On-Line Classification of Human Activities*

J.C. Nascimento[1], M.A.T. Figueiredo[2], and J.S. Marques[1]

[1] ISR-IST
jan@isr.ist.utl.pt, jsm@isr.ist.utl.pt
[2] IT-IST
mario.figueiredo@lx.it.pt

Abstract. In this paper we address the problem of on-line recognition of human activities taking place in a public area such as a shopping center. We consider standard activities; namely, *entering*, *exiting*, *passing* or *browsing*. The problem is motivated by surveillance applications, for which large numbers of cameras have been deployed in recent years. Such systems should be able to detect and recognize human activities, with as little human intervention as possible.

In this work, we model the displacement of a person in consecutive frames using a bank of switched dynamical systems, each of which tailored to the specific motion regimes that each trajectory may contain.

Our experimental results are based on nearly 20,000 images concerning four atomic activities and several complex ones, and demonstrate the effectiveness of the proposed approach.

1 Introduction and Problem Formulation

In this paper, we address the problem of (on-line) recognition of human activities in video sequences. Recently, this has become an active research area in computer vision, mainly driven by a large number of potential applications, such as video surveillance, computer-human interfaces, and contend-based video retrieval.

In a surveillance context, the analysis of the human behavior is often split into two parts: tracking and activity recognition [8]. Considering that tracking has seen tremendous recent progress [2,3,5,6,11,14,16], activity recognition has naturally become the next step to be addressed.

Different methods have been used to recognize human activities from the information extracted from video. The most popular techniques rely on *hidden Markov models* (HMM) and *coupled HMM* [12]. Both approaches are used to characterize the evolution of the person's mass center along the video sequence. A model termed *abstract HMM* was used to recognize human indoor motion patterns [10]. Other types of techniques have also been successfully used for gesture and activity recognition; e.g., Bayesian networks [7], neural networks [15], finite state machines (FSM) [1,4] and syntactic recognition [9].

* This work was supported by Fundação para a Ciência e a Tecnologia (ISR/IST pluri-anual funding) through the POS Conhecimento Program which includes FEDER funds.

J. Martí et al. (Eds.): IbPRIA 2007, Part II, LNCS 4478, pp. 444–451, 2007.

In this work, we consider that a tracking system computes the active region (bounding box) of the person along the video sequence. We also assume that the measurements provided by the tracker are corrected using the image to ground plane projective transformation, thus achieving viewpoint invariance and removing perspective distortion. Fig. 1 shows an example of an observed trajectory, before and after the projective transformation.

(a) (b)

Fig. 1. Original (left) and resulting transformed (right) images of the shopping center scenario

Our fundamental assumption is that the human (motion) activity can be inferred from the sequence of positions of the centroid of the person throughout the video sequence, which is provided by the tracker. After the projective transformation is applied, this sequence is denoted $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, where $\mathbf{x}_t \in \mathbb{R}^2$, for $t = 1, \ldots, n$, is the position at time instant t.

Our approach categorizes human activities using a two-level hierarchical system. At the lower-level, we have *dynamic models*, which are short term coherent units of movement; at the higher level, we consider *activities*, which are linearly ordered sequences of lower level *dynamic models*. In this paper, we consider five low level *dynamic models*: “moving left”, “moving right”, “moving up”, “moving down”, “stopped”. Four activities are considered: “passing”, “entering” “leaving” and “browsing”. Of course, this hierarchy could be extended to more complex arrangements of activities, but this will not be pursued in this paper.

Finally, our problem can be formulated as follows: *given a trajectory* $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$*, observed in a length-n time window, segment it into a sequence of low level dynamic models and classify it into one of the high level activities.*

The paper is organized as follows. Section 2 describes the adopted low level model and the parameter estimation method. Section 3 addresses the segmentation criterion. Section 4 describes the high level classification of sequences. Section 5 describes experimental results and Section 6 concludes the paper.