Reconstruction of 3D Human Movement from Single View Images using Kalman Filter and Fuzzy Reasoning

Kenji Amaya Yuji Hara Shigeru Aoki
Dept.of Mechanical and Environmental Informatics
Tokyo Institute of Technology.
2-12-1 Ookayama, Meguro-ku, Tokyo 152,Japan
e-mail:kamaya@a.mei.titech.ac.jp
phone:+81-3-5734-2856
fax:+81-3-3729-0628

Abstract

A system which reconstructs 3D human movement from a stream of 2D input images is developed. The system finds the 3D shape and motion of the human body by fitting a simple skeleton model to the notable regions found in the 2D image. Surface markers which were robust for obstruction were applied. The fitting is done using a nonlinear optimization approach. This optimization is essentially ill-conditioned because depth information can not be obtained from 2D image. In order to overcome this problem, Kalman filter using Fuzzy reasoning which takes account of following a priori information is applied: 1 The versatile skeleton model exhibiting a human body. 2 Postures at each keyframe for the type of human movement. We reconstructed human movement from a stream of Video images by this method in order to demonstrate its applicability.

1 Introduction

3D reconstruction of human motion poses a challenging vision problem whose solution has a great practical interest. For instance, in the field of computer interface, computer animation, and virtual reality, a motion tracking is an indispensable technology. Various methods have been developed for this purpose. Among these methods, the video based systems have advantages because of the economical aspects and the non-contact measurement. [1] [2] [3] [4]

In this paper, the system which reconstructs 3D human movement from a stream of the 2D input images was developed. This problem is essentially ill-conditioned because depth information can not be obtained from 2D images. To determine the final feasible body structures modified Kalman filter using fuzzy reasoning was applied.

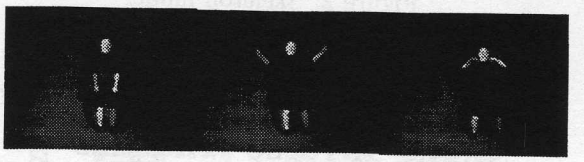


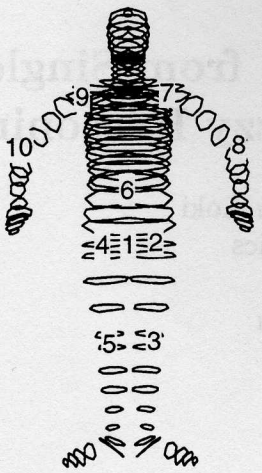
Figure 1: Full color images with some surface markers

Authors had developed the motion capture system which uses pointwise markers. [6] However, the pointwise markers are easy to be occluded by human motion. In order to overcome this problem, the surface-wise maker (See Fig.1) which is robust for occlusion problem was developed.

2 Modelization of Human Body

Since our system uses 2D images as inputs, the depth information is limited. In order to overcome this shortage of the depth information, we introduced 3D human model as a priori information.

Actual human body has over 80 degree of freedoms. For simplicity, our human body model consists of 10 major joints and its degree of freedom is 34. This 34 parameters are denoted as a 34 dimensional posture vector $\boldsymbol{\theta}_t$, where the subscript t is a time. Figure 2 shows our human model. Each joint has its own degree of freedom as indicated in figure 2, and each size of body segment is given.



Parameter θ	D.O.F.
1. Body center	6
2. Left hip	3
3. Left knee	1
4. Right hip	3
5. Right knee	1
6. Lumber	3
7. Left shoulder	3
8. Left elbow	1
9. Right shoulder	3
10. Right elbow	1
Total dimension =	25

