Available online at www.sciencedirect.com
ScienceDirect

Pattern Recognition 41 (2008) 1572 – 1580

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

A daily behavior enabled hidden Markov model for human behavior understanding

Pau-Choo Chung*, Chin-De Liu

Department of Electrical Engineering, Institute of Computer and Communication Engineering, National Cheng Kung University, Tainan 70101, Taiwan, ROC

Received 22 November 2006; received in revised form 5 October 2007; accepted 23 October 2007

Abstract

This paper presents a Hierarchical Context Hidden Markov Model (HC-HMM) for behavior understanding from video streams in a nursing center. The proposed HC-HMM infers elderly behaviors through three contexts which are spatial, activities, and temporal context. By considering the hierarchical architecture, HC-HMM builds three modules composing the three components, reasoning in the primary and the secondary relationship. The spatial contexts are defined from the spatial structure, so that it is placed as the primary inference contexts. The temporal duration is associated to elderly activities, so activities are placed in the following of spatial contexts and the temporal duration is placed after activities. Between the spatial context reasoning and behavior reasoning of activities, a modified duration HMM is applied to extract activities. According to this design, human behaviors different in spatial contexts would be distinguished in first module. The behaviors different in activities would be determined in second module. The third module is to recognize behaviors involving different temporal duration. By this design, an abnormal signaling process corresponding to different situations is also placed for application. The developed approach has been applied for understanding of elder behaviors in a nursing center. Results have indicated the promise of the approach which can accurately interpret 85% of the elderly behaviors. For abnormal detection, the approach was found to have 90% accuracy, with 0% false alarm.

Keywords: Behavior recognition; Duration HMM; Hierarchical HMM; Context

1. Introduction

Due to the lengthening of the human ages, the elderly's daily health care has become one of the most critical issues in our society. As such, how to use the current technology for improving the well being of the elderly daily life has become increasingly important. Due to the increased necessity of assisting elderly care, there are several nursing centers being established, some of which are installed with cameras for monitoring the elderly situation in every bedroom and hallway, for preventing them from unexpected accidence. However, this approach requires a dedicated person watching all of the screens at all time, which is a high human burden and cannot be avoided of the potential of human occasional negligence. Furthermore, monitoring abnormal behaviors should consider past behavior history and contextual environment event occurs. Thus an approach which can understand the elderly behaviors from their daily life based on video sequence would provide great assistance to monitor the elderly situation.

Many literatures have dedicated to human behavior understanding [1–7]. However, most of the results involve only the recognition of primitive action such as, walking, running, sitting, and etc., which are all far from the purpose of understanding daily life behaviors. The approach in Refs. [8,9] performs human behavior recognition through feature matching. Usually human behaviors would be different with personal profile. Thus, a learning mechanism should be applied into the recognition procedure in order to extract the dynamic human behaviors. Carter et al. [10] combined the Bayesian and Markov chain to recognize human behavior. Kumar et al. [11] proposed a framework for behavior understanding from traffic. The approaches in Refs. [12–14] perform behavior recognition through HMM, but they are only based on data sequence representations.

* Corresponding author. Tel.: +886 6 2757575x62373; fax: +886 6 2748678 or +886 6 2345482.

E-mail addresses: pcchung@ee.ncku.edu.tw (P.-C. Chung), dev@ee.ncku.edu.tw (C.-D. Liu).

0031-3203/$30.00 © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2007.10.022

As we know, an activity could be in different speed resulting in different time duration. To solve the problem of speed, duration HMM (DHMM) proposed in Refs. [15–18] included the duration consideration into recognition procedure. But these proposed methods focused only on data representations, without spatial context (SC) consideration. On the other hand, the approach in Ref. [19] performs behavior recognition by only sequence of SCs (denoted as landmark), and [20] performed behavior recognition based on SC and TC (temporal context) only. Both methods do not consider human activities in behavior recognition. As we know, a human behavior is usually perceived from human activities a person performs followed by his interactions with the surrounding environment. Thus, human behavior recognition should simultaneously consider SCs, temporal information and activities. Based on these considerations, a Hierarchical Context Hidden Markov Model (HC-HMM) which is enabled to include these three components into behavior recognition is proposed. The HC-HMM contains three modules, which are devised from the hierarchical architecture in Refs. [21–23]. As SC, activities and temporal information have primary and secondary relationship, these three modules are also constructed into a primary and secondary relationship. The first module plays as the primary by selecting human behaviors considering the SCs. From the selected human behaviors, the second module then screens the candidate behaviors according to human activities. Finally, the third module includes the temporal information in behavior reasoning (BR). In order to encompass the activity speed variations arisen from different people, a DHMM is devised between the first module and the second module, to segment and to extract activities. Thus the second module can perform the recognition relying on only the activity sequence, without being affected by the activity speed variation.

By this design, the HC-HMM is also able to signal abnormal situations. An abnormality could occur in DHMM, behaviors reasoning or temporal reasoning, which correspond to unknown activity, unreasonable activity and abnormal time duration. Relying on the abnormal analysis, the system can be applied into more applications. The remaining parts of the paper are described as follows. Section 2 describes feature extraction for activity recognition from video sequences. Section 3 presents HC-HMM architecture. Section 4 gives the experimental results. Finally the conclusions are drawn in Section 5.

2. Feature extraction and posture recognition

To extract an activity from video stream, it is necessary to detect the foreground objects and extract image features. A simple and common method to detect foreground objects is using the background model which involves subtracting with threshold to determine foreground pixels. The pixel intensity of a completely stationary background can be reasonably modeled as a normal distribution with two parameters: the mean $m(x)$ and variance $\sigma(x)$. The pixel $I(x)$ is detected as a background pixel if the Gaussian probability

$$p(I(x)) = G(I(x), m(x), \sigma(x))$$

is greater than a threshold value. The mean $m(x)$ and variance $\sigma(x)$ are estimated by statistics from testing video and are dynamically updated [1].

