Real-time Human Motion Capturing System

Bau-Cheng Shen, Huang-Chia Shih, and Chung-Lin Huang,
Department of Electrical Engineering
National Tsing-Hua University,
Hsin-Chu, Taiwan.

***Abstract*—**This paper presents a real time vision-based human motion capturing and recognition system using two calibrated CCD cameras. We propose a simple but effective method to estimate the motion parameters (BAPs) of the human object by analyzing the vertical projection profile and the horizontal projection profile in each view to identify different arm and leg postures. With the identified postures, we can apply the Kalman filtering to capture the motion parameters (joint angels). Our method is divided into *macro motion analysis* and *micro motion analysis*. The former identifies certain well-defined postures and the latter traces the variation of joint angle or BAPs. In the experiments, we test 22 different arm and leg postures and show the errors of the estimated BAPs.

Keywords-human tracking; motion parameter; kalman filtering; template matching;posture analysis.

I. INTRODUCTION

Human motion analysis [1] involves tracking human object or objects from a single and multiple views, capturing the motion parameters of the moving object, and recognizing human object activities. In [2], motion information motion information is represented by a feature image. Their idea is to compare the feature images with reference features of different learned actions and look for the best match to find the human posture. Dockstader et al. [3] introduce a near real-time tracking method for multiple articulate objects by using a probabilistic mixing of low-level features and component. In [4], they propose a distributed, real-time computing platform for tracking multiple interacting persons in motion. They use a multi-view implementation to combat the negative effects of occlusion and articulated motion. In [5], their research presents a new model-based approach toward the three-dimensional (3-D) tracking and extraction of gait and human motion. They propose a hierarchical, structural model of the human body and introduce the concept of hard and soft kinematics constraints.

For real-time system, silhouette-based human motion tracking has been proposed [6-9]. Haritaoglu *et al.* [6,7] combine silhouette-based shape model, a motion model, and correlation-based matching method to find multiple people in foreground region and then track each person. Iwasawa *et al.* [8] propose a real-time method to estimate human postures in 3D from trinocular images. They apply the genetic algorithm to estimate the elbow and knee joint positions. Huang *et al.* [9] propose another real-time silhouette-based human motion analysis system using two cameras. They use an overlapped tri-tree model-based matching between the human silhouette and the model to estimate the joint angles. Collin *et al.* [10] introduce a viewpoint dependent technique based on template matching of body silhouette. Their method analyzes the period gate signal based on silhouette width and height for gate identification. Zhao *et al.* [11] propose a similar method based on amplitude spectrum analysis of period gate signals.

This paper introduces a silhouette-based human motion parameter capturing from binocular images. To capture the human motion postures, our system contains the macro motion analysis and the micro motion analysis. Two categories of parameters: Body Definition Parameters (BDPs) and Body Animation Parameters (BAPs) are used to describe human object model and its motion parameters in term of the joint angles. Our system consists of the macro motion analysis and micro motion analysis. The former analyzes the projection profile to identify the postures, whereas the latter uses the Kalman filter to predict the BAPs based on the different posture-to-posture transition. Our system can track 22 different arm and leg postures.

II. PREPROCESSING AND SYSTEM INITIALIZATION

The system which consists of two orthogonal views and the corresponding avatar are shown in Figure 1. The preprocessing module is developed to extract the human object followed by the human motion capturing module. Here, we apply a 3-D articulated human model [1] which consists of ten elliptical cylinders (representing torso, head, arms, and legs) and 10 connecting joints. To extract the silhouette of human object, we apply the well-known background subtraction [2]. This method extracts the foreground object by evaluating the difference between the background model and the current image.

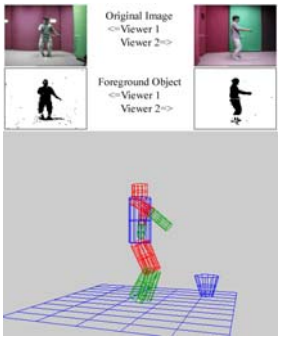


Figure 1. The motion capturing from two views.

