MOVING HUMANS RECOGNITION USING SPATIO-TEMPORAL MODELS*

Winfried Kinzel and Ernst D. Dickmanns

Institut für Systemdynamik und Flugmechanik
Universität der Bundeswehr München (UBM)
Werner-Heisenberg-Weg 39
D-8014 Neubiberg, Germany
phone +49-89/6004-3583
fax +49-89/6004-4092
ISPRS Commission V

ABSTRACT

A computational framework for dynamic scene analysis with respect to 3D-real-time tracking of human motion in typical road environment is presented. In contrast to the practice of many pattern recognition techniques which refrain from considering the underlying signal process, this approach utilizes a modified observer concept from control theory to keep track of changing image features and to aggregate them in a deductive rather than an inductive manner. A procedure of recursive estimation of limb states is derived for humans modelled as mechanical multibody system; it is supported and updated by feature based image sequence processing. For system development with versatile input signals and for an assessment of estimator performance an extensive animation tool has been designed. The proposed approach requires moderate computational power so that the complex recognition task may be accomplished in real-time in the near future. It promises less redundance and may serve as a simulation of perception in general.

Key Words: Human Motion, Computer Graphics, Image Analysis, Object Recognition, Feature Tracking, Recursive State Estimation.

1. INTRODUCTION

1.1. Motivation

The project objective towards automatically recognizing humans and their movements from image sequences was conceived in the scope of general object recognition for the purposes of autonomous road vehicle navigation and driver support in an Autobahn-like environment. It essentially means an extension of the successful work on road following that emerged at UBM over the last decade; on an object detection level, different modules will operate on cars, fixed obstacles, traffic signs etc. [Dickmanns 89, Dickmanns 91]. The motivation for detecting human beings within an autopilot (and driver assistence system) is justified by their need of special protection in road traffic as well as their ability to signal messages by gestures relevant for cooperative vehicle control. Other applications of the system presented here are imaginable in the fields of psychology and movement analysis [Proffitt 86], sports training or kinesiology and rehabilitation medicine where pathological gait patterns are to be analysed.

Activities with a similar scope on the topic of visual body motion detection are known [Hogg 83, Hogg 88, Rohr 90]. Various other studies have been pursued, e.g. emphasizing the occlusion problem, segmentation techniques [Leung 87] or detection by using markers.

1.2. System concept

Figure 1 gives an overview on the system concept. Instead of mere pattern matching of image information with previously stored shape or movement patterns the camera data are compared with internal assumptions of a spatio-temporal process model. This presumes motion states of an external object, i.e. a human body acting in a partially known environment (dynamic movement model). Thereby the static, unmoved state of a figure is treated as a special case of the figure being in motion. The process model maps image features according to its momentary internal state onto the image plane for comparison with data which are associated with the process going on in the real world. It also directs feature extraction after it has been itself initialized by image processing during bootstrapping including first model selection. As a figure's effigy is processed in three spatial dimensions and time, the computer model can be regarded as a means of recursive reconstruction of object states occurring in reality, like in the observer loop concept in standard control theory.
In contrast to this concept, however, a pure recognition task without being able to control the counterpart is at hand here. The process model has no capability to exert an influence on the outside obstacles it is imitating, the adjustment strategy is based merely on visual input. So the task of signal decoding, i.e. signal restoration from a known carrier

* This research project has been supported by the German Federal Ministry of Research and Technology (BMFT) and Daimler-Benz AG, Grant No. ITM 8900A (Prometheus PRO-ART).

and its classification into the initial symbolic meaning prevails. The figure movement parameters over time are considered characteristic for a conveyed message by gestures as part of nonverbal communication.

The idea is also to investigate how far methods of image generation and visualisation (computer graphics area) can positively influence the results that can be acchieved in the recognition field (computer vision area). In order to point out these separate paths, this paper splits into two major parts, a synthesis section and an analysis section. In the first one an animation tool for creating synthetic image sequences for the application domain is described. It uses three different model classes for the calculation of movement trajectories. The second part deals with both state estimation, which is explained in detail, and image analysis (as a prerequisite for obtaining measurement values).

1.3. System organization

The work that has been accomplished is reflected by four coordinated partial process modules all being implemented on a SGI graphics workstation as a simulation facility. Figure 2 gives an overview over the main signal flow. Control values and sychronizing data are exchanged by the UNIX message passing mechanism.

