CVGIP: IMAGE UNDERSTANDING
Vol. 59, No. 1, January, pp. 94–115, 1994

Towards Model-Based Recognition of Human Movements in Image Sequences

K. ROHR*

Arbeitsbereich Kognitive Systeme, FB Informatik, Universität Hamburg, Bodenstedtstr. 16, 22765 Hamburg, Germany

Received August 10, 1992; accepted April 30, 1993

The interpretation of the movements of articulated bodies in image sequences is one of the most challenging problems in computer vision. In this contribution, we introduce a model-based approach for the recognition of pedestrians. We represent the human body by a 3D-model consisting of cylinders, whereas for modelling the movement of walking we use data from medical motion studies. The estimation of model parameters in consecutive images is done by applying a Kalman filter. Experimental results are shown for synthetic as well as for real image data. © 1994 Academic Press, Inc.

1. INTRODUCTION

Vision enables us to perceive the form and spatial arrangements of physical objects as well as to recognize if objects move and how they move. In order to do this, two things seem to be necessary [42]. On one hand there must be a symbolic system to represent shape and motion information and on the other hand the brain must contain a set of processes to derive such informations from images.

When investigating the dynamic aspects of traffic scenes one has to deal with a variety of different objects and motion types. In the sense above, these objects and their motions have to be represented in order to automatically analyze image sequences. Since the observed events are, in general, very complex an efficient symbolic description is important. It is advantageous to describe recognized motions by natural language [4, 49, 50, 64, 47, 48, 22]. We consider the work in this paper as a small step toward this long-term goal.

In the following we introduce a model-based approach for the recognition of pedestrians. We explicitly represent the human body and its motion and use this knowledge to derive 3D-positions and postures from images. The current contribution is restricted to human walking which is the most frequent type of locomotion of pedestrians.

Our aim is to analyze realistic monocular images. Therefore, existing approaches which assume the joints of the human body to be marked (e.g., [57, 73, 23, 34, 19, 3]) or which investigate synthetic images (e.g., [52, 70, 68]) cannot, in general, be applied. Among those approaches which evaluate real-world scenes, Cipolla and Yamamoto [13] use stereo-images. Akita [1], Leung and Yang [36, 37], and Pentland and Horowitz [54] essentially analyze special gymnastic movements, but not locomotion, which in general simplifies the interpretation because the effect of self-occlusion is diminished. Other approaches restrict their analysis of locomotions to the part of the legs [71] or to the part of the arms [77]. The pioneering work for recognizing walking persons in real-world images has been done by Hogg [24, 25]. He uses the model of a human body of Marr and Nishihara [41] in order to estimate postures of walking persons from images. The data for the motion model has been acquired interactively from one prototype image sequence. Our approach also utilizes the cylindrical model of Marr and Nishihara [41]. However, for the motion model we use data from medical motion studies representing an average over a relatively large number of test persons. Hogg [24, 25] compares the contours of the model with grey-value edge points where hidden model contours have not been removed. We instead compare model contours with grey-value edge lines and remove hidden model contours. In contrast to Hogg [24, 25] our approach determines the initial posture automatically and, for estimating the model parameters for a certain image, earlier estimates are taken into account by using a Kalman filter. Therefore, our estimates of the model parameters over time are smoother.

The organization of this paper is as follows. After an overview of our approach in the next section we describe our model of the human body and its movement. Then, we introduce the procedure for estimating the model parameters from single images and extend this approach to image sequences.

2. OVERVIEW OF OUR APPROACH

One problem in recognizing pedestrians arises from the nonrigidity of the human body. When analyzing rigid ob-

* e-mail: rohr@kogs26.informatik.uni-hamburg.de

1049-9660/94 $6.00

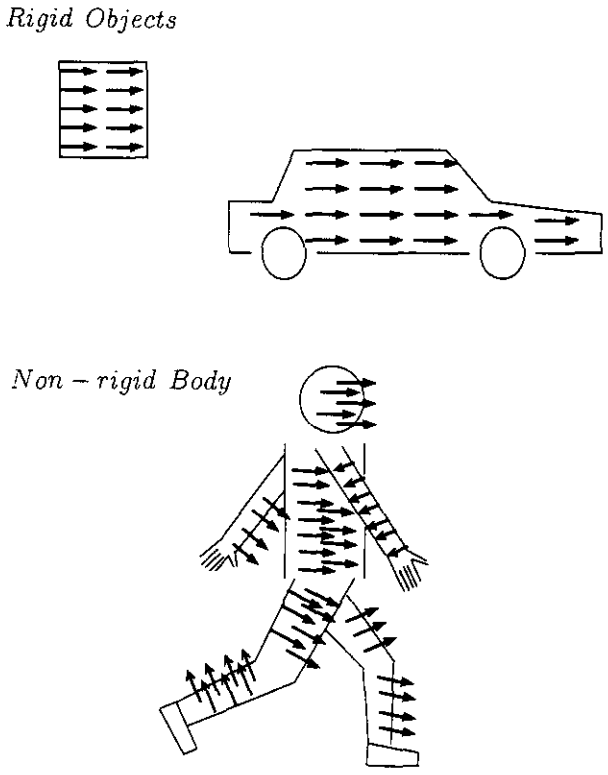

FIG. 1. Comparison between rigid and nonrigid objects.

jects it is possible, for example, to group similar displacement vectors in order to segment image regions belonging to those objects. For nonrigid objects, however, such an approach in general is not feasible since the movements of the body parts are different (Fig. 1). In order to interpret these complex movements it is necessary to represent knowledge about the shape as well as the movement. Therefore, we introduce a model-based approach. Our system will be called MEMO which is an acronym for MEn in MOtion. In addition, this name reminds one of the existence of a MEMOry in which knowledge is stored.

In accordance to Kanade [30] a general scheme for model-based approaches is depicted in Fig. 2. Kanade distinguishes between the picture domain and the scene domain. We have incorporated the $2\frac{1}{2}$D-description of Marr [40] which is a viewer-centered representation of depth, orientation, and discontinuities of visible surfaces. The 2D- and 3D-descriptions represent structures in the image plane and object-centered three-dimensional structures, respectively. As a special case, the interpretation cycle of our approach is shown in Fig. 3. The generic model is a parametric model which represents knowledge about the human body and the movement of walking. By fixing the model parameters we obtain an instantiated 3D-model. The 2D-model description is obtained by a projection onto the image plane and in our case consists

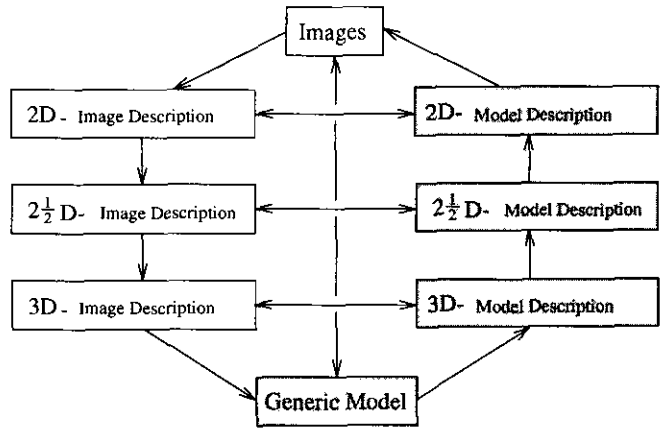

FIG. 2. General interpretation cycle.

of the visible contours of the 3D-model. The following main steps of our approach will be described in detail in the subsequent sections.

For segmenting the image plane in regions corresponding to moving and nonmoving objects we apply a change detection algorithm followed by binary image operations. Using the enclosing rectangle of the detected image region and an assumption about the height of the observed person we estimate the 3D-position. In order to determine more precisely the 3D-position and to estimate the posture we then compare model contours with grey-value edges approximated by straight lines and estimate the model parameters by maximizing a measure of similarity. At the beginning of the image sequence we perform an initialization by applying the procedure above for circa 10–15 images (about the half of a walking cycle) for each image independently. By linear regression over these estimates we obtain starting values for the Kalman filter which is applied afterward. In this subsequent phase, having estimated the parameters for a certain image, we predict the

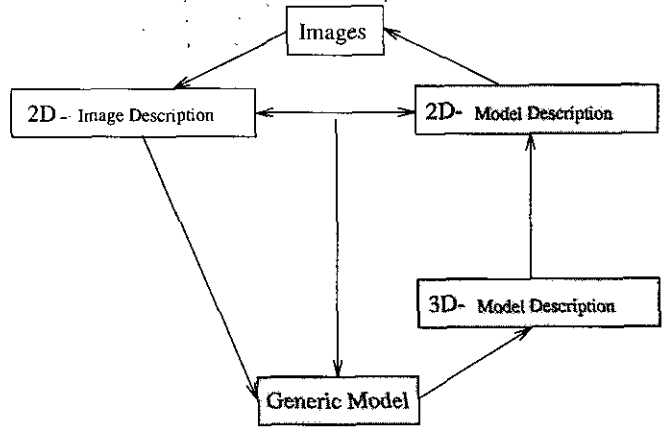

FIG. 3. Interpretation cycle for our approach.

new position and posture for the following image according to our motion model and again adapt the 2D-descriptions of the model and the image. By repeated prediction and estimation we determine the model parameters for the whole image sequence.

3. MODEL OF PEDESTRIANS

Models of the human body and its movements are investigated in the fields of biomechanics, computer graphics, and robotics.

3.1. *Modelling the Human Body*

In order to efficiently recognize shapes from images, Marr and Nishihara [41] laid down three criteria which a static 3D-model should satisfy (see also Vaina [72]). Based on these criteria they proposed to use an object-centered coordinate system, volumetric primitives, and a modular hierarchical organization on the description. As an example, Marr and Nishihara [41] suggest a 3D-model of the human body consisting of cylinders. In our approach we use this model. The human body is represented by 14 cylinders with elliptic cross sections (head, torso, and three primitives for each arm and leg) which are connected by joints (see also Fischer [17]). 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 single 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. Transformations between different coordinate systems are described by homogeneous coordinates $\boldsymbol{X} = (X, Y, Z, 1)^{\mathrm{T}}$,

$$\boldsymbol{X}' = \boldsymbol{A}\,\boldsymbol{X},$$

where $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{R} & \boldsymbol{T} \\ \boldsymbol{0}^{\mathrm{T}} & 1 \end{pmatrix}$, consisting of the 3×3 rotation matrix $\boldsymbol{R}$ and the translation vector $\boldsymbol{T}$. The inverse of $\boldsymbol{A}$ is

$$\boldsymbol{A}^{-1} = \begin{pmatrix} \boldsymbol{R}^{\mathrm{T}} & -\boldsymbol{R}^{\mathrm{T}}\boldsymbol{T} \\ \boldsymbol{0}^{\mathrm{T}} & 1 \end{pmatrix}.$$

If several transformations are applied in series then one has to multiply the single matrices $\boldsymbol{A}_i$:

$$\boldsymbol{A}_i\boldsymbol{A}_{i+1} = \begin{pmatrix} \boldsymbol{R}_i\boldsymbol{R}_{i+1} & \boldsymbol{R}_i\boldsymbol{T}_{i+1} + \boldsymbol{T}_i \\ \boldsymbol{0}^{\mathrm{T}} & 1 \end{pmatrix}.$$

For modelling the human body we use absolute sizes of the body parts. Since pedestrians 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., DIN [15]) is limited. Therefore, we use sizes of the human body parts obtained by direct measurements of a normal person with average clothing. Visualizations of our model are displayed in Figs. 4 and 5 (see also [35]).

