Trajectory Distance Metric Using Hidden Markov Model based Representation

Fatih Porikli
Mitsubishi Electric Research Laboratories
Cambridge, MA 02139, USA
fatih@merl.com

Abstract

In this paper, we introduce a set of novel distance metrics that use model based representations for trajectories. We determine the similarity of trajectories using the conformity of the corresponding HMM models. These metrics enable the comparison of trajectories without any limitations of the conventional measures. They accurately identify the coordinate, orientation, and speed affinity. The proposed HMM based distance metrics can be used not only for ground truth comparisons but for clustering as well. Our experiments prove that they have superior discriminative properties.

1. Introduction

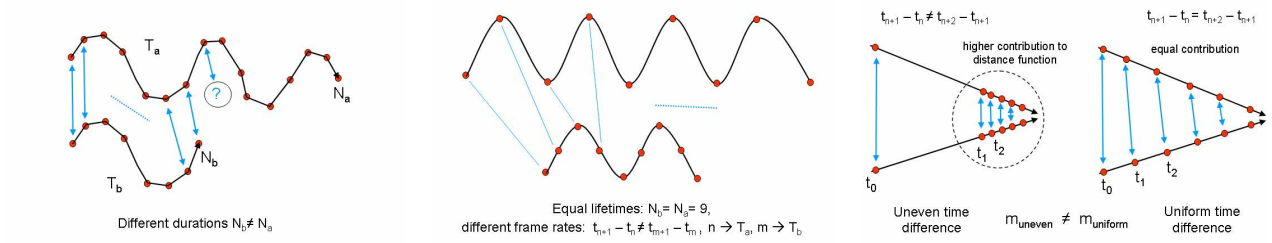
Recent advances in object tracking made it possible to obtain spatiotemporal motion trajectories for further analysis of concealed information. Although the extraction of trajectories is well understood and studied, relatively little investigation on the precise comparison of the trajectories and the secondary outputs of the tracking process is presented in the literature.

A key issue in performance evaluation of tracking results is the distance metric that determines the similarity of the trajectories. Any additional analysis, such as action recognition, event detection, etc., highly depends on the accuracy of the similarity assessment. Most existing measures [2], [6] compute a mean distance of the corresponding positions of two equal duration trajectories. Supplementary statistics such as variance, median, minimum, and maximum distances are also suggested to extend the description of similarity. In a recent work, Needham [4] proposed an alignment based distance metric that reveals the spatial translation and temporal shift between the given trajectories, and introduced an area based metric that calculates the total enclosed area between the trajectories using trajectory intersections. Similarly, Ellis [1] characterized several statistical properties of the tracking performance using the compensated means and standard deviations.

One main disadvantage of the existing approaches is that they are all limited to the equal duration (lifetime) trajectories. By duration we refer the number of coordinate points that constitute the trajectory. These coordinates are sampled at different time instances. Since the existing measures depend on the mutual coordinate correspondences, they cannot be applied to trajectories that have different durations unless the trajectory duration is normalized or parameterized first. However, such a normalization destroy the temporal properties of the trajectory.

Conventional distance measures assume that the temporal sampling rates of the trajectories are equal. For instance, a ground truth trajectory labeled at a certain frame rate can be compared only with the trajectory generated by a tracker working at the same frame rate. These approaches does not cope with the uneven sampling instances, i.e. varying temporal distance between the coordinates, either. This is a common case especially for the real-time object trackers that process streaming video data. A real time tracker works on the next available frame, which may not be the immediate temporal successor of the current, whenever the current frame is processed. Thus, the obtained trajectory coordinates have varying temporal distance.

Therefore, there is a need to develop other alternatives that can effectively measure the difference between unrestricted trajectories. In this paper, we introduce a set of novel distance metrics that use model based representations. We determine the similarity of trajectories using the conformity of the corresponding models. We construct a mixture of continuous Hidden Markov Models (HMM) that capture the dynamic properties of trajectory within a state transition matrix. The HMM based metrics enable the comparison of trajectories without any limitations of the previous measures. We can use the proposed metrics not only for ground truth comparisons but for clustering as well. We measure the similarity of trajectories that have different durations, sampling rates, and temporal properties. We accurately identify coordinate, orientation, and speed affinity. We also propose additional features that are extracted from the trajectories such as object-wise histograms of aspect-



(a) (b) (c)

Figure 1. Ambiguous cases for conventional metrics; (a) unequal durations, (b) equal durations but different frame rates. (c) Effect of uneven frame rates.

ratio, location, orientation, speed, size, etc. to improve the available features.

