2007 IEEE International Conference on
Robotics and Automation
Roma, Italy, 10-14 April 2007

3-D Human Posture Recognition System Using 2-D Shape Features

Jwu-Sheng Hu, *Member, IEEE*, Tzung-Min Su, *Student Member, IEEE*, and Pei-Ching Lin

Abstract—**This paper presents an integrated framework for recognizing 3D human posture from 2D images. A flexible combinational algorithm motivated by the novel view expressed by Cyr and Kimia [1] is proposed to generate the aspects of 3D human postures as the posture prototype using features extracted from the collected 2D images sampled at random intervals from the viewing sphere. Frequency and phase information of the posture are calculated from the Fourier descriptors (FDs) of the sampled points on the posture contour as the main and assistant features to extract the characteristic views as the aspects. Moreover, a modified particle filter is applied to improve the robustness of human posture recognition for continuous monitoring. Experimental trials on synthetic and real sequences have shown the effectiveness of the proposed method.**

I. Introduction

HUMAN posture recognition is an important step towards human behavior analysis, which can be applied to monitor the behavior of people at home, especially for elders with limited autonomy. However, vision systems suffer from the view angle of the human posture. The simplest description of a human posture is a densely sampled collection of independent views. Postures can be described detail with a large number of collected 2D views. However, the computing time for recognizing objects also grows due to huge searching space. Therefore, the main objective of this study is to propose a framework for 3-D human posture recognition with the efficiency in both modeling and search.

Existing theorems [2] about the human posture recognition can be classified as direct and indirect approach based on the description of human body model and be classified as two-dimensional (2-D) and three-dimensional (3-D) representation based on the dimensional of human body model. The first way, direct approach, consists of a detailed human body model. For example, Ghost [3] constructs a silhouette based body model and combines hierarchical body pose estimation, a convex hull analysis of the silhouette, and a partial mapping from the body parts to the silhouette segments. Furthermore, Pfinder [4] used color information to build a multi-class statistical model and then detects the human body parts with the combination of shape detection. However, occlusions and perspective distortion cause the unreliable results.

The second way, indirect approach, extracts features about the human body instead of detailed human body model and combines classifiers to estimate the human posture. The amount of information in the field of indirect approach are less, but more robust than the information of direct approach. For example, Ozer et al. [5] extracted the AC-coefficients as the features and adopted principal component analysis to be the classifier. Furthermore, Ozer et al. [6] extracted color, edge and shape as the features and adopted the hidden markov model to be the classifier. Besides, complex 3-D models are adopted with different equipments. For example, Delamarre et al. [7] proposed a method to build a 3-D human body via three or more cameras and then calculated the projection of silhouette to compare with the 2-D projection in the database. Moreover, 3-D laser scanners [8] or thermal cameras [9] were also adopted to build the 3-D human body model. However, the above 3-D solutions suffer from enormous computing time requirement and high device cost.

In order to reduce the cost and computing time, some methods have been studied to extract a minimal set of object views with a single camera. Aspect-graph representations, which focus on changes in the shape of the projection of the object is one kind of method to achieve the minimal set of object views. The underlying theory that describes 3D objects by aspect-graph was proposed by Koenderink et al. [10]. The traditional aspect-graph method [11] bases on an assumption that an object belonging to limited classes of shapes and characteristic views can be extracted using prior knowledge of the object. In our previous work [12], a similarity-based aspect-graph approach is proposed to recognize free-form objects using the frequency information of FDs and point-to-point lengths. For recognizing human posture in this work, phase information of FDs is adopted to avoid misclassifying human postures between similar body shapes and a modified similarity-based aspect-graph approach is proposed to improve the efficiency of recognizing human posture. The proposed method is applied to video data where a sequence of human motion is observed. To improve the robustness, a modified particle filter is applied as the post-processing stage to reduce the errors caused from the inaccuracy of foreground detection. This method is also useful in monitoring continuous behavior of a person.

The rest of the paper is organized as follows: Section II describes the procedure of extracting the frequency and phase information of FDs . Next, Section III describes the novelty of this work, the procedures of building a modified

Manuscript received September 15, 2006. This work was supported by National Science Council of the R.O.C. under grant no. NSC94−2218−E009064 and DOIT TDPA Program under the project number 95-EC-17-A-04-S1-054.