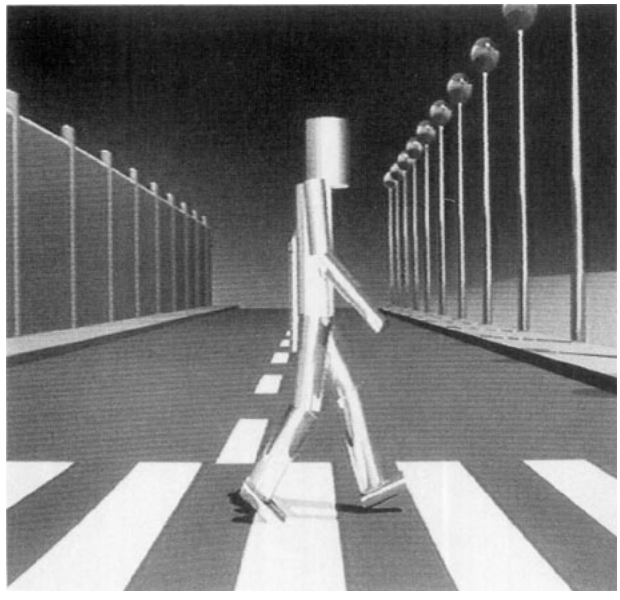

FIG. 4. Model of the human body.

3.2. *Modelling the Movement of Walking*

There are two basic methods for describing bodies in motion: kinematic and dynamic approaches (e.g., [5, 56, 69]). A *kinematic* description explicitly specifies the geometry of objects, i.e., position, orientation, and deforma-

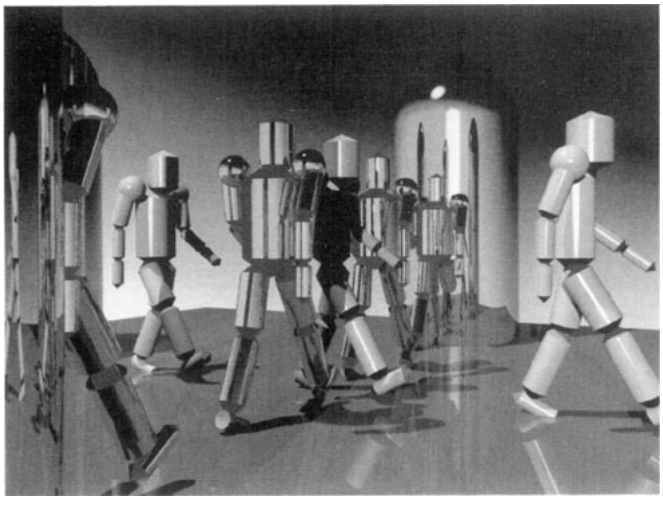

FIG. 5. Several persons.

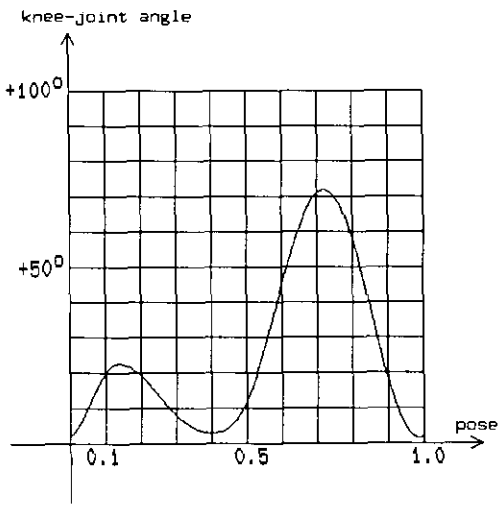

FIG. 6. Motion curve of the knee-joint for the whole walking cycle.

tion without taking into account the cause of the movement (e.g., [21, 12, 78, 79, 5]). 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 the laws of motion in Weber and Weber [74]). Another possibility for simulating human movements is to provide interactively movement positions at certain time instants (keyframe technique) which is time consuming and often does not lead to the desired result. Also, movement positions can be reconstructed interactively from recorded image sequences (rotoscopy). By this, it is possible to model only such movements which have previously been performed. If, however, data from motion studies is already available then these values can be used for simulation.

In contrast to kinematic methods, *dynamic* methods take into account forces and torques and have the potential to produce realistic motions (e.g., [75, 76, 11]). However, these approaches 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 [33]).

Since in our approach for recognizing pedestrians we want to use the model for analyzing real-world images, the agreement with actual movements is important. On the other hand, 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 apply a kinematic approach.

For modelling the movement of walking we use data from human motion studies. Those studies have a long tradition (see [2, 74]). At the end of the 19th century, photographic methods have been developed [46, 39, 9]. For medical purposes in Murray *et al.* [45] (see also Murray [44]) the movements of 60 normal men ranging in age from 20 to 65 years have been analyzed to obtain the basic elements of walking. Astonishingly, the movement patterns of the body parts are very similar for the different persons, although it is often possible to recognize persons by their gait. We use for our model the average data of this investigation (see also [61]). A nice property is that only one parameter is needed to specify the relative positions of all body parts. For each of the joints at the shoulder, elbow, hip, and knee we have taken the values for the angle positions at 10 time instants. These values are interpolated by periodic cubic splines (see [66]). The whole walking cycle has been standardized to one. As an example, the motion curve of the knee-joint is displayed in Fig. 6. Analogously, we modeled the vertical displacement of the whole body. Since walking is a symmetric movement, the motion curves of the joints are only needed for one side of the human body. Movement states of our model for half of the walking cycle are shown in Fig. 7. Depicted are the contours of the cylinders under central projection. Hidden contours have been removed (e.g., [51, 53]). Fast reproduction of these motion states on a screen reveals that our motion model appears to be fairly realistic.

4. ESTIMATING THE MODEL PARAMETERS FROM SINGLE IMAGES

4.1. *Detecting Moving Objects*

We assume the image sequence to be recorded with a stationary camera. In order to segment the image plane in regions corresponding to moving and nonmoving objects we apply a change detection algorithm. For each image point we approximate the grey values in a 5×5 window by a bivariate polynomial of second order and compare polynomial fits in consecutive images (see [63, 8, 7]). Let $\boldsymbol{X}$ be the matrix representing the image coordinates and $\boldsymbol{\beta}$ the parameters of the fit. The n grey values of the window are given by $\boldsymbol{g}_k = \boldsymbol{X}\boldsymbol{\beta}_k$ and we can estimate $\hat{\boldsymbol{\beta}}_k = \boldsymbol{X}^{\#}\boldsymbol{g}_k$ by using the generalized inverse $\boldsymbol{X}^{\#} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}$. We consider a point of an image k to represent systematic changes, if ($\|\cdot\|_2$ denotes the Euclidean norm)

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

holds for the preceding ($j = k - 1$) or for the following ($j = k + 1$) image, where T is a threshold (Figs. 8 and 9). Experimentally we have found that this procedure yields better results than the approach in Hsu *et al.* [26]. The reason for this is that in our approach we directly compare the polynomial fits in consecutive images and therefore we do not need to estimate image noise which, in general,

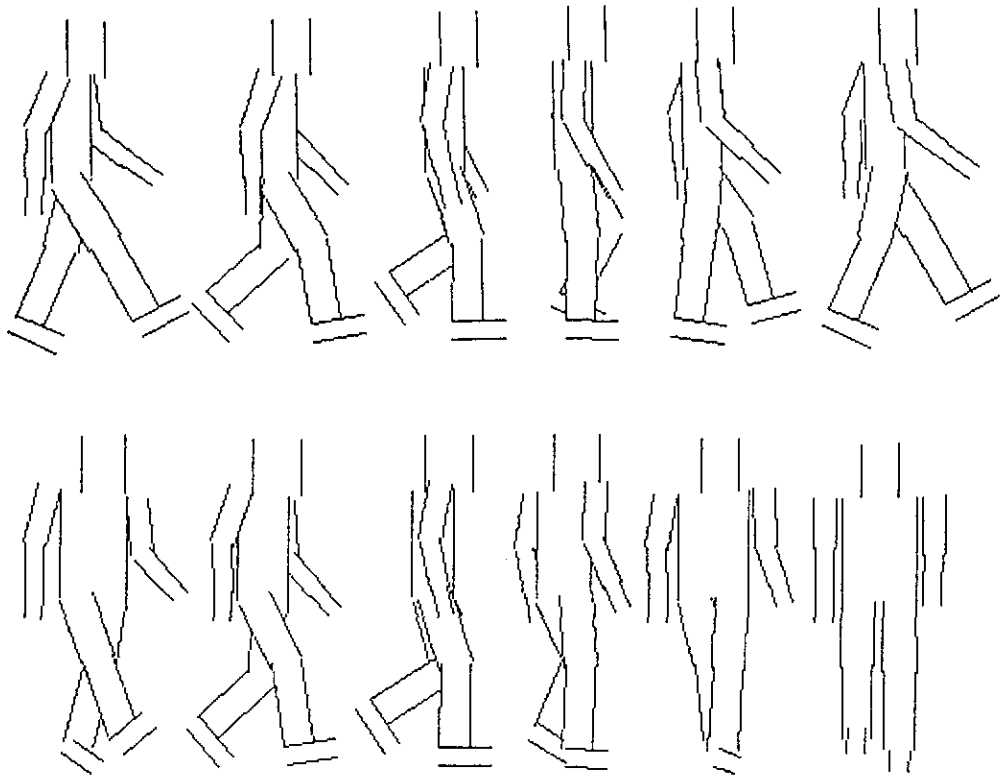

FIG. 7. Movement states for half of the walking cycle.

cannot be done very reliably using only local measurements.

In order to improve the detected image regions we apply binary image operations. Detected image points having a small number of neighbors in a 3×3 window are deleted and nondetected points with a large number of neighbors are marked. Since the relations between neighbors change permanently during this process we implemented these operations recursively. If, at the end of this procedure, there are still holes left (not marked areas surrounded by detected image points) we fill them by an additional algorithm. The application of our approach yields the re-

FIG. 8. Grey-value image of a pedestrian.

FIG. 9. Result of the change detection algorithm.

FIG. 10. Result after applying binary image operations.

FIG. 12. Velocity field.

sult in Fig. 10. Most false detections have been removed and the detected object candidate is a useful representation of the imaged pedestrian.

In addition, we compute a surrounding rectangle and the outline (Figs. 10 and 11). For the object candidate a velocity field is evaluated with the approach in Schnörr [65] (Fig. 12). Based on these descriptions, in Bister *et al.* [8], we introduced an approach to automatically derive trajectories of several simultaneously moving objects without using explicit models. For a street crossing with moving cars, pedestrians, and cyclists observed from bird's-eye view we demonstrated the applicability of our approach.

4.2. *Determining the 3D-Position*

The next step is to estimate the 3D-position of the considered pedestrian. For our approach we suppose the depicted scene to be calibrated (e.g., [60, 18, 58]). We exploit the information of the enclosing rectangle of the detected object candidate and make an assumption about the absolute height of the observed person. Using homogeneous coordinates the projection of a 3D-point $\boldsymbol{X} = (X, Y, Z, 1)^T$ onto the image plane (Fig. 13) with $\boldsymbol{x} = (x, y, *, 1)$, $*$ denoting an arbitrary value, h being nonzero, and the 4×4 matrix $\boldsymbol{T}$ is given by

$$h\boldsymbol{x} = \boldsymbol{T}\boldsymbol{X}. \tag{2}$$

FIG. 11. Outline.

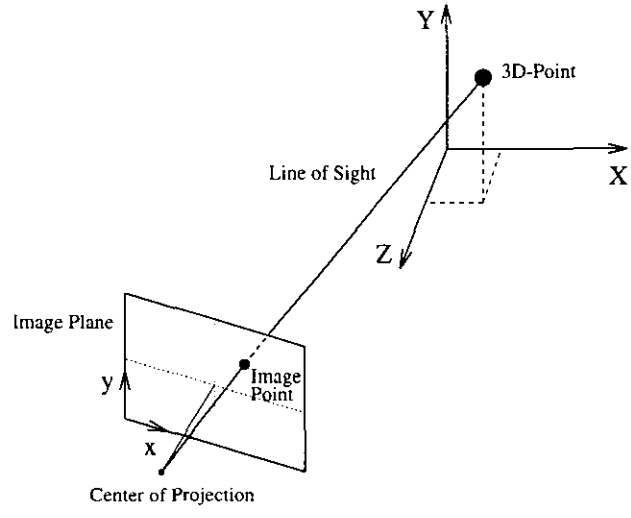

FIG. 13. Used coordinate systems.

