$$

where $\psi_k(\theta_i) = \begin{bmatrix} \cos\theta_i \\ \sin\theta_i \end{bmatrix}$. A comprehensive discussion on the value of the updating parameter $\alpha(i)$ for the pdf of the exponential family is given in [38].

Using the on-line approach leads to a computational time which allows real-time computation of MovM parameters and thus a straightforward representation of T_{new} as the symbol sequence $\overline{T}_{\text{new}}$. We have observed that on a set of 200 real trajectories having an average length of 100 points, the on-line approach is approximately ten times faster than the off-line version (in the order of one tenth of second with respect to one second), and leads to a satisfying learning of mixture parameters.

Thus, given the set $M = \left\{M^1, ..., M^{\widetilde{k}}\right\}$ of current medoids, $\overline{T}_{\text{new}}$ is compared with each medoid to find the most similar one, using the similarity measure Ω: $\widetilde{j} = \arg\max_{j=1,...,\widetilde{k}} \Omega(T_{M^j}, T_{\text{new}})$.

New trajectories can be classified as normal/abnormal depending on the cardinality of the most similar cluster $C^{\widetilde{j}}$. In the surveillance application the behaviors of people can change along time; for instance, if an obstacle is positioned in the scene, people can find alternative trajectories to avoid it, initially resulting in an "abnormal" behavior. However, if this situation persists, the behavior must become normal since it is frequently observed. This can be taken into account by using a "learn-and-predict" paradigm, by updating continuously the clusters.

More operatively, being N the current number of trajectories and T_{new} the new trajectory, if the maximum similarity with all the $\widetilde{k}$ cluster medoids $\Omega_{\max} = \Omega\left(T_{M^{\widetilde{j}}}, T_{\text{new}}\right)$ is lower than a threshold Th_{sim}, a new cluster $C^{\widetilde{k}+1}$ should be created with T_{new}. The normalized cardinality $\mathcal{C}$ of each cluster (which could be considered as the prior for a classification normal/abnormal) is continuously updated to take into account the increased number of samples assigned to the cluster. Conversely, if the new trajectory is similar enough to one of the current medoids, the trajectory is assigned to the corresponding cluster $C^{\widetilde{j}}$.

The medoids are continuously updated as reported above to choose the most representative of the class. Finally, to avoid old and rare trajectories affecting our model, the clusters with small cardinality and with no new trajectories assigned for a fixed-length time window are dropped.

V. Von Mises Versus Wrapped Gaussian Distributions

Section IV described our complete approach for people trajectory shape analysis. Before presenting tests and experiments in Section VI, we motivate the choices we made by providing comparison with possible variants.

The von Mises distribution is not the only pdf defined in circular statistics. Mardia [26] proposed a set of distributions and corresponding tools for handling periodic variables. Among them, the wrapped Gaussian (WG) is one of the most used since it derives from the Gaussian and can thus be easily

manipulated. Recently, a variation named approximated wrapped Gaussian (AWG) has been proposed [29]. The next subsections will briefly introduce these distributions and the corresponding mixtures, namely mixture of wrapped Gaussians (MoWG) and mixture of approximated wrapped Gaussian (MoAWG). In order to provide a fair comparison, the EM for both pdfs have been derived and reported here.

A. The Wrapped Gaussian Distribution

In order to let a pdf $p(x)$ of a linear variable x defined on a line be used for a periodic variable θ, the pdf $p(x)$ can be "wrapped" around the circumference of a circle of unit radius. The wrapped pdf $p^W(\theta)$ of the wrapped variable $\theta = x \bmod 2\pi$ defined in the interval $(-\pi, \pi]$ can be generally obtained tiling different pdfs shifted by multiple of 2π. In the case $p(x)$ is a univariate Gaussian distribution, the wrapped Gaussian (WG) can be written as

$$\begin{aligned}\mathcal{N}^W(\theta|\theta_0, \sigma) &= \sum_{w=-\infty}^{+\infty} \mathcal{N}(\theta - w2\pi|\theta_0, \sigma) \\ &= \sum_{w=-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\theta - w2\pi - \theta_0)^2}{2\sigma^2}}. \end{aligned} \tag{14}$$

$\mathcal{N}^W(\theta|\theta_0, \sigma)$ is unimodal with a single local maximum and symmetric about θ_0. The WG distribution can be approximated by a linear Gaussian distribution at small variances ($\sigma \leq 1$) and by uniform distribution at large variances ($\sigma \geq 2\pi$). Moreover, it can be shown that summation over ± 2 provides a sufficient approximation even for large variances [3].

A mixture model of WG (MoWG) on a periodic variable θ can be derived as follows:

$$\begin{aligned}MoWG(\theta|\boldsymbol{\pi}, \boldsymbol{\theta_0}, \boldsymbol{\sigma}) &= \sum_{k=1}^{K} \pi_k \mathcal{N}^W\left(\theta|\theta_{0,k}, \sigma_k\right) \\ &= \sum_{k=1}^{K} \sum_{w=-\infty}^{+\infty} \pi_k \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{\left(\theta - w2\pi - \theta_{0,k}\right)^2}{2\sigma_k^2}}. \end{aligned} \tag{15}$$

The mixture's parameters of the pdf reported in (15) can be estimated through EM algorithm. Unfortunately, the treatment is not straightforward due to the tiling of wrapped contributions in (15). Similarly to the EM proposed for MovM, we derived a EM algorithm for MoWG; the basic steps can be summarized as

1) *Expectation Step:*

$$\gamma_{ik} = \frac{\pi_k \mathcal{N}^W\left(\theta_i|\theta_{0,k}, \sigma_k\right)}{\sum_{j=1}^{K} \pi_j \mathcal{N}^W\left(\theta_i|\theta_{0,j}, \sigma_j\right)}. \tag{16}$$

2) *Maximization Step:*

$$\pi_k = \frac{1}{n}\sum_{i=1}^{n} \gamma_{ik} \qquad \theta_{0,k} = \frac{\sum_{i=1}^{n} \gamma_{ik} \beta_{ik}^{(\theta_0)}}{\sum_{i=1}^{n} \gamma_{ik} \mathcal{N}^W\left(\theta_i|\theta_{0,k}, \sigma_k\right)} \tag{17}$$

$$\sigma_k^2 = \frac{\sum_{i=1}^{n} \gamma_{ik} \frac{\beta_{ik}^{(\sigma^2)}}{\mathcal{N}^W\left(\theta_i|\theta_{0,k}, \sigma_k\right)}}{\sum_{i=1}^{n} \gamma_{ik}} \tag{18}$$

where, similarly to [3], we define the following quantities:

$$\beta_{ik}^{(\theta_0)} = \sum_{w=-\infty}^{\infty} \mathcal{N}\left(\theta_i - w2\pi|\theta_{0,k}, \sigma_k\right) \cdot (\theta_i - w2\pi) \tag{19}$$

$$\beta_{ik}^{(\sigma^2)} = \sum_{w=-\infty}^{\infty} \mathcal{N}\left(\theta_i - w2\pi|\theta_{0,k}, \sigma_k\right) \cdot \left(\theta_i - w2\pi - \theta_{0,k}\right)^2. \tag{20}$$

The full derivation of the algorithm is provided in Appendix C. Although the steps for parameters' estimation are similar for both mixtures, the EM for MoWG is computationally heavier than MovM due to the presence in (16) and (17) of the terms $\mathcal{N}^W$ and β which include additional infinite summations.

B. The Approximated Wrapped Gaussian Distribution

In order to work in "semi-directional" situations in which some dimensions correspond to circular variables and the others to linear variables, Bahlmann [29] proposed an approximation of the wrapped Gaussian by only one, but the most meaningful wrap of it. Bahlmann demonstrated that, in the presence of sufficiently small variances ($\sigma \lesssim 1$), the sum of wrapped distributions can be approximated with a single periodic distribution since the overlap of neighboring Gaussian wraps is negligible.

