International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:2, No:8, 2008

Human pose estimation using Active Shape Models

Changhyuk Jang and Keechul Jung

***Abstract*—**Human pose estimation can be executed using Active Shape Models. The existing techniques for applying to human-body research using Active Shape Models, such as human detection, primarily take the form of silhouette of human body. This technique is not able to estimate accurately for human pose to concern two arms and legs, as the silhouette of human body represents the shape as out of round. To solve this problem, we applied the human body model as stick-figure, "skeleton". The skeleton model of human body can give consideration to various shapes of human pose. To obtain effective estimation result, we applied background subtraction and deformed matching algorithm of primary Active Shape Models in the fitting process. The images which were used to make the model were 600 human bodies, and the model has 17 landmark points which indicate body junction and key features of human pose. The maximum iteration for the fitting process was 30 times and the execution time was less than .03 sec..

***Keywords*—**Active Shape Models, skeleton, pose estimation.

I. INTRODUCTION

Markerless human motion analysis has been researched over the last few years. The significant research effort in this domain has been motivated that many application areas, including surveillance, Human-Computer Interaction and automatic annotation, will benefit from a robust solution. Human pose estimation is split up into model-based and model-free, depending upon whether a prior information about the object shape is employed [1]. Model-free approaches do not assume a prior human body model and this designates 'bottom-up' or 'data-driven' methods. The image data which includes a human pose is examined at a low level, looking for local structures such as edges or regions, which are assembled into groups in an attempt to identify objects of interest. Without a global model, this approach is difficult and prone to failure [2]. Model-based approaches make use of a prior model of what is expected in the image, and typically attempt to find the best match of the model to the data in a new image. Prior knowledge of the problem can be used to resolve the potential confusion caused by structural complexity, provide tolerance to noisy or missing data, and provide a means of labeling the recovered structures [1],[8].

In this paper, for concerning the merits of model-based method, we propose novel approach about the model-based pose estimation of human body using Active Shape Models that the model was made by stick-figure, "Skeleton". Model-based approaches use a human body model, which includes the kinematic structure (the skeleton), consisting of segments that are linked by joints. D.Shi et al. proposed Handwritten Chinese Radical Recognition using Active Shape Models [4]. This method considered the model as stick-figure for Chinese character which similar to body skeleton. The Active Shape Models, a representative method of model-based approaches, has been applied for many areas, including medical science and face analysis. Cootes et al. proposed Active Shape Models to capture shape variations in an iterative search procedure, capable of locating the modeled structures in noisy, cluttered images even if partially occluded [1].

The method which is proposed in this paper is divided into two processes which are the training and fitting process. The approach follows that described in Cootes et al. to which the reader is directed for details [8]. The proposed method solves the problem that the existing techniques which take the form of silhouette of human body could not estimate accurately the various human poses. We added the background subtraction for the input image that the profile normal of each landmark concerns the pixel information only for the white and black color. And the end part of the arm and leg profile normal has four directions to find the best suited position of the landmark. The maximum iteration for the matching process was 30 times and the execution time was less than .03 sec..

This paper is organized as follow. Section 2 describes the method of making the skeleton model of human body using Active Shape Models. Section 3 describes matching method of transformed primary algorithm of Active Shape Models. Section 4 describes the results of proposed method of human pose estimation. We conclude in section 5.

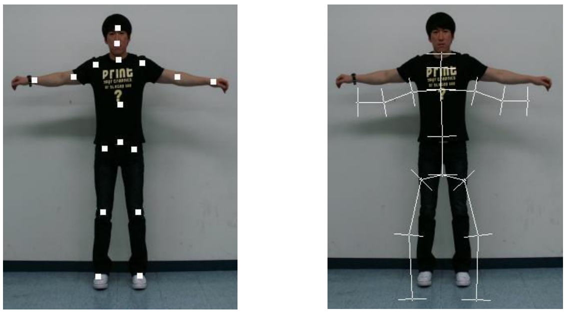
II. SKELETON MODEL

In this section we outline how our body-skeleton models are generated. Model-based approach which estimates the human body has the prior knowledge about the human body. The model which is used to the pose estimation is classed as stick-figure and volumetric ones as its application [1]. The volumetric ones which is formed body silhouette have used body tracking [3], [5]. However, this method is not suited to pose estimation purpose as the human body has two arms and legs which make manifold shape. The silhouette of human body could not recognize accurate human pose. The stick-figure one, "skeleton", estimates the human body accurately than volumetric ones [6]. The pose which shows hand up or legs