In section 2, we discuss the existing trajectory distance metrics. In section 3 we introduce the additional features. In section 4 we present the HMM based distance metrics, and in the last section we discuss the experimental results.

2. Trajectory Distance Measures

A trajectory is a time sequence of coordinates representing the motion path of an object over the duration (lifetime), i.e. number of frames that object exists. These coordinates correspond to marked positions of object shape in consecutive frames. A marked position often indicates the center-of-mass (for pixel model), the intersection of main diagonals (for ellipsoid model), and the average of minimum and maximum on perpendicular axes (for bounding box model) of object region. It is, therefore, possible to view the trajectory as a collection of frame-wise abstractions of object shape. We will adopt the following notation $T : \{p_n\} : \{(x_1, y_1, t_1), (x_2, y_2, t_2), ..., (x_N, y_N, t_N)\}$ where N is the duration.

The simplest metric used for computing the distance between a pair of trajectories is the mean of coordinate distances, which is given as

$$m_1(T^a, T^b) = \frac{1}{N} \sum_{N}^{n=1} d_n^2 \tag{1}$$

where the displacement between the positions is calculated using the Cartesian distance

$$d_n^2 = [(x_n^a - x_n^b)^2 - (y_n^a - y_n^b)^2]^{\frac{1}{2}}, \tag{2}$$

or using other L-norm formulations

$$d_n^L = [(x_n^a - x_n^b)^L - (y_n^a - y_n^b)^L]^{\frac{1}{L}}. \tag{3}$$

Note that, the mean distance metric makes three critical assumptions; 1) the durations of the both trajectories are same $N^a = N^b = N$ (fig. 1-a), 2) the coordinates are synchronized $t_n^a = t_n^b$ (fig. 1-b), and 3) the time sampling rate is constant $t_n^a - t_{n+1}^a = t_m^a - t_{m+1}^a$ since the contribution of each distance component d_n in equation 1 is same as illustrated in fig. 1-c. It is evident that the mean of distances is very sensitive to the partial mismatches and cannot deal with the distortions in time.

To provide more descriptive information, the second order statistics such as median, variance, minimum and maximum distance may be incorporated. The variance is defined as

$$m_2(T^a, T^b) = \frac{1}{N} \sum_{n=1}^{N} (d_n - m_1(T_a, T_b))^2. \tag{4}$$

To find the median, the displacements d_n are ordered with respect to their magnitudes as $d_n \rightarrow d_m$, then the value of the halfway component of the list is assigned

$$m_3(T^a, T^b) = \begin{cases} d_{\frac{N+1}{2}} & N\ odd \\ \frac{1}{2}(d_{\frac{N}{2}} + d_{\frac{N+1}{2}}) & N\ even \end{cases}$$

The minimum and maximum distances are simply defined as

$$m_4(T^a, T^b) = \min d_n \tag{5}$$

$$m_5(T^a, T^b) = \max d_n \tag{6}$$

Although these statistics supply extra information, they inherit (even amplify) the shortcomings of the ordinary mean of distances metric m_1. Besides, none of the above metrics is sufficient enough by itself to make an accurate assessment of the similarity.

An area based distance metric is proposed in [4]. The crossing points $q : T^a(p_i) = T^b(p_j)$ of two paths are used to define regions Q_j $j = 1, ..J$ between the trajectories. For each region, a polygon model is generated and the enclosed area is found by tracing the parameterized shape

$$m_6(T^a, T^b) = \sum_{j=1}^{J} area(Q_j) \tag{7}$$

This metric can handle more complex trajectories, however it is sensitive to entanglements of the trajectory, it discards the time continuity, and fails to the distinguish two trajectories in opposite directions. Although the area between a pair of trajectories is easily apprehended, its computation may demand case-specific analytic solutions that are not always straightforward to formulate.

It is possible to compute an optimal spatiotemporal alignment $(\delta x, \delta y, \delta t)$ for which the mean distance is minimized

$$(\delta x, \delta y, \delta t) = \arg\min m_1(T^a, T^b + (\delta x, \delta y, \delta_t)) \quad (8)$$

and use this alignment to compute a compensated distance

$$m_7(T^a, T^b) = m_1(T^a, T^b + (\delta x, \delta y, \delta_t)). \quad (9)$$

Ellis [1] proposed several statistical measures such as true detection rate, false alarm rate, etc. using the aligned trajectories for comparison of a trajectory with the ground truth. However, not all trajectory distance tasks involve ground truth comparison, i.e. clustering.