The midpoints of the bottom and the top edges of the detected rectangle $\boldsymbol{x}_u$ and $\boldsymbol{x}_o$ 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 $\boldsymbol{H} = (0, H, 0, 0)^T$ we can write

$$h_u \boldsymbol{x}_u = \boldsymbol{T} \boldsymbol{X}_u \tag{3}$$

$$h_o \boldsymbol{x}_o = \boldsymbol{T}(\boldsymbol{X}_u + \boldsymbol{H}). \tag{4}$$

From (3) and (4)

$$h_u \boldsymbol{x}_u - h_o \boldsymbol{x}_o + \boldsymbol{T}\boldsymbol{H} = \boldsymbol{0} \tag{5}$$

follows and with

$$\boldsymbol{T}^* = \begin{pmatrix} t_{11} & t_{12} & t_{13} & -x_u & 0 \\ t_{21} & t_{22} & t_{23} & -y_u & 0 \\ t_{41} & t_{42} & t_{43} & -1 & 0 \\ 0 & 0 & 0 & x_u & -x_o \\ 0 & 0 & 0 & y_u & -y_o \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}, \quad \boldsymbol{X}^* = \begin{pmatrix} X_u \\ Y_u \\ Z_u \\ h_u \\ h_o \end{pmatrix},$$

$$\boldsymbol{b}^* = \begin{pmatrix} t_{14} \\ t_{24} \\ t_{44} \\ Ht_{12} \\ Ht_{22} \\ Ht_{42} \end{pmatrix},$$

we obtain the following linear system of equations:

$$\boldsymbol{T}^* \boldsymbol{X}^* + \boldsymbol{b}^* = \boldsymbol{0}. \tag{6}$$

For the five unknowns of $\boldsymbol{X}^*$ we have six equations; i.e., (6) is overdetermined. An approximate solution could be computed by using a numerical method. However, in our case we require the condition that the connection line of the 3D-positions of the sole and the top of the head is perpendicular to the plane where the pedestrian walks. To satisfy this requirement we allow a shift of the midpoints of the bottom and the top edges of the rectangle. The requirement is satisfied when the last three equations in (6) are exactly fulfilled. Having determined h_u, x_u, and y_u in this manner, the 3D-position of the sole $\boldsymbol{X}_u$ can be evaluated using the first three equations in (6). Then the 3D-position of the top of the head results by adding H to the Y-component of $\boldsymbol{X}_u$.

For a real image sequence the midpoints of the rectangle

FIG. 14. Midpoints of the bottom and top edges of the detected rectangles.

are displayed in Fig. 14. The estimated 3D-positions projected on the XZ-plane and connected by straight lines are shown in Fig. 15. The distance in X-direction covered by the pedestrian is about 4.5 m and the distances of the pedestrian to the camera lie between 10.9 m and 12.6 m. From this result we see that a single measurement is not very reliable because of the variations of the height of the detected rectangles. Therefore, it is advisable to take into account several measurements.

4.3. *Matching Model Contours with Grey-Value Edges*

In order to estimate more precisely the 3D-position and, in addition, to determine the posture of the pedestrian, we compare model contours with grey-value edges (see also [62]). Since the contours of our model consist of

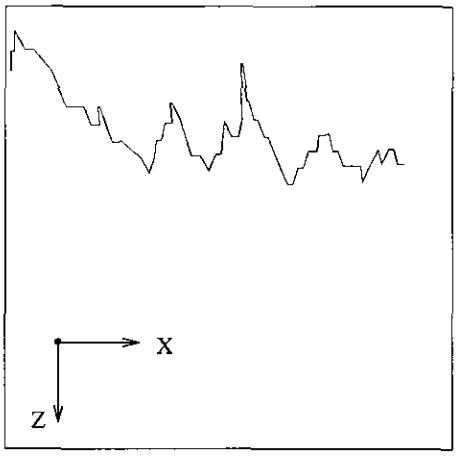

FIG. 15. Estimated 3D-positions projected on the XZ-plane.

FIG. 16. Enclosing rectangle and grey-value edge lines.

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. Inside the detected rectangle we compute edge points with the approach of Korn [32], link them together, and approximate them by eigenvector line fitting as described in Duda and Hart [16, p. 332] (see Fig. 16).

In a certain motion state we compute for each visible model contour a window and cut the grey-value edges to the portion inside the window (Figs. 17 and 18). We use a measure of similarity between model contours and grey-value edges that takes into account three geometric quantities. The length l_i is the projection of the (cut) grey-value edge onto the model contour with length l_{Mi} and measures how much the two edges overlap (since we compare the projected grey-value edge with the starting and endpoints of the model edge, l_i is always smaller than or equal to l_{Mi}). The second quantity is the distance d_i between the midpoint of the (cut) grey-value edge and the corresponding projection onto the model line. The third quantity is the angle between the two edges. The larger l_i and the smaller $(l_{Mi} - l_i)$, d_i, and $\Delta\varphi_i$, the more similar are the two edges. We use the following measure of similarity:

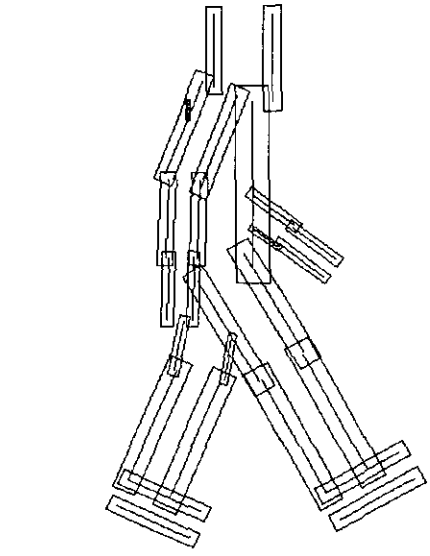

FIG. 17. Search windows of the model.

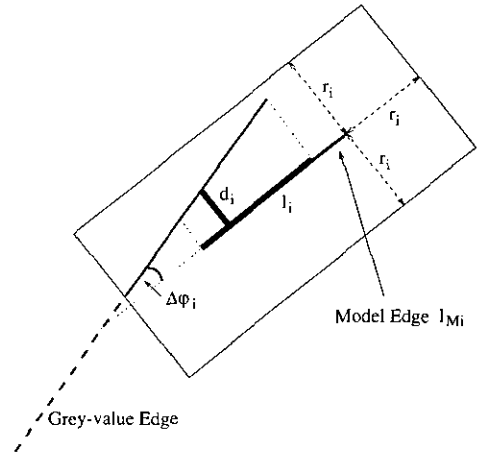

FIG. 18. Comparison between grey-value edge and model contour.

$$s_i = l_i \exp -\frac{1}{2}\left(\frac{(l_{Mi} - l_i)^2}{\sigma_{li}^2} + \frac{d_i^2}{\sigma_{di}^2} + \frac{\Delta\varphi_i^2}{\sigma_{\Delta\varphi}^2}\right). \tag{7}$$

With

$$\Sigma_i = \begin{pmatrix} \sigma_{li}^2 & 0 & 0 \\ 0 & \sigma_{di}^2 & 0 \\ 0 & 0 & \sigma_{\Delta\varphi}^2 \end{pmatrix}, \quad \boldsymbol{k}_i = \begin{pmatrix} l_{Mi} - l_i \\ d_i \\ \Delta\varphi_i \end{pmatrix}$$

we can write

$$s_i = l_i\, e^{-1/2\boldsymbol{k}_i^T \Sigma_i^{-1} \boldsymbol{k}_i}. \tag{8}$$

The parameters σ_{li} and σ_{di} 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 (8) is independent of scaling the model. Since we want grey-value edges with larger values of l_i to have a larger influence on the overall similarity, we weight the exponential function by this value. Alternatively, one could weight by l_{Mi}.

Many approaches for line matching compare the midpoints or the starting and endpoints of the model contour with the grey-value edge supposing that the two edges are similar in length (e.g., [38, 43, 67, 6]). In our application the grey-value edges of the lower and upper parts of the arms and the legs often are connected to one single 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 above.

For qualitatively comparing our measure of similarity with the one in Lowe [38] we write the exponential function in (7) as the multiplication of three single exponential functions $e^{-x_i^2/2}$, where $x_i = (l_{Mi} - l_i)/\sigma_{li}$, d_i/σ_{di} and $\Delta\varphi_i/\sigma_{\Delta\varphi}$ and replace each of these functions by $1/x_i$, which for x_i larger than a minimum value has a comparable course. Then we have

$$s_i = c\, l_i \frac{l_{Mi}^2}{(l_{Mi} - l_i)\, d_i\, \Delta\varphi_i} \tag{9}$$

with c being constant. The measure in Lowe [38] is given by

$$s_{i_L} = c_L \frac{l_{GWi}^2}{(l_{Mi} - l_{GWi})\, d_i\, \Delta\varphi_i} \tag{10}$$

with c_L being constant and l_{GWi} denoting the length of the (uncut) grey-value edge. The comparison shows that the normalization in (10) is done by l_{GWi} and not by l_{Mi} as in (9). In addition, (10) is independent of scale. In our application grey-value edges often are longer than model contours. Besides that, the (visible) model contours in general are highly different in length and we want model edges with larger values l_i to have a larger influence on the overall similarity. Therefore, in our opinion, (9) is the more adequate measure for our case. The use of the exponential function in (7) has the additional advantage that the resulting expression is defined for all values and decreases to zero fast for dissimilar edges. If several grey-value edges overlap the search window, then we take the one with highest similarity value. The overall measure of similarity in dependence of the model parameters $\boldsymbol{p}$ is the sum of the values s_i for all visible model edges of the pedestrian normalized by the sum of their lengths l_{Mi}:

$$s(\boldsymbol{p}) = \frac{\sum_{i=1}^{n} w_{Gi}\, w_{Vi}\, w_{Ki}\, s_i}{\sum_{i=1}^{n} l_{Mi}} \rightarrow \max. \tag{11}$$

The single values s_i can be weighted by the magnitude of the grey-value gradient at the matched grey-value edge w_{Gi} and the magnitude of the corresponding velocity vectors w_{Vi}. By using the w_{Ki} it is possible to weight the significance of the single body parts for recognition.

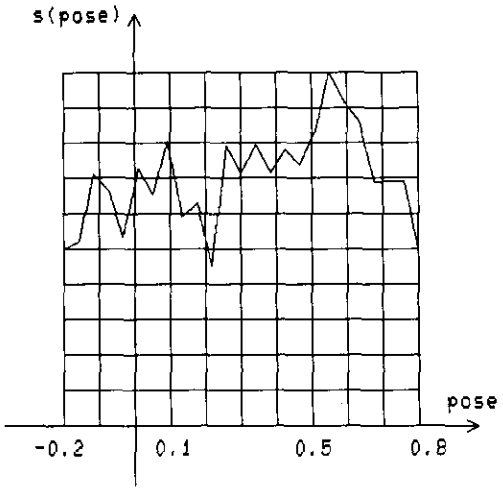

FIG. 19. Similarity curve.

We search those parameters $\boldsymbol{p}$ which maximize $s(\boldsymbol{p})$. In order to obtain a more stable recognition result we remove hidden model contours. However, one disadvantage is that no analytical relation between $\boldsymbol{p}$ and $s(\boldsymbol{p})$ can be determined. Therefore, for maximizing $s(\boldsymbol{p})$ we use a grid search method with equally spaced points and take the state of motion with highest similarity value. Since the computational expense increases exponentially with the number of parameters, in further work possibilities to reduce the computation time should be investigated, e.g., using a smaller number of space points, choosing

FIG. 20. Determined motion state.

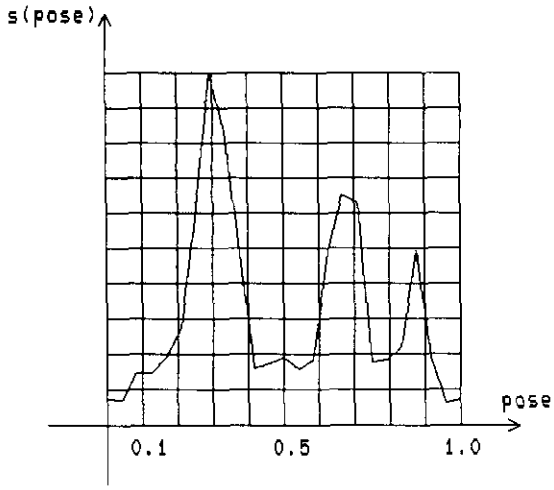

FIG. 21. Similarity curve.