J. S. Hu is with Department of Electrical and Control Engineering, National Chiao Tung University, Hsinchu 300, Taiwan, ROC. (phone: 886-3-5712121 ext. 54424; e-mail: e-mail: jshu@cn.nctu.edu.tw).

T. M. Su and P. C. Lin are with Department of Electrical and Control Engineering, National Chiao Tung University, Hsinchu 300, Taiwan, ROC. (tzungminsu@ieee.org and peiching.ece93g@nctu.edu.tw).

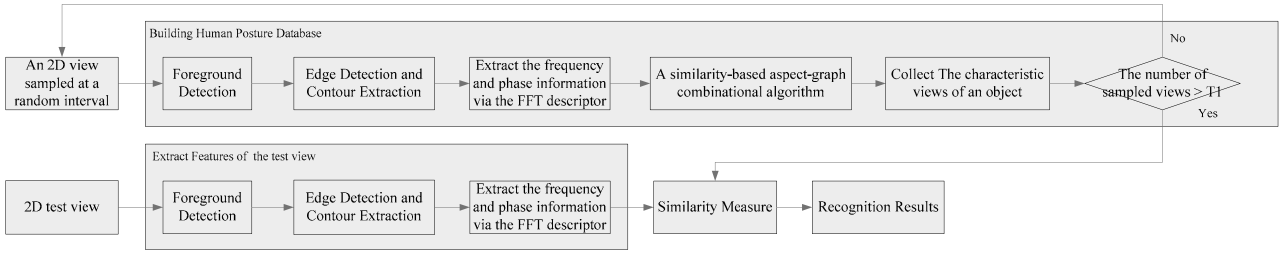

Fig. 1. The flowchart of the proposed method

similarity-based aspect-graph representations and recognizing 3-D objects using 2-D object views. Furthermore, the modified particle filter is described. In Section IV, some experimental results are presented to demonstrate the performance of the proposed method in 3-D human posture recognition. Conclusions are finally made in Section V.

II. Feature Extraction

A. Foreground Detection and Contour Extraction

In this work, the shape feature is used to measure the similarity between two object views. In order to extract the shape information from the foreground object, Canny edge detection [13] is applied to extract the shape edge and Gradient Vector Flow Snake (GVF) [14] is then applied to extract the contour information. Suppose the contour information is included in a set Z, which is composed of N points z_i, where z_i can be described as a complex form with (1).

$$\mathbf{z} = \{z(i)\} = \{x_i + jy_i\},\ 0 \le i \le N \quad (1)$$

B. Fourier Descriptors

In order to avoid the variations in shift and scale, the points inside the set $\mathbf{z}$ are re-sampling by (2).

$$\tilde{\mathbf{Z}} = \{\tilde{z}(i)\} = \{L_c[(x_i - x_c) + j(y_i - y_c)]/L\} \quad (2)$$

where $0 \le i \le N$, L means the contour length, L_c means the expected contour length and (x_c, y_c) means the contour center .

Then the Fourier transform is applied on $\tilde{\mathbf{Z}}$ to calculate the FDs with (3).

$$\tilde{Z}(k) = \sum_{n=0}^{N-1} \tilde{z}(n)\exp(-j2\pi kn/N),\ \text{k} = 0,1,2,...,\text{N-1} \quad (3)$$

The magnitude parts of low frequency information in $\tilde{Z}(k)$ are extracted to describe the human posture without the variations on the high-frequency noises and are defined as FD_m . The method for extracting FD_m is described as (4).

$$FD_m = \{|\tilde{Z}(k)|, |\tilde{Z}(N-k)|,\ 1 \le k \le T_2\} \quad (4)$$

However, although FD_m describes the shape of human posture, the direction information of human posture is lost without phase information of $\tilde{Z}(k)$. For example, the postures of slide standing and lying down are classified as the same posture easily using only FD_m . Therefore, for improving the efficiency of classifying human postures, the phase information in $\tilde{Z}(k)$ is adopted in this work to solve the above problems. Suppose the phase information is θ_z, and θ_z can be calculated using $\tilde{Z}(1)$ and $\tilde{Z}(N-1)$, which is derived from the work of [12]. The representation of $\tilde{Z}(1)$ and $\tilde{Z}(N-1)$ are described as (5) and (6), and θ_z can be calculated using (7).