The first two processes serve as sources for pixel data:

An image sequence recall process is able to display various stored CCD-camera images onto a screen for purposes of testing image processing algorithms, checking their robustness, and for preparing applications outside the simulation environment.
Secondly, an animation process (described in the synsthesis section) synthetically generates image sequences with figures and background objects under user control. It meets the requirements for handy test signals, for having a performance measure for the estimation procedure, and for obtaining multiple training material for future classificators. Figure 3 gives an impression of used input images.
The next two modules belong to the analysis part of the recognition task and are described in detail later in this paper:
The image processing module is the mediator between pixel source and state estimation and picks up screen pixel data and searches for edges as basic image features. It sends information marking search areas to the imaging process and measurement values to the state estimation. This module is the basis for early recognition, proper initialization of state estimation and it supplies the necessary features for estimator update.
The estimation process contains the generic models, synchronizes and controls all other processes and may receive reference values from the animation side.

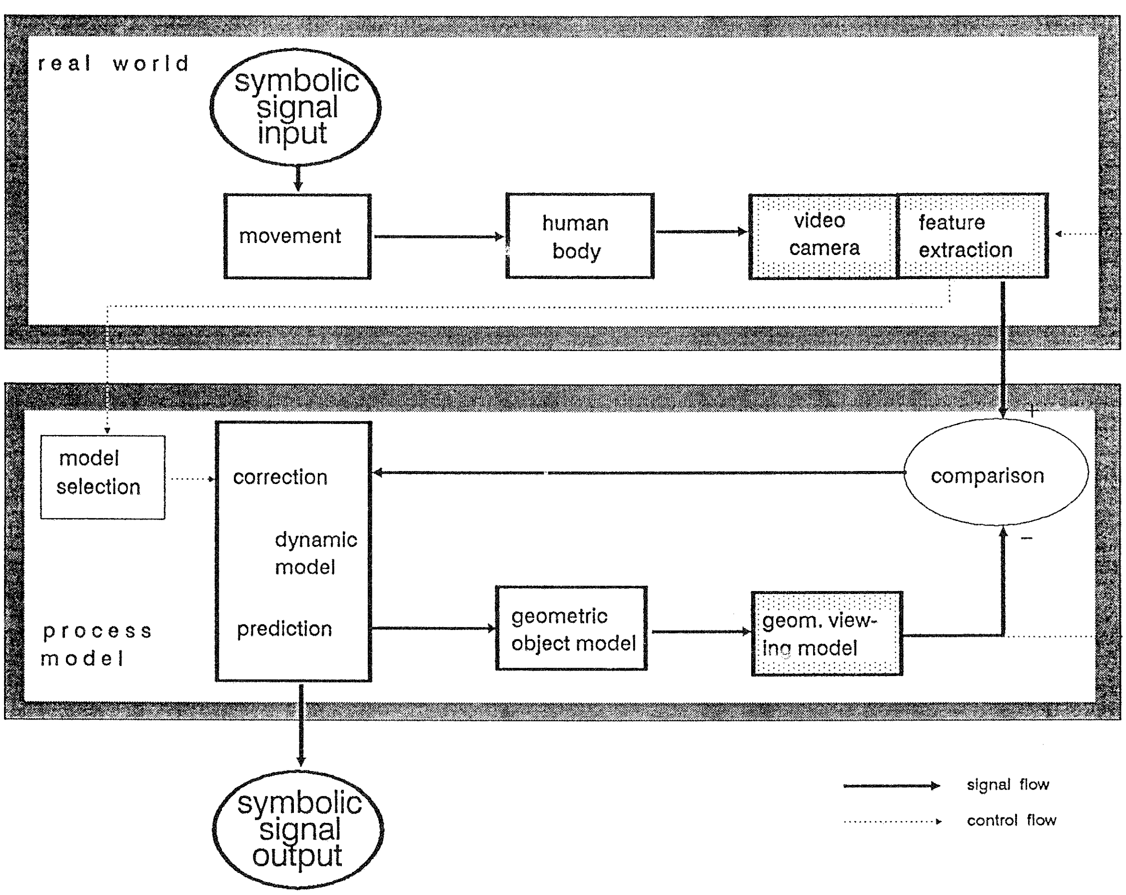

Fig. 1: Signal relation and basic idea

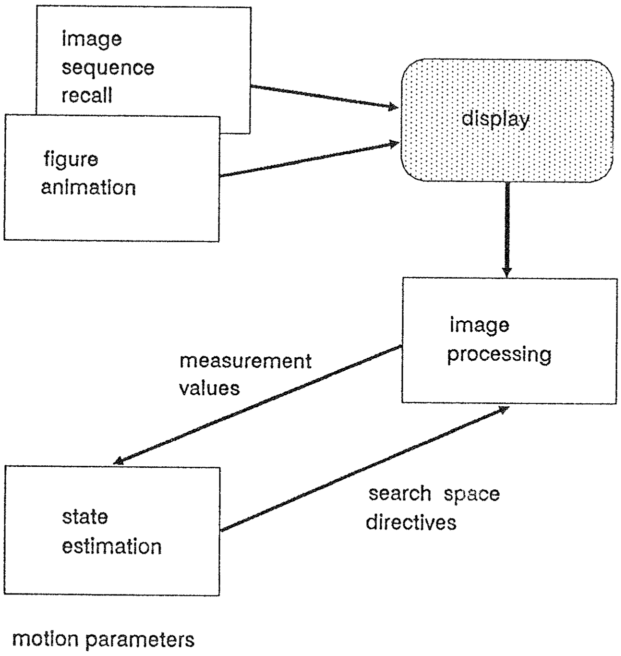

Fig. 2: Inter-process communication

2. SYNTHESIS

The animation tool, which provides flexible means for test signal generation, can be looked at in two aspects, a geometric and a movement generating one (see figure 5).

2.1 Geometry

The figures are regarded as articulated bodies and are geometrically represented by volume models and a limited number of limb joints giving a compromise between realistic shape and modelling effort. The model follows major ideas in the wide-spread field of computer graphics [Dooley 82, Thalmann 90]. As for real human mobility, the body has 107 degrees of freedom according to [Fischer 06], not