The proposed AWG results in [29] as

$$\mathcal{N}^{AW}(\theta|\theta_0, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{((\theta - \theta_0) \bmod 2\pi)^2}{2\sigma^2}} \tag{21}$$

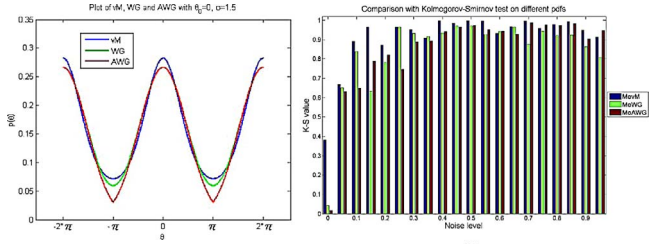
where the operation "mod" represents the remainder of the division and the resulting angle must be in the interval $(-\pi, \pi]$ (as stated at the beginning of Section V-A).

The mixture of AWG (MoAWG) is, as usual, obtained as a weighted sum of AWG pdfs

$$\begin{aligned}MoAWG(\theta|\boldsymbol{\pi}, \boldsymbol{\theta_0}, \boldsymbol{\sigma}) &= \sum_{k=1}^{K} \pi_k \mathcal{N}^{AW}\left(\theta|\theta_{0,k}, \sigma_k\right) \\ &= \sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{\left(\left(\theta - \theta_{0,k}\right) \bmod 2\pi\right)^2}{2\sigma_k^2}}. \end{aligned} \tag{22}$$

The formulation of the EM algorithm for MoAWG is not simple since the mod operator is not derivable. Thanks to the use of a mathematical artifice, we derived the following steps (using $\mathcal{N}^{AW}$ instead of $\mathcal{N}^W$).

1) *Expectation Step:* Same as MoWG (16).



(a) (b)

Fig. 2. (a) Plots of different circular pdfs, with $\sigma = 1.5$ (corresponding to $m = 0.69$). (b) K-S values for MovM, MoWG, and MoAWG with increasing noise level.

2) *Maximization Step:* The mixture weights π_k are computed as in the case of MoWG (17), while (17) becomes

$$\theta_{0,k} = \operatorname{atan}\left(\frac{\sum_{i=1}^{n}\gamma_{ik}\sin\theta_i}{\sum_{i=1}^{n}\gamma_{ik}\cos\theta_i}\right) \sigma_k^2 = \frac{\sum_{i=1}^{n}\gamma_{ik}\left(\left(\theta_i-\theta_{0,k}\right) \mod 2\pi\right)^2}{\sum_{i=1}^{n}\gamma_{ik}}. \tag{23}$$

The full derivation is reported in Appendix D.

In the case of MoAWG, the computational load is lower than in the other cases since the M step requires neither the inversion of $A(m)$ nor the additional summation of the term $\mathcal{N}^W$ mentioned above.

C. Distribution Comparison

Fig. 2(a) shows a comparison of the shapes of the vM, WG, and AWG pdfs. Even though the distributions are very similar, it is evident that the AWG approximation becomes imprecise when the variance increases.

Despite the similar shape, vM seems to be a more suitable pdf than WG and AWG for angular value, although it is more complex mathematically. To give a numerical evidence of this statement, we use a statistical non-parametric goodness-of-fit test, and, in particular, the Kolmogorov-Smirnov (K-S) test which has the remarkable characteristics to not depend on the data origin. To test the goodness of the fit for each pdf, we generated a set of 200 trajectories obtained by a sequence of three angles by iteratively increasing the level of the added noise. Thus, these trajectories should be well modeled with a mixture of 3 pdfs. On each generated trajectory, the EM algorithm for the analyzed pdf is applied and the K-S test is performed on the resulting mixture. The test is repeated five times for each iteration and resulting K-S values are averaged among the five runs. The plot in Fig. 2(b) shows the comparison with an increased level of noise which demonstrates the better performance of vM with respect to the other distributions. The average overall K-S values of vM, WG, and AWG are 91.38%, 84.03%, and 84.58%, respectively. Hence, the AWG is a good approximation of the WG also in case of a mixture. However, MovM results more precise of about 7% than the other two mixtures. For this reason, we adopt MovM, despite its complexity; moreover, the on-line EM algorithm with sufficient statistics presented in Section IV leads to a considerable improvement of the speed.

VI. Experimental Results

We evaluated our approach on different tests carried out on both synthetic and real data for surveillance applications and other publicly-available datasets for people trajectory analysis and handwritten character recognition. Our own datasets have been made available for download in our website http://imagelab.ing.unimore.it/files2/trajectory_dataset. The goal is to test the robustness of this novel approach with respect to other possible solutions.

In all the experiments we consider a mixture of $K = 5$ components (for all the described circular pdfs) since we address a typical open-space scenario where people walk through the scene toward a target (e.g., a gate) with a limited number of changes in their direction. In different applications we could increase the number of the mixture's components accordingly.

A. Experiments on Synthetic Data

In order to analyze synthetic trajectories, we developed a trajectory generator in MATLAB, which allows the user to select graphically the sequence of angles composing the class of trajectories; then, it creates a set of trajectories with added Gaussian noise to both the single angles and their occurrences. Therefore, a set of similar trajectories can be created with trajectories of different lengths but similar directions.

The use of synthetic data enables the collection and evaluation of an adequate amount of data together with their automatic ground truths. Since the system has been conceived bearing the surveillance framework in mind, we used small sets of similar examples affected by strong noise and/or global and local spatial shifts. In fact, in surveillance the detected trajectories for each class are normally few and heavily affected by noise, due to the dynamism of the scene and to the unpredictability of people's behaviors.

Fig. 3 shows some examples of the classes of trajectories used in our tests. The synthetic tests of 22 classes comprehend 5210 trajectories with an average of about 81 points each, but with a high variance (ranging from 50 to 100 points, approximately). This simulates the case of people staying in the surveilled area for about 10–20 s (with five samples per second). Moreover, the approach has another operative parameter, i.e., the threshold on the cardinality of a cluster to be considered representative of abnormal trajectories. In our experiments, we fixed it to 20, i.e., about the 10% of the average cardinality of other clusters. The central graph of Fig. 3 shows several trajectories of each class superimposed.

Table I summarizes the experimental benchmark used in our tests. The first three columns contain the challenge of the test, the mean and the standard deviation of the added Gaussian noise. In the subsequent columns the classes' cardinalities are reported for the training and testing sets. Table II compares our approach versus a similar procedure with different circular pdfs, namely the MoWG or MoAWG instead of MovM, and different alignment method. It is worth clarifying that the comparison is performed by changing one component of the system at a time: for instance, MoWG/MoAWG cases mean that the alignment (N-W), the MAP and the measure (Bhattacharyya) are the same and that only the used circular

