

유비쿼터스 로봇과 휴먼 인터랙션을 위한 제스처 추출

Gesture Extraction for Ubiquitous Robot-Human Interaction

김 문 환, 주 영 훈*, 박 진 배
(Moon Hwan Kim, Young Hoon Joo, and Jin Bae Park)

Abstract : This paper discusses a skeleton feature extraction method for ubiquitous robot system. The skeleton features are used to analyze human motion and pose estimation. In different conventional feature extraction environment, the ubiquitous robot system requires more robust feature extraction method because it has internal vibration and low image quality. The new hybrid silhouette extraction method and adaptive skeleton model are proposed to overcome this constrained environment. The skin color is used to extract more sophisticated feature points. Finally, the experimental results show the superiority of the proposed method.

Keywords : skeleton model, temporal gradient, spatial gradient, hybrid silhouette, active contour

I. Introduction

Human motion analysis is important research subject for man-machine interface. It is receiving increasing attention for a wide spectrum of applications, such as man-machine interface, security surveillance, image retrieval and video indexing, and ubiquitous robot [1]. Especially, it is researched as a new method of human computation interaction (HCI) because human motion analysis concerns with the key techniques of HCI such as pose recognition and motion tracking. To analysis human motion, extracting features of human body from sequential image plays an important role. Without features from human body, it is not easy to analyze human motion.

Human motion analysis can be performed in two approaches. At first, the features of human body are acquired by capturing both position and motion information via sensors fixed on human joints. The sensors fixed on the human joints send major information to a main computer, and then main computer analyze the major information then give the motion paymasters of human body. Without sensor, the color marks can be used instead of sensors. The main advantage of this approach is that it gives accurate features for motion analysis. However, it is not adapt to use in ubiquitous robot system or general situation because it needs sensing equipment and restrict environment.

Another approach extract features of human body from sequential images contain human body. It has been receiving extensive attention from computer vision and HCI researchers because it need only one camera and does not need main computer. Acquiring sequential images is performed in two ways as camera motion. The first one is to use camera fixed on environment. It gives sequential images including motionless background and human motion. The second one is to use camera fixed on moving object such as ubiquitous robot. This method gives sequential images including background and human motion. Most of researched is based on the first method to capture sequential images. In this paper, however, we suppose the sequential image is

acquired from camera fixed on ubiquitous robot which does not move. The sequential images from the supposed environment have following restrictions: 1) there exist background motion because ubiquitous robot has internal vibration. 2) The resolution of image is low. 3) It is not easy to accurate background model. 4) The position of camera is lower than human head. 5) The illumination condition is poor.

It is known that the pose and motion of a human body can be determined by the position and motion information of the joints [2]. How to obtain the position of joint from the sequential image is important to analysis human motion. However, most of the existing method does not consider restrictions occurring ubiquitous robot system. Therefore, it is need to develop a new method to extract key information from images restricted by ubiquitous robot system.

The objective of this work is to give position of the joint in the image as skeleton features. In this paper, we presented a method of skeleton feature extraction method from sequential images restricted by ubiquitous robot system. First, the hybrid silhouette generation method is considered to extract accurate silhouette which is robust to background motion. Then the adaptive skeleton model is applied to extract features of human body. The energy function is defined to find precise features in human body. Finally, color based hand detection method is used to compensate hand position of skeleton model.

In remainder of this paper, we discuss hybrid silhouette generation method in Section 2, and in Section 3 the adaptive skeleton model and energy functions are described. The color based hand detection method is introduced in Section 4. In Section 5 we illustrate the experimental results. Finally, we conclude our paper and give some remarks in Section 6.

II. Hybrid silhouette extraction method

The extraction of skeleton feature is based on silhouette of human body. There are three conventional approaches to silhouette generation: temporal differencing (two-frame or three frame) [7], background subtraction [3-6], and optical flow (see [8] for an excellent discussion).

