Upper Body Pose Recognition with Labeled Depth Body Parts via Random Forests and Support Vector Machines

Myeong-Jun Lim, Jin-Ho Cho, Hee-Sok Han, and Tae-Seong Kim

***Abstract*—Human pose recognition has become an active research topic lately in the field of human computer interface (HCI). However it presents technical challenges due to the complexity of human motion. In this paper, we propose a novel methodology for human upper body pose recognition using labeled (i.e., recognized) human body parts in depth silhouettes. Our proposed method performs human upper body parts labeling using trained random forests (RFs) and utilizes support vector machines (SVMs) to recognize various upper body poses. To train RFs, we create a database of synthetic depth silhouettes of the upper body and their corresponding upper body parts labeled maps using a commercial computer graphics package. Once the body parts get labeled with the trained RFs, a skeletal upper body model is generated from the labeled body parts. Then, SVMs are trained with a set of joint angle features to recognize seven upper body poses. The experimental results show the mean recognition rate of 97.62%. Our proposed method should be useful as a near field HCI technique to be used in applications such as smart computer interfaces.**

***Index Terms*—Upper body pose recognition, body parts labeling, random forests, support vector machines.**

I. Introduction

Human pose recognition has become an active research topic lately in the field of human computer interface (HCI). However it presents technical challenges due to the complexity of human motion.

Most previous studies on human pose recognition are based on color RGB images from which body parts are detected with respect to skin color or shape information. For instance, Lee *et al.* detected head and shoulder contours using Maximum Posteriori Probability from RGB images and estimated the pose using a body outline model [1]. Oh *et al.* proposed upper body pose estimation using a distance transform from human silhouettes in RGB images [2]. Their proposed method worked under a restricted environment with sufficient light. In general, these RGB image based methods are sensitive to light and background conditions. For improved recognition of body poses, stereo cameras have been tested. For instance, J. Mulligan estimated the upper body pose from 3D stereo images [3]. Chu *et al.* also used the disparity maps from a stereo camera and detected the head and hand using Haar features in the pre-populated space [4]. Song *et al.* also proposed a technique for upper body pose estimation in which they detected the hand using the skin color and estimated the upper body poses using depth maps from a stereo camera [5]. Cavin *et al.* segmented the upper body parts into eight regions and tracked a set of joints using likelihood based classification in Bayesian network [6].

Recently, a new type of depth camera has been introduced which utilizes an optical source and depth imaging sensor. This new camera is less sensitive to the lighting conditions. With this camera, Jain *et al.* proposed a method to estimate upper body pose via a weighted distance transformation [7]. However, their method could not overcome a merging problem because their method only used 2D information from the weighted distance transformation. Zhu *et al.* defined eight points as upper body joints, and fitted the torso and head using a likelihood function with initial poses [8]. Then, they estimated the arm pose using the connected regions with the torso. Some comments or discussion about the mentioned studies here, like "if body parts are identified in depth silhouettes, improved pose recognition could be possible."

Recently, Shotton *et al.* have introduced a real-time body parts labeling methodology using random forests (RFs) [9]. They showed the feasibility of recognizing 31 human body parts from a depth whole body silhouette. The presented methodology worked in a little far field (approximately 2~3 meters) due to the limitation of the depth camera and required a database created using optical markers and complicated motion capture setups to train RFs. No work on pose recognition was performed.

In this work, we propose a methodology of upper body pose recognition with a near field supported depth camera and propose a new way of creating a training database. First, we have created a purely synthetic training database without optic markers and motion capture settings. This database is used in training RFs to recognize upper human body parts. Then, a skeletal model is generated from the labeled body parts. Second, Support Vector Machines (SVMs) are trained to recognize seven upper body poses with joint angle features from the skeletal model. We have achieved the mean recognition of 97.62%. Our proposed method works fast and robust, and should be applicable to human computer interface in a near field.