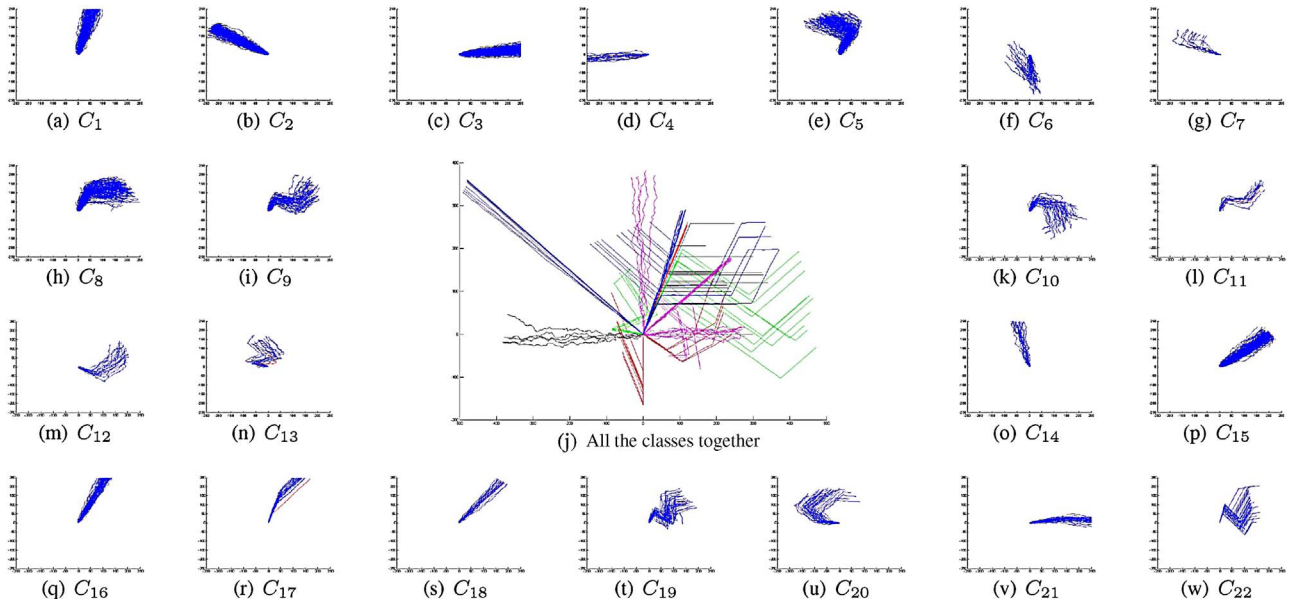


Fig. 3. Examples of the classes of the used trajectories. (a) C_1. (b) C_2. (c) C_3. (d) C_4. (e) C_5. (f) C_6. (g) C_7. (h) C_8. (i) C_9. (j) All the classes together. (k) C_{10}. (l) C_{11}. (m) C_{12}. (n) C_{13}. (o) C_{14}. (p) C_{15}. (q) C_{16}. (r) C_{17}. (s) C_{18}. (t) C_{19}. (u) C_{20}. (v) C_{21}. (w) C_{22}.

TABLE I

SUMMARY OF THE EXPERIMENTAL BENCHMARK

Type of Test	μ	σ	Training Classes		Testing Classes	
			Classes	Total	Classes	Total
Test 1 (periodicity)	0.1	0.5	$C_3(230)$, $C_4^*(20)$	250	$C_3(130)$, $C_4^*(20)$	150
Test 2 (noise)	0.2	0.5	$C_1(200)$, $C_{15}(200)$	400	$C_{14}^*(20)$, $C_{15}(250)$	270
Test 3 (mono-modal)	0.1	0.5	$C_1(215)$, $C_2(215)$, $C_3^*(20)$	450	$C_1(200)$, $C_2(200)$, $C_3^*(25)$, $C_4^*(25)$	450
Test 4 (multi-modal)	0.1	0.5	$C_1(215)$, $C_2(215)$, $C_5(215)$, $C_6^*(35)$	680	$C_1(250)$, $C_5(250)$, $C_6^*(15)$, $C_7^*(15)$	530
Test 5 (sequence)	0.1	0.5	$C_1(215)$, $C_8(215)$, $C_9(215)$, $C_{10}^*(40)$, $C_{11}^*(45)$	730	$C_8(300)$, $C_2^*(20)$, $C_9(20)$, $C_{11}^*(20)$, $C_{12}^*(20)$, $C_{13}^*(20)$	400
Test 6 (learn.norm.)	0.1	0.5	$C_1(150)$, $C_7^*(30)$	180	$C_{13}^*(20)$, $C_7(180)$	200
Test 7 (sim.angles)	0.1	0.3	$C_{16}(100)$, $C_{18}^*(30)$	130	$C_{16}(40)$, $C_{17}(40)$, $C_{18}^*(10)$	90
Test 8 (mixed)	0.1	0.5	$C_{19}(50)$, $C_{20}(50)$, $C_{21}(50)$, $C_{22}^*(4)$	154	$C_{19}(50)$, $C_{20}(50)$, $C_{22}(46)$	146

TABLE II

SUMMARY OF THE EXPERIMENTAL RESULTS WITH DIFFERENT METHODS

	MovM		MoWG		MoAWG		DTW		HMM	
	α_{OV}	α_{NA}	α_{OV}	α_{NA}	α_{OV}	α_{NA}	α_{OV}	α_{NA}	α_{OV}	α_{NA}
Test 1	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	74.7%	84.0%
Test 2	98.9%	98.9%	90.7%	92.6%	90.7%	92.6%	77.8%	94.4%	81.5%	94.1%
Test 3	98.7%	99.6%	88.9%	90.0%	88.0%	90.0%	77.8%	91.1%	92.2%	91.6%
Test 4	97.9%	99.4%	95.3%	97.2%	95.7%	97.5%	83.6%	87.4%	92.1%	99.4%
Test 5	98.3%	100.0%	97.0%	100.0%	97.0%	100.0%	100.0%	100.0%	91.7%	86.3%
Test 6	97.5%	99.0%	93.5%	97.5%	91.0%	98.5%	89.0%	92.0%	94.0%	99.0%
Test 7	96.7%	100.0%	94.4%	96.7%	97.8%	100.0%	96.7%	100.0%	80.0%	87.8%
Test 8	100.0%	100.0%	90.4%	93.2%	87.7%	91.1%	87.0%	91.1%	79.5%	89.7%
Average	98.5%	99.6%	93.8%	95.9%	93.5%	96.2%	89.0%	94.5%	95.7%	91.5%

statistics is changed. Additionally, the HMM-based method with the similarity measure proposed in [24] has been also evaluated. This approach was not proposed for shape analysis but for spatial analysis; thus, to perform a fair comparison it has been adapted by using angles as data, instead of (x, y) coordinates.

Both the *overall classification accuracy* α_{OV} (i.e., the ability to assign the new trajectory to the correct cluster, or in other words, the sum of the diagonal values in the confusion matrix) and the *normal/abnormal accuracy* α_{NA} (i.e., the ability to correctly classify the trajectory as normal or abnormal depending on the cardinality of the specific cluster) are evaluated. Error computation in the case of synthetic data set is based on the ground-truth data available from the trajectory generator.

Test 1 is meant to demonstrate that MovM accounts for the *periodicity* of the features. The test contains class C_3 [Fig. 3(c)] with angles close to 0 deg; the trajectories in this class contain a medium-high level of noise that often makes the directions go from negative to positive around zero: this variance around the origin makes the model with a Gaussian unsuitable. Instead, any circular pdf (MovM, MoWG, or MoAWG) obtains 100% of accuracy. Test 2 analyzes the *robustness to noise* since the considered trajectories (class C_1, C_{14}, and C_{15}) have a high level of noise and have a similar main direction that can make classification more challenging. MovM achieves an accuracy (98.9%) higher than MoWG and MoAWG. In addition, being these trajectories of different lengths, the global alignment outperforms DTW (with an accuracy of 77.8% only).

Tests 3 and 4 give similar results of test 2 accounting for robustness to mono-modality and multi-modality of the trajectories, respectively. In test 5 there are several classes with similar angles but in different order: for instance, trajectories of class C_{11} (abnormal) contain the same angles as those of class C_8 (normal) but with a different sequence; the same holds for class C_{12} and class C_9. Again, MovM is more suitable than the other circular pdfs, but this is the only case where DTW-based approach results better than global alignment-based, thanks to its ability to handle non-linear temporal shifts.

