Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

A probabilistic approach to incorporating domain knowledge for closed-room people monitoring

Ji Tao, Yap-Peng Tan*

*Division of Information Engineering, School of Electrical & Electronic Engineering,
Block S1, Nanyang Avenue, Nanyang Technological University, Singapore*

Received 28 February 2004

Abstract

We propose a novel probabilistic approach to recognizing people entering and leaving a closed room in human work place or living environment. Specifically, people in the view of a monitoring camera are first tracked and represented using low-level color features. Based on a new color similarity measure, optimal recognition of people leaving and entering the room is carried out by probabilistic reasoning under the constraints imposed by the domain knowledge, e.g., a person currently inside a room cannot enter again without first leaving it, and vice versa. The novelty of our work mainly lies in the development of a systematic way to incorporate the correlation and constraint among a sequence of people observations, and the optimality of recognition is achieved by maximizing a joint posterior probability of the observations. Experimental results of real and synthetic data are presented to show the efficacy of the proposed approach.

© 2004 Elsevier B.V. All rights reserved.

Keywords: People monitoring; Domain knowledge; Recognition; Probabilistic reasoning; Joint maximum likelihood

1. Introduction

As automated video surveillance becomes increasingly important [9,10,18], we are endeavoring to develop a video-based system for closed-room people monitoring. In this paper, we present a promising system of our recent effort. The system

consists of two modules: a video analysis module to detect and track people entering or leaving the only entrance/exit of a closed room (e.g., a research lab, class room or meeting room), and a people recognition module to correspond each observed person with a person previously entering the room or to identify him/her as a new person unseen before. The results obtained from the system allow the derivation of such useful information as to how long a person has stayed in the room, how many people are in the room during a

*Corresponding author. Tel.: +65-6790-5872; fax: +65-6793-3318.

E-mail address: eyptan@ntu.edu.sg (Y.-P. Tan).

certain period, and who they are. Potential applications of the system include, for example, identifying who could possibly be infected by a newly identified severe acute respiratory syndrome (SARS) victim [3], and characterizing the regular or abnormal human activities in a monitored work place.

Besides our proposed approach, there exist other solutions to the problems of interest. For example, biometrics have been successfully used for person recognition and very satisfactory results have been achieved. Representative work in biometric-based recognition ranges from fingerprint and iris identification to face and gait recognition [1,4,11,20]. However, many of these methods involve intrusive data collection, i.e., require human proactive action and collaboration in the course of identification or authentication, and thus work mainly in well-controlled environments. Although gait recognition partly addresses this limitation by exploring human motion characteristics, gait information, by itself, has limited discriminating power and only works for people whose motion patterns have been well characterized. In another approach, Nakajima et al. [17] attempt to recognize people by color-based and shape-based low-level features with the use of multi-class support vector machine classifiers (SVMs), where the people under monitoring are restricted to their lab members, for whom a database for recognition can be constructed beforehand.

Regardless of the selection of feature representation, most of the existing approaches accomplish the recognition task based on some maximum likelihood classification rule [7], which is generally simple, effective and fast. Vega et al. [25] model people by their gait-based feature relational distribution and recognize a person by choosing the one with the smallest distribution distance from a test data set. McKenna et al. [14] attempt to recognize people after separation while tracking a group of people. They compare the color models of people after interaction with the models built before the interaction and correspond them to maximize a joint likelihood. It should be noted that these existing methods mainly perform recognition based on observations or features

obtained at a single time instance or duration. The temporal correlation and constraint among observation sequence, however, are seldom utilized although they often exist in some specific contexts; for example, a person currently inside a room cannot be entering the room without first leaving it.

In comparison, we develop a systematic approach to exploit the inherent correlation and constraint among multiple people observations for improving the recognition accuracy. Specifically, an HMM-like, probabilistic inference framework is proposed to incorporate domain-specific knowledge (i.e., a closed-room with a single entrance/exit) for enhanced people recognition. To be unintrusive, we choose color histogram as the feature for people recognition for it can be easily and reliably extracted using a low-cost camera from a distance. Rather than using only a single observation, our method is capable of performing optimal recognition based on multiple observations acquired at different time instances to maximize a joint posterior probability. As a result, it can effectively enhance the limited discriminating power of low-level features, such as color histogram. Experimental results demonstrate that the proposed approach can obtain promising recognition rates, as compared with the existing maximum likelihood approaches.

2. Video analysis

2.1. People detection & tracking

In this module, we extend the color-based people detection and tracking system developed in our previous work [13,24]. To work in a more perceptually uniform color space, we choose the HSV (hue, saturation, and value) color space and represent the color of each pixel with a vector of three color attributes (H, S, V). To have a reference background for change detection, we use a background initialization scheme to construct the background model automatically from a scene that may contain moving objects or light flickers. Specifically, a pixel will not be deemed as a background pixel unless its color/intensity value

remains stable for a sufficiently long period. The color attributes of each pixel during this period are then modeled with a multivariate Gaussian distribution as the pixel's background model.

To accommodate scene change over time, we also update the background model adaptively based on the newly observed background pixels. The updating is carried out as follows:

$$\mu_{x,t} = (1 - \lambda_{x,t})\mu_{x,t-1} + \lambda_{x,t}O_x, \quad (1a)$$

$$\sigma_{x,t}^2 = (1 - \lambda_{x,t})\sigma_{x,t-1}^2 + \lambda_{x,t}(O_x - \mu_{x,t})^2, \quad (1b)$$

where $x = \{H, S, V\}$ and $\mathbf{O} = [O_H, O_S, O_V]^T$ denote the *HSV* values of each newly observed background pixel at time t , and $\lambda_{x,t}$ is an updating parameter proportional to standard deviation $\sigma_{x,t-1}$. In our implementation, we set $\lambda_{x,t} = 0.5 \times \sigma_{x,t-1}$, for the reason that the faster or larger a pixel changes, the faster its background model should be updated.

Since the *HSV* chrominance information becomes less reliable when the scene is too dark, a value threshold Thr_V and a saturation threshold Thr_S are used to decide whether the observed hue information O_H could be used for segmenting the foreground regions. (In our case, the foreground regions of interest are regions containing moving people.) Specifically, the following rules are applied to classify each pixel as either a background or a foreground pixel:

- If the hue information is reliable, the pixel is marked as a foreground pixel if $|\mu_H - O_H| > 3 \times \sigma_H$; otherwise, it is marked as a background pixel.
- If the hue information is not reliable, the pixel is marked as a changed pixel if $|\mu_V - O_V| > 3 \times \sigma_V$; otherwise, it is marked as a background pixel.