C. Jang is with the Department of Media, Graduate school of Soongsil University, Seoul, KOREA (e-mail: luee@ ssu.ac.kr).
K. Jung is with the Department of Media, Graduate school of Soongsil University, Seoul, KOREA (corresponding author to provide phone: 82-2-812-7520; fax: 82-2-822-3622; e-mail: kcjung@ ssu.ac.kr).

International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:2, No:8, 2008

apart can be recognized the use of skeleton. For this reason, we make the human body model like skeleton for the accurate pose estimation. The first step in the skeleton-based 2D matching is computing the skeleton. The term skeleton has many meanings. It generally refers to a "stick-figure" like representation of an object. For 2D objects, the skeleton is related to the medial-axis of the 2D picture [7]. The approach follows that described in Cootes et al. to which the reader is directed for details [8]. The models are generated by a model of body shape variation. The models are trained on 600 body images, each labeled with 17 landmark points representing the positions of key features. Fig.1 shows the whole landmark points and the mean shape of the model. We manually pointed each landmark to obtain the information of human body in the image. The landmark indicates the junction of the body and special position. The point of white color in the Fig. 1(a) shows the landmark points.



(a) (b)

Fig. 1 Making mean shape model using the position information of landmark point: (a) landmark point, (b) mean shape model.

The coordinates of the skeleton points for each training image are stored and analyzed statistically, to extract characteristic shape variations. As the analysis of shape variations, the alignment of the set of training shapes is executed. The alignment is achieved by scaling, rotating and translating the training shapes so that they corresponded as closely as possible with each other. The alignment is performed in a way that minimized the sum of squared distances between equivalent points on different training shapes. Fig. 2 shows the distribution of 600 images in the same frame and the aligned whole landmark points.

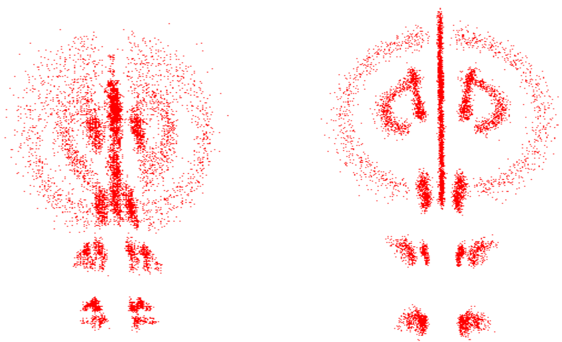


Fig. 2 600 images: landmark points and aligned form.

Each of the aligned training shapes gave rise to a vector X describing the n points. We can represent the 17 landmark points, x_i, for a single example as the *2n* element vector, where

$$\boldsymbol{x}_i = (x_{i,0}, y_{i,0}, x_{i,1}, y_{i,1}, \ldots, x_{i,n-1}, y_{i,n-1})^{\mathbf{T}},\ i = 1,\ldots,N \tag{1}$$

T is the transpose operator; and *N* is the number of training shapes. The mean shape vector $\bar{\boldsymbol{x}}$ is calculated as

$$\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i \tag{2}$$

The modes of variations, *i.e.* the ways in which the points of the shape tend to move together, can be found by applying a principal components analysis to the deviations $\Delta\boldsymbol{x}_i$ from the mean

$$\Delta\boldsymbol{x}_i = \boldsymbol{x}_i - \bar{\boldsymbol{x}}\ \ (i = 1,\ldots,N) \tag{3}$$