FIG. 22. Determined motion state.

a variable density of space points, or using some other optimization procedure.

Fixing the 3D-position of the model and varying the posture parameter *pose* within the whole walking cycle, for the grey-value edges in Fig. 16, we obtain the similarity curve in Fig. 19 using equal weights w_i for all the model edges in (11). The state of motion with the highest similarity (*pose* ≈ 0.55) is superimposed onto the original image in Fig. 20 and agrees with the observation. However, the similarity curve has several secondary maxima and we see that in general it is not a good choice to use a downhill method. The relatively large values around *pose* ≈ 0.05 which stem from the fact that the walking cycle is symmetric with respect to time are interesting. Even for a human observer it is hard to decide whether the right leg is in front of the left leg or vice versa and surely there are cases in which our algorithm determines the posture displaced by half of the walking cycle from the actual posture. Another example in Figs. 21 and 22 shows qualitatively the same result.

5. ESTIMATING THE MODEL PARAMETERS FROM CONSECUTIVE IMAGES

In this section, the approach for evaluating single images as described above is extended to image sequences. We first perform an initialization and then incrementally estimate the model parameters by using a Kalman filter approach.

5.1. *Initialization*

A human observer only needs 0.2 s (less than a quarter of the walking cycle) to recognize a walking human represented by moving light displays [28, 55]. Therefore, automatically analyzing a number of images which show about half a walking cycle (circa 10–15 images) should be sufficient to obtain starting values. In the initialization phase for each image independently we apply the change detection algorithm and the procedure for estimating the 3D-position in Section 4. When matching image edges to model contours the search space for the posture is the whole walking cycle ($2\,\Delta pose = 1$), whereas after initialization (because of the knowledge available now) the search space can be reduced (e.g., $2\,\Delta pose = 0.4$). In addition, we vary X and Y around the estimated 3D-position.

For a synthetic image sequence (grey-value images can be seen in Fig. 27) the determined postures are marked in Fig. 23. The dashed line indicates the actual movement. With the exception of the outlier at image number 10 which is displaced about the half of the walking cycle

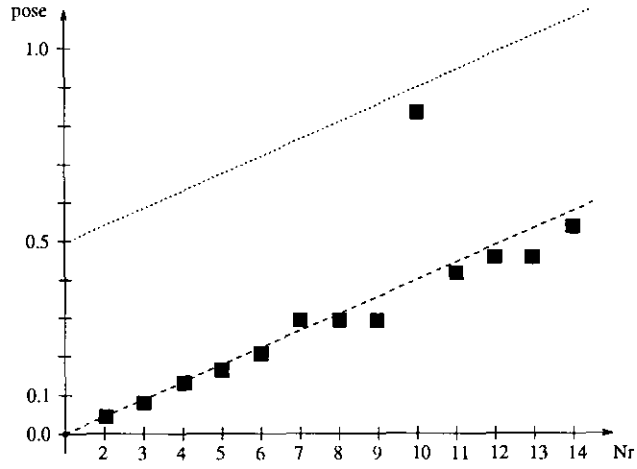

FIG. 23. Initialization for a synthetic image sequence.

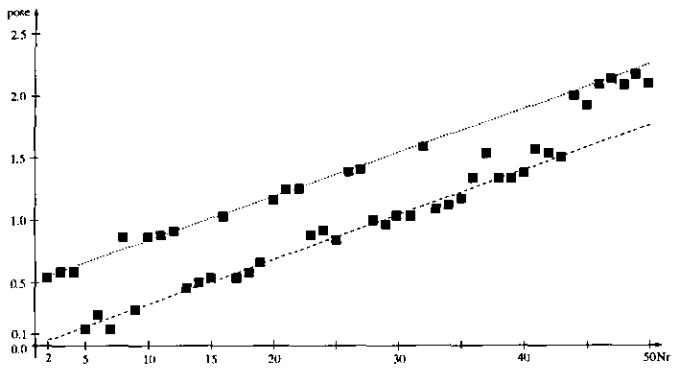

FIG. 24. Initialization for a real image sequence (f2).

(dotted line) the estimates agree with the actual movement. The estimates for a real image sequence named f2 (Figs. 24 and 29) also lie relatively well on the line of the actual movement and the one displaced by half of the walking cycle. However, the number of false estimates displaced about $\Delta pose \approx 0.5$ is much higher. It should be noted that for this image sequence the image portion covered by the pedestrian is relatively small and therefore it is hard to decide whether the right leg is in front of the left leg or vice versa. For another image sequence (Figs. 25, 31), where the pedestrian covers a larger portion of the image plane, the number of false estimates displaced about $\Delta pose \approx 0.5$ is smaller.

5.2. *Incremental Estimation*

After initialization we apply the recursive equations of the Kalman filter [29, 20]. In the field of computer vision, Kalman filter approaches have been introduced, for example, by Broida and Chellappa [10] and Deriche and Faugeras [14]. Broida and Chellappa [10] describe an approach for incrementally estimating model parameters of rigid objects from measured image points, whereas in Deriche and Faugeras [14] grey-value edge lines are tracked in the image plane. Our aim is to recognize human movements. Since the body parts of pedestrians often are occluded by one another (self-occlusion), tracking of image lines in general would lead to severe problems.

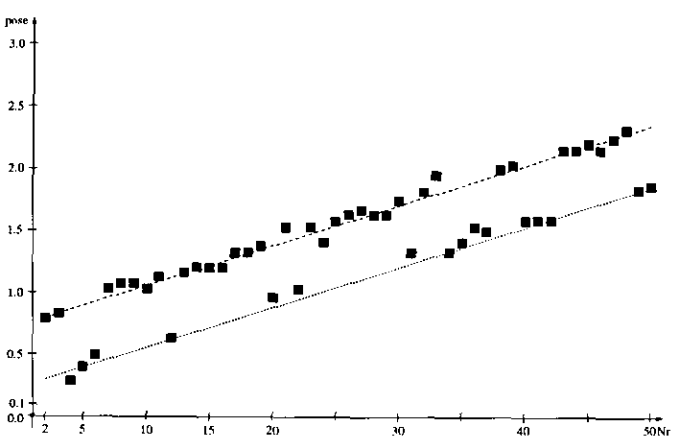

FIG. 25. Initialization for a real image sequence (f3).

In the following we assume the pedestrian to move with constant velocity. By interpreting the parameter values for the best fit in a single image as measurements for the corresponding time instant, the system description and the measurement model for 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

$$\boldsymbol{p}_k = \boldsymbol{\Phi}_{k,k-1}\boldsymbol{p}_{k-1} + \boldsymbol{\Gamma}_{k-1}\boldsymbol{w}_{k-1} \tag{12}$$

$$\boldsymbol{z}_k = \boldsymbol{H}_k\boldsymbol{p}_k + \boldsymbol{v}_k. \tag{13}$$

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

Predicting the parameters and the covariance matrix is done by

$$\boldsymbol{p}_k^* = \boldsymbol{\Phi}_{k,k-1}\hat{\boldsymbol{p}}_{k-1} \tag{14}$$

$$\boldsymbol{P}_k^* = \boldsymbol{\Phi}_{k,k-1}\hat{\boldsymbol{P}}_{k-1}\boldsymbol{\Phi}_{k,k-1}^T + \boldsymbol{Q}_{k-1}. \tag{15}$$

With these predictions and the current measurement $\boldsymbol{z}_k$ the estimates $\hat{\boldsymbol{p}}_k$ and $\hat{\boldsymbol{P}}_k$ in the current image are computed by ($\boldsymbol{I}$ is the identity matrix)

$$\hat{\boldsymbol{p}}_k = \boldsymbol{p}_k^* + \hat{\boldsymbol{K}}_k(\boldsymbol{z}_k - \boldsymbol{H}_k\boldsymbol{p}_k^*) \tag{16}$$

$$\hat{\boldsymbol{P}}_k = (\boldsymbol{I} - \hat{\boldsymbol{K}}_k\boldsymbol{H}_k)\boldsymbol{P}_k^* \tag{17}$$

$$\hat{\boldsymbol{K}}_k = \boldsymbol{P}_k^*\boldsymbol{H}_k^T(\boldsymbol{H}_k\boldsymbol{P}_k^*\boldsymbol{H}_k^T + \boldsymbol{R}_k)^{-1}. \tag{18}$$

At the beginning, a starting vector $\boldsymbol{p}_0$ and a covariance matrix $\boldsymbol{P}_0$ are needed.

In our application, we want to estimate the 3D-position and posture of the observed pedestrians. The state vector is $\boldsymbol{p}_k = (X_k, \dot{X}_k, Y_k, \dot{Y}_k, Z_k, \dot{Z}_k, pose_k, \dot{pose}_k)^T$ and the time difference between two images is Δt. Supposing constant velocities, we have

$$\Phi_{k,k-1} = \begin{pmatrix} 1 & \Delta t & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \Delta t & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \Delta t & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (19)$$

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

$$\boldsymbol{Q}_k = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_{\dot{Q}X} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{\dot{Q}Y} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{\dot{Q}Z} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{Q\dot{p}ose} \end{pmatrix}. \quad (20)$$

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

$$\boldsymbol{H}_k = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad (21)$$

and uncertainties of the measurements are characterized by

$$\boldsymbol{R}_k = \begin{pmatrix} \sigma^2_{RX} & 0 & 0 & 0 \\ 0 & \sigma^2_{RY} & 0 & 0 \\ 0 & 0 & \sigma^2_{RZ} & 0 \\ 0 & 0 & 0 & \sigma^2_{Rpose} \end{pmatrix}. \quad (22)$$

6. EXPERIMENTAL RESULTS

We show experimental results for a synthetic as well as for real-world image sequences. In the initialization phase we evaluate the images number 2 to 14 (see Figs. 23, 24, and 25) and determine starting values by linear regression. The 3D-position $\boldsymbol{X}$ is an estimate for the center of the torso (origin of the overall coordinate system). Since the estimation for *pose*, especially in real scenes, often is displaced by half of the walking cycle, for this parameter the estimates are first grouped together. In general the estimates lie approximately on two lines. It is possible that the number of estimates for the actual movement is smaller than for the displaced movement. Therefore, if for one line the number of estimates is higher than half of the number for the other line, we take the line with highest mean value of similarity. We note that, although in our experiments it was possible to automatically estimate the actual initial posture with this procedure, in general this is a very hard problem and should be investigated further.

For incrementally estimating the searched parameters we predict the model for the following image and compare the model contours with the grey-value edges. The search space for the parameters is placed symmetrically around the predicted values. In this phase we do not apply the change detection algorithm. The assumption that the pedestrian moves parallel to the image plane on a plane in the scene simplifies the expressions (19)–(22) for the Kalman filter. The scene coordinate Z describing the depth of the observed human is supposed to remain constant and we take over the value from the initialization phase. Incrementally we estimate the height of the person above the plane of movement (Y) assuming that the (normalized) height remains constant, i.e., $\dot{Y} = 0$ and therefore $\sigma^2_{\dot{Q}Y} = 0$. In addition, the coordinate X in the direction of the movement and the posture *pose* are estimated. The initial uncertainties for the velocities of these parameters are derived from the average velocities of pedestrians. Assuming 1s time duration for a walking cycle and the covered distance to be 1.6 m (e.g., Murray *et al.* [45, 44] and Inman [27]), the average velocities of X and *pose* at an image rate of 25 images per second are $\bar{\dot{X}} = 6.4$ cm/image and $\overline{\dot{pose}} = 0.04$/image. A chosen deviation of 30% then leads to $\sigma^2_{\dot{Q}X} = 3.69$ cm^2/image^2 and $\sigma^2_{Q\dot{p}ose} = 0.000144$/image^2. The other elements for the covariance matrices have been chosen heuristically and in such a way that we obtain reasonable values for automatically controlling the search spaces for the parameters. With the intervals $2\Delta X = 40$ cm, $2\Delta Y = 20$ cm, and $2\Delta pose = 0.4$ we have $\sigma^2_{RX} = 400$ cm^2/image^2, $\sigma^2_{RY} = 100$ cm^2/image^2, and $\sigma^2_{Rpose} = 0.04$/image^2. For the corresponding elements of the initial covariance matrix we chose the same values, such that at the beginning the initial values and the new measurements have the same uncertainties. If we subdivide the search intervals for X and Y every 2 cm and the interval for *pose* every 0.02 then the number of space points are given by $n_X = [\Delta X]$, $n_Y = [\Delta Y]$, and $n_{pose} = [100\ \Delta pose]$, where [ ] denotes rounded natural (odd) number.