The temporal differencing is very adaptive to dynamic environments, but generally does a poor job of extracting all relevant feature pixels. The background subtraction provide accurate silhouette of human body, but is extremely sensitive to change of image due to lighting and extraneous events. Optical flow can be used to detect

* 책임저자(Corresponding Author)

논문접수 : 2005. 9. 15., 채택확정 : 2005. 10. 25.

김문환, 박진배 : 연세대학교 전기전자공학과

(jmcs@control.yonsei.ac.kr/jbpark@control.yonsei.ac.kr)

주영훈 : 군산대학교 전자정보공학부(yhjoo@kunsan.ac.kr)

※ This work is supported by a project "Ubiquitous Robotic Companion (URC)," and was funded from the Ministry of Information and Communications (MIC) Republic of Korea.

features from image with background motion, but most optical flow computation methods are very complex and are inapplicable to embedded system such as ubiquitous robot system. However, these conventional approaches are not adaptable to extract silhouette from images restricted within ubiquitous robot system.

The hybrid silhouette extraction method is based on temporal differencing method. To overcome the poor extraction of relevant silhouette and restriction of ubiquitous robot system, we propose motion region model and compensation method for missing silhouette information.

1. Hybrid silhouette extraction

In this subsection, we will discuss about hybrid silhouette extraction method. Let sequential image is define as

$$I(x, y; t) = [I_r(x, y; t), I_g(x, y; t), I_b(x, y; t)]^T = [p(t)]^T. \quad (1)$$

Temporal gradient is temporal difference between consecutive two images. The temporal gradient and the spatial gradient including edge information are defined as

$$I_t = \frac{\partial I}{\partial t} \quad (2)$$

$$I_s = \left[\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right].$$

When we want to gradient information as key value of edge, gradient information should be mapped into natural number. In this approach, we simply use following mapping strategy,

$$\bar{I}_t = \|I_t\| \quad (3)$$

$$\bar{I}_s = \|I_s\| \quad (4)$$

In ubiquitous robot system, the temporal gradient \bar{I}_t can have three kind of undesired values: 1) \bar{I}_t can disappear because of camera problem. 2) \bar{I}_t has insufficient silhouette information. 3) \bar{I}_t has unnecessary motion information due to internal vibration of ubiquitous robot. To overcome first undesired situation, we should check sum of temporal gradient S_t defined as

$$S_t = \int_{\forall p \in I} f(\bar{I}_t, \gamma_t) dp \quad (5)$$

$$f(\bar{I}_t, \gamma_t) = \begin{cases} 1, & \text{If } \bar{I}_t > \gamma_t \\ 0, & \text{else} \end{cases}$$

where γ_t is minimum value to transform color image as binary image.

When S_t is zero, the old temporal gradient is is replace to current temporal gradient.

The second and third undesired situation can be solved by using motion region model. To remove unnecessary motion information and complete insufficient silhouette information, we should use spatial gradient and history information of temporal gradient. History

information of temporal gradient is the sum of consecutive some old temporal gradients. How to add spatial gradient and history information is key problem.

The convex sum of temporal and spatial gradient can generate reliable accurate silhouette. The convex sum denotes as

$$\eta \bar{I}_s(x, y) + (1 - \eta) \bar{I}_t(x, y) \quad (6)$$

where η is the convex sum parameter. By adjusting η , we can get accurate silhouette or imprecise silhouette. Generally, the convex sum parameter has static value. It is determined by experimental or manual method, generally. In this paper, but, convex parameter is determined by motion region model.