taken into account the vertebrae and jaw. Nevertheless, in order to achieve a somewhat natural appearance and movability less effort is required, but the interdepenceny between good-looking movements and pleasing body representation has to be borne in mind. An adjustable figure repertoire with different shapes and scale factors has been created, also clothing artefacts are implemented. Figure 4 gives an impression of the degree of modelling detail. The limb ends are thoroughly modelled such that under no circumstances undesireable gaps at joints occur.

2.2. Movements

The actual issue of animation - animating objects, i.e. let them move - can be performed threefold with the existing tool:
by a kinetik approach (implemented for the particularly critical simulation of human gait),
by kinematic movement description (for less complex definition of cyclic movement of arms) and
by key-frame interpolation (for being able to uniformly animate a whole figure and to switch smoothly between single movement patterns).
Firstly, an exclusively kinetic approach is assigned for leg movement during normal walking only. Stiffed leg locomotion was enhanced by including knee flexion. Regarding the stance leg as an inverted pendulum and

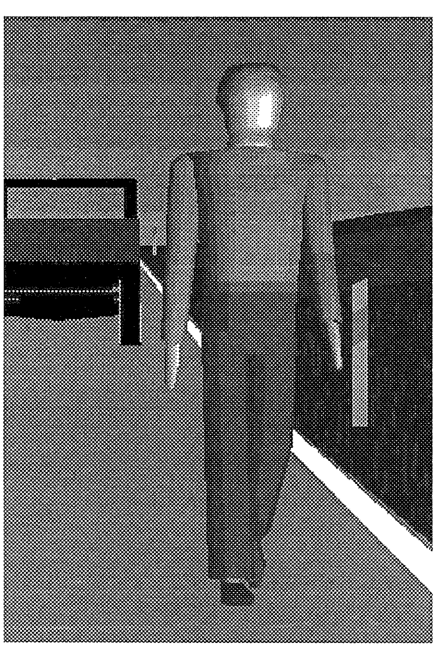

Fig. 3: Sample input image

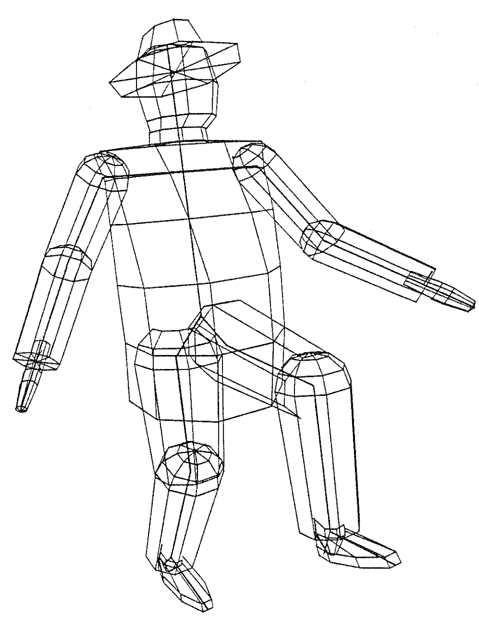

Fig. 4: Wire frame rendering of figure specimen

geometry	movements
• 3-D volume model • 15 single objects for limbs, torso, head forming figure with fixed topology	• dynamic (kinetic control) based on Lagrange equations and biomechanical studies
• 13 joints with degrees of freedom from 1-3 (resulting in a total of 22 d.o.f.) • uniformly described, flexibly generated partial objects with octagonal sectional area, adaptable shape and different size (morphology)	• procedural (kinematic description) with generator functions of sinusoidal or triangular shape
• anthropometric data for suitable body properties • material definitions and light model for realistic rendering	• key-framing using cubic splines for separate interpolation in space and time between chosen body poses derived from movement studies