The application of our approach for a synthetic image

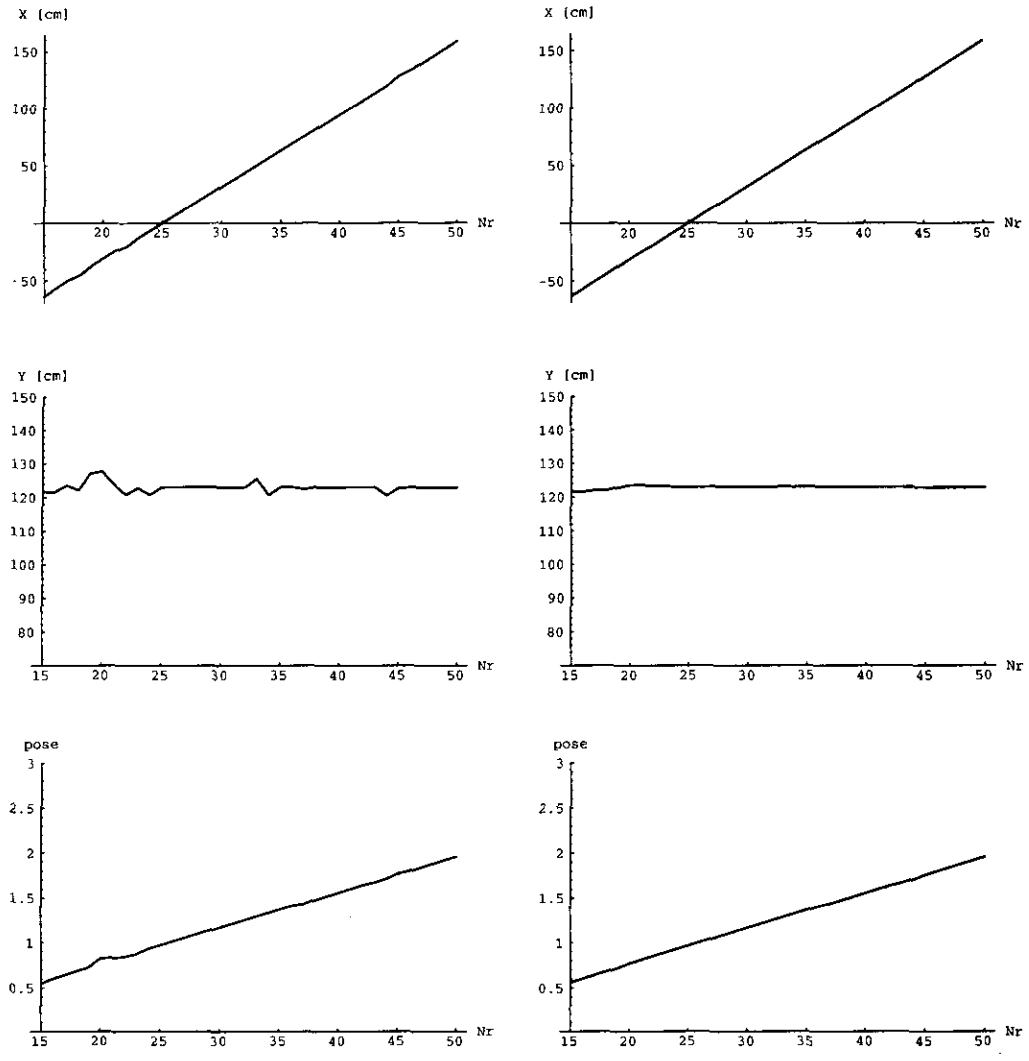

FIG. 26. Estimation result (X, Y, *pose*) for a synthetic image sequence: left, measurements; right, Kalman filter estimates.

sequence leads to the results for X, Y, and *pose* (in dependence of the image number) in Fig. 26. On the left side are the determined parameter values for the best fit and on the right side is the estimation result of the Kalman filter. The pedestrian has been tracked over the whole sequence. For some of the original images we have superimposed the estimated models. The agreement is fairly good (Fig. 27).

For the real scene named f2 we obtain the diagrams in Fig. 28. The pedestrian also has been tracked over the whole sequence, although the image portion covered by the pedestrian is relatively small. Since real images in contrast to synthetic images are disturbed by noise, here the smoothing properties of the Kalman filter can be seen more clearly (see especially the result for Y). The estimated models in Fig. 29 agree with the observation. However, in some images the model deviates a bit from the imaged pedestrian.

The results for the real image sequence f3 are similar (Figs. 30 and 31). Since for this sequence the image portion covered by the pedestrian is larger, the deviations between the model and the imaged pedestrian are easier to recognize. In order to analyze more precisely these deviations, the estimated motion curves of the hip- and knee-joint for the right and left sides are depicted in Figs. 32 and 33 (solid lines). For comparison, we also displayed corresponding curves obtained by interactively adjusting the model such that for a human observer the agreement is good (dashed lines). It should be noted that such an adjustment is not very easy since even small changes of the angle at the hip-joint (e.g., 3°–5°) lead to relatively large changes when superimposing the model onto the image. Nevertheless, from the figures we see that the estimated curves generally agree with the interactively obtained curves. However, at the hip-joint, especially for positive angle values, there are larger deviations which

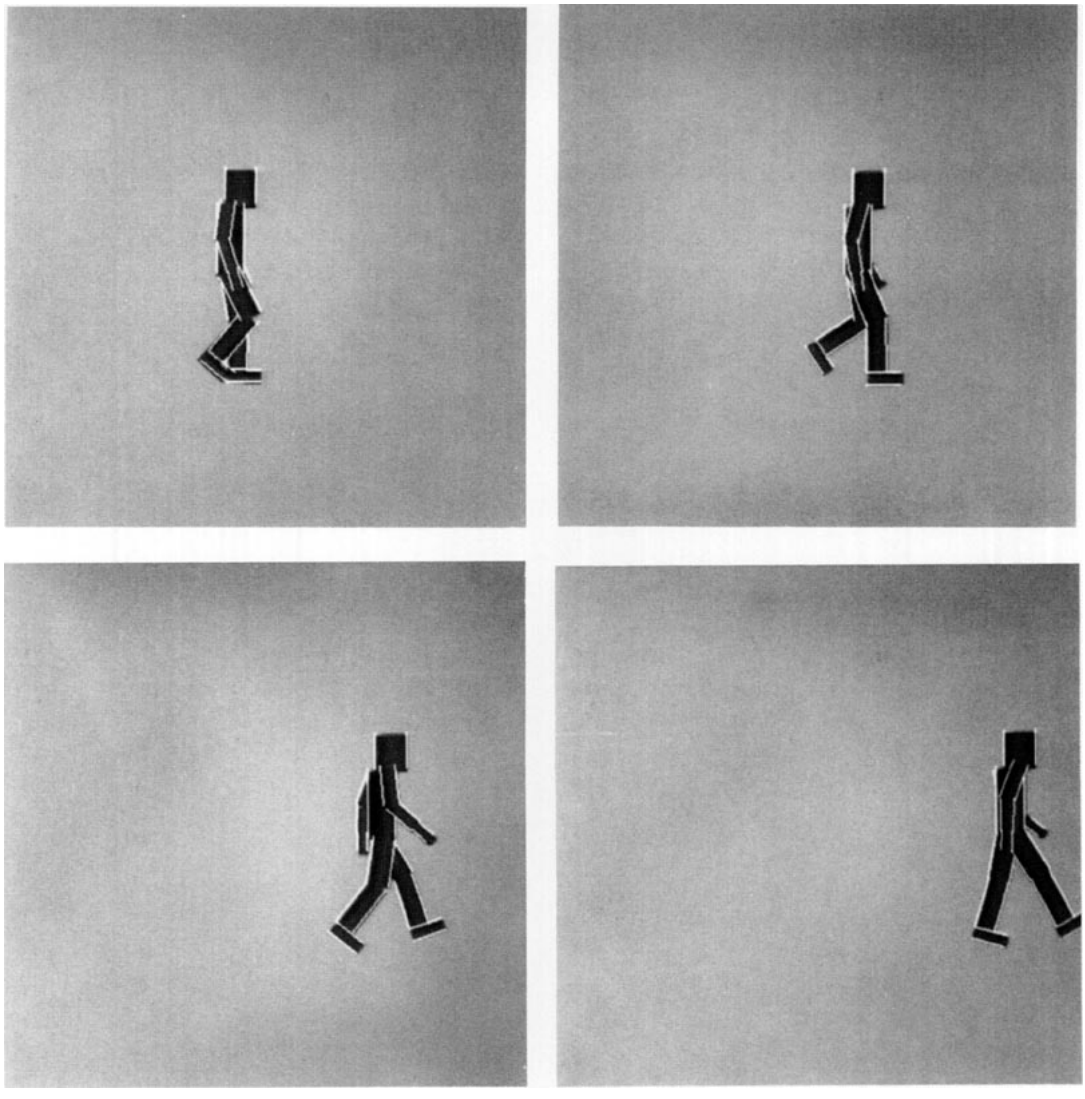

FIG. 27. Estimated models for a synthetic image sequence (image numbers 20, 30, 40, and 50).

result in the overall difference between the model contours and the image data. For comparison, in Figs. 32 and 33 we also draw those curves which for the corresponding velocity represent the optimal movement with our motion model. These curves also generally agree with the other motion curves.

7. CONCLUSION AND FUTURE WORK

Toward the long-term goal of recognizing and describing human movements in image sequences we introduced an approach for the recognition of pedestrians. For a synthetic as well as for real image sequences we were able to track the imaged pedestrian over the whole sequences. Starting values are determined automatically. Incrementally we estimate the 3D-position and the posture, assuming that the observed person walks parallel to the image plane. By applying a Kalman filter we obtain smooth estimation results.

