Reconstruction of 3D Human Movement Using Inverse Analysis

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@mes.titech.ac.jp

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 markers found in the 2D image. The fitting is done using a nonlinear optimization approach. This optimization is essentially difficult because depth information can not be obtained from 2D image. In order to overcome this problem, inverse analysis technique which takes account of the following four kinds of a priori information are applied: 1 Human body can be exhibited by versatile skeleton model. 2 Joint angles have their limits. 3 Human motion is smooth. 4 Human keeps one's balance in slow motion. We reconstructed human movement from a real video images by this method in order to demonstrate its applicability.

1 Introduction

The human body motion is an "object" that can assume a great variety of complex postures. 3D reconstruction of these postures 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 (see Fig.1) 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 a priori information of human motion, e.g. physical and motion constraints were implemented in the inverse analysis.

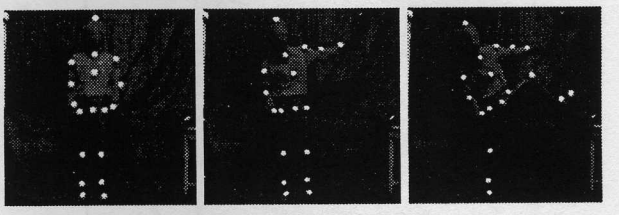


Figure 1: Sample of original Images

2 Reconstruction Procedure Using Inverse Analysis

2.1 Human body model

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.

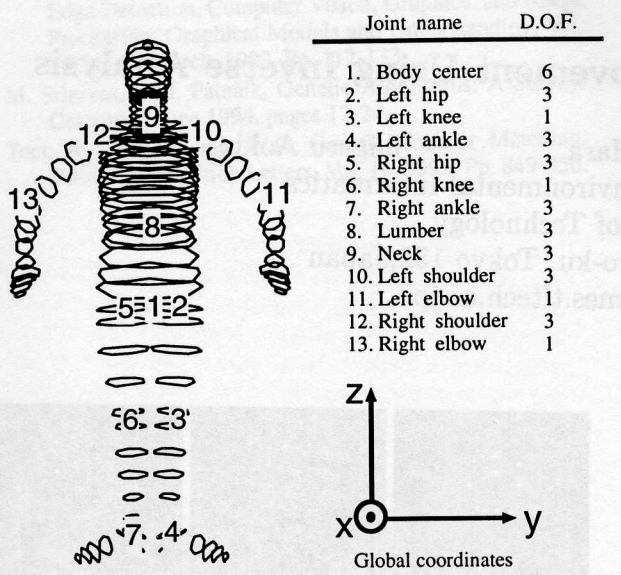
Figure 2 shows our human model. Each joint has its own degree of freedom as indicated in figure 2, and each size of body segment is given.

Using direct kinematics, locations of all markers can be calculated in 3D space from the posture vector $\boldsymbol{\theta}_t$. Applying a knowledge of computer graphics analysis, 2D coordinates of markers on screen $\boldsymbol{g}(\boldsymbol{\theta}_t)$ can be obtained easily[5].

2.2 Improvement of accuracy by singular value decomposition

Let $\bar{\boldsymbol{g}}_t$ be the 2D coordinates of markers in source image at t. Therefore, the problem considered in this paper is to find posture vector $\boldsymbol{\theta}_t$ for each t by solving the following non-linear equation.

$$\bar{\boldsymbol{g}}_t = \boldsymbol{g}(\boldsymbol{\theta}_t) \quad (1)$$



Joint name	D.O.F.
1. Body center	6
2. Left hip	3
3. Left knee	1
4. Left ankle	3
5. Right hip	3
6. Right knee	1
7. Right ankle	3
8. Lumber	3
9. Neck	3
10. Left shoulder	3
11. Left elbow	1
12. Right shoulder	3
13. Right elbow	1

Figure 2: Skeleton model

For simplicity, the subscript t is omitted in the following except the necessary case.