The motion region model is defined as $R \subset \mathbb{R}^2$. The initial motion region model has zero. Then the motion region mode is updated as following,

$$R(x, y; t+1) = \begin{cases} R(x, y; t) + \gamma_i, \bar{I}_i(x, y; t+1) > \gamma_i \\ R(x, y; t) + \gamma_d, \bar{I}_i(x, y; t+1) < \gamma_i \end{cases}$$

$$\gamma_i = \begin{cases} \bar{\gamma}_i, \bar{I}_s(x, y; t+1) > \gamma_s \\ \underline{\gamma}_i, \bar{I}_s(x, y; t+1) < \gamma_s \end{cases} \quad (7)$$

$$\gamma_d = \begin{cases} \bar{\gamma}_d, \bar{I}_s(x, y; t+1) > \gamma_s \\ \underline{\gamma}_d, \bar{I}_s(x, y; t+1) < \gamma_s \end{cases}$$

where γ_i is the binarization parameter of spatial image. $\bar{\gamma}_i$ and $\underline{\gamma}_i$ are upper and lower increasing parameters. $\bar{\gamma}_d$ and $\underline{\gamma}_d$ are upper and lower decreasing parameters. In subsection 2.2, the detailed description for parameters will be presented. Finally, the hybrid silhouette is calculated as

$$\bar{I}_b(x, y) = \eta \bar{I}_s(x, y) + (1 - \eta) \bar{I}_t(x, y) \quad (8)$$

$$\eta = \begin{cases} \eta_R, R(x, y) > \gamma_R \\ \eta_B, R(x, y) < \gamma_R \end{cases}$$

where η_R is convex sum parameter for motion region. η_B is the convex sum parameter for background. Determination of these parameters is also discussed in section 2.2. Figure 1 shows the procedure of hybrid silhouette extraction.

2. Choices of parameters

There exist many parameters in hybrid silhouette extraction method. Unfortunately, these parameters are not easy to be determined via automatic or intelligent method. Therefore, we provide a guideline on appropriate relative choices of parameters.

The upper and lower increasing parameters have the following constraint:

$$\bar{\gamma}_i > \underline{\gamma}_i > 0. \quad (9)$$

Similarly, the upper and lower decreasing parameters have the following constraint

$$0 < \bar{\gamma}_d < \underline{\gamma}_d. \quad (10)$$

The increasing and decreasing parameters have to comply to the following constraint

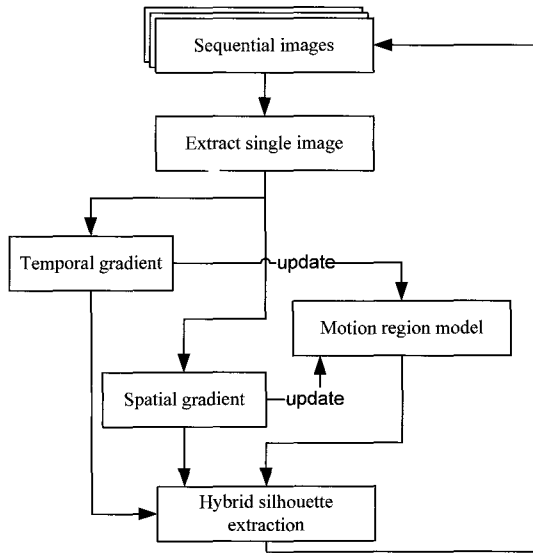


Fig. 1. Hybrid silhouette extraction process.

$$\begin{aligned} |\bar{\gamma}_i| &> \alpha |\bar{\gamma}_d| \\ |\underline{\gamma}_i| &< |\underline{\gamma}_d| \end{aligned} \quad (11)$$

where α means temporal memory length. When α is two, the motion region model have recent two past motion information. The convex sum parameter η_R and η_B have following constraint

$$\eta_R > 0.5 > \eta_B. \quad (12)$$

Large $\bar{\gamma}_i$ values lead to make spatial gradient complete insufficient silhouette whereas small $\bar{\gamma}_i$ values yield less complete silhouette. However, large $\bar{\gamma}_i$ values also yield redundant silhouette information.

III. Adaptive skeleton model

The skeleton model is one of frequently used model to represent motion. The skeleton model gives the smallest key information of human motion. Figure 2 shows the position of features in skeleton model used in this paper. The skeleton model has 15 features on each important joint. From F_1 to F_{15} mean feature points in body.

Lengths of between features are described as S_1, \dots, S_{13} .