Whereas for synthetic images (generated with the same model as used for interpretation) the agreement between the estimated models and the image data is fairly good, in real-world images differences can be observed. These differences are due to individual deviations of the analyzed pedestrian from our movement model which represents an average over a relatively large number of 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 as in our approach but then to continue the evaluation by adaptively estimating

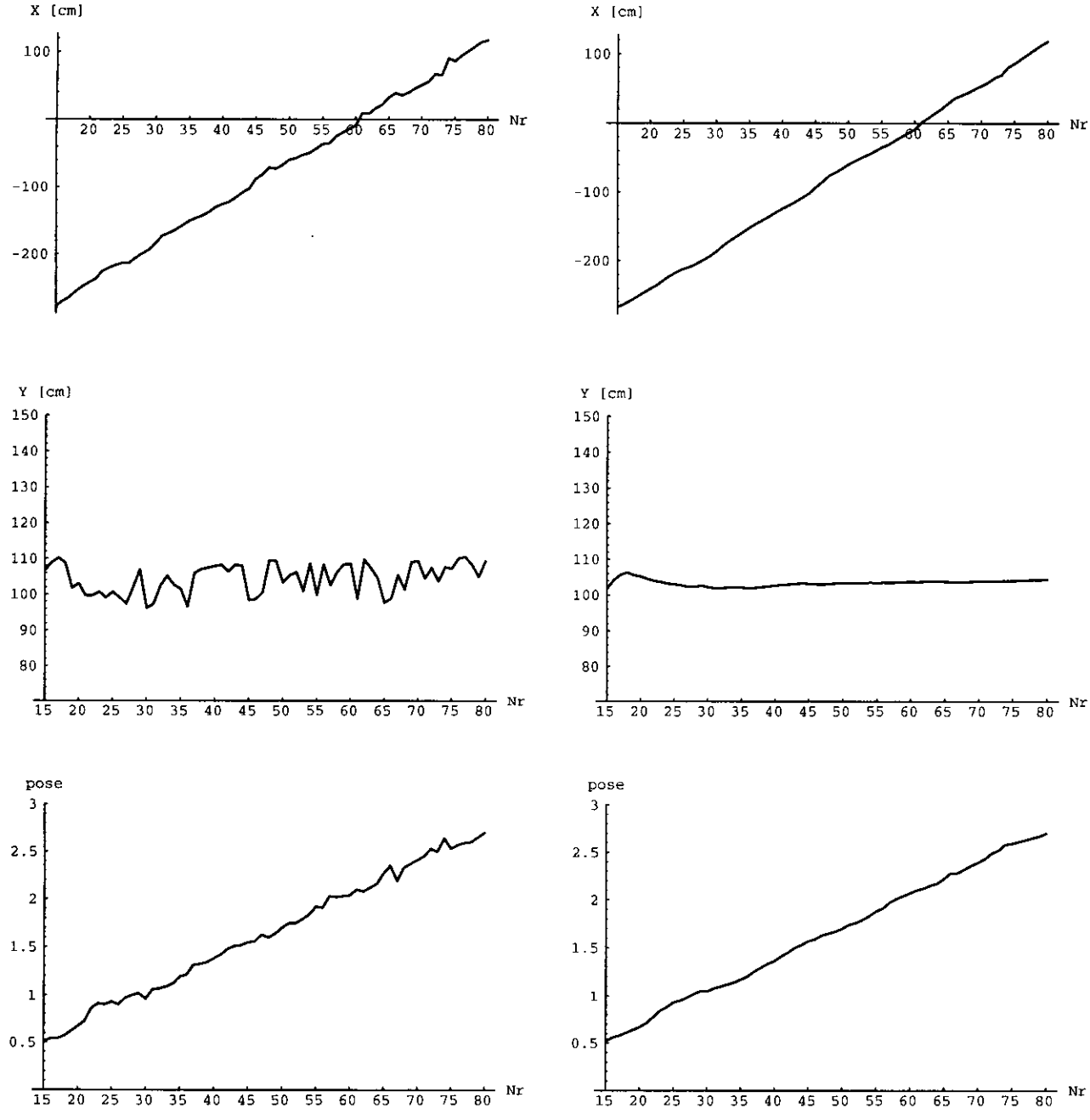

FIG. 28. Same as Fig. 26 for a real image sequence (f2).

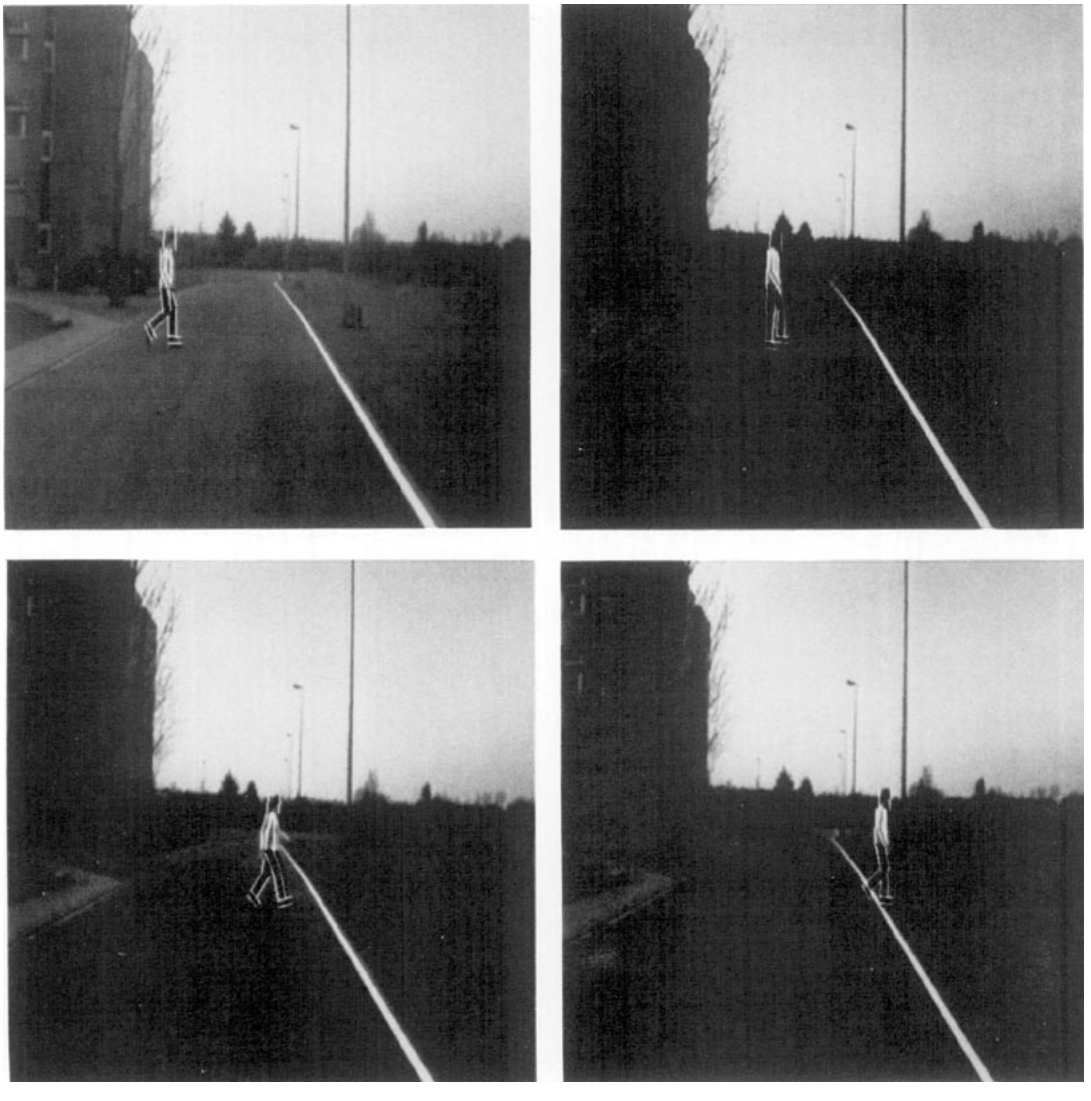

FIG. 29. Estimated models for a real image sequence (f2) (image numbers 20, 40, 60, and 80).

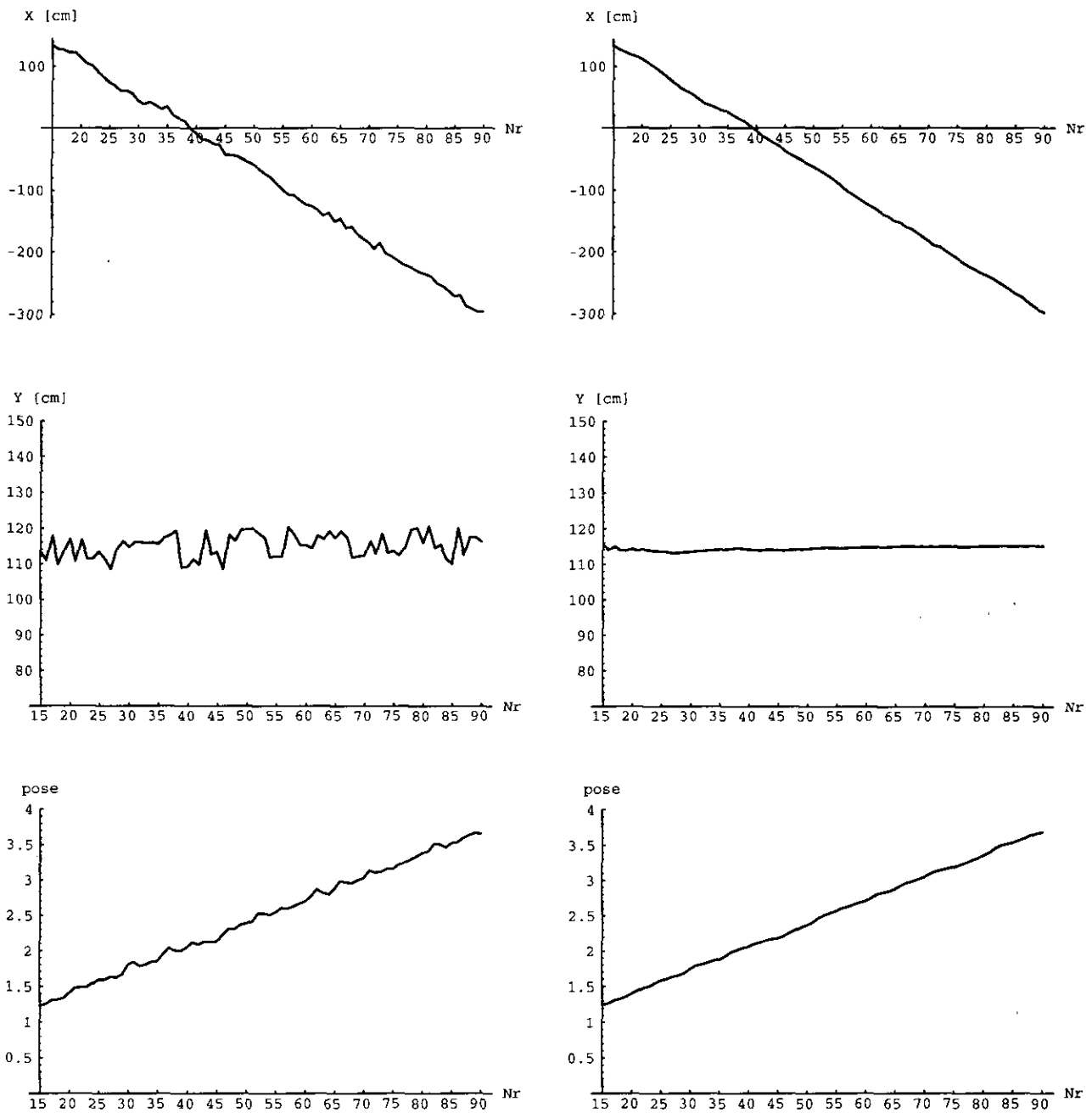

FIG. 30. Same as Fig. 26 for a real image sequence (f3).