Here, we assume the function $\boldsymbol{g(\theta)}$ written in(1) is smooth and we obtain the solution by iterative calculation with linearization around the estimated solution. Equation (1) can be linearized around j-th estimated solution $\boldsymbol{\theta^j}$ as the following equation.

$$\boldsymbol{g(\theta^j) - \bar{g} = H^j(\theta^{j+1} - \theta^j)} \quad (2)$$

where, $\boldsymbol{H^j}$ is a matrix which is given by:

$$\boldsymbol{H^j = \nabla g(\theta^j)}. \quad (3)$$

Applying standard *NewtonRaphson* method, j+1-th estimated solution can be calculated by:

$$\boldsymbol{\theta^{j+1} = \theta^j - \alpha^j [H^j]^{-1}(g(\theta^j) - \bar{g})} \quad (4)$$

where, α^j is a correction coefficient.

However, the calculation of $[\boldsymbol{H^j}]^{-1}$ in eqn.(4) is unstable, because the shortage of depth information makes this problem ill-conditioned. In order to avoid this unstable calculation, we introduced the singular value decomposition.

First, the singular value decomposition[6] is performed on $\boldsymbol{H^j}$ as the following equation:

$$\boldsymbol{H^j = U^j D^j V^{jt}} \quad (5)$$

where, $\boldsymbol{U^j, V^j}$ are 36 and 34 dimensional orthonormal matrices, respectively, and D is (36 × 34) diagonal matrix which diagonal components $\sigma_1, \sigma_2, \cdots, \sigma_R (R = Rank(A))$ are singular values.

Some singular values become small in the case of ill-condition, and they cause magnification of errors. In order to avoid such magnification of errors, small singular values are omitted.

By using these matrices, the solution of *Moore Penrose* can be represented as the following formula:

$$\begin{aligned} \boldsymbol{\theta^{j+1}} &= \boldsymbol{\theta^j - \alpha^j H^{j\dagger}(g(\theta^j) - \bar{g})} \\ &= \boldsymbol{\theta^j - \alpha^j V^j D^{j\dagger} U^{jt}(g(\theta^j) - \bar{g})} \end{aligned} \quad (6)$$

where, † is the general inverse of *Moore Penrose*, $D^\dagger$ is (34 × 36) diagonal matrix which diagonal components are inverses of singular values: $1/\sigma_1, 1/\sigma_2, \cdots, 1/\sigma_r$(r is the reduced rank by omitting the small singular values).

3 Example Analysis

3.1 Extraction of markers

Figure 1 shows the sample of employed source image. 18 markers, which were forming plastic balls with a diameter of 4cm, were attached to human body.

We used a camcorder (SONY-DCR-VX1000) to record a subject performs "Sumo Movement"for 4 seconds. The recorded image data were input into workstation. Its frame rate was $30 frames/sec$ and its resolution was $300 \times 300 pixel$.

In order to obtain 2D coordinates of 18 markers in source image, the following procedures were performed: (1)The image was smoothed and binarized to eliminate noise and to extract the marker area[7]. (2)All 18 marker areas were labeled, and center of each marker area was calculated by integration. We defined $\bar{\boldsymbol{g}}$ as a 36 dimensional vector which components are 2D coordinates of 18 markers in source image.

3.2 Reconstruction using SVD

The 3D motion was reconstructed from video images by using eqn.(6).

Figure 3 shows the singular values of matrix H in equation (6). The singular values are shown after normalized by its maximum value. It can be seen that the singular value drastically decreases around 25th dimension. We truncated the rank to keep the condition number under 200 (see Cut-line in Fig.3). Around 8 ranks were truncated for this example.

Iterative calculation of equation (6) was stopped when the following criterion was satisfied

$$|\bar{\boldsymbol{g}} - \boldsymbol{g(\theta)}|^2 < 200[\text{pixels}^2]$$

The converged solution $\boldsymbol{\theta_t}$ was used as an initial estimation of next frame $\boldsymbol{\theta^0_{t+1}}$.

