Manuscript published in “Pattern Recognition” 44 (2011)

Adaptive Human Motion Analysis and Prediction

Zhuo Chen*,
Laboratory for Intelligent Transportation Systems Research,
Department of Electrical and Electronic Engineering,
The University of Hong Kong, Pokfulam Road, Hong Kong.
Email: zchen@eee.hku.hk

Lu Wang,
Laboratory for Intelligent Transportation Systems Research,
Department of Electrical and Electronic Engineering,
The University of Hong Kong, Pokfulam Road, Hong Kong.
Email: wanglu@eee.hku.hk

Nelson H.C. Yung,
Laboratory for Intelligent Transportation Systems Research,
Department of Electrical and Electronic Engineering,
The University of Hong Kong, Pokfulam Road, Hong Kong.
Email: nyung@eee.hku.hk

*Corresponding author

Abstract

Human motion analysis and prediction is an active research area where predicting human motion is often performed for a single time step based on historical motion. In recent years, longer term human motion prediction has been attempted over a number of future time steps. Most current methods learn Motion Patterns (MPs) from observed trajectories and then use them for prediction. However, these learned MPs may not be indicative due to inadequate observation, which naturally affects on the reliability of motion prediction. In this paper, we present an adaptive human motion analysis and prediction method. It adaptively predicts motion based on the classified MPs in terms of their credibility, which refers to how indicative the learned MPs are for the specific environment. The main contributions of the proposed method are: First, it provides a comprehensive description of MPs including not only the learned MPs but also their evaluated credibility. Second, it predicts long-term future motion with reasonable accuracy. A number of experiments have been conducted in simulated scenes and real-world scenes and the prediction results have been quantitatively evaluated. The results show that the proposed method is effective and superior in its performance when compared with a recursively applied Auto-Regressive (AR) model which is called the Recursive Short-term Predictor (RSP) for long-term prediction. The proposed method has 17.73% of improvement over the RSP in prediction accuracy in the experiment with the best performance. On average, the proposed method has 5% improvement over the RSP in prediction accuracy over 10 experiments.

Keywords: Motion pattern, pattern clustering, pattern classification, prediction.

I. Introduction

Human motion analysis is essential for understanding movement of human objects and how they interact with each other in applications such as pedestrian surveillance, crowd control and social signal processing. The study of related spatial trajectories, velocities and changing angles under specific conditions builds knowledge of motion characteristics, which helps predict future movement, object interaction as well as assess the likelihood of potential dangerous and/or abnormal scenarios in dynamically changing situations.

In cases such as along corridors or sidewalks, pedestrians tend to follow well-defined paths, and the resultant motions are consistent and almost known a priori. Thus, motion analysis is trivial and prediction can be reasonably accurate. In other more generic environments however, there may not be any defined paths to walk along, e.g., shopping mall, trains station, playground and field. In these cases, the spatial trajectory, walking speed and how often direction is changed reflect the broad intention of the particular human concerned. Therefore, human motion analysis aims to extract such information, which can be very useful for predicting future movements.

The main challenge of human motion analysis and prediction is that human beings move according to their intentions, which can be rather difficult to be modeled or predicted [1]. Conventionally, a prediction of the human's location in the next time step is made based on his/her current and previous positions [2]. In such short-term prediction, motion characteristic is assumed to be consistent, i.e., a certain trend runs through the past, current and future steps. Given the historical motion patterns, the next location may be predicted based on techniques such as neural network [3], Markov models [4, 5], Kalman filter [6, 7] or collision/velocity cones [8, 9].

Although these techniques are able to produce reasonably good single-step prediction, their performance degrades rapidly when they are used to make longer term prediction [10], especially when motion trend is not obvious or consistent. To address this issue, recent attempts have been made to predict human motion over a number of future time-steps. For instance, the method presented in [11] manually defined some points of interest in the environment first, where human may likely visit, and one of these points is then selected as the destination position. The criterion used is that the selected destination is the closest to the tangential vector presenting the human positions of the last two time steps. It then treats the destination as the result of long-term motion prediction, although it really is a substantial simplification as there are many possible routes and MPs towards the same destination. In [12], the authors clustered observed trajectories into MPs using Expectation Maximization (EM), then derived hidden Markov models (HMM) from the learned MPs and used these HMMs to record human positions. However, the limitation of this approach is that the observed trajectories must include one or more of the so-called resting places where humans are assumed to stop and stay for a certain period of time. It requires the locations of these resting places be known a priori for the formation of MPs, which may not be readily available in reality. Other researchers further tackled the MP learning problem based on more general observable trajectories that do not need prior information. Often, a raw trajectory in these approaches is represented by a sequence of positions, describing the human's state at consecutive discrete time steps. These descriptive models focus on the physical state of the object without taking semantics into account and the learning problem is dealt with by unsupervised clustering algorithms to extract a number of 'typical' MPs from a set of raw trajectories [1]. One example is the approach presented in [13], which used fuzzy *k*-means to find MP clusters. In [14], the authors performed a geometrical analysis that compares the separation distance between trajectories and then hierarchically grouped trajectories to learn

MPs. Some other clustering approaches for learning MPs have also been proposed, including divisive clustering [15], graph cutting [16, 17], and spectral clustering [18, 19].

In this paper, we propose a method based on clustering and classifying MPs from the observable trajectories. Compared with the other methods for learning MPs, we not only obtain MPs by clustering accumulated human trajectories but also evaluate learned MPs in terms of their characteristics and adaptively predict long-term motions according to evaluated credibility of MPs. In the proposed method, observable trajectories are first derived from key frames in a video based on detected humans in a single frame and data association across frames [20-22]. Then, the derived trajectories are clustered using Constrained Gravitational Clustering (CGC) [23] to form MPs. This algorithm belongs to a class of agglomerative hierarchical clustering algorithms that are widely utilized in recent research on learning patterns [24]. For each clustered MP, it is further evaluated for credibility. The criterion of the credibility is to separate all clustered MPs into several credibility levels based on analyzing the mass and size information of each MP. The number of credibility levels is adaptively determined according to the characteristic of MPs. The MPs at the top level are most credible since each of such MP represents a class that has a considerable number of members that are strongly consistent with each other. The MPs at the bottom level have the lowest credibility because each of such MP represents a class that has a small number of members that are reasonably consistent but variation is evident. Based on credibility levels of the clustered MPs, an adaptive prediction model is developed. As long as a matched MP can be found for the current trajectory, its future motion is predicted to be similar to the matched MP. If the current trajectory is matched with a MP at the top credibility level, the predicted future motion covers the most number of time steps. On the other hand, matching with MP at lower credibility level implies a prediction of future motion with less number of time steps. Generally, the number of time steps of the predicted motion is determined by the credibility

level, i.e., the lower it is, the less number of time steps is predicted for future motion. If the current trajectory is not matched with any existing MPs, then prediction is made by an Auto-Regressive (AR) model [25] which is used to predict its action in the next time step only. This adaptive prediction approach is more reliable than those that apply the learned MPs equally without knowing their credibility. The proposed method has been realized and extensively tested in both simulated scenes and real-world scenes. It has been evaluated quantitatively by calculating prediction error based on predicted and actual motions. The results show that the proposed method makes reasonably accurate long-term prediction with acceptable error. It is further proved that it is superior in its performance, i.e., has lower prediction error, when compared with an AR model applied recursively for long-term prediction.

The rest of this paper is organized as follows. In Section II, the generalized framework of the proposed method is outlined. Section III describes the main functions in the proposed method. Section IV depicts the experimental results together with some related evaluation and analysis. Section V concludes the paper with a brief discussion of future research direction.

II. GENERALIZED FRAMEWORK

The generalized framework of the proposed method is depicted in Fig. 1, which consists of four main functions: (1) Trajectory Extraction; (2) MP Clustering; (3) MP Classification; and (4) Motion Prediction.

