Features for Optical Flow Based Gait Classification Using HMMs

Dorthe Meyer * and Heinrich Niemann

University of Erlangen–Nuremberg,
Chair for Pattern Recognition (Informatik 5)
Martenstr. 3, D-91058 Erlangen, Germany
Email: demeyer@informatik.uni-erlangen.de

Abstract

This paper describes several features used in a system for automatic gait analysis. In most clinical systems markers are used to determine the trajectories. We use features derived from the optical flow field to recognize different kinds of gait like walking, hopping, running, and limping. This method does not presume any markers. Every sequence of images produces a sequence of feature vectors. These vectors are regarded as random variables. They are used to train hidden Markov models for different kinds of gait. The models are used for gait classification.

1 Introduction

Application of gait analysis can be found in several fields, for example medical diagnosis, physical therapy and sports. It is used to receive information about gait disorders of patients with knee or hip pain, or tumors. It is also possible to control cycles of motion for rehabilitation or training.

In most medical examination systems the trajectories are determined by markers which are attached to several points of the body. The major problems using markers are exact positioning and the shifting of the skin surface when the person is moving, which causes variations of the marker positions. Patients also may feel obstructed walking with stickers all over their body.

The evaluation in clinics is mostly done by the doctor using his experience. Our aim is to develop a system which classifies the gait automatically without the use of markers.

Other approaches for action recognition use just the local motion information for classification. [2] computes a binary motion image and a motion history image to describe the history of motion in a blurred sequence. [5] computes local motion statistics in xyt–cells. The feature vector consists of the summed normal flow in each cell. The classification of periodic action is done by a 3–D template match. [3] computes features from the optical flow field. The difference of the phase of these features in periodic actions are used for recognizing people by their characteristic gait.

[6] uses the gray level values of rows and columns in an image sequence or in difference images to extract features for gesture recognition. He uses hidden Markov models (HMMs) and a neural net for classification. [1] obtains 3–D data of body parts from a stereo blob tracker and classifies gestures by coupled HMMs.

The approach presented in this contribution does not presume any markers. We use features extracted from the optical flow field. To reduce the effect of background noise we detect the person by the difference of two succeeding frames. The size of the person is used for normalization. The recognition is done by hidden Markov models. In this paper we investigate different kinds of features and their use for gait classification.

The paper is structured as follows. In section 2 we introduce the approach and describe the use of hidden Markov models for gait classification. We give an overview on the system. Section 3 describes the different features we used for experiments which we present in section 4. An outlook on future work is given in section 5.

*The author is member of the Graduiertenkolleg 3–D Image Analysis and Synthesis sponsored by the German Research Foundation (DFG).

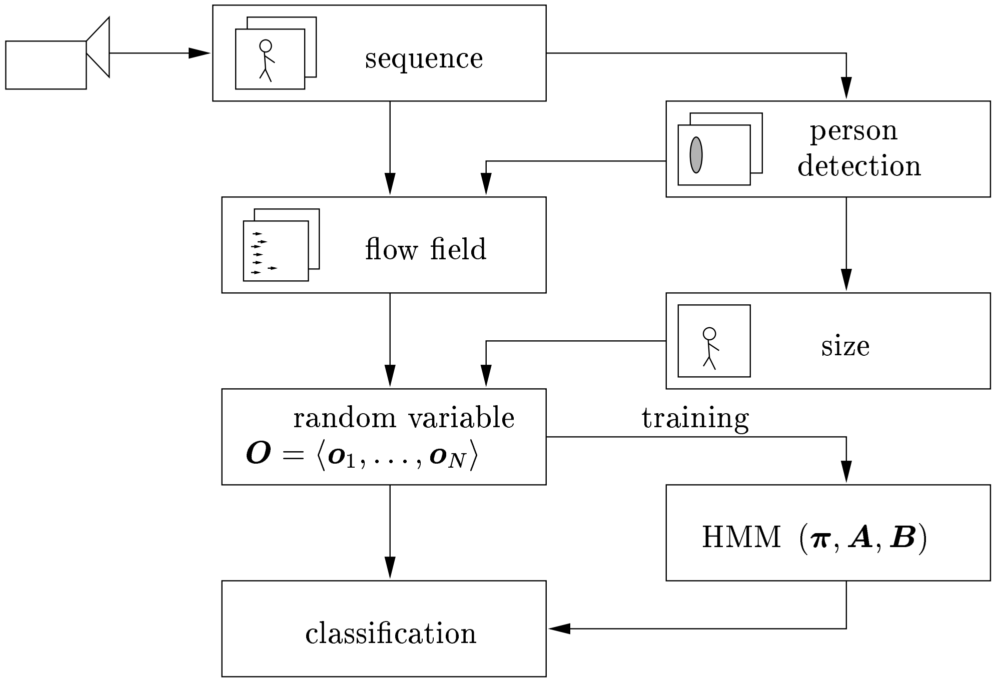

Figure 1: Overview on the system

2 Overview

The system classifies different kinds of gait like walking, running, hopping and limping without detecting body parts or using markers. The person is detected in an image sequence using the difference image of two succeeding frames. A region where the person is moving is computed. From the first difference image we obtain the size of the person. An image and the used sub image are shown in Figure 2.

