

# Alignment-based Similarity of People Trajectories using Semi-directional Statistics

Simone Calderara  
D.I.I.

Univ. of Modena and Reggio Emilia  
Modena, Italy  
simone.calderara@unimore.it

Andrea Prati  
Di.S.M.I.

Univ. of Modena and Reggio Emilia  
Reggio Emilia, Italy  
andrea.prati@unimore.it

Rita Cucchiara  
D.I.I.

Univ. of Modena and Reggio Emilia  
Modena, Italy  
rita.cucchiara@unimore.it

**Abstract**—This paper presents a method for comparing people trajectories for video surveillance applications, based on semi-directional statistics. In fact, the modelling of a trajectory as a sequence of angles, speeds and time lags, requires the use of a statistical tool capable to jointly consider periodic and linear variables. Our statistical method is compared with two state-of-the-art methods.

**Keywords**-Trajectory analysis, circular statistics, people video surveillance

## I. INTRODUCTION AND RELATED WORKS

People trajectories represent a rich source of information which can be robustly extracted from single or multiple fixed cameras, and exploited in many video-surveillance and forensics applications, and in particular for people behavior understanding [1].

Here, we address the approaches which compare complete and long trajectories for classifying or clustering them into groups of similarity in order to infer specific behaviors or detect anomalies. Trajectories can be considered as time series of features and compared with inexact matching techniques (to cope with their different length) which perform data alignment based on a certain cost function or distance measure, such as DTW (Dynamic Time Warping) [2] or LCSS (Longest Common Sub Sequence) [3].

Many papers addressed people trajectories [4], [5], [6]; a recent survey is [1]. In this work we discuss several methods to manage trajectory data, with different choices on the type of feature used and the type of comparison. Then, we propose a new statistical method which, in spite of its high computational time, has a very high accuracy in trajectory clustering and is very robust to noise and non-idealities.

As for most of the problem of pattern recognition, the choice of the feature set is very critical. In most of the approaches, the trajectories are modeled as a sequence of  $\langle x, y \rangle$  coordinates on the ground plane, often associated with the motion status, e.g. the punctual velocity  $v$  or the time lag  $t$  [1]. The addition of speed and time (duration) can indeed provide a more complete analysis of the trajectory, distinguishing, for instance, the same path covered by a walk or by a run. In some applications the analysis of *trajectory shape* is more suitable than that of the *trajectory points*. For instance, for a human observer, people that are walking

straight have the same behavior even if their path is more at right or at left side of a corridor. To describe trajectory shape, the sequence of directions between two consecutive tracking samples can be adopted [7], [5]. Unfortunately, the use of a sequence of directions to model the trajectories poses the problem of non-homogeneous variables, since angles are periodic or directional variables, while speed, time and many others are linear.

Regardless of the features (defined with a heuristic choice or with a feature-selection mechanism), the approaches can be classified into categories depending on three characteristics:

- 1) whether they use a point-to-point comparison or exploit statistical data representation and a correspondent pattern recognition approach (*point-to-point* versus *statistical*);
- 2) whether they use the original feature space or provide a feature extraction step after a more or less sophisticated space transformation (*original* versus *transformed*);
- 3) whether they use all the temporal data or select a subset of them (*complete* versus *selected*).

The first dichotomy is very strong: many approaches treat trajectories as time series and, especially if large amount of data are available as in very-large-database (VLDB) community, they provide a point-to-point comparison, on each feature or the feature set, after an adequate alignment. Many different measures exist that can be applied in the point-wise comparison of people trajectories: Euclidean Distance is the simplest one but its efficacy is demonstrated only when data presents a similar length. However, point-to-point comparison is likely to decrease its effectiveness in the presence of noise and uncertainty [5]. Hence, in surveillance contexts the similarity measures have been typically defined in a statistical framework. In [8], Basharat *et al.* proposed 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. Hu *et al.* [4] reported a complete system for learning statistical motion patterns with a two-stage fuzzy k-means. Porikli [6] proposed the use of a HMM-based similarity measure where

each trajectory is modelled with a HMM and compared using the cross likelihood.

Regarding the second issue, most of the point-to-point methods work on the original data but some papers exist to work into a more suitable transformed space. For instance, by working on angles we proposed to use a symbol representing a quantization of the angle into main directions. This method has been extended in a recent work by [5]: the authors proposed an edit distance with a global alignment approach on a sequence of symbols representing angles, speeds and discontinuities. Finally, also an adequate reduction of the representative data in the dataset is possible: while many approaches use the complete time series, others [5] select a subset of representative points only by automatically identifying the discontinuities in the trajectory.