The skeleton model is applied bases on the position of head. At first, the position and width of head are determined. Various methods to find head are proposed by many researchers. In this paper, however, we use already developed method presented in [] since finding head is not scope of research. The distance between features and positions of feature in skeleton model is normalized base on the width of face. Then we could estimate approximate position of features located in torso ($F_1, F_2, F_3, F_4, F_9, F_{10}$, and F_{11}).

The energy functions are defined for features in torso. Similar to snake algorithm, we define internal and external energy function. Then we could find good features by minimizing energy functions.

The internal energy function contains the distance and geometrical relationships between features. The internal energy function representing distance between features is defined as

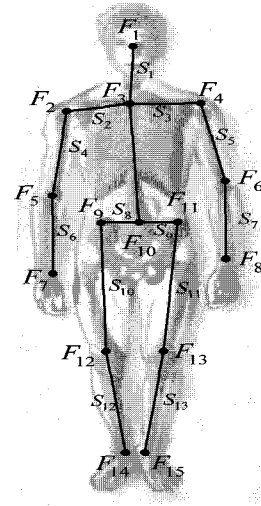


Fig. 2. Features on skeleton model.

$$E_i^{int} = |S_i - \|f_j - f_k\|| \quad (14)$$

where f_j and f_k are the estimated features corresponding to S_i . For example, the internal energy function of f_3 is defined as $E_3^{int} = |S_3 - \|f_3 - f_4\||$. The estimated feature f_i is defined as $f_i = [f_{ix}, f_{iy}]$. Some features need to defined geometrical relationship between features. For example, F_{10} and F_{11} must locate right side of F_9 . The geometrical relationship between features is defined as $g(l, f_i, f_j)$. The function g return difference results as parameter l ,

$$g(1, f_i, f_j) = \begin{cases} 1, & f_{ix} > f_{jx} \\ 0, & f_{ix} < f_{jx} \end{cases} \quad (15)$$

$$g(0, f_i, f_j) = \begin{cases} 1, & f_{iy} > f_{jy} \\ 0, & f_{iy} < f_{jy} \end{cases} \quad (16)$$

Finally, the internal energy functions for features in torso are defined as

$$\begin{aligned} E_1^{int} &= |S_1 - \|f_1 - f_3\|| + g(0, f_3, f_1) \\ &\quad + g(1, f_2, f_1) + g(1, f_1, f_4) \\ E_2^{int} &= |S_2 - \|f_2 - f_3\|| + g(1, f_2, f_3) + g(0, f_1, f_2) \\ E_3^{int} &= |S_2 - \|f_2 - f_3\|| + |S_3 - \|f_3 - f_4\|| \\ &\quad + |S_3 - \|f_3 - f_4\|| + |S_8 - \|f_3 - f_{10}\|| \\ &\quad + g(1, f_2, f_3) + g(1, f_3, f_4) + g(0, f_1, f_3) \\ E_4^{int} &= |S_3 - \|f_3 - f_4\|| + g(1, f_3, f_4) \\ &\quad + g(0, f_1, f_4) \\ E_9^{int} &= |S_9 - \|f_9 - f_{10}\|| + g(1, f_9, f_{10}) \\ E_{10}^{int} &= |S_9 - \|f_9 - f_{10}\|| + |S_{10} - \|f_{10} - f_{11}\|| \end{aligned}$$

$$\begin{aligned}
 &+ |S_8 - \|f_3 - f_{10}\| + g(1, f_9, f_{10}) \\
 &+ g(1, f_{10}, f_{11}) + g(0, f_3, f_{10}) \\
 E_{11}^{int} &= |S_{10} - \|f_{10} - f_{11}\| + g(1, f_{10}, f_{11}).
 \end{aligned}$$