The extraction of image features is done by the optical flow field. We compute several features which are described in the following section. We extract one observation vector from two succeeding frames, so $N+1$ images lead to N vectors. The dimension of the vectors is determined by the number of extracted features.

The observed feature vector in the n–th frame is denoted by $\boldsymbol{o}_n$, so the whole observed sequence will be described by a random variable $\boldsymbol{O} = \langle \boldsymbol{o}_1, \ldots, \boldsymbol{o}_N \rangle$. Human gait contains a periodic and a translatoric component, but only the periodic component will be used for classification. One period of walking consists of a left and a right step. A sequence usually consists of more than two steps.

The classification is done by hidden Markov models [4]. We use discrete HMMs. The output vectors $\boldsymbol{o}_n$ are quantized before training the HMMs and testing the sequences. The HMMs $(\pi, \boldsymbol{A}, \boldsymbol{B})$ consist of I states $\boldsymbol{S} = \{S_1, \ldots, S_I\}$. In the training phase the initial state probability π, the state transition probability matrix $\boldsymbol{A}$ and the output probability $\boldsymbol{B}$ are computed. We consider HMMs of degree one, the actual state just depends on the preceding state.

For every kind of gait Ω_κ, $\kappa = \{1, \ldots, 4\}$, one HMM is trained. In the classification phase the probabilities for each HMM to generate the observation sequence are computed and maximized, which means $\operatorname{argmax}_\kappa p(\boldsymbol{O}|(\pi, \boldsymbol{A}, \boldsymbol{B})_\kappa)$ has to be found. This is the optimal classifier if we consider the apriori probability as equal for all classes.

3 Features

For classification of gait without determine the speed of someone walking, we use periodic features. We compute several features and use combinations of them for the classification. All of them are derived directly from the optical flow field and the difference image. In order to reduce background noise only the region of the image where someone moving is detected is taken into consideration.

The velocity vector $\boldsymbol{v}(x, y)$ consists of components in x– and y–direction which are denoted

Figure 2: Image of a walking person. The sub image shown by the box is used for feature extraction.

$u(x,y)$ and $v(x,y)$. The mean of these vectors is

$$\bar{u} = \frac{\sum u(x,y)}{\text{number of vectors}},$$

$$\bar{v} = \frac{\sum v(x,y)}{\text{number of vectors}}.$$

The kinetic energy depends on the square of the velocity and we consider the mass of the persons to be distributed equally. Another feature is the mean of the energy

$$\bar{w}_{kin} = \frac{\sum |\boldsymbol{v}(x,y)|^2}{\text{number of vectors}}.$$

The center of gravity of the velocity and the energy is denoted as

$$y_{S,|\bar{\boldsymbol{v}}|} = \frac{\sum |\boldsymbol{v}(x,y)| \cdot y}{\sum |\boldsymbol{v}(x,y)|},$$

$$y_{S,w_{kin}} = \frac{\sum |\boldsymbol{v}(x,y)|^2 \cdot y}{\sum |\boldsymbol{v}(x,y)|^2}.$$

It varies more if for example the foot is moving fast and the trunk does not move and less if the whole body moves slowly.

We also consider the main direction which varies in the whole sequence periodically

$$\tan\phi = \frac{\bar{v}}{\bar{u}}.$$

The variance is

$${\sigma_u}^2 = \frac{\sum (u(x,y) - \bar{u})^2}{(\text{number of vectors}) - 1},$$

$${\sigma_v}^2 = \frac{\sum (v(x,y) - \bar{v})^2}{(\text{number of vectors}) - 1}.$$

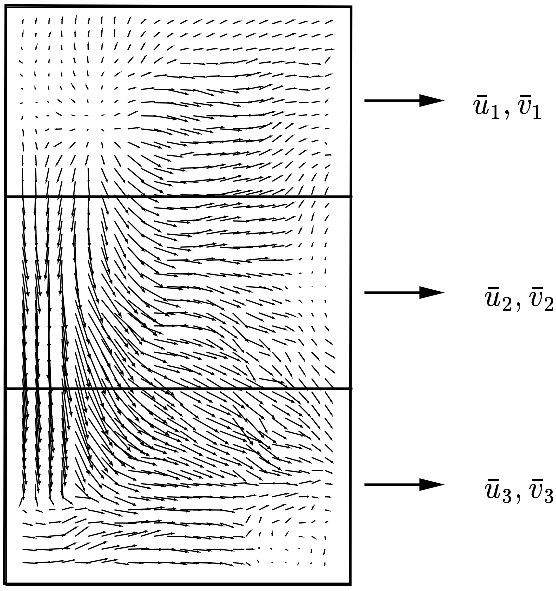

Figure 3: Flow image of a walking person. Three regions in y–direction are shown.