Fig. 5: Figure model for animation

the swinging leg as a double pendulum system for the swing phase of walking [Alexander 84, Mochon 80] without external forces and friction, leads to 3 coupled differential equations yielding typical gait patterns. In order to solve these equations, which contain in great number sine- and consine-terms of angle differences as coefficients, linearizations around the reference state (figure standing upright) is done, giving dependent eq. 1a-c:

$$K_1 \ddot{\Theta} - C_2 \ddot{\Phi} - C_3 \ddot{\Psi} = W_1 \Theta \tag{1a}$$

$$K_2 \ddot{\Phi} + C_1 \ddot{\Psi} - C_2 \ddot{\Theta} = -W_2 \Phi \tag{1b}$$

$$K_3 \ddot{\Phi} + C_1 \ddot{\Psi} - C_2 \ddot{\Theta} = -W_3 \Psi \tag{1c}$$

with Θ, Φ, Ψ each representing the flexion angle of stance leg, thigh and shank of the swing leg relative to the vertical (see figure 6), and K_i, C_i, W_i as mechanical constants and body parameters.
Further approximations after comparing the sizes of remaining coefficients relative to each other, leads to a direct solution for three joint angles that still very well specify the leg motion during one swing cycle:

$$\Theta = A_{11}\, exp(T_1\, t) + A_{12}\, exp(-T_1\, t) \tag{2a}$$

$$\Phi = A_{21}\, cos(T_2\, t) + A_{22}\, sin(T_2\, t) \tag{2b}$$

$$\Psi = A_{31}\, cos(T_3\, t) + A_{32}\, sin(T_3\, t) \tag{2c}$$

with three individual time constants (T_1, T_2 and T_3). Constants A_{ij}, together with another parameter T_S (swing time) are to be calculated from seven initial and boundary (final) conditions that are imposed, i.e. geometric conditions (a-b), experimental results (c), suitable prescriptions (d-g)

$$l \cos \Theta_0 - l_1 \cos \Phi_0 - l_2 \cos \Psi_0 - d \sin \alpha = 0 \tag{3a}$$

$$l \sin \Theta_0 + l_1 \sin \Phi_0 + l_2 \sin \Psi_0 - d \cos \alpha = s_L - d \tag{3b}$$

$$l \sin \Phi(T_s) - \sin \Theta_0 = 0.9\, d \tag{3c}$$

$$\Psi(T_s) = \Phi(T_s) \tag{3d}$$

$$\Phi(T_s) = \Theta(T_s) = -arcsin\left(\frac{s_L}{2l}\right) \tag{3e}$$

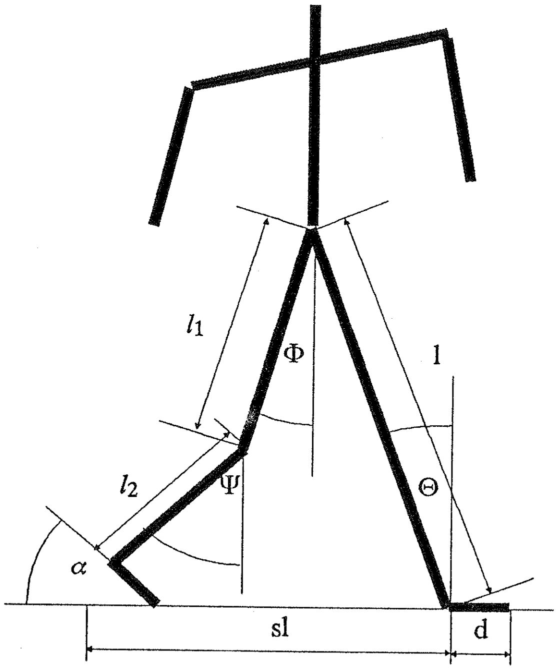

Fig.6: Leg skeleton geometry

$$s_L = const \tag{3f}$$

$$\Theta_0 = const \tag{3g}$$

with l, l_1, l_2, d, s_L length of whole leg, thigh, shank, foot and step length, α angle between foot and the horizontal, and Θ_0, Φ_0, Ψ_0 angles at the beginning of swing phase (see figure 6).