Based on the extracted foreground pixel, a posture is represented by a pair of histogram projection both in horizontal and vertical. Then the posture estimation can be calculated by using the horizontal histogram projection $H(x)$ and vertical histogram projection $V(x)$, computed through $i^* = \min_{1 \leqslant i \leqslant k}\{D^i\}$, where

$$D^i = \frac{1}{2}\left(\sum_x H(x)\log\left(\frac{H(x)}{h^i(x)}\right) + \sum_x h^i(x)\log\left(\frac{h^i(x)}{H(x)}\right)\right) + \frac{1}{2}\left(\sum_x V(x)\log\left(\frac{V(x)}{v^i(x)}\right) + \sum_x v^i(x)\log\left(\frac{v^i(x)}{V(x)}\right)\right),$$

the i^* is the obtained posture, which has the minimum Kullback–Leibler (KL) distance. The k here is the number of postures in the database. $h^i(x)$ and $v^i(x)$ are the horizontal and vertical projection, respectively, of the ith posture.

Besides the postures, an activity should also be determined by the composition of a sequence of motions. The motion computed from the motion history map (MHS) [24] is also used as the features in determining the activity. The MHS is computed as

$$MHS_t(x) = \begin{cases} 255 & x \in M, \\ \max(MHS_{t-1}(x) - 1, 0) & \text{otherwise}, \end{cases}$$

where M represents the set containing motion pixels involving frame subtraction through a threshold value. The max function here is to ensure that the values in the motion history are always larger than or equal to zero. The posture and the MHS are applied as the features in the following for activity recognition.

3. HC-HMM

Human behavior is composed of three components which are surrounding environment, human activities and temporal information. Thus, the same activities may represent entirely different behaviors under different contexts. For instances, an activity “walking” with SC sequence “door, sidewalk, bed” and “bed, sidewalk, toilet”. The former behavior could be “a person goes to bed” and the later behavior could be “a person goes to toilet”.

According to above descriptions, behavior understanding should take all of these contexts into considerations. Context plays an important role in behavior understanding. The contexts considered should include SC such as location, interaction equipments, etc., and TC such as time, duration, etc. In this paper, a HC-HMM is devised for taking contexts into behavior understanding. The HC-HMM includes 3 reasoning components which are spatial context reasoning (SCR) module, the BR module, and temporal context reasoning (TCR) module. As we know, the three components of environment, activities and temporal, usually have the primary and the secondary relationship. Under one SC, there are only some specific behaviors. In other words, the same activity sequence under different SCs may

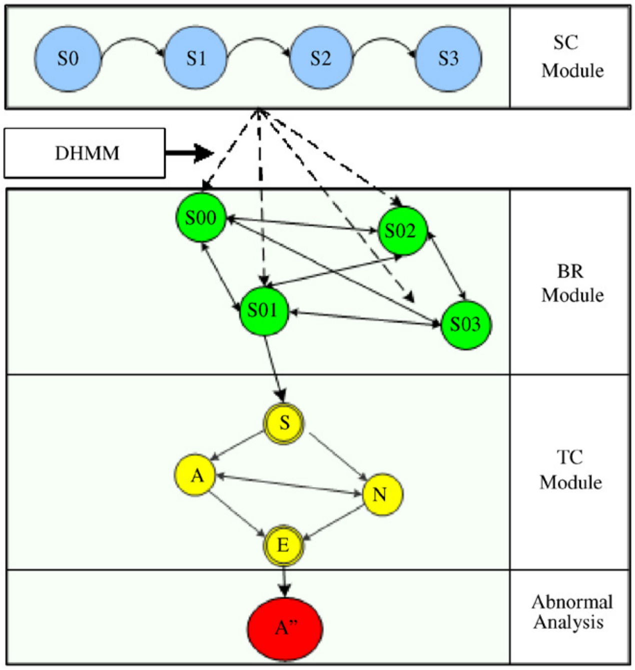

Fig. 1. The architecture of the daily behavior enabled HC-HMM, which contains the module of SCR, BR, TCR, and DHMM layer.

represent different behaviors. So SCR is designed to be the first module of HC-HMM for supporting the BR under SC. Under the SCR module is the BR, which is designed to infer human behaviors under a specific SC, based on a sequence of human activities. Finally, the third layer is the TCR module, which is designed to take in the time constraint for each activity. In order to extract human activities, while considering different people may perform activities at different speed, resulting in different time duration for each activity, a DHMM is designed as the middle layer between the SCR module and BR module, so as to compose human activities without being affected by the variation of time duration of activities resulted from different people. Consequently, the HC-HMM is constructed into a hierarchical inference structure. The upper levels serve as the primary of the lower levels. Each module in this structure is responsible for one component inference. With these three modules and DHMM, the architecture is shown in Fig. 1. Fig. 2 shows the implementation diagram. The foreground object is extracted from video sequence. History map and histogram projection are computed from the foreground objects. By taking the history map and histogram projection, the activity is recognized by a DHMM. The temporal information is also extracted by the history map. The detected spatial locations of the foreground object are taken into SC module for inference. In each inference state of SC module, the activities are taken into BR module for inference. The temporal information is taken into the TC module for inference in each BR state. By the three layer inference, the daily behavior would be recognized in the final state of SC module.