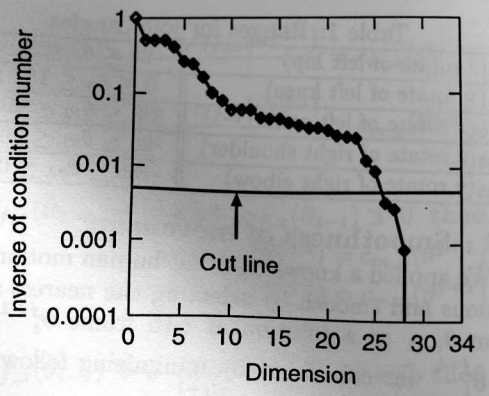


Figure 3: Normalized singular values

Figure 4 and 5 show the front and side view of the reconstructed motion respectively. It is seen that the front view shows the believable "Sumo Motion", but side view contains some strange frames.

A solution of equation (1) is not unique, since 2D image is not sufficient for general recovery. In this section, we employed a solution of *Moore Penrose* to form a solution set of equation (1). A solution of *Moore Penrose* is numerically stable, but does not promise that it is the true solution.

4 Application of A Priori Information

In this section, we consider a priori information about the human motion to improve the estimation accuracy. In other words, we use a set of rules to verify the correctness of body postures.

In order to apply the a priori information, it is necessary to obtain the solution set of equation (1) quantitatively as follows. By using V and D in equation(5), the elements in the null space $\boldsymbol{k^j(z)}$of linear transform $\boldsymbol{H^j}$ can be represented as the following formula:

$$\boldsymbol{k^j(z) = V^j(I - D^{j\dagger}D^j)z}, \quad \forall z \qquad (7)$$

where, $\boldsymbol{z}$ denotes 34 dimensional arbitrary vector, $\boldsymbol{I}$ is the 34×34 identity matrix and small singular values of D^j has already truncated. We will use r as the number of ranks after this truncating procedure.

Because from 0 to r-th components of vector $\boldsymbol{(I - D^{j\dagger}D^j)z}$ are 0, all elements in the null space can be represented as the following linear combination of $34 - r$ number of vectors:

$$\boldsymbol{k^j(z) = v^j_{r+1}z_{r+1} + v^j_{r+2}z_{r+2} + \cdots + v_{34}{}^j z_{34}} \quad (8)$$

where column vector $\boldsymbol{v^j_i}$ is the i-th column of the matrix $\boldsymbol{V^j}$, z_i the i-th component of $\boldsymbol{z}$. Thus the sum of the solution of *Moore Penrose* and an element of the

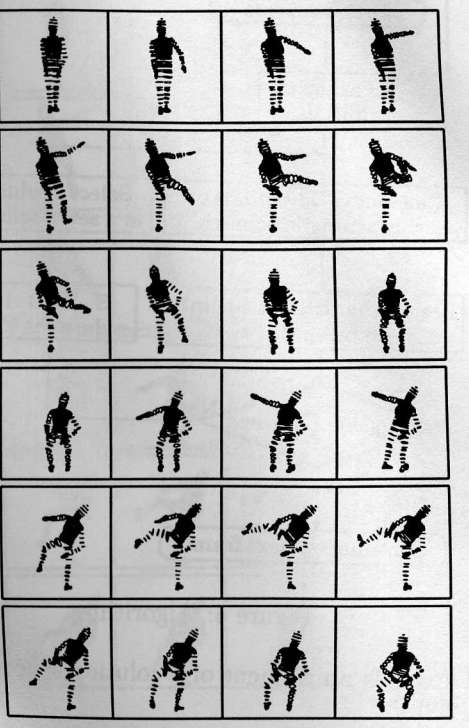


Figure 4: Reconstructed motion (front view)