Based on the vertical projection profiles of the human silhouette from two views, we may construct 3-D human model. The variation of the joint angles is the prominent features for analyzing human motion. The 3-D articulated model [1] has been used as a top-down process for the human motion tracking. Each cylinder is described by three parameters: the length of the axis, the major and the minor axes of the ellipse cross- section. The cylinders in the model are connected by joints with different degree-of-freedoms (DOFs). The BAPs (motion parameters) represent the ten joints of human model. These joints are located at navel, neck, right shoulder, left shoulder, right elbow, left elbow, right hip, left hip, right knee, and left knee. The human joints are either

flexion or spherical. The former has 1 DOF, whereas the latter has 3 DOFs. Totally, there are four spherical and five flexion joints, and the human model has 17 DOFs.

We assume the initial posture to be standing up with stretched-arms. To evaluate Body Definition Parameters (BDPs), we construct the vertical projection profile of the binary human silhouette in façade as $P_{v1}(x_i)$ and $x_r \leqq x_i \leqq x_l$, where x_l and x_r are the left and right boundary of the profile. Second, we remove the head by applying morphological filtering operations, which consist of the morphological closing and then opening operations by a structuring element (a horizontal rectangle with size $0.8W_{torso} \times 1$). Third, we define the center of body in y-axis. Let (x_c, y_c) be the location of center of body, which happens to be the position of navel. Define the long radius of torso, the long radius of head, the short radius of head, the long radius of leg, and the short radius of leg. Finally, we average the location of the pixels on the left (or right) boundary of body to find the extreme position of right (or left) arm as y_l (or y_r). Having extracted the extreme positions, we may define the position of shoulder joints.

III. Macro Motion Analysis

There are two main posture categories: arm postures and leg postures. Once our system locates the human object, the macro motion analysis may identify the possible bending posture, arm postures, or leg postures by analyzing the two projection profiles. The macro motion analysis recognizes the possible postures, whereas the micro motion analysis finds the BAPs of the human object. The macro motion analysis consists of (1) bending posture, (2) arm postures, (3) over-head arm-raising postures, and (4) leg postures. We illustrate the 2nd and the 4th processes as follows.

A. Arm Posture Analysis

The arm posture analysis determines 17 different arm postures. Based on the projection profiles of human silhouette in two views, it may decide which arms are raised, and it is forward or sideward. In the façade view, we analyze the $P_{v1}(x_i)$ outside the range of the torso, *i.e.,* $x_i < x_{lb}$ or $x_i > x_{rb}$. "Raised-sideward" means the right/left arm stretches on *x-y* plane, whereas "raised-forward" means the right/left arm stretches on *z-y* plane. Different combinations of arm-raising-leftward/sideward for the left or right arm are used to define different postures (as shown in Table 1).

TABLE 1. DIFFERENT COMBINATIONS OF TWO RAISING ARMS

Posture	Sideward (Right arm)	Sideward (Left arm)	Forward (Right arm)	Forward (Left arm)
0	False	False	False	False
1-a	False	False	True	False
1-b	False	False	False	True
1-c	False	False	True	True
2	False	True	False	False
3	False	True	True	False
4	True	False	False	False
5	True	False	True	False
6	True	True	False	False
7	True	True	True	True

We may determine which hand is raised sideward by analyzing the $P_{v1}(x_i)$ in facade view. If the right arm is raised sideward, we will find $x_l < x_{lb}$. To verify the arm-raising-sideward posture, we find the set of locations of the projectiles as $X_{facade} = \{x_i \mid P_{v1}(x_i) > 0.2W_{torso,1}, x_l \leq x_i \leq x_r\}$. If $\mid X_{facade} \mid \geq 0.5\, W_{torso,1}$, then the right arm is raised sideward. Similarly, we identify whether the left arm is raised sideward by verifying $x_{rb} < x_r$. If the arm is not raised sideward, we may determine whether the arm is raised forward by analyzing the $P_{v2}(x_i)$ and verifying $x_{rb} < x_r$ in the flank view. Similarly, to verify the arm-raising-forward posture, we find the set of projectiles, *i.e.*, $X_{flank} = \{x_i \mid P_{v2}(x_i) > 0.1W_{torso,2}\ ,\ x_l \leq x_i \leq x_r\}$ and then compare $\mid X_{facade} \mid$ with $W_{torso,2}$.

