Motion Analysis of Human Based on Coplanar Constraint

Chunhong PAN and Songde MA

Sino-French Laboratory in Computer Science, Automation and Applied Mathematics
National Laboratory Of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences
email panch@lotus.ia.ac.cn

March 29, 2000

Abstract

In this paper we show that based on coplanar constraint the motion and structure of the jointed links can be determined within a scale factor from a monocular sequence of images. We analyze the relationship among the number of the frames and links and the number of the equations and unknowns, and achieve the unique robust numerical solution to motion estimation of the coplanar links by Genetic Algorithm. Then we apply the techniques to human motion analysis, and obtained the $3D$ motion data of joints, could successfully re-animate the data. The experiments with simulated data and real images are included to demonstrate the validity of the theoretic results.

1 Introduction

Generating appealing human motion is the central problem in virtual reality and computer animation[6][3]. A wide variety of techniques are presented in the process of creating a complex animation. Generally speaking, these techniques can be grouped into three main classes: keyframing[9], procedure[2], and motion capture[8]. The generated animations by these techniques are so-called keyframe animation, procedural animation, and motion capture animation. Motion capture is a very popular technique because of the relative ease with which many human motions can be recorded. Furthermore, up to now motion capture is the only effective method for the human arbitrary motion. Motion capture employs some special sensors or markers to record the motion of a human performer by multiple cameras from different directions. The recorded data is then used to generate the motion for an animation. The system is able to estimate position with an accuracy of 0.1% of the workspace diameter. However, to achieve such accuracy, it is necessary to have a complicated system composed of many special markers, and 4-8 cameras that need to be accurately calibrated. It is stressful and hard to handle in many applications which limits the use of the system.

An articulated object means an object consisting of some rigid links connected together by joints[1]. It is well known that human body can be approximately modeled by articulated objects. In articulated model, the motion of each constituent part is rigid, but the motion of the whole object is nonrigid. Since there exist multiple motions in the articulated objects, to recover their $3D$ structure from motion, some special motion or projection models must be assumed. The typical motion models have: the coplanar motion with a point known or fixed[7], the fixed axis motion[10], and at least one known point[2]; the projection model has: orthogonal projection[5]. Without these models it is seldom possible to determine 3D structure of the articulated objects from the monocular perspective views by correspondences of the joints of the links although provided with more than two frames.

Some observations reveal that in general arms or legs of human do not move about arbitrarily[4]. Rather, for anatomical reasons, each leg tends to move approximately in a single plane for extended periods of time. Here to obtain the unique 3D human motion we assume that over a small number of image frames some parts such as leg and arm move in a unknown fixed plane. In many real human motions the constraint appears reasonable. If we can determine motion of the coplanar parts of the articulated human, it is possible to estimate $3D$ structure of all human. Thus motion and structure determination of some coplanar rigid links from many frames play important role in the motion estimation of the articulated objects. In this paper, we analyze the $3D$ structure of multi-link from motion, and investigate the relationship among the number of the frames and links and the number of the equations and unknowns. By Genetic Algorithm, the robust numerical solutions can be achieved. The experiments with the simulated data and the real images are included to demonstrate the validity of the theoretic results.

This paper is organized as follows: section 2 analyzes theoretically the motion and structure of the coplanar multi-link, section 3 describes the robust motion estimation of the articulated objects due to noise. Experimental results and conclusions are presented in section 4 and 5.

2 Motion Analysis of Coplanar Multi-link

2.1 Coplanar Multi-link

To determine 3D structure of an articulated object from a monocular sequence of images, and further to make the solution of 3D motion unique within a global scale factor, in addition to the rigidity conditions we need constraints on the motion. For the articulated objects we assume that there exist some links which move in a fixed plane in some frames. In real motion the constraint is reasonable. For example in human walking or running, each of the foot is alternately fixed on the ground during stance and swing phase, but movement of any leg can be considered in a plane in a short time because of only one degree of freedom at the knee. Fig.1(left) shows the some coplanar links connected together by joints. Our aim is to determine $3D$ structure of the rigid links from a monocular sequence of images based on coplanar constraint.