In the following sections, we introduce an extended set of trajectory based features, and then we present the details of the HMM based metrics.

3. Trajectory Based Features

A set of coordinates is not the only available trajectory feature. In spite of its simplicity, duration (lifetime) is a distinctive feature. For instance, at a hallway camera in a surveillance setting the suspicious event may be a left behind unattended bag, which can be easily detected since human objects do not stay still for extended periods of time. The total length of the trajectory is defined as $\sum_{n=2}^{N} |T(p_n) - T(p_{n-1})|$. This is different from the total displacement of the object, which is equal to $|T(p_N) - T(p_1)|$. By assuming a ground plane of the camera imaging system is available, the trajectory may be projected to obtain the respective 3D length. A total orientation descriptor keeps the global direction of the object. Depending on the camera setup, the length related descriptors may be used to differentiate unusual paths. The length/duration ratio gives the average speed.

Dynamic properties of an object such as orientation, aspect ratio, size, instantaneous speed, and location are represented by histograms. The location histogram keeps track of the image coordinates where object stays most. Using the size histogram, dynamic properties of the object size are captured, e.g. we can separate an object moving towards the camera (assuming the size will get larger) from another object moving parallel. An object moves at different speeds during tracking, therefore the instantaneous speed of an object is accumulated into a histogram. Speed is the key aspect of some events, e.g. a running person where everybody walks. The speed histogram may be used to interpret the regularity of the movement such as erratically moving objects. An accident can be detected using the speed histogram; the speed components will accumulate around zero and high velocities rather than being distributed uniformly.

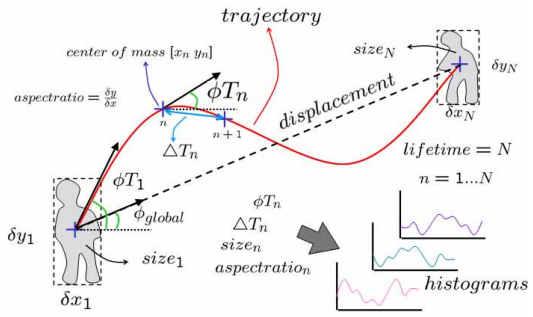


Figure 2. Additional trajectory features.

The orientation histogram is one of the important descriptors. For instance, it becomes possible to distinguish objects moving on a certain path, making circular movements, etc. It is possible to find a vehicle backing up on a wrong lane then driving correctly again, which may not be detected using a global orientation. The aspect ratio is a good descriptor to distinguish between human objects and vehicles. The aspect ratio histogram can capture whether a person crouches and stands up during its lifetime. Figure 2 illustrates some of the object features.

4. Hidden Markov Model Based Metric

Due to the shortcomings of the existing metrics, we propose a model based representation that captures the dynamic properties of trajectories. We project each trajectory T into a parameter space λ that is characterized by a set of HMM parameters.

An HMM is a probabilistic model composed of a number of interconnected states a directed graph, each of which emits an observable output. Each state is characterized by two probability distributions: the transition distribution over states and the emission distribution over the output symbols. A random source described by such a model generates a sequence of output symbols as follows: at each time step the source is in one state, and after emitting an output symbol according to the emission distribution of the current state, the source jumps to a next state according to the transition distribution of its current state. Since the activity of the source is observed indirectly, through the sequence of output symbols, and the sequence of states is not directly observable, the states are said to be hidden.

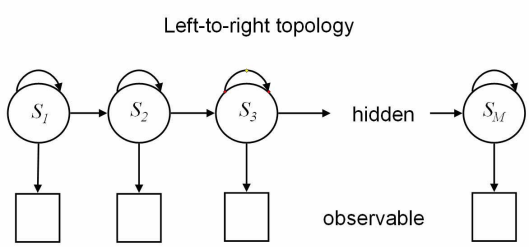


Figure 3. Left-to-right topology.

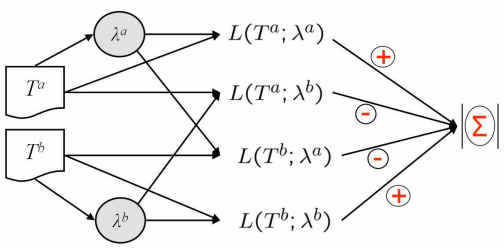


Figure 4. Cross-fitness distance.

