Gait Classification with HMMs for Trajectories of Body Parts Extracted by Mixture Densities

Dorthe Meyer*, Josef Pösl* and Heinrich Niemann
Universität Erlangen–Nürnberg
Lehrstuhl für Mustererkennung (Informatik 5)
Martensstr. 3, D–91058 Erlangen, Germany
[demeyer|poesl]@informatik.uni-erlangen.de

Abstract

In this paper we describe a system for automatic gait analysis. Different kinds of human gait are recognized using sequences of grey–level images. No markers are needed to get the trajectories of different body parts. The tracking of body parts and the classification are based on statistical models. We model several body parts and the background as mixture densities. The positions are determined iteratively, we begin with the most stable part to find. The anatomy of a human body restricts the area to search for the next one. From the trajectories, features for gait analysis are derived. These are used to train hidden Markov models (HMMs), one HMM for each kind of gait.

1 Introduction

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

In most medical examination systems the trajectories which are the curves the body parts describe 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. In gesture recognition people often have to wear coloured gloves. We develop a system which works without any markers and does not presume special clothing. The classification of gait is done automatically.

One example for motion analysis using markers is given in [11]. Leung attaches LEDs to the body, tracks them and computes the trajectories. The periodicy of the motion is used by matching the curvature of one period of the trajectory with a model trajectory. In contrast to this we do not presume any markers and train hidden Markov models for each kind of gait.

*The authors are members of the Graduiertenkolleg 3–D image analysis and synthesis sponsored by the Deutsche Forschungsgemeinschaft (DFG).

Figure 1: Sample images from sequences of people walking, limping, hopping and running

There are several approaches for motion tracking and the recognition of human body parts using a geometric model. Most of them presume contour detection which is not necessary in our system. [10] uses a model of the human body consisting of 14 cylinders with elliptic cross sections. He matches the lines of the image with the contours of the projected model. Hidden contours of the model are removed. [4] generates a 3–D model of the human body consisting of tapered super–quadrics. He uses several orthogonal views to track humans in action. [5] detects different body parts by an iterative approach using multiple views. Starting with a single deformable model, this is segmented into two parts if the model does not fit the following frame.

Head and hand position are tracked by a stereo Blob tracker [1]. This system is used to recognize Tai Chi gestures by HMMs using the motion of the head and the hands as features [3]. [2] computes blobs based on motion and colour similarity, spatial proximity and groupings in earlier frames. The limbs belong to one blob. For training HMMs for gait recognition he needs hand–labeled sequences or tracked markers.

Other approaches for action recognition use just the local motion information for classification. [7] 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. [9] uses the grey–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. In contrast to these systems we consider the motion of body parts, which we think is more detailed.

The paper is structured as follows. In section 2 we give a short overview on the system. This consists of two parts. The first one determines the trajectories of different body parts which is described in section 3. These are used to derive features for classification of gait based on HMMs afterwards, see section 4. In section 5 we show our experiments and conclude with an outlook on future work.

2 Overview

The system performs an automatic classification of different gaits from grey–level sequences. The classes are walking, running, hopping, and limping. The train and test images are sequences of people moving parallel to the camera in front of a static background. There are no restrictions concerning clothing. Example frames are shown in Figure 1.