Test 6 evaluates the capacity of our approach to "learn normality," in the sense that an initially-abnormal class of trajectory (class C_7) becomes normal since its cardinality grows in the testing phase. Our approach is still the most accurate. Test 7 tries to classify classes C_{16}, C_{17} and C_{18} in different cluster trajectories with very *similar angles*: our approach achieves a good accuracy, even though worse than in the other tests. In this case the choice between global alignment and DTW is irrelevant.

Finally, test 8 is a more complex case accounting for clusters with different cardinalities and shapes: MovM is the only approach capable to reach an accuracy of 100%, even in the case of class C_{22} containing initially only four samples. This demonstrates the capability of a statistical approach to work properly also when few samples per class are available.

In general, MoWG and MoAWG obtained very similar performances, apart from test 7 in which MoAWG obtained the best overall accuracy. This is due to the fact that MoAWG tends to be more discriminative in fitting the data, thus assigning different, but close, angles to different mixture's components. Conversely, EM for MovM will merge two close vM components into one, resulting slightly less discriminant in the case of very similar angles.

TABLE III
COMPARISON WITH STATE-OF-THE-ART ONE-NEAREST-NEIGHBOR (1NN) APPROACHES

	α_{OV} (MovM)	α_{OV} (Eucl/1NN)	α_{OV} (DTW/1NN)
Test 1R	100.0%	86.7%	100.0%
Test 2R	94.5%	100.0%	100.0%
Test 3R	98.7%	100.0%	100.0%
Test 4R	98.3%	96.5%	100.0%
Test 5R	98.5%	94.4%	91.8%
Test 6R	98.9%	96.7%	99.4%
Test 7R	96.0%	80.0%	80.0%
Test 8R	100.0%	45.2%	67.1%
Average	98.1%	87.4%	92.3%

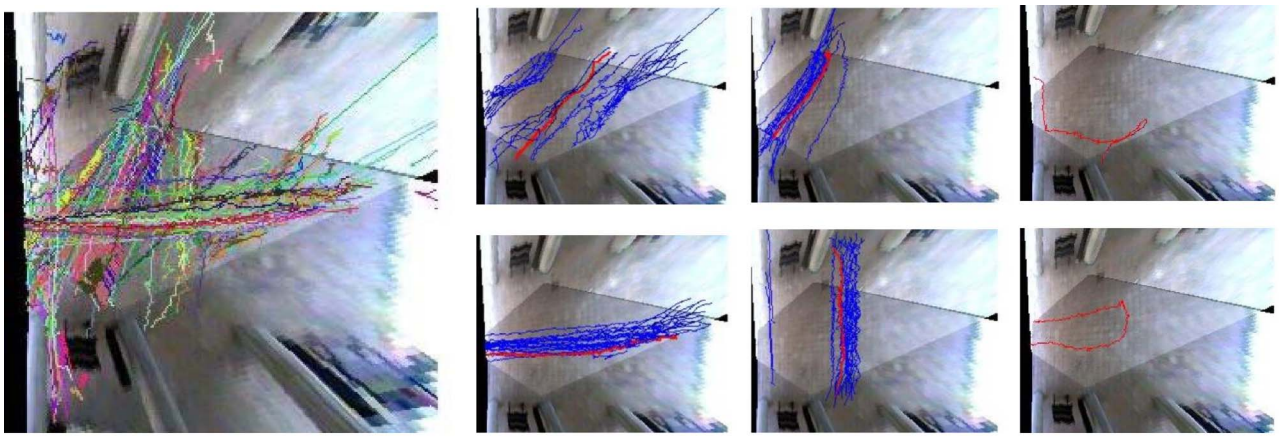
In all the examples, the approach based on HMMs presents the lowest classification performance. This is mainly due to the over-fitting problem: when few data are available, the HMM training stage fails to correctly estimate all the parameters. HMMs work well with classification tasks with defined classes of data, but, in unsupervised classification, they require a large amount of data that are not always available in real scenarios. Conversely, our approach is not strongly affected by the number of data available because the number of parameters to estimate is significantly lower; it can thus be applied profitably in many different situations.

In these synthetic tests we also evaluate classical approaches with point-to-point measures without the need of a statistical model of the trajectory, using either a point-to-point Euclidean distance (properly adapted to deal with angles) or a DTW-based distance measure. Thanks to their simplicity, these methods are often associated with brute-force classifiers, such the one-nearest-neighbor (1NN) classifier which can be used to both cluster and classify trajectories [12]. These measures have been suggested to be state-of-the-art approaches when the amount of data is significant, as, for instance, in string matching, time series analysis or database applications.

These approaches are actually effective if no continuous updating of the classes must be performed. As a proof of concept, Table III compares Euclidean/1NN and DTW/1NN with our approach. To produce a fair comparison, tests are provided on reduced datasets (indicated with "R" in Table III) without testing elements of those classes which are not present in the training set in order to allow also Euclidean/1NN and DTW/1NN to work properly. In our approach we disabled the class updating process described in Section IV. In most of the cases, DTW/1NN performs as well as (and often even better than) our approach, but fails when the training set contains classes with few trajectories (as is the case of test 8). Nevertheless, the average accuracy is 98.1% for our approach, compared to 87.4% and 92.3% for Eucl/1NN and DTW/1NN.

B. Experiments on Real Data

The considerations we have made for synthetic data are still valid and even more evident in the case of real surveillance



(a) (b) (c)

Fig. 4. Examples of real data classes representing a dataset of 980 trajectories with lengths from 50 to 100 points. The clusters created in the training set are 18 and becomes 22 slightly different classes during the testing phase. Some examples of normal and abnormal classes are reported in (b) and (c), respectively. (a) Entire dataset. (b) Normal classes. (c) Abnormal classes.

TABLE IV
ACCURACY ON REAL DATA

	Avg Len	# Train.	# Test.	MovM		MoWG		MoAWG		DTW		HMM	
Test Re1	66	550	430	α_{OV}	α_{NA}	α_{OV}	α_{NA}	α_{OV}	α_{NA}	α_{OV}	α_{NA}	α_{OV}	α_{NA}
(real)				94.9%	100.0%	89.7%	100.0%	89.7%	100.0%	79.5%	100.0%	79.4%	66.0%
Test Re2	100	232	248	α_{OV} (MovM)		α_{OV} (Eucl/1NN)				α_{OV} (DTW/1NN)			
(real)				87.1%		52.8%				57.7%			

data. Therefore, the use of this complex approach based on circular statistics becomes particularly suitable.

Examples of such real trajectories have been reported in Fig. 4,[1] where normal and abnormal classes are also reported. Real data are extracted from the scene using a multi-camera system called homography and epipolar-based consistent labeling (HECOL) [39]. This system is capable of detecting and tracking multiple moving people by using multiple cameras with partially-overlapped fields of view; it exploits ground-plane homography and epipolar geometry for assigning a consistent label to different instances of the same person seen from different cameras/views. Having multiple cameras allows the collection of longer trajectories and the partially-overlapped views improve both the robustness to occlusions and the tracking accuracy without any perspective distortion. Directions are computed with an average sampling rate of ten points per second. Thus, we applied a running average filter of fixed size to smooth the segmentation errors and discretization effects on the direction computation. Indeed, more sophisticated approaches could be used to compute the directions. However, the experimental results (Table IV) show that the non-idealities introduced by this approximation are filtered out by the employed statistical framework. All circular statistics are well suitable for the problem but MovM outperforms the other methods and the use of gaps in the global alignment brings to an improvement in accuracy of about 15% with respect to DTW. Moreover, the second experiment (test Re2) considers a reduced set of trajectories (as did previously for the synthetic tests) in order to disable the class updating and compare with 1NN approaches. The performances confirm the good results achieved in the case of synthetic data and demonstrate the goodness of adopting a statistical model and its broad range of applicability.