The imaging geometry of the multiple frame motion analysis problem is shown in fig.1(right). The origin of the coordinate system is the focal point. The image plane is at a distance f from the camera and is orthogonal to the optical axis. The x-axis and y-axis of the image plane are respectively parallel to the X-axis and Y-axis of the camera coordinate system. A pinhole camera model is used. Images of a moving object are taken at different times. By processing these images, we attempt to determine the $3D$ motion of the object.

The feature point $A^{(j)}$ denotes one of the endpoints of the link at time t_j, and $\vec{a^{(j)}}$ is the unit direction vector along the line of sight of point $A^{(j)}$. The feature space point $A^{(j)}$ and the corresponding image point $a^{(j)}$ are denoted A and a at time t_0. If the image $a^{(j)}$ of $A^{(j)}$ is located at $(x_{a^{(j)}}, y_{a^{(j)}})$, then the unit vector is given by

$$\vec{a^{(j)}} = \frac{1}{\sqrt{x^2_{a^{(j)}} + y^2_{a^{(j)}} + f^2}}(x_{a^{(j)}}, y_{a^{(j)}}, f) \quad (1)$$

where f is the focal length. Let the distance from the focal point to point $A^{(j)}$ be $\lambda_{A^{(j)}}$. $\lambda_{A^{(j)}}$ can be given by

$$\vec{A^{(j)}} = \lambda_{A^{(j)}}\vec{a^{(j)}} \quad (2)$$

2.2 Rigidity Constraints

For a single rigid link with two endpoints A_1 and A_2, the angle $\angle A_1OA_2$ is the same as the angle $\angle a_1oa_2$ which is denoted as $\alpha_{a_1a_2}$. Obviously the angle can be expressed as

$$\cos\alpha_{a_1a_2} = \vec{a_1} \bullet \vec{a_2} \quad (3)$$

where A_1 and A_2 are two endpoints of single link, and a_1 and a_2 denote the corresponding image points.

Let l denote the length of the link. From the triangle $\triangle A_1OA_2$ we obtain

$$| A_1 - A_2 | = l \;\Rightarrow\; \lambda^2_{A_1} + \lambda^2_{A_2} - 2\lambda_{A_1}\lambda_{A_2}\cos\alpha_{a_1a_2} = l^2 \quad (4)$$

The distance between two endpoints of link must be the same for all links in all views which yields the following rigidity constrain equations

$$\lambda^2_{A_i^{(j)}} + \lambda^2_{A_{i+1}^{(j)}} - 2\lambda_{A_i^{(j)}}\lambda_{A_{i+1}^{(j)}}\cos\alpha_{a_i^{(j)}a_{i+1}^{(j)}} = l^2$$
$$for \quad j = 0, 1, 2...n \quad and \quad i = 1, 2...k \quad (5)$$

2.3 Coplanar Motion Constraints

For an arbitrary space plane in coordinate system $OXYZ$. The unit normal vector of the plane can be written as

$$\vec{n} = \frac{1}{\sqrt{n_x^2 + n_y^2 + 1}}(n_x, n_y, 1) \quad (6)$$

Let $\phi_{a_i^{(j)}}$ and $\phi_{a_{i+1}^{(j)}}$ denote the angles between the unit normal vector of the plane and $\vec{a_i^{(j)}}$ and $\vec{a_{i+1}^{(j)}}$ respectively, we have:

$$\cos\phi_{a_i^{(j)}} = \vec{n} \bullet \vec{a_i^{(j)}}; \qquad \cos\phi_{a_{i+1}^{(j)}} = \vec{n} \bullet \vec{a_{i+1}^{(j)}} \quad (7)$$

Since movement of the link is always in a fixed plane which means that all endpoints of links at different times are fixed in a same plane. It leads to the following coplanar constraint equations

$$D = \lambda_{A_i^{(j)}}\cos\phi_{a_i^{(j)}} \quad (8)$$

Here $\lambda_{A_i^{(j)}}$ are unknown variables for $i = 1, 2...k$ and $j = 1, 2...n$ and D is the distance from the origin of the coordinate system to the fixed plane. By combining equation (1),(7) and (8), the above equations can be written as

$$k_{a_i^{(j)}}\lambda_{A_i^{(j)}}(x_{a_i^{(j)}}n_x + y_{a_i^{(j)}}n_y + f) = D \quad (9)$$

