Chapter 8 in *Motion-Based Recognition,* M. Shah and R. Jain (Eds.),
Kluwer Academic Publishers, Dordrecht Boston 1997, pp. 171-198

HUMAN MOVEMENT ANALYSIS BASED ON EXPLICIT MOTION MODELS

K. ROHR
Arbeitsbereich Kognitive Systeme
Fachbereich Informatik, Universität Hamburg
Vogt-Kölln-Str. 30, D-22527 Hamburg, Germany

1. Introduction

Within the field of computer vision the automatic interpretation of human movements is one of the most challenging tasks. A central problem in analyzing such movements is due to the fact that the human body consists of body parts linked to each other at joints which allows different movements of the parts. Therefore, the human body generally has to be treated as a nonrigid or more precisely as an articulated body. In addition, for general camera positions always some of the body parts are occluded. Although occlusions can provide important cues in a recognition task, the automatic interpretation is more difficult. Another problem that has to be dealt with is the clothing which can have a large influence on the appearence of a person (wide or tight trousers, different jackets, etc.). Clothing can also cause complex illumination phenomena that, in addition, change during movement (compare with efforts in the field of computer graphics to simulate cloth objects, e.g., [83]).

Because of these difficulties most existing approaches for analyzing human movements assume the joints of the human body to be marked (e.g., [66],[33],[81],[3],[27],[18],[65],[74]), or they are applied to synthetic images only (e.g., [61],[78],[76]). When using real-world images often special gymnastic movements are analyzed but not locomotion. In this case, the interpretation is generally less difficult because the effect of self-occlusion is diminished (e.g., [1],[47],[62]). Other approaches use stereo-images (e.g., [19]), restrict their analysis to certain parts of the body (e.g., [79],[85],[41],[25]), or analyze image sequences with more or less homogeneous background to diminish the segmentation problem (e.g., [49],[45],[28]). Often, image analysis is not carried out on an incremental basis but uses the entire sequence

(e.g., [60]). For references, see also [40] and the recent survey on motion-based recognition in [17].

In this chapter, we describe a motion-based approach for analyzing human movements in monocular real-world image sequences. We explicitly represent the human body as well as its movement and use this knowledge to estimate 3D positions and postures of persons from images. Central to our approach is the use of an explicit motion model which is based on analytically given motion curves for the body parts. These motion curves exploit data from medical motion studies and represent an average over a relatively large number of test persons. Additionally, we use a Kalman filter to incrementally estimate the model parameters from consecutive images. Since estimates from previous images are taken into account we obtain a smooth and robust result. Moreover, an initialization phase is performed to automatically estimate the initial model parameters. Our algorithm is designed for analyzing the movement of human walking which is the most frequent type of locomotion of persons. However, a generalization to other movement types such as, for example, running is straightforward.

Hogg [34],[35] in his pioneering work also has introduced a model-based approach for determining the 3D positions and postures of walking persons from real-world images. However, the motion model he uses does not exploit data from medical motion studies but has been acquired interactively from one prototype image sequence. Also, the initial model parameters are not provided completely automatically. Moreover, the estimates of the model parameters for the current image do not take into account previous estimates by using a Kalman filter scheme.

Automatic analysis of human movements has a wide spectrum of potential applications. For example, in street-traffic scenes it is important to early recognize situations which might lead to accidents. In the medical area it is important to analyze abnormal movement patterns.

The organization of this chapter is as follows. First, we give an overview of our motion-based approach. Then, we describe in more detail the human body model and the motion model of walking. After that, we show how this knowledge can be used to incrementally estimate the 3D positions and postures of persons from image data. The applicability of our approach will be demonstrated for real-world images. To indicate a possible extension of our work we finally describe a model of a cyclist.

2. Overview of our Motion-Based Approach

In our approach to the recognition of human movements from images we represent the shape of the human body by a volume model which is build of cylinders connected by joints ([51],[52]). Our motion model consists of

a set of analytically given motion curves which represent the postures of a person, i.e., the relative positions of all body parts. A nice property is that only one parameter, named *pose*, is needed to fully specify all postures during walking. Together with the 3D space coordinates $\vec{X} = (X, Y, Z)$ of the center of the torso (the origin of the person's coordinate system) and assuming the person to move parallel to the image plane we generally have to estimate four parameters from consecutive images. This is done using a Kalman filter where the system description represents the overall movement of the person, i.e., the movement of the center of the torso. The system description together with the motion curves of the body parts constitute our motion model. The model is explicitly given in an analytic form.

Our algorithm can be subdivided into two phases: initialization and incremental estimation. Whereas in the first phase the images are evaluated in a batch type manner, in the second phase processing is done on an incremental basis. The main parts can be summarized as follows.

1. *Initialization*

   Independent evaluation of about 10-15 images:

   - Detection of image regions corresponding to moving persons (using a change detection algorithm and binary image operations)
   - Estimation of the movement states, i.e., 3D positions and postures (using a calibration matrix for central projection and matching contours of the 3D model with grey-value edges)
   - Determination of starting values for the Kalman filter (using linear regression and the estimates from above)

2. *Incremental estimation*

   After initialization a Kalman filter scheme is applied to each image:

   - Prediction of the movement state (using estimation results from previous images)
   - Determination of measurements (using matching results of the 3D model to the current image)
   - Estimation of the current movement state (using the predicted movement state and the measurements)

The part of our work which deals with modelling the human body and its movement is a central subject within the fields of computer graphics, computer animation, or image synthesis. Thus, our approach for image sequence analysis utilizes methods from image synthesis, and therefore we can speak of an analysis-by-synthesis approach. The general relation between image analysis and image synthesis has been depicted in Fig. 1 (see also,

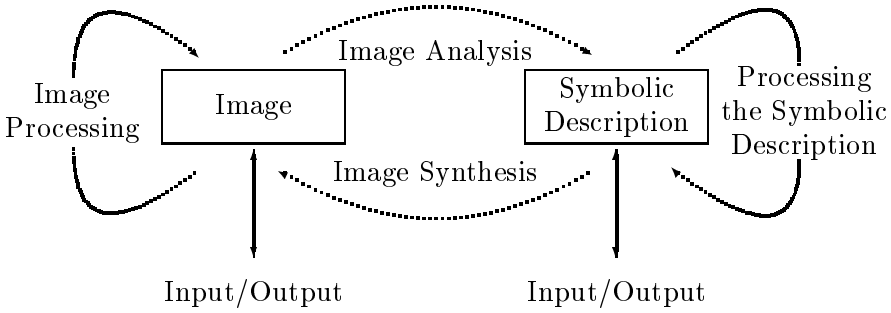

Figure 1. Relation between image analysis and image synthesis

e.g., [80],[46],[64],[15]). Whereas the aim of image analysis consists in deriving a symbolic description from real images, it is the aim of image synthesis to produce realistic images from a symbolic description. By using certain operations images can be transformed into images (image processing) or symbolic descriptions may be transformed into new symbolic descriptions.

3. Model of the Human Body

To represent static 3D models for the purpose of efficient recognition Marr and Nishihara [51] proposed to use volumetric primitives, an object-centered coordinate system, and a modular hierarchical organization on the description. As an example, they considered a 3D model of the human body consisting of cylinders.

In our approach we use this representation and describe the human body by 14 cylinders with elliptic cross-sections (head, torso and three primitives for each arm and leg) which are connected by joints (see also [23],[8],[6]). Cylinders are a good compromise between the number of parameters and the quality of representation of the human body. Each cylinder is described by three parameters: one for the length and two for the sizes of the semi-axes. The coordinate systems for the body parts are aligned with the natural axes. The origin of the coordinate system of the whole body is at the center of the torso (see Fig. 2). Transformations between different coordinate systems are described by homogeneous coordinates $\vec{X} = (X, Y, Z, 1)^T$:

$$\vec{X}' = \underline{A}\,\vec{X}, \qquad \underline{A} = \begin{pmatrix} \underline{R} & \vec{T} \\ \vec{0}^T & 1 \end{pmatrix}, \tag{1}$$

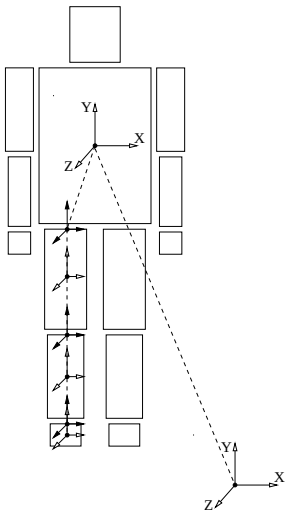

Figure 2. Human body representation

Figure 3. Rendered 3D model of the human body

where $\underline{R}$ is the 3×3 rotation matrix and $\vec{T}$ the translation vector. The inverse of $\underline{A}$ is given by

$$\underline{A}^{-1} = \begin{pmatrix} \underline{R}^T & -\underline{R}^T\vec{T} \\ \vec{0}^T & 1 \end{pmatrix} \tag{2}$$

If several transformations are applied in series then we can multiply the corresponding matrices $\underline{A}_i$, for example,

$$\underline{A}_i\underline{A}_{i+1} = \begin{pmatrix} \underline{R}_i\underline{R}_{i+1} & \underline{R}_i\vec{T}_{i+1} + \vec{T}_i \\ \vec{0}^T & 1 \end{pmatrix} \tag{3}$$

In our model of the human body we use absolute sizes of the body parts. Since persons are "in general" dressed, and since clothing can strongly influence the appearance of a person, the usefulness of existing catalogues of body measurements of unclothed persons (e.g., [21]) is limited. Therefore, we use sizes of the human body parts obtained by direct measurements of a normal person with average clothing. A visualization of our 3D model is displayed in Fig. 3 (see also [46]).