[1]The real trajectories are collected in an unsupervised way on a walking area of our university campus. Most of these videos are available in the Video Surveillance Online Repository (http://www.openvisor.org).

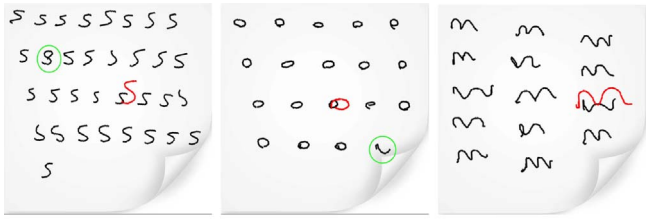
C. Additional Tests

We want to demonstrate its extendibility by testing it on a real scenario not developed by us. As a natural extension of the tests, we downloaded a publicly-available dataset from Edinburgh Informatics Forum Pedestrian Database (http://homepages.inf.ed.ac.uk/rbf/FORUMTRACKING/). This dataset contains several days of people trajectories taken from a bird-eye view camera. The results of our classification on 453 trajectories (the whole 12 September set and 200 trajectories of the 24 August set) are shown in Fig. 5(a), where two abnormal trajectories are also highlighted with thicker lines. Fig. 5(b) and (c) shows two separate clusters to better highlight that our approach also considers the direction with which a path is taken, resulting in two different classes (which can also be one normal and the other abnormal).

Finally, as a further proof of the broader range of applicability of the proposed approach, we have also tested it on a completely different dataset which contains handwritten characters developed by Universitat Jaume I, Castellón, Spain, and available on the University of California Irvine Machine Learning Repository (http://archive.ics.uci.edu/ml/).

(a) (b) (c)

Fig. 5. Clustering results on Edinburgh surveillance dataset. (a) Classification result on the dataset with bolded trajectories representing abnormal paths. (b) and (c) Two example clusters (note that the direction of movement is opposite).



(a) (b) (c)

Fig. 6. Examples of classes in UCI-UPEN dataset test. Green circles indicate erroneous classifications. Red indicates the medoid of the cluster. (a) Char "S." (b) Char "o." (c) Char "n."

This dataset consists of the characters from "a" to "z," from "A" to "Z" and from "0" to "9" written for two times by 11 different writers with a special device (UPEN). We have tested our approach by using the leaving-one-writer-out test, where ten writers are used for training and the remaining writer for testing. Some examples of correct classification can be observed in Fig. 6, where classification errors are circled in green. Red indicates the medoid of the cluster. Many works have been reported in the literature on character recognition using this dataset. One of the state-of-the-art approaches [40], which achieved 90% of accuracy, exploits a neural network classifier specifically tuned on this dataset. We tested our approach without changing any parameter. Nevertheless, our approach is capable to obtain a very promising classification accuracy of 91.3%. This result proves the suitability of our approach for shape recognition which could also be used in other contexts.

VII. Conclusion

This paper presented a deep discussion on the exploitation of circular statistics for modeling sets and sequences of angles, including its numerical solution, and its usage for building a complete system for trajectory shape analysis in people surveillance. We proved our approach to be very accurate in classifying normal and abnormal trajectories by testing it in specific challenging situations.

The use of a MovM distribution has been motivated by comparing it with other circular pdfs, namely a MoWG and a MoAWG. MovM proves to be more accurate, even though MoAWG tends to be faster and could be used for fast implementations.

However, computational complexity of EM on MovM is limited to the training phase, which can be performed off-line. Conversely, in order to meet real-time constraints, in the testing phase we derived an on-line version of the EM algorithm for MovM which does not introduce errors in the classification.

The defined EM algorithms for circular statistics, the Bhattacharyya distance for vM pdfs and the global alignment for sequence of directional data can be potentially exploited in many other contexts, such as handwritten character recognition, texture analysis or action analysis.

Appendix A
Derivation of the EM for a Mixture of von Mises (MovM)

Given the set $\boldsymbol{\theta} = \{\theta_1, \cdots, \theta_n\}$ of i.i.d. observations and the set of corresponding latent points $\boldsymbol{Z}$ (introduced with a 1-of-K variable z_i), we can write the likelihood of the *complete data set* $\{\boldsymbol{\theta}, \boldsymbol{Z}\}$ for the MovM of (3) as follows:

$$p\left(\boldsymbol{\theta}, \boldsymbol{Z}|\boldsymbol{\pi}, \boldsymbol{\theta_0}, \boldsymbol{m}\right) = \prod_{i=1}^{n}\prod_{k=1}^{K} \pi_k^{z_{ik}} \mathcal{V}\left(\theta_i|\theta_{0,k}, m_k\right)^{z_{ik}} \tag{24}$$

where z_{ik} denotes the kth component of z_i and the set of parameters to be estimated through EM is $\{\boldsymbol{\pi}, \boldsymbol{\theta_0}, \boldsymbol{m}\}$, where $\boldsymbol{\pi}$ represents the weights, $\boldsymbol{\theta_0}$ the mean values, and $\boldsymbol{m}$ the precisions of the mixture's components. Thus, the expected value of the log likelihood for the complete data set $\mathbb{E}_{\boldsymbol{Z}} \equiv \mathbb{E}_{\boldsymbol{Z}}\left[\ln p\left(\boldsymbol{\theta}, \boldsymbol{Z}|\boldsymbol{\pi}, \boldsymbol{\theta_0}, \boldsymbol{m}\right)\right]$ results to be

$$\begin{aligned}\mathbb{E}_{\boldsymbol{Z}} &= \sum_{i=1}^{n}\sum_{k=1}^{K} \gamma_{ik}\left\{\ln \pi_k + \ln \mathcal{V}\left(\theta_i|\theta_{0,k}, m_k\right)\right\} \\ &= \sum_{i=1}^{n}\sum_{k=1}^{K} \gamma_{ik}\left\{\ln \pi_k - \ln 2\pi I_0\left(m_k\right) + m_k \cos\left(\theta_i - \theta_{0,k}\right)\right\}\end{aligned} \tag{25}$$

where $\gamma_{ik} \equiv \mathbb{E}\left[z_{ik}\right]$ is the responsibility of component k for data point θ_i, and can be estimated as reported in Section III-A, (4).

The M step maximizes the log likelihood of (25) with respect to each of the parameters $\{\boldsymbol{\pi}, \boldsymbol{\theta_0}, \boldsymbol{m}\}$. Computing the derivative of (25) with respect to π_k by making use of a Lagrange multiplier, we can simply obtain (5). Deriving (25) on $\theta_{0,k}$ and setting the result equal to zero, leads to the following equation:

$$\sum_{i=1}^{n} \gamma_{ik} m_k \sin\left(\theta_i - \theta_{0,k}\right) = 0 \tag{26}$$

which easily can lead to (6) of Section III-A.

Maximization with respect to m_k is not trivial due to the presence of the Bessel function

$$\frac{\partial \mathbb{E}_{\boldsymbol{Z}}}{\partial m_k} = \frac{I_1\left(m_k\right)}{I_0\left(m_k\right)} \sum_{i=1}^{n} \gamma_{ik} + \sum_{i=1}^{n} \gamma_{ik} \cos\left(\theta_i - \theta_{0,k}\right) = 0. \tag{27}$$

Equation (27) can be solved in m_k by introducing, similarly to Section III-A, a function $A\left(m_k\right)$ and obtaining (7).