Here $k_{a_i^{(j)}} = \frac{1}{\sqrt{n_x^2+n_y^2+1}} \bullet \frac{1}{\sqrt{x^2_{a_i^{(j)}}+y^2_{a_i^{(j)}}+f^2}}$. In above equations the $\lambda_{A_i^{(j)}}$,n_x and n_y are unknown variables for $j = 0, 1, 2, ...n$ and $i = 1, 2, ...k$

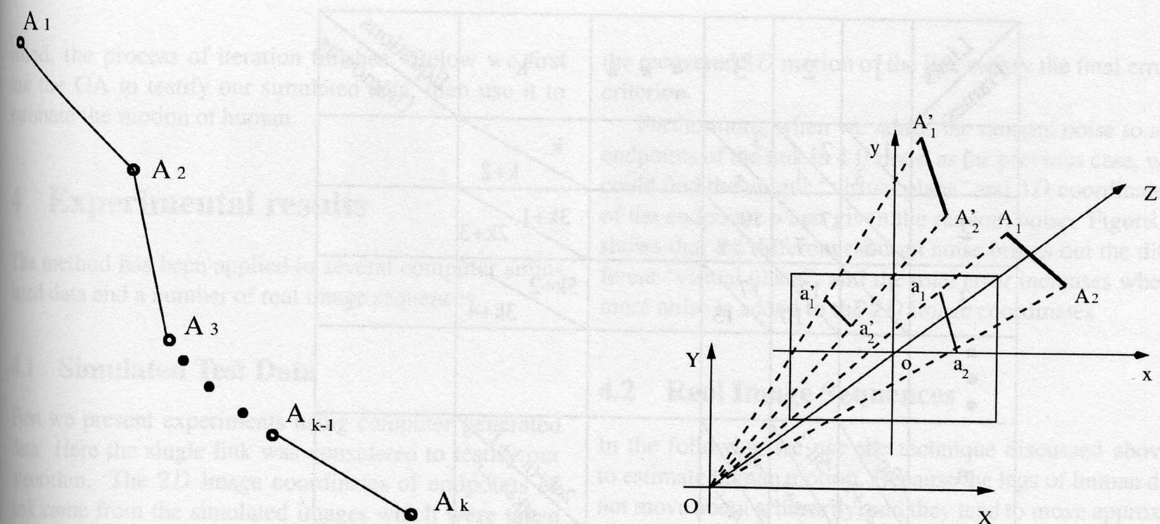

Figure 1: Left: Coplanar multi-link of articulated object
Right:Perspective imaging geometry for motion analysis

2.4 Analysis of Solutions

We assume that the articulated object has k connected links which rotate and translate in a fixed plane for all frames. For the fixed arbitrary space plane when given n frames the unknown variables have $\lambda_{A_i^{(j)}}$, n_x and n_y, and the number of the unknowns is $n(k+1)+1$ within a scale factor. From Equ(4) we obtain $k(n-1)$ rigidity constraint equations, and from coplanar constraint we have $n(k+1)-1$ equations, and therefore the number of the equations is $2nk+n-k-1$.

when $k=1$, which means only one link of the articulated object is in a fixed plane, the number of the equations is $3n-2$, and the number of the unknowns is $2n+1$. Obviously when provided with 3 frames the number of the equations is 7 which is equal to the number of the unknowns. Thus it is possible to recover the 3D motion of the single link from above equations when given 3 frames. When $k=2$, the number of the unknowns and equations are $3n+1$ and $5n-3$ respectively. To obtain the 3D information of the endpoints of the links only two frames are needed. Table.1 gives the relationship among the number of the frames and links and the number of the equations and unknowns. From Table.1 we see that: when $k>2$, there are more equations than unknown variables although given 2 frames.

Furthermore we assume that the links are fixed in a plane for each frame, but in the different plane for the different frame. The assumption seems more reasonable for human motion. For example: because of only one degree of freedom at the elbow the forearm and upperarm are always in same plane for each frame but maybe in the different plane for the different frame. In this case the unit normal vector of the plane can be written as $\vec{n^{(j)}} = \frac{1}{\sqrt{n_x^{(j)2}+n_y^{(j)2}+1}}(n_x^{(j)}, n_y^{(j)}, 1)$. For k links when given n frames the number of the unknown variables is $n(k+1)+2n+1$, while the number of the equations is similar to the previous case. When $k<2$, the number of equations is always less than that of the unknowns, and only when $k \geq 3$ and $n \geq 5$ there are more equations than unknowns.