3.1. SC reasoning

Human behavior can be more or less interpreted through the support of a sequence of SC. For example, a person is initially in the room, then walks across the sidewalk to the bathroom, finally back to the room. With the SC, the person possibly goes to the toilet or has a shower instead. To use the SC for human behavior understanding, let $O^{SC} = O_1^{SC}, O_2^{SC}, \ldots, O_{T_{SC}}^{SC}$ be the context sequence observed in the SC layer, where each O_t^{SC} contains the context at time instant t. To reason the human activity from observed contexts, HMM is applied in this layer. The HMM is denoted as $\lambda^{SC} = \{\pi^{SC}, A^{SC}, B^{SC}\}$, where π^{SC} is the initial distribution, A^{SC} is the transition distribution and B^{SC} is the observation distribution. Let $A^{SC} = \{a_{ij}^{SC}\}$ be the state transition from state i to state j, and $B^{SC} = \{b_j^{SC}(O_t^{SC})\}$ represents the probability of the person approaching the context j which is computed as $b_j^{SC}(O_t^{SC}) = G(O_t^{SC}, \mu_j^{SC}, \sigma_j^{SC})$, where G is a Gaussian distribution with mean μ_j^{SC} and variance σ_j^{SC}. Let

$$\alpha_t^{SC}(i) = P(O_1^{SC}, O_2^{SC}, \ldots, O_t^{SC}, q_t^{SC} = S_i | \lambda^{SC})$$

be the forward estimator of observation sequence O^{SC} in state i at time t, where q_t^{SC} represents the state at time t. With these notations, the probability of human behavior with an SC sequence can be determined as $P(O^{SC}|\lambda^{SC}) = \sum_{i=1}^{N} \alpha_{T_{SC}}^{SC}(i)$, where N is the number of states of λ^{SC}.

Under the constraint condition, it is then possible to give more precise BR, since each rational behavior is performed under some SC situations. As behavior is the composition sequence of activities, between the SCR and the BR, the system is embedded with a middle layer for activity extraction.

3.1.1. Activity feature extraction by duration-like HMM

As we know, different people may have different activity speeds. An activity can be represented by a sequence of postures and motions with orders. Thus how to use the relatively fixed postures and motions as the sequence landmark for accommodating duration difference is one approach solving the time duration diversity. Here a duration-like DHMM is adopted for solving this problem. Let $\lambda^A = (\pi^A, A^A, B^A, P^A)$ represent a human activity model with observations $O^A = O_1^A, O_2^A, \ldots, O_{T_A}^A$, where O_t^A is the input observation at time t. Each state in λ^A represents a key posture of an activity. $A^A = \{a_{ij}^A\}$ is the state transition from state i to state j, $B^A = \{b_i^A(O_t^A)\}$ is the probability of observation O_t^A in state i, $\pi^A = \{\pi_i^A\}$ is the initial probability of state i, and $P^A = \{P_i^A(d)\}$ is the probability of state i with duration d. Also let $O = \{O_1, O_2, \ldots, O_T\}$ be the frame sequence for activity recognition, where each $O_t = [\rho_t, v_t]$, $1 \leqslant t \leqslant T$, contains the detected posture ρ_t and motion history v_t, in frame t. To take $O = \{O_1, O_2, \ldots, O_T\}$ into λ^A for activity recognition, the frame sequence $O_1, O_2, \ldots, O_T$ is divided into T_A segments $O_1^A, O_2^A, \ldots, O_{T_A}^A$ based on the same postures in a sequence. Each observation O_t^A represents a sub-sequence $O_{s_t}, O_{s_{t+1}}, \ldots, O_{s_t+d_t-1}$ in O having the same posture, where s_t is the starting point of the sub-sequence and d_t is the length of the sub-sequence. The state observation for DHMM is then computed as $O_t^A = [\rho_{s_t}, \frac{1}{d_t}\sum_{i=s_t}^{s_t+d_t-1} v_i]$, where ρ_{s_t} is the

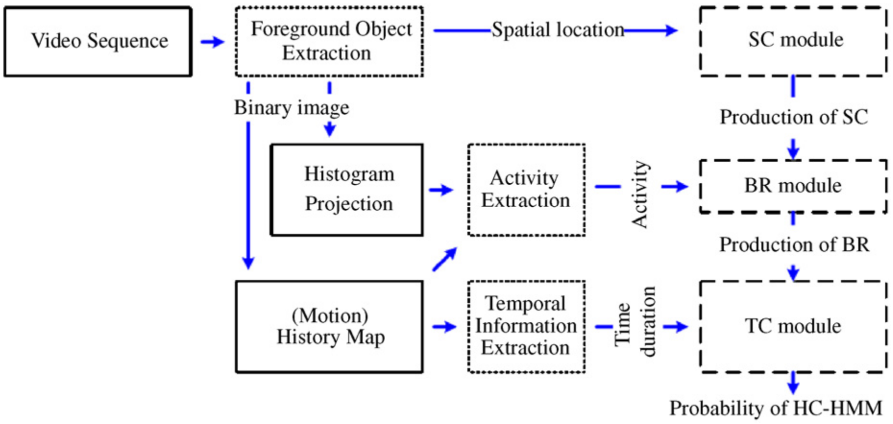

Fig. 2. The implementation step diagram.

posture symbol of O_t^A. By the forward estimator, the probability of observation $O_1^A, O_2^A, \ldots, O_{T_A}^A$ is computed by:

$$P(O^A|\lambda^A) = P(O_1^A, O_2^A, \ldots, O_{T_A}^A|\lambda^A) = \sum_{j=1}^{N^A} \alpha_{T_A}^A(j),$$

where N^A represents the number of state for the key postures of an activity. The probability of $\alpha_t^A(j)$ in forward–backward learning is computed as:

$$\begin{aligned}\alpha_t^A(j) &= P(O_1^A, O_2^A, \ldots, O_t^A, q_t^A = S_j^A|\lambda^A)\\ &= \left[\sum_{i=1}^{N^A} \alpha_{t-1}^A(i) * a_{ij}^A * G(d_i, \mu_i^A, \sigma_i^A)\right] * b_j(O_t^A),\end{aligned}$$

where S_j^A is the state j in model λ^A, q_t^A is the state in time t, G is a Gaussian estimator to determine the probability with time duration d_i staying in state i, μ_i^A and σ_i^A are the mean and variance, respectively, of time duration in state i.