Using the binary segmentation results, the module then localizes a moving person within a minimum bounding rectangle (MBR) and represents the person with a set of identity attributes, including a label, the size, and the centroid of the MBR.

For people tracking, we extend the continuously adaptive mean shift (CAMSHIFT) algorithm [2] to work with the density function of the binary segmentation results. The extended CAMSHIFT algorithm searches within the current frame for the

nearest foreground region of each moving person tracked in the previous frame. This is because each moving person in the current frame cannot move far away from his/her location in the previous frame in a video sequence of high frame rate—an assumption that holds for many general applications.

During tracking, color histogram, a computationally easy color feature that is known to be rather invariant to changes of shapes and sizes [23], is acquired to represent the color appearance of each person. In our implementation, we quantize the coordinates of *HSV* color space uniformly into 12(hue), 4(saturation) and 4(value) bins, respectively, resulting in a total of 192 quantized colors. To account for varying person size, the color histogram $H_p(k)$ for each tracked person p is normalized by the total number of foreground pixels of the person. Furthermore, to accommodate changes due to different lighting condition or camera view angle, the color histogram is updated as follows:

$$H_p^t(k) = (1 - l) \cdot H_p^{t-1}(k) + l \cdot H_p^{\text{new}}(k), \quad 1 \leq k \leq 192, \quad (2)$$

where l is a learning parameter, and $H_p^{\text{new}}(k)$ is the color histogram obtained from the new frame acquired at time t . Note that as the color histogram of a tracked person could be severely affected by other pixels, such as the shadow pixels or noise clutters due to occlusion, we also make use of a shadow-removal scheme in the proposed system to enhance the quality of each extracted color histogram [24].

The information acquired during the tracking of a person, including the size variation and motion trajectory of a tracked person, is then used to judge whether the person is entering or leaving a room as seen by the camera (Figs. 1 and 2).

2.2. Similarity measure

Color histogram intersection has been widely used as a similarity measure for image recognition and retrieval [6,8,15]; however, direct application of this measure to gauge the likeness of people in our work poses some potential problems. For

Fig. 1. System overview with two sample frames illustrating a person entering and leaving a closed room with a single entrance/exit.

Fig. 2. Sample images of the people detection and tracking results: (a) a person is leaving the lab, and (b) another person is entering the lab.

example, color histogram intersection of two different people is generally larger than zero for some of their appearance features, such as hair and skin, could share similar colors. On the other hand, the same person observed at different time instances may not have identical color histograms due to the differences in lighting conditions, camera view angles, and segmentation results. A function is therefore required to map the value of color histogram intersection of two observed people c_i and c_j , defined as $I(H_i, H_j) = \sum_{k=1}^{192} \min(H_i(k), H_j(k))$, onto a similarity probability, $P(c_i \sim c_j)$, indicating whether c_i and c_j correspond to the same person.

Conceivably, the function $P(I)$ needs to possess the following properties: (1) it should be non-decreasing; (2) $P(I)$ should be approaching 0 and 1 as I takes values near the ends of the interval $[0, 1]$, respectively; (3) the transition of $P(I)$ from a low value to a high value should take place at where the value of I becomes evident to support that the two observed people are identical. After some subjective studies and comparisons, we select the *sigmoid* function [12] to perform the required mapping, which is shown in Fig. 3 and can be expressed as

$$P[c_i \sim c_j] = \frac{1}{1 + \exp[-\alpha(I(H_i, H_j) - \beta)]}. \quad (3)$$

There are two parameters determining the shape of a sigmoid function curve: α controlling the steepness of the transition and β defining the

Fig. 3. The *Sigmoid* function to map color histogram intersection onto the similarity probability of two observed people.

center of the transition point. To assign proper values to the two parameters α and β , we have examined a large number of people segmentation results with different similarity degrees; some of them are indicated in Fig. 3 using the crosses, whose x-coordinates are color histogram intersection values and y-coordinates are the similarity degrees judged by human observers. We have found that setting the parameter pair (α, β) to (16, 0.55) can fit the curve well.

It should be noted that other features, such as face or gait features, for which a similarity measure is defined can be used in our proposed people monitoring system. To make the system less intrusive, we have only made use of color histogram in the work reported in this paper.

3. Probabilistic reasoning framework

3.1. Problem formulation

As hidden Markov models (HMMs) could be used to integrate prior knowledge and incomplete observation data to perform probabilistic reasoning [16,22,26], our first attempt is to establish a suitable HMM for the recognition task in our problem. In general, an HMM can be fully characterized by a set of parameters $\lambda = \{A, B, \pi\}$, where A denotes the state transition probabilities, B , the observation symbol probabilities, and C , initial state probabilities. These parameters are usually pre-learned from a set of representative data based on a fixed number of states. This situation, however, is not applicable to our case, for the number of people as well as their activity patterns, i.e., the frequencies of entering and leaving, are generally different from place to place or time to time, and hard to estimate from prior data.

To construct a framework that is well suited for the problem of our concern, we make use of the lattice structure and parameter setting of HMMs and formulate the problem of people recognition as follows. Assume that the room being monitored has only a single entrance/exit (referred to as a closed room) and is empty when the system is first activated. When a person is entering the room, we

append a new state to the state set (database) to represent his/her identity; when a person leaves at time t , he/she will be recorded as a new observation O_t . Thus the states in the state set at time t correspond to the current people identities in the database (i.e., people that possibly stay in the room at time t), and are denoted as $S(t) = \{S_1 \cdots S_{N_t}\}$. When an observation sequence $O = \{O_1 \cdots O_T\}$ has been obtained over a period of time, the identities of the people leaving the room can be recognized from the state sets, $S(1), S(2), \dots$, and $S(T)$, recorded over the same period. Characterizing this reasoning framework with a time-variant parameter set $\lambda(t) = \{A(t), B(t), \pi(t)\}$, we can use the *Viterbi* algorithm [19] to find an optimal state sequence $Q = \{q_1 \cdots q_T\}$ associated with the observation sequence to maximize a joint posterior probability $P(Q, O|\lambda(t))$, so that each leaving person can correspond to one of those who are judged still inside the closed room.