In fact the number of the needed frames are relevant to not only the number of the links but also the position of the unknown plane. Especially when the optical axis of the camera is perpendicular to the fixed plane, we have $n_x = n_y = 0$, thus the ranges of endpoints of the link are only the unknowns. When we assume the the fixed plane is parallel to Y axis which means $n_y = 0$. the normal vector of the plane can be written as $\vec{n} = \frac{1}{\sqrt{n_x^2+1}}(n_x, 0, 1)$. in this case the number of the unknowns is $n(k+1)$ for n frames, while the number of the equations is unchangeable, which means that less frames are needed than when the unknown plane is arbitrary.

3 Robust estimation:

Due to noise, correspondences with error, and approximate assumption on articulated model, the exact solutions of $3D$ motion of the articulated model, which satisfy meanwhile the coplanar constraints and rigidity con-

Links \ Frames	1	2	3	• • •	k	Equations / Unknowns
1	1 / 3	2 / 4	3 / 5		k / k+2	
2	4 / 5	7 / 7	10 / 9		3k+1 / 2k+3	
3	7 / 7	12 / 10	17 / 13		5k+2 / 3k+4	
• • •						
n	3n-2 / 2n+1	5n-3 / 3n+1	7n-4 / 4n+1		2nk+(n-k)-1 / n(k+1)+1	

Table 1 : Relationship among the number of the frames and links and the number of the equations and unknown

straints, do not exist. Generally the solutions which make the articulated objects coplanar can not have the length of its each link same in these views, and vice verse. In this case, although not all the endpoints of the articulated link are in a single plane and meanwhile the length of the links may be not equal in these frames, there exists a "virtual plane". In fact the "virtual plane" can be an appropriate selection of the solutions in terms of a tradeoff between the coplanar constraint error and the rigid constraint error.

If the endpoints of links are not in the "virtual plane" for some frames. The distances between these points and this plane can be written as:

$$d_{ij} = \frac{|\lambda_{A_i^{(j)}} \vec{a_i^{(j)}} \cdot \vec{n}|}{\sqrt{n_x^2 + n_y^2 + 1}} \quad (10)$$

We have the following distance error criterion:

$$\xi_1 = \sum_{i=0}^{n} \sum_{j=1}^{k} d_{ij} \quad (11)$$

In addition the rigidity constraints may not be satisfied exactly in these views, i.e $\| \vec{A_{1j}^{(i)}} - \vec{A_{2j}^{(i)}} | - | \vec{A_{1j}^{(l)}} - \vec{A_{2j}^{(l)}} \| \neq 0$. Thus the length error criterion can be considered as

$$\xi_2 = \sum_{i \neq l; i,l=0}^{n} \sum_{j=1}^{k} \| \vec{A_{1j}^{(i)}} - \vec{A_{2j}^{(i)}} | - | \vec{A_{1j}^{(l)}} - \vec{A_{2j}^{(l)}} \| \quad (12)$$

The final error criterion can be a tradeoff between ξ_1 and ξ_2

$$\xi = \alpha\xi_1 + \beta\xi_2 \quad (13)$$

Where $\alpha, \beta \subseteq 0, 1$ and $\alpha + \beta = 1$. It is a typical optimization problem which can be solved by dynamic programming.

Genetic Algorithm: It is well known that Genetic Algorithm is the robust global search and optimization technique which imitate the mechanism of life evolution in the nature. In this paper, we use Genetic Algorithm to estimate $3D$ motion of the coplanar rigid links. First the final error criterion above is established as the objective function. According to the equation below, the objective function is mapped into fitness function value.

$$f = C + \frac{\xi}{1 + \xi} \quad (14)$$

Where C is the constant factor. The unknown parameters: $\lambda_{A_i^{(j)}}, n_x$ and n_y $(j = 0, 1, 2, ...n,\ i = 1, 2, ...k)$ are converted into binary codes which are concatenated together in one order and construct an individual(or chromosome), then generate an initial population of n individuals. Based on the initial population, according to certain probability, m individuals, the most highly chromosome(in accordance with the evaluation function) in the current generation are reproduced in the new generation, and copies(or parents) are put into a mating pool for further genetic optimizing operations: cross over and mutation, according to corresponding probability. Through trial-and-error process, the chromosomes with greater fitness survive to the next generation, thus the population with the great fitness increase, when the termination conditions(the predefined threshold of the final error)are satisfied, the process of iteration finishes. Below we first use the GA to testify our simulated data, then use it to estimate the motion of human.