$$\tilde{Z}(1) = |\tilde{Z}(1)|.\exp(j\theta_1) = R_1 + jI_1 \quad (5)$$

$$\tilde{Z}(N-1) = |\tilde{Z}(N-1)|.\exp(j\theta_{N-1}) = R_{N-1} + jI_{N-1} \quad (6)$$

$$\theta_z = (\theta_1 + \theta_{N-1})/2 = (arc\tan(I_1/R_1) + arc\tan(I_{N-1}/R_{N-1}))/2 \quad (7)$$

where R_1 and R_{N-1} means the real parts of $\tilde{Z}(1)$ and $\tilde{Z}(N-1)$, I_1 and I_{N-1} means the imaginary parts of $\tilde{Z}(1)$ and $\tilde{Z}(N-1)$, and θ_1 and θ_{N-1} means the phase of $\tilde{Z}(1)$ and $\tilde{Z}(N-1)$.

III. Similarity-Based Aspect-Graph

A. Similarity Function

In order to calculate the similarity between human postures, a similarity measure metric is necessary to apply on the extracted features. Suppose the magnitude parts of low frequency information extracted from the test image and database are FD_m^T and FD_m^D , and the phase information extracted from the test image and database are θ_z^T and ϑ_z^D Then the similarity between FD_m^T and FD_m^D is calculated using one-norm distance, which is described as (8) and the similarity between θ_z^T and ϑ_z^D is calculated using one-norm distance, which is described as (9).

$$d_F = \sum_{k=1}^{T_2} |FD_m^T(k) - FD_m^D(k)| + |FD_m^T(N-k) - FD_m^D(N-k)| \quad (8)$$

$$d_p = |\theta_z^T(k) - \theta_z^D(k)| \quad (9)$$

B. Generation of Aspects and Characteristic Views

In our previous work [12], a similarity-based aspect-graph approach, are proposed to extract the characteristic views of complex free-form objects via two features, FD_m and point-to-point lengths. However, point-to-point length is not suitable for human posture recognition due to the rotation invariant, huge computing time and appearance variations. In

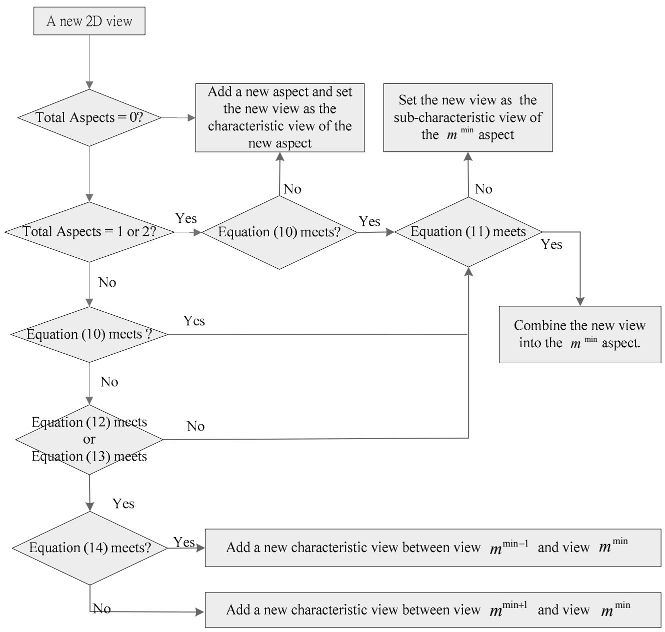

Fig. 2. The procedure of the proposed combinational algorithm

this work, phase information of FDs θ_z is adopted to replace point-to-point lengths to avoid misclassifying human postures between similar body shapes. Moreover, a modified similarity-based aspect-graph approach is proposed to improve the efficiency of human posture recognition with FD_m and θ_z. One or one above characteristic view can be combined into the same aspect to maintain the differences between similar human shapes via keeping the phase information in the database.

Suppose V_{new}^n means the new sampled view of the n_{th} object, $C_m^n(i)$ means the i_{th} characteristic view of the m_{th} aspects of the n_{th} object, $C_{m^{\min}-1}^n$ and $C_{m^{\min}+1}^n$ means the neighbor views of $C_{m^{\min}}^n$ that has the minimum distance with V_{new}^n, $A_{m^{\min}}$ means the aspects that has the minimum distance with V_{new}^n, where $m^{\min}$ means the index of $A_{m^{\min}}$. Then four steps are imposed to form aspects and characteristic views as Step A-1 to A-4 and the flowchart of the modified aspect-graph representation is illustrated as Fig. 2.