The double support phase is modelled by extrapolation of the swing phase backwards in the direction of the negative time axis and linearly incrementing the angle between ground and toeing-off foot during double support time. By varying mechanical constants or modifying the boundary conditions - where applicable - different gait patterns can be produced.
Although much effort has gone into biomechanical analyses [Schmid-Schönbein 86, Baumann 86] and dynamic gait descriptions are applied widely in the field of computer animation [Bruderlin 89, Wilhelms 87] and also in robotics, particularly in designing walking machines [Miura 84], fully kinetic modelling of all figure movements continues to become rapidly complicated. Numerous material parameters, external and reactive forces constraining the possible movement range are to be assessed. In fact, the difficulty of movement definition is shifted to a wise choice of generalized forces the knowledge of which is sparse.

Therefore, a second method, the procedural description of movements, is implemented mainly for waving arms in a simple manner. For upper limb animation, generator functions produce joint angle curves of sinusoidal and triangular shapes. In addition the leg movement of a cyclist riding a bike has been modelled employing inverse kinematics equations. The movements of this method appear less sophisticated but are realistic enough to serve their purposes.

In order to unify all limb and body movements and to be able to switch smoothly between single movement patterns, a third approach utilizes fairly popular key-frame interpolation technique together with cinematographic movement studies [Muybridge 55]. Here, a movement pattern is defined by designing a set of n body poses

$$S(x) = \{S_0, S_1, \dots, S_{n-1}\} \text{ with } x \in [x_0, x_n]. \tag{4a}$$

Using the Hermite form of cubic parametric splines (as a piecewise approximation of cubic polynomial functions)

$$S_i(x) = a_i + b_i(x-x_i) + c_i(x-x_i)^2 + d_i(x-x_i)^3, \tag{4b}$$

an interpolation between joint angle control points is performed. For calculating a_i, b_i, c_i, d_i the control point position and either first or second derivatives at the two interval boundaries are utilized (from interval to interval continuity of theses values is provided). For joint movements of the model this means a direct definition of angular velocity or acceleraction which allows to account for physical boundary conditions in the context of movement definition.
Subsequently, the movement description is factored into shape and motion; a separate manipulation of positional and time splines according to [Steketee 85] gives the user various possibilities to alter the movement appearance and to create complicated patterns. These also can be blended with partial movements generated by both methods described above. Figure 7 depicts a typical movement pass defined by 9 key-frames.

3. ANALYSIS

3.1. Image processing

For extracting image information edges are the feature primitives. They are detected by differential operations performed in limited search areas on a grey scale intensity image, alternatively with homogenous background or with distracting edges. As edge detector one mask out of a set of five individual gradient masks - 5x5-picture elements (pixels) each - is applied in two orthogonal search directions. The search paths are a few 10 pixels long yielding several correlation extrema in the image area in question.

As a human figure in motion exhibits a non-rigid shape a meaningful local connection of single measurement values (in a gestalt-sense of psychology) within a limb finder supplies orientation, position, length and variance (as a confidence value) for straight line segment approximations of limb parts. Criteria for connecting single data values for a least squares fit are their absolute and relative sizes, their standard deviation estimated from a correlation function discussion, and the relationship with expected line properties assumed by the dynamic process model. Especially vertical lines as contour elements and double edges are searched for. The successful operation of the image analysis can be monitored by the reconstruction of a stick figure (see figure 7). Apart from the two vertical lines, short lines

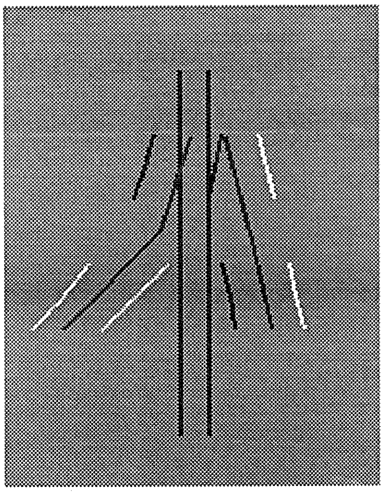

Fig. 7: Reconstructed limb orientations

representing edges of thighs and shanks can be seen; they are used to calculate the leg skeletons.

In the initialization phase, the operation of image analysis with the goal of forming the inductive step towards an initial object hypothesis still needs some improvement. Provided that a valid object model has been found, feature tracking subsequently can be supported by the process model that helps in systematically deriving search regions and test features from one hypothesis. The measured features either verify the supposition, and the model will be updated, or falsify it with the rejected model being replaced by another one. In the future, the search area will be confined by road recognition because only in the vicinity of lanes objects are presumed to be obstacles.