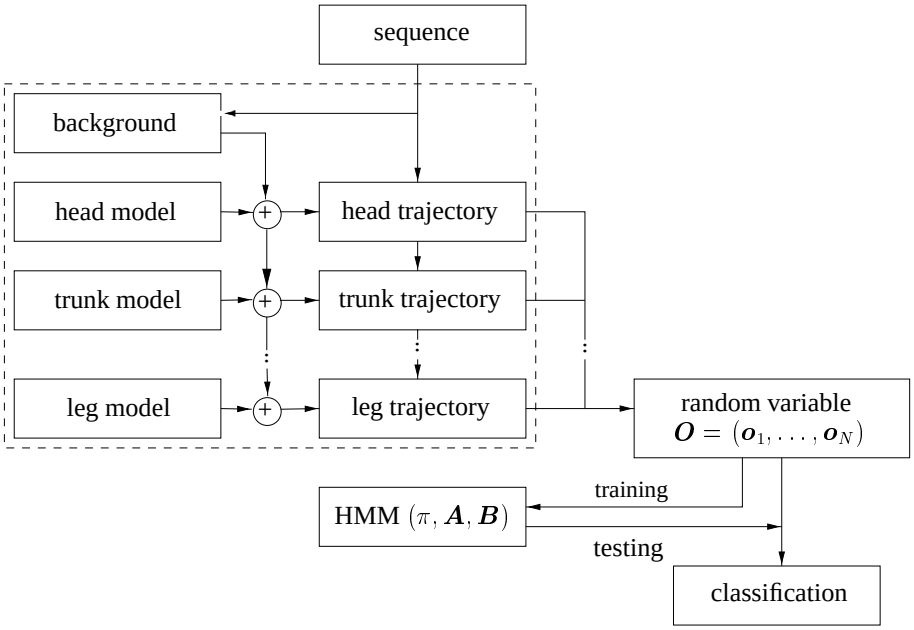

Figure 2: Overview on the system. The determination of trajectories is shown in the dashed box

An overview on the system is given in Figure 2. There are two major parts, the determination of trajectories and the classification of gait, both of them are based on statistical models. The first one is shown in the dashed box. We define densities for different body parts (head, trunk, leg) and the background. These are general models for the body parts derived from images of different people, the background is trained for each image sequence. We localize the body parts in every frame by mixture densities taking into account the anatomic relationships between the parts.

The trajectories are used to extract features from two succeeding frames. The features describe the periodic component of the motion of the body parts. They define random variables which are used to train an HMM for each kind of gait.

3 Tracking of body parts

The tracking and recognition of body parts has been addressed by many researchers. They either define geometric object models for the positions of segmentation features like edges or track the parts by correlation with object templates which may be adapted as the person moves in an image sequence. At least for the human face more sophisticated appearance based models have been evaluated [12].

We use a statistical approach for modelling a scene with static background (see also [8]) and a person moving parallel to the image plane. Local features are extracted for a lattice of rectangular image positions. For the body parts which are tracked we define independent density functions which are — together with the background — combined in a

mixture density for the complete image. Thereby a priori knowledge of the relative positions of the parts is incorporated. The object positions are determined by an expectation maximization approach.

3.1 Features

We apply a discrete wavelet transform to each image in order to extract local feature vectors. Given an image $f(x,y)$ with $x \in \{0,1,\ldots,D_x-1\}$, $y \in \{0,1,\ldots,D_y-1\}$ we combine the wavelet coefficients of different resolutions at each grid location $\boldsymbol{x}_m \in X$, $\boldsymbol{x}_m \in \mathbb{R}^2$ of a quadratic grid $X = [\boldsymbol{x}_m]_{m=0,\ldots,M-1}$ with resolution 2. Let $c_{m,r,i}$ ($i = 0,1,2$) be the high–pass and $d_{m,s}$ the low pass coefficients of a tensor product wavelet transform on scale $s \in \mathbb{Z}$ with resolution $r_s = 2^{1+s}$ at location $\boldsymbol{x}_m$. The wavelet analysis is performed up to level s_{N-2}: $s \in \{0,\ldots,N-2\}$. To reduce the directional dependency of the high–pass coefficients we compute the feature vectors $\boldsymbol{c}(\boldsymbol{x}_m) = (c_{m,0},\ldots,c_{m,N-1})^{\mathrm{T}}$ as:

$$\begin{aligned} c_{m,s} &= \log\left(\sum_{i=0,1,2}|c_{m,s,i}|\right), s = 0,\ldots,N-2 \\ c_{m,N-1} &= \log|d_{m,N-2}|\,. \end{aligned}$$

Feature vectors $\boldsymbol{c}(\boldsymbol{x})$ for arbitrary $\boldsymbol{x}$ are calculated by linear interpolation.