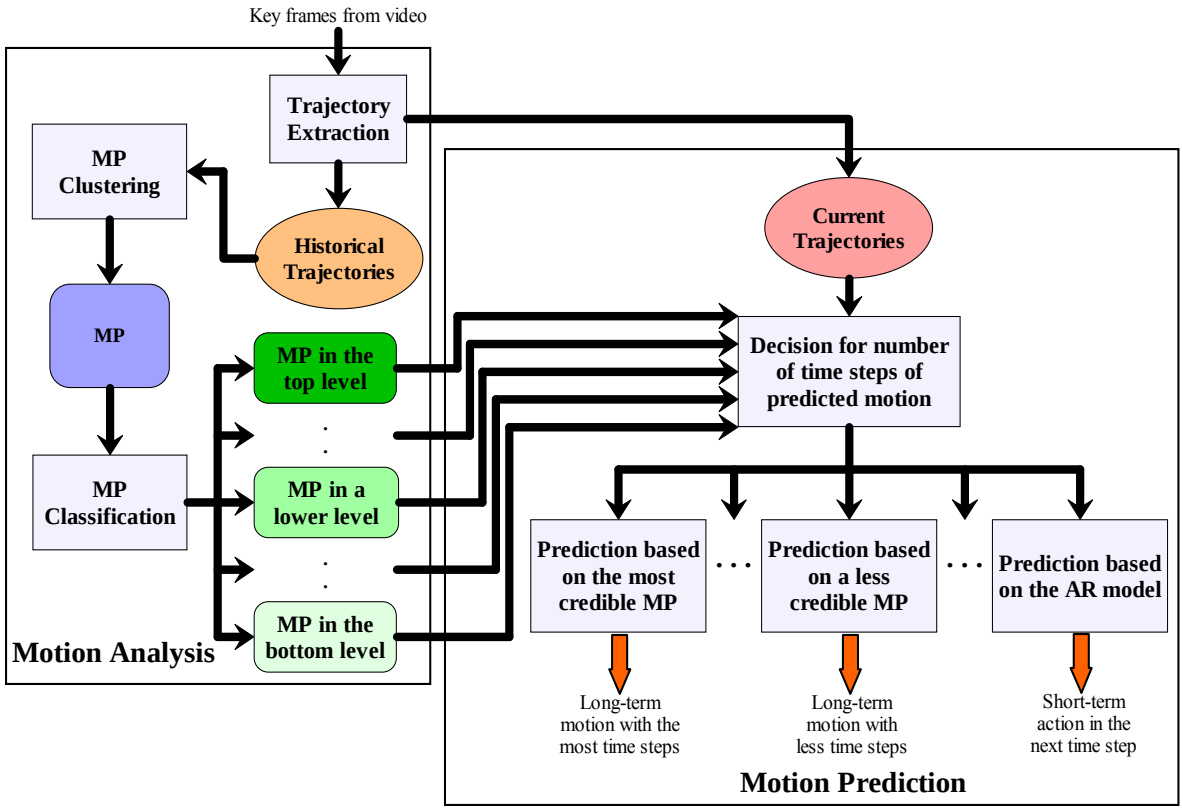

Fig. 1: Generalized framework of the proposed method

Given key frames from a video input, the observable trajectories are first extracted to form historical trajectories by using a model-based human detection and association method (3.1). MPs are then clustered using CGC as a general representation of a sub-group of trajectories (3.2). Other clustering methods may be used as well. After analyzing the mass and size information of clustered MPs, each clustered MP is further classified into a credibility level (3.3).

Given a current trajectory, it is matched with all the available MPs and the decision for number of time steps of predicted motion is made based on a probability model which measures the similarity between the current trajectory and the classified MP (3.4). If there is a match between the current trajectory and a MP at the top credibility level, then the prediction based on the most credible MP is performed and a long-term motion with most time steps is predicted to be similar to the matched MP. If there is a match between the current trajectory and a MP in a lower credibility level, then a prediction based on a less credible MP is

performed and a long-term motion with less time steps is predicted to be similar to the matched MP. If all else fail, which means the current trajectory is not matched to any existing MP, then prediction is performed for a single action in the next time step by using the AR model.

III. PROPOSED METHOD

3.1 Trajectory Extraction

For detecting human objects in a single frame [20], we formulate the problem into a Bayesian framework and design 3D human shape models to represent various postures. We calculate the image likelihood of a human model as the product of region likelihood, which measures how well the region covered by the model overlaps with the foreground, and shape likelihood, which measures the probability of the pixels covered by the model's boundary being real boundary pixels, with image evidence being provided by foreground extraction [21] and probability of boundary [26]. The prior requires that two humans cannot stand on the same location and the likelihood of a validated candidate is large enough, e.g. at least the human head is visible. To find the optimal solution, human candidates are first nominated by a head detector, which is an upper-semi circle detector; and a foot detector, which detects lower extrema on the foreground boundary. Then an iterative model fitting and candidate validation/rejection step follows. In each iteration, only human candidates that are possible to be un-occluded or whose occluding humans are likely to have been validated are selected for model fitting, and then a minimum description length based candidate validation and rejection strategy is applied on the fitted models to determine which should be validated or rejected. It ends when the status of every candidate has been determined. The output of human detection is a best fit model for each confirmed candidate.

To associate models across frames, the idea is to avoid identity switch by approximating the complicated 2nd-order Markov Chain with a simple and fast method. The proposed method consists of five steps. First, reliable initial tracklets are obtained by connecting two detection responses from consecutive frames using a two-threshold strategy [27], i.e. they are connected only if their affinity is high enough and significantly higher than the affinity of any other conflicting pairs (if two links share the same start point or end point, they are conflicting). Second, ambiguous tracklets that tend to introduce identity switch are explicitly detected and are not allowed to link temporarily. Third, the remaining tracklets are linked by applying the Hungarian algorithm [27]. Fourth, ambiguous tracklets are inserted into tracks formed by reliable tracklets by applying the Hungarian algorithm again. Finally, we iteratively link all the tracklets and tracks, and break the link with lower link probability if an ambiguous tracklet has two ends linked simultaneously. The iteration ends when there is no link to break. Due to occlusion and missed detection, some tracks may not be complete, i.e. not both ends are at the entrances/exits of the scene. We simply discard incomplete tracks and further manually check complete tracks in which the checked correct ones are extracted for the later trajectory clustering.

Based on human detection and model association, we could derive the original location information of the human, which is represented in the form of discrete time location information $\mathbf{r}^o_k[n]$ where $\mathbf{r}^o_k[n]=(x^o_k[n], y^o_k[n])$. Since human beings could walk at different speeds, which result in different distances covered in the same time interval, a re-sampling step is required such that a more reasonable comparison and match can be performed in the clustering and prediction stage. The re-sampled trajectory is obtained by using a circular moving window along the motion direction of the original trajectory. The intersection coordinate of the moving window and the trajectory is orderly recorded, and the re-sampled trajectory is defined by the set of corresponding sequential coordinates

$\mathbf{T}_k=\mathbf{t}_k(n_1,n_2)=\{\mathbf{r}_k[n]\}=\{(x_k[n], y_k[n])\}$ where n_1 and n_2 are the starting and ending time steps of the trajectory, respectively, and $n_1 \le n \le n_2$.

3.2 MP Clustering

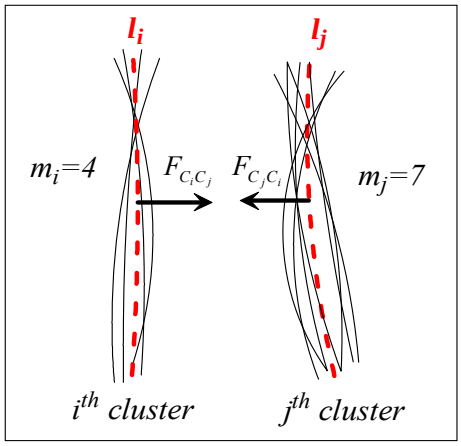

Fig. 2: Parameters for gravitational force

To cluster MPs from observable trajectory data, we employ the CGC method as described in [23]. It imposes a clustering constraint per iteration to control the formation of multiple clusters, without the need to assign a termination condition. At the start, each trajectory is regarded as the initial mean location vector of a cluster. In principle, the clustering method is completely controlled by the attraction between existing clusters. Analogy to gravitational force, existing clusters separated by a short distance are more likely to form a new cluster compared with those separated by a long distance. The 'gravitational force', $F_{C_iC_j}$ between the i^{th} and j^{th} clusters is given as

$$F_{C_iC_j} = G\frac{m_i \times m_j}{\left|l_i - l_j\right|^3}(l_i - l_j), \tag{1}$$

where G is the gravitational constant, m_i and m_j are the masses represented by the numbers of trajectories in the i^{th} and j^{th} clusters respectively, and l_i and l_j are the mean vectors of trajectory data in the i^{th} and j^{th} clusters respectively, as depicted in Fig. 2.