The activity recognition is determined through the maximization of the observation probability of activity models under the prior knowledge constraints of key postures and motion orders for each activity. To apply the prior knowledge for activity recognition, let $Q^A = (q_1^A, q_2^A, \ldots, q_{T_A}^A)$ be the state sequence of activity observation. Then the detected activity Ω with the best path Q^A is determined by Viterbi Algorithm under prior knowledge as follows:

$$\Omega = \arg\max_k (P(q_1^A, q_2^A, \ldots, q_{T_A}^A, O^A|\lambda_k^A, W)),$$

where W is the prior knowledge of key postures and motion orders, and k is the activity index.

With the detected activity Ω and activity probability $P(O^A|\lambda^A)$, the next is to carry out the BR based on the detected activities.

3.2. Behavior reasoning (BR) with activity sequence under the spatial context (SC)

The purpose of BR is to inference human behavior with a sequence of activities under SC. Let $\Omega_1\Omega_2\ldots\Omega_{T_{BR}}$ be the detected activity sequence by DHMM under SC, and $P(\Omega_1)P(\Omega_2)\ldots P(\Omega_{T_{BR}})$ is the probability sequence corresponding to $\Omega_1\Omega_2\ldots\Omega_{T_{BR}}$. Then the BR is to take the activity sequence for behavior inference by computing the probability of $P(\Omega_1, \Omega_2, \ldots, \Omega_{T_{BR}}|\lambda^{BR})$ with activity sequence $\Omega_1, \Omega_2, \ldots, \Omega_{T_{BR}}$ in behavior model λ^{BR}. Let $\lambda^{BR} = (\pi^{BR}, A^{BR}, B^{BR})$ represents a behavior model under SC, where $\pi^{BR} = \{\pi_i^{BR}\}$ is the initial probability in state i, $A^{BR} = \{a_{ij}^{BR}\}$ is the transition probability from state i to state j in BR, and $B^{BR} = \{b_i^{BR}(\Omega_t)\}$ is the probability of observation Ω_t in state i. The observation probability is determined by computing the conditional probability of activity Ω_t at time t in state i as $b_i^{BR}(\Omega_t) = P(\Omega_t|S_i)$. Then the probability of BR under SC is computed as

$$P(\Omega_1, \Omega_2, \ldots, \Omega_{T_{BR}}|\lambda^{BR}) = \sum_{i=1}^{N^{BR}} \alpha_{T_{BR}}^{BR}(i),$$

where N^{BR} is the number of states in BR layer.

Once the probability of activity-based BR is computed, the behaviors of high probabilities are retained for recognition by combining the SCs. Let $\overline{a_k^{BR}}$ be the indicator showing whether the behavior model k is retaining. Then $\overline{a_k^{BR}}$ can be defined as

$$\overline{a_k^{BR}} = \begin{cases} 1 & \text{if } P(\Omega_1, \Omega_2, \ldots, \Omega_{T_{BR}}|\lambda_k^{BR}) \geqslant \alpha,\\ 0 & \text{otherwise,}\end{cases}$$

where α is a threshold value. Then the human BR combined with SC is computed as

$$\begin{aligned}&P(O^{SC}, \Omega_1\Omega_2\ldots\Omega_{T_{BR}}|\lambda^{SC}, \lambda_k^{BR})\\ &\quad= \sum_{i=1}^{N}\left[\alpha_{T_{SC}}^{SC}(i) * \left[\sum_{j=1}^{N^{BR}} \alpha_{T_{BR}}^{BR}(j)\right] * \overline{a_k^{BR}}\right],\end{aligned}$$

where $\alpha_t^{BR}(j) = [\sum_{i=1}^{N^{BR}} \alpha_{t-1}^{BR}(i) * a_{ij}^{BR}] * b_j^{BR}(\Omega_t)$.

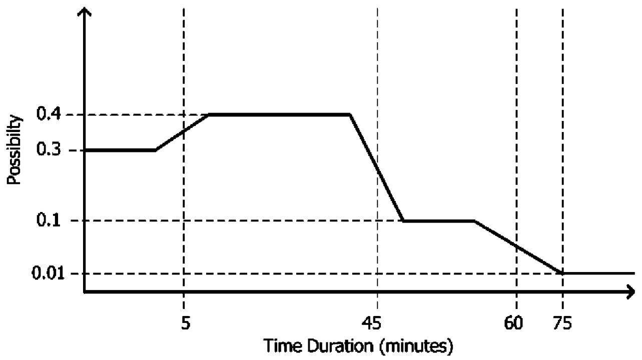

Fig. 3. The possibility time duration function of activity "sitting" in behavior "watching TV".

3.3. TC reasoning

Besides SC and activities, behaviors involving different time periods of activities could also indicate different meanings. This TCR module is designed for taking time periods into BR. To consider the time duration in BR, each activity k in each behavior l is associated with a TCR model $\lambda_{k,l}^{TC}$ and a possibility time duration function $b_{k,l}^{TC}(T_{k,l})$ derived from personal routine behaviors. One example of possibility time distribution function associated with activity "sitting" in behavior "watching TV" is shown in Fig. 3.

Each TCR model $\lambda_{k,l}^{TC}$ contains four states which are the initial state $S_{k,l}^{S}$, abnormal state $S_{k,l}^{A}$, normal state $S_{k,l}^{N}$, and final state $S_{k,l}^{E}$. $\lambda_{k,l}^{TC}$ would automatically enter the initial state when activity k is detected. Then $\lambda_{k,l}^{TC}$ would remain in initial state for a time period T_{TC}. After that, $\lambda_{k,l}^{TC}$ transits to normal state $S_{k,l}^{N}$ if $b_{k,l}^{TC}\left(T_{k,l}\right) \geqslant \alpha$, or abnormal state $S_{k,l}^{A}$ if $b_{k,l}^{TC}\left(T_{k,l}\right) < \alpha$ for each time click. Once a TCR has been in $S_{k,l}^{N}$ or $S_{k,l}^{A}$, it would transit into final state $S_{k,l}^{E}$ and terminate when BR module has a state transition, namely the change of activity. When the TCR terminates, the $b_{k,l}^{TC}\left(T_{k,l}\right)$ representing the probability of $T_{k,l}$ for activity k under behavior l as well as the last state will be returned to the BR module. With the results, human behavior understanding is computed by considering the SC, TC and activities as follows:

$$P(O^{SC}, O^{BR}, O^{TC}|\lambda^{SC}, \lambda_{l}^{BR}, \lambda_{k,l}^{TC}) = \sum_{i=1}^{N}\left[\alpha_{T_{SC}}^{SC}(i) * \left[\sum_{j=1}^{N^{BR}}[\alpha_{T_{BR}}^{BR}(j) * b_{k,l}^{TC}(T_{k,l})]\right] * \overline{a_{l}^{BR}}\right].$$

