# Probabilistic Spatio-temporal 2D-Model for Pedestrian Motion Analysis in Monocular Sequences*

Grégory Rogez**, Carlos Orrite, Jesús Martínez***, and J. Elías Herrero

CVLab, Aragon Institute for Engineering Research, University of Zaragoza, Spain
{grogez, corrite, jesmar, jelias}@unizar.es
http://www.cv.i3a.unizar.es

**Abstract.** This paper addresses the problem of probabilistic modelling of human motion by combining several 2D views. This method takes advantage of 3D information avoiding the use of a complex 3D model. Considering that the main disadvantage of 2D models is their restriction to the camera angle, a solution to this limitation is proposed in this paper. A multi-view Gaussian Mixture Model (GMM) is therefore fitted to a feature space made of Shapes and Stick figures manually labelled. Temporal and spatial constraints are considered to build a probabilistic transition matrix. During the fitting, this matrix limits the feature space only to the most probable models from the GMM. Preliminary results have demonstrated the ability of this approach to adequately estimate postures independently of the direction of motion during the sequence.

## 1 Introduction

In recent years, human motion analysis has grown to become one of the most active research areas in computer vision [1]. It has a wide spectrum of promising applications in many fields, especially in video-surveillance where the possibility of automatic video understanding and activity recognition would enable a single human operator to monitor wide areas. The most efficient systems are based on the use of a model [2], which is, most of the time, a representation of the human body. The election of an appropriate model is a critical issue. The use of an explicit body model is not simple, given the high number of degrees of freedom of the human body and the self-occlusions, direct consequences of the monocular observation. In previous works, the structure of human body has been represented as 2D or 3D Stick figure [3], 2D (Active) Contour or Shape [4] or 3D volumetric model [5]. The benefits from using a more sophisticated and appropriate model can be reduced or annihilated by poor parameter estimates.

In this paper we present a probabilistic 2D model for pedestrian motion analysis in monocular sequences. The disadvantage of 2D models is their restriction

* This work is supported by grant TIC2003-08382-C05-05 from Spanish Ministry of Science and Technology (MCyT) and FEDER.

** Funded by FPU grant AP2003-2257 from Spanish Ministry of Education.

*** Supported by Spanish Ministry of Education under FPI grant BES-2004-3741.

F.J. Perales and R.B. Fisher (Eds.): AMDO 2006, LNCS 4069, pp. 175–184, 2006.

to the camera's angle. We therefore propose to construct 2D dynamical models independent of the orientation of the person with respect to the camera and that can respond robustly to any change of direction during the sequence.

To carry out this goal, we follow the methodology proposed by Bowden [6]. We construct a human model encapsulating within a Point Distribution Model (PDM) the information of the full body silhouette (given by the 2D Shape made of a series of landmarks located along the human contour) and the structural information (given by the corresponding 2D Stick figure). Both training and testing sets comprise of hand-labelled data. The CMU Mobo database [7] has been used for training and real video-surveillance sequences for testing.

The method is based on learning dynamical models. A series of local motion models is learnt by clustering the Stick figure subspace. Using this structure-based partitioning, correspondences between several different views of the same walking sequences are established. This leads to a clustering in the global Shape-Skeleton feature space where all the views considered are projected together. The different clusters correspond in terms of dynamic or view-point. We consider in this work the use of Gaussian Mixture Models (GMM) to cope with the problem of non-linearity of the model as proposed in various papers [8,9]. GMM are fitted to the total Shape-Skeleton training data using the Expectation Maximization (EM) algorithm [8,9]. Temporal and spatial constraints are considered to build a probabilistic transition matrix. This enables a frame to frame prediction of the most probable local models from the GMM that have to be considered.

Once the model has been generated (off-line), it can be applied (on-line) to real sequences. Given an input human blob provided by a motion detection algorithm, the model is fitted for inferring both body shape and posture.

The structure of the paper is as follows: in Sect. 2, we introduce probabilistic modelling. Model construction and fitting are respectively explained in Sect. 3 and Sect. 4. Results are presented in Sect. 5 and conclusions drawn in Sect. 6.

## 2 Probabilistic Modelling

Our Point Distribution Model (PDM) consists of 2D Shape landmarks concatenated with 2D Skeleton joints. The total space will be clustered following temporal approach (clusters $C_j$) as well as spatial approach (clusters $R_j$) as described in Section 3. The first one will partition the dynamic of the motion, and the second one, the direction of motion. The purpose of this probabilistic dynamic model is to obtain a transition matrix combining both constraints.

### 2.1 Markov Chain for Modelling Temporal Constraint

Following the standard formulation of probabilistic motion model [3], the temporal prior $p(S_t|S_{t-1})$ satisfies a first-order Markov assumption where the choice of the present state $S_t$ is made upon the basis of the previous state $S_{t-1}$. In the same way, if we partition the state space into $N$ clusters $\mathcal{C} = \{C_1, ..., C_N\}$, the conditional probability mass function defined as $p(C_j^t|C_k^{t-1})$ corresponds to