A recognition example of the proposed framework is illustrated in Fig. 4, where the optimal state sequence obtained is highlighted in bold (red). From this path, we can identify O_1 as S_1, O_2 as S_3, O_3 as S_1 , etc. We can see that the framework has a lattice structure similar to HMMs; however, there are several key differences that distinguish

our framework from conventional HMMs. First, and most importantly, the parameter set $\lambda(t)$ is time variant and needs to be derived at each observation time instance based on all the previous possible states and current observations rather than from some training data. Second, the number of states in our model is not fixed but can increase over time before a decision is made. Third, the states can be indefinite because more than one states could be associated with the same person, e.g., states S_1 and S_4 in the example shown in Fig. 4.

It should be noted that in our framework a state could represent the feature model and the identity of a person simultaneously. For clarity, we shall use s_i to denote a person's identity and S_i his/her feature model.

3.2. Framework construction

It can be seen from our problem formulation that the main task in constructing the proposed framework lies in the estimation of the time-variant parameter set $\lambda(t)$. Once $\lambda(t)$ is known, we can find the optimal path indicating the recognized people by a “turn-the-crank” procedure given by the *Viterbi* algorithm [5]. Our solutions are given as follows.

- **The initial state distribution, π :**

According to the definition of the state set $S(t) = \{S_1 \cdots S_{N_t}\}$, there are N_1 states (people) in the database when the first observation of exit is made at time $t = 1$. Without any other prior knowledge, we intuitively assume that everyone inside the room has an equal probability to leave the room. Hence,

$$\pi_i = \frac{1}{N_1}, \quad 1 \leq i \leq N_1. \quad (4)$$

- **The output probability of state i at time t , $b_i(O_t)$:**

In analogy with the definition in HMMs, let $b_i(O_t)$ be the probability of observation O_t generated by state i . We regard this probability as how likely an observed O_t is due to person s_i , and simply approximate it by Eq. (3) using their

Fig. 4. A recognition example of the proposed probabilistic reasoning framework.

feature similarity as

$$b_i(O_t) = P[O_t \sim S_i], \quad 1 \leq i \leq N_t. \quad (5)$$

• **The state transition probability, $a_{ij}(t)$:**

Before proceeding to the next step, we define a set of probabilities between observation times t and $t+1$ as shown in Fig. 5. For concision, we shall refer to “observation time” as “time” hereafter without causing confusion. Let $P[s_{i,t^+/-} = 1]$ be the probability of person s_i staying in the room at time instance $t^+/-$ (straightforwardly $P[s_{i,t^+/-} = 0] = 1 - P[s_{i,t^+/-} = 1]$), where t^+ and t^- denote the time instances right after and before the observation O_t being made. Let M be a likelihood matrix measuring the similarities between people entering between t and $t+1$ (i.e., $S_{N_t+1} \cdots S_{N_{t+1}}$) and the people staying in the room at time t (i.e., $S_1 \cdots S_{N_t}$), defined as

$$M = \{m_{uv} | m_{uv} = P[S_{N_t+v} \sim S_u]\}, \quad 1 \leq u \leq N_t, \quad 1 \leq v \leq N_t^{t+1}, \quad (6)$$

where N_t is the number of states (people who possibly stay in the room) at time t and $N_t^{t+1} = N_{t+1} - N_t$ is the number of people entering between times t and $t+1$. With these auxiliary probabilities, we can now derive the transition probability using the following three steps:

(i) *Transition probability, $a_{ij}(t)$:*

Fig. 5. The likelihood matrix M between times t and $t+1$ and the probability of the existence of a person associated with a certain state at time instances t^+ and $t+1^-$.

The transition probability $a_{ij}(t)$ measures the odds of person s_j leaving the room at time $t+1$, given that person s_i left at time t , i.e., $q_t = s_i$ and $q_{t+1} = s_j$. Without other prior knowledge, it is reasonable to assume that the probability for s_j to leave at time $t+1$ is proportional to that of his/her existence in the room at time instance $t+1^-$. Conceivably, a person cannot leave a room if he/she is not in the room at all. Thus, the transition probability can be computed as

$$a_{ij}(t) = \frac{P[s_{j,t+1^-} = 1 | q_t = s_i]}{\sum_j P[s_{j,t+1^-} = 1 | q_t = s_i]}, \quad 1 \leq i \leq N_t, \quad 1 \leq j \leq N_{t+1}, \quad (7)$$

where the denominator is a normalization factor to make $\sum_j a_{ij}(t) = 1$ for each i . The term in Eq. (7) can be further expanded for each particular state s_i by using Bayesian rules as

$$P[s_{j,t+1^-} = 1 | q_t = s_i] = \sum_{\text{all } cond} P[s_{j,t+1^-} = 1 | cond] \cdot P[cond | q_t = s_i], \quad (8)$$

where $cond = \{s_{1,t^+} = \theta_1 \cdots s_{N_t,t^+} = \theta_{N_t}\}$ is one of the possible realizations of $\{s_{1,t^+} \cdots s_{N_t,t^+}\}$ over all θ , and $\theta_i = 1$ or 0 represents the *status* of person s_i being in or out of the room. By assuming that the status of each person is statistically independent of the others, the second term on the right hand side of Eq. (8) can be expressed as

$$P[cond | q_t = s_i] = \prod_{i=1}^{N_t} P[s_{i,t^+} = \theta_i | q_t = s_i]. \quad (9)$$

From Eqs. (7)–(9), we can see that the transition probability $a_{ij}(t)$ depends on two probabilities: $P[s_{j,t+1^-} = 1 | cond]$ and $P[s_{i,t^+} = \theta_i | q_t = s_i]$; the former is the conditional odds of person s_j staying in the room at time instance $t+1^-$ given the status of all the people observed up to time instance t^+ , and the latter is the probability of person s_i being status θ_i at time instance t^+ given that person s_i leaves the room at time t . The calculation of these two probabilities is discussed below.