II. Methods

To recognize upper body parts, At first, we create a database of depth silhouettes of various upper body poses and their corresponding body parts labeled maps using a computer graphics commercial package, 3Ds MAX [10] This database is used to train random forests (RFs). From the labeled body parts from the trained RFs, we generate the skeleton model for joint angle feature vectors and apply

Manuscript received October 30, 2012; revised December 13, 2012.

Myeong-Jun Lim, Jin-Ho Cho, Hee-Sok Han, and Tae-Seong Kim are with the Department of Biomedical Engineering, Kyung Hee University, Yong In, Republic of Korea (e-mail: tskim@khu.ac.kr).

DOI: 10.7763/IJIET.2013.V3.236

linear discriminant analysis (LDA) to reduce feature dimensions and separate feature vectors more clearly. Then, we recognize the upper body poses using support vector machines (SVMs). Fig. 1 shows the overall flow of our algorithms. Fig. 1 (a) shows the flow of body parts labeling via training RFs and Fig. 1 (b) shows the flow of training SVMs. Fig. 1 (c) shows the testing process using trained RFs and SVMs.

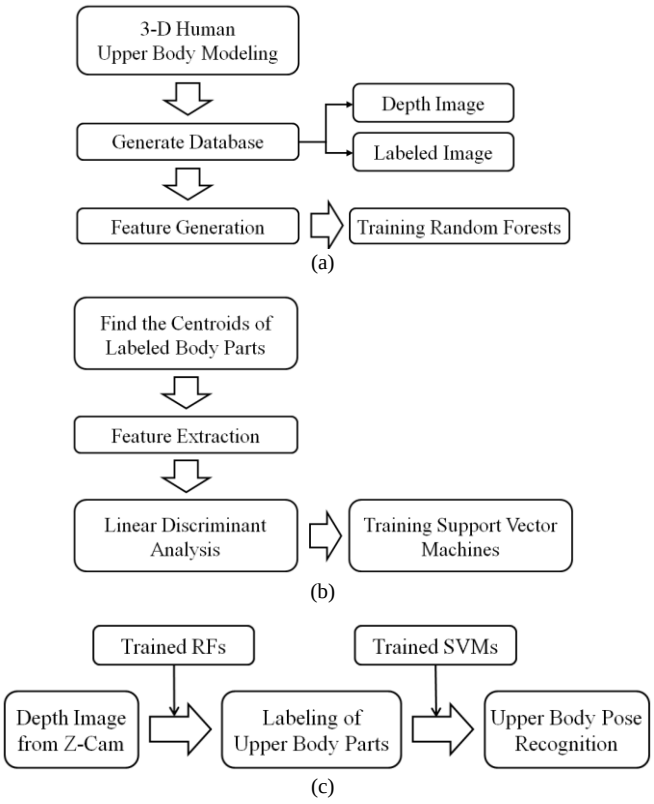

Fig. 1. Overall flow of our recognition system: (a) processes of body parts labeling via RFs, (b) processes of training SVMs for upper body pose recognition, and (c) processes of real-time upper body pose recognition.

A. Depth Imaging

In this study, we utilize a Z-cam which can capture depth images in the near field [11]. The imaging parameters were set to be an image size of 240x320, field of view of 60 degrees, and frame speed of 30fps. The distance from the camera to a subject is in a range of 0.5m~1.5m in which a subject's upper body can be captured in the field of view of the camera.

B. 3D Upper Body Modeling and Database Generation

To train RFs, we create a database (DB) using 3Ds MAX [10]. According to the human body ratio information, we create the upper body skin model which is a set of polygons, and do the same for the bone model. Our bone model consists of twelve bones (i.e., head, neck, spine, spine1, right shoulder, right upper arm, right forearm, right hand, left shoulder, left upper arm, left forearm, and left hand). Then, we modify the bone model to match the skin model in size and synchronize both the bone and skin models. To create a body parts labeled model, the upper body gets divided into twelve body parts and each body part is assigned with a different color. Our upper body model has the following parts (i.e., head, torso, right shoulder, right upper arm, right elbow, right forearm, right hand, left shoulder, left upper arm, left elbow, left forearm, and left hand).