With above description, the human behaviors would be recognized according to three components of SCs, activities, and temporal information.

3.4. Discussion of abnormal TC reasoning

An abnormal state returned from TCR module could be due to the occurrence of (a) an unknown activity in this system, which results in low probability of behavior, (b) an abnormal activity which is under SC consideration, and (c) the abnormal time duration for the activity. These three situations correspond to error occurring during DHMM recognition, BR, and TCR. In order to understand the situation, when an abnormal state is returned, the behaviors are further analyzed by the following sequence. First, the motion history of the activity is compared with the activity model in D which contains unexpected accident samples. The abnormal situation is taken as an unknown activity situation (a) if (max $P(\upsilon_t|D) \geqslant \alpha$), where υ_t is the acquired motion history and α is a threshold value. If (a) is not satisfied, the abnormal situation is considered as an abnormity under SC if $(P(\Omega_t|O_t^{SC}) \leqslant \beta)$, where $P(\Omega_t|O_t^{SC})$ represents the probability of activity Ω_t under SC O_t^{SC}, and β is a threshold value. When both situations are not satisfied, the abnormality is the temporal duration error as (c).

4. Experiment results

The developed system is applied in a nursing home (or nursing center) for tests. A video sequence captured from a week of daily life is applied to train HC-HMM parameters. The video sequence contains 5 types of daily behaviors—"an elderly walks to toilet and comes back for a rest", "an elderly takes a walk and comes back for a rest", "an elderly watches TV and comes back for a rest", "a elderly goes to eat meal and comes back for a rest", "an elderly takes a shower and comes back for rest" which occur under 6 types of SCs including "door", "bed", "toilet", "sidewalk 1", "sidewalk 2", and "window". The video sequence contains two elderly with these 5 types of daily behaviors. From the video sequence 23 segments are manually extracted, where each video segment is composed of one daily behavior occurred under a sequence of SCs, starting on the elderly getting up from the bed and ending in coming back to the bed. There are 5 video segments for training the daily behavior "an elderly walks to toilet and comes back for a rest", 4 video segments for training the daily behavior "an elderly takes a walk and comes back for a rest", 4 video segments for training the daily behavior "an elderly watches TV and comes back for a rest", 4 video segments for training the daily behavior "an elderly goes to eat meal and comes back for a rest", 6 video segments for training the daily behavior "an elderly walks to toilet and comes back for a rest". The video segment of each daily behavior is further partitioned into sub-video segments (called spatial segment) based on the differences of SCs. In our experiments, a total of 64 spatial segments are extracted. From each spatial segment, activity segments, each composed of one activity, are extracted. The segments of the same activity under the same behavior in the training sequences also reveal slightly different time durations. A total of 117 activity segments (in 3510 frames) are obtained from the 64 spatial segments in the experiment. From each frame of these 117 activities, the histogram projection of motion history [24] is extracted and applied to train the parameters of DHMM. Each DHMM is composed of four states and its observations are the histogram projections representing four stages of the activity: "initial stage", "first stage", "second stage", and "final stage" of an activity. In our experiment, a total of five activities—"running", "walking",

Table 1
Activity recognition rate for elderly in nursing center

Activity	Our method		Tao Zhao		Vili Kellokumpu	
	Matched/total	False positive	Matched/total	False positive	Matched/total	False positive
Walking	9/12	2	10/12	1	–	–
Running	4/6	3	5/6	2	–	–
Lying	7/8	2	–	–	7/8	2
Sitting	11/15	4	–	–	11/15	4
Standing	13/16	2	–	–	13/16	2
Total	44/57		15/18		31/39	

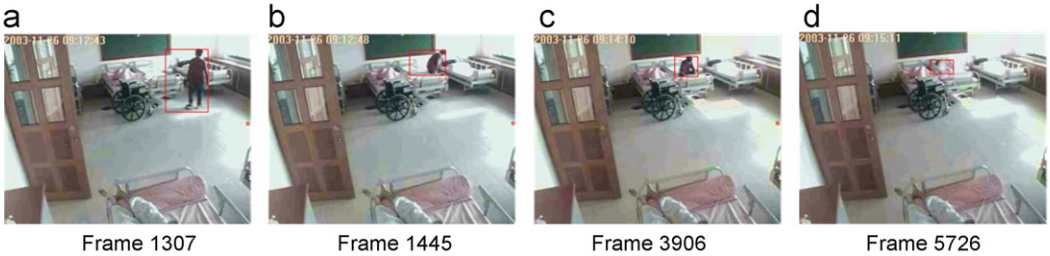

Fig. 4. In (a) a behavior “walking to bed” is detected, in (b) a behavior “sitting on the bed” is detected, in (c) an behavior “resting on bed” is detected, in (d) a behavior “lying on bed” is detected, and a behavior “go to sleep” is detected by time limitation.

“sitting”, “lying”, and “standing”—are to be recognized and therefore five DHMMs are designed in the system.