4 Experimental results

The method has been applied to several computer simulated data and a number of real image sequences.

4.1 Simulated Test Data

First we present experiments using computer generated data. Here the single link was considered to testify our algorithm. The $2D$ image coordinates of endpoints of link came from the simulated images which were taken from some snapshots of a link confined to coplanar motion. Without loss of generality the focal length f is set to 1.

As discussed in the previous section, it is possible to recover $3D$ motion of the single link when given 3 frames, but obviously the solutions are not unique. Due to the multiplicity of the solutions, we need more frames for the unique solution. For $n \geq 4$, an over-constraint system is obtained(See Table 1: 10 equations, 7 unknowns), generally there are the unique solutions when given 4 frames.

We used computer to generate 4 frames of single link which is in a fixed plane , and obtained the $2D$ observations of two endpoints of the single link: ($a_1^{(1)}$:(2.0, 2.0), $a_2^{(1)}$:(0.403, 0.62)); ($a_1^{(2)}$:(5.0, 5.0), $a_2^{(2)}$:(1.39, 0.788)); ($a_1^{(3)}$:(4.0, 6.0), $a_2^{(3)}$:(0.667, 1.33)); ($a_1^{(4)}$:(8.0, 2.0), $a_2^{(4)}$:(2.0, 0.0)). By GA optimization approach discussed above, we got the unique solutions of the normal vector of the fixed plane: n_x = 0.4986, n_y = 0.4986(the corresponding normal angle $\theta_x = 26.5°$, $\theta_y = 26.5°$). Once the normal vector is determined, we can obtain the $3D$ coordinates of endpoints at different frames when length of the link is fixed. In this case, the yielded normal vector and $3D$ coordinates can satisfy exactly the coplanar equations and rigidity equations.

We added some random noise (percentage of the $2D$ image coordinates of the corresponding endpoints) to $2D$ image coordinates of endpoints in the different frames. First the random noise is added to the image coordinates of any one endpoint. Using the same method, we obtained the normal vector, $3D$ coordinates of the endpoints with random noise in $2D$ image coordinates, and total error. Figure 2 gives relationship between the normal angle(left), $3D$ coordinates of the endpoint with the random noise(middle), total error(right) and the random noise. Although the coplanar equations and rigid equations can not be satisfied exactly due to the noise in $2D$ images, there does exist "a virtual plane" which makes the recovered $3D$ motion of the link satisfy the final error criterion.

Furthermore, when we added the random noise to all endpoints of the link in 4 frames, as the previous case, we could find the unique "virtual plane" and $3D$ coordinates of the endpoints when given the random noise. Figure 3 shows that the different random noise brings out the different "virtual plane", and the total error increases when more noise is added to the $2D$ image coordinates

4.2 Real Image Sequences

In the following we use the technique discussed above to estimate human motion. Because the legs of human do not move about arbitrarily, and they tend to move approximately in a single plane in a short time, here we attempt to estimate $3D$ motion of legs from a sequence of images when a person is walking. It is known that good correspondences of feature points from images is of importance for 3D motion estimation, while extracting and tracking feature points in the grayscale sequence images are very difficult. For similarity, we attached some markers at the joints(see fig.4). By using video camera we took the sequence of human walking, obtained the frames by sampling the sequence at the rate of 30 frames per-second. Then we tracked these markers in the sequence images, and got the $2D$ image coordinates of these markers. Finally we used the $2D$ image coordinates to estimate $3D$ coordinates of the joints by above approach.