In this paper we propose a statistical, transformed and complete approach where all the data are transformed by mapping each original point into a symbol representing the most suitable pdf with an EM to estimate the pdf parameters. In order to have a complete approach, we aim at jointly working on both periodic (angle) and linear (speed and time) data; but their statistical formulation became less trivial since the space is not homogeneous. To take different nature of these features into account and to use a statistical model, the resulting trivariate joint probability  $p(\theta, v, t)$  can be easily modelled as the product  $p(\theta) \cdot p(v) \cdot p(t)$  if and only if the three variables result to be statistically independent. Since the velocity is computed using both space and time the statistical independence assumption is not feasible, additionally when the sampling rate is not constant the time variable is also not negligible and must be accounted in comparing sequences. As a consequence the joint probability must be modelled by using a directional (univariate) pdf for  $\theta$  and a linear (bivariate) pdf for  $v$  and  $t$ . The estimation of the covariance matrix for this trivariate joint pdf can be quite challenging since the dependency between  $\theta, v$  and  $t$  must be modelled properly. When a directional or periodic variable is combined with linear ones the term *semi-directional* is often used [7].

Hence, we define a mixture of semi-directional pdfs, named AWLG (Approximated Wrapped and Linear Gaussian) to model and analyze people trajectory shapes. Each semi-directional pdf (representative in the mixture of a given trajectory) is mapped to a symbol in order to the trajectory in a sequence of symbols.

In the following sections, we will detail the method and then provide a comparison with two state-of-the-art methods falling into other two categories. The method proposed by [2] has been selected as representative of the methods used in time series comparison in large database, without the use of statistics, and with a very-high-speed point-to point comparison in the original feature space without any sub-sampling. The method [5] is a very recent proposal that still works without statistical models, but on a transformed space

with a quantization of both the time and the feature domain.

## II. OBSERVATION MODEL

Among the possible pdfs from directional statistics to model angles, the *approximated wrapped Gaussian* (AWG) is one the most used since it derives from the Gaussian and can be thus easily treated:

$$AWG(\theta|\theta_0, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{((\theta-\theta_0) \bmod 2\pi)^2}{2\sigma^2}} \quad (1)$$

where the operation “mod” represents the remainder of the division and resulting angle must be in the interval  $(-\pi, \pi]$ , being the input angle  $\theta$  in this interval.

Dealing with semi-directional data, the joint pdf we used is called Approximated Wrapped and Linear Gaussian (AWLG) and is defined as:

$$AWLG(X|\mu, \Sigma) = \frac{1}{\sqrt{2\pi} \cdot |\Sigma|} e^{-\frac{1}{2}(X-\mu)^T \Sigma^{-1} (X-\mu)} \quad (2)$$

Here  $X = \begin{bmatrix} \theta \\ v \\ t \end{bmatrix}$  is the observation,  $\mu = \begin{bmatrix} \theta_0 \\ v_0 \\ t_0 \end{bmatrix}$  is the mean,  $X - \mu = \begin{bmatrix} (\theta - \theta_0) \bmod 2\pi \\ v - v_0 \\ t - t_0 \end{bmatrix}$  is the “difference”

between them,  $\Sigma$  is the covariance matrix and  $|\Sigma|$  denotes its determinant. This expression is similar to a trivariate Gaussian pdf, but one of the three dimensions is represented by a AWG.

In order to account for trajectories with different main directions and speeds, a mixture model (called MoAWLG) can be defined and the corresponding EM steps can be found (for bivariate version of it) in [7].

## III. COMPARING TRAJECTORIES

A sequence  $T_j$  of  $n_j$  semi-directional data ( $\theta, v$  and  $t$ ) can be written as:  $T_j = \{\langle \theta_{1,j}, v_{1,j}, t_{1,j} \rangle, \dots, \langle \theta_{n_j,j}, v_{n_j,j}, t_{n_j,j} \rangle\}$ . In order to cluster or classify these sequences, a method for comparing them by means of a suitable similarity or distance measure is mandatory. Assuming that the three variables are statistically dependent, the EM algorithm of the corresponding distribution can be run on every trajectory of the training set to learn the MoAWLG’s parameters. After obtaining the main components of the associated distribution, it is possible to encode each data  $\langle \theta_i, v_i, t_i \rangle$  with a symbol  $S_i$  with a Maximum-A-Posteriori (MAP) approach, which associates a single value to the most similar component of the mixture distribution representing the data. With this MAP approach each trajectory  $T$  is encoded with a sequence of symbols  $\bar{T} = \{S_1, S_2, \dots, S_n\}$ . The symbol  $S_i$  corresponds to a AWLG component ( $S_i \leftrightarrow \langle (\mu_r, \Sigma_r) \rangle$  as in eq. (2)), where  $\mu_r$  and  $\Sigma_r$  are the parameters of the  $r^{th}$  components of the MoAWLG that maximizes the posteriori.