Step A-1:

When the number of existed aspects of the n_{th} object equals zero, V_{new}^n is regarded as a characteristic view of a new aspect.

Step A-2:

When the number of existed aspects of the n_{th} object equals one or two,

(A-2.1) If (10) and (11) both meet, V_{new}^n is combined into the $m^{\min}$ aspect and the characteristic view of the keep the same,

(A-2.2) Otherwise, if (10) meets but (11) conflicts , V_{new}^n is combined into the $m^{\min}$ aspect and is regarded as a new characteristic view of the $m^{\min}$ aspect.

(A-2.3) Otherwise, if (10) and (11) both conflict, a new aspect of the n_{th} object is built, and V_{new}^n is regarded as the new characteristic view of the new aspect.

$$\min_{all\ C_m^n \in A_{m^{\min}}} d_F\left(V_{new}^n, C_m^n\right) < T_3 \tag{10}$$

$$\min_{all\ C_m^n \in A_{m^{\min}}} d_p\left(V_{new}^n, C_m^n\right) < T_5 \tag{11}$$

where T_3 and T_5 are both predefined threshold value.

Step A-3:

When the number of existed aspects of the n_{th} object equals three or above three,

(A-3.1) If (12) or (13) meet and (11) conflicts, a new aspect is built up and V_{new}^n is regarded as the characteristic view of the new aspect.

(A-3.2) Otherwise, if (12) and (13) both conflict and (11) meets, V_{new}^n is combined into the $m^{\min}$ aspect and the characteristic view of the $m^{\min}$ aspect keeps the same.

(A-3.3) Otherwise, if (12) and (13) both conflict and (11) conflicts, V_{new}^n is combined into the $m^{\min}$ aspect and is regarded as a new characteristic view of the $m^{\min}$ aspect.

$$\min_{all\ C_m^n \in A_{m^{\min}}} d_F\left(V_{new}^n, C_m^n\right) > T_4 \tag{12}$$

$$T_3 < \min_{all\ C_m^n \in A_{m^{\min}}} d_F\left(V_{new}^n, C_m^n\right) < T_4 \quad and \quad d_F\left(V_{new}^n, C_{m^{\min}\pm 1}^n\right) > T_4 \tag{13}$$

Moreover, if a new aspect is built, the aspect order can be decided using (14). If the similarity distance between V_{new}^n and $C_{m^{\min}+1}^n$ is larger than the similarity distance between V_{new}^n and $C_{m^{\min}-1}^n$, the new aspect is inserted between aspect $m^{\min}$ and aspect $m^{\min-1}$. Otherwise, the new aspect is inserted between aspect $m^{\min}$ and aspect $m^{\min+1}$. Therefore, the similar aspects are close to each other.

$$d_F\left(V_{new}^n, C_{m^{\min}+1}^n\right) > \ d_F\left(V_{new}^n, C_{m^{\min}-1}^n\right) \tag{14}$$

Besides, T_3 and T_4 are two predefined threshold values and $T_4 > T_3$. If T_3 and T_4 are defined with both small values, the criterion of combing 2D views becomes strict and thus the aspect number becomes more numerous. Moreover, if the difference between T_3 and T_4 becomes smaller, the tolerance of difference between 2D views inside an aspect becomes smaller and thus aspect number becomes more numerous.

C. Posture Recognition using 2D Characteristic Views

After building the aspects-graph representation of each human posture in the database, a test view of an unknown posture can be recognized using the similarity measure with the main features and the assistant features. Two steps are imposed as follows:

Step B-1:

The test view of an unknown human posture is compared with the characteristic views of the database via main features. Then, the first T_6 2D characteristic views in the database having the smallest similarity distance with the test 2D view via main features are preserved to be further recognized.

Step B-2:

Suppose A_{T_6} is defined as the set that contains the T_6 2D characteristic views described at the Step B-2 , then the final similarity distance can be calculated with the assistant features by (15).

$$d(V_j^i, C_m^n) = \sqrt{(d_p((V_j^i, C_m^n))^2 + (90 \cdot (L_{main} / L_{Max})^2} \quad (15)$$

where V_j^i means the 2D view of an unknown human posture, C_m^n denotes the m_{th} characteristic view of the n_{th} human posture in the database , L_{main} denotes the similarity distance calculated using FD_m between the unknown human posture and the m_{th} characteristic view of the n_{th} human posture, which is defined as $L_{main} = d_F(V_j^i, C_m^n))$, where $C_m^n \in A_{T_6}$, and $L_{Max} = \arg \max_{all\ C_m^n \in A_{T_6}} (d_F(V_j^i, C_m^n))$.