3.2 Statistical model

The feature vectors of an image result from static background and different body parts. This means that we have to consider a static background class Ω_1 and several object classes $\Omega_2,\Omega_3,\ldots$. Furthermore we model not all body parts and image distortions occur. These unknowns are modelled by arbitrary background Ω_0.

All features are assumed as independent. The local feature vectors $\boldsymbol{c}(\boldsymbol{x}_m)$ are concatenated in the image feature vector $\boldsymbol{c}$. Let $p(\boldsymbol{c}|\boldsymbol{B},\Omega_i,\boldsymbol{R}_i,\boldsymbol{t}_i)$ denote the density for the image features $\boldsymbol{c}$ for object Ω_i with model parameters $\boldsymbol{B}$. The 2–D rotation $\boldsymbol{R}_i = \boldsymbol{R}(\phi_i)$ and translation $\boldsymbol{t}_i$ describe the position of the object parts ($i \geq 2$) and are not necessary for background.

To define the statistical model of the complete image we first introduce independent models for each body part Ω_i ($i \geq 2$). The model object is composed of local feature vectors on a rectangular grid. The grid is identical to X if no object transformation is applied. Let $A_i \subset X$ be a small region which contains the object as shown in Figure 3. The features outside A_i are considered as arbitrary position independent background with normal density $p(\boldsymbol{c}(\boldsymbol{x})|\boldsymbol{B},\Omega_0) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)$. This leads to

$$\begin{aligned} \hat{p}(\boldsymbol{c}|\boldsymbol{B},\Omega_i,\boldsymbol{R}_i,\boldsymbol{t}_i) &= \prod_{\boldsymbol{x}_m\in A_i} p(\boldsymbol{c}|\boldsymbol{B},\boldsymbol{x}_m,\Omega_i,\boldsymbol{R}_i,\boldsymbol{t}_i) \prod_{\boldsymbol{x}_m\notin A_i} p(\boldsymbol{c}|\boldsymbol{B},\boldsymbol{x}_m,\Omega_0) \\ &= \prod_{\boldsymbol{x}_m\in A_i} \mathcal{N}\left(\boldsymbol{c}\left(\boldsymbol{R}_i\boldsymbol{x}_m+\boldsymbol{t}_i\right)|\boldsymbol{\mu}_{i,m},\boldsymbol{\Sigma}_{i,m}\right) \\ &\quad \prod_{\boldsymbol{x}_m\notin A_i} \mathcal{N}\left(\boldsymbol{c}\left(\boldsymbol{R}_i\boldsymbol{x}_m+\boldsymbol{t}_i\right)|\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0\right). \end{aligned}$$

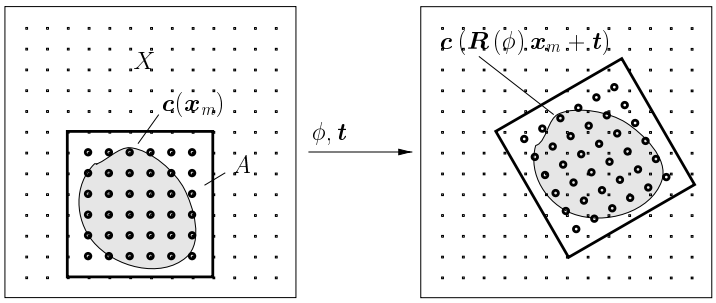

Figure 3: Object covered with grid for feature extraction

The value $\boldsymbol{c}\left(\boldsymbol{R}_i\boldsymbol{x}_m + \boldsymbol{t}_i\right)$ at each transformed location is computed by interpolation from the image feature vector $\boldsymbol{c}$. Thereby we assume the transformation of all local feature vectors as orthonormal. This is not exactly true for linear interpolation but experimental results show that this assumption is not critical.