Other features are computed by dividing the region of a person into several sections along the $x-$ or the $y-$ axes and compute the features for every region. This results in more features per sequence and a higher dimensional feature vector. An example for 3 regions in y–direction is shown in Figure 3. The vector image is computed from the sub image in Figure 2. Using the mean of all three regions we get a six dimensional vector.

All features described above are periodic, but there is one disadvantage: they depend on the distance of the person to the camera. The flow of someone walking with the same speed in different distances parallel to the camera will be different in each case. A small person also moves with smaller displacements of his body parts than a large one. A normalization concerning the size of a person is needed.

The normalized features of the mean depend on the size of the person which is denoted by G and is derived from the first difference image of the sequence. The mean and scattering become

$$\bar{u}' = \frac{\bar{u}}{G}, \quad \bar{v}' = \frac{\bar{v}}{G},$$

$$S_u' = \frac{S_u}{G}, \quad S_v' = \frac{S_v}{G}.$$

The feature describing the center of gravity should be invariant against shifting the image in

	recognition rate
$\bar{u}$, $\bar{v}$	42 %
$\bar{u}'$, $\bar{v}'$	55 %
w_{kin}	51 %
w'_{kin}	51 %

Table 1: Recognition rates for normalized and non–normalized features

	recognition rate
$\tan\phi$	57 %
S'_u, S'_v	43 %
$y_{S,\lvert\boldsymbol{v}\rvert}, y_{S,w_{kin}}$	56 %
$u_1, v_1, u_2, v_2, u_3, v_3$	55 %
$\bar{u}, \bar{v}, S'_u, S'_v, y_{S,\lvert\boldsymbol{v}\rvert}, y_{S,w_{kin}}$	63 %

Table 2: Recognition rates for several features

y–direction, so this is also taken into consideration. y_0 is the y–coordinate of the foot, the lowest detected point of a person,

$$y'_S = \frac{y_S - y_0}{G}.$$

4 Experiments

We performed experiments with 243 sequences of 18 different persons. 64 of them show someone walking, 52 limping, 64 hopping and 63 running. The persons have different size and different distances to the camera. We trained several HMMs for every kind of gait, with different training sets. So every sequence was tested with a HMM not containing the sequence itself.

Table 1 shows the effects of normalization. Using the mean of the x– and y–component of the vector field, the recognition rates increase from 42 % to 55 %. Using non–normalized features results in person dependent and distance–dependent recognition. Small people and persons in large distance to the camera are for example often detected as limping. The kinetic energy does not show this effect.

The results for other features are shown in table 2. Regarding just the scattering without the mean of the velocities does not include much information on the motion. Combining several features results in the highest recognition rates. This will be taken into consideration by dividing the region into more than 3 regions.

5 Future

The flow field we used is very dense, but it has the disadvantage that there is smoothing in space. This means that different body parts moving into different directions cannot be distinguished in the field. We will perform experiments with other kinds of vector fields. It is also necessary to use more sequences of images of different people to cover the wide variety of human gait.

References

[1] M. Brand, N. Oliver, and A. Pentland. Coupled hidden markov models for complex action recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 994–999, 1997.

[2] J.W. Davis. Appearence-based motion recognition of human actions. Technical Report TR No. 387, M.I.T. Media Lab Perceptual Computing Group, Massachusetts, 1996.

[3] J.J. Little and J.E. Boyd. Describing motion for recognition. In *International Symposium on Computer Vision*, pages 235–240, Coral Gables, 1995.

[4] D. Meyer. Human Gait Classification Based on Hidden Markov Models. In H.-P. Seidel, B. Girod, and H. Niemann, editors, *3D Image Analysis and Synthesis '97*, pages 139–146, Erlangen, November 1997. Infix.

[5] R. Polana and R. Nelson. Recognizing activities. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 815–818, Seattle, Washington, 1994.

[6] G. Rigoll, A. Kosmala, and M. Schuster. A new approach to video sequence recognition based on statistical methods. In *Proceedings of the International Conference on Image Processing (ICIP)*, volume 3, pages 839–842, Lausanne, Schweiz, September 1996. IEEE Computer Society Press.