This representation allows to compare sequences if a suitable measure  $\Omega(\overline{T}_i, \overline{T}_j)$ , which must be obtained through an alignment process, is defined. Among the various techniques, we propose to adopt the *global alignment*, specifically the well-known Needleman-Wunsch algorithm which also accounts for *gaps*. A global alignment (over the entire sequence) is preferable (instead of a local one) because the former preserves both global and local shape characteristics.

In global alignment a cost  $C(S_i, S_j)$  of the alignment between two symbols  $S_i$  and  $S_j$  must be defined:

$$C(S_i, S_j) = KL(AWLG(X|\mu_{r_i}, \Sigma_{r_i})|AWLG(X|\mu_{r_j}, \Sigma_{r_j})) \quad (3)$$

where  $r_i$  and  $r_j$  are the selected components for symbols  $S_i$  and  $S_j$ , respectively. KL corresponds to the Kullback-Leibler divergences between two AWLG distributions. Since AWLG is an approximation of a Gaussian, the analytical formulation of the KL divergence between normal distributions can be used. Given  $AWLG(X|\mu_1, \Sigma_1)$  and  $AWLG(X|\mu_2, \Sigma_2)$ , their KL divergence is equal to:

$$KL = \frac{1}{2} \ln \frac{|\Sigma_2|}{|\Sigma_1|} - \frac{N_1}{2} + \frac{1}{2} \text{tr}(\Sigma_2^{-1} \Sigma_1) + \frac{1}{2} \left( (\mu_1 - \mu_2)^T \Sigma_2^{-1} (\mu_1 - \mu_2) \right) \quad (4)$$

where  $N_1$  is the dimension of  $\Sigma_1$ .

#### IV. EXPERIMENTAL RESULTS AND DISCUSSION

To test the proposed measure a corpus of trajectories acquired in a typical campus scenario during several working days has been collected and manually ground-truthed. We decided to evaluate the efficacy of the similarity measure in clustering similar trajectories based on direction, velocity and time observations. Four different campaigns have been carried out. The first two (T1 and T2 in Tab. I) consist in clustering trajectories acquired during ordinary days. The third test (T3 in Tab. I) campaign consists in trajectories performed by authors instructed to follow specific paths. The last campaign (T4 in Tab. I) aims to test the influence of speed in the clustering process; more precisely, the trajectories in the data set share the same direction but are performed at significantly different speeds.

As said in the introduction, we compared our method with those proposed in [2] and [5]. Without going into details, the method in [2] simply performs a point-to-point comparison adopting the Euclidean distance in conjunction with the DTW algorithm to align sequences of different length. Conversely, the approach in [5] first performs a subsampling of the trajectories by finding some interesting points (breakpoints) that represent temporal (start or stop events) or spatial (sudden or slow direction changes) discontinuities. Once the breakpoints are detected, the time series are described by the sequences of their breakpoints, in particular by a triplet (direction, speed and elapsed time)

computed between two consecutive breakpoints. The similarity measure between two sequences is then obtained by associating, through quantization, every breakpoints' triplet to a symbol, by aligning separately the three symbol sequences, one for every quantity (direction, speed and time), using the Needleman-Wunsch algorithm, and finally by summing up all the alignment scores with a weighted sum.

Test ID	# traj	[2]	[5]	AWLG
T1	140	78%	73%	95%
T2	108	80%	87%	99%
T3	195	94%	86%	96%
T4	100	90%	80%	97%

Table I  
SUMMARY OF THE RESULTS.

To perform a fair comparison we adopt the same clustering algorithm for the three considered similarity measures, computing the distance matrix between the time series with the corresponding approach. The choice of the clustering algorithm is not critical: we selected the iterative variant of the *k-medoids* algorithm [7]: the algorithm iteratively merges similar clusters until convergence to obtain the “best” number of clusters.

The accuracy have been evaluated comparing the clustering results against a manual ground truth and the results are summarized in Tab. I.

By observing the results it is evident that our proposal outperforms the compared methods in all the tests. This is mainly due to both the statistical quantization of the data, that tends to filter out the measurement noise and tracking errors, and the adoption of linear and angular distributions to jointly model the multi-dimensional observation (direction, speed and time).