The evaluation of this density is on a transformed grid. In order to evaluate the density on the image grid we have to apply the inverse transformation to the normal density parameters. The image features are not independent then but have local correlations. We ignore those correlations and use only diagonal variances $\boldsymbol{\Sigma}_{i,m}(\boldsymbol{R}, \boldsymbol{t})$, which constitute the model parameters $\boldsymbol{B}$ together with the mean vectors:

$$\begin{aligned} p(\boldsymbol{c}|\boldsymbol{B}, \Omega_i, \boldsymbol{R}_i, \boldsymbol{t}_i) &= \prod_{\boldsymbol{x}_m \in X} p(\boldsymbol{c}|\boldsymbol{B}, \boldsymbol{x}_m, \Omega_i, \boldsymbol{R}_i, \boldsymbol{t}_i) \\ &= \prod_{\boldsymbol{x}_m \in X} \mathcal{N}\left(\boldsymbol{c}(\boldsymbol{x}_m)|\boldsymbol{\mu}_{i,m}(\boldsymbol{R}_i, \boldsymbol{t}_i), \boldsymbol{\Sigma}_{i,m}(\boldsymbol{R}_i, \boldsymbol{t}_i)\right). \end{aligned}$$

The static background has no transformation parameters but is position dependent:

$$p(\boldsymbol{c}(\boldsymbol{x})|\boldsymbol{B}, \Omega_1) = \prod_{\boldsymbol{x}_m \in X} \mathcal{N}(\boldsymbol{c}(\boldsymbol{x}_m)|\boldsymbol{\mu}_{1,m}, \boldsymbol{\Sigma}_{1,m}).$$

Based on the assignment of each image location to one object class we can define a mixture density for the observation. Let $\zeta : X \to \{0, 1, 2, \ldots\}$ denote the assignment function which bears the hidden information to which class $\Omega_{\zeta(\boldsymbol{x}_m)}$ location $\boldsymbol{x}_m$ belongs. For a shorter notation we write $\zeta_m = \zeta(m) := \zeta(\boldsymbol{x}_m)$. Then the mixture density is

$$\begin{aligned} p(\boldsymbol{c}|\boldsymbol{B}, \boldsymbol{R}, \boldsymbol{t}) &= \sum_{\boldsymbol{\zeta}} p(\boldsymbol{c}, \boldsymbol{\zeta}|\boldsymbol{B}, \boldsymbol{R}, \boldsymbol{t}) \\ &= \sum_{\boldsymbol{\zeta}} p(\boldsymbol{c}|\boldsymbol{\zeta}, \boldsymbol{B}, \boldsymbol{R}, \boldsymbol{t}) p(\boldsymbol{\zeta}|\boldsymbol{B}, \boldsymbol{R}, \boldsymbol{t}) \end{aligned}$$

with $\boldsymbol{\zeta} = (\zeta(m))_{\boldsymbol{x}_m \in X}$ and $(\boldsymbol{R}, \boldsymbol{t}) = \left((\boldsymbol{R}_i, \boldsymbol{t}_i)_{i=2,\ldots}\right)$.

We model the a priori density of the assignment based on the body part positions as $p(\boldsymbol{\zeta}|\boldsymbol{B}, \boldsymbol{R}, \boldsymbol{t}) = \prod_{\boldsymbol{x}_m \in X} p(\boldsymbol{\zeta}_m|\boldsymbol{B}, \boldsymbol{R}, \boldsymbol{t})$. For each local assignment we only distinguish between background and body assignment $\hat{\boldsymbol{\zeta}} : X \to \{0, 1\}$:

$$p(\boldsymbol{\zeta}_m|\boldsymbol{B}, \boldsymbol{R}, \boldsymbol{t}) = k \cdot p(\hat{\boldsymbol{\zeta}}_m|\boldsymbol{B}, \boldsymbol{R}, \boldsymbol{t}),$$