The external energy function is defined by using pixel values in silhouette images. The external function of features in head is defined as

$$E_i^{ext} = - \sum_{x=f_{ix}-k}^{f_{ix}+m} \sum_{y=f_{iy}-k}^{f_{iy}+m} \bar{I}_b(x, y) \quad (17)$$

where k is distance between feature and silhouette and $2m$ is the width of searching area. Finally, the external functions for feature in torso are defined as

$$E_1^{ext} = - \sum_{x=f_{1x}}^{f_{1x}+m} \sum_{y=f_{1y}}^{f_{1y}+m} \bar{I}_b(x, y) \quad (18)$$

$$E_2^{ext} = - \sum_{x=f_{2x}-m}^{f_{2x}+m} \sum_{y=f_{2y}-k-m}^{f_{2y}-k+m} \bar{I}_b(x, y) \quad (19)$$

$$\begin{aligned}
 E_3^{ext} &= - \sum_{x=(f_{2x}+f_{3x})/2-m}^{(f_{2x}+f_{3x})/2+m} \sum_{y=f_{3y}-k-m}^{f_{3y}-k+m} \bar{I}_b(x, y) \\
 &- \sum_{x=(f_{3x}+f_{4x})/2-m}^{(f_{3x}+f_{4x})/2+m} \sum_{y=f_{3y}-k-m}^{f_{3y}-k+m} \bar{I}_b(x, y) \quad (20)
 \end{aligned}$$

$$E_4^{ext} = - \sum_{x=f_{4x}-m}^{f_{4x}+m} \sum_{y=f_{4y}-k-m}^{f_{4y}-k+m} \bar{I}_b(x, y) \quad (21)$$

$$E_9^{ext} = - \sum_{x=f_{9x}-k-m}^{f_{9x}-k+m} \sum_{y=f_{9y}-m}^{f_{9y}+m} \bar{I}_b(x, y) \quad (22)$$

$$\begin{aligned}
 E_{10}^{ext} &= - \sum_{x=f_{10x}-S_0-k-m}^{f_{10x}-S_0-k+m} \sum_{y=f_{10y}-m}^{f_{10y}+m} \bar{I}_b(x, y) \\
 &- \sum_{x=f_{10x}+S_{10}-k-m}^{f_{10x}+S_{10}-k+m} \sum_{y=f_{10y}-m}^{f_{10y}+m} \bar{I}_b(x, y) \quad (23)
 \end{aligned}$$

$$E_{11}^{ext} = - \sum_{x=f_{11x}-k-m}^{f_{11x}-k+m} \sum_{y=f_{11y}-m}^{f_{11y}+m} \bar{I}_b(x, y) \quad (24)$$

Final energy function is composed as sum of internal and external energy function with convex parameter \mathcal{K} , $E_i = \mathcal{K} \bar{E}_i^{ext} + (1 - \mathcal{K}) E_i^{int}$.

where \bar{E}_i^{ext} is the normalized external energy function calculated as

$$\bar{E}_i^{ext} = \frac{E_i^{ext}}{\sum E_k^{ext}} \quad (25)$$

The convex parameter \mathcal{K} is determined manually. When \mathcal{K} is large, the estimated features tend to move to silhouette. However, the shape of features can be deformed. When \mathcal{K} is small, the feature try to keep basic shape.

The greedy like algorithm is used to minimize energy functions. At

first, the energy function for all pixels near to feature are calculated. Then, the pixel has minimum energy function is selected as a new feature. When all pixels have same energy value, the new feature is selected as nearby pixel which is the farthest from center of body.

IV. Color based feature extraction

It is not enough to find features in body by using only silhouette information. Especially, some body part which has the great change of motion such as hand and elbow are not easy to detect by using silhouette information since silhouette is sensitive to change of motion. In this paper, we use color information as additional information to find position of hand. After finding hand, positions of elbows are calculated as geometrical relationship. Peer's color model is used to detect color in sequential images. The peer's model has wide range to detect skin color. Therefore, redundant region which has similar color to skin is also detected as skin color. To overcome this difficulty, we merge color information with temporal gradient. Let the image contains skin color region denote I_c . Then the merged skin region can be calculated as

$$\bar{I}_c = I_c \bar{I}_t.$$

To detect position of hand, we need to remove skin color region appeared in the face. The position and size of facial region can be obtained from adaptive skeleton model. After removing skin color in facial region, the positions of hand are searched by using mean shift searching algorithm. Mean shift searching algorithm tries to find minimum region contains maximum histogram values. Detailed algorithm is described in the follows.