To take pictures of the body models, we regulate the camera direction and distance to the upper body model. Finally we generate images of the depth and labeled body parts. Fig. 2 (a)-(b) show our 3-D upper body model, Fig. 2 (c) a depth image, and Fig. 2 (d) the color labeled map of the body parts corresponding to Figs. 2 (a) and (b) respectively.

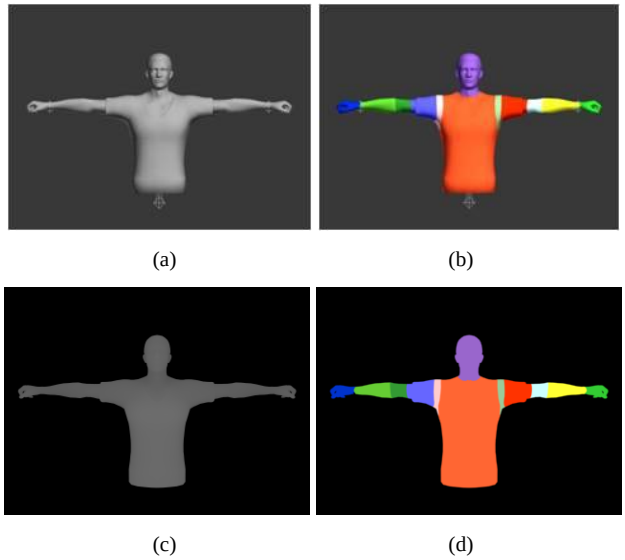

Fig. 2. Synthetic database generation: (a) front view of our 3D upper body model, (b) front view of our 3D upper body model with color labeled body parts, (c) depth image of our model, and (d) color labeled body parts map corresponding to (c)

C. Body Parts Labeling via Random Forests

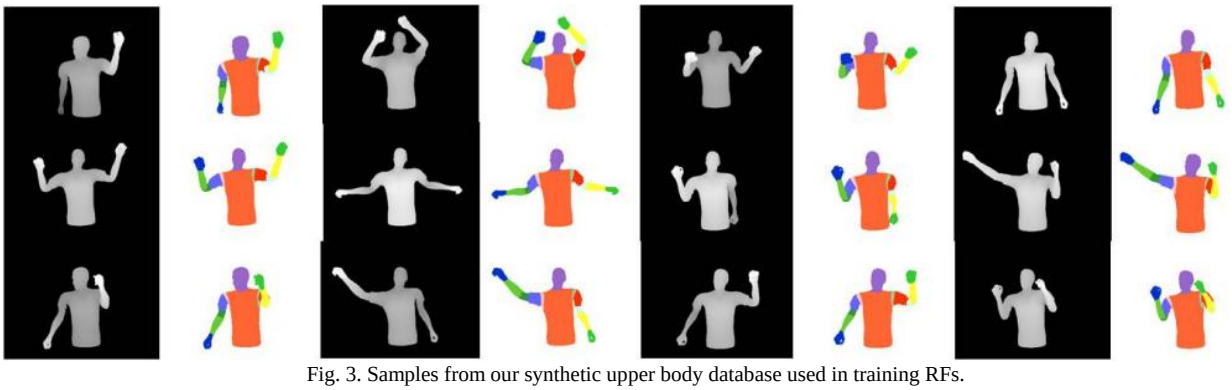

Fig. 3. Samples from our synthetic upper body database used in training RFs.

To recognize the body parts, we use RFs as a classifier [9]-[12]. The RFs consist of several number of decision trees the ensemble of trees votes to get the favorable decisions [12].

To generate features for the RFs, a set of 2,000 pixels is randomly selected from each of upper body depth silhouettes (as shown in Fig. 2 (c)) in the DB. A window with a size of 160x160 is utilized to select the 2,000 random vector pairs. The features are generated by taking differences in depth values as given below,

$$f_d(I_{pi}, n) = D(y + u_{n1}, x + u_{n2}) - D(y + v_{n1}, x + v_{n2}) \quad (1)$$

where $D(y, x)$ is depth value at the location of (y, x) and (u_{n1}, u_{n2}), (v_{n1}, v_{n2}) are the random vector pairs. At the same time, the label of selected pixel is stored as a ground truth of that feature vector.