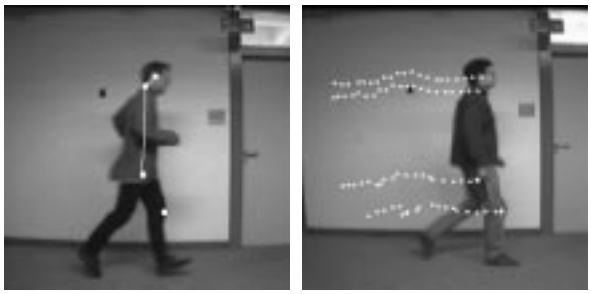

Figure 4: Left: Reference locations for the head (ear), trunk (neck) and leg (hip). The distances are marked. The fourth point is the knee, which is determined by the rotation of the leg and the length. Right: Trajectories of these four points out of 21 frames of a sequence.

where k is a norming constant. The body probability $p(\hat{\boldsymbol{\zeta}}_m = 1|\boldsymbol{B}, \boldsymbol{R}, \boldsymbol{t})$ is derived from the relative positions of reference locations $\tilde{\boldsymbol{x}}_i$ on each body part Ω_i. The reference positions are transformed to $\tilde{\boldsymbol{x}}_i(\boldsymbol{R}_i, \boldsymbol{t}_i)$. For two subsequent parts Ω_i and Ω_{i+1} we define the body probability

$$p_i(\hat{\boldsymbol{\zeta}}_m|\boldsymbol{B}, \boldsymbol{R}, \boldsymbol{t}) = \frac{p\left(|\tilde{\boldsymbol{x}}_i(\boldsymbol{R}_i, \boldsymbol{t}_i) - \tilde{\boldsymbol{x}}_{i+1}(\boldsymbol{R}_{i+1}, \boldsymbol{t}_{i+1})|\,|\hat{\boldsymbol{\zeta}}_m\right)}{\sum_{\hat{\boldsymbol{\zeta}}_m=0,1} p\left(|\tilde{\boldsymbol{x}}_i(\boldsymbol{R}_i, \boldsymbol{t}_i) - \tilde{\boldsymbol{x}}_{i+1}(\boldsymbol{R}_{i+1}, \boldsymbol{t}_{i+1})|\,|\hat{\boldsymbol{\zeta}}_m\right)}.$$

Thereby we assume normal distribution for the distance of body reference locations and a uniform distribution of the distance otherwise. The body probabilities are combined by $p(\hat{\boldsymbol{\zeta}}_m = 1|\boldsymbol{B}, \boldsymbol{R}, \boldsymbol{t}) = \prod_i p_i(\hat{\boldsymbol{\zeta}}_m = 1|\boldsymbol{B}, \boldsymbol{R}, \boldsymbol{t})$.

3.3 Pose search

The simultaneous pose estimation of all body parts is too time consuming. Therefore we choose an iterative approach in this paper where at each iteration step we search for one additional body part. All body parts are trained on different people to get general part models. The first object of the iteration is searched in the first image of the image sequence globally. The other parts and all subsequent images are only searched locally then.

The search itself is conducted by an expectation maximization approach, where the expectation term (Kullback–Leibler)

$$\mathcal{E}_{\boldsymbol{\zeta}}\left(\log\, p(\boldsymbol{c}, \boldsymbol{\zeta}|\boldsymbol{B}, \boldsymbol{R}, \boldsymbol{t})\,|\,\boldsymbol{c}, \boldsymbol{B}, \boldsymbol{R}, \boldsymbol{t}\right) = \sum_{\boldsymbol{x}_m \in X} \sum_{\boldsymbol{\zeta}_m} p(\boldsymbol{\zeta}_m|\boldsymbol{c}, \boldsymbol{x}_m, \boldsymbol{B}, \boldsymbol{R}, \boldsymbol{t}) \log p(\boldsymbol{c}|\boldsymbol{\zeta}_m, \boldsymbol{x}_m, \boldsymbol{B}, \boldsymbol{R}, \boldsymbol{t})$$

is maximized with respect to $(\boldsymbol{R}, \boldsymbol{t})$.

4 Classification of gait

4.1 Features for motion recognition

Trajectories of body parts contain a lot of information on someone moving. We can receive data about the velocities, acceleration and rotation of different body parts. Especially the information derived from the legs, feet and the centre of mass are important in clinical application. Analyzing gait by trajectories, physicians interpret the symmetry, amplitude and frequency of these curves.