Step 1: Histograms of x axis and y axis are obtained.

$$H_x(y) = \sum_{\forall x} \bar{I}_c(x, y)$$

$$H_y(x) = \sum_{\forall y} \bar{I}_c(x, y).$$

Step 2: The mean shift algorithm is applied to histogram $H_x(y)$ and obtains skin region in x axis. The searching space is constrained by skeleton model. By analysis skeleton model, we can estimate searching space.

Step 3: The mean shift algorithm is applied to histogram $H_y(x)$ and obtained skin region in y axis. The searching space is constrained by skeleton model and obtained skin region in x axis.

Step 4: The center of each hand is calculated from the extracted skin regions.

The positions of hands are record as feature values f_7 and f_2 . When features f_7 and f_2 are given, we can calculate feature f_5 . Let the length between estimated features f_7 and f_2 be d calculated as

$$d = \sqrt{(f_{2x} - f_{7x})^2 + (f_{2y} - f_{7y})^2} \quad (26)$$

Then position of elbow is calculated as

$$f_{5x} = \frac{S_4 f_{7x} + S_6 f_{2x}}{S_4 + S_6} - \frac{1}{\sqrt{1+a^2}} \frac{2b}{d}$$

$$f_{5y} = \frac{S_4 f_{7y} + S_6 f_{2y}}{S_4 + S_6} - \frac{a}{\sqrt{1+a^2}} \frac{2b}{d}$$

$$a = -\frac{f_{2y} - f_{7y}}{f_{2x} - f_{7x}}$$

$$b = \sqrt{(S_4 + S_7 + d)(S_7 + d)(S_4 + d)(S_4 + S_7)}$$

Right side elbow f_6 is also calculated similar to f_5 . The sequential images from ubiquitous robot system are obtained. The sequential images are captured as 320x240 resolution with 24 bit color depth. Two kind experiments are performed to check the performance of proposed method. At first, hybrid silhouette extraction method is applied to sequential image. The parameters of extraction method are chosen by following guidance presented in section 2.2.

V. Experimental results

We perform experiments for three undesired situation as described

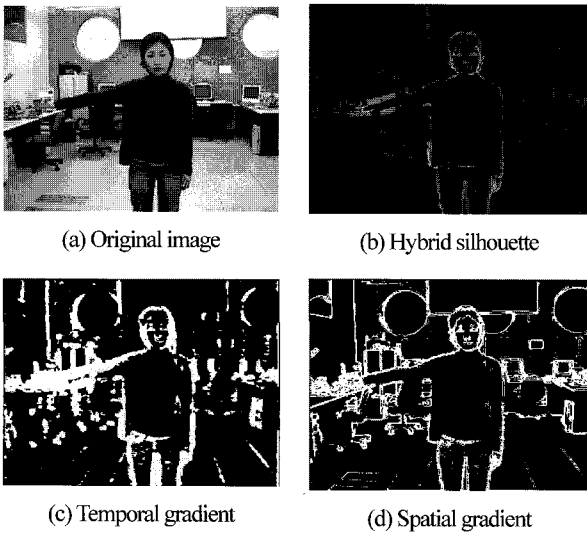


Fig. 3. Hybrid silhouette extraction: \bar{I}_t has unnecessary motion information.

