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