Most of the errors of the method in [2] are due to the absence of a statistical approach. Regarding [5], most of the errors are due to the high number of score weights used to compute the similarity measure. It is indeed difficult to find the optimal values that fit for all the considered scenarios. Additionally, another important issue is that the method considers the speed as disjoint from the direction, by performing different alignments and then summing up the similarity scores. In this manner, two time series may be similar in direction but dissimilar in speed, but still result to be similar depending on the weights assigned to the two measures. Fig. 1 shows some examples of the trajectories used in the tests and the most numerous clusters obtained by our method.

As said before, the high accuracy achieved by our proposal comes at the price of time performances, mainly due to the EM algorithm that iterates until convergence. In fact, the average time (computed implementing all the algorithms in Matlab on a Core 2 Duo PC running Windows Vista) needed to perform a single comparison between two time

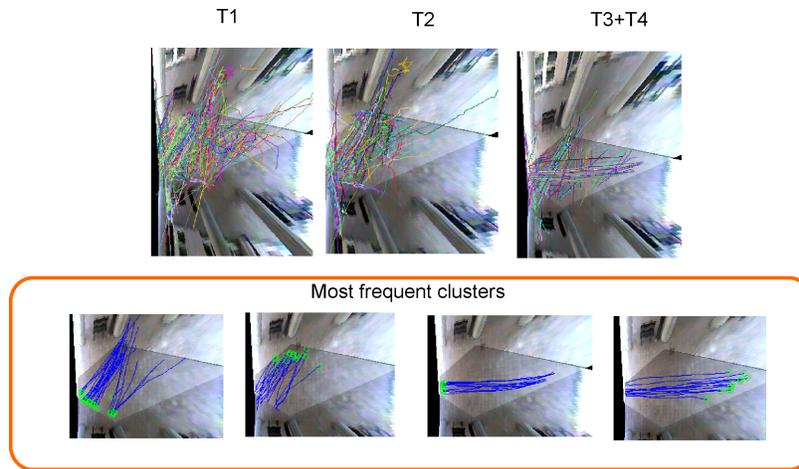


Figure 1. Trajectories used for test (top) and most frequent clusters automatically extracted (bottom).

series is approximately 3.2 seconds for our approach, while in the order of 0.08 seconds for both [2] and [5]. Despite the slow computational time, our approach is by far more precise than the other algorithms and requires to tune very few parameters that can be left unchanged for most of the real video-surveillance scenarios. Thus, we suggest to adopt the proposed similarity measure in all the applications where the processing time is not a strict constraint (such as forensic analysis) or as a preliminary step to obtain a robust clustering of the training data, using more performant approaches for classification.

## V. CONCLUSIONS

In this paper, we propose a similarity measure that is suitable for comparing and clustering time series of multidimensional data using a joint statistical distribution that can model in a natural way both periodical and linear data. The results underline the high degree of precision of the proposed measure applied to video surveillance data where noise and errors strongly affect the quality of the data themselves.

## REFERENCES

- [1] B. Morris and M. Trivedi, "A survey of vision-based trajectory learning and analysis for surveillance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 8, pp. 1114–1127, Aug. 2008.
- [2] E. J. Keogh and M. J. Pazzani, "Scaling up dynamic time warping for datamining application," in *Proc. of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2000)*, 2000, pp. 285–289.
- [3] M. Vlachos, D. Gunopulos, and G. Das, "Rotation invariant distance measures for trajectories," in *Proceedings of SIGKDD*, 2002.
- [4] W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, and S. Maybank, "A system for learning statistical motion patterns," *IEEE Trans. on PAMI*, vol. 28, no. 9, pp. 1450–1464, September 2006.
- [5] N. Piatto, N. Conci, and F. De Natale, "Syntactic matching of trajectories for ambient intelligence applications," *IEEE Transactions on Multimedia*, vol. 11, no. 7, pp. 1266–1275, Nov. 2009.
- [6] F. Porikli and T. Haga, "Event detection by eigenvector decomposition using object and frame features," in *Proc. of Computer Vision and Pattern Recognition (CVPR) Workshop*, vol. 7, 2004, pp. 114–121.
- [7] S. Calderara, A. Prati, and R. Cucchiara, "Learning people trajectories using semi-directional statistics," in *Proceedings of IEEE International Conference on Advanced Video and Signal Based Surveillance (IEEE AVSS 2009)*, Genova, Italy, Sep. 2009.
- [8] A. Basharat, A. Gritai, and M. Shah, "Learning object motion patterns for anomaly detection and improved object detection," in *Proc. of IEEE Int'l Conference on Computer Vision and Pattern Recognition*, 2008.