D. Particle Filter

For dealing with human behavior analysis, combining the temporal information from the video sequences with the results of human posture recognition is adopted in this work. A modified particle filter is proposed from the novel view expressed by Kwartra et al. [16] to compute the statistics for the distribution of each human posture and then the statistics are used to estimate a confidence measure for each posture.

The particle filter uses particles to represent the posteriori possibility distribution. Suppose the state of each time-instant t is the discrete posture state x_t, the measurement of each time-instant t is z_t, and the measurement in a period of time t is $Z_t = \{z_1, z_2, ..., z_t\}$. Then $p(x_t \mid Z_t)$ is defined as the posteriori density probability given the measurement Z_t , $p(x_t \mid Z_{t-1})$ is defined as the priori probability, $p(x_t \mid x_{t-1})$ is defined as the process density probability describing the dynamics, and $p(z_t \mid x_t)$ is defined as the observation density probability.

Furthermore, a set of M samples are defined as S_M to represent $p(x_t \mid Z_t)$, which is described as (16). Each sample s_k consists of a weighting parameter π_k , a accumulating parameter a_k and a class parameter c_k. The weighting of each sample is initialized as $1/M$ and each class has equal samples.

$$S_M = \{s_k = (\pi_k, c_k, a_k), k = 1,2,3,...,M\} \quad (16)$$

Moreover, $p(z_t \mid x_t)$ can be calculated using the similarity distance between z_t and the images of database via a zero-mean Gaussian distribution. A discrete state translation matrix T is being substituted for $p(x_t \mid x_{t-1})$.

The following four steps are the basic steps of the modified particle filter proposed in this work and are implied to generate a probabilistic estimation at each time-instant.

Step C-1: (Prediction)

(1) Calculate $p(z_t \mid x)\ \forall x \in X$ (X :the set of all postures in the database)

(2) Calculate $p(x_t \mid Z_{t-1}) = T_{ij}^{'} = p(z_t \mid x = j).T_{ij}$, and let $\sum_j T_{ij} = 1$.

(3) Calculate the class parameter c_k^t for each sample s_k by (17)

$$c_k^t = \max_{\forall j}(T_{ij}^{'} \mid i = c_k^{t-1}) \quad (17)$$

Step C-2: (Update)

Update the weighting parameter π_k^t and the accumulating parameter a_k^t for each sample s_k by (18)-(19)

$$\pi_k^t = p(z_t \mid x_t = c_k^t) \text{ and let } \sum_k \pi_k^t = 1. \quad (18)$$

$$a_0^t = 0, a_k^t = a_{k-1}^t + \pi_k^t, k = 1,2,3,...,M \quad (19)$$

Step C-3: (Output):

Calculate the summation of the posteriori density probability $p(x \mid Z_t)$ for each posture class of the database and select the posture having the maximum value as the estimated posture, which is described as (20).

$$\hat{x} = \arg \max_{\forall j} p(x_t = j \mid Z) = \arg \max_{\forall j} \sum_{k \in \gamma_j} \pi_k^t \quad (20)$$

where $\gamma_j = \{k \mid c_k^t = j\}$

Step C-4: (Selection):

Relace $p(x \mid Z_{t-1})$ by $p(x \mid Z_t)$ and approximate $p(x \mid Z_t)$ by resampling the samples of S_M

(1) Generate a random valuable $r \in [0,1]$ by uniform distribution.

(2) For each sample, find the minimum integer j that meets $a_j^t \geq r$ and assign $c_k^{t+1} = c_j^t, k = 1,2,3,...,M$

IV. Experimental Results

This section describes several experiments that demonstrate the effectiveness of the proposed method. SONY EVI-D30 PTZ camera is used to capture the video sequences containing human postures. Fig.3(a) shows the indoor environment and Fig.3(b) shows a sample frame of video sequences. Eight basic human postures are built in the database, and are illustrated in the Fig. 4. The training views of human postures are captured with 5-degree increment intervals and are collected as W_d^n, where contains 72 views for each human posture. The extra views of each object are captured from the trisection-points between each 5-degree point and are collected as W_t^n, where contains 216 views for each human posture. The descriptions of the captured views are list as (21)-(22)

$$W_d^n = \{W^n(d)\}, 1 \leq n \leq 8,\ 1 \leq d \leq 72 \quad (21)$$

$$W_t^n = \{W^n(d)\}, 1 \leq n \leq 8,\ 1 \leq t \leq 216 \quad (22)$$

Furthermore, in the following experiments, T_1 is defined as 72, as and T_2 is defined as 20, T_3 is defined as 1450, T_4 is