(ii) *Conditional probability, $P[s_{j,t+1^-} = 1 | cond]$:*

We compute this conditional probability with the aid of the likelihood matrix M defined in

Eq. (6). The specific calculation is given by

$$P[s_{j,t+1} = 1 | \text{cond}] = \begin{cases} \theta_j + \sum_v \tilde{m}_{jv}, & 1 \leq j \leq N_t, \\ 1 - \sum_u \tilde{m}_{u(j-N_t)}, & N_t \leq j \leq N_{t+1}, \end{cases} \quad (10)$$

where $\tilde{m}_{uv}$ is the entry of a modified matrix $\tilde{M}$ which is derived from M to indicate how likely person s_{N_t+v} is in fact person s_u returning under a certain *cond*. We introduce the matrix $\tilde{M}$ in order to incorporate the domain knowledge in the context of the *cond* and improve our estimation. In particular, the value $\sum_v \tilde{m}_{jv}$ can be considered as the gain of the existence odds for person s_j due to newly entered people between times t and $t+1$. This gain, along with the existence odds of person s_j at time instance t^+ , which is the binary value of θ_j , constitutes his/her existence odds at time instance $t+1^-$. For those newly added states s_j where $N_t \leq j \leq N_{t+1}$, the value $1 - \sum_u \tilde{m}_{u(j-N_t)}$ can therefore be considered as the odds for person s_j to exist as a new person. Note that in this framework we attempt to estimate these odds in an approximate manner, which is intuitive and has been found to be effective for improving the recognition rates.

Specifically, $\tilde{m}_{uv}$ is derived from m_{uv} as

$$\tilde{m}_{uv} = \eta^{2D}[\varepsilon(\theta_u)\rho(S_u)m_{uv}], \quad (11)$$

where η^{2D} is a 2-D normalization operation to be discussed later and

$$\varepsilon(\theta_u) = \begin{cases} 0 & \text{if } \theta_u = 1 \text{ in } \text{cond}, \\ 1 & \text{otherwise}, \end{cases} \quad (12)$$

$$\rho(S_u) = \begin{cases} 0 & \text{if } S_u \notin PT(S_k, t), \\ 1 & \text{otherwise}. \end{cases} \quad (13)$$

$\varepsilon(\theta_u)$ and $\rho(S_u)$ are two indication factors incorporating the domain knowledge of a particular *cond*. Specifically, $\varepsilon(\theta_u)$ shows that it is impossible for one to enter a closed room if he/she is currently inside, while $\rho(S_u)$ accounts for the fact that a person staying in the room could not enter the room again if he/she has not been observed leaving the room. $PT(S_i, t)$ is the most likely path ending in state S_i at time t , i.e., the most probable state sequence that is obtained by the Viterbi algorithm

and includes all the people that have been observed leaving the room up to time t .

To illustrate it, we provide in the following a numerical example, in which five people enter a room with three of them leaving, forming a process of three observations as shown in Fig. 6.

For the time pair $\{t, t+1\}$ shown in Fig. 6, suppose that the likelihood matrix obtained by Eq. (6) is equal to

$$M = \begin{matrix} & \begin{matrix} s_4 & s_5 \end{matrix} \\ \begin{matrix} s_1 \\ s_2 \\ s_3 \end{matrix} & \begin{bmatrix} 0.2 & 0.8 \\ 0.5 & 0.3 \\ 0.4 & 0.2 \end{bmatrix} \end{matrix}.$$

Consider one of the possible realizations of people's status at time instance t^+ : *cond1* = $\{s_{1,t^+} = 0, s_{2,t^+} = 0, s_{3,t^+} = 1\}$; that is, $\{\theta_1 = 0, \theta_2 = 0, \theta_3 = 1\}$ or all people are outside the room at time instance t^+ except for person s_3 . The conditional probabilities that need to be computed are $P[s_{j,t+1} = 1 | \text{cond1}]$, $1 \leq j \leq 5$.

First, we incorporate the knowledge of the people status into the likelihood matrix M by using the two indication factors. From Eq. (12), $\varepsilon(\theta_3) = 0$ and the third row of M is set to zero. This is because if s_3 is known to be inside the room, then neither s_4 nor s_5 can be s_3 regardless of their similarities. The second indication factor relies on the most likely path ending in the state for which the transition probability needs to be calculated. Assume that S_1 at time t is the state and the most likely path $PT(S_1, t)$ is as shown in Fig. 6. It can be seen that S_3 is not on this optimal

Fig. 6. An example illustrating the estimation of conditional probability.

path, which means person s_3 has not left the room since he/she entered. So none of s_4 and s_5 could be s_3 and $\rho(S_3) = 0$ according to Eq. (13). With this domain knowledge, the original likelihood matrix M can be modified as

$$M' = \begin{array}{cc} & \begin{array}{cc} s_4 & s_5 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{bmatrix} 0.2 & 0.8 \\ 0.5 & 0.3 \\ 0 & 0 \end{bmatrix} \end{array}.$$

In this example, the two indication factors affect the same row of the likelihood matrix M . In general, the adjustments could be made on different rows depending on the *cond* as well as the state chosen for evaluating the transition probability.

Next, the adjusted likelihood matrix M' is normalized so that the summations of the likelihoods corresponding to all possible circumstances are equal to one. To do this, we introduce a normalization operation, referred to as the correlated normalization operation or η operation, which works as follows. Consider a person $\mathcal{T}$ and a group of candidates $\mathcal{C}$ consisting of N people, and define the similarity measure $w_n = P[\mathcal{C}_n \sim \mathcal{T}]$. Under the constraint that at most one person (or none) of $\mathcal{C}$ could be $\mathcal{T}$, we can obtain the following new similarity measures

$$P[\mathcal{C}_n \sim \mathcal{T} | \text{constr}] \propto w_n \cdot \prod_{j \neq n} (1 - w_j), \quad 1 \leq n \leq N, \quad (14)$$

$$P[\mathcal{T} \notin \mathcal{C} | \text{constr}] \propto \prod_{i=1}^N (1 - w_i). \quad (15)$$

To make the probabilities corresponding to all possible circumstances sum up to one, the η operation is defined to normalize the above new similarity measures as

$$\eta[w_n] = \frac{w_n \cdot \prod_{j \neq n} (1 - w_j)}{\sum_{i=1}^N w_i \prod_{j \neq i} (1 - w_j) + \prod_{i=1}^N (1 - w_i)}, \quad 1 \leq n \leq N. \quad (16)$$

This normalization operation aims to correlate likelihoods that are measured independently by imposing the constraint mentioned before