In our case, we replace the trajectory information as the emitted observable output for the above directed graph. The hidden states then capture the transitive properties of the consecutive coordinates of the spatiotemporal trajectory. The state sequence that maximizes the probability becomes the corresponding model for the given trajectory.

A simple specification of an K-state $\{S_1, S_2, ..., S_K\}$ continuous HMM with a Gaussian observation is given by:

1. A set of prior probabilities $\pi = \{\pi_i\}$ where $\pi_i = P(q_1 = S_i)$ and $1 \leq i \leq K$.
2. A set of state transition probabilities $B = \{b_{ij}\}$, where $b_{ij} = P(q_{t+1} = S_j | q_t = S_i)$ and $1 \leq i, j \leq K$.
3. Mean, variance and weights of mixture models $\mathcal{N}(O_t; \mu_j, \Sigma_j)$ where μ_j and Σ_j are the mean and covariance of the state j.

Above, q_t and O_t are the state and observation at time t, respectively.

For each trajectory T^a, we fit an M-mixture HMM $\lambda^a = (\pi, B, \mu, \Sigma)^a$ that has left-to-right topology using the Baum-Welch algorithm. We chose the left-to-right topology since it can efficiently describe continuous processes.

We train a HMM model using the trajectory as the training data after we initialize the state transition and prior probability matrices with random variables, thus we make no assumptions on the trajectory. Initialization can be adapted for specific applications as well. Finally, each trajectory is assigned to a separate model.

The optimum number of states and mixtures depend on the complexity and duration of the trajectories. To provide sufficient evidence to every Gaussian of every state in the training stage, the lifetime of the trajectory should be much larger than the number of mixtures times number of states $N \gg M \times K$. A state can be viewed as a basic pattern of the trajectory, thus depending the trajectory, the number of states should be large enough to conveniently characterize such distinct patterns but small enough to prevent from overfitting.

A priori knowledge about tracking scenario may be used to impose a structure on an HMM and a meaning for the values of the state variable. It is known that each state may be associated with a certain label. Furthermore, the topology of the HMM can be strongly constrained: most transition probabilities are forced to be zero. Since the number of free parameters and the amount of computation are directly dependent on the number of non-zero transition probabilities, imposing such structure is very useful when it is appropriate. The most basic structure that is often imposed on HMM's is the left-to-right structure: states are ordered sequentially and transitions go from the "left" to the "right", or from a state to consecutive state or itself, as in fig. 3.

We search an optimal number of states of the HMM network for the given trajectory while repeating the generation and evaluation of the topology. At the beginning of the search, possible HMM's up to a maximum number of states are generated randomly. In general, the likelihood of HMM increases with the complexity of the topology. However, it is known that over representation is frequently observed as the complexity increases. Therefore, in order to balance the likelihood and the complexity, we have adopted a score [3] as

$$v_i = [-\log L(T; \lambda_i) + i\alpha]^{-1} \tag{10}$$

where $i = 2, ..., M_{max}$ is the number of states, $L(T; \lambda_i) = P(T|\lambda_i)$ is the likelihood obtained for HMM with i-states, and α is a constant balancing factor. The number of states is then chosen as the one that given the highest score.

We define the distance between two trajectories in terms of their HMM parameterizations:

$$\begin{aligned} m_8(T^a, T^b) \quad = \quad & |L(T^a; \lambda^a) + L(T^b; \lambda^b) \\ & -L(T^a; \lambda^b) - L(T^b; \lambda^a)| \end{aligned} \tag{11}$$

which corresponds the cross-fitness of the trajectories to each other's models as illustrated in fig. 4. The $L(T^a; \lambda^a)$ and $L(T^b; \lambda^b)$ terms indicate the likelihood of the trajectories to their own fitted model, i.e. we obtain the maximum likelihood response for the models. The cross terms $L(T^a; \lambda^b), L(T^b; \lambda^a)$ reveal the likelihood of a trajectory generated by the other trajectories model. In other words, if two trajectories are identical, the cross terms will have a