Fig. 3. The indoor environment for capturing the test videos of human posture recognition and human behavior analysis

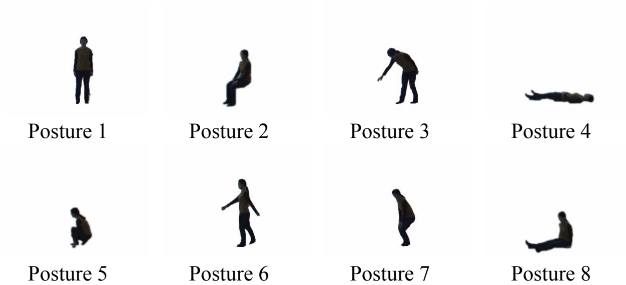

Fig. 4. The image database that contains 8 3-D human postures

defined as 1800, T_5 is defined as 10 and T_6 is defined as 10. The computing time of calculating the similarity between a test view and a view in the database is about 0.004 seconds with P4 2.8G CPU and 512MB RAM.

A. 3D human posture recognition using 2D views via FDs

In the first experiment, the efficiency of the modified aspect-graph representation using 2D views is demonstrated with synthetic video sequences, where the background was removed manually and automatically using the proposed method in our previous work [17]. The 2D views in W_d^n were used for training the database and the 2D views in W_t^n were used for estimating the performance. Table I and II shows the information about the mean of the aspect numbers, the Top1, Top2 and Top3 matching rates, where the number of aspects of each human posture is fewer than the training views. Although the matching rates decrease while adopting automatic background subtraction, the Top3 matching rates are still above 90%. Moreover, the computing time for recognizing objects is reduced using fewer aspects.

B. 3D human posture recognition using 2D views via FDs and the modified particle filter

In the second experiment, the modified particle filter is applied to combine the temporal information and the human recognition result using FDs. Table III shows the human posture recognition results after applying the modified particle filter. The recognition results are all better than the Top1 matching rates listed in Table I and II. Moreover, Table IV shows the discrete state translation matrix T, which is applied for replace the process density probability distribution $p(x_t \mid x_{t-1})$. Suppose the column index i means the previous state x_{t-1} and the row index j means the present state x_t. Then the criterion for setting the discrete state translation matrix T_{ij} is list below:

TABLE I
THE RESULTS OF HUMAN POSTURE RECOGNITION USING 2D VIEWS VIA FDS WITH BACKGROUND SUBTRACTION MANUALLY

Result	The Index of Postures								
	1	2	3	4	5	6	7	8	A.*
Aspect	8	25	37	41	42	38	8	38	29.6
Top 1(%)	94.9	99.1	98.2	100	99.1	96.3	99.5	100	98.4
Top 2(%)	99.1	99.5	100	100	100	99.5	100	100	99.8
Top 3(%)	100	99.5	100	100	100	99.5	100	100	99.9

*: A. means the average number of aspect

TABLE II
THE RESULTS OF HUMAN POSTURE RECOGNITION USING 2D VIEWS VIA FDS WITH BACKGROUND SUBTRACTION AUTOMATICALLY

Result	The Index of Postures								
	1	2	3	4	5	6	7	8	A.*
Aspect	8	25	37	41	42	38	8	38	29.6
Top 1(%)	90.8	84.9	89.4	86.1	96.3	93.5	81.5	94.0	89.6
Top 2(%)	96.8	85.5	93.5	87.5	97.2	97.7	93.1	95.4	93.3
Top 3(%)	99.5	90.2	96.3	90.2	98.6	97.7	95.4	98.2	95.8

TABLE III
THE RESULTS OF HUMAN POSTURE RECOGNITION USING 2D VIEWS VIA FDS AND THE MODIFIED PARTICLE FILTER

Result	The Index of Postures								
	1	2	3	4	5	6	7	8	A.*
C1**(%)	100	100	99.5	100	100	98.6	100	100	99.8
C2**(%)	94.0	92.2	90.7	95.8	97.2	94.9	90.7	94.0	93.7

**: C1 and C2 means the recognition results with background subtraction manually and automatically

(1) When $i = j$, the coefficient of T_{ij} is given as the highest value, 10.