From these deviations, the $2n\times 2n$ covariance matrix $\boldsymbol{S}$ can be calculated as

$$\boldsymbol{S} = \frac{1}{N}\sum_{i=1}^{N} \Delta\boldsymbol{x}_i\, \Delta\boldsymbol{x}_i^{\mathbf{T}} \tag{4}$$

The modes of variation of the point of the shape can be described by the *2n* unit eigenvectors, $P_1, \ldots, P_{2n}$ and the corresponding *2n* eigenvalues , $\lambda_1, \ldots, \lambda_{2n}$, of $\boldsymbol{S}$. The k^{th} principal component corresponding to the vector $\boldsymbol{x}_i$ is defined as a weighed sum of the elements of this vector

$$\boldsymbol{b}_{i,k} = \boldsymbol{P}_k^{\mathbf{T}}(\boldsymbol{x}_i - \bar{\boldsymbol{x}}),\ \ i=1,\ldots,N,\ k=1,\ldots,2n \tag{5}$$

The principal components represent a linear independent decomposition of the variation of the training shape. The first principal component, which is associated with the largest eigenvalue λ_1, describes the largest part of the shape variation. The proportion of the total shape variance described by the k^{th} principal component is equal to the λ_k. Most of the variation can usually be represented by a small number of principal components, say *t* *(t<2n)*. The value of *t* can be chosen in such a way that the first *t* principal components explain a sufficiently large proportion (98%) of the total variance of the training shapes. The total variance λ_T is defined as

$$\lambda_T = \sum_{k=1}^{2n} \lambda_k \tag{6}$$

and the proportion of 3 principal components is obtained

The shape in the aligned training set can be approximated as a sum of the mean shape:

$$\boldsymbol{x}_i \cong \bar{\boldsymbol{x}} + \boldsymbol{P}\boldsymbol{b}_i,\ i = 1,\ldots,N \tag{7}$$

where $\boldsymbol{P} = (\boldsymbol{p}_1 \boldsymbol{p}_2 \ldots \boldsymbol{p}_t)$ is a matrix of the first *t* eigenvectors, and $\boldsymbol{b}_i = (\boldsymbol{b}_{i,1}, \boldsymbol{b}_{i,2} \ldots \boldsymbol{b}_{i,t})^{\mathbf{T}}$ is a vector of principal components, which is calculated as

$$\boldsymbol{b}_i = \mathbf{P}^{\mathbf{T}}(\boldsymbol{x}_i - \bar{\boldsymbol{x}}),\ \ i = 1,\ldots,N \tag{8}$$

Equation (8) permits the generation of a new shape example by replacing $\boldsymbol{b_i}$ with a new vector of weight values, $\boldsymbol{b_i} = (\boldsymbol{b_1}, \boldsymbol{b_2} \dots \boldsymbol{b_r})^{\mathbf{T}}$.

III. Skeleton Matching

The Active Shape Models has its own algorithm to match the shape model for input image. In the various researches, we can find the application of Active Shape Models. In the medical science, the Active Shape Models was used to segment or find special objects [9]-[11]. Also the model is applied to the tracking system in the image process [3], [5]. These research applied the primary algorithm which was originally composed of the Active Shape Models.

The shape model has profile normal for each landmark that it can find the best suited position of each landmark. The profile normal is a perpendicular line of the landmark which indicates pixel information to the boundary (Fig. 1 (b)). In the face research, the Active Shape Models can use the primary method to match the model for the input image [12], [13]. However, it is a laborious task for applying the pose estimation of human body that the method which is used to the case of face has regard for the pixel information of near landmark. The human body has various pixel information rather than the face as the human can put on the cloth as different color. Also, the profile normal has a limitation to find the best suited position of the landmark for the pose estimation as the skeleton model. In the face experiment, the model which forms an oval figure generally transforms one direction that the landmarks change the position for inside to out or reverse direction following the line of profile normal. Fig.3 shows the inaccurate results using the primary algorithm of Active Shape Models. In this figure, we applied the primary algorithm of Active Shape Models, and the end part of the arm and leg of the skeleton model was not accurately fitted the human body.