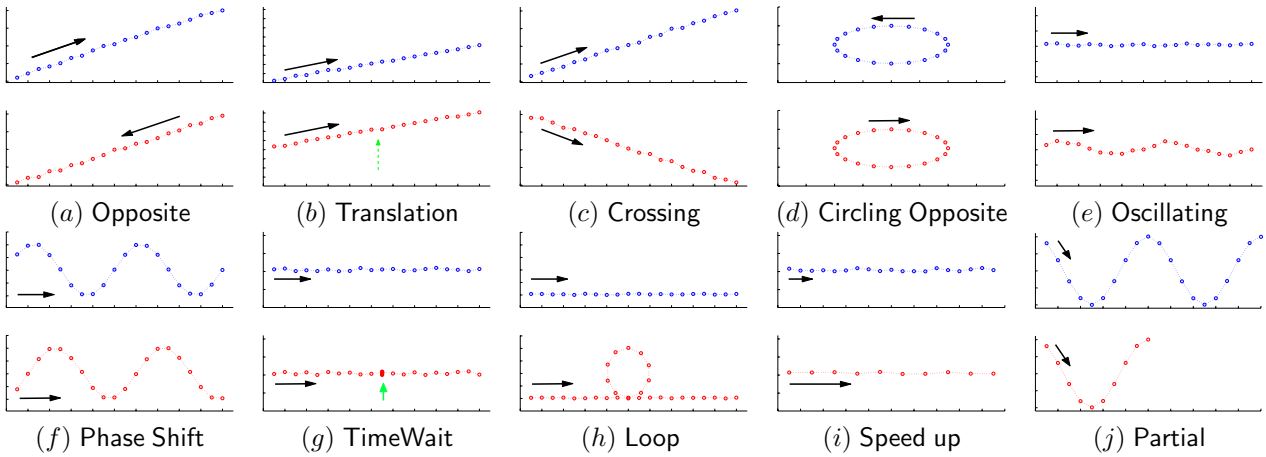


Figure 5. Different trajectory pattern pairs.

maximum value, thus eq. 12 will be equal to zero. On the other hand, if trajectories are different, their likelihood of being generated from each others model will be small, thus the distance will be high.

Up to now, we employed trajectory coordinates as our feature sequence. Using coordinates reveals spatial correlation between trajectories, however in some situations it is more important to distinguish shape similarity of the trajectories independent of the spatial coordinates. The instantaneous speed and orientation sequences are potential features that establish shape similarity even if there is a spatial translation. Thus, we define two other sequential features and corresponding distances; the orientation sequence as

$$\begin{aligned} \phi T(p_n) &= \tan^{-1} \frac{y_n - y_{n-1}}{x_n - x_{n-1}} \\ m_9(T^a, T^b) &= m_8(\phi T^a, \phi T^b) \end{aligned} \quad (12)$$

and the speed sequence

$$\begin{aligned} \triangle T(p_n) &= [(x_n - x_{n-1})^2 + (y_n - y_{n-1})^2]^{\frac{1}{2}} \\ m_{10}(T^a, T^b) &= m_8(\triangle T^a, \triangle T^b). \end{aligned} \quad (13)$$

The mentioned HMM distance is also applicable to histogram features such as orientation histogram, speed histogram, etc. However, since these features discard the temporal ordering of the points, they are more suitable to evaluate the statistical properties of trajectories rather than measuring the similarity of their shape and coordinates.

5. Comparisons

To compare the proposed metrics m_8, m_9, m_{10} and the referenced conventional metrics $m_1, .., m_7$, we computed the distances of several distinct trajectory pattern pairs as presented in fig. 5. Equal (a-f) and unequal (g-j) duration trajectories are among these patterns. Each equal duration trajectory consists of 100 points. To make the comparison more realistic, we added a random white noise to all patterns. The first set of equal duration patterns include the trajectory pairs that are in opposite direction, spatially shifted trajectories, trajectories that are crossing each other, trajectories that have the same circling path but in opposite direction, trajectories that their global orientation is same but their paths have small perturbations, and trajectories that the form is same except a time shift. The second set of trajectory pairs have different durations. For instance, fig. 5-g shows a pair that have same spatial path but one of the trajectory has a several frames long waiting period as shown with the green arrow. Fig. 5-h shows a pair that are same spatial form except one trajectory has a loop. In fig. 5-i the second trajectory has the same form but its duration is half of the first one. In fig. 5-j a partially matching pair is given.

After we computed the distances of all pairs for a given metric, we normalized the distances using the maximum distance obtained for that metric since there is no a common normalization factor that can applied to all the metrics. For instance, the numerical values of the variance (m_3) and the area (m_6) metrics are clearly incommensurate. Thus, we evaluate the sensitivity based on the given pattern set. We listed the normalized responses of all metrics in table 1. The highest score at each column indicates the pattern that the metric is most sensitive. Note that, an ideal metric should be applicable to all diverse patterns regardless of the trajectory duration, frame-rate, and other limitations.