4. Motion Model of Walking

Models of human movements are studied within the fields of biomechanics, computer graphics, and robotics. Generally, one can distinguish between kinematic and dynamic methods for representing movements (e.g., [5],[77]).

A *kinematic* description explicitly specifies the geometry of objects, i.e., position, orientation, and deformation without taking into account the cause of the movement (e.g., [30],[14],[86]). If the movement is explicitly given by time-dependent functions then it is very easy to simulate movements. However, there are hardly any functions known for describing the complex movements of the human body (but see, for example, the laws of motion in [82] published early in 1836). Another possibility for simulating human movements is to interactively provide movement positions at certain time instants (keyframe-technique). However, this technique is time consuming and often does not lead to the desired result. Also, movements can be reconstructed interactively from recorded image sequences (rotoscopie). By this approach it is only possible to model such movements which have previously been performed. If, however, data from motion studies is already available then it is advantageous to use this data for simulation.

Dynamic methods, in contrast to kinematic schemes, take into account forces and torques (e.g., [84],[13]). These methods have the potential to produce realistic motions, however, they are computationally expensive and specifying forces and torques can be difficult. In addition, the resulting movements are not always satisfying and sometimes they are improved by kinematic adjustment (see also [44]).

In our approach to the recognition of human movements in real-world image sequences the agreement of the model with actual movements is important. Also, the number of parameters needed for specifying the model should be kept small in order to facilitate their estimation from images. Therefore, we decided to use a kinematic approach which exploits data from human motion studies. Studies of the human motion have a long tradition (see [2],[82]). At the end of the 19th century, photographic methods have been developed ([56],[50],[11]). For medical purposes Murray et al. [55],[54] have analyzed the movements of sixty normal men ranging in age from twenty to sixty-five years to obtain the basic elements of walking. In this study it is demonstrated that the motion curves of the body parts for different persons are very similar. Note, however, that this similarity is very astonishing if one imagines that it is often possible to identify persons by their gait. The fact that the motion curves are very similar opens us the possibility to use this data as knowledge source.

To represent the movement of walking in our motion model we have taken the average data from [55],[54]. For each of the joints at the shoulder, elbow, hip, and knee within one walking cycle we have taken function values which measure the relative angles between connected body parts, e.g., for the knee-joint the angle between the thigh and the lower leg. Additionally, we have used the values for the vertical displacement of the whole body which describe the periodic ups and downs of the center of the torso during

walking (see also [70]). Note, that with our motion model the overall movement is simplified in such a way that we have only modelled the movement of four joints and only considered joint rotations within the sagittal plane (which is spanned by the walking direction and the vertical direction). Since the joint movements of one side of the human body agree with those on the other side but are displaced about half of the walking cycle, we only need four (plus one) 1D motion curves for describing the overall movement of walking.

The motion curves of the joints at the shoulder, elbow, hip, and knee as well as for the vertical displacement of the whole body are based on the function values y_ν at certain points of time $t_\nu, \nu = 0, .., N-1$. Since walking is a periodic movement these values have been interpolated by periodic cubic splines $B(t)$ (see [73] and also [35]). Each interval $[t_\nu, t_{\nu+1}]$ is described by

$$B_\nu(t) = b_{1\nu}(t - t_\nu)^3 + b_{2\nu}(t - t_\nu)^2 + b_{3\nu}(t - t_\nu) + b_{4\nu}. \tag{4}$$

For periodic splines there is $t_N = t_0 + T$ (with T being the period). At the boundaries the function values as well as the first and the second derivatives agree, i.e., $y_0 = y_N$, $y_0^{'} = y_N^{'}$ and $y_0^{''} = y_N^{''}$. In the resulting system of N linear equations with the N unknowns $\vec{y}^{''} = (y_0^{''}, .., y_{N-1}^{''})$:

$$\underline{H}\,\vec{y}^{''} + \vec{d} = \vec{0}, \tag{5}$$

the matrix $\underline{H}$ represents the intervals $h_\nu = t_{\nu+1} - t_\nu$ between consecutive points of time t_ν, and the vector $\vec{d}$ represents the function values y_ν in conjunction with the h_ν. $\underline{H}$ is symmetric, diagonal dominant and positive definite. We solve (5) by Cholesky decomposition. With $\vec{y}^{''}$ we can compute $b_{1\nu}$, $b_{2\nu}$, $b_{3\nu}$, $b_{4\nu}$ and therefore $B_\nu(t)$ for each interval. In our case we took $N = 10$ function values at equidistant points of time and standardized the walking cycle to $T = 1$. The interval between two consecutive points of time therefore is $h_\nu = 0.1$.

The resulting motion curves of the joints and the vertical displacement are shown in Figs. 4 and 5, respectively. We use the parameter *pose* as some kind of time parameter which specifies the relative positions of all body parts within one walking cycle. A nice property is that we only need one parameter for this specification. Movement states for half of the walking cycle are shown in Fig. 6 ($pose = 0, 0.1, 0.2, 0.3, 0.4, 0.5$). Depicted are the contours of the cylinders under central projection. Hidden contours have been removed (see, e.g., [59],[29]). Fast reproduction of these motion states on a screen reveals that our motion model appears to be fairly realistic.

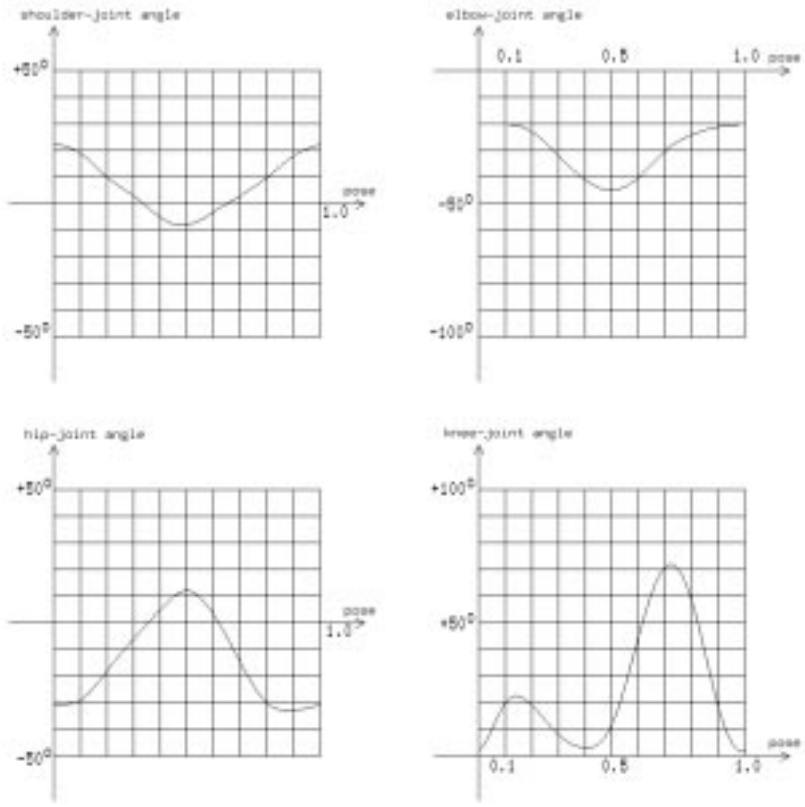

Figure 4. Motion curves of the shoulder, elbow, hip and knee joints

5. Incremental Estimation of the Model Parameters

In this section, we describe how our model of the human body and its motion is used to incrementally estimate the 3D positions and postures of persons from image sequences. Before applying a Kalman filter scheme we first perform an initialization phase. The main parts of this phase are change detection, 3D position estimation, and contour matching.