The activity sequence from each spatial segment is used to train the parameters of the BR module under the corresponding SC. In our experiment, there are three states designed for each BR model, and each state has five observations which are the activities “running”, “walking”, “sitting”, “lying”, and “standing”. Then the SC sequence from each daily behavior is used to train the parameters of SC layer. In our experiment for each daily behavior there are six states and six observations of the SCs “door”, “bed”, “toilet”, “sidewalk 1”, “sidewalk 2”, and “window”.

To evaluate the performance of the system in the recognition of human activities, behaviors, and abnormal behaviors, we take 7 video segments which are a total of about 120 min for testing. Although we have different elderly in training data, we still select 3 video segments containing the different elderly in training data, and 4 testing segments from one elderly in training data. These videos contain the daily life behaviors such as “going to sleep”, “going to bathroom”, “watching TV”, and “taking a shower”, etc., We first classify the videos manually into our defined behaviors and activities, to obtain a total of 15 daily behavior segments under different SC(s), total 49 behavior video segments in the 12 different types of behaviors shown in Table 3 under different activities in a SC, and a total of 57 activity segments of 5 different types of activities shown in Table 1. Fig. 4 shows video sample frames for behavior detection. In this video sequence, the sequence of SCs “door”, “sidewalk”, and “bed” is extracted from the SCR. The DHMM extracts human activities “walking” from frame 1004 to 1307, “sitting” from frame 1911 to 2021, and “lying” from frame 5762 to 5842. These activities are taken into BR under SC “door”, “sidewalk”, and “bed”. For each activity, a TCR module is created for the temporal information reasoning. Each TCR model is used to determine the probability of time duration of an activity under the spatial consideration, and the probability is returned to BR. According to the detected time duration of “lying”, the behavior “sleep” reveals the highest probability. Thus, the behavior “go to sleep” is recognized from the video frames.

Table 2
Activity recognition rate of “Fast Walking” and “Running” for normal people

Activity	Our method		Tao Zhao	
	Matched/ total	False positive	Matched total	False positive
Fast Walking	3/4	2	1/4	1
Running	4/6	1	5/6	3

The human activities detected for human behaviors recognition in this paper include “walking”, “lying”, “sitting”, “running”, and “standing”. These activities are the essential activities of normal life at home. In this paper, the activity recognition is performed by DHMM with a left–right model. To reduce the noise effect to activity recognition, the video segments for activity recognition have a minimum of 30 frames which equals to 3 s in 10 frames per second (fps). The activity recognition results for the test cases acquired from the nursing center are shown in Table 1 which reveals that among the 57 activities, 44 are correctly recognized, with the recognition rate more than 77%. The activity “running” has the lowest

Table 3
Behavior recognition rate

Behavior	Our method		Nam T. Nguyen		Thi V. Duong	
	Matched/total	False-positive	Matched/total	False-positive	Matched/total	False- positive
Goes to toilet	4/4	0	4/4	2	4/4	0
Takes a shower	2/2	0	#	#	2/2	0
(Stand) Stay by the side of chair for telephone	2/3	1	#	#	#	#
Sit on chair and watch TV	4/5	1	5/5	3	4/5	3
Sit on chair to rest awhile	5/6	1	#	#	5/6	1
Lie on bed to sleep	7/8	2	8/8	4	7/8	2
Sit on bed to read a book	4/5	1	#	#	#	#
Lie on bed to rest awhile	5/6	1	#	#	5/6	1
Eat Breakfast	2/3	1	3/3	3	2/3	1
Take a walk	3/3	1	2/3	1	3/3	1
Walk to window, and in a daze	2/2	0	2/2	0	2/2	0
Walk to window, and call a nurse	2/2	0	#	#	2/2	0
Total	42/49		24/25		36/41	

recognition rate for two reasons. First, “running” has small amount of testing data. Secondly, the activities “walking” and “running” of elderly are similar in test data. Therefore running is relatively less statistic and has less distinguishable features. Table 1 also shows that among the 15 “sitting” activity, 11 are correctly recognized, giving a relative smaller recognition rate of 73% compared with activities “standing” and “lying”. This is due to that activity “sitting” lies between activities “standing” and “lying”, and therefore has postures similar to “lying” and “standing”. The testing result obtained using Vili Kellokumpu’s [14] and Tao Zhao’s [9] methods on the same videos are also shown in Table 1 for comparison. The Vili Kellokumpu’s method used only posture sequence for activity recognition, without considering motions. Tao Zhao’s focuses on outdoor activity recognition, and therefore adopted the speed as a basic prior knowledge for activity recognition. Both methods were unable to recognize several activities, as shown by the symbol “-” in Table 1.

Besides the testing data of nursing center, the video sequences containing the activities “Running” and “Fast Walking” for normal people are also applied to compare our method and Tao Zhao’s method. The results are shown in Table 2. Although the recognition rate of Tao Zhao’s method is better than ours in “walking” and “running” in Table 1, Tao Zhao’s method is worse in their recognition of activity “Fast Walking”, as shown in Table 2.

The behavior recognition result in Table 3 shows that among the 49 behaviors, 42 are correctly recognized, with the recognition rate more than 85%. Table 3 also shows the recognition rate of our method, Nam’s method [19] and Thi’s method [20] on the same videos for comparison. Symbol ‘#’ represents that the proposed method is unable to recognize the behaviors. As mentioned, Nam’s method employed SCs for behavior recognition, and Thi’s method used both SCs and temporal information. The behaviors “Goes to toilet”, “Sit on chair and watch TV”, “Lie on bed for sleep”, “Eat Breakfast in dining room”, “Take a walk”, and “Walk to window and abstracted” are different in SCs. Thus, all of the proposed methods can distinguish these behaviors.