(2) When $i \neq j$, the coefficient of T_{ij} is given by (23), where T_s means the number of translation steps from state i to state j.

$$\mathrm{T}_{ij} = 10 - T_s * 2 \tag{23}$$

(3) Adjust the coefficients of T_{ij} with ± 1 by the similarities between each state manually.

(4) Normailze the discrete state translation matrix T_{ij} and let $\sum_j \mathrm{T}_{ij} = 1$.

C. Human behavior analysis via FDs and particle filter

In the final experiment, one real video sequence is performed using the proposed method in this work to analyze the human behavior. Fig. 5 shows the results of human behavior analysis and each figure contains three figures, where the top figure means the result with only FDs, the middle figure means the result with FDs and the modified particle filter, and the bottom figure means the results decided manually. The video sequence listed in Fig. 5 has total 300 frames and represent the human behavior of “Stand, lie down and then standing up”. From the observation of Fig. 5, some misclassifications happened while the extracted posture contour is not accurate enough to classify the captured human posture, e.g. at the 700th frame, the seventh posture (sit-to-stand) was regarded as the second posture (sit), and at the 870th frame, the eighth posture (sit on the floor) was

TABLE IV
THE DISCRETE STATE TRANSLATION MATRIX T

	1	2	3	4	5	6	7	8
1	0.24	0.15	0.22	0.04	0.03	0.19	0.14	0.025
2	0.16	0.25	0.11	0.04	0.15	0.15	0.16	0.125
3	0.16	0.09	0.26	0.04	0.03	0.19	0.10	0.025
4	0.02	0.02	0.02	0.38	0.10	0.02	0.02	0.20
5	0.02	0.15	0.02	0.15	0.26	0.02	0.16	0.20
6	0.19	0.12	0.22	0.04	0.03	0.24	0.12	0.025
7	0.19	0.20	0.13	0.04	0.20	0.17	0.20	0.15
8	0.02	0.02	0.02	0.27	0.20	0.02	0.10	0.25

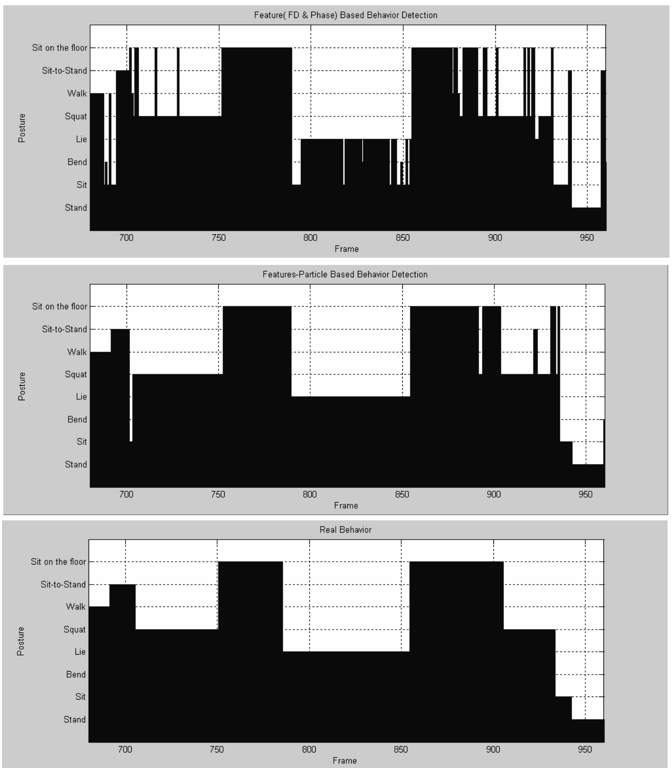

Fig.5. Human behavior analysis: Stand, lie down and then standing up

regarded as the fifth posture (squat), etc. However, applying particle filter to human behavior analysis is demonstrated to be useful with real video sequences.