5.1. INITIALIZATION

5.1.1. *Change detection*

We assume the image sequences to be recorded with a stationary camera. To segment the images into regions corresponding to moving and nonmoving objects we apply a change detection algorithm. In each image point the intensities are approximated locally by a polynomial of second order and

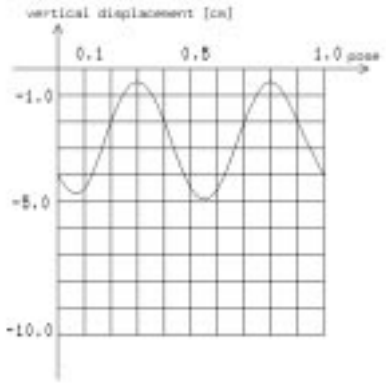

Figure 5. Vertical displacement of the whole body

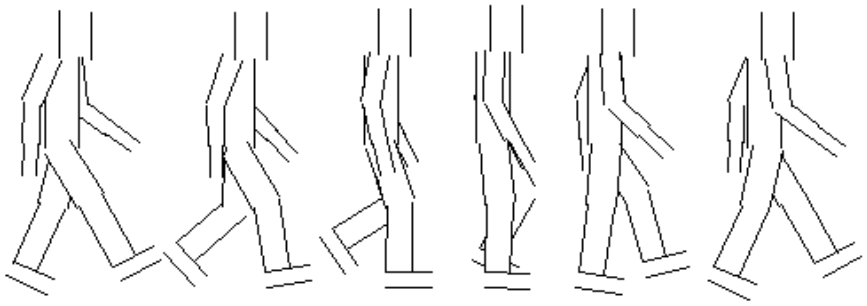

Figure 6. Movement states of walking

these polynomial fits are compared for consecutive images (see [10],[9]). Let $\underline{X}$ be the matrix representing the image coordinates within a 5×5 window and $\vec{\beta}$ the parameters of the fit. The n grey values within the window are given by $\vec{g}_k = \underline{X}\vec{\beta}_k$ and we can estimate $\hat{\vec{\beta}}_k = \underline{X}^{\#}\vec{g}_k$ by using the generalized inverse $\underline{X}^{\#} = (\underline{X}^T\underline{X})^{-1}\underline{X}^T$. We consider a point of an image k to represent systematic changes, if

$$\frac{1}{\sqrt{n}}\|\underline{X}\hat{\vec{\beta}}_k - \underline{X}\hat{\vec{\beta}}_j\|_2 = \frac{1}{\sqrt{n}}\|\underline{X}\ \underline{X}^{\#}(\vec{g}_k - \vec{g}_j)\|_2 > T \tag{6}$$

holds for the preceding ($j = k - 1$) or for the following ($j = k + 1$) image, where $\|\cdot\|_2$ denotes the Euclidean norm and T is a threshold. Experimentally we have found that this procedure yields better results than the approach in [36]. To remove false detections due to noise as well as to fill

Figure 7. Image of a walking person

Figure 8. Result after change detection and applying binary image operations

in falsely nondetected image points we subsequently apply binary image operations. The result for the image in Fig. 7 can be seen in Fig. 8. The applicability of this approach has also been demonstrated in [10]. There, the trajectories of several simultaneously moving objects observed from bird's-eye view (namely, cars, pedestrians, and cyclists) have automatically been derived without using explicit models.

5.1.2. *3D position estimation*

To estimate the 3D position of an observed person we compute the enclosing rectangle of the detected object candidate (see Fig. 8) and make an assumption about the absolute height of the person. The camera is supposed to be upright w.r.t. gravity and a 4×4 calibration matrix $\underline{T}$ is assumed to be available (e.g., [69],[24],[67]). Using homogeneous coordinates the projection of a 3D point $\vec{X} = (X, Y, Z, 1)^T$ onto the image plane is then given by

$$h\,\vec{x} = \underline{T}\,\vec{X}, \tag{7}$$

where $\vec{x} = (x, y, *, 1)$, $*$ denotes an arbitrary value, and h being nonzero (Fig. 9). The midpoints of the bottom and the top edges of the detected rectangle $\vec{x}_u$ and $\vec{x}_o$ represent the sole and the top of the human body, respectively. The connection line between the corresponding 3D positions is supposed to lie perpendicular to the plane upon which the person moves. With H denoting the assumed height of the person and $\vec{H} = (0, H, 0, 0)^T$ we can write

$$h_u\vec{x}_u \;=\; \underline{T}\vec{X}_u \tag{8}$$

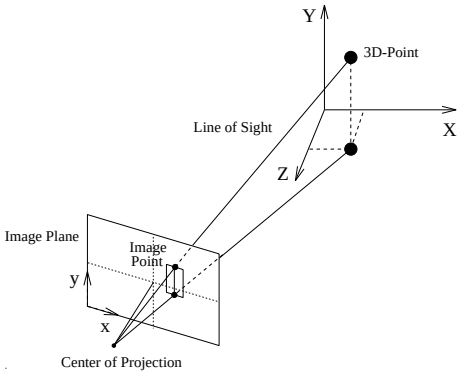

Figure 9. Coordinate systems

Figure 10. Enclosing rectangle and extracted grey-value edge lines

$$h_o \vec{x}_o \quad = \quad \underline{T}(\vec{X}_u + \vec{H}). \tag{9}$$

Based on (8) and (9) we can derive a linear system of equations which is solved to obtain the 3D position of the person (for details, see [72]).

5.1.3. *Contour matching*

After change detection and 3D position estimation we then apply a contour matching algorithm to estimate the posture of the person. Since the contours of our model consist of straight lines we decided to compare them with edge lines. In comparison to edge points the number of image features is smaller and the influence of noise is reduced. Since we compute the edges only within the detected rectangle the search space for matching has been considerably reduced. Edge detection is done using the approach in [43] which is similar to that in [16]. After that, we use an edge linking procedure followed by an Eigenvector line fitting as described in [22] (see Fig. 10).

For matching the model with the detected grey-value edges we compute a search window for each visible model contour. The size of the search window depends on the length of the model contour (Figs. 11, 12). If a grey-value edge overlaps a window we first cut this edge to the portion inside the window. Then, we compute a measure of similarity between the grey-value edge and the model contour. This measure takes into account the length l_i as the projection of the grey-value edge onto the model contour l_{Mi}, second, the distance d_i between the midpoint of the grey-value edge and its corresponding projection onto the model contour, and third, the angle between the two edges $\Delta\varphi_i$. The larger l_i and the smaller $(l_{Mi} - l_i)$, d_i and

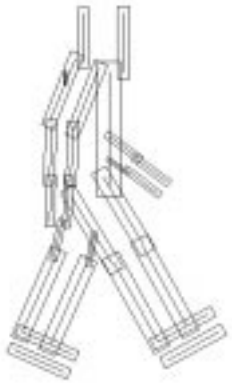

Figure 11. Search windows for the model

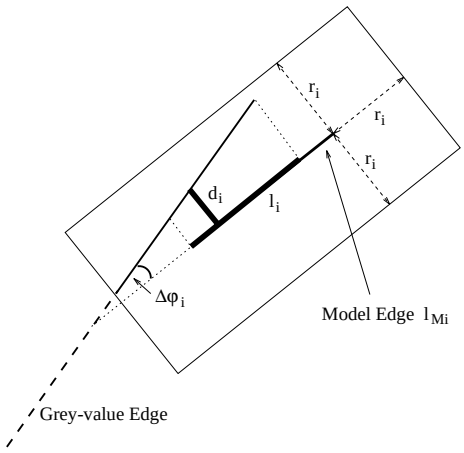

Figure 12. Comparison between grey-value edge and model contour

$\Delta\varphi_i$ the more similar are the two edges. We use the following measure of similarity:

$$s_i = l_i \; e^{-\frac{1}{2}\left(\frac{(l_{Mi}-l_i)^2}{\sigma_{li}^2} + \frac{d_i^2}{\sigma_{di}^2} + \frac{\Delta\varphi_i^2}{\sigma_{\Delta\varphi}^2} \right)} \tag{10}$$

With

$$\underline{\Sigma}_i = \begin{pmatrix} \sigma_{li}^2 & 0 & 0 \\ 0 & \sigma_{di}^2 & 0 \\ 0 & 0 & \sigma_{\Delta\varphi}^2 \end{pmatrix}, \qquad \vec{k}_i = \begin{pmatrix} l_{Mi}-l_i \\ d_i \\ \Delta\varphi_i \end{pmatrix} \tag{11}$$

the measure can also be written as

$$s_i = l_i \; e^{-\frac{1}{2}\vec{k}_i^T \underline{\Sigma}_i^{-1} \vec{k}_i}. \tag{12}$$

The parameters σ_{li} and σ_{di} of the measure are determined in dependence of the length of the model contour: $\sigma_{li} = c_l \; l_{Mi}$ and $\sigma_{di} = c_d \; r_i = c_d \; c_r \; l_{Mi}$, where c_l, c_d and c_r are constant for all model contours; $\sigma_{\Delta\varphi}$ is constant, too. Therefore, the exponent in (12) is independent of scaling the model. Since we want grey-value edges with larger values l_i to have a larger influence on the overall similarity we weight the exponential function by this value. Alternatively, we could weight by l_{Mi}.