The feature vector and ground truth are used as a training set of RFs. Random subsample are selected from the training set to train each decision tree via bootstrap sampling [12]. Then, each tree is grown to the fullest extent possible without pruning. At each internal node, the best split is determined using the Gini index among the randomized selection of features [12]. Classification is performed with the majority vote from all individually trained trees.

D. Body Parts Labeling via Random Forests

To recognize the upper body pose, we generate a body skeleton model as shown in Fig. 4 (c). The body skeleton model is created by connecting the centroids of the labeled regions [13]. From the skeleton model, the body joint (i.e., shoulder, elbow, and wrist) points are derived. Then from the body joint points, we derive directional unit vectors as features to be used in the upper body pose recognition. The equations for these features are shown below,

$$d(x, y, z) = J_a(x, y, z) - J_b(x, y, z) \quad (2)$$

$$f(i, v) = d(x, y, z) / \sqrt{x^2 + y^2 + z^2} \quad (3)$$

where d is the difference between two joint point, $f(i,v)$ is a feature vector, and J_a and J_b joint points of the upper body respectively.

Then, we apply linear discriminant analysis (LDA) to the feature vector to reduce the dimension of the feature space and make our features more compact and robust. It can be expressed as,

$$D_{opt} = \arg\max_D \frac{|D^T S_b D|}{|D^T S_w D|} = [d_1, d_2, ..., d_t]^T \quad (4)$$

where D_{opt} is the optical discrimination projection matrix, S_w and S_b are the within and between classes scatter matrices respectively [14].

To recognize upper body poses, we use SVMs. A library for SVM, LIBSVM [15]-[16] is utilized which is open and available on the web.

III. Results

In our experiments, we created pairs of 350 synthetic depth and labeled images in our DB using the upper body model. Fig. 3 shows the synthetic database (i.e., depth and labeled images) used in training RFs. Then, with the trained RFs, every incoming upper body depth silhouette gets labeled into each body parts.

Then, we performed the experiments with ten subjects in twenties. The subjects were asked to make a pose in front our system and our system recognize their poses. We divided the subjects into two groups: a group of five subjects deriving 100 images per activity to train SVMs and another group of five subjects deriving 100 images per activity to test. In the experiment, we recognized seven poses: namely stand, right hand lift (RHL), left hand lift (LHL), both hands lift (BHL), upward stretch (US), side stretch (SS), and love sign (LS). Fig. 4 (a) shows a depth silhouette of both hands lift from the depth camera and Fig. 4 (b) is an upper body parts labeled image using the trained RFs. After labeling, we found the centroids of each labeled body part. Fig. 4 (c) show the centroids of the labeled body parts and the skeleton body model superimposed on the labeled silhouettes. Fig. 4 (d) show the joint proposal.

Fig. 5 shows the features after LDA. Fig. 4 (d) and Fig. 6 show the labeling results and the skeletal model. The mean recognition rate of 97.62% was achieved. Table I shows the recognition results of the seven poses in a confusion matrix.

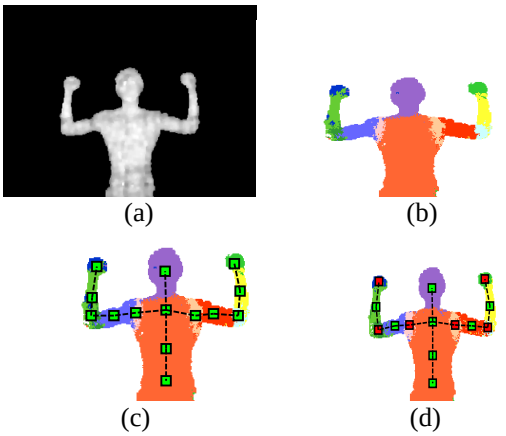

Fig. 4. Features for upper body pose recognition: (a) a depth silhouette of both hand lift, (b) labeled body parts using the trained random forest, (c) a skeletal model superimposed on (b) with the centroids, and (d) a skeletal model superimposed on (b) with the joints in red boxes.