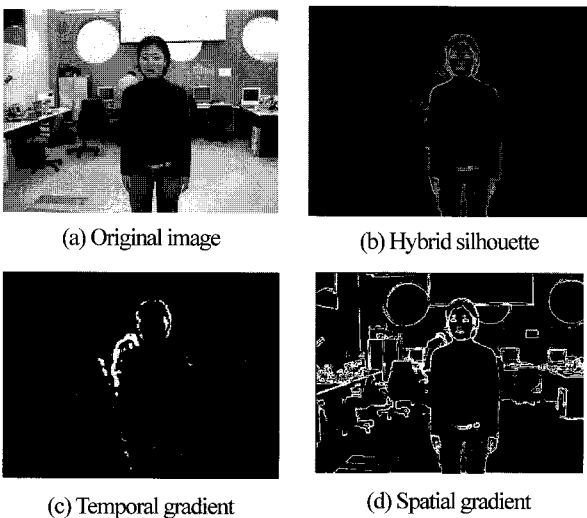


Fig. 4. Hybrid silhouette extraction: \bar{I}_t has insufficient silhouette.

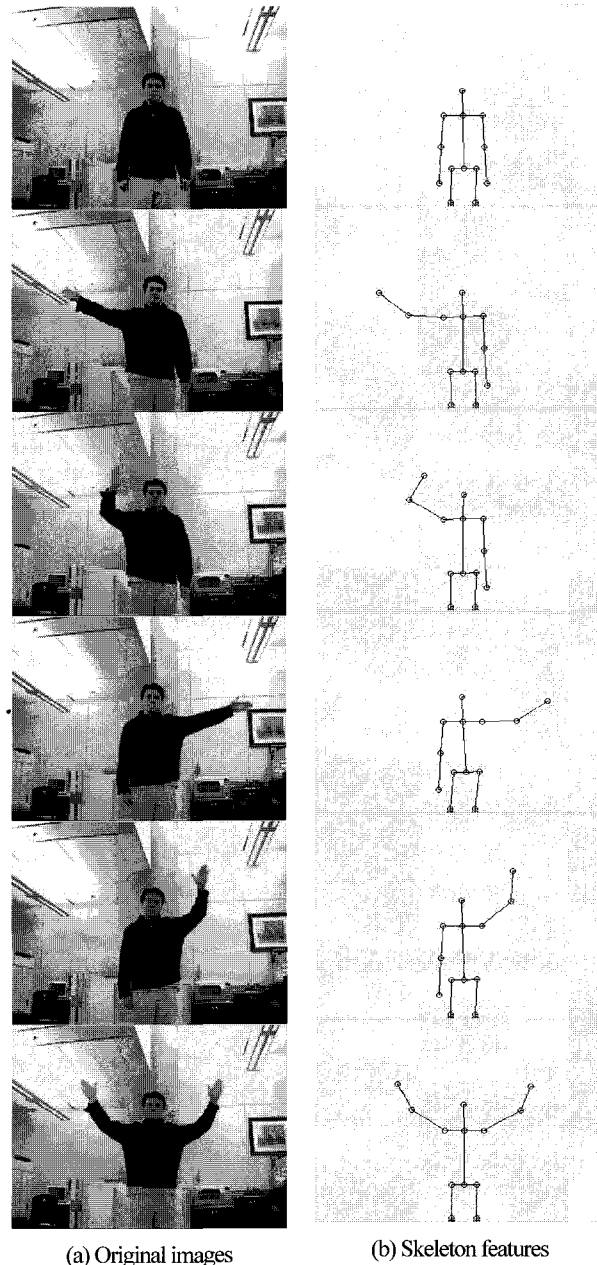


Fig. 5. Skeleton features in various motion.

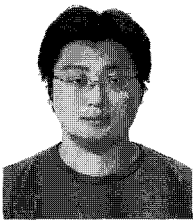
in Section 2.1. Fig. 3-4 show the results of hybrid silhouette extraction method under three undesired situation. We can check the accurate hybrid silhouette is extract in spite of unnecessary or insufficient silhouette information. Secondly, the skeleton features are extracted by using adaptive skeleton model and skin information. Fig. 5 shows the various motion and corresponding skeleton features. We could check that the skeleton features are well extracted.

VI. Conclusion

In this paper, we proposed skeleton feature extraction method by using hybrid silhouette with color information in ubiquitous robot system. The hybrid silhouette is proposed to overcome the drawback occurring in ubiquitous robot system. The adaptive skeleton model is used to find accurate features in torso. In addition, color information is used to detect the position of hand. In the experiment, we could check the superiority of the proposed method.