Many approaches for line matching compare the midpoints or the starting and endpoints of model contours with the grey-value edge lines supposing that the two edges are similar in length (e.g., [48],[53],[75],[7]). In

our application the grey-value edges of the lower and upper part of the arms and the legs often are connected to one grey-value edge. Then, a comparison between midpoints or starting and endpoints would lead to large discrepancies. Therefore, we first cut the grey-value edges and then use the quantities of similarity as described above. The use of the exponential function in (12) has the advantage that the resulting expression is defined for all values and decreases fast to zero for less similar edges. If several grey-value edges overlap the search window, then we take the one with highest similarity value. The overall measure of similarity $s(\vec{p})$ in dependence of the model parameters $\vec{p}$ is the sum of the values s_i for all visible contours of the human body model normalized by the sum of their lengths l_{Mi}:

$$s(\vec{p}) = \frac{\sum_{i=1}^{n} w_i \; s_i}{\sum_{i=1}^{n} l_{Mi}} \longrightarrow max. \tag{13}$$

The parameters w_i can be used to weight the values s_i by other information, e.g., by the magnitude of the grey-value gradient at the matched grey-value edge or by values derived from prior knowledge about the recognition significance of the body parts. With (13) we search those parameters $\vec{p}$ which maximize $s(\vec{p})$. To obtain a more stable recognition result we remove hidden model contours. However, one disadvantage is that no analytic relation between $s(\vec{p})$ and $\vec{p}$ can be determined. Therefore, for maximizing $s(\vec{p})$ we use a grid search method with equally spaced points and take those parameters which yield the highest similarity value.

The procedure decribed above has been tested on single images. We have fixed the 3D position of the model and have varied the posture parameter *pose* within the whole walking cycle (see [72]). It turned out, that the estimated postures generally agree with the observation. However, the similarity curve has several secondary maxima and therefore it is no good choice to use a downhill optimization procedure. Also, we obtained relatively large similarity values for values of *pose* displaced about half of the walking cycle from the correct posture. However, this is what we expect if we imagine that walking is a symmetric movement w.r.t. to time. Even for a human observer it is generally hard to decide which of the two legs is in front of the other when seeing a walking person in some distance. In either case, to make the estimation result more robust it seems to be important to take into account information derived from several consecutive images.

5.1.4. *Starting values for the Kalman filter*
In the initialization phase of the Kalman filter we apply the procedures described above for each image independently. We use a number of 10-15

images which represent about half of a walking cycle. In comparison to that, a human observer only needs $0.2s$ (less than a quarter of a walking cycle) to recognize a walking human represented by moving light displays ([38],[63]). Starting values for the model parameters are obtained by applying linear regression to the estimated values from the initialization phase. Since the estimates of *pose* for real images often are displaced by about half of the walking cycle, some preprocessing is necessary for this parameter. The values of *pose* in dependence of the image number approximately lie on two lines. Therefore, to exclude systematic errors, we first group those values together which approximately lie on the two lines, and then apply linear regression. The selection of one of the two lines is then done by taking into account the overall number of estimates as well as the mean value of similarity. In our experiments, it was possible to automatically estimate the correct initial posture with this procedure. However, it should be noted, that generally this is a very hard problem which should be investigated further.

5.2. INCREMENTAL ESTIMATION

After initialization we apply a Kalman filter scheme ([39],[26]) to incrementally estimate the model parameters in consecutive images. In the field of computer vision, Kalman filter approaches have been introduced, for example, in [12] and [20]. In [12] an approach is described for incrementally estimating the model parameters of rigid objects from measured image points, whereas in [20] grey-value edge lines are tracked in the image plane. Our aim is to analyze walking persons. Since for general camera positions the body parts during walking often are occluded by one another (self-occlusion) tracking of single grey-value edge lines in general would lead to severe problems. Therefore, to cope with the problem of self-occlusion, in our case we fit the model as a whole and use a hidden-line algorithm to remove occluded model contours.

In the following, we assume the observed person to walk with constant velocity. By interpreting the parameter values for the best fit of the model contours with the grey-value edges as measurements for the corresponding time instant, the system description as well as the measurement model of the Kalman filter approach can be expressed by a linear relation. Therefore, we use a discrete linear Kalman filter.

The general discrete linear model is given by:

$$\vec{p}_k = \underline{\Phi}_{k,k-1}\,\vec{p}_{k-1} + \underline{\ }_{k-1}\vec{w}_{k-1} \tag{14}$$

$$\vec{z}_k = \underline{H}_k\,\vec{p}_k + \vec{v}_k, \tag{15}$$

where the searched parameters are repesented by the state vector $\vec{p}_k$ at the time instant k, and $\underline{\Phi}_{k,k-1}$ is the transition matrix. $\underline{.}_k\vec{w}_k$ represents modelling errors, where $\underline{.}_k$ often is chosen to be the unity matrix and $\vec{w}_k$ is assumed to be Gaussian distributed with expected value $E\{\vec{w}_k\} = \vec{0}$ and covariance matrix $E\{\vec{w}_k\vec{w}_k^T\} = \underline{Q}_k$, i.e., $\vec{w}_k \sim N(\vec{0}, \underline{Q}_k)$. The current measurements are represented by $\vec{z}_k$. $\underline{H}_k$ denotes the measurement matrix and $\vec{v}_k \sim N(\vec{0}, \underline{R}_k)$ the measurement errors. Often it is reasonable to assume that the errors $\vec{w}_k$ and $\vec{v}_j$ are uncorrelated, i.e., $E\{\vec{w}_k\vec{v}_j^T\} = \underline{0}$ for all j, k. The prediction of the parameters and the covariance matrix is given by:

$$\vec{p}_k^* = \underline{\Phi}_{k,k-1}\hat{\vec{p}}_{k-1} \tag{16}$$

$$\underline{P}_k^* = \underline{\Phi}_{k,k-1}\underline{\hat{P}}_{k-1}\underline{\Phi}_{k,k-1}^T + \underline{Q}_{k-1} \tag{17}$$

With these predictions and the current measurement $\vec{z}_k$ the estimates $\hat{\vec{p}}_k$ and $\underline{\hat{P}}_k$ in the current image can be computed by

$$\hat{\vec{p}}_k = \vec{p}_k^* + \underline{\hat{K}}_k(\vec{z}_k - \underline{H}_k\,\vec{p}_k^*) \tag{18}$$

$$\underline{\hat{P}}_k = (\underline{I} - \underline{\hat{K}}_k\underline{H}_k)\underline{P}_k^* \tag{19}$$

$$\underline{\hat{K}}_k = \underline{P}_k^*\underline{H}_k^T(\underline{H}_k\underline{P}_k^*\underline{H}_k^T + \underline{R}_k)^{-1}, \tag{20}$$

where $\underline{I}$ is the unity matrix.

In our application, we want to estimate the 3D position $\vec{X} = (X, Y, Z)$ and the posture *pose* of walking persons. The state vector is $\vec{p}_k = (X_k,\ \dot{X}_k,\ Y_k, \dot{Y}_k,\ Z_k, \dot{Z}_k,\ pose_k, \dot{pose}_k)^T$, where $\dot{X}_k, \dot{Y}_k, \dot{Z}_k$, and $\dot{pose}_k$ are the velocities. The time difference between two succesive images is denoted by Δt. Supposing constant velocities we have

$$\underline{\Phi}_{k,k-1} = \begin{pmatrix} 1 & \Delta t & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \Delta t & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \Delta t & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{21}$$

Errors for the velocities are taken into account by the covariance matrix $\underline{Q}_k$. At the beginning, cross-correlations between the single parameters are

set to zero.

$$\underline{Q}_k = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_{Q\dot{X}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{Q\dot{Y}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{Q\dot{Z}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{Qp\dot{o}se} \end{pmatrix} \tag{22}$$

In our case, we have measurements for $\vec{X} = (X, Y, Z)$ and *pose*. The measurement matrix is given by

$$\underline{H}_k = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \tag{23}$$

and uncertainties of the measurements are characterized by

$$\underline{R}_k = \begin{pmatrix} \sigma^2_{RX} & 0 & 0 & 0 \\ 0 & \sigma^2_{RY} & 0 & 0 \\ 0 & 0 & \sigma^2_{RZ} & 0 \\ 0 & 0 & 0 & \sigma^2_{Rpose} \end{pmatrix} \tag{24}$$

6. Experimental Results

Our approach has been tested on synthetic as well as on real-world image sequences. To reduce the complexity of the recognition task we assume the observed pedestrian to move parallel to the image plane. Therefore, the scene coordinate Z describing the depth of the walking person is supposed to remain constant and we take over the value from the initialization phase. Incrementally we estimate the height Y of the person above the plane of movement taking into account the vertical displacements of the whole body as represented by the motion curve in Fig. 5. Note, that since we normalize the values for Y to the height of a standing person we can assume that the velocity for this parameter is zero and also the velocity's uncertainty can be set to zero. In addition to Y, we estimate the coordinate X in the direction of the movement as well as the posture *pose*. The initial uncertainties for the velocities of these parameters are derived from prior known average velocities of pedestrians. We assume $1s$ time duration for a walking cycle and