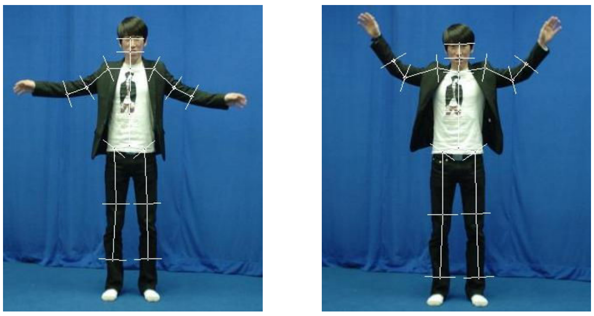


Fig. 3 Inaccurate results.

As the pose estimation, the human body has various pose (hands up and down, legs apart and stretch out, etc.). Each landmark has to be considered various directions than the case of face experiment. For this reason, we added the background subtraction for the input image that the profile normal of each landmark finds the pixel information only for the white and black color. And the end part of the arm and leg profile normal has four directions to find the best suited position of the landmark. The algorithm for fitting is:

Step 1 initialize the shape parameters, , to zero (the mean skeleton model).

Step 2 accord the center (background subtraction position and the mean skeleton model).

Step 3 generate the model instance $\boldsymbol{x_i}$ $\bar{\boldsymbol{x}} + \boldsymbol{P}$

Step 4 find each landmark, .in background subtraction region (find the pose parameters).

Step 5 update the model parameters to match to $\boldsymbol{b_i} = \mathbf{P}^{\mathbf{T}}(\boldsymbol{x_i} - \bar{\boldsymbol{x}})$

Step 6 apply constraints on .

Step 7 If not fitted, return to Step 4.

Fig. 4 shows the fitting process of the input image using skeleton model. We did experiment background subtraction manually.

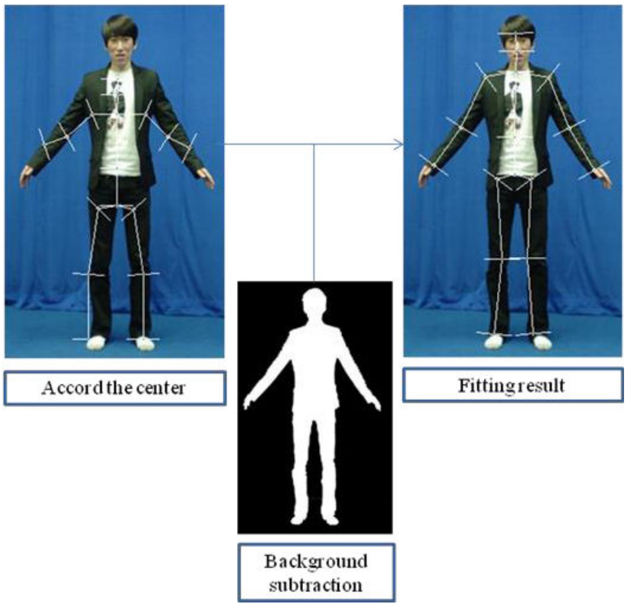


Fig. 4 Fitting process.

IV. Experimental Result

We concerned frontal image of human body and experimented this research using Pentium(R) 4 CPU 3.00GHz, 2.00GB RAM. We made the 600 images for this research and used the implementation code to make the skeleton model in the training process [14]. In the fitting process, we did manually experiment background subtraction and made the fitting algorithm deformed to obtain effective results. From the experiment, we obtained 3 principal components. Table I shows the proportion of principal axis.

TABLE I
3 PRINCIPAL COMPONENTS.