From the table, it is evident that the sum of coordinate distances m_1, the variance of coordinate distances m_2, and the median coordinate distance m_3 have all similar properties. Their fusion would not improve the overall discriminative capability. The maximum distance m_4 and minimum

Table 1. Comparison of Distance Metrics

	m_8	m_9	m_{10}	m_1	m_2	m_3	m_4	m_5	m_6	m_7
Opposite (ED)	0.123	1.000	0.001	1.000	1.000	1.000	0.055	1.000	1.000	0.001
Translation (ED)	0.356	0.001	0.006	0.283	0.001	0.287	1.000	0.148	0.002	0.573
Crossing (ED)	1.000	0.370	0.002	0.707	0.502	0.721	0.016	0.707	0.677	1.000
Circling (ED)	0.008	0.105	0.000	0.449	0.143	0.491	0.000	0.355	0.403	0.000
Perturbation (ED)	0.001	0.027	0.012	0.017	0.000	0.018	0.001	0.014	0.417	0.139
Phase shift (ED)	0.073	0.001	0.002	0.107	0.008	0.123	0.029	0.085	0.020	0.226
Wait (VD)	0.069	0.071	0.316	-	-	-	-	-	-	0.001
Loop (VD)	0.389	0.529	0.775	-	-	-	-	-	-	0.001
Speed up(VD)	0.001	0.214	1.000	-	-	-	-	-	-	0.003
Partial (VD)	0.198	0.001	0.002	-	-	-	-	-	-	0.000

(Each column is normalized within itself, ED: equal duration, VD: variable duration)

distance m_5 are very sensitive to singularities, for instance the maximum distance can be very high even a the trajectories have matching well except a single coordinate. The minimum distance fails if a single crossing exits. The spatiotemporal alignment metric m_7 is insensitive to shifting, otherwise it is similar to m_1. These metrics cannot handle different duration trajectories. The area metric m_6 fails for patterns that have same path but opposite direction. It cannot distinguish the temporal deformations either.

On the other hand, the HMM based metrics are applicable to trajectories that have different durations. It is shown that these metrics can successfully identify various temporal deformations including the time waiting, partial match, different speed, time loop, etc. Each topology has 3 states and 3 Gaussian models. The coordinate based HMM metric m_8 is sensitive towards the spatial positioning of the trajectories, and it can identify the crossing, translation, phase shift, time loop, partial, and opposite directions. The orientation based HMM metric m_9 is responsive towards the orientation variances, i.e. it gave the highest score to opposite direction pattern, and it can recognize crossing, time loop, and circling patterns. The speed based HMM m_{10} detects the speed changes and time loops most effectively, and it can identify the uneven frame-rates as well.

We observed that the three possible HMM metrics are responsive towards the different patterns, thus their mixture is a perfect candidate for measuring the trajectory distance.

We conducted another experiment using the PETS-2004 benchmark sequences. For the sequences that the ground truth is not given, we obtained the trajectories by our mean-shift based tracker [5]. The trajectories, which have duration ranging from 30 to 800 points, are presented in fig. 6. We determined the most similar and most different trajectories to a given trajectory using the m_8 metric as shown in fig. 7. In the graphs, the red is the given trajectory. The blue is the most similar and green is the most different trajectory among the all trajectories. As visible, the proposed HMM based metric accurately identified the most similar and dissimilar trajectories at each case.

6. Conclusion

We proposed a set of HMM based trajectory distance metrics that can accurately measure the coordinate, orient, and speed similarity of a pair of trajectories. These metrics measure different duration trajectories without destroying the temporal properties. They can be used not only for ground truth comparisons but also for further analysis of the tracking results, e.g. clustering and event analysis. Our experiments prove that the HMM distance metrics have superior discriminative properties than conventional metrics.

References

[1] T. Ellis. Performance metrics and methods for tracking in surveillance. *Proc. of PETS*, pages 26–31, Copenhagen, Denmark, June 2002.

[2] C. Jaynes, S. Webb, R. Steele, and Q. Xiong. An open development environment for evaluation of video surveillance systems. *Proc. of PETS*, Copenhagen, Denmark, June 2002.