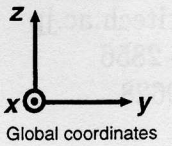


Figure 2: Skeleton model

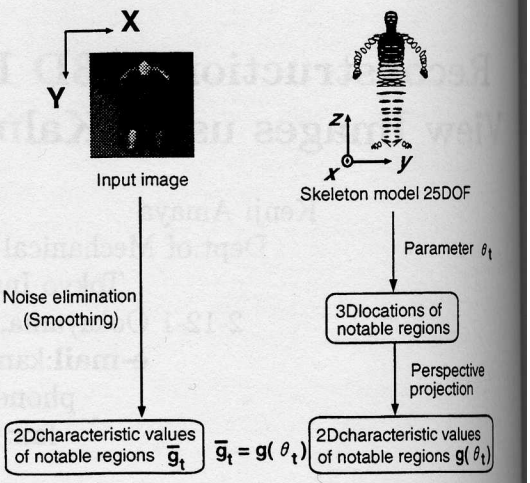


Figure 3: The procedure of the reconstruction

3 Reconstruction Procedure Using Inverse Analysis

3.1 Acquisition of Characteristics $\bar{g}_t$ from Input Image

Figure 1 shows the sample of employed source image. Five parts of body, e.g., the trunk, arms and lower legs, are given by different colors. The images were smoothed and binarized to eliminate noise and to extract the 5 marker areas[7].

Zero-order, primary-order and secondary-order moments of each marker area$S$ are calculated with following equation:

$$m_n = m_{(i,j)} = \int_S x^i y^j dS \qquad (n = i + j) \qquad (1)$$

where, $i, j = 0, 1, 2$ and in case of secondary-order moments calculation is performed under local coordinates which origin is at the center of S,(S_X, S_Y). We defined matrix A as the following.

$$A = \frac{1}{m_{(0,0)}} \begin{pmatrix} m_{(2,0)} & m_{(1,1)} \\ m_{(1,1)} & m_{(0,2)} \end{pmatrix} \qquad (2)$$

The characteristics (center,direction, aspect ratio) of marker area S in Fig.4 can be defined as follows:

$$S_X = \frac{m_{(1,0)}}{m_{(0,0)}} \qquad (3)$$

$$S_Y = \frac{m_{(0,1)}}{m_{(0,0)}} \qquad (4)$$

$$S_\theta = \arctan(\frac{a_Y}{a_X}) \qquad (5)$$

$$S_\lambda = \frac{\lambda_1}{\lambda_2} \qquad (6)$$

where, λ_1, λ_2 are eigen values of matrix A $(\lambda_1 > \lambda_2)$ and $\boldsymbol{a} = (a_X, a_Y)$ is eigen vector which corresponds λ_2. The vector $\boldsymbol{a}$ directs the the major axis of area S.

The 20 components of characteristic vector $\bar{g}_t$ consists of S_X, S_Y, S_θ and S_λ of all 5 marker area.

3.2 Calculation of characteristics $g(\theta_t)$ by Model

Using direct kinematics, locations of all body parts can be calculated in 3D space from the posture vector $\boldsymbol{\theta_t}$. Applying a knowledge of computer graphics analysis, 2D coordinates of body parts on screen $\boldsymbol{g(\theta_t)}$ can be obtained[8].

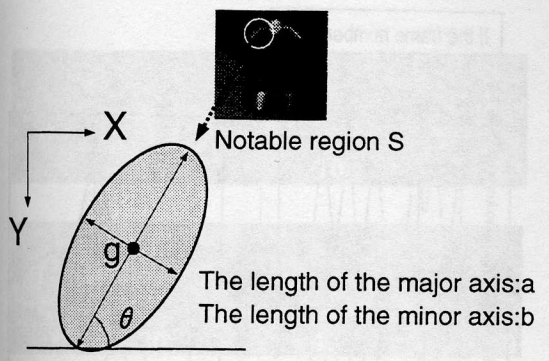
The characteristics of marker areas S are expressed with 2D coordinates of both ends of the body parts as the function $\boldsymbol{g(\theta_t)}$. We denote that $\boldsymbol{g(\theta_t)}$ is the function of the posture vector $\boldsymbol{\theta_t}$.