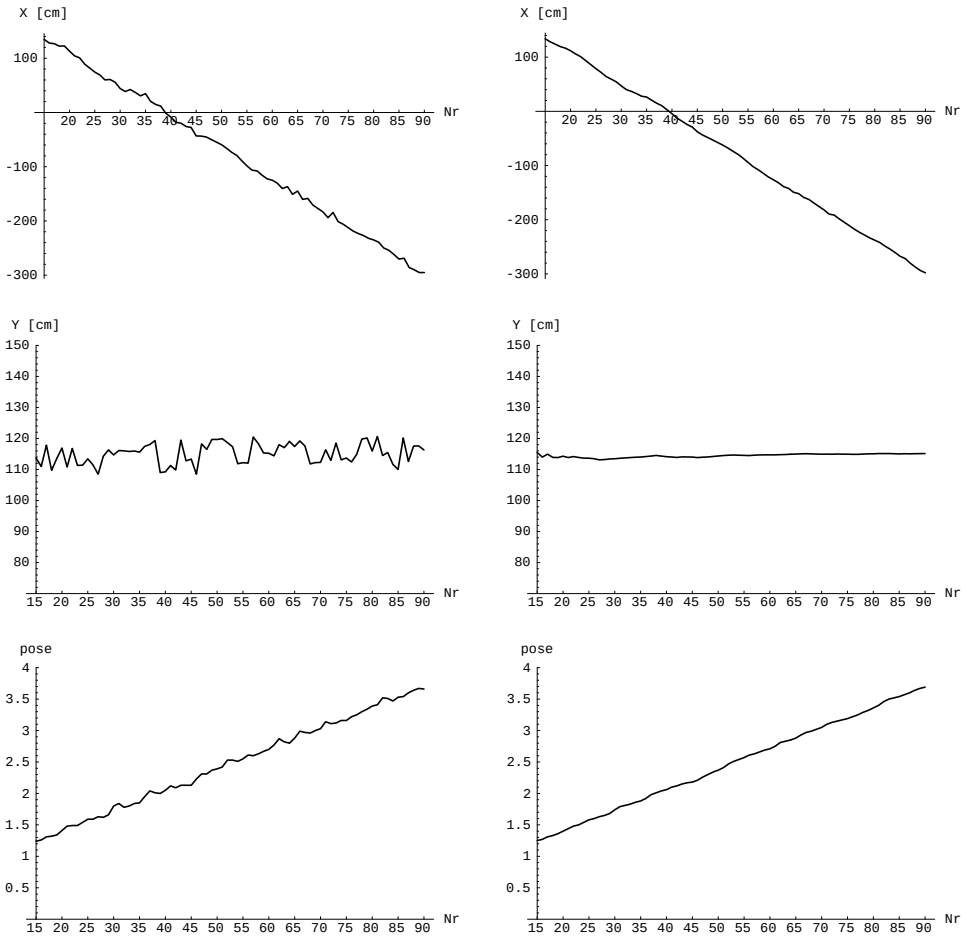

Figure 13. Estimation result of X, Y and *pose*; left: measurements; right: Kalman filter estimates

the covered distance to be $1.6m$ (e.g., [55],[54],[37]). The other elements for the covariance matrices have been chosen heuristically in such a way that we get reasonable values for automatically controlling the search spaces for the parameters. Note, that for incrementally estimating the model parameters we only exploit the matching results of the model contours with the grey-value edges. In this phase, we do not use the change detection and 3D position estimation procedures as described in Section 5.1. Note also, that based on the predictions of the Kalman filter the search space for matching the model contours with grey-value edges is considerably be reduced in comparison to the initialization phase.

Application of our approach to a real-world image sequence consisting of

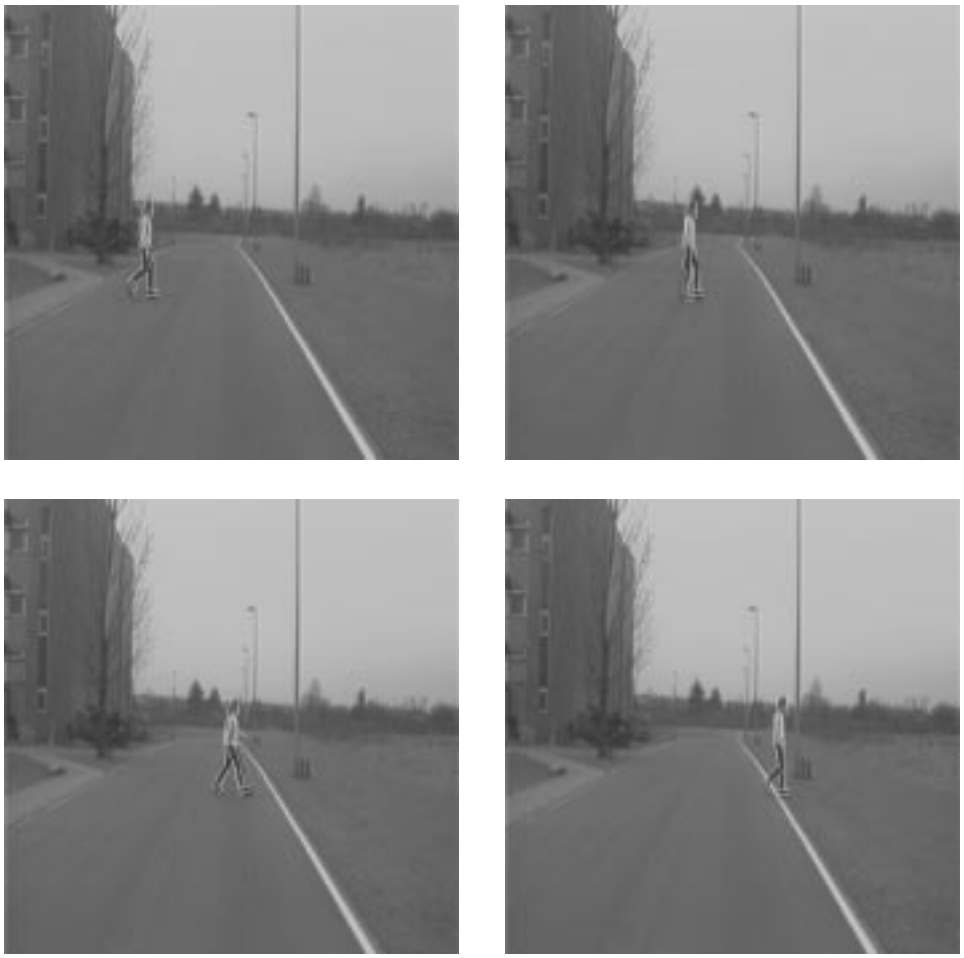

Figure 14. Estimated movement states superimposed onto the original images (images number 20, 40, 60 and 80)

80 images yields the estimation results shown in Fig. 13. For the similarity measure in (13) we have used equal weights $w_i = 1, i = 1, ..., n$. On the left side are the measurements for X, Y and *pose* due to contour matching, where X and Y specify the center of the person's torso. On the right side are the Kalman filter estimates. Especially for Y the smoothing properties of the Kalman filter can clearly be seen. Although the image portion covered by the pedestrian is relatively small, the person has been tracked over the whole sequence. Estimated movement states superimposed onto some of the original images are shown in Fig. 14. Generally, the agreement is fairly well. However, in some images we can observe slight deviations.

To analyze these deviations more precisely we have investigated the re-

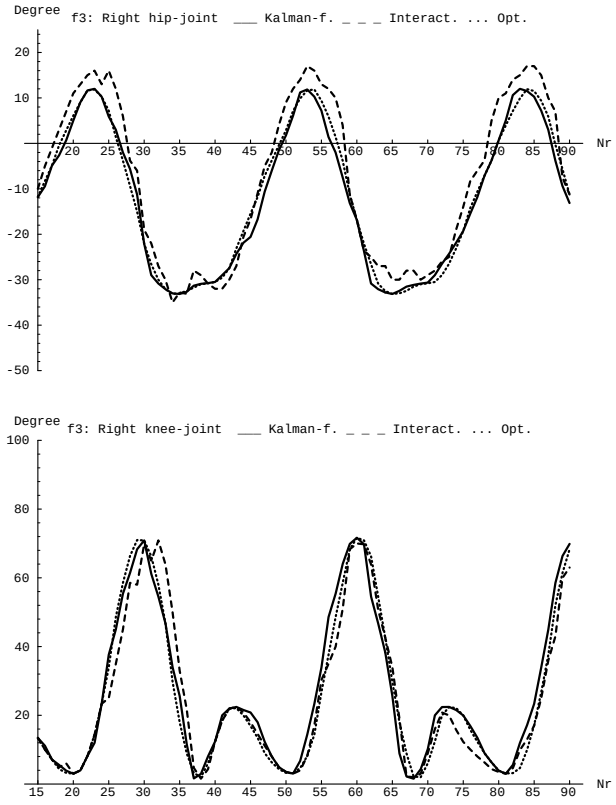

Figure 15. Motion curves for hip and knee joints on the right side of the human body

sult of our approach for another real-world image sequence where the image portion covered by the pedestrian is larger and therefore the deviations are easier to recognize. We have used the same parameter setting as before (see also [72]). For this image sequence the estimated motion curves of the hip and knee joint for the right side of the human body can be seen in Fig. 15 (solid lines). Since we have no ground truth for these images we have interactively determined reference values which we regard as ground truth. This has been done by manually adjusting the model for each image in such a way that for a human observer the agreement is best (see the dashed lines in Fig. 15). Note, that such an adjustment is not very easy since even small changes of the angle at the hip-joint (e.g., $3^o - 5^o$) lead to relatively large posture changes of the whole model. Nevertheless, from Fig. 15 we see that the estimated curves generally agree with the interactively obtained curves. However, at the hip-joint especially for positive angles there are larger deviations which result in the overall difference between the model and the image data. For a further comparison we have also drawn those curves