The features we extracted from the trajectories are periodic, which is the important component of a movement. The amplitude of the trajectories are related to the size of the person. The form and symmetry of the trajectory does not depend on it.

As features we use displacements of body parts in x– and y–direction which are denoted $v_{i,x}$ and $v_{i,y}$ for the i–th body part. They are derived from two succeeding frames:

$$v_{i,x} = \frac{x_{i,n+1} - x_{i,n}}{\Delta t}$$

$$v_{i,y} = \frac{y_{i,n+1} - y_{i,n}}{\Delta t}.$$

$(x_{i,n}, y_{i,n})$ is the position $\tilde{\boldsymbol{x}}_i(\boldsymbol{R}_i, \boldsymbol{t}_i)$ of the i–th body part in the n–th frame. Δt denotes the frame rate of a sequence. The features are independent of sequences taken with different rates. These features are used especially for body parts which do not perform rotation, like the head and the trunk. The trunk contains the centre of mass, but the position of the head is more exact because of the trunk's deformation. The trajectories are similar.

Concerning the leg the rotation angle ϕ is a useful feature. It describes the flexibility of the joint which is important especially in medical applications.

4.2 Hidden Markov models

The classification is done by hidden Markov models [6]. These have been used in speech recognition successfully. We extract one observation vector from two succeeding frames, $N + 1$ images will lead to N feature vectors. 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 $\boldsymbol{O} = (\boldsymbol{o}_1, \ldots, \boldsymbol{o}_N)$. The dimension of the vector $\boldsymbol{o}_n$ is determined by the number of extracted features. This random variable will describe in general more than one period T which is one step with the right leg and one with the left leg. The sequence may start anywhere in the walking cycle. 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. The training is expected to end up in a cyclic left–to–right model, as the state transitions do not go backward in time, but include a periodic motion.

For each kind of gait one HMM is trained. The different kinds of gait, the classes are denoted as $\tilde{\Omega}_\kappa$, $\kappa = (1, \ldots, 4)$. In the classification phase the probabilities for each HMM to generate the observation sequence is computed and maximized, which means $\mathrm{argmax}_\kappa\, p(\boldsymbol{O}|(\pi, \boldsymbol{A}, \boldsymbol{B})_\kappa)$ has to be found.

Figure 5: Some result images. The four positions are marked. The last image shows a person whose body parts could not be found successfully.

5 Experiments

5.1 Trajectories

We performed our experiments with sequences similar to the sample frames shown in Figure 1. 110 frames of 29 people are used to train general models of three body parts, the head, the trunk and the leg, including different clothing and hair colour. The features for the localization of body parts (section 3.1) are extracted by the Johnston wavelet transform. To train the static background, the whole sequence is used as training material.

The recognition of body parts is done iteratively. The head is the first part to find as it is the most stable one, there is no occlusion or deformation. In the first frame of a sequence the head is searched globally, in the whole image. Then we consider the trunk. We define a reference location for every body part. For the position of the parts see Figure 4. For the head, the location is the ear. For the trunk, it is the neck position. The relationship between these two points can be considered by a constant distance between them, see section 3.2. This restricts the search area for the trunk to a region determined by the distance to the head. We assume normal distribution of distances as described in section 3.2, it is determined by sample frames when training the general models. For the leg we chose a reference location at the hip. The bigger variance concerning the distance to the neck causes a larger search area.

We tested the system for 96 sequences of 12 different people. Some localization results are shown in Figure 5. The positions for the head, neck and hip are marked. The fourth position shows the rotation angle of the thigh.

Table 5.1 shows the results. In 18 of 96 sequences the head could not be localized correctly in the first frame or later. Most of them are sequences of a person larger than the others. The person itself was found, but the assignment of body parts was not successful. The heads of all other persons are tracked well, if they are detected correctly in the first

	Localization correct	Recognition rate
Head tracking	78/96 (sequences)	81 %
Head position	2492/2632 (frames)	95 %
Trunk position	2492/2632 (frames)	95 %
Leg position	1704/2410 (frames)	71 %

Table 1: Results for the tracking of body parts.