Shape mode variation		
1.	80.62%	(80.62%)
2.	12.68%	(93.3%)
3.	4.02%	(97.32%)

Fig. 5 shows the results of pose estimation using Active Shape Models. We made the skeleton model concerning the frontal pose images (e.g., "two hands up or two legs apart in the same time"). The proposed approach has been successfully tested on the frontal human poses. Also our research does not undergo influences of the illumination as the use of background subtraction not the texture or skin color of human body in the fitting process. In the fitting process, the maximum iteration was 30 times and the execution time was less than 0.03 sec.. However, we did not concern the whole human poses that stoop down, raise one hand, and raise one leg. As this reason, the proposed method could not estimate the pose which is not included in the training process. The matching results of the model can be used for further processing, such as for measurement or classification. Ongoing work focuses on the various human poses and the different view point of human pose.

V. CONCLUTION

This paper proposed novel approach about the model-based pose estimation of human body using Active Shape Models that the model was made by stick-figure, "Skeleton". Model-based approaches use a human body model, which includes the kinematic structure (the skeleton), consisting of segments that are linked by joints. We obtained the potentiality applying for the representative model-based method, "Active Shape Models", for human pose estimation. However, our research did not experiment the various human poses. Future work will explore the various view points of human pose and more effective matching algorithm of pose estimation.

REFERENCES

[1] R. Poppe, "Vision-based human motion analysis: An overview," *Computer Vision and Image Understanding*, vol. 108, 2007, pp. 4-18.

[2] T. F. Cootes, C. J. Taylor, "Statistical Models of Appearance for Computer Vision," http://personalpages.manchester.ac.uk/staff/timothy.f.cootes/Models/app_models.pdf.

[3] D. Kim, V. Maik, D. Lee, J. Shin, and J. Paik, "Active Shape model-based object tracking in panoramic Video," *Proc. ICCS, LNCS*, vol. 3994, pp. 922-929, 2006.

[4] D. Shi, S. R. Gumm, and R. I. Damper, "Handwritten Chinese Radical Recognition Using Nonlinear Active Shape Models," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE,* vol. 25, no. 2, pp. 277-280, 2003.

[5] A. M. Baumberg, and D. C. Hogg, "An Efficient Method for Contour Tracking using Active Shape Models," *Motion of Non-Rigid and Articulated Objects, Proceedings of the 1994 IEEE*, pp. 194-199.

[6] H. Sunder, D. Silver, N. Gagvani, and S. Dickinson, "Skeleton Based Shape Matching and Retrieval," *Shape Modeling International*, pp. 130-139, 2003

[7] H. Blum, "A Transformation for Extracting New Descriptors of Shape," *Computer Vision and Image Understanding: MIT press*, pp. 362-380, 1967.

[8] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active Shape Models-Their Training and Application," *Computer Vision and Image Understanding*, vol. 61, pp. 38-59, 1995.

[9] B. V. Ginneken, A. F. Frangi, J. J. Staal, B. H. Romeny, and M. A. Viergever, "Active Shape Model Segmentation With Optimal Features," *IEEE TRANSACTIONS ON MEDICAL IMAGING,* vol. 21, no. 8, pp. 924-933, 2002

[10] H. H. Thodberg, and A. Rosholm, "Application of the active shape model in a commercial medical device for bone densitometry," *Image and Vision Computing*, vol. 21, pp. 1155-1161, 2003.

[11] T. F. Cootes, A. Hill, C. T. Taylor, and J. Haslam, "The use of Active Shape Models For Locating Structures in Medical Images," *Image and Vision Computing,* vol. 12, no. 6, pp. 355-366, 1994.

[12] W. Wang, S. Shan, W. Gao, B. Cao, and B. Yin, "An Improved Active Shape Model for Face Alignment," *Proceedings of the Fourth IEEE International Conference on Multimodal Interfaces*, pp. 523-528, 2002.

[13] S. Romdhani, S. Gong, and A. Psarrou, "A Multi-View Nonlinear Active Shape Model Using Kernel PCA," *BMVC99*, pp. 483-492.