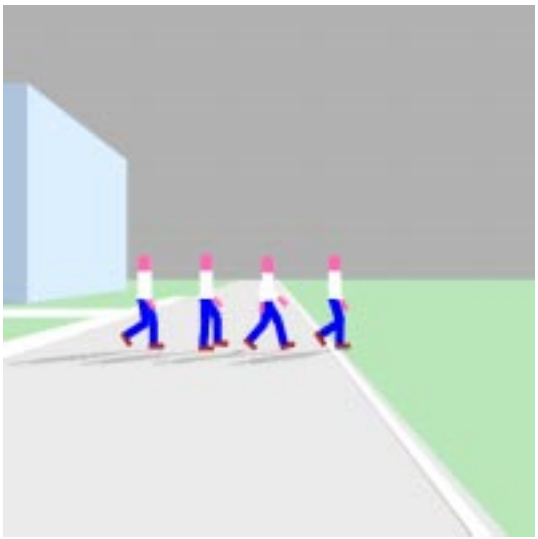

Figure 16. Visualization of the geometrical scene description (GSD) for images number 20, 40, 60 and 80 of the image sequence shown in Fig. 14

which represent the optimal movement according to our motion model for the corresponding velocity. These curves also generally agree with the other motion curves.

Recently, we have also investigated our approach in conjunction with the natural language access system VITRA [32] to derive natural language descriptions such as "The pedestrian walks across the street" (see [31]). As intermediate representation between our computer vision system and the natural language access system we have used the 'geometrical scene description' (GSD) as proposed in [58],[57]. Generally, the GSD represents all available information about the visible objects and their locations over time. In our case, the GSD consists of the 3D positions and postures of the pedestrian for each image. Additionally, we have coarsely represented the stationary background. The GSD corresponding to the real image sequence in Fig. 14 has been visualized in Fig. 16.

7. Model of a Cyclist

In this section, we indicate a possible extension of our work dealing with articulated movements in street-traffic scenes. We consider cyclists which, besides pedestrians and cars, belong to the most often occuring class of objects in this kind of domain.

To model a cyclist we combine the representation of the human body together with a geometric description of a bicycle. As with our motion model

Figure 17. Movement states of a cyclist

of walking the movement of the cyclist is modelled by using a kinematic method. However, for the cyclist we do not need any data from motion studies. This is possible because the geometric relations of the person and the bicycle fully constrain the movement if we consider "normal" driving straight ahead. With the movement of walking there are more degrees of freedom. For the cyclist we assume that the person sits on the saddle and that the upper part of the body remains unchanged during the movement. Note, however, that this model is a strong simplification in comparison to the complex movements that can be observed, for example, in a finish of a bicycle race. Anyway, the relative positions of the human legs are generally determined by the positions of the pedals of the bicycle. If we assume that the ankle-joint does not move, then the positions of the legs can be computed by intersecting two circles. Since only one solution is physically possible we thus obtain a unique position for the legs. This is because for the knee-joint only positive angles can occur (compare with the motion curves of walking in Fig. 4). The overall movement of the cyclist is periodic and symmetric w.r.t. the different body sides. Therefore, we can apply the same kinematic scheme as we have used before for modelling human walking. For the motion model of the cyclist we have evaluated ten positions of the pedals within one cycle to obtain function values for the joints. These values have been interpolated by periodic cubic splines as described in Section 4 above. Movement states of the cyclist for about half of the cycle are shown in Fig. 17. In Fig. 18 some images from an animated sequence of the cyclist together with a walking person can be seen.

Analogously to our approach for recognizing walking persons, the model of the cyclist, i.e., the geometric and the motion model, could be used to analyze the movements of cyclists in image sequences. To efficiently recognize the wheels of the bicycle a matching algorithm for circles (or ellipses) seems to be necessary.

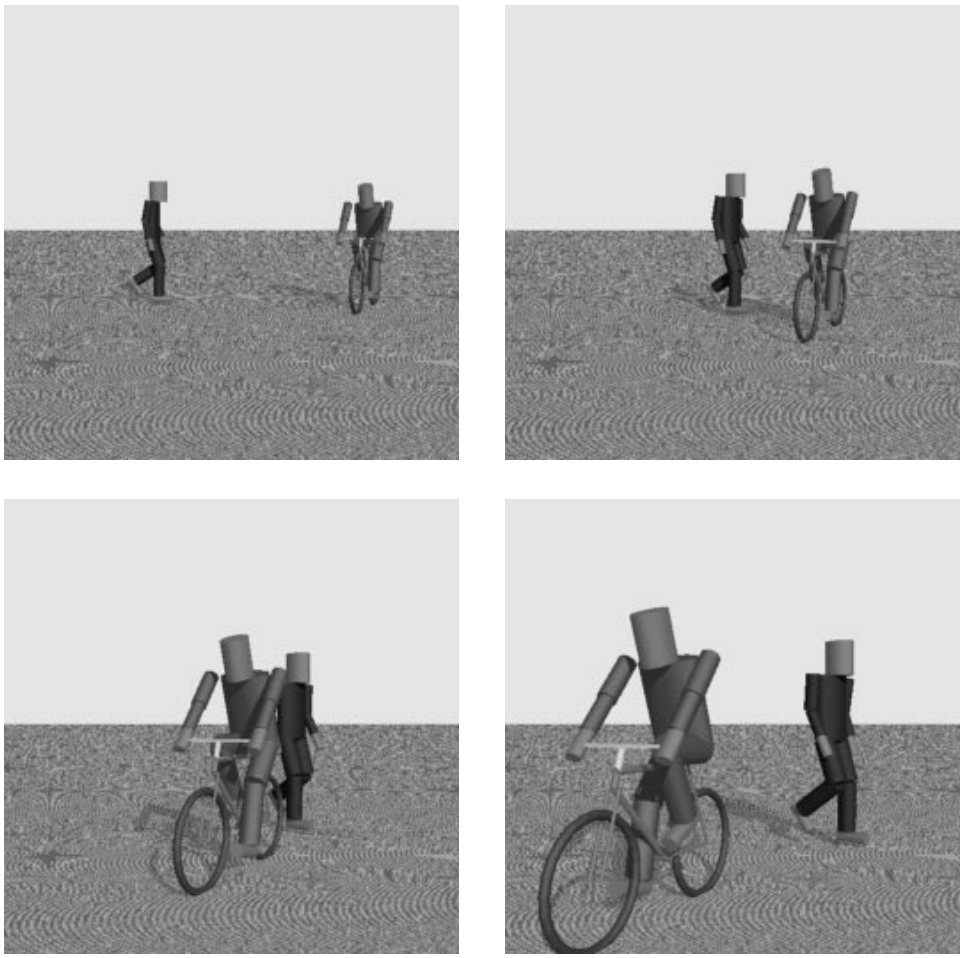

Figure 18. Moving cyclist and pedestrian

8. Summary and Future Work

The movement of articulated bodies such as the human body is enabled by the coordinated movement of its rigid body parts. The body parts are connected by joints and, in general, move differently. For interpreting these movements as the movement of one single body, it seems to be necessary to incorporate knowledge in the image analysis process.

In our approach for analyzing human movements in real-world image sequences we exploit knowledge about the human body as well as its movement. Central to this approach is the use of an explicit motion model which is based on analytically given motion curves for the body parts. These motion curves represent data from medical motion studies. The fact that the

Figure 19. Several walking persons

Figure 20. Crossing pedestrian and car

motion curves are very similar for different persons opens us the possibility to use this data as knowledge source. A nice property is that only one parameter is needed to fully specify the relative positions of all body part. To incrementally estimate the 3D positions and postures of a walking person in consecutive images we have applied a Kalman filter scheme yielding smooth and robust results. We have assumed that the person moves parallel to the image plane with constant velocity. Starting values for the Kalman filter have automatically been determined through an initialization phase. Additionally, we have described a model of a cyclist to give an idea of extending our work for more complex street-traffic scences. Estimation on an incremental basis rather than using the entire sequence is important, for example, in street-traffic scenes for the purpose of early recognition of situations which might lead to accidents, followed by giving a warning to the involved driver.

Future work on the recognition of walking persons should address the problem of reducing the deviations between the observed individual movements and those of our motion model which represents an average over a relatively large number of test persons walking at about the same velocity. An important point is that the motion curves of the body parts in general depend on the speed of movement. One possibility for improvement is to start the algorithm with the predefined motion curves and then try to adapt those curves to the observed movement. In this way, an incremental update of the model curves would be used for further evaluation. Also, to make the estimation result more robust, additional features could be taken into account, e.g., one could compare the model velocity fields with image velocity fields. A more efficient match of the model to the image data could possibly be achieved by using an aspect graph (view graph) for the model (e.g., [42],[68]).