[3] A. Konagaya and Y. Kondo. Stereo person tracking with adaptive plan-view statistical templates. *Hawaii Int. Conf. on System Sciences*, pages 746–755, 1993.

[4] C. Needham and R. Boyle. Performance evaluation metrics and statistics for positional tracker evaluation. *Proc. of ICVS*, pages 278–289, Graz, Austria, April 2003.

[5] F. Porikli and O. Tuzel. Performance evaluation metrics and statistics for positional tracker evaluation. *Proc. of PETS*, pages 37–45, Graz, Austria, April 2003.

[6] A. Senior, A. Hampapur, Y. Tian, L. Brown, S. Pankanti, and R. Bolle. Appearance models for occlusion handling. *Proc. of PETS*, Hawaii, Kauai, December 2001.

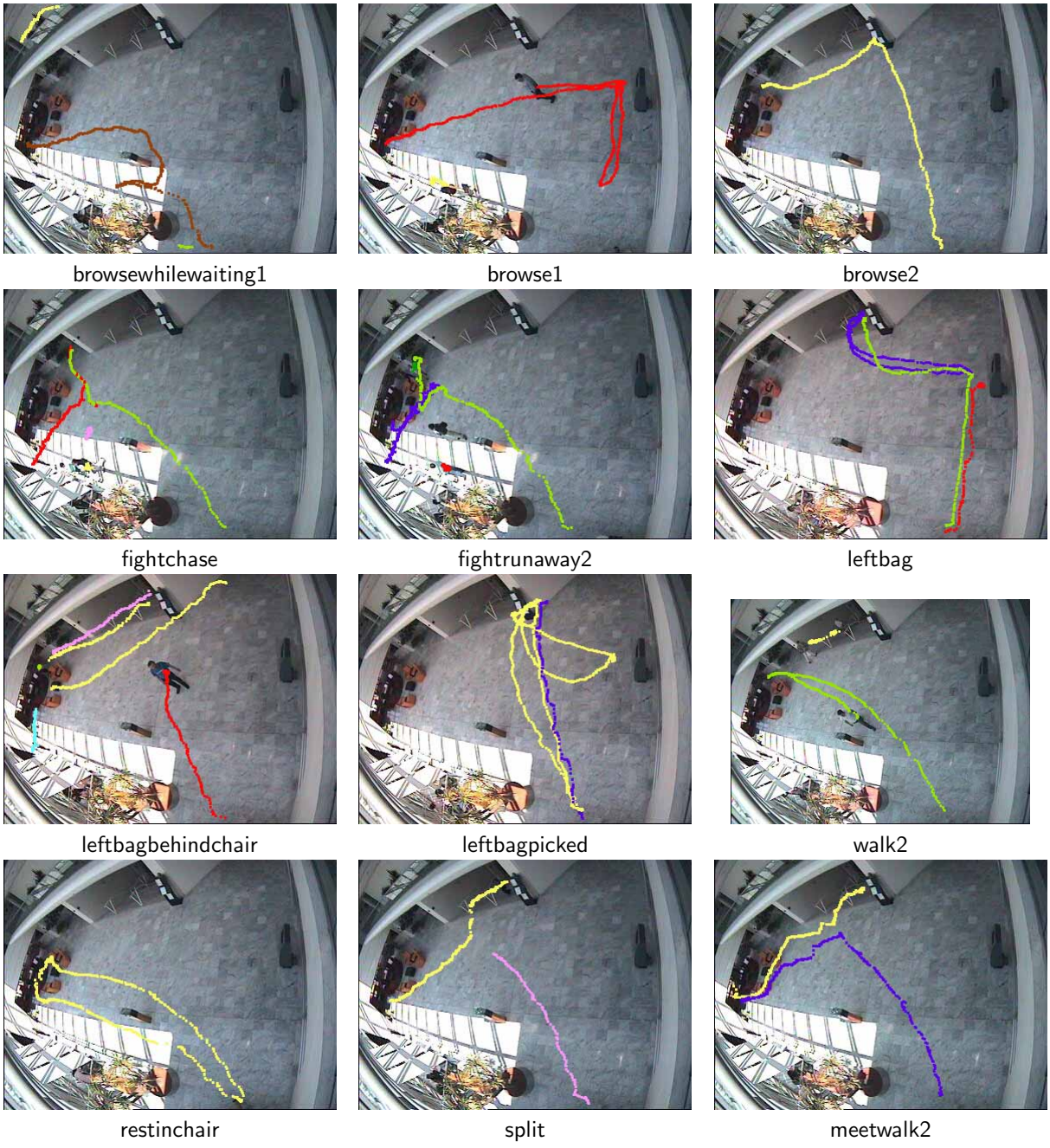


Figure 6. Detected trajectories for the PETS-2004 sequences.