Due to only one degree of freedom at the knee, the leg can be considered as the coplanar articulated objects with two links. In fact, when attaching the markers to the joints, it is very difficult to make the three markers at one of the legs in the same plane so that leg can be considered as the multi-link, thus in this experiment we consider each leg as two different single link, and estimate the $3D$ motion of two single link respectively. In addition, in this experiment it is dispensable to have a person walking along the line path. Because when the sequence is sampled at the rate 30 frames per-second, one step have more than 4 frames. We give a pre-defined total error threshold as the criterion, and use this criterion to decide which frames are in the same "virtual plane" by trial-and-error process. First the initial four frames are selected to find the "virtual plane" and meanwhile calculate the total error, then next frame is added to calculate the total error using the same plane, when the total error is less than the error threshold, the frame are considered as in the same "virtual plane", otherwise the frame belongs to next one. We repeat the process up to the end of the sequence. By the method, we obtained the $3D$ coordinates of the joints of the legs. Fig.5 gives the normalized $3D$ coordinates of the waist, knee and ankle, the results are very satis-

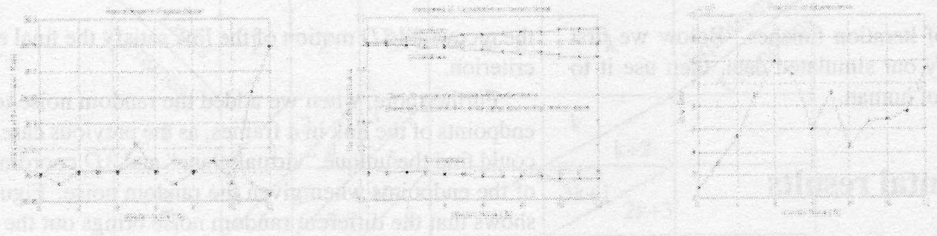

Figure 2: The normal angle(left), $3D$ coordinates of the endpoint with the random noise(middle), total error(right) vs the random noise

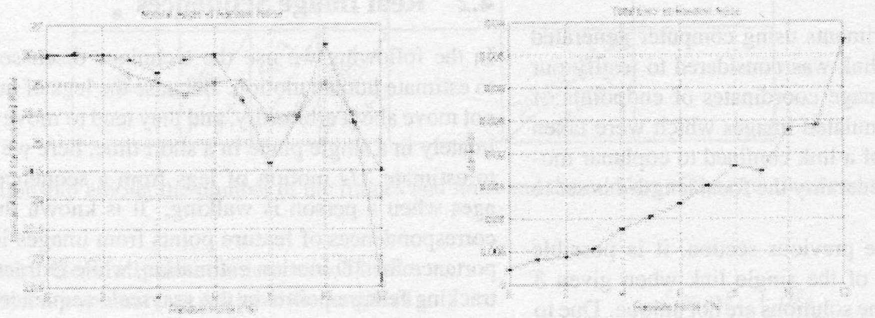

Figure 3: The normal angle(left), total error(right) vs the random noise when add some random noise to all endpoints

factory. Then the obtained $3D$ motion data of joints of leg is attuned to generate the animation of walking. Fig.6 shows the stimulated walking of articulated model at a given viewpoint.

5 Conclusions

Motion and structure determination of articulated objects from multiple frames is important in a large number of applications such as human gait analysis and the monitoring of robot motion. In this paper based on assumptions of the articulated model and coplanar constraint, we estimate $3D$ motion of the human walking. The experimental results are very satisfactory. Our further research work is to track automatically the features of human such as joint points in the sequence images by constraint fusion, then use decomposition approach to estimate $3D$ motion of the total human. We have gotten some results.

References

[1] J.K. Aggarwal, Q. Cai, W. Liao, and B. Sabata. Nonrigid motion analysis: Articulated and elastic motion. *Computer Vision and Image Understanding*, 70:142–156, 1998.

[2] D.T.Lawton. Constraint-based inference from image motion. In *Proc. 1st Annu. Nat. Conf. Artificial Intelligence*, pages 300–309, Stanford Univ., 1980.

[3] J.K. Hodgins and J.F. O'Brien. *Computer Animation. The Encyclopedia of Computer Science*. International Thomson Business Press, 1998.

[4] D.D. Hoffman and B.E.Flinchbaugh. The interpretation of biological motion. *Biological Cybernetics*, 42:195–204, 1982.