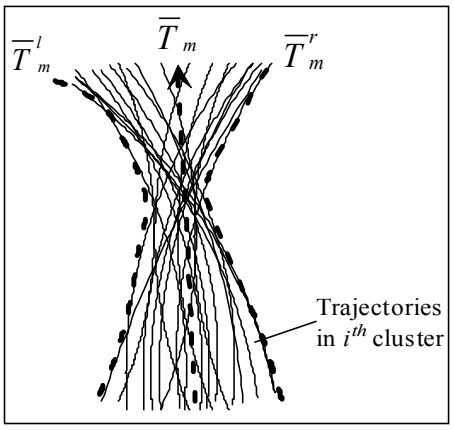

Fig. 3: Description of MP

After MP clustering, each cluster contains a number of trajectories with similar spatial location information. An MP is defined for each cluster. In the definition, we use three elements: 1) The mean vector of trajectories in the cluster; 2) The left boundary of trajectories in the cluster; 3) The right boundary of trajectories in the cluster. For example, the m^{th} clustered MP is given by $C_m=\{\overline{\mathbf{T}}^{l}_{m}, \overline{\mathbf{T}}_{m}, \overline{\mathbf{T}}^{r}_{m}\}$, as depicted in Fig. 3. The mean vector $\overline{\mathbf{T}}_{m}$ is used to represent the mean location characteristics of the m^{th} cluster. The left boundary $\overline{\mathbf{T}}^{l}_{m}$ and the right boundary $\overline{\mathbf{T}}^{r}_{m}$ describe the maximum distance deviations on the left and right of $\overline{\mathbf{T}}_{m}$, respectively. $\overline{\mathbf{T}}^{l}_{m}$ and $\overline{\mathbf{T}}^{r}_{m}$ are determined by connecting the location $(x_k(t), y_k(t))$ that has the largest distance deviation on the left (or right) of $\overline{\mathbf{T}}_{m}$ at each time step. As $\mathbf{T}_k$, $\overline{\mathbf{T}}_{m}$, $\overline{\mathbf{T}}^{l}_{m}$ and $\overline{\mathbf{T}}^{r}_{m}$ are also defined by the set of sequential states, which are given as $\overline{\mathbf{T}}_{m}=\bar{\mathbf{t}}_{m}(n_1,n_2)=\{\bar{\mathbf{r}}_{m}[n]\}$, $\overline{\mathbf{T}}^{l}_{m}=\bar{\mathbf{t}}^{l}_{m}(n_1,n_2)=\{\bar{\mathbf{r}}^{l}_{m}[n]\}$ and $\overline{\mathbf{T}}^{r}_{m}=\bar{\mathbf{t}}^{r}_{m}(n_1,n_2)=\{\bar{\mathbf{r}}^{r}_{m}[n]\}$, respectively, where $n_1 \le n \le n_2$.

3.3 MP Classification

The credibility levels of clustered MPs are determined based on the mass and size information of each MP cluster. For the m^{th} clustered MP C_m, let W_m be the mass value which is defined by the number of trajectories in the MP, and Z_m be the size value which is represented by the distance between $\overline{\mathbf{T}}^{l}_{m}$ and $\overline{\mathbf{T}}^{r}_{m}$ of the MP. We propose a credibility index

$cred_m$ for representing how credible the C_m is for the specific environment, which is calculated as:

$$cred_m = \begin{cases} \dfrac{W_m / Z_m}{\max_m \left(W_m / Z_m \right)} & Z_m > 0 \\ 0 & Z_m = 0 \end{cases}. \tag{2}$$

In this case, a clustered MP with a larger mass value and a smaller size value refers to a more credible MP. Thus, a clustered MP has the highest credibility when it satisfies $cred_m$=1.0, and a clustered MP with a zero credibility index value is least credible. Suppose there are totally M MPs and they are then sorted in a descending order in terms of the credibility index value. Let C_1 denote the clustered MP with the largest credibility index value and C_M denote the clustered MP with the smallest credibility index value. Based on the difference of credibility index values of adjacent clustered MPs, we define a reverse credibility assignment $RCA(C_m)$ for each MP C_m ($1 \le m \le M$) which is detailed described as:

$$RCA(C_{m+1}) = \begin{cases} RCA(C_m) & C\_Diff(C_m, C_{m+1}) < \overline{D_C} \\ RCA(C_m) + 1 & C\_Diff(C_m, C_{m+1}) \ge \overline{D_C} \end{cases}. \tag{3}$$

The initial value $RCA(C_1)$ is 1 which means the top level with the largest credibility index value. $C\_Diff(C_m,C_{m+1})$ is the absolute difference of $cred_m$ and $cred_{m+1}$, and $\overline{D_C}$ is the average value of all $C\_Diff(C_m,C_{m+1})$. From Equation (3), each $RCA(C_{m+1})$ is obtained from the previously known $RCA(C_m)$. The last one $RCA(C_M)$ has the maximal numerical value which means the bottom level with the smallest credibility index value. All resultant $RCA(C_m)$ values are represented by the number as shown in Fig. 4. It is noted that $RCA(C_M)$ also represents the number of credibility levels. In this approach, we do not require the knowledge of the number of credibility levels which can be adaptively and automatically determined.

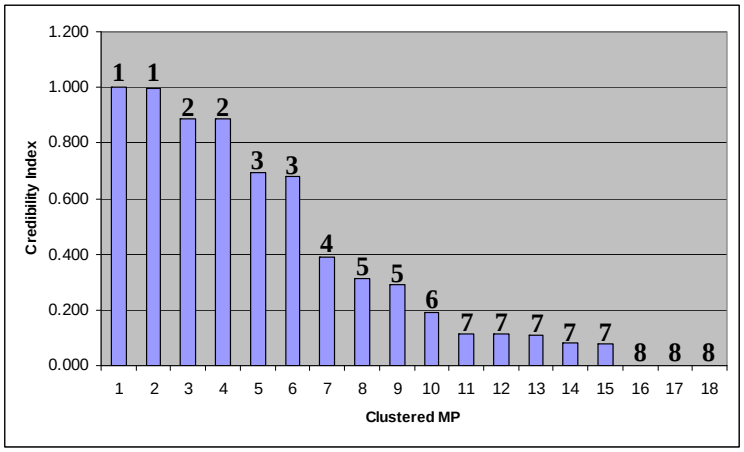

Fig. 4: Reverse credibility assignment of clustered MPs

3.4 Motion Prediction

3.4.1 General Concept

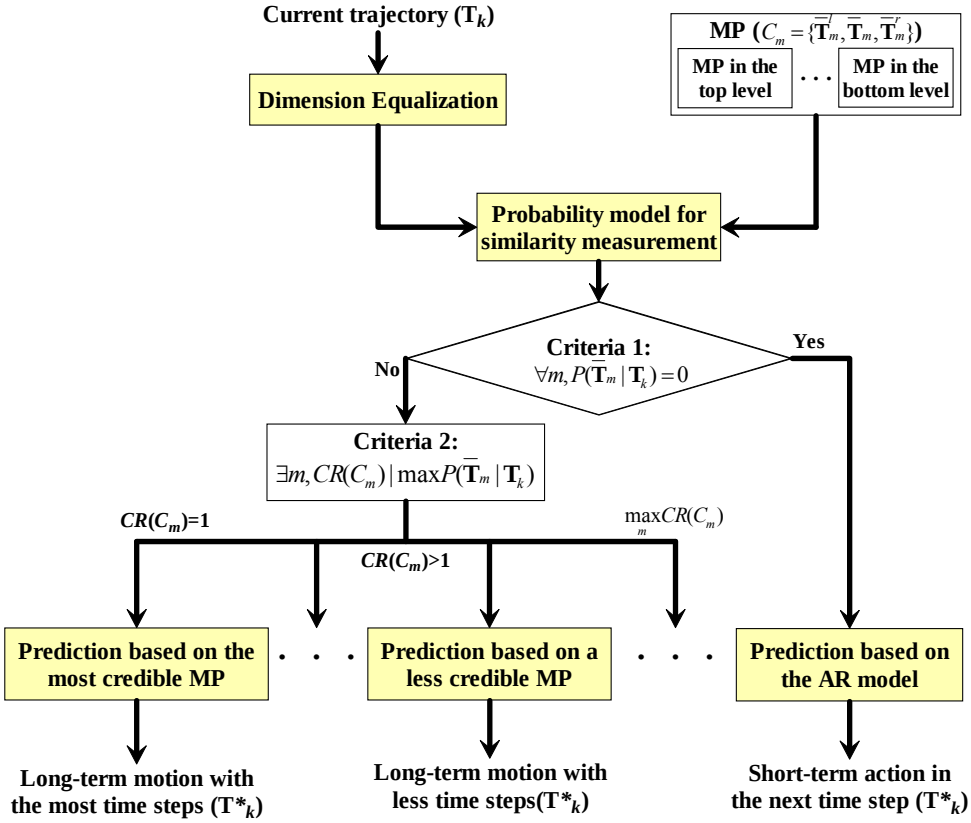

Fig. 5: Algorithmic flow of the adaptive motion prediction model

The focus of the proposed method is to predict human motion in the most appropriate manner based on the classified MPs and the current trajectory, through an adaptive prediction hierarchy as depicted in Fig. 5. $\mathbf{T}_k$ describes the current trajectory, $\bar{\mathbf{T}}_m$ represents the mean

location characteristics of the m^{th} MP. Let $\mathbf{T}^*_k$ denotes the predicted motion of $\mathbf{T}_k$. If $\mathbf{T}_k$ is defined up to t, then $\mathbf{T}^*_k$ is defined from t+1 onward. Suppose there are N observable trajectories and M MPs. $\overline{\mathbf{T}}_m$ needs to be first equalized with the current trajectory $\mathbf{T}_k$ in terms of dimension. Then, both $\mathbf{T}_k$ and the equalized portions of $\overline{\mathbf{T}}_m$ are input to a probability model in which a probability is calculated for measuring the similarity between $\mathbf{T}_k$ and $\overline{\mathbf{T}}_m$. Based on the probability value, we propose two criteria for deciding the number of time steps of the predicted motion. Finally, the corresponding different kind of long-term motion or short-term action $\mathbf{T}^*_k$ is predicted for the current trajectory.