[14] M. B. Stegmann,"The AAM-API: An Open Source Active Appearance Model Implementation," *Informatics and Mathematical Modelling, Technical University of Denmark*, http://www.imm.dtu.dk/~mbs/

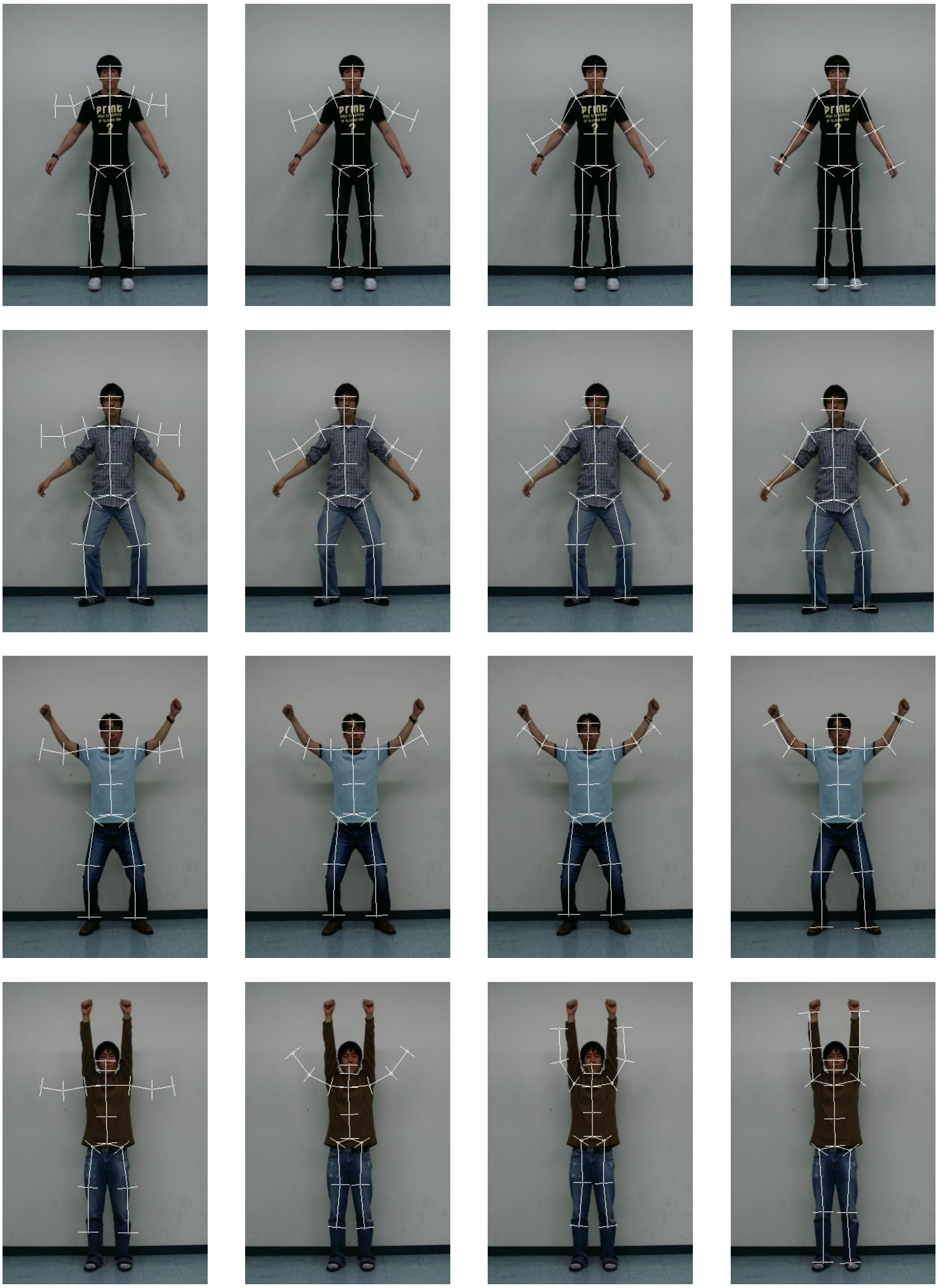


Fig. 5 Experimental results.