APPENDIX B
DERIVATION OF THE CLOSED-FORM EQUATION OF BHATTACHARYYA DISTANCE FOR TWO VON MISES DISTRIBUTIONS

Let $p(\theta)$ and $q(\theta)$ be two distributions on a variable θ, then we can compute the quantity, called *Bhattacharyya coefficient*

$$c_B(p,q) = \int_0^{2\pi} \sqrt{p(\theta)\, q(\theta)} d\theta. \tag{28}$$

The Bhattacharyya distance can be computed by either $\widetilde{B} = -\ln c_B$ or $B = \sqrt{1 - c_B}$. Since the first expression is not actually a distance (hence called *similarity* more properly), i.e., does not satisfy the triangular inequality, we used the second expression.

By defining two von Mises distributions $p(\theta)$ and $q(\theta)$ as

$$\begin{aligned} p(\theta) &= p\left(\theta|\theta_{0,1}, m_1\right) = \frac{1}{2\pi I_0(m_1)} e^{m_1 \cos(\theta-\theta_{0,1})} \\ q(\theta) &= q\left(\theta|\theta_{0,2}, m_2\right) = \frac{1}{2\pi I_0(m_2)} e^{m_2 \cos(\theta-\theta_{0,2})} \end{aligned} \tag{29}$$

and substituting these two expressions in the Bhattacharyya distance and then isolating the terms that do not depend on θ and defining $\Lambda(m_1, m_2) = \sqrt{\frac{1}{I_0(m_1) I_0(m_2)}}$ and $\xi = \theta - \theta_{0,1}$, we obtain

$$B(p,q) = \sqrt{1 - \frac{1}{2\pi}\Lambda(m_1, m_2) \int_0^{2\pi} e^{\frac{m_1 \cos(\xi) + m_2 \cos\left(\xi + \theta_{0,1} - \theta_{0,2}\right)}{2}} d\xi}. \tag{30}$$

Note that we have not changed the integral's limits because we are assuming that the variable θ is periodic with period 2π and the same applies for variable ξ. Assuming $\Delta\theta = \theta_{0,1} - \theta_{0,2}$ and by rearranging the terms at numerator, we can rewrite (30) as follows:

$$B(p,q) = \sqrt{1 - \Lambda(m_1, m_2) \frac{1}{2\pi} \int_0^{2\pi} e^{\alpha \cos\xi + \beta \sin\xi} d\xi} \tag{31}$$

where $\alpha = \frac{m_1 + m_2 \cos\Delta\theta}{2}$ and $\beta = -\frac{m_2 \sin\Delta\theta}{2}$.

The linear combination of sines and cosines can be written as a cosine with a different phase, that is

$$a \sin x + b \cos x = \sqrt{a^2 + b^2} \cos(x + \varphi) \tag{32}$$

with

$$\varphi = \begin{cases} \operatorname{arccot}\left(\frac{b}{a}\right) & a \geq 0 \\ \operatorname{arccot}\left(\frac{b}{a}\right) \pm \pi & a < 0. \end{cases}$$

By substituting the result obtained in (32) into (31) we can write the second term of (31) as follows:

$$\frac{1}{2\pi} \int_0^{2\pi} e^{\alpha \cos\xi + \beta \sin\xi} d\xi = \frac{1}{2\pi} \int_0^{2\pi} e^{\sqrt{\alpha^2+\beta^2} \cos(\xi+\varphi)} d\xi. \tag{33}$$

Again, with a simple variable substitution with $\chi = \xi + \varphi$ we obtain

$$\frac{1}{2\pi} \int_0^{2\pi} e^{\sqrt{\alpha^2+\beta^2} \cos\chi} d\chi = I_0\left(\sqrt{\alpha^2 + \beta^2}\right). \tag{34}$$

By exploiting the definitions of α and β we can write

$$\begin{aligned} \sqrt{\alpha^2 + \beta^2} &= \sqrt{\left(\frac{m_1 + m_2 \cos\Delta\theta}{2}\right)^2 + \frac{m_2^2 \sin^2\Delta\theta}{4}} \\ &= \frac{\sqrt{m_1^2 + m_2^2 + 2 m_1 m_2 \cos\Delta\theta}}{2} \end{aligned} \tag{35}$$

which leads to the required result (9).

APPENDIX C
DERIVATION OF THE EM FOR A MIXTURE OF WRAPPED GAUSSIAN (MOWG)

Given the same notation and derivation scheme reported in Appendix A, the $\mathbb{E}_{\mathbf{Z}}$ value in the case of MoWG (15) can be written as $\mathbb{E}_{\mathbf{Z}} = \sum_{i=1}^{n} \sum_{k=1}^{K} \gamma_{ik} \left\{\ln \pi_k + \ln \mathcal{N}^W\left(\theta_i|\theta_{0,k}, \sigma_k\right)\right\}$.

The values γ_{ik} in the E step and π_k in the M step can be computed similarly to Appendix A [(16), (17)].

The maximum with respect to $\theta_{0,k}$, instead, can be computed as follows:

$$\begin{aligned} &\frac{d}{d\theta_{0,k}} \mathbb{E}_{\mathbf{Z}} = \sum_{i=1}^{n} \frac{d}{d\theta_{0,k}} \left(\gamma_{ik} \ln \mathcal{N}^W\left(\theta_i|\theta_{0,k}, \sigma_k\right)\right) = 0 \\ &\sum_{i=1}^{n} \gamma_{ik} \frac{1}{\mathcal{N}^W\left(\theta_i|\theta_{0,k}, \sigma_k\right)} \frac{d}{d\theta_{0,k}} \left(\mathcal{N}^W\left(\theta_i|\theta_{0,k}, \sigma_k\right)\right) = 0. \end{aligned} \tag{36}$$

The derivative of the wrapped Gaussian $\mathcal{N}^W$ with respect to $\theta_{0,k}$ can be computed as follows:

$$\begin{aligned} &\frac{d}{d\theta_{0,k}} \left(\sum_{w=-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{\left(\theta_i - w2\pi - \theta_{0,k}\right)^2}{2\sigma_k^2}} \right) = \\ &= \sum_{w=-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_k} \left(e^{-\frac{\left(\theta_i - w2\pi - \theta_{0,k}\right)^2}{2\sigma_k^2}} \cdot \frac{\left(\theta_i - w2\pi - \theta_{0,k}\right)}{\sigma_k^2} \right) \\ &= \sum_{w=-\infty}^{+\infty} \mathcal{N}\left(\theta_i - w2\pi|\theta_{0,k}, \sigma_k\right) \left(\frac{\left(\theta_i - w2\pi - \theta_{0,k}\right)}{\sigma_k^2} \right). \end{aligned} \tag{37}$$

Inserting (37) in (36) and eliminating the denominators of both equations since they do not change the zeros of the derivative and are never equal to zero, we can obtain the following equation:

$$\sum_{i=1}^{n}\gamma_{ik}\sum_{w=-\infty}^{+\infty}\mathcal{N}\left(\theta_i - w2\pi|\theta_{0,k},\sigma_k\right)\left(\theta_i - w2\pi - \theta_{0,k}\right) = 0. \tag{38}$$

This equation can be split in two and, accounting that $\theta_{0,k}$ is independent of w and i, we can write

$$\sum_{i=1}^{n}\gamma_{ik}\sum_{w=-\infty}^{+\infty}\mathcal{N}\left(\theta_i - w2\pi|\theta_{0,k},\sigma_k\right)\cdot(\theta_i - w2\pi) = \theta_{0,k}\sum_{i=1}^{n}\gamma_{ik}\sum_{w=-\infty}^{+\infty}\mathcal{N}\left(\theta_i - w2\pi|\theta_{0,k},\sigma_k\right) \tag{39}$$

which can easily lead to (17).