3.3 Model Fitting

The posture vector $\boldsymbol{\theta_t}$ for each t can be considered as the observation vector for the following non-linear observation equation. (See Fig.3)

$$\bar{g}_t = \boldsymbol{g(\theta_t)} + \boldsymbol{v_g} \qquad (7)$$

where, $\bar{g}_t$ and $\boldsymbol{g(\theta_t)}$ are 20-dimensional vector and $\boldsymbol{v_g}$ is a 20-dimensional vector corresponding to Gaussian noise which mean is 0, deviation depends on each components.



Characteristic values in notable region

1.X coordinates of the center point g : S_X
2.Y coordinates of the center point g : S_Y
3.Slope angle of the major axis : S_θ
4.Aspect ratio (=a/b) : S_λ

Figure 4: The characteristic values of the notable region

4 Application of A Priori Information

Since Eqn.(7) doesn't carry any depth information, this problem is ill-conditioned[5]. In order to overcome this problem, we consider to implement a priori information about a target action. For example, in case capturing walking motion, we teach the system "This motion is walking motion". Practically,a keyframe library is constructed in advance [12]. The keyframe library is consists of assumed posture vectors $\boldsymbol{\theta_k}$ at keyframes. Each keyframe locates at the frame which the square sum of the velocity of markers S in images as shown in Fig.5. The square sum V_{sum} is defined as the following.

$$V_{sum} = \sum_{i=1}^{5} \sqrt{(\dot{S}_{Xti})^2 + (\dot{S}_{Yti})^2} \qquad (8)$$

where, the subscript i indicates each marker area and t is discrete time. For example, 10 circled frames in Fig.5 are selected out of 118 frames. The keyframe library is produced using computer animation design software Life Form [9]. A priori posture vector $\boldsymbol{\theta_{p,t}}$ can be assumed by interpolating the keyframes.

By using a priori posture vector $\boldsymbol{\theta_{p,t}}$ The depth locations of major 13 points (see Fig.6) on human model $\bar{\boldsymbol{d}}_t$ can be calculated.

On the other hand this depth locations of major 13 points can be expressed with a function $\boldsymbol{d(\theta_t)}$. The function $\boldsymbol{d(\theta_t)}$ can be expressed applying the knowledge of direct kinematics and computer graphics with the same idea of expressing $\boldsymbol{g(\theta_t)}$.

We take account the difference between a priori value and exact value as a noise as the following:

$$\bar{\boldsymbol{d}}_t = \boldsymbol{d(\theta_t)} + \boldsymbol{v_d} \qquad (9)$$

where, $\boldsymbol{v_d}$ is a Gaussian noise which mean is 0 and variance is $10^5[mm^2]$.