Eq. (14). It should be noted that this constraint can be applied to the adjusted likelihood matrix M' both row-wise and column-wise. For instance, at most one of s_4 and s_5 (or none of them) could be s_1 , while s_4 could be at most one of s_1 , s_2 , and s_3 (or none of them, i.e., s_4 is a new person) in the example shown in Fig. 6. For simplicity, we apply a 2-D η operation (η^{2D}), a row-wise normalization followed by a column-wise normalization, to normalize the likelihood matrix M' and obtain

$$\tilde{M} = \begin{array}{cc} & \begin{array}{cc} s_4 & s_5 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} & \begin{bmatrix} 0.05 & 0.72 \\ 0.40 & 0.05 \\ 0 & 0 \end{bmatrix} \end{array}.$$

From this normalized likelihood matrix $\tilde{M}$, we can obtain useful information such as the existence odds of s_1 is increased by $0.05 + 0.72 = 0.77$, the odds that s_4 is a new person is $1 - 0.05 - 0.40 - 0 = 0.55$, etc. Consequently, the existence odds at time instance $t + 1^-$ under *cond1* can be obtained using Eq. (10) as

$$\begin{aligned} P[s_{1,t+1^-} = 1 | \text{cond1}] &= 0 + 0.05 + 0.72 = 0.77, \\ P[s_{2,t+1^-} = 1 | \text{cond1}] &= 0 + 0.40 + 0.05 = 0.45, \\ P[s_{3,t+1^-} = 1 | \text{cond1}] &= 1 + 0 + 0 = 1, \\ P[s_{4,t+1^-} = 1 | \text{cond1}] &= 1 - 0.05 - 0.40 - 0 = 0.55, \\ P[s_{5,t+1^-} = 1 | \text{cond1}] &= 1 - 0.72 - 0.05 - 0 = 0.23. \end{aligned}$$

Applying the same procedure to all possible *cond*'s, we can obtain all the conditional probabilities $P[s_{i,t+1^-} = 1 | \text{cond}]$ required in Eq. (8), one of the two prerequisites for computing the transition probability as discussed in step (i).

(iii) *Probability*: $P[s_{i,t^+} = 1 | q_t = s_{\hat{j}}]$

The remaining unknown for calculating the transition probability $a_{ij}(t)$ is $P[s_{i,t^+} = 1 | q_t = s_{\hat{j}}]$, which is the existence odds of person s_i at time instance t^+ given that the person leaving at time t is $s_{\hat{j}}$. To illustrate how to compute this unknown, we shall extend the example of Fig. 6 to time $t + 1$ and show instead how to compute the unknown $P[s_{j,t+1^+} = 1 | q_{t+1} = s_{\hat{j}}]$ by making use of the probabilities obtained so far and assuming the full knowledge of $P[s_{i,t^+} = 1 | q_t = s_{\hat{j}}]$, where $S_{\hat{j}}$ is the previous state of S_j in the most likely path $PT(S_j, t + 1)$. By the same procedure,

$P[s_{i,t^+} = 1|q_t = s_i]$ can be similarly estimated based on $P[s_{h,t-1^+} = 1|q_{t-1} = s_{\hat{i}}]$. Note that, h, i , and j are indices of the states at times $t-1$, t , and $t+1$, respectively.

As shown in Fig. 7, assume that person s_1 leaves the room both at time t and at time $t+1$, i.e., $(\hat{j}) = 1$ and $\hat{j} = 1$ (denoted by the shaded circles). We manifest the values of $P[s_{i,t^+} = 1|q_t = s_1]$, which are assumed to be known, on the right hand side of the observation made at time t , and the computed values of $P[s_{j,t+1^-} = 1|q_t = s_1]$ on the left hand side of the observation made at time $t+1$. The estimation process of the probabilities on the right hand side of the observation O_{t+1} and the final results are shown in Table 1.

We first examine the gain of existence odds from time instance t^+ to $t+1^-$ for each person: $\gamma(j) = P[s_{j,t+1^-} = 1|q_t = s_1] - P[s_{j,t^+} = 1|q_t = s_1]$. Clearly, $\sum_j \gamma(j) = 2$ because the gain in existence odds is due to the two newly entered people. However, if we know for sure that s_1 leaves the room at time $t+1$, then he/she must stay in the

room at time instance $t+1^-$, and thus, $\gamma(1)$ should be equal to 1 rather than 0.77. In other words, the gain for each person needs to be re-adjusted ($\tilde{\gamma}$ in the table) to incorporate the knowledge of this new assumption to ensure that $\tilde{\gamma}(1) = 1$. That $\tilde{\gamma}(1)$ is equal to one can be also explained as follows: since s_1 leaves at time t and time $t+1$ successively, one of the entering persons s_4 and s_5 must be s_1 . As there is no reason to favor any person other than s_1 , our approach is to increase s_1 's existence odds from 0.77 to 1 and decrease the others' proportionally. This is sensible as once we know that the person who leaves the room at time $t+1$ is s_1 , then s_4 and s_5 should be more likely to be s_1 than what we originally estimate, and as a result, less likely to be the other people. At time instance $t+1^+$, the existence odds of s_1 becomes zero due to his exit, and the others' can be obtained as the summation of their existence odds at time instance t^+ ($P[s_{j,t^+} = 1|q_t = s_1]$) and the re-adjusted gain ($\tilde{\gamma}(j)$).

The above analysis and derivation can be generalized into the formulas below:

$$P[s_{j,t+1^+} = 1|q_{t+1} = S_j] = \begin{cases} 0 & \text{if } j = \hat{j}, \\ P[s_{j,t^+} = 1|q_t = S_{\hat{j}}] + \tilde{\gamma}(j) & \text{otherwise,} \end{cases} \quad (17)$$

$$\tilde{\gamma}(j) = \begin{cases} 1 & \text{if } j = \hat{j}, \\ \gamma(j) \left(1 - \frac{1 - P[s_{\hat{j},t+1^-} = 1|q_t = S_{\hat{j}}]}{\sum_{j \neq \hat{j}} \gamma(j)} \right) & \text{otherwise,} \end{cases} \quad (18)$$

$$\gamma(j) = P[s_{j,t+1^-} = 1|q_t = S_{\hat{j}}] - P[s_{j,t^+} = 1|q_t = S_{\hat{j}}]. \quad (19)$$

Fig. 7. An example to illustrate the estimation of existence odds.