[5] I.A. Kakadiaris, D.Metaxas, and R.Bajcsy. Active part-decomposition, shape and motion estimation of articulated objects: A physics-based approach. In *in Proceedings of CVPR*, pages 980–984, 1994.

[6] V.I. Pavlovic, R. Sharma, and T.S. Huang. Visual interpretation of hand gestures from human-computer interaction, a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:677–617, 1997.

[7] Holt R.J., Huang T.S., Netravali A.N., and Gian R.J. Determining articulated motion from perspective views: A decomposition approach. *Pattern Recognition*, 30:1435–1449, 1997.

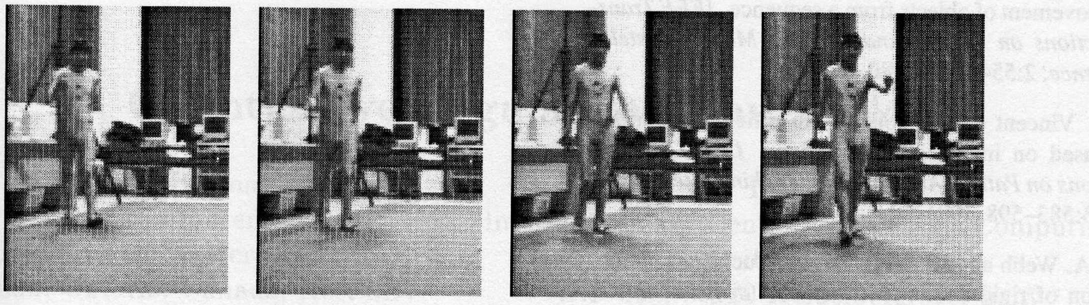

Figure 4: A sampled image sequence with human walking

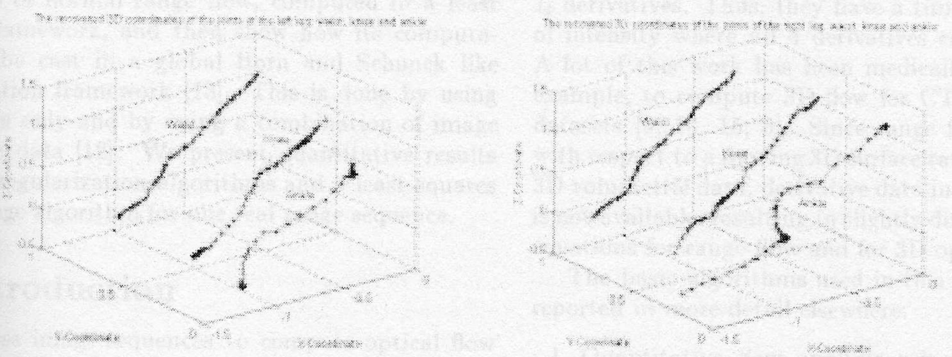

Figure 5: The recovered 3D motion trajectories of leg: waist, knee, ankle

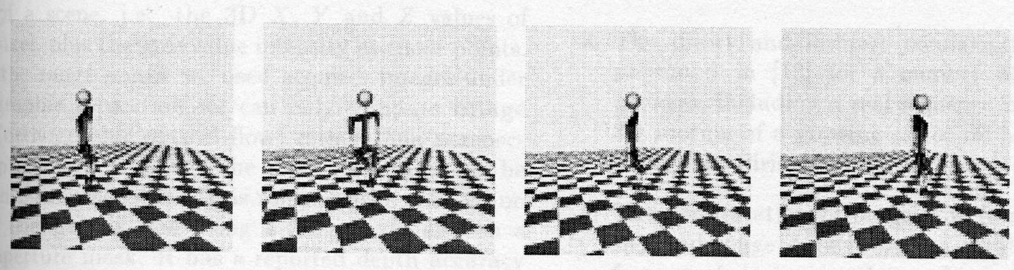

Figure 6: The generated human walking at any viewpoint

[8] J.K. Roach and J.K. Aggarwal. Determining the movement of objects from a sequence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2:554–562, 1980.

[9] L. Vincent and P. Soille. An efficient algorithm based on immersion simulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:583–598, 1991.

[10] J.A. Webb and J.K. Aggarwal. Structure from motion of rigid and jointed objects. *artificial Intell.*, 19:107–130, 1982.