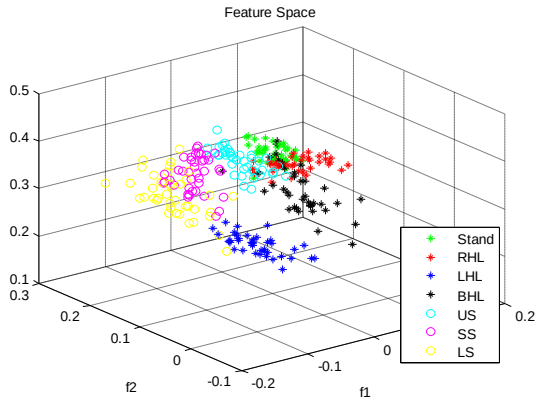

Fig. 5. A 3D feature plot from the joint points after LDA.

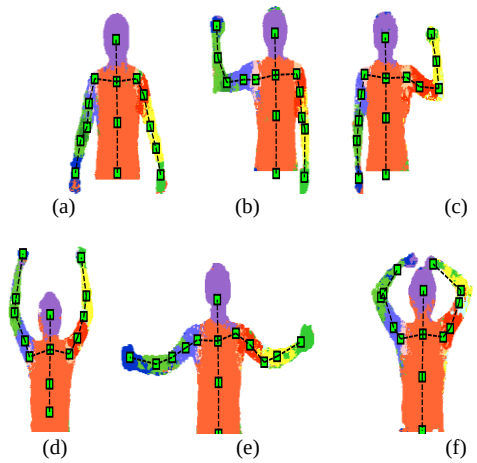

Fig. 6. Skeletal models of six different poses superimposed on their labeled body parts: (a) stand, (b) right hand lift, (c) left hand lift, (d) upward stretch, (e) side stretch, and (f) love sign.

TABLE I: A CONFUSION MATRIX OF UPPER BODY POSE RECOGNITION RESULTS

Mean [%]	Stand	RHL	LHL	BHL	US	SS	LS
Stand	**100**	0	0	0	0	0	0
RHL	0	**96.67**	0	0	0	0	3.33
LHL	0	0	**100**	0	0	0	0
BHL	0	0	0	**100**	0	0	0
US	0	0	3.33	0	**93.33**	0	3.33
SS	0	0	0	0	0	**100**	0
LS	0	3.33	0	3.33	0	0	**93.33**

IV. CONCLUSION

In this paper, we have implemented an upper body pose recognition system via labeling upper body parts of depth silhouettes using random forests and support vector machines. Our system recognizes seven different upper body poses with the mean recognition rate of 97.62%. We expect that the proposed upper body pose recognition system which works in a near view field should be useful to HCI applications for smart TV, PC and smart home applications.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST) (No. 2012-0000609). This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency) (NIPA-2012-(H0301-12-2001)).

REFERENCES

[1] M. Lee and R. Nevatia, "Body part detection for human pose estimation and tracking," *Workshop on Motion and Video Computing*, DC. USA, Feb. 2007

[2] C. M. Oh, M. Z. Islam, and C. W. Lee, "A Gesture Recognition Interface with Upper Body Model-based Pose Tracking," *International Conference on Computer Engineering and Technology*, vol. 7, pp. 531-534, 2010.

[3] J. Mulligan, "Upper body pose estimation from stereo and hand-face tracking," *Canadian Conference on Computer and Robot Vision*, British Columbia, Canada, pp. 9–11, May 2005.

[4] C. T. Chu and R. Green, "Robust Upper Body Pose Recognition in Unconstrained Environments Using Haar-Disparity," in *Proceedings of Image and Vision Computing New Zealand 2007*, Hamilton, New Zealand, pp. 97–102, Dec. 2007.