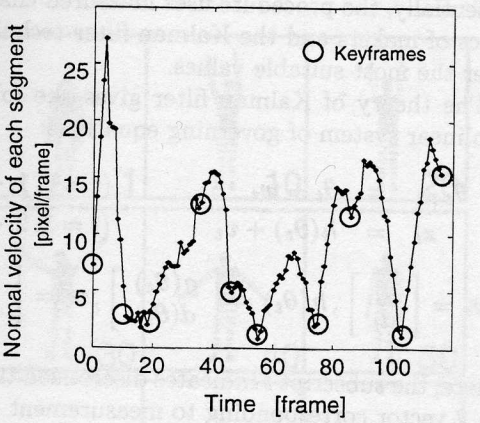


Figure 5: Normal velocity of each segments

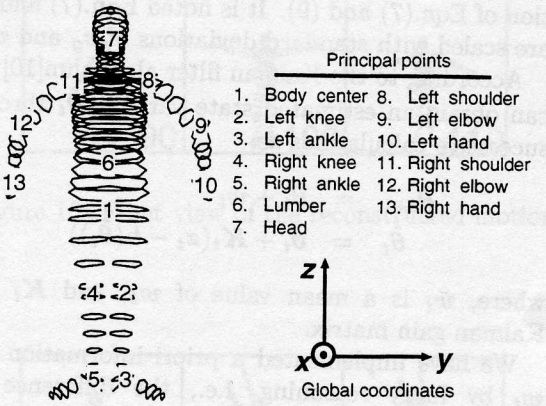


Figure 6: Principal points of the skeleton model

5 Application of Kalman filter using Fuzzy Logic

Here, we propose an effective method based on the inverse analysis to estimate the posture vectors $\boldsymbol{\theta_t}$.

Essentially, the procedure uses measured characteristics of makers and the Kalman filter technique to infer the most suitable values.

The theory of Kalman filter gives the following nonlinear system of governing equations:

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \boldsymbol{w}_t \quad (t = 0, 1, \cdots) \quad (10)$$

$$\boldsymbol{z}_t = \boldsymbol{h}(\boldsymbol{\theta}_t) + \boldsymbol{v}_t \quad (t = 1, 2, \cdots) \quad (11)$$

$$\boldsymbol{z}_t = \begin{bmatrix} \boldsymbol{g}_t \\ \boldsymbol{d}_t \end{bmatrix}, \boldsymbol{h}(\boldsymbol{\theta}_t) = \begin{bmatrix} \boldsymbol{g}(\boldsymbol{\theta}_t) \\ \boldsymbol{d}(\boldsymbol{\theta}_t) \end{bmatrix}, \boldsymbol{v}_t = \begin{bmatrix} \boldsymbol{v}_g \\ \boldsymbol{v}_d \end{bmatrix} \quad (12)$$

where, the subscript t indicates discretized time and $\boldsymbol{w}_t$ 2 vector corresponding to measurement error.

In general, Eqn.(10) is known as the state transition equation, Eqn.(11) is the observation equation and $\boldsymbol{v}_t$,$\boldsymbol{w}_t$, $\boldsymbol{\theta}_t$, $\boldsymbol{z}_t$ are called the random variable vectors. Equation (11) is formed by the combination of Eqn.(7) and (9). It is noted Eqn.(7) and (9) are scaled with standard deviations of $\boldsymbol{v}_g$ and $\boldsymbol{v}_d$.

According to the Kalman filter algorithm[10] one can obtain an estimated state variable $\hat{\boldsymbol{\theta}}_t$ through successive calculations as;

$$\bar{\boldsymbol{\theta}}_{t+1} = \hat{\boldsymbol{\theta}}_t + \bar{\boldsymbol{w}}_t \quad (13)$$

$$\hat{\boldsymbol{\theta}}_t = \bar{\boldsymbol{\theta}}_t + \boldsymbol{K}_t(\boldsymbol{z}_t - \boldsymbol{h}(\bar{\boldsymbol{\theta}}_t)) \quad (14)$$

where, $\bar{\boldsymbol{w}}_t$ is a mean value of $\boldsymbol{w}_t$, and $\boldsymbol{K}_t$ is a Kalman gain matrix.

We have implemented a priori information into $\boldsymbol{w}_t$ by fuzzy reasoning, i.e., the difference between a priori posture vector $\boldsymbol{\theta}_{p,t}$ and the posture vector$\bar{\boldsymbol{\theta}}_t$ increases as the current frame is far from a keyframe.