3.2. State estimation

The second part of the analysis task resides in the estimation process incorporating the two major steps of recursive estimation [Kalman 60, Maybeck 79]: correction and prediction (see figure 8). Estimation is supported and updated by pre-interpreted features

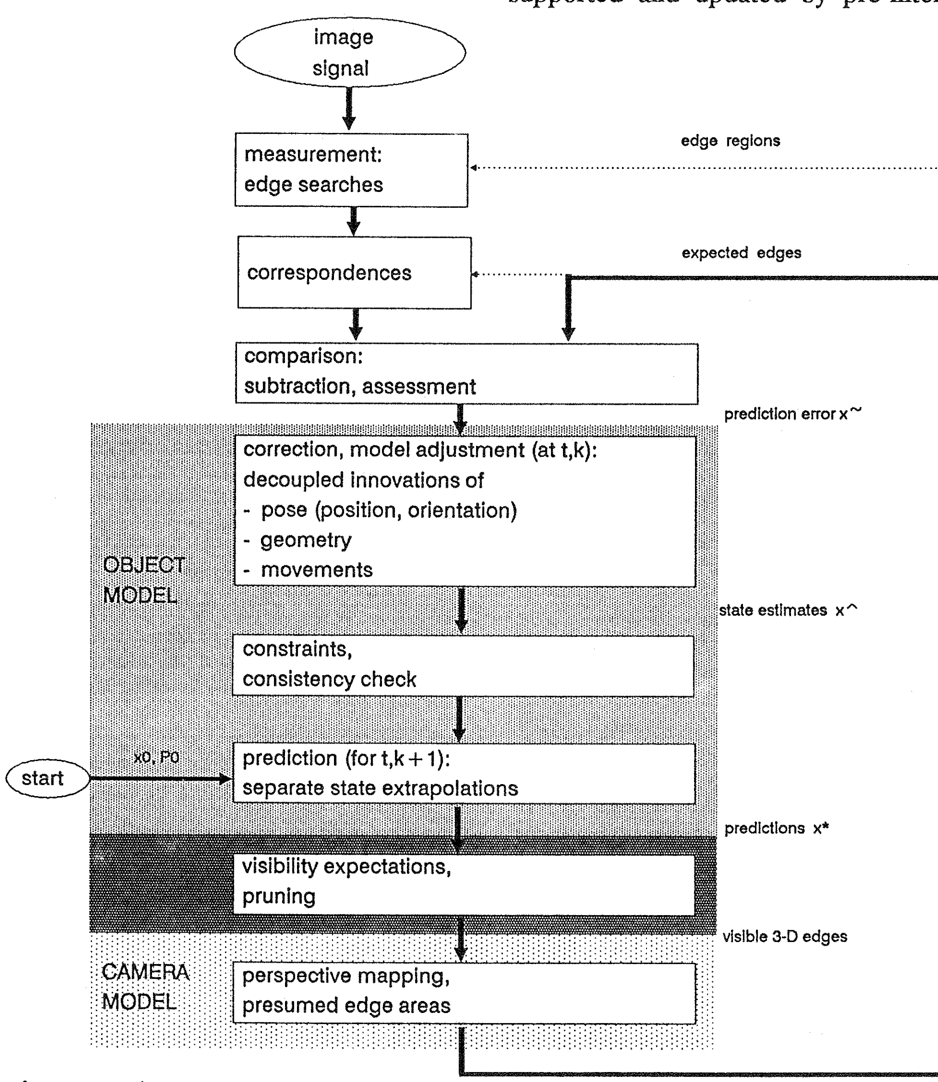

Fig. 8: Estimation procedure

collected from image processing, i.e measured edges. These noisy measurement values are filtered according to the presently supposed set of state variables. It is distinguished between the rigid motion of the overall figure (inertial 3-D pose), its geometic properties (3-D shape), and the movement of the articulated body (multiple limb motion). Individual estimators are in operation simultanously for determining the best fitting state variables belonging to these three classes; all together maintain geometric coherence of the semi-independent partial volume models, and temporal coherence of motion and movement variables. So, in a deductive step expected edge properties derived from an instantiated generic model (having become a specific model for a time instant) are used to direct the image measurement process and to assess measurement quality.

The essential information that the models are to deliver constists of the motion states in all joints (joint angles) in order to solve the pattern recognition task. The figure movement is regarded as fully recoverable from theses state variables as psychological recognition experiments [Johansson 73] show that joint positions traced over time are sufficient for that purpose.