In further investigations also other human movement types, such as run-

ning, should be analyzed. Additionally, one should investigate the recognition of several simultaneously moving bodies as well as the classification between different movement types (see Figs. 18, 19, and 20 for simulations of such scenes). Another future challenge is the automatic natural language description of observed human movements. Recently, investigations towards this long-term goal have been reported in [31].

Acknowledgement

For discussions and critical comments I thank D. Bister, C. Bregler, K. Daniilidis, G. Herzog, W. Leister, B. Neumann, J.H. Rieger, C. Schnörr, H.S. Stiehl, J. Weisbrod, and G. Zimmermann. This work has been supported by the Deutsche Forschungsgemeinschaft (DFG), Sonderforschungsbereich 314 "Künstliche Intelligenz – Wissensbasierte Systeme".

References

1. Akita, K. Image sequence analysis of real world human motion, *Pattern Recognition* 17 (1984) 73-83
2. Aristoteles, *Über die Bewegung der Lebewesen; Über die Fortbewegung der Lebewesen,* Datierung um 330 v.Chr., Teil II und III, J. Kollesch, Aristoteles Werke in deutscher Übersetzung, Band 17, Zoologische Schriften II, E. Grumach, H. Flashar (Hrsg.), Wissenschaftliche Buchgesellschaft, Darmstadt 1985
3. Attwood, C.I., Sullivan, G.D. and Baker, K.D. Model-based Recognition of Human Posture Using Single Synthetic Images, *Proc. Fifth Alvey Vision Conf.,* Univ. of Reading, Reading/UK, 25-28 Sept. 1989, 25-30
4. Badler, N.I. Temporal scene analysis: conceptual descriptions of object movements, Tech. Rep. No. 80, Dept. Computer Science, Univ. Toronto, Feb. 1975
5. Badler, N.I. and Manoochehri, K.H. and Walters, G. Articulated Figure Positioning by Multiple Constraints, *IEEE Computer Graphics & Appl.* 7:6 (June 1987) 28-38
6. Baum, L.F. and Denslow, W.W. *The Wonderful Wizard of Oz,* Geo. M. Hill Co. Chicago New York 1900, Justin Knowles Publishing Group London 1987
7. Beveridge, J.R., Weiss, R. and Riseman, E.M. Optimization of 2-Dimensional Model Matching, *Proc. Image Understanding Workshop,* Palo Alto, California, May 23-26, 1989, 815-830
8. Binford, T.O. Visual perception by computer, *IEEE Conf. on Systems and Control,* Dec. 1971
9. Bister, D. Bestimmung der Trajektorien von zeitweise verdeckten Objekten aus einer Bildfolge, Diplomarbeit, Institut für Algorithmen und Kognitive Systeme, Fakultät für Informatik der Universität Karlsruhe (TH), April 1991
10. Bister, D., Rohr, K. and Schnörr, C. Automatische Bestimmung der Trajektorien von sich bewegenden Objekten aus einer Bildfolge, *12. DAGM - Symposium Mustererkennung,* 24.-26. Sept. 1990, Oberkochen-Aalen, *Informatik-Fachberichte* 254, R.E. Großkopf (Hrsg.), Springer-Verlag Berlin Heidelberg 1990, 44-51
11. Braune, W. and Fischer, O. *Der Gang des Menschen, 1. Theil: Versuche am unbelasteten und belasteten Menschen,* Abhandlungen der Mathematisch-Physischen Classe der königlich sächsischen Gesellschaft der Wissenschaften, Einundzwanzigster Band, S. Hirzel Leipzig 1895
12. Broida, T.J. and Chellappa, R. Estimation of Object Motion Parameters from Noisy Images, *IEEE Trans. on Pattern Anal. and Machine Intell.* 8 (1986) 90-99

13. Bruderlin, A. and Calvert, T.W. Goal-Directed, Dynamic Animation of Human Walking, *Computer Graphics* 23:3 (July 1989) 233-242
14. Calvert, T.W. and Chapman, J. Aspects of the Kinematic Simulation of Human Movement, *IEEE Computer Graphics & Appl.* 2:9 (Nov. 1982) 41-49
15. Calvert, T.W. and Chapman, A.E. Analysis and synthesis of human movement, in T.Y. Young (Ed.), Handbook of Pattern Recognition and Image Processing (Vol. 2): Computer Vision, Academic Press, SanDiego, CA, 1994, 431-474
16. Canny, F. "A computational approach to edge detection", *IEEE Trans. on Pattern Anal. and Machine Intell.* 8 (1986) 679-698
17. Cédras, C. and Shah, M. Motion-based recognition: a survey, *Image and Vision Computing* 13:2 (1995) 129-155
18. Chen, Z. and Lee, H.-J. Knowledge-Guided Visual Perception of 3-D Human Gait from a Single Image Sequence, *IEEE Trans. on Systems, Man and Cyb.* 22:2 (1992) 336-342
19. Cipolla, R. and Yamamoto, M. Stereoscopic Tracking of Bodies in Motion, *Proc. Fifth Alvey Vision Conf.,* Univ. of Reading, Reading/UK, 25-28 Sept. 1989, 109-114
20. Deriche, R. and Faugeras, O. Tracking line segments, *Image and Vision Computing* 8:4 (Nov. 1990) 261-270
21. DIN 33402, Deutsche Normen, *Körpermaße des Menschen,* Beuth Verlag Berlin 1987
22. Duda, R.O. and Hart, P.E. *Pattern classification and scene analysis,* Wiley New York 1973
23. Fischer, O. *Theoretische Grundlagen für eine Mechanik der lebenden Körper mit speziellen Anwendungen auf den Menschen, sowie auf einige Bewegungsvorgänge an Menschen,* B.G. Teubner Leipzig Berlin 1906
24. Ganapathy, S. Decomposition of transformation matrices for robot vision, *Pattern Recognition Letters* 2:6 (1984) 401-412
25. Gavrila, D.M. and Davis, L.S. 3-D model-based tracking of human upper body movement: a multi-view approach, *Proc. Intern. Symposium Computer Vision (ISCV'95),* Los Alamitos, 21-23 Nov. 1995, IEEE Computer Society Press 1995, 253-258
26. Gelb, A. *Applied Optimal Estimation,* MIT Press, Cambridge/MA, 1974
27. Goddard, N.H. The Interpretation of Visual Motion: Recognizing Moving Light Displays, *Proc. Workshop on Visual Motion,* Irvine, CA, March 20-22, 1989, 212-220
28. Guo, Y., Xu, G. and Tsuji, S. Tracking human body motion based on a stick figure model, *J. of Visual Communication and Image Representation* 5:1 (1994) 1-9
29. Hartmann, E. *Computerunterstützte Darstellende Geometrie,* B.G. Teubner Stuttgart 1988
30. Hartrum, T.C. Computer implementation of a parametric model for biped locomotion kinematics, Ph.D. Thesis, School of the Ohio State University, Columbus, Ohio, 1973
31. Herzog, G. and Rohr, K. "Integrating Vision and Language: Towards Automatic Description of Human Movements", *Proc. 19th Conf. on Artificial Intelligence, KI-95: Advances in Artificial Intelligence*, Sept. 1995, Bielefeld/Germany, *Lecture Notes in Artificial Intelligence* 981, I. Wachsmuth, C.-R. Rollinger, and W. Brauer (Eds.), Springer-Verlag Berlin Heidelberg 1995, 259-268
32. Herzog, G. and Wazinski, P. VIsual TRAnslator: Linking Perceptions and Natural Language Descriptions, *Artificial Intelligence Review* 8:2/3 (1994) 175-187
33. Hoffman, D.D. and Flinchbaugh, B.E. Interpretation of Biological Motion, *Biol. Cybern.* 42 (1982) 195-204
34. Hogg, D. Model based vision: a program to see a walking person, *Image and Vision Computing* 1:1 (1983) 5-20
35. Hogg, D. Interpreting Images of a Known Moving Object, PhD dissertation, University of Sussex, Brighton/UK 1984
36. Hsu, Y.Z., Nagel, H.-H. and Rekers, G. New Likelihood Test Methods for Change