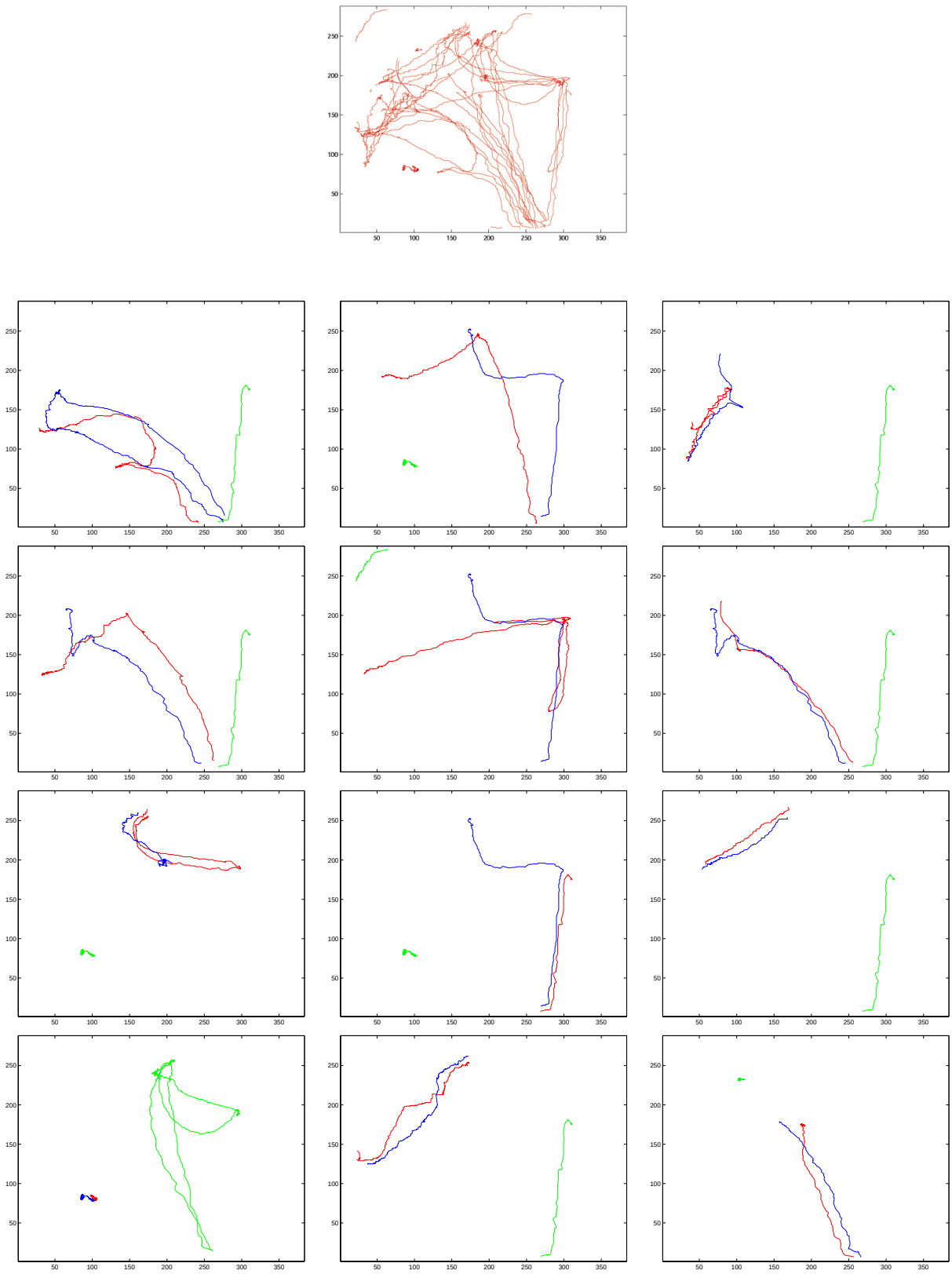


Figure 7. (Top graph) All trajectories mapped together. (Other graphs) Red: given trajectory, blue: most similar trajectory, green: most different trajectory obtained by the coordinate HMM metric (m_8) for the given trajectory.