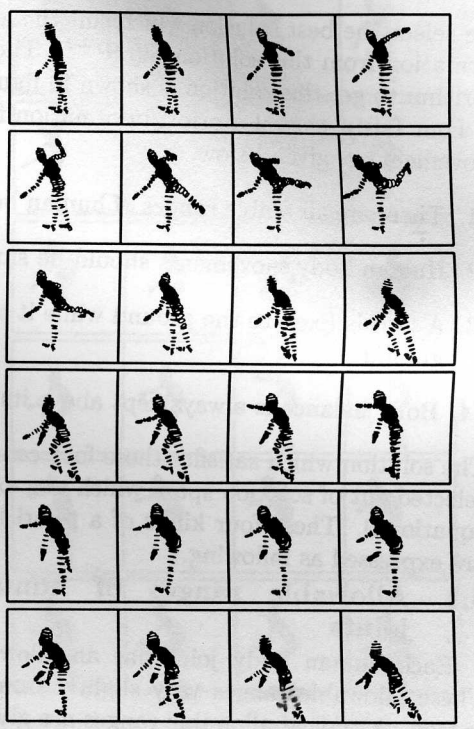


Figure 5: Reconstructed motion (side view)

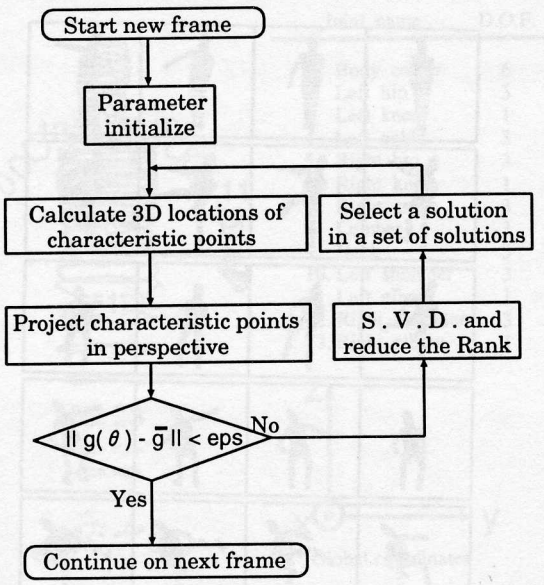


Figure 6: Algorithm

null space is an element of a solution space S^{j+1} for eqn.2[8].

$$S^{j+1} = \{\theta^{j+1} | \theta^{j+1}(z) = \theta^j - H^{j\dagger}(g(\theta^j) - \bar{g}) + k^j(z), \ \forall z\}$$

We select the best solution which suit the a priori information from this solution set S^{j+1}. The basic algorithm to get this solution is shown in figure 6.

Four fairly general a priori information for human movement are given below.

1. There are allowable ranges of human body joints.
2. Human body movements should be smooth.
3. A foot is fixed to the ground while it touches the ground.
4. Body balance is always kept above its stance.

The solution which satisfies these four conditions was selected out of solution space which was expressed by equation 9. These four kinds of a priori information are expressed as following.

4.1 Allowable ranges of human body joints

Each human body joint has an allowable range. These allowable ranges vary slightly from person to person. A typical allowable ranges are given in Table 1 for some major joints. These data are employed for constrains of posture vector θ

Table 1: Ranges for joint angles

θ_8(z rotate of left hip)	-90 $\leq \theta_8 \leq$ 90 (deg)
θ_9(y rotate of left knee)	0 $\leq \theta_9 \leq$ 180 (deg)
θ_{10}(x rotate of left ankle)	-90 $\leq \theta_{10} \leq$ 90 (deg)
θ_{32}(z rotate of right shoulder)	-90 $\leq \theta_{32} \leq$ 90 (deg)
θ_{33}(y rotate of right elbow)	-180 $\leq \theta_{33} \leq$ 0 (deg)

4.2 Smoothness of movement

We applied a knowledge that human motion is continuous and smooth by selecting the nearest solution from θ_{t-1} as a solution at t-th frame θ_t^{j+1}. That is, $\hat{\theta}_t^{j+1}$ was calculated by minimizing following cost function e in solution set S_t^{j+1}.

$$e(z) = \left\| \hat{\theta}_{t-1} - \theta_t^{j+1}(z) \right\|^2 \tag{9}$$

where, $\hat{\theta}_{t-1}$ is a estimated vector at (t-1)-th frame. Because, this cost function e can be expressed with quadratic formula of z, the solution which minimizes e can be obtained by calculating z by the following equation:

$$\frac{\partial e(z)}{\partial z_i} = 0, \qquad \text{i} = 1,2,\cdots,34 \tag{10}$$

where, z_i is the i-th component of z.