For arm-raising-forward cases, we may have similar $P_{v1}(x_i)$ and $P_{v2}(x_i)$ that makes it very difficult to detect the hand-raise-forward postures. From $P_{v1}(x_i)$, we find the projectiles of both arms as $\{P_{v1}(x_i) \mid x_l \leq x_i \leq x_{lb}\}$ and $\{P_{v1}(x_i) \mid x_{rb} \leq x_i \leq x_r\}$. Then, we compute the centroid of both arms, *i.e.*, $y_{cl}(k)$ and $y_{cr}(k)$ for the *k*th frame. If the position of centroid of arm moving up for a small distance in a period of 5 consecutive frames, *i.e.,* $(y_{cl}(k+j) - y_{cl}(k)) > \theta_{thr}$ and $(y_{cr}(k+j) - y_{cr}(k)) > \theta_{thr}$, where $j=5$, then we determine the left/right arm moves forward. Based on the displacement of the centroid, we may identify the different postures. Similarly, we may detect the arms moving downward by using the same method. Therefore, the three hand raising-forward postures (with similar $P_{v1}(x_i)$ and $P_{v2}(x_i)$ as shown in Figure 2), can be differentiated into three different arm postures.

However, the profile analysis may not identify some hand-raising postures correctly. For *Posture 7* (shown in Table 1), we can find the projectiles outside the torso range in both views, *i.e.*, $\{P_{v1}(x_i) > 0.2W_{torso,1},\ x_{l,1} \leq x_i \leq x_{lb,1}\}$, $\{P_{v1}(x_i) > 0.2W_{torso,1},\ x_{rb,1} \leq x_i \leq x_{r,1}\}$, $\{P_{v2}(x_i) > 0.2W_{torso,2},\ x_{rb,2} \leq x_i \leq x_{r,2}\}$. It indicates that both arms are raised forward and sideward at the same time. To solve this ambiguity, we find the width of projectiles outside the torso range (L_1 and L_3) and compare the ratio between L_1 and L_3 to determine the left shoulder joint angle.

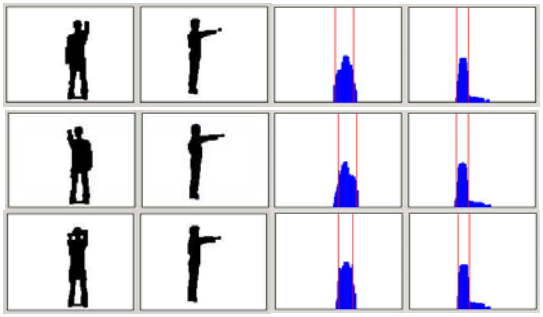


Figure 2. The similar projection profiles are differentiated as Posture 1-a, 1-b, and 1-c.

We find that some transitions between two arm postures are not allowed. We may define all the possible transitions between two arm postures. We assume the variation of joint angles on the *x-y* plane or *y-z* plane is linear so that we can apply the micro motion analysis to track the movement of the arm in terms of the variations of joint angles. If we know the initial and terminating postures, we can decide the motion plane and find out how the arm moves. We consider the arm-raising postures and other postures (*e.g.*, posture 8-15) to construct the posture transition map (Figure 3), in which we show all allowable transitions between every two postures.

Based on the map, the micro motion analysis may tracked arm motion effectively by using Kalman filter.

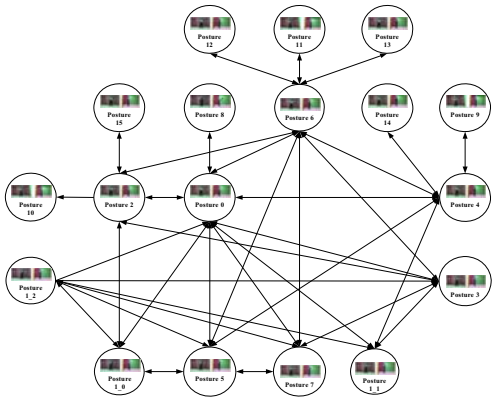


Figure 3. The transition map between two postures.

B. Leg Posture Analysis

We define four kinds of leg postures: two leg lifting (*i.e.*, Postures 16 and 17) and two squatting (*i.e.*, squatting forward or sideward as Postures 18 and 19). To identify squatting, we analyze the $P_h(y_i)$, where $y_t \leq y \leq y_b$ and y_t is the up boundary and y_b is the lower boundary of the profile. First, we find the average of the profile as $\Sigma_i P_h(y_i)/N=avg$, where $N=\mid b\text{-}t \mid$. Second, we scan the $P_h(y_i)$ of the facade view from top to bottom to find the smallest y_j where $P_h(y_j)>avg$. We average the identified location from 10 consecutive frames as the squatting threshold, *i.e.*, $y_{th}=\Sigma_k\, y_j(k)/10$, with $k=1,..10$. We assume that human object does not squat down in the 10 consecutive frames, so that we may find the location of shoulder $y_{shoulder}$ and y_{th}. Then, we may compare $y_{shoulder}$ with y_{th} to determine whether the human object is squatting or not. If $y_{shoulder} > y_{th}+4$, then the squatting is identified, where $y_{shoulder}$ is the position of the shoulder at current frame.