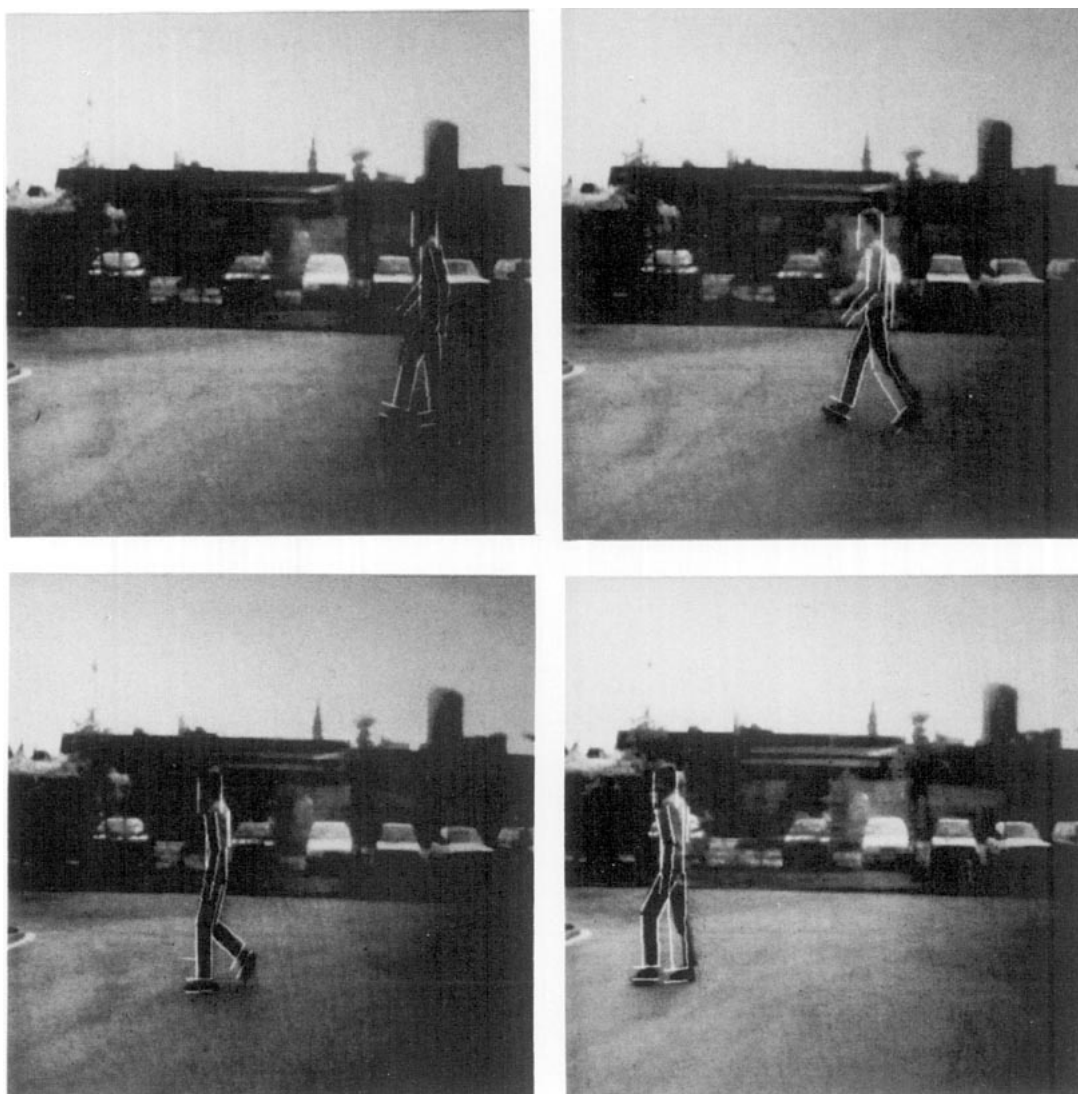

FIG. 31. Estimated models for a real image sequence (f3) (image numbers 20, 40, 60, and 80).

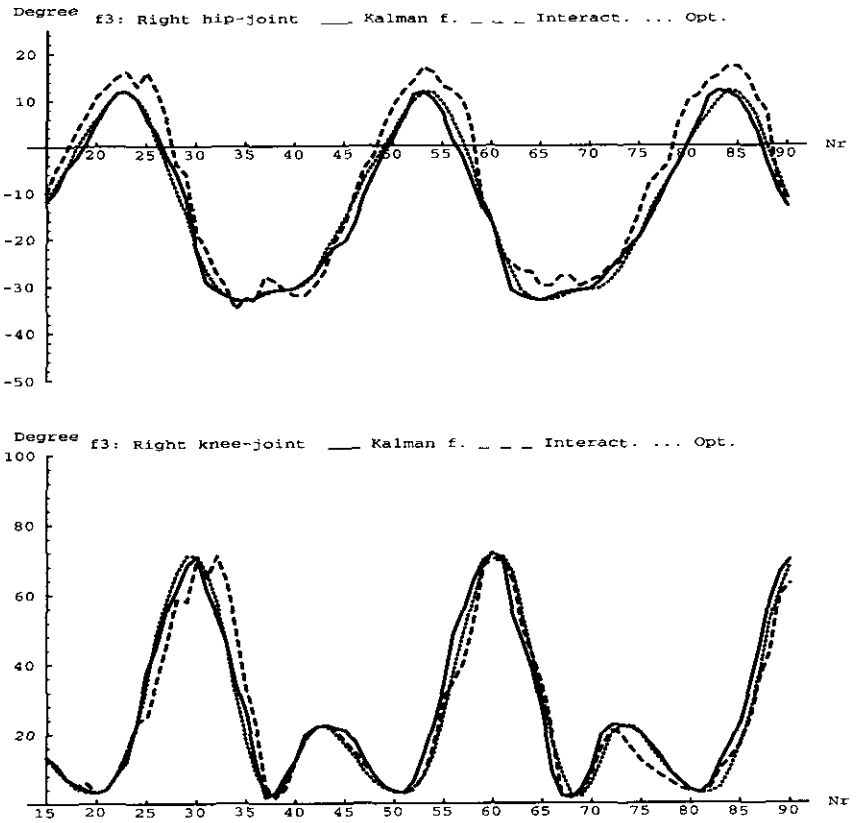

FIG. 32. Motion curves (right side) for the real scene f3: top, hip; bottom, knee.

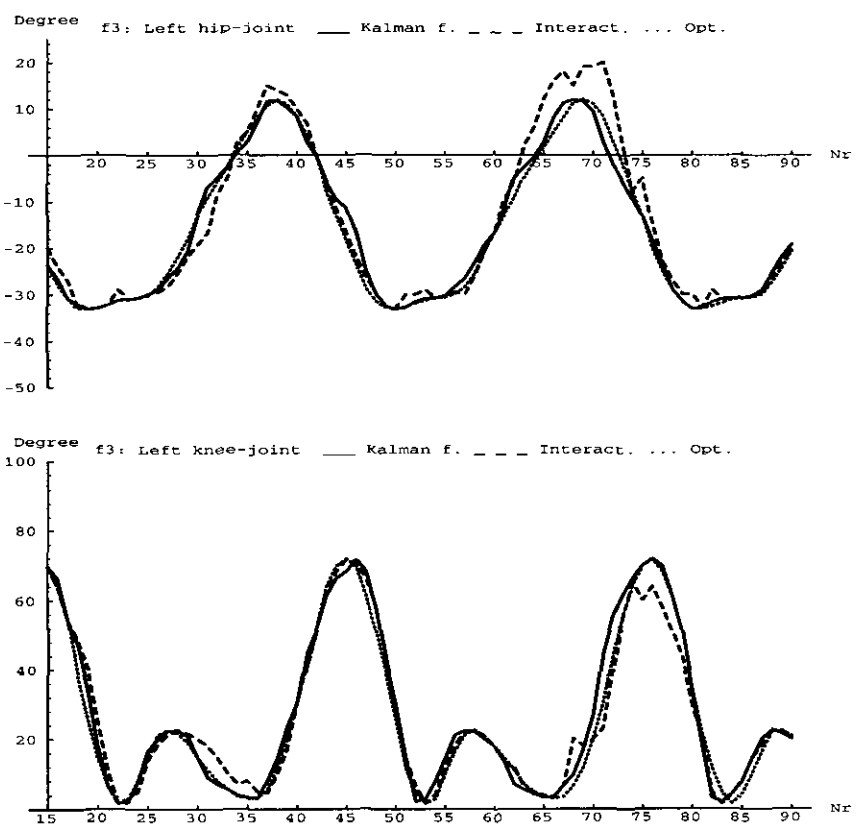

FIG. 33. Motion curves (left side) for the real scene f3: top, hip; bottom, knee.

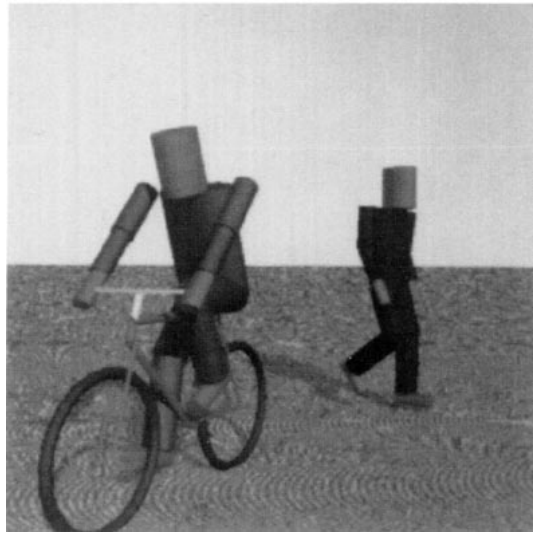

FIG. 34. Moving pedestrian and cyclist.

these curves. Also, additional image features could be taken into account; e.g., one could compare the image velocity field with the model velocity field. For estimating the posture more precisely it is also possible to introduce additional parameters for the body model. In either case a good compromise between the quality of representation and the number of parameters has to be found. A very hard problem is the decision which of the two legs is in front of the other and in our case the incremental estimation with the Kalman filter depends on this decision. With the extensions above it is perhaps possible to determine the initial posture more robustly.

In order to match more efficiently the model to the image data a view graph for the model of the pedestrian could be used (see, e.g., [31, 59]).

FIG. 35. Moving pedestrian and car.

Further investigations should also analyze movements that are not parallel to the image plane and movements with changing orientations. A possible extension of our approach is the recognition of simultaneously moving (rigid and nonrigid) bodies (single images from simulations are shown in Figs. 5, 34, and 35). When analyzing these scenes the occlusion problem has to be treated explicitly.

ACKNOWLEDGMENTS

For discussions and critical comments I thank D. Bister, C. Bregler, K. Daniilidis, G. Herzog, W. Leister, B. Neumann, J. Rieger, C. Schnörr, W. Wahlster, G. Winkler, and G. Zimmermann. This work has been done while the author was with the Institut für Algorithmen und Kognitive Systeme, Fakultät für Informatik, Universität Karlsruhe (TH). The support of the DFG, Sonderforschungsbereich 314 "Künstliche Intelligenz—Wissensbasierte Systeme" is gratefully acknowledged.

REFERENCES

1. K. Akita, Image sequence analysis of real world human motion, *Pattern Recognit.* **17,** 1984, 73–83.
2. Aristoteles, Über die Bewegung der Lebewesen; Über die Fortbewegung der Lebewesen, Datierung um 330 v.Chr., Teil II, III, J. Kollesch, Aristoteles Werke in deutscher Übersetzung, Band 17, Zoologische Schriften II, (E. Grumach and H. Flashar, Eds.), Wissenschaftliche Buchgesellschaft, Darmstadt, 1985.
3. C. I. Attwood, G. D. Sullivan, and K. D. Baker, Model-based recognition of human posture using single synthetic images, in *Proc. Fifth Alvey Vision Conf., University of Reading, Reading, UK, September 25–28, 1989,* pp. 25–30.
4. N. I. Badler, *Temporal Scene Analysis: Conceptual Descriptions of Object Movements,* Tech. Rep. No. 80, Dept. of Computer Science, University of Toronto, Feb. 1975.
5. N. I. Badler, K. H. Manoochehri, and G. Walters, Articulated figure positioning by multiple constraints, *IEEE Comput. Graphics Appl.* **7**(6), 1987, 28–38.
6. J. R. Beveridge, R. Weiss, and E. M. Riseman, Optimization of 2-dimensional model matching in *Proc. Image Understanding Workshop, Palo Alto, California, May 23–26, 1989,* pp. 815–830.
7. D. Bister, *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.
8. D. Bister, K. Rohr, and C. Schnörr, Automatische Bestimmung der Trajektorien von sich bewegenden Objekten aus einer Bildfolge, in *12. DAGM—Symposium Mustererkennung, 24.–26. September 1990,* Oberkochen-Aalen, Informatik-Fachberichte 254 (R. E. Grosskopf, Ed.), pp. 44–51, Springer-Verlag, Berlin/Heidelberg, 1990.
9. W. Braune and O. Fischer, Der Gang des Menschen, 1. Theil: Versuche am unbelasteten und belasteten Menschen, *Abh. Math. Phys. Cl. König. sächsisch. Ges. Wiss.,* Band 21, S. Hirzel, Leipzig, 1895.
10. T. J. Broida and R. Chellappa, Estimation of object motion parameters from noisy images, *IEEE Trans. Pattern Anal. Mach. Intell.* **8,** 1986, 90–99.
11. A. Bruderlin and T. W. Calvert, Goal-directed, dynamic animation of human walking, *Comput. Graphics* **23**(3), 1989, 233–242.