V. CONCLUSIONS

This work proposes an integrated framework for recognizing 3D human postures with 2D images. Frequency and phase information of the posture are calculated from the FDs of the sampled points on the posture contour as the main and assistant features to extract the characteristic views as the aspects. A modified aspect-graph representation from our previous work [12] is proposed to improve the efficiency in human posture recognition. Experimental trials on synthetic and real video sequences have shown the effectiveness of the proposed method. Moreover, the computing time for recognizing an unknown object increases while more human postures are adopted in the database. The real-time issue is still the future work of this study. Besides, although the integrity of the database increases while more 2-D views of a posture are collected, the extracted characteristic views of a postures are limited with the consideration of similarity measure between 2-D views. Furthermore, for combing the temporal information between video frames, a modified particle filter is proposed to improve the efficiency in human behavior analysis. A real video sequence that contains human behavior of "Stand, lie down and then standing up" is used to demonstrate the effectiveness of the modified particle filter. The experimental results show that the proposed framework is efficient in both human posture recognition and human behavior analysis.

REFERENCES

[1] C. M. Cyr and B. Kimia, "A Similarity-Based Aspect-Graph Approach to 3D Object Recognition," *in Int'l Journal of Computer Vision*, vol. 57, no. 1, pp. 5–22, 2004.

[2] R. Cucchiara, C. Grana, A. Prati, and R. Vezzani, "Probabilistic Posture Classification for Human-Behavior Analysis," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 35, no.1, January 2005.

[3] I. Haritaoglu ,D. Harwood and L. S. Davis, "Ghost : A Human Body Part Labeling System Using Silhouettes," *in 14th Int'l Conf. on Pattern Recognition*, Brisbane,Australia,1998.

[4] C. Wren ,A. Azarbayejani , T. Darrell and A. Pentland, "Pfinder : Real-Time Tracking of the Human Body," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no 7, pp. 780-785, 1997.

[5] L.B.Ozer, W.Wolf, "Real-time posture and activity recognition", *in Proc. IEEE Workshop, Motion and Video Computing*, pp. 133–138, Dec. 2002.

[6] L.B.Ozer, T. Lu and W. Wolf, "Design of a real-time gesture recognition system: high performance through algorithms and software" *IEEE Signal Processing Magazine*, vol. 22, pp. 57-64, May 2005.

[7] Q. Delamarre and O. Faugeras, "3D Articulated Models and Multi-View Tracking with Silhouettes," *Proc. 17th Int"l Conf. Computer Vision*, pp. 716–721, Sept. 1999.

[8] N. Werghi and Y. Xiao, "Recognition of Human Body Posture from a Cloud of 3d Data Points using Wavelet Transform Coefficients," *IEEE Int'l Conf. on Automatic Face and. Gesture Recognition*, pp. 70–75, 2002.

[9] S. Iwasawa, K. Ebihara, J. Ohya and S. Morishima, "Real-time human posture estimation using monocular thermal images," *IEEE Int'l Conf. on Automatic Face and Gesture Recognition*, pp.492-497, 1998.

[10] J.J. Koenderink, and A.J. van Doorn, "The singularities of the visual mapping," *Biol. Cyber.*, vol.24, pp. 51-59, 1976.

[11] I. Shimshoni, J. Ponce, "Finite-Resolution Aspect Graphs of Polyhedral Objects," *IEEE Trans. on Pattern Anal. Mach. Intell.*, pp.315-327, 1997.

[12] J.S. Hu, T.M. Su, and C.C. Lin, "Shape Memorization and Recognition of 3-D Objects Using a Similarity-Based Aspect-Graph Approach," *IEEE Int'l Conf. on Systems, Man, and Cybernetics*, Oct. 2006.

[13] J. Canny, "A Computational Approach to Edge Detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 8, no.6, pp. 679-698, 1986.

[14] C. Xu and J. L. Prince, "Gradient Vector Flow: A New External Force for Snakes," *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 66-71, 1997.

[15] A. Folkers and H. Samet, "Content-based Image Retrieval Using Fourier Descriptors on a Logo Database," *Proceedings of the 16th Int'l Conf. on Pattern Recognition*, vol. 3, pp. 521-524, August 2002.

[16] V. Kwatra, A. Bobick, A. Johnson, "Temporal integration of multiple silhouette-based body-part hypotheses," *IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, pp. 758-764, Dec. 2001.

[17] J.S. Hu, T.M. Su and S.C. Jeng, "Robust Background Subtraction with Shadow and Highlight Removal for Indoor Environment Surveillance," *IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems*, Oct. 2006.