On the other hand, the behavior “Goes to toilet” versus “Takes a shower”, “Sit on chair and watch TV” versus “Sit on chair to rest awhile”, “Lie on bed to sleep” versus “Lie on bed to rest awhile”, “Walk to window, and in a daze” versus “Walk to window, and call a nurse” are different in temporal time duration. Thus, Nam’s method which takes only SCs in the reasoning cannot properly differentiate the behaviors. Since Thi’s and our method have considered the temporal time duration, both methods can distinguish the behaviors and also have similar performance. Although the performance is similar, but the behavior “Sit on chair and watch TV” has more false-positive in Thi’s method. This is due to that the time duration in Thi’s method is associated with SCs instead of activities. Thus the time durations of “staying aside by chair” and the followed “sitting on chair” are considered in one continuous period in BR. Due to this reason, if the elderly stands aside the chair for sometime before the behavior “Sit on chair to rest awhile“, this behavior “Sit on chair to rest awhile” would be classified as behavior “Sit on chair and watch TV”. In our proposed method, since the time duration is associated with the activities, the time duration when the elderly stands aside the chair for sometime is associated to activity “stand”, and time duration when elderly sits on chair to rest awhile is associated to activity “sit”. Thus the time durations for activities in the behavior “Sit on chair to rest awhile” can be accurately determined in the BR. Consequently we have less false-positive in behavior “Sit on chair and watch TV”.

Even so, the behavior “(Stand) Stay by the side of chair for telephone” versus “Sit on chair and watch TV”, “Sit on bed to read a book” versus “Lie on bed to rest awhile” differ mainly in activities. Since Nam’s and Thi’s method do not consider the activities in behavior recognition, they both cannot distinguish these behaviors as shown in Table 3.

In order to evaluate the capability of abnormality detection, we also test videos containing different abnormal situations.

Table 4
Abnormal behavior recognition rate

Abnormal behavior	Our method		Sebastian Luhr/ Kosuke Hara	
	Detected/ total	False positive	Detected/ total	False positive
Accident (faint)	3/3	0	3/3	0
Unreasonable activity (standing on bed)	2/2	0	–	–
Unreasonable activity (Running on sidewalk)	2/2	1	–	–
Abnormal time duration (sleeping too long)	1/1	0	1/1	0

The abnormal situations include accidents such as “faint”, unreasonable activity under SC such as “standing on bed” and “running on sidewalk” (the nursing center does not allow running on sidewalk), and abnormal time duration such as “sleeping too long”. Table 4 also shows the abnormal behavior detection rate, among the 8 abnormalities and 1 false alarm, 8 abnormalities are correctly detected, with more than 90% accuracy, and 100% of true-positive alarm detection, which reveals that all abnormal behaviors can be detected by our method. The results of Sebastian’s method [25] and Kosuke’s method [26] on the same videos are also shown on Table 4 for comparison. The symbol “–” in Table 4 represents that the method was unable to detect the behaviors. Despite that Sebastian’s [25] and Kosuke’s methods also include abnormal behaviors detection, they use only the time duration in the detection. As such, although the accident “faint” is detected to be abnormal by their method, they cannot further reason the abnormal situations. But due to applying the motion history on abnormality analysis, our method can detect the accident. On the other hand, from Table 4, we can also find that Sebastian’s and Kosuke’s methods cannot detect the abnormal behaviors “standing on bed” and “running on sidewalk” which contain the unreasonable activities in SCs. However, Table 4 shows that our method can recognize these types of abnormal behaviors with relatively high performance.

5. Conclusions

In view of the increasing necessity of elderly care, a hierarchical-context based human behavior understanding from video streams for a nursing center has been developed. In contrast to existing approaches which base on only postures and activities, our system embeds activities and contexts into a HC-HMM for behavior recognition. In HC-HMM, a behavior is established by a sequence of SCs which contains activities each of which is imposed with temporal reasoning. In this design, the activities, SCs and TCs are integrated in BR in the different modules. In order to obtain accurate activity recognition with speed variation, a DHMM is adopted. The speed variation is handled by the DHMM state duration which is determined by a sequence of the same postures. While high level behaviors exist only under certain contextual environments combined with activities, this approach provides a higher potential for behavior understanding. The developed approach has been applied to the understanding of the elderly daily life behaviors in a nursing center. Results have indicated the promise of the approach which can accurately interpret 42 behaviors among 49, with 85% accuracy rate. The approach is also employed for abnormal detection which was found to have accuracy of 8 among 9 abnormalities, with 90% accuracy rate, and with 100% true-positive alarm. The proposed method is currently applied in 2D video sequence, so the camera view and motion directions become an important issue of feature extractions. The system currently works in side-view with camera installed such that the optical axis is nearly perpendicular to the entrance direction of the bed room. Also the video segments for behavior recognition start from the patient leaving the bed and end in coming to the bed. In the future, we will improve the system to recognize daily behaviors with any starting and ending points, under which situations the identification of video segments for behavior recognition becomes much more challenging. The system is tested in the video sequences captured from the nursing center, and a life application installation is in the planning stage.

Acknowledgment

This work was supported by National Science Council, Taiwan, under Grant NSC 94-2213-E-006-012. The authors are corresponding with the Smart Media and Intelligent Living Excellence (SMILE) Lab.

References

[1] I. Haritaoglu, D. Harwood, L.S. David, W4: Real-time surveillance of people and their activities, IEEE Trans. PAMI 24 (8) (2000).

[2] J. Ben-Arie, Z. Wang, P. Pandit, S. Rajaram, Human activity recognition using multidimensional indexing, IEEE Trans. PAMI 24 (8) (2002).

[3] A.F. Bobick, J.W. Davis, The recognition of human movement using temporal templates, IEEE Trans. PAMI 23 (3) (2001).

[4] R. Polana, R. Nelson, Recognizing activities, in: Proceedings of the International Conference on IAPR, vol. 1, October 1994, pp. 815–818.

[5] H. Fujiyoshi, A.J. Lipton, Real-time human motion analysis by image skeletonization, in: Proceedings of the IEEE Workshop Applications of Computer Vision, October 1998, pp. 15–21.

[6] M.M. Rahman, K. Nakamura, S. Ishikawa, Recognizing, human behavior using universal eigenspace, in: Proceedings of the International Conference on Pattern Recognition, vol. 1, August 2002, pp. 295–298.