If "The current frame locates near a keyframe"then "The posture vector $\bar{\boldsymbol{\theta}}_{t+1}$ is like a priori posture vector $\boldsymbol{\theta}_{p,t+1}$ " ($\bar{\boldsymbol{w}}_t \simeq \boldsymbol{\theta}_{p,t+1}$ - $\hat{\boldsymbol{\theta}}_t$)

If "The current frame is far from a key frame" then "The vector $\bar{\boldsymbol{w}}_t$ is like the gradient of $\boldsymbol{\theta}_{p,t}$ " ($\bar{\boldsymbol{w}}_t \simeq \boldsymbol{\theta}_{p,t+1}$ - $\boldsymbol{\theta}_{p,t}$)

Membership functions based on above two fuzzy rules are shown in Fig.7.[11] The first step is to take the inputs and determine the degree to which they belong to each of the fuzzy appropriate fuzzy sets. The input is a current frame number and output is a fuzzy degree of membership (in this case "near" or "far"). Once the inputs have been fuzzified, we know the degree to which each part of the antecedent has been satisfied for each rule. In order to unify the outputs of each rule by joining the parallel threads, aggregation is performed. In Fig.7, all two rules have been placed together to show how the out put of each rule is combined by logical sum (OR). Since, this output membership function f_m can be regarded as a statistical character of $\boldsymbol{w}_t$, we

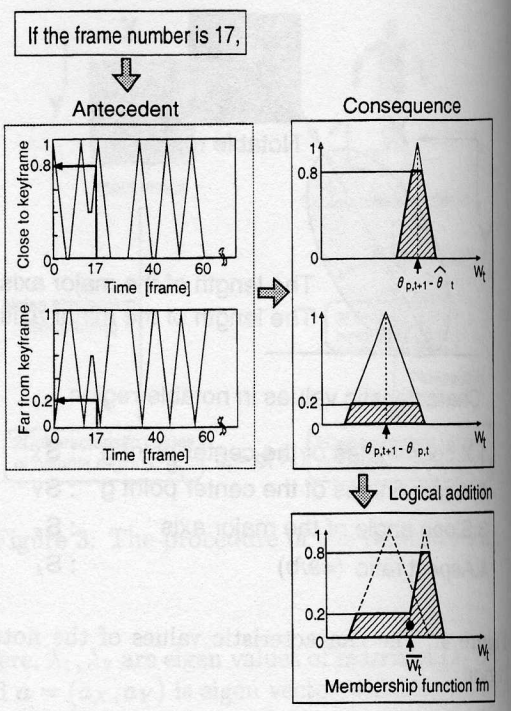


Figure 7: Fuzzy reasoning

considered the mean and the covariance of f_m for the calculation of Kalman gain $\boldsymbol{K}_t$ as follows:

$$\boldsymbol{K}_t = \boldsymbol{P}_t\boldsymbol{H}_t^T\boldsymbol{R}_t^{-1} \quad (15)$$

$$\boldsymbol{P}_t = (\boldsymbol{M}_t^{-1} + \boldsymbol{H}_t^T\boldsymbol{R}_t^{-1}\boldsymbol{H}_t)^{-1} \quad (16)$$

$$= \boldsymbol{M}_t - \boldsymbol{M}_t\boldsymbol{H}_t^T(\boldsymbol{H}_t\boldsymbol{M}_t\boldsymbol{H}_t^T + \boldsymbol{R}_t)^{-1}\boldsymbol{H}_t\boldsymbol{M}_t$$

$$\boldsymbol{M}_{t+1} = \boldsymbol{P}_t + \boldsymbol{Q}_t \quad (17)$$

$$\boldsymbol{H}_t = \frac{\partial \boldsymbol{h}(\boldsymbol{\theta}_t)}{\partial \boldsymbol{\theta}_t} \quad (18)$$

where $\boldsymbol{Q}_t$ and $\boldsymbol{R}_t$ are the matrices of covariance of $\boldsymbol{w}_t$ and $\boldsymbol{v}_t$ respectively.

6 Example Reconstruction

Figure 8 shows the sample of employed source images. Each number which attached at bottom right indicates the frame number. We used a camcorder (SONY-DCR-VX1000) to record a subject performs "aerobic exercise" for 4 seconds.