To follow one exemplary estimation cycle in discrete time (see figure 8), one set of knee state varibles, flexion angle and angular velocity, is propagated from time instance k to k + 1 through the following transition equation (prediction block in figure 8)

$$\begin{pmatrix} \theta \\ \dot{\theta} \end{pmatrix}_{k+1} = \begin{pmatrix} \cos\omega\tau & \sin\omega\tau/\omega \\ -\omega\sin\omega\tau & \cos\omega\tau \end{pmatrix} \begin{pmatrix} \theta \\ \dot{\theta} \end{pmatrix}_{k} \qquad (5)$$

where ω is the cycle frequency and τ the sample time. This transition matrix is well suited to predict cyclic changes. Accepting other state variables to be valid that influence the image features (here: shank skeleton line segment properties), an approximate perspective mapping equation for edge orientation β for the shank,

$$\beta = arctan \left[\frac{K_Y}{K_Z} \sin(\varphi_F) \frac{sin(\theta_P + \theta_K)}{cos(\theta_P + \theta_K)} \right] \qquad (6)$$

is applied, with K_Y ,K_Z camera parameters, φ_F , θ_P , θ_K jaw angle of whole figure, flexion relativ angles in pelvis and knee joints, respectively.

Figure orientation is expected to be greater than 0 deg. in this case (90 deg. is walking direction perpendicular to the viewer). When deviations to the real values increase considerably due to decrease of figure orientation from 90 deg. the knee angles are not observable any more and no innovation takes place. Provided that corresponding image features are found, mapped edge and measured edge are compared and the internal model is adjusted. After application of constraints which check whether the recent values are reasonable or not, i.e. whether they are within the physiologically possible ranges or are consistent with movement semantics, the next estimation cycle begins.

Figure 9 shows curves of estimated pelvis and knee flexion angles recovered from the same animation sequence (broken lines) against the reference values. The pelvis estimates anticipate the true values about one cycle in advance. The knee angles have not been limited to the 0°-level here. The estimation is robust also when invalid measurement values fail to up-date the state estimates, e.g. in the interval from 2.64 - 3.12 sec. for the knee angle. All curves quickly converge towards their true values in the beginning.

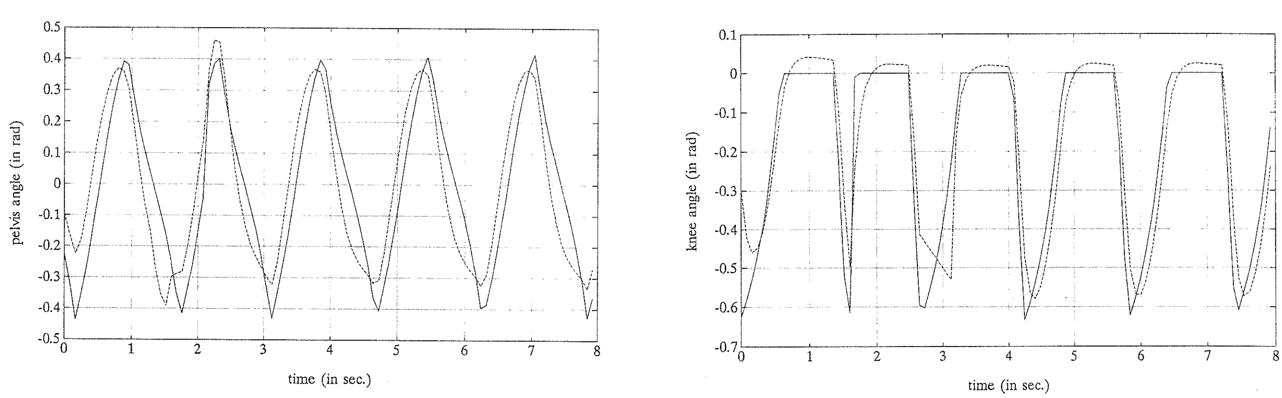

Fig. 9: Curves of estimated pelvis and knee angles

4. CONCLUSION

A simulation environment has been created to achieve two aims: first to allow for a convenient developement of obstacle detection algorithms for practical real-time operation with 'VaMoRs', the autonomous vehicle of UBM; and secondly to study the practicabiliy of the design of fully automated recognition and classification of gait patterns (lower extremities, locomotion) and waving movements (upper limbs, signalling of messages). Not applying sophisticated image feature detectors but adequatly combining single measurement values under model control has been attempted. This approach also reflects both the synthesis and analysis aspects of computer vision.

Preliminary results concerning the detection of lower limbs show the possiblity of tracking relevant image features and being able to reconstruct selected movement parameters for future decision of class membership. The presented rudimentary study with its computer realization for a specific application domain may also evolve to a contribution to the endeavour of revealing principles of visual perception through analysis by synthesis.