[7] N. Robertson, I. Reid, M. Brady, Behaviour recognition and explanation for video surveillance, in: The Institution of Engineering and Technology Conference on IEEE Crime and Security, 2006.

[8] H. Miyamori, S.-i. Iisaku, Video annotation for content-based retrieval using human behavior analysis and domain knowledge, in: Proceedings of the International Conference on Automatic Face and Gesture Recognition, March 2000.

[9] T. Zhao, R. Nevatia, Tracking multiple humans in complex situations, IEEE Trans. PAMI 26 (9) (2004).

[10] N. Carter, D. Young, J. Ferryman, A combined Bayesian Markovian approach for behaviour recognition, in: 18th International Conference on Pattern Recognition, 2006 (ICPR 2006).

[11] P. Kumar, S. Ranganath, H. Weimin, K. Sengupta, Framework for real-time behavior interpretation from traffic video, IEEE Transactions on Intelligent Transportation Systems, March 2005, vol. 6, pp. 43–53.

[12] J. Yamato, J. Ohya, K. Ishii, Recognizing human action in time-sequential images using hidden Markov model, in: Proceedings of the

IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1992 (CVPR ’92), 15–18 June 1992, pp. 379–385.

[13] A. Galata, N. Johnson, D. Hogg, Learning variable-length Markov models of behavior, Comput. Vision Image Understanding 81 (3) (2001) 398–413.

[14] V. Kellokumpu, M. Pietikäinen, J. Heikkilä, Human activity recognition using sequences of postures, in: Proceedings of the IAPR Conference on Machine Vision Applications (MVA 2005), Tsukuba Science City, Japan, pp. 570–573.

[15] L.R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proc. IEEE 77 (2) (1989) 257–286.

[16] M.-Y. Chen, A. Kundu, A complement to variable duration hidden Markov model in handwritten word recognition, in: Proceedings of the IEEE International Conference on Image Processing, 1994 (ICIP-94), vol. 1, 13–16 November 1994, pp. 174–178.

[17] M. Russell, A segmental HMM for speech pattern modeling, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, 1993 (ICASSP-93), vol. 2, 27–30 April 1993, pp. 499–502.

[18] X. Zhang, J.S. Mason, Improved training using semi-hidden Markov models in speech recognition, in: International Conference on Acoustics, Speech, and Signal Processing, 1989 (ICASSP-89), vol. 1, 23–26 May 1989, pp. 306–309.

[19] N.T. Nguyen, D.Q. Phung, S. Venkatesh, H. Bui, Learning and detection activities from movement trajectories using the hierarchical hidden Markov model, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), vol. 2, 2005, pp. 955–960.

[20] T.V. Duong, H.H. Bui, D.Q. Phung, S. Venkatesh, Activity recognition and abnormality detection with the switching hidden semi-Markov model, in: IEEE Computer Society Conference on CVPR 2005, vol. 1, pp. 838–845.

[21] S. Fine, Y. Singer, N. Tishby, The hierarchical hidden Markov model analysis and applications, Mach. Learn. 32 (1) (1998) 41–62.

[22] H.H. Bui, D.Q. Phung, S. Venkatesh, Hierarchical hidden Markov models with general state hierarchy, in: Proceedings of the Nineteenth National Conference on Artificial Intelligence, San Jose, CA, 2004, pp. 324–329.

[23] N. Oliver, E. Horvitz, A. Garg, Layered representations for human activity recognition, in: Proceedings of the Fourth IEEE International Conference on Multimodal Interfaces, 14–16 October 2002, pp. 3–8.

[24] A.F. Bobick, J.W. Davis, The recognition of human movement using temporal templates, IEEE Trans. PAMI 23 (3) (2001).

[25] S. Luhr, S. Venkatesh, G. West, H.H. Bui, Duration abnormality detection in sequence of human activity, Technical Report, Department of Computing, Curtin University of Technology, May 2004.

[26] K. Hara, T. Omori, R. Ueno, Detection of unusual human behavior in intelligent house, in: Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing, 4–6 September 2002, pp. 697–706.

About the Author—PAU-CHOO CHUNG received the B.S. and M.S. degrees in electrical engineering from National Cheng Kung University, Taiwan, Republic of China, in 1981 and 1983, respectively, and the Ph.D. degree in electrical engineering from Texas Tech University in 1991. In 1991, she joined the Department of Electrical Engineering, National Cheng Kung University, and has become a full professor since 1996. Currently, she also serves as the vice director of the Center for Research of E-life Digital Technology, National Cheng Kung University and the associate editor of Journal of Information Science and Engineering. Dr. Chung’s research interests include image analysis and pattern recognition, neural networks, video image processing/analysis, computer vision, and multimedia transmission. Particularly she applies most of her research results on medical applications, such as for medical image analysis, broadband telemedicine, pervasive home care, and multimedia-based behavior analysis. She received many awards, such as the annual best paper award in Chinese Journal Radiology 2001, the best paper awards from World Multiconference on Systemics, Cybernetics, and Informatics (SCI) 2001 and International Computer Symposium (ICS) 1998, Acer’s Best Research Award in 1994 and 1995, the best paper awards from the Conference of Computer Vision, Graphics, and Image Processing (CVGIP), in 1993, 1996, 1997, 1999, and 2001, Best Research Young Innovator Award (NSC) 1999. Dr. Chung has served as the program committee member in many international conferences. She is a senior member of IEEE and Member of Phi Tau Phi honor society.

About the Author—CHIN-DE LIU received the B.S. from the Department of Computer Science and Information Engineering, Tamkang University, Taiwan, in 1998, and the M.S. degrees from the Department of Electrical Engineering, National Cheng Kung University, Taiwan, in 2000. Currently, he is a Ph.D. degree student with the Department of Electrical Engineering, National Cheng Kung University. His current research interests are image processing and video image analysis.