3.4.2 Dimension Equalization

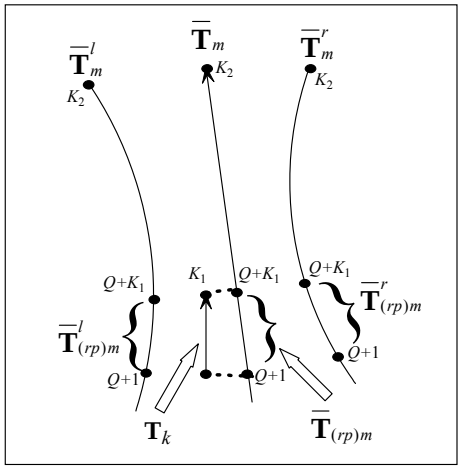

Fig. 6: Dimension equalization (K_2>K_1)

Since the current trajectories and the MPs consist of spatial locations of different number of time steps, before matching is performed, their dimensions need to be equalized. To do that, we first segment $\overline{\mathbf{T}}_m$, $\overline{\mathbf{T}}^l_m$ and $\overline{\mathbf{T}}^r_m$ to the same data dimension as $\mathbf{T}_k$. For example, if $\mathbf{T}_k$ has K_1 time steps and $\overline{\mathbf{T}}_m$ has K_2 time steps (K_2>K_1), as depicted in Fig. 6, we first select the portion on $\overline{\mathbf{T}}_m$ which has the smallest Euclidean distance to $\mathbf{T}_k$ as the representative $\overline{\mathbf{T}}_{(rp)m}$ of the whole $\overline{\mathbf{T}}_m$. Thus $\overline{\mathbf{T}}_{(rp)m}$ can be represented as

$$\overline{\mathbf{T}}_{(rp)m} = \bar{\mathbf{t}}_m(Q+1,Q+K_1)=\{\bar{\mathbf{r}}_m[n]\}, \quad 0\le Q\le K_2-K_1,\ Q+1\le n\le Q+K_1, \tag{4}$$

where Q+1 is the corresponding time step on $\overline{\mathbf{T}}_m$ that marks where $\overline{\mathbf{T}}_{(rp)m}$ starts from. Then we obtain the representatives $\overline{\mathbf{T}}^l_{(rp)m}$ and $\overline{\mathbf{T}}^r_{(rp)m}$ of the whole $\overline{\mathbf{T}}^l_m$ and $\overline{\mathbf{T}}^r_m$ by selecting the portions on $\overline{\mathbf{T}}^l_m$ and $\overline{\mathbf{T}}^r_m$ from the same starting to ending time step as $\overline{\mathbf{T}}_m$. So $\overline{\mathbf{T}}^l_{(rp)m}$ and $\overline{\mathbf{T}}^r_{(rp)m}$ can be similarly described as

$$\overline{\mathbf{T}}^l_{(rp)m} = \overline{\mathbf{t}}^l_m(Q+1,Q+K_1)=\{\overline{\mathbf{r}}^l_m[n]\}, \quad 0\le Q\le K_2-K_1,\ Q+1\le n\le Q+K_1,$$

$$\overline{\mathbf{T}}^r_{(rp)m} = \overline{\mathbf{t}}^r_m(Q+1,Q+K_1)=\{\overline{\mathbf{r}}^r_m[n]\}, \quad 0\le Q\le K_2-K_1,\ Q+1\le n\le Q+K_1. \tag{5}$$

After dimension equalization, $\{\overline{\mathbf{T}}^l_{(rp)m}, \overline{\mathbf{T}}_{(rp)m}, \overline{\mathbf{T}}^r_{(rp)m}\}$ is used in the following prediction step.

3.4.3 Probability Model for Similarity Measurement

A Bayesian probability model is proposed for similarity measurement between the current trajectory and the MP cluster. In the Bayes model, a posterior probability $P(\overline{\mathbf{T}}_m|\mathbf{T}_k)$ is calculated, which is the probability of the MP cluster $\overline{\mathbf{T}}_m$ that being followed by the current trajectory $\mathbf{T}_k$. $P(\overline{\mathbf{T}}_m|\mathbf{T}_k)$ is given by

$$P(\overline{\mathbf{T}}_m \mid \mathbf{T}_k) = \frac{p(\mathbf{T}_k \mid \overline{\mathbf{T}}_m)\times P(\overline{\mathbf{T}}_m)}{\sum_{m=1}^{M} p(\mathbf{T}_k \mid \overline{\mathbf{T}}_m)\times P(\overline{\mathbf{T}}_m)}. \tag{6}$$

The calculation of $P(\overline{\mathbf{T}}_m|\mathbf{T}_k)$ is based on the prior probability $P(\overline{\mathbf{T}}_m)$ for $\overline{\mathbf{T}}_m$, the conditional probability density $p(\mathbf{T}_k|\overline{\mathbf{T}}_m)$ which is the likelihood of $\overline{\mathbf{T}}_m$ with respect to $\mathbf{T}_k$, and the evidence factor $p(\mathbf{T}_k)=\sum_{m=1}^{M} p(\mathbf{T}_k \mid \overline{\mathbf{T}}_m)\times P(\overline{\mathbf{T}}_m)$ which can be viewed as merely a scale factor that guarantees that the posterior probabilities sum to 1.

The prior probability $P(\overline{\mathbf{T}}_m)$ is calculated based on the mass W_m which is given as

$$L(\overline{\mathbf{T}}_m) = \frac{W_m}{\sum_{n=1}^{M} W_n}, \tag{7}$$

$$P(\overline{\mathbf{T}}_m) = \frac{L(\overline{\mathbf{T}}_m)}{\sum_{n=1}^{M} L(\overline{\mathbf{T}}_n)}. \tag{8}$$

In (7), $L(\overline{\mathbf{T}}_m)$ denotes the likelihood of $\overline{\mathbf{T}}_m$ based on considering both the likelihood of $\overline{\mathbf{T}}_m$ among all the MPs. In (8), $P(\overline{\mathbf{T}}_m)$ is obtained by a normalization operation based on the likelihood value $L(\overline{\mathbf{T}}_m)$ thus the summation of all prior probability values is 1.

The conditional probability density $p(\mathbf{T}_k|\overline{\mathbf{T}}_m)$ represents the likelihood of $\mathbf{T}_k$ belonging to $\overline{\mathbf{T}}_m$. $p(\mathbf{T}_k|\overline{\mathbf{T}}_m)$ is calculated as

$$p(\mathbf{T}_k \mid \overline{\mathbf{T}}_m) = p(d, \Delta\theta \mid \overline{\mathbf{T}}_m) = p(d \mid \overline{\mathbf{T}}_m) \cdot p(\Delta\theta \mid \overline{\mathbf{T}}_m). \tag{9}$$

The calculation of $p(\mathbf{T}_k|\overline{\mathbf{T}}_m)$ is based on the product of two independent conditional probability density functions: $p(d|\overline{\mathbf{T}}_m)$ and $p(\Delta\theta|\overline{\mathbf{T}}_m)$. $p(d|\overline{\mathbf{T}}_m)$ is the probability density function for d given $\overline{\mathbf{T}}_m$ in which d means the distance between $\mathbf{T}_k$ and $\overline{\mathbf{T}}_m$. d is calculated as

$$d = \sum_{i=1}^{K_1} \left(a_i \times D(\mathbf{r}_k[i], \bar{\mathbf{r}}_m[Q+i]) \right), \tag{10}$$

$$a_i = \frac{i}{\left(K_1 \times (K_1+1)\right)/2}, \tag{11}$$

where $D(\mathbf{r}_k[i], \bar{\mathbf{r}}_m[Q+i])$ refers to the Euclidean distance between the corresponding coordinate pair $\mathbf{r}_k[i]$ on $\mathbf{T}_k$ and $\bar{\mathbf{r}}_m[Q+i]$ on $\overline{\mathbf{T}}_m$, and a_i is a weight factor for each time step, which means an "older" time step has less impact when matching is performed, and K_1 is the time steps of $\mathbf{T}_k$. $p(d|\overline{\mathbf{T}}_m)$ is calculated under the condition that $d<d^B$, in which d^B is defined as the acceptable distance range which is obtained based on the left and right boundary of $\overline{\mathbf{T}}_m$

$$d^B = \sum_{i=1}^{K_1} \left(a_i \times \max\left\{ D(\bar{\mathbf{r}}_m^l[Q+i], \bar{\mathbf{r}}_m[Q+i]), D(\bar{\mathbf{r}}_m^r[Q+i], \bar{\mathbf{r}}_m[Q+i]) \right\} \right), \tag{12}$$

where a_i is the same weight factor calculated by (11), $D(\bar{\mathbf{r}}_m^l[Q+i], \bar{\mathbf{r}}_m[Q+i])$ and $D(\bar{\mathbf{r}}_m^r[Q+i], \bar{\mathbf{r}}_m[Q+i])$ refer to the Euclidean distances between $\bar{\mathbf{r}}_m^l[Q+i]$ on $\overline{\mathbf{T}}_m^l$ and $\bar{\mathbf{r}}_m[Q+i]$ on