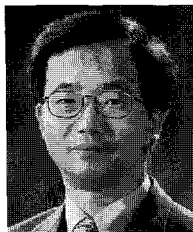
References

- [1] J. K. Aggarwal and Q. Cai, "Human motion analysis: a review," *Computer Vision and Image Understanding*, pp. 428-440, 1999
- [2] B. Fan; Z.-F. Wang, "Pose estimation of human body based on silhouette images," *International Conference on Information Acquisition Proceedings.*, pp. 296-300, June, 2004.
- [3] I. Haritaoglu, R. Cutler, D. Hawood and L. Davis, "Backpack: detection of people carrying objects using silhouettes," *Computer Vision and Image Understanding*, pp. 385-397, no. 3, 2001.
- [4] I. Haritaoglu, D. Harwood, and L. Davis, "A real time system for detection and tracking people" *Journal of Image and Vision Computing*, 1999.
- [5] I. Haritaoglu, D. Hawood and L. Davis, "Who? When? Where? What? a real time system for detecting and tracking people," *Automatic Face and Gesture Recognition*, pp. 222-227, 1998.
- [6] A. Blake, M. Isard, and D. Reynard, "Learning to track curves in motion of contours," *Artificial Intelligence*, pp. 101-133, 1995.
- [7] C. Anderson, P. Burt, and G. van der Wal, "Change detection and tracking using pyramid transformation techniques," *In Proceedings of SPIE - Ubiquitous robots and Computer Vision*, vol. 579, pp. 72-78, 1985.
- [8] J. Barron, D. Fleet, and S. Beauchemin, "Performance of optical flow techniques," *International Journal of Computer Vision*, pp. 42-77, 1994.



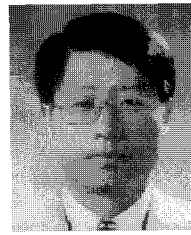
Moon Hwan Lee

He received the B.S. degrees in Electrical and Electronic Engineering from Yonsei University, Seoul, Korea, in 2004. He is currently working toward M.S. degree in Electrical and Electronic Engineering from Yonsei University. His interests include fuzzy and intelligent systems, pattern recognition, artificial intelligence, emotion recognition, and evolutionary algorithm.



Jin Bae Park

He received the B.E. degree in Electrical Engineering from Yonsei University, Seoul, Korea, and the M.S. and Ph.D. degrees in Electrical Engineering from Kansas State University, Manhattan, in 1977, 1985, and 1990, respectively. Since 1992, he has been with the Department of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea, where he is currently a Professor. His research interests include robust control and filtering, nonlinear control, mobile robot, fuzzy logic control, neural networks, genetic algorithms, and Hadamard transform spectroscopy. He served as the Director for the Transactions of the Korean Institute of Electrical Engineers (KIEE) (1998-2003) and as the Vice-President for the Institute of Control, Automation, and Systems Engineers (2004).



Young Hoon Joo

He received the B.S., M.S., and Ph.D. degrees in Electrical Engineering from Yonsei University, Seoul, Korea, in 1982, 1984, and 1995, respectively. He worked with Samsung Electronics Company, Seoul, Korea, from 1986 to 1995, as a Project Manager. He was with the University of Houston, Houston, TX, from 1998 to 1999, as a Visiting Professor in the Department of Electrical and Computer Engineering. He is currently an Associate Professor in the School of Electronic and Information Engineering, Kunsan National University, Korea. His major interest is mainly in the field of intelligent robot, intelligent control, genetic algorithms, and nonlinear systems control. He is serving as Editor-in-Chief for the Journal of Fuzzy Logic and Intelligent Systems (KFIS) (2002-2005) and Director for the Transactions of the Korean Institute of Electrical Engineers (KIEE) (2005) and for the Institute of Control, Automation and Systems Engineers (ICASE) (2005).