12. T. W. Calvert and J. Chapman, Aspects of the kinematic simulation of human movement, *IEEE Comput. Graphics Appl.* **2**(9), 1982, 41–49.
13. R. Cipolla and M. Yamamoto, Stereoscopic tracking of bodies in motion, in *Proc. Fifth Alvey Vision Conf., University of Reading, Reading/UK, September 25–28, 1989*, pp. 109–114.
14. R. Deriche and O. Faugeras, Tracking line segments, *Image Vision Comput.* **8**(4), 1990, 261–270.
15. DIN 33402, Deutsche Normen, *Körpermasse des Menschen*, Beuth Verlag, Berlin, 1987.
16. R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.
17. O. Fischer, *Theoretische Grundlagen für eine Mechanik der lebenden Körper mit speziellen Anwendungen auf den Menschen, sowie auf einige Bewegungsvorgänge an Menschen*, Teubner, Leipzig/Berlin, 1906.
18. S. Ganapathy, Decomposition of transformation matrices for robot vision, *Pattern Recognit. Lett.* **2**(6), 1984, 401–412.
19. N. H. Goddard, The interpretation of visual motion: Recognizing moving light displays, in *Proc. Workshop on Visual Motion, Irvine, CA, March 20–22, 1989*, pp. 212–220.
20. A. Gelb (Ed.), *Applied Optimal Estimation*, MIT Press, Cambridge, MA, 1974.
21. T. C. Hartrum, *Computer Implementation of a Parametric Model for Biped Locomotion Kinematics*, Ph.D. thesis, Ohio State University, Columbus, Ohio, 1973.
22. G. Herzog, C.-K. Sung, E. André, W. Enkelmann, H.-H. Nagel, T. Rist, W. Wahlster, and G. Zimmermann, Incremental natural language description of dynamic imagery, in *Wissensbasierte Systeme*, Informatik-Fachberichte 227 (W. Brauer and C. Freksa, Eds.), pp. 153–162, Springer-Verlag, Berlin/Heidelberg, 1989.
23. D. D. Hoffman and B. E. Flinchbaugh, Interpretation of biological motion, *Biol. Cybern.* **42**, 1982, 195–204.
24. D. Hogg, Model based vision: A program to see a walking person, *Image Vision Comput.* **1**(1), 1983, 5–20.
25. D. Hogg, *Interpreting Images of a Known Moving Object*, Ph.D. dissertation, University of Sussex, Brighton, UK, 1984.
26. Y. Z. Hsu, H.-H. Nagel, and G. Rekers, New likelihood test methods for change detection in image sequences, *Comput. Vision Graphics Image Process.* **26**, 1984, 73–106.
27. H. V. T. Inman, H. J. Ralston, and F. Todd, *Human Walking*, Williams & Wilkins, Baltimore/London, 1980.
28. G. Johansson, Spatio-temporal differentiation and integration in visual motion perception, *Psychol. Res.* **38**, 1976, 379–396.
29. R. E. Kalman, A new approach to linear filtering and prediction problems, *Trans. ASME J. Basic Eng. Ser.* **82D**, March 1960, 35–45.
30. T. Kanade, Region segmentation: Signal vs. semantics, in *Proc. Int. Joint Conf. on Pattern Recognition, Kyoto/Japan, November 7–10, 1978*, pp. 95–105; *Comput. Graphics Image Process.* **13**, 1980, 279–297.
31. J. J. Koenderink and A. J. van Doorn, The international representation of solid shape with respect to vision, *Biol. Cybern.* **32**, 1979, 211–216.
32. A. F. Korn, Toward a symbolic representation of intensity changes in images, *IEEE Trans. Pattern Anal. Mach. Intell.* **10**, 1988, 610–625.
33. T. L. Kunii and L. Sun, Dynamic analysis-based human animation, in *CG International '90* (T. S. Chua and T. L. Kunii, Eds.), pp. 3–15. Springer-Verlag, Tokyo/Berlin/Heidelberg, 1990.
34. H.-J. Lee and Z. Chen, Determination of 3D human body postures from a single view, *Comput. Vision Graphics Image Process.* **30**, 1985, 148–168.
35. W. Leister and K. Rohr, *Voruntersuchungen von Bildsynthesesmethoden zur Analyse von Bildfolgen*, Interner Bericht Nr. 25/90, Universität Karlsruhe (TH), Fakultät für Informatik, September 1990.
36. M. K. Leung and Y. H. Yang, Human body motion segmentation in a complex scene, *Pattern Recognition* **20**(1), 1987, 55–64.
37. M. K. Leung and Y. H. Yang, A region based approach for human body motion analysis, *Pattern Recognit.* **20**(3), 1987, 321–339.
38. D. G. Lowe, The viewpoint consistency constraint, *Int. J. Comput. Vision* **1**, 1987, 57–72.
39. É.-J. Marey, *Movement*, Heine, London, 1895.
40. D. Marr, Representing visual information—A computational approach, in *Computer Vision Systems* (A. R. Hanson and E. M. Riseman, Eds.), Academic Press, New York/San Francisco/London, 1978.
41. D. Marr and H. K. Nishihara, Representation and recognition of the spatial organization of three-dimensional shapes, *Proc. R. Soc. London B* **200**, 1978, 269–294.
42. D. Marr and L. Vaina, *Representation and Recognition of the Movements of Shapes*, AI Memo 597, Artificial Intelligence Laboratory, MIT, Cambridge, MA, October 1980.
43. J. H. McIntosh and K. M. Mutch, Matching straight lines, *Comput. Vision Graphics Image Process.* **43**, 1988, 386–408.
44. M. P. Murray, Gait as a total pattern of movement, *Am. J. Phys. Med.* **46**(1), 1967, 290–332.
45. M. P. Murray, A. B. Drought, and R. C. Kory, Walking patterns of normal men, *J. Bone Joint Surgery* **46-A**(2), 1964, 335–360.
46. E. Muybridge, *Muybridge's Complete Human and Animal Locomotion. All 781 Plates from the 1887 "Animal Locomotion,"* Vol. 1, Dover, New York, 1979.
47. H.-H. Nagel, From image sequences towards conceptual descriptions, *Image Vision Comput.* **6**(2), 1988, 59–74.
48. H.-H. Nagel, A perspective on machine vision, submitted for publication.
49. B. Neumann and H.-J. Novak, Event models for recognition and natural language description of events in real-world image sequences, in *Proc. IJCAI-83, Karlsruhe/FRG, August 8.–12. 1983*, pp. 724–726.
50. B. Neumann and H.-J. Novak, NAOS: Ein System zur natürlichsprachlichen Beschreibung zeitveränderlicher Szenen, *Inform. Forsch. Entw.* **1**, 1986, 83–92.
51. W. M. Newman and R. F. Sproull, *Grundzüge der interaktiven Computergrafik*, McGraw-Hill, Hamburg, 1986.
52. J. O'Rourke and N. I. Badler, Model-based image analysis of human motion using constraint propagation, *IEEE Trans. Pattern Anal. Mach. Intell.* **2**(6), 1980, 522–536.
53. T. Pavlidis, *Algorithmen zur Grafik und Bildverarbeitung*, Verlag Heinz Heise, Hannover, 1990.
54. A. Pentland and B. Horowitz, Recovery of non-rigid motion and structure, *IEEE Trans. Pattern Anal. Mach. Intell.* **13**(6), 1991, 730–742.
55. D. I. Perrett, M. H. Harries, P. J. Benson, A. J. Chitty, and A. J. Mistlin, 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.), pp. 181–200, Wiley, Chichester/New York, 1990.
56. X. Pueyo and D. Tost, A survey of computer animation, *Comput. Graphics Forum* **7** (1988), 281–300.

57. R. F. Rashid, Towards a system for the interpretation of moving light displays, *IEEE Trans. Pattern Anal. Mach. Intell.* **2**(6) 1980, 574–581.

58. N. Rehfeld, *Auswertung von Stereobildfolgen mit Kantenmerkmalen*, Dissertation, Fakultät für Informatik der Universität Karlsruhe (TH), June 1990.

59. J. H. Rieger, The geometry of view space of opaque objects bounded by smooth surfaces, *Artif. Intell.* **44,** 1990, 1–40.

60. D. F. Rogers and J. A. Adams, *Mathematical Elements for Computer Graphics*, McGraw–Hill, New York, 1976.

61. K. Rohr, Auf dem Wege zu modellgestütztem Erkennen von bewegten nicht-starren Körpern in Realweltbildfolgen, in *11. DAGM—Symposium Mustererkennung, Oktober 2.–4., 1989, Hamburg*, Informatik-Fachberichte 219 (H. Burkhardt, K. H. Höhne, and B. Neumann, Eds.), pp. 324–328, Springer-Verlag, Berlin/Heidelberg, 1989.

62. K. Rohr and H.-H. Nagel, Modellgestützte Bestimmung des Bewegungszustandes von Fussgängern in Realweltbildfolgen, *12. DAGM—Symposium Mustererkennung, September 24.–26., 1990, Oberkochen-Aalen*, Informatik-Fachberichte 254 (R. E. Grosskopf, Ed.), pp. 52–58, Springer-Verlag, Berlin/Heidelberg, 1990.

63. K. Rohr and C. Schnörr, *Ein Ansatz zur Entwicklung eines Verfahrens zur automatischen Ermittlung der Trajektorien sich bewegender Objekte bei stationärer Kamera*, Hausbericht Nr. 10174, Fraunhofer-Institut für Informations- und Datenverarbeitung (IITB), Karlsruhe, June 1989.

64. J. R. J. Schirra, G. Bosch, C.-K. Sung, and G. Zimmermann, From image sequences to natural language: A first step toward automatic perception and description of motion, *Appl. Artif. Intell.* **1,** 1987, 287–305.

65. C. Schnörr, Determining optical flow for irregular domains by minimizing quadratic functionals of a certain class, *Int. J. Comput. Vision* **6**(1), 1991, 25–38.

66. H. R. Schwarz, *Numerische Mathematik*,Teubner, Stuttgart, 1986.

67. M. Sester and W. Förstner, Object location based on uncertain models, in *11. DAGM—Symposium Mustererkennung 1989, Hamburg, Oktober 2.–4., 1989*, Informatik-Fachberichte 219 (H. Burkhardt, K. H. Höhne, and B. Neumann, Eds.), pp. 457–464, Springer-Verlag, Berlin/Heidelberg, 1989.

68. M. Shiohara, T. Gotoh, Y. Nakagawa, and M. Yoshida, Surface Correspondence based on three-dimensional structure inference in animation images, in *10th Int. Conf. on Pattern Recognition, June 16–21, 1990, Atlantic City, New Jersey*, pp. 194–197.

69. D. Tost and X. Pueyo, Human body animation: A survey, *Visual Comput.* **3,** 1988, 254–264.

70. S. Tsuji, M. Osada, and M. Yachida, Tracking and segmentation of moving objects in dynamic line images, *IEEE Trans. Pattern Anal. Mach. Intell.* **2**(6), 1980, 516–522.

71. V. T. Tsukiyama and Y. Shirai, Detection of the movements of persons from a sparse sequence of TV images, *Pattern Recognit.* **18,** 1985, 207–213.

72. L. Vaina, From shapes and movements to objects and actions—Design constraints on the representation, *Synthese* **54,** 1983, 3–36.

73. J. A. Webb and J. K. Aggarwal, Structure from motion of rigid and jointed objects, *Artif. Intell.* **19,** 1982, 107–130.

74. W. Weber and E. Weber, *Mechanik der menschlichen Gehwerkeuge*, Dietrichsche Buchhandlung, *Göttingen*, 1836.

75. J. Wilhelms, Toward automatic motion control, *IEEE Comput. Graphics Appl.* **7**(4), 1987, 11–22.

76. J. Wilhelms, Using dynamic analysis for realistic animation of articulated bodies, *IEEE Comput. Graphics Appl.* **7**(6), 1987, 12–27.

77. M. Yamamoto and K. Koshikawa, Human motion analysis based on a robot arm model in *Proc. Computer Vision and Pattern Recognition, Lahaina, Maui, Hawaii, June 3–6, 1991*, pp. 664–665.

78. D. Zeltzer, Representation of complex animated figures, *Proc. Graphics Interface, May 1982, Toronto, Ontario*, pp. 205–211.

79. D. Zeltzer, Motor control techniques for figure animation, *IEEE Comput. Graphics Appl.* **2**(9), 1982, 53–59.