The recorded image data were input into workstation. Its frame rate was $30 frames/sec$ and its resolution was $320 \times 240 pixels$. Keyframes were selected based on the data shown in Fig.5. Figure 7 shows the employed membership functions. Figure 10 and 11 show the front and the side view of the reconstructed motion respectively.

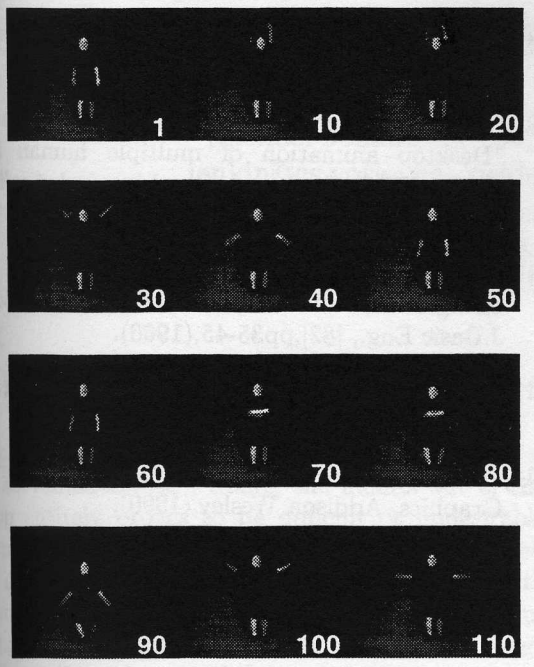


Figure 8: Input images

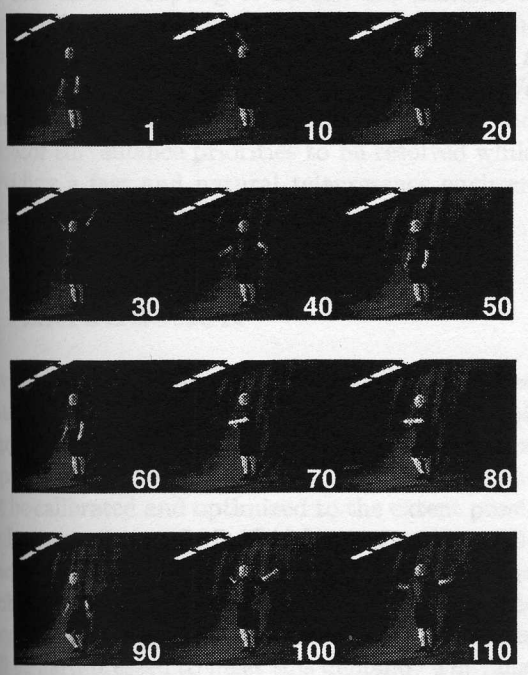


Figure 9: Reference images

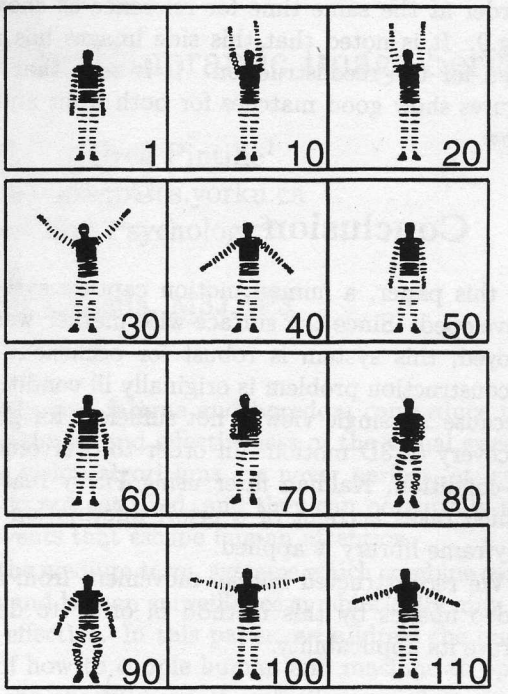


Figure 10: Front view of the reconstructed motion

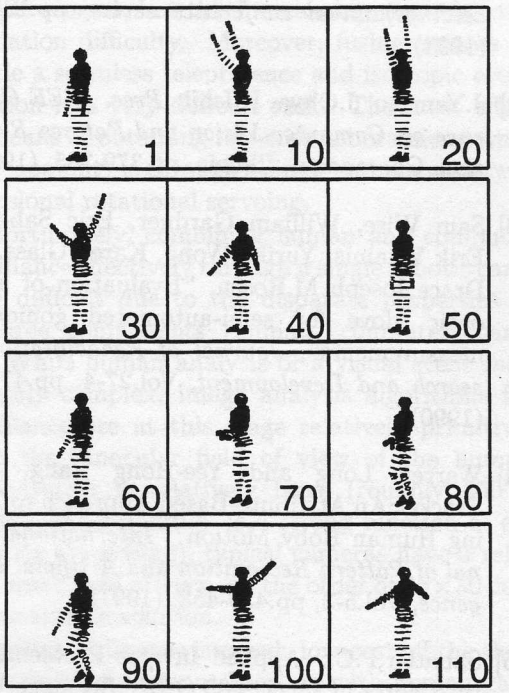


Figure 11: Side view of the reconstructed motion

We have recorded the side view with another camcorder at the same time for reference as shown in Fig.9. It is noted that this side images has never used for the reconstruction. It is seen that these figures show good matches for both front and side view.

7 Conclusion

In this paper, a human motion capture system is developed. Since the surface-wise marker was employed, this system is robust for occlusion. The reconstruction problem is originally ill-conditioned, because 2D single view is not sufficient for general recovery of 3D motion. In order to overcome this ill-condition, Kalman filter using Fuzzy reasoning which takes account of a priori information using keyframe library is applied.

We reconstructed human movement from a real video images by this method in order to demonstrate its applicability.

References