Similarly, the log likelihood equation can be derived with respect to σ_k, and in accordance with the procedure above, the derivative of the wrapped Gaussian with respect to σ_k must be computed

$$\frac{d}{d\sigma_k}\left(\sum_{w=-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma_k}e^{-\frac{\left(\theta_i - w2\pi - \theta_{0,k}\right)^2}{2\sigma_k^2}}\right) = \sum_{w=-\infty}^{+\infty}\left(-\frac{\mathcal{N}\left(\theta_i - w2\pi|\theta_{0,k},\sigma_k\right)}{\sigma_k} + \mathcal{N}\left(\theta_i - w2\pi|\theta_{0,k},\sigma_k\right)\cdot\frac{\left(\theta_i - w2\pi - \theta_{0,k}\right)^2}{\sigma_k^3}\right) \tag{40}$$

which can be obtained by using the product rule of the derivative. Including this equation in the (36), making use of (14) and (20) and simplifying σ_k at the denominator, we obtain

$$\sum_{i=1}^{n}\gamma_{ik}\frac{1}{\mathcal{N}^W\left(\theta_i|\theta_{0,k},\sigma_k\right)}\cdot\left(-\frac{\mathcal{N}^W\left(\theta_i|\theta_{0,k},\sigma_k\right)}{\sigma_k} + \frac{\beta_{ik}^{(\sigma^2)}}{\sigma_k^3}\right) = 0$$
$$\sum_{i=1}^{n}\gamma_{ik} = \frac{1}{\sigma_k^2}\sum_{i=1}^{n}\frac{\gamma_{ik}\beta_{ik}^{(\sigma^2)}}{\mathcal{N}^W\left(\theta_i|\theta_{0,k},\sigma_k\right)} \tag{41}$$

which leads to (18).

Appendix D
Derivation of the EM for a Mixture of Approximated Wrapped Gaussian (MoAWG)

Once again (see Appendix C), we start our derivation from the expected value of the log likelihood of (22)

$$\mathbb{E}_{\mathbf{Z}} = \sum_{i=1}^{n}\sum_{k=1}^{K}\gamma_{ik}\left\{\ln\pi_k + \ln\mathcal{N}^{AW}\left(\theta_i|\theta_{0,k},\sigma_k\right)\right\}. \tag{42}$$

Deriving with respect to $\theta_{0,k}$ and finding zeros will result in the following equation:

$$-\frac{1}{2\sigma_k^2}\sum_{i=1}^{n}\gamma_{ik}\frac{d}{d\theta_{0,k}}\left(\left(\theta_i - \theta_{0,k}\right)\mod 2\pi\right)^2 = 0. \tag{43}$$

The mod operator is not derivable. However, the function $(x \bmod 2\pi)$ has the extrema points in $k\pi$ with $k \in \mathbb{Z}$. By substituting the non-derivable function $(x \bmod 2\pi)$ with a derivable function with the same extrema, such as $\cos\left(\frac{x}{2}\right)$, we can easily find the zeros of the cos function as follows:

$$\sum_{i=1}^{n}\gamma_{ik}\frac{d}{d\theta_{0,k}}\left(\frac{1}{2} + \frac{\cos\left(\theta_i - \theta_{0,k}\right)}{2}\right) = 0$$
$$\cos\theta_{0,k}\cdot\sum_{i=1}^{n}\gamma_{ik}\sin\theta_i - \sin\theta_{0,k}\cdot\sum_{i=1}^{n}\gamma_{ik}\cos\theta_i = 0 \tag{44}$$

which brings to the solution reported in (23).

Deriving the expected log likelihood of (42) with respect to σ_k^2 we obtain

$$\sum_{i=1}^{n}\gamma_{ik}\frac{d}{d\sigma_k^2}\left\{\left(\frac{\ln\left(2\pi\sigma_k^2\right)}{2}\right) + \left(\frac{\left(\left(\theta_i - \theta_{0,k}\right)\mod 2\pi\right)^2}{2\sigma_k^2}\right)\right\} = 0$$
$$\sum_{i=1}^{n}\gamma_{ik}\left\{-\frac{1}{2}\frac{2\pi}{2\pi\sigma_k^2} + \frac{1}{2}\frac{\left(\left(\theta_i - \theta_{0,k}\right)\mod 2\pi\right)^2}{\left(\sigma_k^2\right)^2}\right\} = 0$$
$$\sum_{i=1}^{n}\gamma_{ik}\frac{\left(\left(\theta_i - \theta_{0,k}\right)\mod 2\pi\right)^2}{\sigma_k^2} = \sum_{i=1}^{n}\gamma_{ik} \tag{45}$$

which corresponds to (23).

References

[1] B. Morris and M. Trivedi, "A survey of vision-based trajectory learning and analysis for surveillance," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 8, pp. 1114–1127, Aug. 2008.

[2] J. Carta, C. Bueno, and P. Ramires, "Statistical modeling of directional wind speeds using mixtures of von Mises distributions: Case study," *Energy Conversion Manage.*, vol. 49, no. 5, pp. 897–907, 2008.

[3] Y. Agiomyrgiannakis and Y. Stylianou, "Stochastic modeling and quantization of harmonic phases in speech using wrapped Gaussian mixture models," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Apr. 2007, pp. IV-1121–IV-1124.

[4] D. Berndt and J. Clifford, *Finding Patterns in Time Series: A Dynamic Programming Approach*. Menlo Park, CA: American Association for Artificial Intelligence, 1996, pp. 229–248.

[5] C. Ratanamahatana and E. Keogh, "Three myths about dynamic time warping data mining," in *Proc. SIAM Int. Conf. Data Mining*, 2005, pp. 506–510.

[6] M. Vlachos, D. Gunopulos, and G. Das, "Rotation invariant distance measures for trajectories," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2004, pp. 707–712.

[7] N. Pelekis, I. Kopanakis, G. Marketos, I. Ntoutsi, G. Andrienko, and Y. Theodoridis, "Similarity search in trajectory databases," in *Proc. 14th Int. Symp. Temporal Represent. Reasoning*, Jun. 2007, pp. 129–140.

[8] Z. Zhang, K. Huang, and T. Tan, "Comparison of similarity measures for trajectory clustering in outdoor surveillance scenes," in *Proc. Int. Conf. Pattern Recognit.*, vol. 3. 2006, pp. 1135–1138.

[9] N. Piotto, N. Conci, and F. De Natale, "Syntactic matching of trajectories for ambient intelligence applications," *IEEE Trans. Multimedia*, vol. 11, no. 7, pp. 1266–1275, Nov. 2009.

[10] L. Chen, M. Özsu, and V. Oria, "Robust and fast similarity search for moving object trajectories," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2005, pp. 491–502.

[11] P.-F. Marteau, "Time warp edit distance with stiffness adjustment for time series matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 306–318, Feb. 2009.

[12] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang, and E. J. Keogh, "Querying and mining of time series data: Experimental comparison of representations and distance measures," in *Proc. VLDB Endowment*, vol. 1, no. 2. 2008, pp. 1542–1552.