$\overline{\mathbf{T}}_m$, and $\bar{\mathbf{r}}_m^r[Q+i]$ on $\overline{\mathbf{T}}_m^r$ and $\bar{\mathbf{r}}_m[Q+i]$ on $\overline{\mathbf{T}}_m$, respectively. We regard MP($\{\overline{\mathbf{T}}_m^l, \overline{\mathbf{T}}_m, \overline{\mathbf{T}}_m^r\}$) as a Gaussian distribution model where the *Mean* locates at $\overline{\mathbf{T}}_m$, From $\overline{\mathbf{T}}_m$ to $\overline{\mathbf{T}}_m^l$ (or $\overline{\mathbf{T}}_m^r$), a larger distance of $\mathbf{T}_k$ away from $\overline{\mathbf{T}}_m$ means a less likely matching between $\mathbf{T}_k$ and $\overline{\mathbf{T}}_m$. If $\mathbf{T}_k$ goes outside of $\overline{\mathbf{T}}_m^l$ or $\overline{\mathbf{T}}_m^r$, the matching fails as depicted in Fig. 7. From the symmetry attribute of the Gaussian model, the larger value of either $D(\bar{\mathbf{r}}_m^l[Q+i], \bar{\mathbf{r}}_m[Q+i])$ or $D(\bar{\mathbf{r}}_m^r[Q+i], \bar{\mathbf{r}}_m[Q+i])$ is selected to be $3\sigma_i$ for representing the maximal match-able range at each time step. The function $p(d|\overline{T}_m)$ is given as

$$p(d \mid \overline{\mathbf{T}}_m) = \begin{cases} \sum_{i=1}^{K_1} \left(a_i \times \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{1}{2} \left(\frac{D(\mathbf{r}_k[i], \bar{\mathbf{r}}_m[Q+i])}{\sigma_i} \right)^2 \right] \right), & d < d^B \\ 0, & d \geq d^B \end{cases}. \quad (13)$$

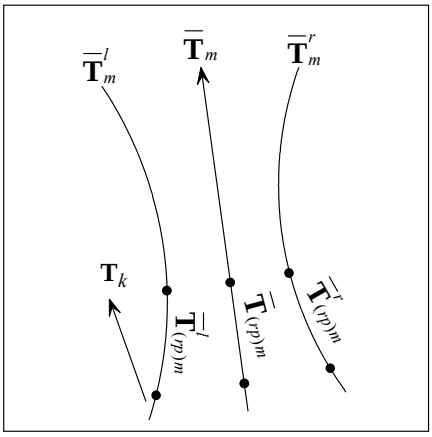

Fig. 7: Failed matching case for $p(d|\overline{\mathbf{T}}_m) = 0$, where $\overline{\mathbf{T}}_{(rp)m}$ is the representative of whole $\overline{\mathbf{T}}_m$ for matching with $\mathbf{T}_k$, and $\overline{\mathbf{T}}_{(rp)m}^l$ and $\overline{\mathbf{T}}_{(rp)m}^r$ are defined similarly

The other conditional probability density $p(\Delta\theta|\overline{\mathbf{T}}_m)$, which is considered for the calculation of $p(\mathbf{T}_k|\overline{\mathbf{T}}_m)$, is the probability density function for $\Delta\theta$ given $\overline{\mathbf{T}}_m$ in which $\Delta\theta$ depicts the changing angle from $\mathbf{T}_k$ to $\overline{\mathbf{T}}_m$ at the last time step of $\mathbf{T}_k$ that the prediction is performed. It is believed that a smaller changing angle means a higher similarity between $\mathbf{T}_k$ and $\overline{\mathbf{T}}_m$ since there is less change in motion direction. The calculation of $p(\Delta\theta|\overline{\mathbf{T}}_m)$ is given by

$$p(\Delta\theta \mid \overline{\mathbf{T}}_m) = \begin{cases} \frac{2}{\Delta\theta^{(\max)}} \times \left(1 - \frac{\Delta\theta}{\Delta\theta^{(\max)}} \right), & \Delta\theta \leq \Delta\theta^{(\max)} \\ 0, & \Delta\theta > \Delta\theta^{(\max)} \end{cases}, \quad (14)$$

$$\Delta\theta^{(\max)} = \arg\max_{1\le i\le K_1-1} \Delta\theta_i , \tag{15}$$

where $\Delta\theta^{(\max)}$ depicts the maximal changing angle in the historical time steps of $\mathbf{T}_k$. Here, $\frac{2}{\Delta\theta^{(\max)}}$ is a scale factor that guarantees that the integral of the density over the set is 1.

Based on this probability model for adaptive prediction, we quantify the posterior probability that $\mathbf{T}_k$ follows each $\overline{\mathbf{T}}_m$ and then propose the criteria for deciding the number of time steps which is to be predicted for $\mathbf{T}_k$. As presented in Fig. 5, if $P(\overline{\mathbf{T}}_m|\mathbf{T}_k)=0$ is satisfied for any $\overline{\mathbf{T}}_m$, which means the current trajectory $\mathbf{T}_k$ fails to match with all existing MPs, $\mathbf{T}_k$ is performed short-term prediction for one single time step action. Otherwise, $\mathbf{T}_k$ is predicted for a long-term motion over a number of future time steps.

3.4.4 Adaptive Prediction

For $\mathbf{T}_k$ with the matched $\overline{\mathbf{T}}_m$, a long-term future motion is predicted based on $\overline{\mathbf{T}}_m$. To do this, we first calculate the number of time steps of the predicted motion $\mathbf{T}^*_k$. Let S denote the corresponding time step of $\overline{\mathbf{T}}_m$ which is closest to the last time step (K_1^{th} time step) of $\mathbf{T}_k$ for performing prediction, and K_2 denotes the total number of time steps of $\overline{\mathbf{T}}_m$. The number of the predicted time steps $NPM(\mathbf{T}^*_k)$ is adaptively calculated in terms of the reverse credibility assignment $RCA(C_m)$ ($C_m=\{\overline{\mathbf{T}}^l_m, \overline{\mathbf{T}}_m, \overline{\mathbf{T}}^r_m\}$) which is given as:

$$NPM(\mathbf{T}^*_k) = (K_2 - S)\times\left[\left(1-\frac{RCA(C_m)-1}{RCA(C_M)}\right)\times 100\%\right]. \tag{16}$$

Here, (K_2-S) represents the possible maximal time steps of $\mathbf{T}^*_k$. $\left(1-\frac{RCA(C_m)-1}{RCA(C_M)}\right)$ denotes a confidence coefficient for prediction based on $RCA(C_m)$. If $RCA(C_m)=1$, it can be determined from Equation (16) that the corresponding confidence coefficient is 1, and that the exactly maximal (K_2-S) time steps are predicted. Along with an increasing $RCA(C_m)$ value referring to a descending credibility level of C_m, the corresponding confidence coefficient also decreases

and less future time steps are predicted for $\mathbf{T}_k$. If the matched C_m for $\mathbf{T}_k$ has the maximal $RCA(C_m)(=RCA(C_M))$ value which means a lowest credibility level, a long-term future motion with the least time steps will be predicted for $\mathbf{T}_k$. Based on the number of time steps determined for the predicted motion, $\mathbf{T}^*_k$ can be represented as

$$\mathbf{T}^*_k=\mathbf{t}^*_k(K_1+1,\ K_1+NPM(\mathbf{T}^*_k))=\{\mathbf{r}^*_k[n]\},K_1+1\le n\le K_1+NPM(\mathbf{T}^*_k), \tag{17}$$

where $\mathbf{r}^*_k[n]$ means the predicted spatial location of $\mathbf{T}_k$ at each time step after K_1, which is defined as

$$\mathbf{r}^*_k[n] = \bar{\mathbf{r}}_{m_1}[S+n-K_1] + (\mathbf{r}_k[K_1] - \bar{\mathbf{r}}_{m_1}[S]),\ K_1+1\le n\le K_1+NPM(\mathbf{T}^*_k). \tag{18}$$

When $\mathbf{T}_k$ is not matched with any existing MP, a single time step action is predicted, which is achieved by the following equation

$$\mathbf{w}(t+1) = \mathbf{w}(t) + \mathbf{v}(t)T^s + B_t\mathbf{a}(t)T^{s^2}, \tag{19}$$

where $\mathbf{w}(t)$ means the position at time step t, and $\mathbf{v}(t)$ and $\mathbf{a}(t)$ are the corresponding velocity and acceleration values, respectively. B_t is time-dependent and is updated by the adaptive algorithm in [25].