In squatting motion, we may differentiate two different squatting down postures by analyzing the projection profile $P'_{v1,k}(x_i)$ of the lower part of the human silhouette of the kth frame in façade view. We can locate the two peaks located at $p1$ and $p2$ in the $P'_{v1,k}(x_i)$ as $h_{p1}=P'_{v1,k}(x_i)\mid_{x_i=p_1}$ and $h_{p2}=P'_{v1,k}(x_i)\mid_{x_i=p_2}$. First, we find the outer boundary of the $P'_{v1,k}(x_i)$, $b1(k)$ and $b4(k)$, with $P'_{v1,k}(x_i)\mid_{x_i=b1} > 0.8\ h_{p1}$ and $P'_{v1,k}(x_i)\mid_{x_i=b4} > 0.8\ h_{p2}$ and have $d_{\text{width}}(k)=b4(k)-b1(k)$. With 10 projection profiles, $\{P'_{v1,k}(x_i), k=1,\ldots 10\}$, of a standing human object, we calculate the average width $d_2=\Sigma_k\, d_{\text{width}}(k)/10$. Similarly, we compute inner boundary location of $b_2(k)$ and $b_3(k)$, and calculate the average width $d_1 = b_2(k)-b_3(k)$ of $P'_{v1,k}(x_i)$ for squatting sequence. If $d_2>1.5d_1$, we identify the posture as the squatting down (*i.e.*, Posture 19).

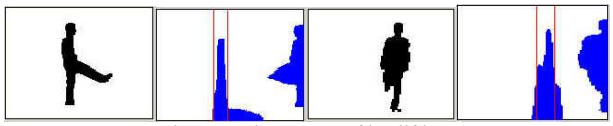


Figure 4. The posture of leg lifting.

To identify leg-lifting posture, we analyze the $P_{v2}(x_i)$ in flank view (as shown in Figure 4). Since the leg is lifted forward, we only analyze the right section of the $P_{v2}(x_i) \mid_{x_i>x_{rb}}$ Similar to arm posture analysis, leg posture analysis finds the leg-lifting, *i.e.*, if $X=\{x_i \mid P_{v2}(x_i)>0.2W_{torso,2},\ x_{rb}\leq x_i\}$ and $\mid X \mid \geq 0.4W_{torso,2}$,, then it finds the leg-lifting posture. After leg-lifting has been identified, we need to determine which leg is lifted.

If the human object is lifting his right leg, he needs to lean leftward to balance his body, as shown in Figure 4. Therefore, we compute the centroid of the profile $P_v(x_i) \mid_{x_{lb}\leq x_i\leq x_{rb}}$, *i.e.*, $x_c=\Sigma_i P_v(x_i)\bullet x_i / \Sigma_i P_v(x_i)$. If the entroid is located at the left side of the torso, *i.e.*, $(x_c-x_{lb}) < (x_{rb}-x_c)$ and $(x_{rb}-x_c)-(x_c-x_{lb})) > \theta_{thres}$, the object is lifting his right leg, else if $(x_{rb}-x_c) < (x_c-x_{lb})$ and $((x_c-x_{lb})-(x_{rb}-x_c))>\theta_{thres}$, he is lifting his left leg. If neither case is found, *i.e.*, $(x_c-x_{lb})\approx(x_{rb}-x_c)$, he is standing. Similar to the arm postures, we can also define the allowable linkage between every two leg postures.