5. REFERENCES

[Alexander 84] Alexander, R. McN.: The gaits of bipedal and quadrupedal animals. Journal of Robotics Research, Vol.3, No.2, 1984.

[Baumann 86] Baumann, W.; Schwirtz, A.; Groß, V.: Biomechanik des Kurzstreckenlaufs. In: Biomechanik der Sportarten, Bd.1 Leichtathletik, pp.1-15. Verl. Ferdinand Enke, Stuttgart, 1986.

[Bruderlin 89] Bruderlin, Armin; Calvert, Thomas W.: Goal-directed, dynamic animation of human walking. Computer Graphics, Vol.23, No.3, ACM SIGGRAPH '89, Boston, July 1989.

[Dickmanns 89] Dickmanns, Ernst D.: Active vision through prediction-error minimization. NATO-Advanced Study Institute, Maratea, Italy, July 1989.

[Dickmanns 91] Dickmanns, Ernst D.; Christians, Th.: Relative 3D-state estimation for autonomous visual guidance of road vehicles. Robotics and Autonomous Systems, Vol.7, pp.113-123, 1991.

[Dooley 82] Dooley, Marianne: Anthropometric Modeling programs - survey. IEEE Computer Graphics and Applications, vol.2, no.9, pp.17-25, November 1982.

[Fischer 09] Fischer, Otto: Theoretische Grundlagen für eine Mechanik der lebenden Körper. Lehrbücher der mathematischen Wissenschaften XXII, Teubner, Leipzig, 1906.

[Hogg 83] Hogg, David C.: Model-based vision - a program to see a walking person. Image and Vision Computing, vol.1, no.1, pp.5-20, Feb. 1983.

[Hogg 88] Hogg, David C.: Finding a known object using a generate and test strategy. In Ian Page (ed.): Parallel architectures and computer vision. Clarendon Press, Oxford, pp.119-132, 1988.

[Johansson 73] Johansson, Gunnar: Visual perception of biological motion and a model for its analysis. Perception and Psychophysics, Vol.14, No.2, pp.201-211, 1973.

[Kalman 60] Kalman, R. E.: A new approach to linear filtering and prediction problems. Transactions of ASME, Series D, Journal of Basic Engineering Vol.82, pp.35-45, March 1960.

[Leung 87] Leung, Maylor K.; Yang, Yee-Hong: Human body motion segmentation in a complex scene. Pattern Recognition, Vol.20, No.1, pp.55-64, Pergamon, 1987.

[Maybeck 79] Maybeck, Peter S.: Stochastic models, estimation, and control. Vol.1. Mathematics in science and engineering 141-1, Academic Press, 1979.

[Miura 84] Miura, Hirofumi; Shimoyama, Isao: Dynamic walk of a biped. Jour. of Robotics Research, Vol.3, No.2, 1984.

[Mochon 80] Mochon, Simon; MacMahon, Thomas A.: Ballistic walking. Journal of Biomechnics. Vol.13, pp.49-57, 1980.

[Muybridge 55] Muybridge, Eadweard: Human Figure in Motion. Dover-Publ., New York, 1955.

[Proffitt 86] Proffitt, Dennis R.; Kaiser, Mary R.: Applications of computer-graphics animation for motion-perception research. NASA Technical Memorandum 88335, Ames Research Center, Moffett Field, CA. Sept. 1986.

[Rohr 90] Rohr, K.; Nagel, Hans-Hellmut: Modellgestützte Bestimmung des Bewegungszustandes von Fußgängern in Realweltbildfolgen. In R.E. Großkopf (ed.): Mustererkennung '90, Proc. 12. DAGM-Symposium, pp.52-58, 1990.

[Schmid-Schönbein 86] Schmid-Schönbein, G. W.; et. al. (eds.): Frontiers in biomechanics. Springer, New York, 1986.

[Steketee 85] Steketee, Scott N.; Badler, Norman I.: Parametric keyframe interpolation incorporating kinetic adjustment and phrasing control. ACM SIGGRAPH '85, Vol.19, No.3, pp.255-262, 1985.

[Thalmann 90] Thalmann, Daniel; Magnenat-Thalmann, Nadia (eds.): Computer animation '90. Second Int. Workshop on Computer Animation, Geneva, Apr. 25-27. Springer, Tokyo, 1990.

[Wilhelms 87] Wilhelms, Jane: Using dynamic analysis for realistic animation of articulated bodies. IEEE Computer Graphics and Applications, pp.12-27, June 1987.