	Classification correct	Recognition rate
Walking	40/53	75.5 %
Limping	23/40	57.5 %
Hopping	26/53	49.0 %
Running	31/47	66.0 %
Sum	120/193	62.2 %

Table 2: Results for the classification.

image. 95 % (just considering the 78 sequences) of the localization was exact. The right position of the leg is much harder to find. Just 71 % of the position were found correctly. Here we just considered 2410 frames, because in the beginning and and frames it was not completely in the frame. Correct positioning of the leg means that the left or right leg was localized correctly. We will get along with mixing up the two legs if we consider both of them.

5.2 Classification

We performed the experiments for classification using 193 sequences of 16 different people, 53 walking, 40 limping, 53 running and 47 jumping. 10 HMMs for each kind of gait are trained. Each sequence is tested by a model not containing itself for training. We used the trajectories of the head and the trunk. The feature vector is four–dimensional. We consider HMMs of degree one. The choice of four states is motivated by the gait cycle itself, representing the stance and swing phase of each leg.

The results are shown in Table 5.2. Walking people are recognized well. Most of the wrong results occur from hopping people to be detected as running. The confusion matrix is shown in Table 5.2. There are some possibilities to improve classification. The most obvious are using more sequences for training and features from other body parts. There is still a lot of unused information in the leg trajectory, we will use this in further experiments.

6 Future

In future work more body parts will be incorporated to develop a general model for the human body. Especially the second thigh and the feet are important for gait recognition. The system will be tested with a larger data set. Scaling and other views (3–D trajectories) will be taken into account.

	Walking	Limping	Hopping	Running
Walking	75.5 %	18.9 %	5.6 %	0 %
Limping	37.5 %	57.5 %	0 %	5.0 %
Hopping	9.4 %	0 %	49.0 %	41.5 %
Running	12.8 %	2.1 %	19.1 %	66.0 %

Table 3: Confusion matrix, it describes to which kinds of gait (columns) the sequences (rows) are assigned.

References

[1] A. Azarbayejani and A. Pentland. Real-time self-calibrating stereo person tracking using 3–D shape estimation from blob features. In *Proceedings of the* 13^{th} *International Conference on Pattern Recognition (ICPR)*, Vienna, 1996.

[2] C. Bregler. Learning and recognizing human dynamics in video sequences. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 568–574, 1997.

[3] L.W. Campbell, D.A. Becker, A. Azarbayejani, A.F. Bobick, and A. Pentland. Invariant features for 3–d gesture recognition. In *Second International Workshop on Face and Gesture Recognition*, Killington VT, 1996.

[4] D.M. Gavrila and L.S. Davis. Tracking of humans in action: A 3D model-based approach. In *ARPA Image Understanding Workshop*, Palm Springs, 1996.

[5] I.A. Kakadiaris and D. Metaxas. 3–D human body model acquisition from multiple views. In *Proceedings of the* 5^{th} *International Conference on Computer Vision (ICCV)*, pages 618–623, Boston, June 1995. IEEE Computer Society Press.

[6] D. Meyer. Human gait classification based on hidden Markov models. In B. Girod, H. Niemann, and H.-P. Seidel, editors, *3D Image Analysis and Synthesis '97*, pages 139–146, Sankt Augustin, November 1997. Infix.

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

[8] J. Pösl and H. Niemann. Object localization with mixture densities of wavelet features. In *International Wavelets Conference*, Tangier, Marocco, April 1998. INRIA.

[9] 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.

[10] K. Rohr. Towards model-based recognition of human movements in image sequences. *Computer Vision Graphics and Image Processing*, 59(1):94–115, 1994.

[11] P.S. Tsai, M. Shah, K. Keiter, and T. Kasparis. Cyclic motion detection for motion based recognition. *Pattern Recognition*, 27(12):1591–1603, 1994.

[12] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, March 1991.