[1] Andrew Blake and Michael Isard, "3D position, attitude and shape input using video tracking of hands and lips", *Computer graphics proceedings,annual conference series,* pp.185-192, (1994).

[2] J.Yamato, J.Ohya, K.Ishii: *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Champaign, Illinois, pp.379-385, (1992).

[3] Sam Wise, William Gardner, Eric Sabelman, Erik Valainis, Yuriko Wong, Karen Glass, John Drace,Joseph M.Rosen, "Evaluation of a fiber optic glove for semi-automated goniometric measurements", *Journal of Rehabilitation Research and Development*, Vol.27-4, pp.411-424, (1990).

[4] Warren Long and Yee-Hong Yang, "Log-Tracker:An Attribute-Based Approach to Tracking Human Boby Motion", *International Journal of Pattern Recognition and Artificial Intelligence*, Vol.5-3, pp.439-458, (1991).

[5] Sabatier,P.C.: Applied Inverse Problems, Lecture Notes in Physics 85 (1978) Springer-Verlag.

[6] K.Amaya, Y.Hara, S.Aoki: "Reconstruction of 3D Human Movement Using Inverse Analysis"; Vision Interface 97, pp.183-188(1997).

[7] A.Rosenfeld, A.C.Kak: *Digital Picture Processing* Academic Press, (1976).

[8] John Craig: *Introduction to robotics*, Addison Wesley, (1955).

[9] Calvert, Bruderlin, Dill, Schiphorst, Welman, "Desktop animation of multiple human figures", *IEEE Computer Graphics and Applications,* May pp18-26, (1993).

[10] Kalman,R.E.: A new approach to linear filtering and prediction problems, Trans. ASME, J.Basic Eng., [82],pp35-45,(1960).

[11] L.A.Zadeh:"Fuzzy Logic," Computer,[1] 4,pp83-93 (1988)

[12] Foley,vanDam, Feiner, Hughes:Computer Graphics, Addison Wesley,(1990).