IV. Validation of the Proposed Method

In order to validate the proposed method, we have conducted a number of experiments in several simulated scenes and real-world scenes. In this section, we present the results of these experiments to demonstrate the performance of the proposed method. First of all, a scene of people walking in a simulated shopping mall [28] is depicted in Fig. 8, in which the numbers label the entrances/exits in the scene. The training trajectories (in grey) for MP clustering and the learned MPs (in red) are shown in Fig. 9 where a red solid-curve and a red dot-curve are used for differentiating double-directional trajectories in the scene, and multi-level prediction results for the current trajectories (none of them were used for training) are shown in Fig. 10. 8 trajectories were predicted at different levels (in blue), since they match with MPs at

different credibility levels. Fig. 11 depicts the predicted motions and actual motions. It can be seen that the proposed method is reasonably effective in deploying adaptive predictions for different trajectories and the predicted motions (in blue) are consistent with the actual motions (in magenta). We have conducted 5 separated experiments in this scene and analyzed the deviation rate d_r of prediction by calculating the ratio between the deviated distance of the predicted destination and the actual destination, and the actual total traversed distance. As depicted in Table I, d_r_{min} , d_r_{max} and d_r_{avg} of each experiment are listed, in which d_r_{min} and d_r_{max} mean the minimal and the maximal deviation rate, respectively, and d_r_{avg} means the average deviation rate for all. It is found that the average deviation rate for each experiment is around 8% and the prediction accuracy can be considered as acceptable.

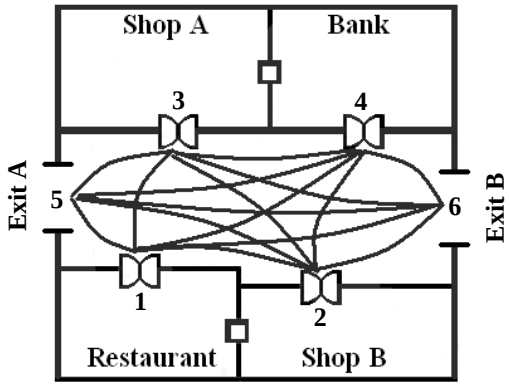

Fig. 8: Simulated scene

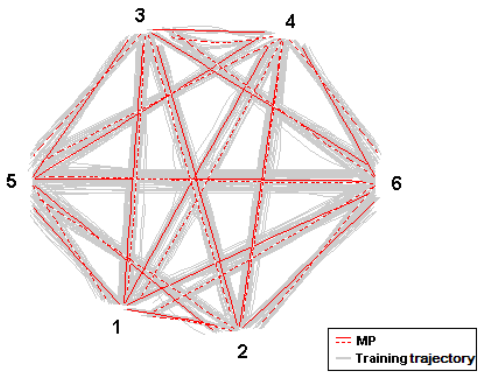

Fig. 9: Training trajectories and clustered MP

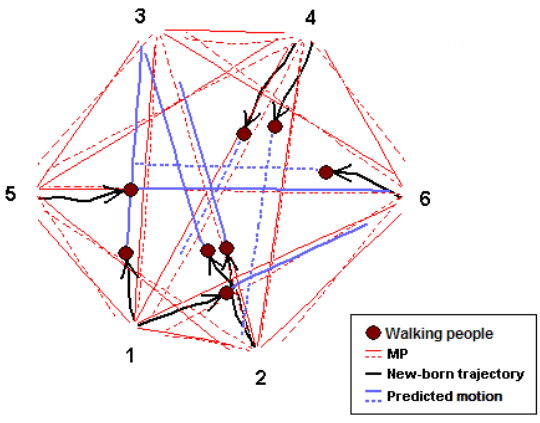

Fig. 10: Prediction results

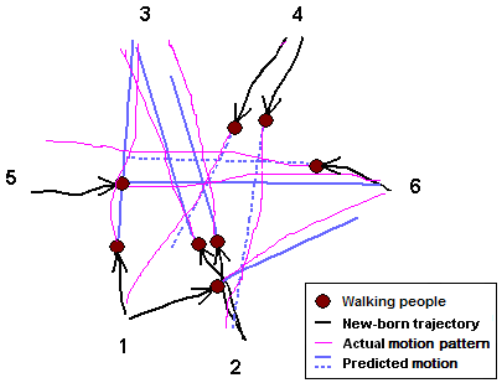

Fig. 11: Predicted and actual motions

TABLE I
DEVIATION RATE OF PREDICTION FOR SIMULATION

No.	Number of humans for prediction	Minimal deviation rate d_r_{min}	Maximal deviation rate d_r_{max}	Average deviation rate d_r_{avg}
1	8	5.693%	10.127%	8.296%
2	12	5.372%	10.009%	8.192%
3	9	5.921%	10.426%	8.397%
4	7	4.925%	9.481%	7.975%
5	13	5.097%	9.659%	8.022%

Obviously, simulations do not necessarily indicate how the proposed method works in real-world. For this reason, we captured a sequence of frames from a video camera looking down a shopping mall with people walking freely without any defined or agreed trajectories as shown in Fig. 12. There are several entrances and exits to the scene as depicted: entrance/exit '1' connects to some shops; entrances/exit '2' connects to neighboring buildings; entrance/exit '3' connects to up/down escalators. The frames were extracted at a rate of 25 frames per second and a total of 19786 frames were obtained. Through Trajectory Extraction described in Section 3.1, a total of 326 observable human trajectories were extracted. Some of the extracted trajectories are depicted in Fig. 13, each of which is represented by a series of discrete positions which were recorded at sampled time steps. In Fig. 13(a), it could be seen that some humans shown in the top and bottom regions of the frame are not detected because of their incompleteness. They are not included in this frame, but may subsequently be considered if their view is improved. The false positive and negative rates of human detection are 12.1% and 15.8% respectively, and the accuracy rate of data association between frames is 79.5% [22].

Fig. 12: A frame of the scenario


(a) (b)

Fig. 13: Extracted trajectories (a) Human detection results (b) Data association results


Fig. 14: Trajectories for MP clustering


Fig. 15: Clustered MP


(a) (b)

Fig. 16: Classified MPs at different credibility levels (a) higher level (b) lower level

Out of all extracted trajectories, we randomly select 296 trajectories for MP clustering and leave the remaining 30 trajectories for testing the prediction performance. Fig. 14 illustrates the selected 296 extracted trajectories, in which red curves and green curves represent bi-directional trajectories between each pair of entrance and exit. There are altogether 20 MPs clustered as a result from MP Clustering (Fig. 15), in which the arrows describe motion directions of the clustered MPs. These 20 clustered MPs are classified into 8 credibility levels. Based on classification criteria of the clustered MPs as described in Section 3.3, it should be noted that compared with the MPs at a lower credibility level, the MPs at a higher credibility level are more reliable because they are clustered from more collective trajectories, as shown in Fig. 16, in which green solid-lines represent the trajectories, and magenta and blue solid-lines represent MPs in higher and lower credibility levels, respectively. In addition, the left and right bounds of the MPs are also shown for illustration in Fig. 16.

(a) (b)

(c)

Fig. 17: Adaptive prediction results (a) long-term motion with more time steps (b) long-term motion with less time steps (c) single time step action

To further illustrate how prediction is adaptively made, we use the classified MPs for prediction on the remaining 30 trajectories. Fig. 17 depicts some adaptive prediction results in which the long-term motions are predicted based on the MPs at different credibility levels as shown in Fig. 16, respectively. In Fig. 17(a) and (b), the corresponding predicted long-term motions with more time steps and less time steps are shown by green lines. Black lines represent the actual motions and black dots denote the time step when the predictions were performed. Fig. 17(c) depicts the predicted single time step action by the short green lines.