4.3 Non-slip constraint of foot

Applying the knowledge of direct kinematics, the 3D coordinates of both foot c_{lf} and c_{rf} can be calculated, where, c_{lf} and c_{rf} are coordinates of the left foot and the right foot, respectively.

A priori information: "A foot is fixed to the ground while it touches the ground " is expressed as below:

$$\textbf{if } c_{lf\_z}(\hat{\theta}_{t-1}) \simeq 0 \textbf{ then}$$
$$c_{lf}(\hat{\theta}_{t-1}) = c_{lf}(\theta^{j+1}) \tag{11}$$

$$\textbf{if } c_{rf\_z}(\hat{\theta}_{t-1}) \simeq 0 \textbf{ then}$$
$$c_{rf}(\hat{\theta}_{t-1}) = c_{rf}(\theta^{j+1}) \tag{12}$$

4.4 Body balance constraint

In case human movement is rather slow, the dynamics of human body can be assumed semi-static. Under this assumption we can introduce the following information:

- A body center is located above the stance.

Strictly, body center should be taken as a center of gravity, but we just used a center of pelvis for simplicity.

if $c_{rf\_z}(\hat{\theta}_{t-1}) \simeq 0$ **and** $c_{lf\_z}(\hat{\theta}_{t-1}) \simeq 0$ **then**

$$\frac{c_{rf\_x}(\theta^{j+1}) + c_{lf\_x}(\theta^{j+1})}{2} = c_{m\_x}(\theta^{j+1})$$
$$\frac{c_{rf\_y}(\theta^{j+1}) + c_{lf\_y}(\theta^{j+1})}{2} = c_{m\_y}(\theta^{j+1}) \quad (13)$$

if $c_{rf\_z}(\hat{\theta}_{t-1}) \simeq 0$ **and** $c_{lf\_z}(\hat{\theta}_{t-1}) \gg 0$ **then**

$$c_{rf\_x}(\theta^{j+1}) = c_{m\_x}(\theta^{j+1})$$
$$c_{rf\_y}(\theta^{j+1}) = c_{m\_y}(\theta^{j+1}) \quad (14)$$

if $c_{rf\_z}(\hat{\theta}_{t-1}) \gg 0$ **and** $c_{lf\_z}(\hat{\theta}_{t-1}) \simeq 0$ **then**

$$c_{lf\_x}(\theta^{j+1}) = c_{m\_x}(\theta^{j+1})$$
$$c_{lf\_y}(\theta^{j+1}) = c_{m\_y}(\theta^{j+1}) \quad (15)$$

where, c_{rf_i} , c_{lf_i} , c_{m_i} are 3D coordinates of the left foot, the right foot and the body center, and subscripts $i = x, yandz$ correspond to each X,Y and Z, respectively. dimension. It is noted that Z is a vertical axis.

5 Example Reconstruction using A Priori Information

A 3D motion was reconstructed from video images which recorded in section 3.1 with four kinds of a priori information.

Figure 7 and 8 show the front and the side view of the reconstructed motion, respectively.

It is seen that the both figures show a believable "Sumo Motion". It is noted that an iterative calculation converges around 7-th step for each frame.