IV. Micro Motion Analysis

To track the movement between the transition of two identified postures we apply the Kalman filtering[3,4] which consists of two groups: *time update equations* and *measure update equations*. The time update equations are responsible for predicting the current state and error covariance estimates to obtain a priori estimates for the next time step, *i.e.*, $\boldsymbol{x}_{k+1} = \mathbf{F}_{k+1,k}\boldsymbol{x}_k + \boldsymbol{w}_k$, where $\mathbf{F}_{k+1,k}$ is the transition matrix taking the state $\boldsymbol{x}_k$ from time k to time $k+1$. The measurement update equations are responsible for the feedback, *i.e.*, for incorporating a new measurement into a prior estimate to obtain an improved posteriori estimate such as $\boldsymbol{y}_k = \mathbf{H}_k\boldsymbol{x}_k + \boldsymbol{v}_k$, where $\mathbf{y}_k$ is the observable at time k and $\mathbf{H}_k$ is the measurement matrix.

In the absence of any observed data at time $k=0$, we may choose the initial estimate of the state. The first task during the measurement update is to compute the Kalman gain, $\mathbf{G}_k$. The next step is to actually measure the process to obtain $\boldsymbol{y}_k$, and then to generate a posteriori state estimate by incorporating the measurement. The final step is to obtain a posteriori error covariance estimate. After each time and measurement update, the process is repeated with the previous a posteriori estimates used to predict the new a priori estimates.

We assume the state vector which represents the joint angles of the human is defined as $\boldsymbol{x}_i[k] =[\theta_i(x, k), \theta_i(y, k), \theta_i(z, k), \Delta\theta_i(x, k), \Delta\theta_i(y, k), \Delta\theta_i(z, k)]^T$, where, $\theta_i(x, k)$, $\theta_i(y, k)$, $\theta_i(z, k)$ represent the angles of the i-th joint in the k-th frame. $\Delta\theta_i(x, k)$, $\Delta\theta_i(y, k)$, $\Delta\theta_i(z, k)$ represent the small joint angle difference between two frames. In our system, we assume the initial value as $\Delta\theta_i(x,0)=\Delta\theta_i(y,0)=\Delta\theta_i(z,0)=5$. We use Kalman filter as a linear prediction-correction filter, of which our purpose is not to solve occlusion, but to predict and correct the joint angles in motion-plane.

To develop the most accurate prediction for the current frame, we construct $\mathbf{F}_{k,k-1}$ based on a simple motion model $\theta(i, k)=\theta(i, k-1)+a\Delta\theta(i, k-1)$ with $\Delta\theta(i, k)=b\Delta\theta(i, k-1)$ where $i = x, y, z$. We assume that if we know the current joint angle, then we can predict the next joint angle by adding a small angle. Meanwhile, we also correct the joint angle difference between two estimates. Using the above motion model, we can construct the transition matrix. After the training phase, a and b are determined and can be used in tracking phase. To observe

every element in state vector, we let the measurement matrix **H** be an identity matrix.

V. Experimental Results

In our experiment, the color image frame is 160 × 120 and the frame rate is 15 frames/sec. The micro motion analysis can track the human motion real time. The human object is allowed to make certain pre-defined articulated human motion, each one consists of a sequence of different postures. As shown in Table 1, there are totally 18 different actions (Action A ~ Action R) tested in our experiment.

In Figures 5-6, we also show the animated human object which does the same action as the real human object. If we move only one arm at one time, we can have good tracking accuracy. we also show the joint angle curves of the arm movement. The green curve represents the joint angle we measured manually. The blue curve represents joint angle we tracked. The red curve represents the errors between the joint angles measured and tracked. We can see the green curve and the blue curve are almost the same. Using Kalman filtering, we may find the tracking results in the following frames will become more accurate.

Here, we find some time lag between the green curve and blue curve that generate the errors in red curve. The phenomenon is due to the posture identification process is non-sensitive to the small movement of the arm. The macro motion analysis does not identify the beginning of the transition of arm posture 1-0 → arm posture 4 in frames 25 but in frame 27.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named "Heading 1", "Heading 2", "Heading 3", and "Heading 4" are prescribed.

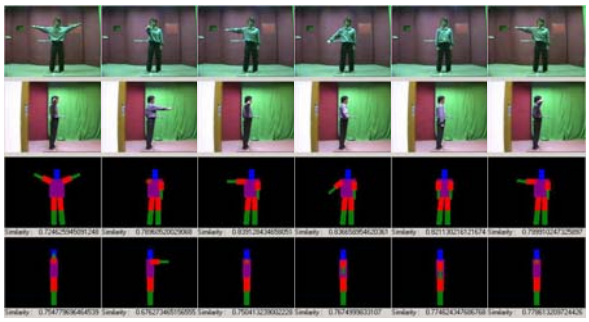


Figure 5. Action A: Posture6→Posture1-0→Posture4→ Posture0→Posture4 (frame 10~frame 54).