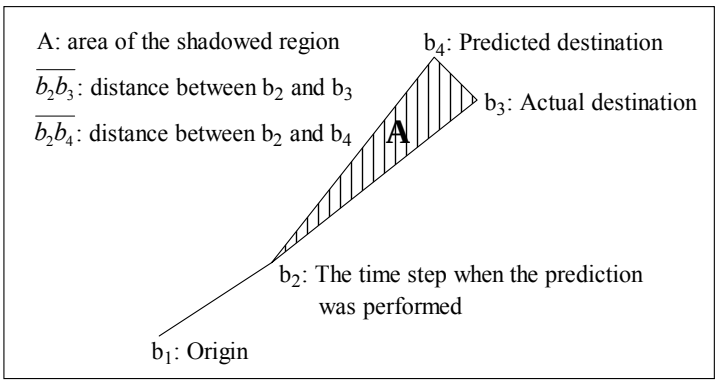

Fig. 18: Calculation of prediction error

In order to evaluate the performance of the proposed method, we quantitatively compare the predicted motion of each trajectory with the corresponding actual motion. For the predicted motion $\mathbf{T}^*_k$ of each $\mathbf{T}_k$ at the time step t, we calculate an absolute error $e_{k(t)}$, which is defined as the deviated distance between the predicted motion and the actual motion after time step t. Fig. 18 illustrates how the absolute prediction error is calculated, which is given as

$$e_{k(t)} = \frac{A}{0.5 \times (\overline{b_2b_3} + \overline{b_2b_4})}, \tag{20}$$

where $\overline{b_2b_3}$ and $\overline{b_2b_4}$ are the actual and predicted traversed distances, respectively, and A represents the area of the region between the actual motion and the predicted motion. In order to work out an overall prediction error for each trajectory, we calculate a series of $e_{k(t)}$, to generate a global prediction error ε_k of $\mathbf{T}^*_k$ at all possible time steps t when a prediction can be performed. t is set from 3 because changing angle information at historical time steps is

necessary for prediction and there is no changing angle information before 3 time steps. The calculation of ε_k is performed as

$$\varepsilon_k = \frac{\sum_{t=3}^{U-1} e_{k(t)}}{U-3}, \tag{21}$$

where U is the total number of time steps of the human trajectory from the origin to the destination.

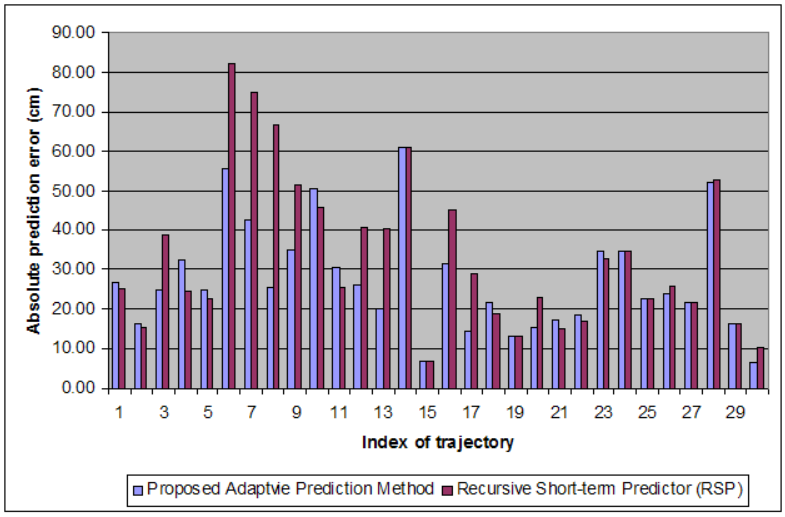

Fig. 19: Absolute prediction error of the proposed method and RSP for 30 testing trajectories

Besides calculating the absolute prediction error of the proposed adaptive prediction method, we also calculate the absolute prediction error of the AR method for comparison. Since the AR method is for short-term prediction of one time-step, we apply it recursively, which results in a predicted MP with a number of future time steps. The recursively applied AR model is called the Recursive Short-term Predictor (RSP). It should be noted that other short-term predictors can replace the AR predictor just as well. Fig. 19 depicts absolute prediction errors of all 30 trajectories by using the proposed method and the RSP. For long-term future motion predicted by the proposed method, the RSP generates a MP with the same number of future time steps for comparison. From Fig. 19, it can be seen that (1) the proposed method has lower absolute prediction errors in more trajectories than otherwise; (2) the proposed method produces significantly less errors in better-performed cases while produces

slightly more errors in worse-performed cases; (3) the proposed method produces a smaller sum error. In general, the proposed method improved 17.73% over the RSP in prediction of all 30 trajectories.

(a) (b)

Fig. 20: Prediction results of No.8 trajectory at the 8th time step (a) proposed method (b) RSP

(a) (b)

Fig. 21: Prediction results of No.11 trajectory at the 7th time step (a) proposed method (b) RSP

For illustration purpose, the best case and the worst case predictions are depicted in Fig. 20 and Fig. 21 for comparison, in which blue lines represent the MPs, and green lines and black lines represent the predicted and actual motions, respectively. In the best case in Fig. 20 where the largest improvement is obtained by the proposed method, compared with the result of the RSP, the prediction was made at the time step when the trajectory started to change direction. The RSP (Fig. 20(b)) had worse performance because its predicted motion would keep the previous direction while the proposed method (Fig. 20(a)) made a better prediction based on the MP. In the worst case in Fig. 21, the current trajectory was acceptably similar to the matched MP, but certain deviation is also evident. As such, the proposed method (Fig. 21(a)) produced a larger error while the RSP (Fig. 21(b)) produced a smaller error because the

trajectory had very little change in direction. Clearly, the proposed method performs better than the RSP when the trajectory changes direction frequently.

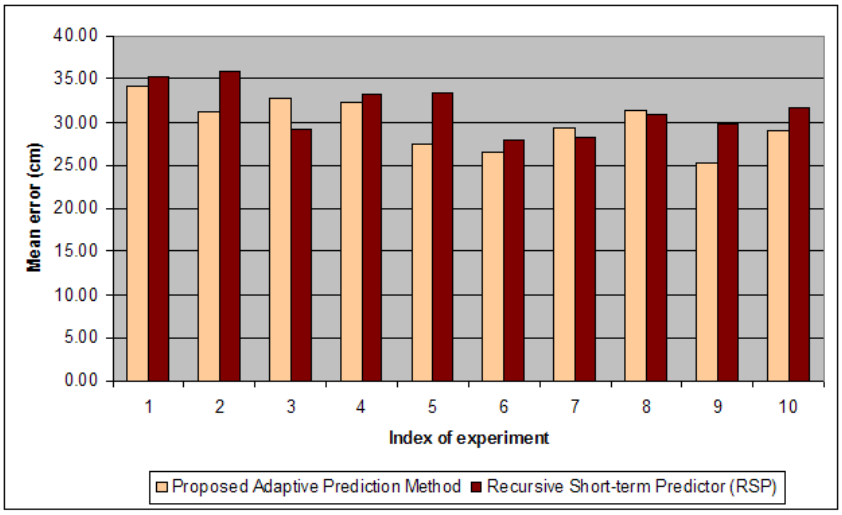

Fig. 22: Mean error of the proposed method and RSP for 10 real experiments

In order to further test whether the proposed method can be reliably applied in different cases, we perform a number of experiments by orderly dividing the 326 extracted trajectories into groups of 296 and 30 trajectories according to a randomly assigned index to each trajectory. There are totally 10 sets of data altogether. In each set, the large group was used for analysis while the small group was used for testing. After the prediction, we calculate the average of the prediction error of all trajectories in each experiment as a mean error. In Fig. 22, we plot the average prediction error of all 10 experiments. From Fig. 22, we can see that the proposed method performed better than the RSP in 7 out of 10 experiments. On average, the proposed method has a 5% improvement over the RSP in prediction accuracy. Fig. 23 depicts the curvature of all 326 extracted trajectories. The curvature of each trajectory is calculated as the ratio between the total traversed distance from the origin to the destination and the straight line connecting the origin to the destination. We can see that most trajectories are very well-defined since most curvature values in the distribution are close to 1. In other words, most of the extracted trajectories do not change direction frequently. In fact, this is in

favor of RSP rather than the proposed method. If the trajectories change directions frequently, we would expect an even larger improvement by the proposed method.