[13] F. I. Bashir, A. A. Khokhar, and D. Schonfeld, "Segmented trajectory based indexing and retrieval of video data," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2003, pp. 623–626.
[14] J. Lou, Q. Liu, T. Tan, and W. Hu, "Semantic interpretation of object activities in a surveillance system," in *Proc. Int. Conf. Pattern Recognit.*, vol. 2. 2002, pp. 777–780.
[15] I. Junejo, O. Javed, and M. Shah, "Multifeature path modeling for video surveillance," in *Proc. Int. Conf. Pattern Recognit.*, vol. 2. Aug. 2004, pp. 716–719.
[16] X. Chen, D. Schonfeld, and A. Khokhar, "Robust null space representation and sampling for view invariant motion trajectory analysis," in *Proc. IEEE Int. Conf. Comput. Vision Pattern Recognit.*, Jun. 2008, pp. 1–6.
[17] Y. Zhou, S. Yan, and T. Huang, "Detecting anomaly in videos from trajectory similarity analysis," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2007, pp. 1087–1090.
[18] S. Needleman and C. Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins," *J. Mol. Biol.*, vol. 48, no. 3, pp. 443–453, 1970.
[19] S. Calderara, R. Cucchiara, and A. Prati, "A dynamic programming technique for classifying trajectories," in *Proc. IEEE Int. Conf. Image Anal. Process.*, Sep. 2007, pp. 137–142.
[20] H. Fashandi and A. M. Eftekhari Moghaddam, "A new rotation invariant similarity measure for trajectories," in *Proc. IEEE Int. Symp. Comput. Intell. Robot. Autom.*, 2005, pp. 631–634.
[21] M. Bennewitz, W. Burgard, and S. Thrun, "Using EM to learn motion behaviors of persons with mobile robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2002, pp. 502–507.
[22] A. Mecocci and M. Pannozzo, "A completely autonomous system that learns anomalous movements in advanced videosurveillance applications," in *Proc. IEEE Int. Conf. Image Process.*, vol. 2. Sep. 2005, pp. 586–589.
[23] W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, and S. Maybank, "A system for learning statistical motion patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1450–1464, Sep. 2006.
[24] F. Porikli and T. Haga, "Event detection by eigenvector decomposition using object and frame features," in *Proc. CVPR Workshop*, vol. 7. 2004, pp. 114–121.
[25] A. Basharat, A. Gritai, and M. Shah, "Learning object motion patterns for anomaly detection and improved object detection," in *Proc. IEEE Int. Conf. Comput. Vision Pattern Recognit.*, May 2008, pp. 1–8.
[26] K. Mardia, *Statistics of Directional Data*. New York: Academic Press, 1972.
[27] P. Smaragdis and P. Boufounos, "Learning source trajectories using wrapped-phase hidden Markov models," in *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust.*, Oct. 2005, pp. 114–117.
[28] F. Seitner and A. Hanbury, "Fast pedestrian tracking based on spatial features and color," in *Proc. Comput. Vision Winter Workshop*, 2006, pp. 105–110.
[29] C. Bahlmann, "Directional features in online handwriting recognition," *Pattern Recognit.*, vol. 39, no. 1, pp. 115–125, Jan. 2006.
[30] A. Prati, S. Calderara, and R. Cucchiara, "Using circular statistics for trajectory analysis," in *Proc. IEEE Int. Conf. Comput. Vision Pattern Recognit.*, Jun. 2008, pp. 1–8.
[31] R. Fisher, "Dispersion on a sphere," *Proc. Roy. Soc. London Ser. A*, vol. 217, pp. 295–305, 1953.
[32] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 381–396, Mar. 2002.
[33] T. H. Cormen and R. L. Rivest, *Introduction to Algorithms*. Cambridge, MA: MIT Press, 2001.
[34] A. Reynolds, G. Richards, and V. Rayward-Smith, *The Application of K-Medoids and PAM to the Clustering of Rules*, vol. 3177. Berlin/Heidelberg, Germany: Springer, 2004, pp. 173–178.
[35] C. Stauffer and W. Grimson, "Learning pattern of activity using real-time tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 747–757, Aug. 2000.
[36] G. Casella and R. Berger, *Statistical Inference*, 2nd ed. Pacific Grove, CA: Duxbury Press, 2002.
[37] C. Bishop, *Pattern Recognition and Machine Learning*. Berlin, Germany: Springer-Verlag, 2006.
[38] M. Sato, "Fast learning of on-line EM algorithm," ATR Hum. Inform. Process. Res. Labs., Tech. Rep. TR-H-281, 1999 [Online]. Available: citeseer.ist.psu.edu/330617.html
[39] S. Calderara, R. Cucchiara, and A. Prati, "Bayesian-competitive consistent labeling for people surveillance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 354–360, Feb. 2008.
[40] D. Llorens, F. Prat, A. Marzal, J. M. Vilar, M. J. Castro, J. C. Amengual, S. Barrachina, A. Castellanos, S. España, J. A. Gómez, J. Gorbe, A. Gordo, V. Palazón, G. Peris, R. Ramos-Garijo, and F. Zamora, "The ujipenchars database: A pen-based database of isolated handwritten characters," in *Proc. 6th Int. Conf. Language Resources Eval.*, May 2008, pp. 2647–2651.

Simone Calderara received the M.S.I. degree in computer science from the University of Modena and Reggio Emilia, Reggio Emilia, Italy, in 2004, and the Ph.D. degree in computer engineering and science in 2009.

He is currently a Research Assistant with the Imagelab Computer Vision and Pattern Recognition Laboratory, University of Modena and Reggio Emilia. His current research interests include computer vision and pattern recognition applied to human behavior analysis, video surveillance and tracking, and time series analysis for forensic applications.

Andrea Prati (SM'09) received the Laurea degree in computer engineering and the Ph.D. degree in information engineering from the University of Modena and Reggio Emilia, Reggio Emilia, Italy, in 1998 and 2002, respectively.

During the last year of his Ph.D. studies, he was with the Computer Vision and Robotics Research laboratory, University of California, San Diego, for 6 months. Since 2005, he has been an Assistant Professor and has recently become an Associate Professor with the Faculty of Engineering. He collaborates on research projects at regional, national, and European levels. He is the author or co-author of more than 100 papers in national and international journals and conference proceedings. He has been an invited speaker, organizer of workshops and journal's special issues, and reviewer for many international journals in the field of computer vision and multimedia.

Dr. Prati was the Program Chair of the 14th International Conference on Image Analysis and Processing, Modena, Italy, in September 2007, and of the upcoming ACM/IEEE International Conference on Distributed Smart Cameras in 2011. He is a member of ACM and Gruppo Italiano di Ricercatori in Pattern Recognition (International Association for Pattern Recognition Italy).

Rita Cucchiara (M'99) received the Laurea degree in electronic engineering and the Ph.D. degree in computer engineering from the University of Bologna, Bologna, Italy, in 1989 and 1992, respectively.

Since 2005, she has been a Full Professor with the University of Modena and Reggio Emilia, Reggio Emilia, Italy. Since 2008, she has been the Deputy Dean of the Faculty of Engineering. She heads the Imagelab Laboratory (http://imagelab.ing.unimore.it). In the field of video surveillance she works on models for people tracking and behavior analysis by coordinating several projects, including FREE-SURF (PRIN MIUR 2006–2008) on distributed video surveillance, and behavioral learning in surveilled areas with feature extraction (BE SAFE) (2006–2009—Nato Science for Peace). She collaborates with the Modena's Police for forensic analysis and with several local companies for urban security. In the field of multimedia, she works on annotation, retrieval, and human-centered searching of images and videos in different EU and national projects. In the field of robot vision, she is responsible for research projects on 3-D reconstruction and visual inspection. She is the author of more than 200 papers in journals and international proceedings, and is a reviewer for several international journals. Her current research interests include computer vision, pattern recognition, and multimedia systems.

Prof. Cucchiara is a member of the Editorial Boards of *Multimedia Tools and Applications* and *Machine Vision and Applications* journals and Chair of several workshops and conferences. She has been the General Chair of the 14th International Conference on ICIAP in 2007 and the 11th National Congress AI*IA in 2009. She is a member of the IEEE Computer Society, ACM, GIRPR, and AI*IA. Since 2006, she has been a Fellow of IAPR.