Table 1
The estimation of the existence odds

	$P[s_{j,t^+} = 1 q_t = s_1]$	$P[s_{j,t+1^-} = 1 q_t = s_1]$	γ	$\tilde{\gamma}$	$P[s_{j,t+1^+} = 1 q_{t+1} = s_j]$
s_1	0	0.77	0.77	1	0
s_2	0.30	0.50	0.20	0.16	0.46
s_3	0.70	0.75	0.05	0.04	0.74
s_4	N/A	0.75	0.75	0.61	0.61
s_5	N/A	0.23	0.23	0.19	0.19

It can be seen from these formulas that the estimation of existence odds is recursive; that is, one's existence odds at time instance $t+1^+$ depends on that at t^+ . To initialize the estimation, we set $P[s_{i,1^+} = 1 | q_1 = s_i] = 1$ for all $i \neq \hat{i}$ and $P[s_{i,1^+} = 1 | q_1 = s_i] = 0$ if s_i is the first person leaving the room (i.e. $q_1 = s_i$), since all the people who have entered the room, except for s_i , should stay inside at time instance 1^+ .

It should also be noted that although the derivation may appear somewhat complex, it leads to an important underlying property of the proposed framework: the summation of the existence odds of all states at any time instance $t^{+/-}$, i.e., $\sum_i P[s_{i,t^{+/-}} = 1 | q_t = s_j]$ for any t and j , is always equal to the number of people who stay in the room at that time.

Batch recognition of observations (people who have left the room) can be performed at any time when necessary by retrieving the state sequence with the maximum score of joint posterior probability as the *optimal state sequence*.

To use the proposed framework for recognizing people re-entering the room is equivalent to finding and merging those states associated with a same person in the *optimal state sequence*. This can be accomplished by the following local maximum likelihood scheme. Let $q_t = S_i$ and search backward in the *optimal state sequence* for $q_{t'} = S_i$, where $t' \in \{1, \dots, t-1\}$ and $t-t'$ is minimized. If such $q_{t'}$ exists, the person s_j will be assumed as the person s_i re-entering, where $\hat{j}$ is obtained as

$$\hat{j} = \arg \max_{j \in S(t) \setminus S(t')} P[S_j \sim S_i], \quad (20)$$

and $S(t)$ is the state set at time t . Then, $S_{\hat{j}}$ should be merged into S_i and removed from all the state sets containing it thereafter. The backward search is performed from $t=2$ to the end of the *optimal state sequence* until all the states possibly representing a same person are disclosed and merged.

4. Case studies

In this section, we study two special cases in which certain extreme conditions are assumed so

that intuitive analysis can be performed. By comparing the results obtained by our model and that of the intuitive analysis, we demonstrate that the proposed model can produce reasonable results even under such extreme conditions.

4.1. Case 1: Successive exits

In this case, we assume that three people enter a room initially and then leave successively without any other new entries among them. The model established is shown in Fig. 8, where the likelihood matrices M are null as nobody enters between two observations.

The transition matrix is first computed at time $t=1$ by using Eqs. (7)–(19) derived in Section 3.2 and obtained as

$$A(t=1) = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}.$$

The diagonal entry $a_{ii}(t=1)$ is zero, which is in accordance with the intuitive analysis that if s_i leaves at time $t=1$ and nobody enters before the next observation time $t=2$, s_i could not leave at time $t=2$ again.

Moving on to the next observation time $t=2$, we further assume that the previous states of S_1 , S_2 and S_3 in their own most likely paths are S_2 , S_3 and S_1 at time $t=1$, respectively, which can be obtained by using the *Viterbi* algorithm based on $P[O_1 \sim S_i]$ and $A(t=1)$. Hence, the transition

Fig. 8. A model illustrating the case in which three people enter and leave a closed room successively, where the dash arrows indicate the transitions that are not allowed.

matrix at time $t = 2$ can be computed as

$$A(t = 2) = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

That means, from time $t = 2$ to time $t = 3$, each state has only one possible choice of transition. Because, for instance, if s_1 leaves at time $t = 2$ and its previous state s_2 has already left at time $t = 1$, the person who leaves the room at time $t = 3$ must be s_3 . More specifically, to select which state at time t depends on not only the states at the previous time $t - 1$ but also those prior to $t - 1$.

4.2. Case 2: Ideal similarity measure

In this case, we shall assume that the ideal similarity measure between two people can be a binary value of either one (identical) or zero (distinct). In Fig. 9, two people enter first and one of them leaves afterward. After that one person enters again followed by one exit. We suppose that S_3 looks the same as S_1 but has different appearance from S_2 , i.e. $P[S_3 \sim S_1] = 1$ and $P[S_3 \sim S_2] = 0$ (likelihood matrix $M = [1 \ 0]^T$). The transition matrix can be computed as

$$A(t = 1) = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix}.$$

We can interpret this process based on our general knowledge. If s_1 leaves at time $t = 1$, s_3 will be viewed as s_1 re-entering due to their same appearance. Therefore, at time instance $t = 2^-$, s_1 and s_2 must stay in the room while s_3 cannot be a new person. s_1 and s_2 then take the equal

Fig. 9. A model illustrating the case in which S_3 has perfect similarity with S_1 and S_2 .

probability to leave at time $t = 2$, which is reflected in the transition matrix as $\{a_{11} = a_{12} = 1/2, a_{13} = 0\}$. If s_2 leaves at time $t = 1$, it is interesting to see that s_3 has to be a new person although he/she has the same appearance with S_1 since s_1 is already known to be inside. (One possibility is that s_3 and s_1 are identical twins wearing the same clothes.) In this case, s_1 and s_3 are equally possible to leave at time $t = 2$ while s_2 could not leave again. In the transition matrix, it is shown as $\{a_{21} = a_{23} = 1/2, a_{22} = 0\}$. Once again, the proposed model obtains the same result as that derived by the intuitive analysis based on our general knowledge.

5. Experimental results

We have tested the proposed people monitoring system with two videos acquired in a research laboratory using a camera to monitor the lab's only entrance. During 1-h monitoring period, video-I recorded eleven different people who were unaware of the experiment, of which four entered and left twice, another four entered and left only once, and three entered without leaving. Video-II simulated the process of people entering and leaving with the help of eight different students, among whom seven entered and left the lab for three times and another one twice. To further increase the test samples, a synthetic process generator is also designed to randomly re-arrange the entries and exits from the two videos based on the rule that one cannot enter unless he/she is outside the lab, and vice versa. This generator allows us to simulate a wide variety of combinations of entries and exits over time from the same group of people.