[5] Y. Song, D. Demirdjian, and R. Davis, "Tracking Body and Hands for Gesture Recognition: NATOPS Aircraft Handling Signals Database," *IEEE International Conference on Automatic Face & Gesture Recognition and Workshops*, MA, USA, pp. 500-506, March 2011

[6] R. D. Cavin, A. T. Nefian, and N. Goef, "A Bayesian Formulation for 3D Articulated Upper Body Segmentation and Tracking from Dense Disparity Maps," *International Conference on Image Processing*, Catalonia, Spain, pp. 97-100, Sep. 2003.

[7] H. Jain and A. Subramanian, "Real-time upper-body human pose estimation using a depth camera," *HP Technical Reports*, 2010.

[8] Y. Zhu, B. Dariush, and K. Fujimura, "Controlled human pose estimation from depth image streams," *CVPR Workshop on Time of Flight Computer Vision*, Anchorage, Alaska, 2008.

[9] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-Time Human Pose Recognition in Parts from Single Depth Images," *Computer Vision and Pattern Recognition*, pp. 1297-1304, June 2011.

[10] Autodesk 3ds MAX, 2012.

[11] Z-Cam. 3DV System. [Online]. Available: http://www.3dvzcam.com.

[12] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, pp. 5-32, 2001.

[13] J. A. M. Henk and Heijmans, "Connected Morphological Operators for Binary Images," *Computer Vision and Image Understanding*, vol. 73, pp. 99-120, 1999.

[14] G. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, New York: John Wiley & Sons, 1992.

[15] C. C. Chang and C. J. Lin. "LIBSVM: a library for support vector machines," *Journal of ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1-27, April 2011.

[16] E. Mayoraz and E. Alpaydin, "Support vector machines for multi-class classification," in *Proc. of International Work-Conference on Artificial Neural Network*, vol. 2, pp. 833–842, 1999,

Myeong-Jun Lim received his B.S. degree in Biomedical Engineering from Kyung Hee University, South Korea. He is currently working toward his M.S. degree in the Department of Biomedical Engineering at Kyung Hee University, Republic of Korea. His research interests include image processing, pattern recognition, artificial intelligence, and computer vision.

Jin-Ho Cho received his B.S. degree in Biomedical Engineering from Kyung Hee University, Republic of Korea. He is currently working toward his M.S. degree in the Department of Computer Engineering at Kyung Hee University, Republic of Korea. His research interests include image processing, pattern recognition, artificial intelligence, and machine learning.

Hee-Sok Han received his B.S. degree in Biomedical Engineering from Kyung Hee University, Republic of Korea. He is currently working toward his M.S. degree in the Department of Biomedical Engineering at Kyung Hee University, Republic of Korea. His research interests include image processing, pattern recognition, ultrasound signal processing.

Tae-Seong Kim received the B.S. degree in Biomedical Engineering from the University of Southern California (USC) in 1991, M.S. degrees in Biomedical and Electrical Engineering from USC in 1993 and 1998 respectively, and Ph.D. in Biomedical Engineering from USC in 1999. After his postdoctoral work in cognitive sciences at the University of California, Irvine in 2000, he joined the Alfred E. Mann Institute for Biomedical Engineering and Dept. of Biomedical Engineering at USC as a Research Scientist and Research Assistant Professor. In 2004, he moved to Kyung Hee University in South Korea where he is currently an Associate Professor in the Biomedical Engineering Department. His research interests have spanned various areas of biomedical imaging including Magnetic Resonance Imaging (MRI), functional MRI, E/MEG imaging, DT-MRI, transmission ultrasonic CT, and Magnetic Resonance Electrical Impedance Imaging. Lately he has started research work in proactive computing at the u-Lifecare Research Center where he serves as Vice Director. Dr. Kim has been developing advanced signal and image processing methods, pattern classification and machine learning methods, and novel medical imaging and rehabilitation instruments and technologies. Dr. Kim has published more than 60 peer reviewed papers and 100 proceedings, and holds 3 international patents. He is a member of IEEE, KOSOMBE, and Tau Beta Pi, and listed in Who's Who in the World ('09-'12).