Detection in Image Sequences, *Computer Vision, Graphics, and Image Processing* 26 (1984) 73-106
37. Inman, H.V.T., Ralston, H.J. and Todd, F. *Human walking,* Williams & Wilkins Baltimore/London 1980
38. Johansson, G. Spatio-temporal differentiation and integration in visual motion perception, *Psychological Research* 38 (1976) 379-396
39. Kalman, R.E. A New Approach to Linear Filtering and Prediction Problems, *Trans. ASME J. Basic Eng.,* Series 82D (March 1960) 35-45
40. Kambhamettu, C., Goldgof, D.B., Terzopoulos, D. and Huang, T.S. Nonrigid motion analysis, in T.Y. Young (Ed.), Handbook of Pattern Recognition and Image Processing (Vol. 2): Computer Vision, Academic Press, SanDiego, CA, 1994, 405-430.
41. Kinzel, W. and Dickmanns, E.D. Moving Humans Recognition using Spatio-Temporal Models, *Proc. ISPRS'92,* Washington, D.C., *Internat. Archives for Photogrammetry and Remote Sensing,* Vol. XXIX, Part B5, L.W. Fritz and J.R. Lucas (Eds.), 1992, 885-892
42. Koenderink, J.J. and van Doorn, A.J. The Internal Representation of Solid Shape with Respect to Vision, *Biol. Cybernetics* 32 (1979) 211-216
43. Korn, A.F. Toward a Symbolic Representation of Intensity Changes in Images, *IEEE Trans. on Pattern Analysis and Machine Intelligence* 10 (1988) 610-625
44. Kunii, T.L. and Sun, L. Dynamic Analysis-Based Human Animation, *CG International'90*, T.S. Chua, T.L. Kunii (Eds.), Springer-Verlag Tokyo Berlin Heidelberg 1990, 3-15
45. Kurakake, S. and Nevatia, R. Description and Tracking of Moving Articulated Objects, *Proc. 11th Intern. Conf. on Pattern Recogn.,* The Hague, The Netherlands, Aug. 30 - Sept. 3, 1992, Vol. I, 491-495
46. Leister, W. and Rohr, K. Voruntersuchungen von Bildsynthesesmethoden zur Analyse von Bildfolgen, Techn. Report Nr. 25/90, Universität Karlsruhe (TH), Fakultät für Informatik, Sept. 1990
47. Leung, M.K. and Yang, Y.H. A region based approach for human body motion analysis, *Pattern Recognition* 20:3 (1987) 321-339
48. Lowe, D.G. The Viewpoint Consistency Constraint, *Internat. J. of Computer Vision* 1 (1987) 57-72
49. Luo, Y., Perales, F.J. and Villanueva, J.J. An Automatic Rotoscopy System for Human Motion Based on a Biomechanic Graphical Model, *Comput. & Graphics* 16:4 (1992) 355-362
50. Marey, É.-J. *Movement,* William Heine London 1895
51. Marr, D. and Nishihara, H.K. Representation and recognition of the spatial organization of three-dimensional shapes, *Proc. R. Soc. Lond.* B 200 (1978) 269-294
52. Marr, D. and Vaina, L. Representation and recognition of the movements of shapes, *Proc. R. Soc. Lond.* B 214 (1982) 501-524
53. McIntosh, J.H. and Mutch, K.M. Matching straight Lines, *Computer Vision, Graphics, and Image Processing* 43 (1988) 386-408
54. Murray, M.P. Gait as a total pattern of movement, *American J. of Physical Med.* 46:1 (1967) 290-332
55. Murray, M.P., Drought, A.B. and Kory, R.C. Walking Patterns of Normal Men, *The J. of Bone and Joint Surgery* 46-A:2 (1964) 335-360
56. Muybridge, E. *Muybridge's complete Human and Animal Locomotion. All 781 Plates from the 1887 'Animal Locomotion',* Vol. 1, Dover Publications, Inc., New York 1979
57. Neumann, B. Natural Language Description of Time-Varying Scenes, in *Semantic Structures, Advances in Natural Language Processing,* D.L. Waltz (Ed.), Lawrence Erlbaum, Hillsdale/NJ, 1989, 167-206
58. Neumann, B. and Novak, H.-J. Event models for recognition and natural language description of events in real-world sequences, *Proc. Internat. Joint Conf. on Artifi-*

cial Intell. (IJCAI'83), 1983, 724-726
59. Newman, W.M. and Sproull, R.F. *Grundzüge der interaktiven Computergrafik,* McGraw-Hill Book Company Hamburg 1986
60. Niyogi, S.A. and Adelson, E.H. Analyzing and Recognizing Walking Figures in XYT, *Proc. IEEE Conf. on Computer Vision & Pattern Recognition,* Seattle, WA, June 21-23, 1994, 469-474
61. O'Rourke, J. and Badler, N.I. Model-based image analysis of human motion using constraint propagation, *IEEE Trans. on Pattern Anal. and Machine Intell.* 2:6 (1980) 522-536
62. Pentland, A. and Horowitz, B. Recovery of Non-Rigid Motion and Structure, *IEEE Trans. on Pattern Anal. and Machine Intell.* 13:7 (1991) 730-742
63. Perrett, D.I., Harries, M.H., Benson, P.J., Chitty, A.J. and Mistlin, A.J. *Retrieval of Structure from Rigid and Biological Motion: An Analysis of the Visual Responses of Neurones in the Macaque Temporal Cortex,* in *AI and the Eye,* A. Blake and T. Troscianko (Eds.), John Wiley & Sons Chichester/UK New York/NY 1990, 181-200
64. Pun, T. and Blake, E. Relationships between Image Synthesis and Analysis: Towards Unification, *Computer Graphics Forum* 9:2 (1990) 149-163
65. Qian, R.J. and Huang, T.S. Motion Analysis of Human Ambulatory Patterns, *Proc. 11th Intern. Conf. on Pattern Recogn.,* The Hague, The Netherlands, Aug. 30 - Sept. 3, 1992, Vol. I, 220-223
66. Rashid, R.F. Towards a System for the Interpretation of Moving Light Displays, *IEEE Trans. on Pattern Anal. and Machine Intell.* 2:6 (Nov. 1980) 574-581
67. Rehfeld, N. Auswertung von Stereobildfolgen mit Kantenmerkmalen, Dissertation, Fakultät für Informatik der Universität Karlsruhe (TH), Juni 1990
68. Rieger, J.H. On the complexity and computation of view graphs of piecewise smooth algebraic surfaces, *Phil. Trans. R. Soc. Lond. A* 354 (1996) 1899-1940
69. Rogers, D.F. and Adams, J.A. *Mathematical Elements for Computer Graphics,* McGraw-Hill Book Company New York 1976
70. Rohr, K. Auf dem Wege zu modellgestütztem Erkennen von bewegten nicht-starren Körpern in Realweltbildfolgen, 11. *DAGM - Symposium Mustererkennung,* 2.-4. Okt. 1989, Hamburg, *Informatik-Fachberichte* 219, H. Burkhardt, K.H. Höhne, B. Neumann (Hrsg.), Springer-Verlag Berlin Heidelberg 1989, 324-328
71. Rohr, K. Incremental Recognition of Pedestrians from Image Sequences, *Proc. IEEE Conf. on Computer Vision & Pattern Recognition,* New York/NY, USA, June 15-17, 1993, 8-13
72. Rohr, K. Towards Model-Based Recognition of Human Movements in Image Sequences, *Computer Vision, Graphics, and Image Processing: Image Understanding* 59:1 (1994) 94-115
73. Schwarz, H.R. *Numerische Mathematik,* B.G. Teubner Stuttgart 1986
74. Seitz, S.M. and Dyer, C.R. Affine Invariant Detection of Periodic Motion, *Proc. IEEE Conf. on Computer Vision & Pattern Recognition,* Seattle, WA, June 21-23, 1994, 970-975
75. Sester, M. and Förstner, W. Object Location Based on Uncertain Models, *11. DAGM-Symposium Mustererkennung,* Hamburg, 2.-4. Okt. 1989, *Informatik-Fachberichte* 219, H. Burkhardt, K.H. Höhne, and B. Neumann (Hrsg.), Springer-Verlag Berlin Heidelberg 1989, 457-464
76. Shiohara, M., Gotoh, T., Nakagawa, Y.M. and Yoshida, Surface Correspondence Based on Three-dimensional Structure Inference In Animation Images, *10th Intern. Conf. on Pattern Recognition,* 16-21 June 1990, Atlantic City, New Jersey, USA, 194-197
77. Tost, D. and Pueyo, X. Human body animation: a survey, *The Visual Computer* 3 (1988) 254-264
78. Tsuji, S., Osada, M. and Yachida, M. Tracking and Segmentation of Moving Objects in Dynamic Line Images, *IEEE Trans. on Pattern Anal. and Machine Intell.* 2:6

(1980) 516-522
79. Tsukiyama, V.T. and Shirai, Y. Detection of the movements of persons from a sparse sequence of TV images, *Pattern Recognition* 18 (1985) 207-213
80. Voss, K. *Theoretische Grundlagen der digitalen Bildverarbeitung,* Akademie Verlag Berlin 1988
81. Webb, J.A. and Aggarwal, J.K. Structure from Motion of Rigid and Jointed Objects, *Artificial Intelligence* 19 (1982) 107-130
82. Weber, W. and Weber, E. *Mechanik der menschlichen Gehwerkzeuge,* Dietrichsche Buchhandlung, Göttingen 1836
83. Weil, J. The Synthesis of Cloth Objects, *Computer Graphics* 20:4 (1986) 49-54
84. Wilhelms, J. Using Dynamic Analysis for Realistic Animation of Articulated Bodies, *IEEE Computer Graphics & Appl.* 7:6 (June 1987) 12-27
85. Yamamoto, M. and Koshikawa, K. Human Motion Analysis Based on a Robot Arm Model, *Proc. Computer Vision and Pattern Recogn.,* Lahaina, Maui, Hawaii, June 3-6, 1991, 664-665
86. Zeltzer, D. Motor Control Techniques for Figure Animation, *IEEE Computer Graphics & Appl.* 2:9 (Nov. 1982) 53-59