For comparison, we also implement a recognition approach based on maximum likelihood (ML) classification [14,21] as follows. When a person is detected entering a closed room, he/she is compared with people who are in the system's database and currently labeled as 'out'. If the maximum likelihood with respect to an 'out' person is larger than a threshold T_m (set to 0.7 in our implementation), they are considered the same person. Then, the observed person is labeled as 'in' and his/her corresponding color histogram

in the database is updated with the latest one. Otherwise, the person is assumed to be new and then labeled as ‘in’ with a new identity label. When there are multiple exits without entries among them, the leaving people are recognized from the people with label ‘in’ by maximizing a joint likelihood.

To ensure that the system can perform as desired, we assume that each person has consistent color appearance (both frontal and back views) and it remains the same during the period of observation. This is because our proposed system uses only the color appearance of each person to perform recognition and the camera looks only at a fixed direction. This limitation can likely be circumvented by incorporating other possible features, such as face and thermal infrared imagery, or using more than one camera looking at different directions. We consider this as a potential future research direction.

Fig. 10 shows the sample images of the eight people ($P1$ – $P8$) recorded in video-II after foreground segmentation. Table 2 shows an example sequence of entries and exits obtained by the synthetic process generator based on these eight people, where ‘I’ and ‘O’ represents in (enter) and out (exit), respectively. A total of 44 entries and exits are observed, among them 22 are entering while the other 22 are exiting.

The recognition results of the eight people (at observation time) are given in Table 3. The results obtained by our proposed approach and the ML approach are compared with the ground truth. In this particular process, our approach achieves 100 percent recognition rates, outperforming the ML approach. In the ML approach, the system wrongly recognizes $P8$ as $P1$ at time 7 and reversely at time 19. Furthermore, $P7$ is incorrectly identified as $P3$ at time 18 and on the contrary at time 20. The errors of recognizing different people as the same one are mainly due to that the similarities between the color histograms of these different people exceed the preset threshold T_m . On the contrary, the reason for the errors at time 10, 13 and 15, where $P6$ and $P7$ are identified as new people when they are in fact just re-entering, is that the similarities of their color histograms observed at different times are lower than T_m . In other words, the results of the ML approach are rather sensitive to the threshold T_m , inappropriate selection of which often causes such false recognitions as mentioned above. In contrast, our proposed approach benefits from using a threshold-free scheme; therefore, it is more robust to less-than-perfect segmentation results as well as changes in lighting conditions or view angles.

At each observation time, there is a most likely path ending in each individual state with the

Fig. 10. The segmentation results of eight different people recorded in video-II, with their entering (frontal) images in the first row and leaving (backward) ones in the second row.

Table 2

An example sequence of entries (I) and exits (O) produced by the synthetic process generator using the eight people observed in video-II

Person	P6	P2	P6	P6	P1	P1	P3	P2	P1	P8	P2	P1	P7	P5	P3
In/out	I	I	O	I	I	O	I	O	I	I	I	O	I	I	O
Person	P5	P1	P8	P7	P3	P5	P2	P6	P4	P6	P8	P5	P8	P7	P2
Cont'd	O	I	O	O	I	I	O	O	I	I	I	O	O	I	I
Person	P7	P2	P6	P5	P7	P4	P4	P4	P7	P1	P3	P8	P8	P5	
Cont'd	O	O	O	I	I	O	I	O	O	O	O	I	O	O	

Table 3

Comparison of the recognition results obtained by the maximum likelihood (ML) approach and our proposed approach

Obs. time	1	2	3	4	5	6	7	8	9	10	11
Ground truth	S_1	S_4	S_2	S_4	S_5	S_{10}	S_7	S_9	S_2	S_1	S_{10}
Proposed approach	S_1	S_4	S_2	S_4	S_5	S_{10}	S_7	S_9	S_2	S_1	S_{10}
ML approach	S_1	S_4	S_2	S_4	S_5	S_{10}	<u>S_4</u>	S_9	S_2	<u>S_3</u>	S_{10}
cont'd	12	13	14	15	16	17	18	19	20	21	22
Ground truth	S_7	S_9	S_2	S_1	S_{14}	S_{14}	S_9	S_4	S_5	S_7	S_{10}
Proposed approach	S_7	S_9	S_2	S_1	S_{14}	S_{14}	S_9	S_4	S_5	S_7	S_{10}
ML approach	S_7	<u>S_{17}</u>	S_2	<u>S_3</u>	S_{14}	S_{14}	<u>S_{12}</u>	<u>S_{16}</u>	<u>S_9</u>	S_7	S_{10}

probability of $\delta_t(i)$, which can be iteratively computed as $\delta_{t+1}(j) = [\max_j \delta_t(i) a_{ij}] b_j(O_{t+1})$ using the *Viterbi* algorithm. To make a recognition decision at time t is to choose from all the most likely paths the one with the largest value of $\delta_t(i)$. Hence, a confidence index can be defined in the range of $[0, 1]$ to evaluate the reliability of a decision made at each time

$$Conf(t) = \frac{\max(\delta_t(i))}{\sum_i \delta_t(i)}. \quad (21)$$

The variation of the confidence index over time for the above example is shown in Fig. 11.

At the early monitoring stage, the system has only few choices for making decision so that the confidence index is usually high. With the increase of the state number, i.e., more possible paths to choose from, the reliability of a decision may decrease. However, the advantage of our approach lies in that it is capable of maintaining the decision reliability at a later observation time by collectively considering all the available observations. In Fig. 11, when the confidence index is lower than 0.8, the ML approach is likely to make

a wrong decision at the corresponding time (as indicated by the (red) circles). Meanwhile, the proposed approach can still find the right choice since the probabilities (or scores) of the other paths are even lower. Note that as the recognition decision made at time t depends upon all the observations (not the decisions) obtained prior to time t in a sense of maximizing the joint posterior probability of multiple observations—a merit of the *Viterbi* algorithm—our proposed system can minimize the chance of error propagation. However, if there is a substantial change in the lighting condition or a person changes his/her color appearance (e.g., clothes) during the period of observation, the proposed framework may generate fallacious state set and degrade the recognition performance.