If the human object moves both arms at the same time, we find that the Postures 1-c, 3, 5, and 7 are not always identified correctly. For arm postures tracking, if the arm of the human object moves on *xy*-plane, the tracking is accurate, and if it also moves on *zy*-plane simultaneously, the tracking may be inaccurate. Since, we analyze the 2-D silhouette from two views without using any depth information, we cannot estimate any 3-D articulated motion correctly.

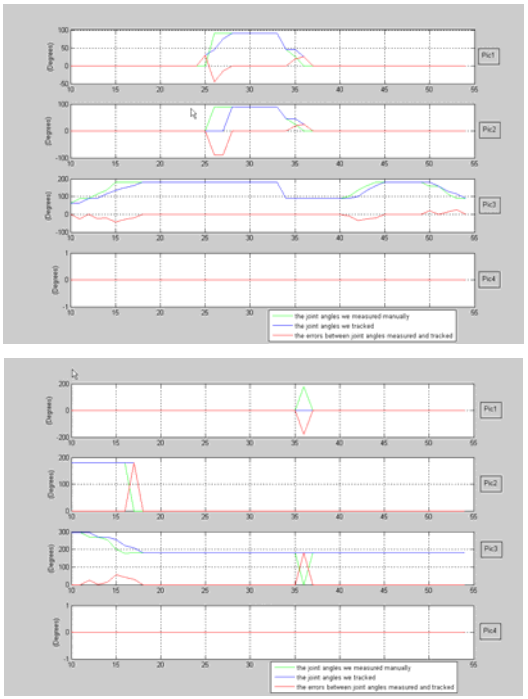


Figure 6. (a) the θx, θy, and θz, of right shoulder and θx of right elbow. (b) the θx, θy, and θz, of left shoulder and θx of left elbow.

References

[1] J. K. Aggarwal and Q. Cai, "Human motion analysis: a review," *Computer Vision and Image Understanding,* Vol. 73, No.3, pp.428-440, 1999.

[2] O. Masoud and N. Papanikolopoulos, "A method for human action recognition," *Image and Vision Computing 21 (2003)*, 729-743

[3] S. L. Dockstader and A. M. Tekalp, "On the tracking of articulated and occluded video object motion," *Real-Time Imaging 7*, 415-432 (2001).

[4] S. L. Dockstader and A. M. Tekalp, "Multiple camera tracking of interacting and occluded human motion," *Proc. IEEE*, vol. 89, no. 10, pp.1441-1455, Oct. 2001.

[5] S. L. Dockstader, M. J. Berg, and A. M. Tekalp, "Stochastic kinematic modeling and feature extraction for gait analysis," IEEE *Transactions on Image Processing*, vol. 12, no. 8, August 2003.

[6] I. Haritaoglu, D. Harwood, and L. S. Davis, "Hydra: Multiple People Detection and Tracking Using Silhouettes," Proc. of the Second IEEE Workshop on Visual Surveillance, June 26, 1999.

[7] I. Haritaoglu, D. Harwood, and L. S. Davis,"Ghost: A human body part labeling system using silhouette," Fourteenth Int. Conf. On Pattern Recognition, Brisbane, Aug. 1998.

[8] C. L. Huang, T. H. Tzeng, and H. C. Shih, "A model-based articulated human motion tracking system," Proc. of 6th *Asian Conf. on Computer Vision*, Jan. 2004, Jeju, Korea.

[9] R. T. Collins, R. Cross, and J. Shi, "Silhouette-based Human Identification from Body shape and Gait," Proc. of IEEE 5th Int. Conf. Automatic Face and Gesture Recognition, FGR'02, 2002.

[10] G. Zhao, R. Chen, G. Liu, and H. Li, "Amplitude Spectrum-based gait Recognition," Proc. of IEEE 6th Int. Conf. Automatic Face and Gesture Recognition, FGR'04, 2004.

[11] S. Iwasawa, J. Ohya, K. Takahashi, and T. Sakaguchi, "Human Body Postures mfrom Triocular Camera Images," Proc. of IEEE 4th Int. Conf. Automatic Face and Gesture Recognition, FGR'00, 2000.