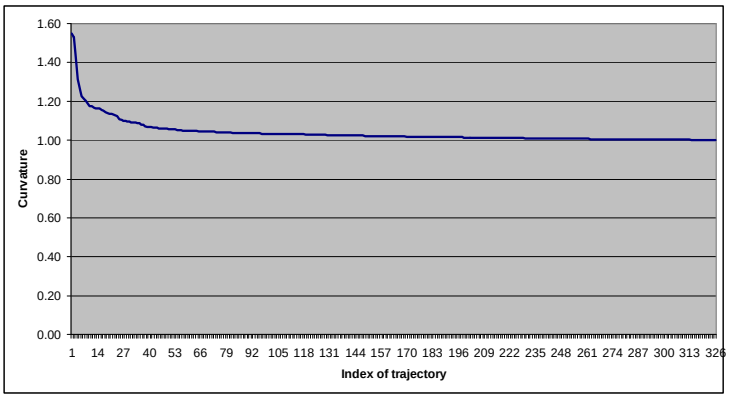

Fig. 23: Curvature of extracted trajectories

Besides comparing the performance and robustness of the proposed method with the RSP, we further compare the proposed method with another related method. In the method presented in [14], spatial and probabilistic models are built sequentially and the learned models are used for trajectory labeling and atypical behavior detection in an automatic video surveillance system. It is noted that different experiment results are generated since the proposed method in this paper focuses more on human motion prediction based on learning and classifying MPs. Thus, a directly quantitative analysis in experiment results may not be feasible. As a result, we have conducted a qualitative comparison and three issues are identified which imply that the proposed method in this paper is superior to the method in [14]. First, the proposed method in this paper learns MPs and further classifies MPs into different credibility levels by analyzing characteristic of each MP, while the method in [14] only learns MPs and applies the learned MPs equally without knowing their credibility. Second, the method in [14] performs behavior recognition by classifying the new trajectory into one of the learned MP. The proposed method in this paper not only classifies the new trajectory into some existing MP but also predicts long-term future motion for incomplete trajectory based on the MP into which the trajectory is classified. Third, the proposed method has been evaluated both qualitatively and quantitatively. The reliability of the proposed

method for different cases has also been evaluated by experiments. However, only one experimental case is presented in [14], and the results are only qualitatively described without any quantitative evaluation.

V. CONCLUSIONS

In this paper, we present a novel adaptive human motion analysis and prediction method. The proposed method clusters MPs based on observable trajectories extracted from a series of frames of a taken video and then classifies them into different credibility levels according to their mass and size information. It then predicts future motion based on matching current trajectories to classified MPs. The main contributions of the proposed method are that (1) it offers a viable approach for analyzing and adaptively predicting human motions of different number of time steps; (2) it provides a more comprehensive description of MPs including not only the learned MPs but also their evaluated credibility; (3) it adaptively makes long-term predictions of human future motion according to the credibility of the learned MPs. From the experiments based on simulated and real-world data, it can be concluded that the proposed method is effective in performing adaptive predictions for different future motions. For the proposed method, our future research work will focus on three aspects: (1) to investigate online learning of MPs for discerning the change of MPs more accurately; (2) to research the fusion of multiple feature dimensions for generating an integrated prediction result for human future behavior or intention; (3) to analyze human predicted behavior for incident analysis or anomaly detection in crowd surveillance.

ACKNOWLEDGEMENT

The work on MP prediction was supported in part by a grant from the Research Grant Council of the Hong Kong Special Administrative Region, China, under Project HKU719406E and in part by the Postgraduate Studentship of The University of Hong Kong. The work on trajectory extraction was supported in part by a grant from the Research Grant

Council of the Hong Kong Special Administrative Region, China, under Project HKU719608E and in part by the Postgraduate Studentship of the University of Hong Kong.

REFERENCES

[1] D. Vasquez, T. Fraichard and C. Laugier, "Incremental learning of statistical motion patterns with growing hidden Markov models," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 10, No. 3, pp. 403-416, September 2009.

[2] V.S. Rajpurohit and M.M.M. Pai, "An Optimized Fuzzy Based Short Term Object Motion Prediction for Real-life Robot Navigation Environment," *Proceedings of the 10th International Conference on Visual Information Systems: Web-Based Visual Information Search and Management.* Salerno, Italy, pp. 114-125, 2008.

[3] J. Tani, "Model-based learning for mobile robot navigation from the dynamical systems perspective," *IEEE Transactions on Systems, Man and Cybernetics - Part B*, Vol. 26, No. 3, pp. 421-436, June 1996.

[4] Q. Zhu, "Hidden Markov model for dynamic obstacle avoidance of mobile robot navigation," *IEEE Transactions on Robotics and Automation*, Vol. 7, No. 3, pp. 390-397, 1991.

[5] R. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME – Journal of Basic Engineering*, No. 82 (Series D), pp. 35-45, 1960.

[6] A. Elnagar, "Prediction of moving objects in dynamic environments using Kalman Filters," *Proceedings of IEEE International Symposium on Computational Intelligence in Robotics and Automation*. Banff, Alberta, Canada. pp. 414-419, July 29-August 1, 2001.

[7] E. D. Dickmanns, B. Mysliwetz and T. Christians, "An integrated spatio-temporal approach to automatic visual guidance of autonomous vehicles," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 20, No. 6, pp. 1273-1284, Nov./Dec. 1990.

[8] A. Chakravarthy and D. Ghose, "Pedestrian avoidance in a dynamic environment: A collision cone approach," *IEEE Transactions on Systems, Man and Cybernetics - Part B*, Vol. 28, No. 5, pp. 562-574, 1998.

[9] Z. Qu, J. Wang and C. E. Plaisted, "A new analytical solution to mobile robot trajectory generation in the presence of moving pedestrians," *IEEE Transactions on Robotics*, Vol. 20, No. 6, pp. 978-993, December 2004.

[10] D. Vasquez, T. Fraichard, O. Aycard and C. Laugier, "Intentional Motion On-line Learning and Prediction," *Machine vision and applications*, Vol. 19 Issue 5/6, pp. 411-425, December 2008.

[11] A.F. Foka and P.E. Trahanias, "Predictive autonomous robot navigation," *Proceedings of the 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems*. EPFL, Lausanne, Switzerland, pp. 490-495, October 2002.

[12] M. Bennewitz, W. Burgard, G. Cielniak and S. Thrun, "Learning Motion Patterns of People for Compliant Robot Motion," *The International Journal of Robotics Research*, Vol. 24, No. 1, pp. 31-48, January 2005.

[13] W. Hu, X.Xiao, Z. Fu, D. Xie, T. Tan and S. Maybank, "A system for learning statistical motion patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 9, pp. 1450-1464, September 2006.

[14] D. Markris and T. Ellis, "Spatial and probabilistic modeling of pedestrian behavior," *Proceedings of the 13th British Machine Vision Conference*. Cardiff, UK, pp. 557-566, September 2002.

[15] D. Biliotti, G. Anotonini and J.P. Thiran, "Multi-layer hierarchical clustering of pedestrian trajectories for automatic counting of people in video sequences," *Proceedings of IEEE Workshop on Motion and Video Computing*. Brekenridge, Colorado, U.S.A., pp. 50-57, January 2005.

[16] I. Junejo, O. Javed and M. Shah, "Multi feature path modeling for video surveillance," *Proceedings of the 17th International Conference on Pattern Recognition*, Cambridge, UK, pp. 716-719, August 2004.

[17] X. Li, W. M. Hu and W. Hu, "A coarse-to-fine strategy for vehicle motion trajectory clustering," *Proceedings of the International Conference on Pattern Recognition*, Hong Kong, pp. 591-594, August 2006.

[18] W.M. Hu, Z.Y. Fu, W.R. Zeng and S. Maybank, "Semantic-based surveillance video," *IEEE Transaction on Image Processing*, Vol. 16, No. 4, pp. 1168-1181, March 2007.

[19] X.G. Wang, K. Tieu and E. Grimson, "Learning semantic scene models by trajectory analysis," *ECCV 2006*, Part III, LNCS 3953, pp. 110–123, 2006.

[20] L. Wang and N. H. C. Yung, "Crowd Counting and Segmentation in Visual Surveillance," *Proceedings of International Conference on Image Processing*, Cairo, Egypt, pp. 2573-2576, November, 2009.

[21] L. Wang and N. H. C. Yung, "Extraction of moving objects from their background based on multiple adaptive thresholds and boundary evaluation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 1, pp 40-51, 2010.
[22] L. Wang and N. H. C. Yung, "Detection based low frame rate human tracking," *Proceedings of International Conference on Pattern Recognition* (ICPR2010), Istanbul, Turkey, pp. 3529-3532, 2010.
[23] N. H. C. Yung and A. H. S. Lai, "Segmentation of color images based on the gravitational clustering concept," *Optical Engineering*, Vol. 37, No. 3, pp. 989-1000, March 1998.
[24] B. Morris and M. Trivedi, "Learning Trajectory Patterns by Clustering: experimental Studies and Comparative Evaluation," *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. Maimi, Florida, pp. 312-319, June 2009.
[25] Ye Cang, "Behavior-Based Fuzzy Navigation of Mobile Vehicle in Unknown and Dynamically Changing Environment," *PhD thesis*. September 1999.
[26] D. Martin, C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 26, no. 5, pp. 530-549, 2004.
[27] C. Huang, B. Wu and R. Nevatia, "Robust object tracking by hierarchical association of detection responses," *Proceedings of European Conference of Computer Vision*, pp. 788-801, 2008.
[28] Z. Chen, D. C. K. Ngai and N. H. C. Yung, "Behavior prediction based on obstacle motion patterns in dynamically changing environments," *Proceeding of 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, Volume 2, Sydney, Australia, pp. 132-135, December 2008.