Overall, 20 synthesized processes of entries and exits are generated from each of the two test videos. We set T_m to 0.7 to obtain the best results of the ML approach. The ground truth is obtained manually. It is evident from Table 4 that our proposed approach can improve the recognition rates notably.

Fig. 11. The confidence index over decision time.

Table 4

Comparison of the recognition rates obtained by the maximum likelihood (ML) approach and our proposed approach

Data	ML approach	Proposed approach
Video-I (%)	83.3	99.6
Video-II (%)	75.0	97.5

It should also be noted that the computational complexity of our algorithm may increase rapidly when more states are generated. However, it is possible for one to make a definite decision when the confidence index is sufficiently high so that states associated with the same person can be merged to reduce the total number of the states, e.g., at time 16 in Fig. 11. We shall investigate this in our future work.

6. Conclusions

We have presented in this paper a novel probabilistic reasoning framework for people monitoring in a closed environment. The main contribution of our work lies in the development of a systematic way to exploit the temporal correlation and domain constraint of multiple people observations for improving the recognition

accuracy compared with that of conventional maximum-likelihood approaches. Rather than identifying each single observation from a database, the framework is devised to recognize people based on multiple observations acquired at different time instances. Experimental results demonstrate that the proposed approach outperforms the conventional maximum-likelihood approach when using the same feature and being tested with the same test videos.

It should also be noted that although we focus mainly on a system suitable for closed-room monitoring, we expect such correlation and constraint to be present in many application scenarios, e.g., an interconnected network of cameras monitoring multiple entrances/exits or adjacent areas of a large site, and our proposed approach can be readily extended to these applications.

References

- [1] B. Bhanu, Ju Han, Bayesian-based performance prediction for gait recognition, in: Proceedings of the Workshop on Motion and Video Computing, December 2002, pp. 145–150.
- [2] G.R. Bradski, Computer vision face tracking for use in a perceptual user interface, Intel Technol. J. Q2 (April 1998).
- [3] CDC SARS Investigative Team, A.T. Fleischauer, Outbreak of Severe Acute Respiratory Syndrome—Worldwide, 2003, MMWR 52 (11) (March 2003) 226–228.

- [4] R. Chellappa, C.L. Wilson, S. Sirohey, Human and machine recognition of faces: a survey, *Proc. IEEE* 83 (5) (May 1995) 705–741.
- [5] K. Church, Comments on computational learning model for metrical phonology, in: R. Levine (Ed.), *Formal Grammar: Theory and Implementation*, Vancouver Studies in Cognitive Science Series, vol. 2, UBC Press, 1992.
- [6] Y. Deng, B.S. Manjunath, C. Kenney, M.S. Moore, H. Shin, An efficient color representation for image retrieval, *IEEE Trans. Image Process.* 10 (1) (January 2001) 140–147.
- [7] R. Duda, P. Hart, D. Stork, *Pattern Classification*, second ed., Wiley, New York, 2001.
- [8] A.M. Ferman, A.M. Tekalp, R. Mehrotra, Robust color histogram descriptors for video segment retrieval and identification, *IEEE Trans. Image Process.* 11 (5) (May 2002) 497–508.
- [9] G.L. Foresti, Object recognition and tracking for remote video surveillance, *IEEE Trans. Circuits Systems Video Technol.* 9 (7) (October 1999) 1045–1062.
- [10] I. Haritaoglu, D. Harwood, L.S. Davis, W^4 : Real time surveillance of people and their activities, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (8) (August 2000) 809–830.
- [11] A. Kale, A. Rajagopalan, N. Cuntoor, V. Kruger, Gait-based recognition of humans using continuous HMMs, in: *Proceedings of the ICFGR*, 2002.
- [12] B.L. Kalman, S.C. Kwasny, Why tanh: choosing a sigmoidal function, in: *Proceedings of the International Joint Conference on Neural Networks*, vol. 4, June 1992, pp. 7–11.
- [13] W. Lu, Y.-P. Tan, A color histogram based people tracking system, in: *Proceedings of the IEEE International Symposium on Circuits and Systems*, vol. 2, 2001, pp. 137–140.
- [14] S. McKenna, et al., Tracking groups of people, *Comput. Vision Image Understanding* 80 (2000) 42–56.
- [15] F. Moreno, A. Tarrida, J. Andrade-Cetto, A. Sanfeliu, 3D real-time head tracking using color histograms and stereovision, in: *Proceedings of the 16th International Conference on Pattern Recognition*, Quebec, Canada, August 2002, pp. 368–371.
- [16] V. Nair, J.J. Clark, Automated visual surveillance using hidden Markov models, in: *Proceedings of the 15th Vision Interface Conference*, Calgary, May 2002, pp. 88–92.
- [17] C. Nakajima, M. Pontil, B. Heisele, T. Poggio, People recognition in image sequences by supervised learning, MIT Technical Report, June 2000.
- [18] N.M. Oliver, B. Rosario, A.P. Pentland, A bayesian computer vision system for modeling human interactions, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (8) (August 2000) 831–843.
- [19] L.R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, *Proc. IEEE* 77 (22) (1989) 257–285.
- [20] W. Shen, M. Surette, R. Khanna, Evaluation of automated biometrics-based identification and verification systems, *Proc. IEEE* 85 (1997).
- [21] J.M. Siskind, Q. Morris, A maximum-likelihood approach to visual event classification, in: *Proceedings of ECCV*, 1996.
- [22] T.E. Starner, A.P. Pentland, Visual recognition of American sign language using hidden Markov models, in: *Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition*, 1995, pp. 189–194.
- [23] M. Swain et al., Color indexing, *Int. J. Comput. Vision* (1991).
- [24] J. Tao, Y.-P. Tan, Color appearance-based approach to robust tracking and recognition of multiple people, in: *Proceedings of the Fourth International Conference on Information, Communications & Signal Processing and Fourth Pacific-Rim Conference on Multimedia*, December 2003.
- [25] I.R. Vega, S. Sarkar, Statistical motion model based on the change of feature relationships: human gait-based recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 25 (10) (October 2003).
- [26] J. Yang, Y. Xu, S.C. Chen, Human action learning via hidden Markov model, *IEEE Trans. Systems Man Cybern.* 27 (1) (January 1997) 34–44.