6 Conclusion

In this paper, we have provided a method to reconstruct human motion from a single view. This 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, inverse analysis method with a priori information is applied. The procedures is summarized as below:

1. Markers on feature points on the body are extracted from the processed input image.
2. Human body is modeled with 34 major joint angles.
3. A inverse analysis is applied to fit this human skeleton model to the marker locations found in the image. The truncated singular value decomposition is performed to obtain a solution space. A unique solution which satisfies a priori information: ranges of joint angles, a smoothness of motion and a body balance is selected from this solution set.

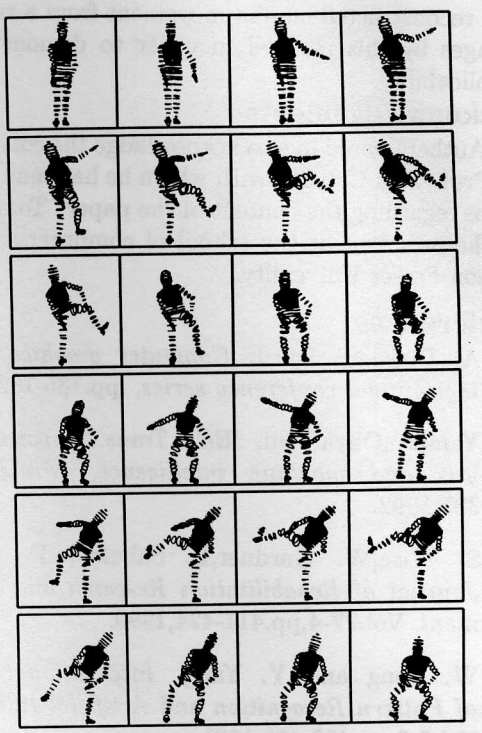


Figure 7: Reconstructed motion (front view)

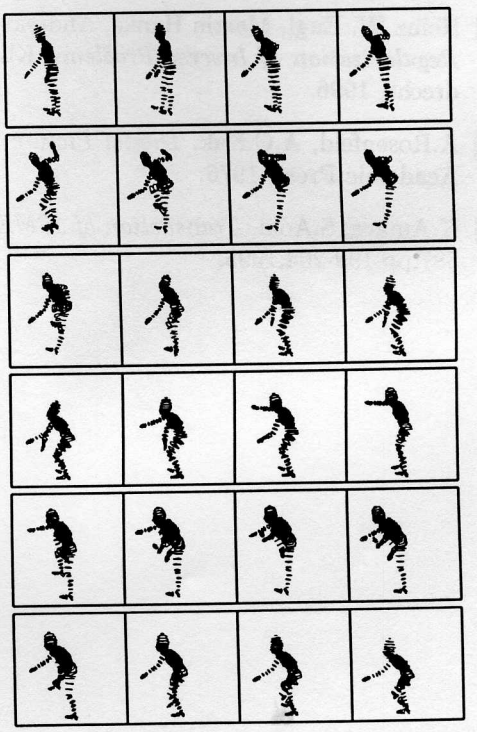


Figure 8: Reconstructed motion (side view)

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

Acknowledgements

Authers would like to acknowladge the contribution of Prof. Tom Calvert, with whom he had many discussions regarding the content of the paper. Tom Calvert is the professor at the school of computer science at Simon Fraser University.

References

[1] A. Blake,M. Isard. *Computer graphics proceedings,annual conference series*, ,pp.185-192,1994.

[2] Yamato,Ohya,Ishii, *IEEE Trans. on pattern analysis and machine intelligence*, Vol.??,pp.379-385,1992.

[3] S. Wise,W. Gardner,E. Sabelman,E. Valainis, *Journal of Rehabilitation Research and Development*, Vol.27-4,pp.411-424,1990.

[4] W. Long and Y. Yang, *International Journal of Pattern Recognition and Artificial Intelligence*, Vol.5-3,pp.439-458,1991.

[5] John Craig, *Introduction to robotics*, Addison Wesley, 1955.

[6] Heinz W. Engl, Martin Hanke, Andreas Neubauer *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.

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

[8] K.Amaya, S.Aoki, *Transaction of JSME A*,Vol.61-587,pp.199-204,1995.