International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:2, No:8, 2008

View-Point Insensitive Human Pose Recognition using Neural Network

Sanghyeok Oh, Yunli Lee, Kwangjin Hong, Kirak Kim, and Keechul Jung

***Abstract*—**This paper proposes view-point insensitive human pose recognition system using neural network. Recognition system consists of silhouette image capturing module, data driven database, and neural network. The advantages of our system are first, it is possible to capture multiple view-point silhouette images of 3D human model automatically. This automatic capture module is helpful to reduce time consuming task of database construction. Second, we develop huge feature database to offer view-point insensitivity at pose recognition. Third, we use neural network to recognize human pose from multiple-view because every pose from each model have similar feature patterns, even though each model has different appearance and view-point. To construct database, we need to create 3D human model using 3D manipulate tools. Contour shape is used to convert silhouette image to feature vector of 12 degree. This extraction task is processed semi-automatically, which benefits in that capturing images and converting to silhouette images from the real capturing environment is needless. We demonstrate the effectiveness of our approach with experiments on virtual environment.

***Keywords*—**Computer vision, neural network, pose recognition, view-point insensitive.

I. Introduction

HUMAN activity recognition has received much attention from the computer vision community since it leads to several important applications such as video surveillance for security, human-computer interaction, entertainment systems, monitoring of patients in hospitals, and elderly people in their homes.

Image-based human pose analysis has been a hot trend in the computer vision domain [1], but still remain difficult problems. First of problems, to reduce the processing time we extract features from 2D silhouette image and to find similar feature in database we make lots of feature data.

There are two approaches in making motion capture data from human action. First approach uses several sensors to sense action of human. But it is uncomfortable, because wearing sensing devices bring troubles in daily life. Second approach uses multiple cameras and captures the human action. But this approach have calibration problem such as modifying angle of multiple cameras. The main problem of these two approaches is generates three dimensional information from real environment. But in this paper, 2D information is used to recognize pose due to the input of contour shape method is a silhouette image. Although lack of information causes ambiguous problem we do not consider this with avoid using ambiguous images.

There are three approaches to recognize human pose by 2D information such as data driven approach, real-time processing, using neural network.

A. Data Driven Approach

The meaning of multiple views is there exist infinite position of camera. It is impossible to handle every position. For that reason we simplify this problem by using virtual orbit. We divide orbit into 16 view-points basis 3D human, and capture silhouette image and extract feature. And construction of capturing environment and take picture by human labor. These tasks are simple but time consuming. In this paper, capture module and feature extract module are proposed to get image and generate feature automatically. Therefore we can handle a lot of position of view easily.

B. Real-Time Processing

Real-time processing is possible through contour shape feature method because it uses two dimensional information. It does not need calibration of camera position and 3D model generation.

C. Neural Network

We use neural network for human pose recognition. Because of the feature, extracted by each pose, has patterns we use neural network to recognize many kind of features. Various features are extracted from 3D human model, and use those data to train the neural network. In the phase of test feature, which was not used at train the neural network, from virtual model is also used.

The paper is organized as follows. Section II introduces overview of our human pose recognition system. In section III, we propose the method to construct feature vector database. Using of neural network is presented in section IV. Then we show the experimental results in section V. At last, conclusion is presented in section VI.

S. Oh is with the Department of Media, Graduate School of Soongsil University, Seoul, Korea (e-mail:hyeok@ ssu.ac.kr).

Y. Lee is with the Department of Media, Graduate School of Soongsil University, Seoul, Korea (e-mail:yunli@ssu.ac.kr)

K. Hong is with the Department of Media, Graduate school of Soongsil University, Seoul, Korea (e-mail: hongmsz@ssu.ac.kr).

K. Kim is with the Department of Media, Graduate School of Soongsil University, Seoul, Korea (e-mail: raks@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).

II. Human Pose Recognition System Overview

The system architecture in proposed system consists of three parts. There are 3D human pose capture module, feature extract module and neural network. The overview of recognition system is illustrated in Fig. 1.

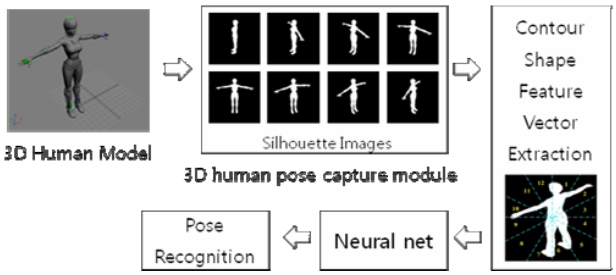

Fig. 1 Human Pose Recognition System

In 3D human pose capture module, 16 cameras are used to capture the silhouette image from 3D human model. 16 cameras are placed at around of model, difference of each camera is 11.25 degree. Silhouette images are used as input of feature extract module, output is contour shape feature vector. And 2/3 features are selected to training the neural network. After training, it is possible to test the recognition rate by using non-training data.

III. Feature Vector Database Construction

In this paper, it is proposed that human pose recognition by contour shape feature which is extracted from 2D silhouette image. To construct the feature database, we need automatic method is needed because of it takes many times and simple task is repeated. Time consuming process can be reduced by pose capture module. This module generates silhouette images from 3D human model, and feature extract module is processed to convert image to feature. After making feature database, it is possible to select set of data by simple option. In order to training neural network we choose various features which have same patterns such as same pose.

A. 2D Silhouette Image Generation

By using the 3D human model, we can generate a lot of silhouette images automatically and easily. For the automatic process we download 3D human model [5] and adjust pose what we need to use. And this model is exported to x file to use as input of pose capture module. Before use of this module number of camera should be set, then it calculates interval of each camera. To eliminate error pixel such as separate from body we use labeling. As you can see Fig. 2, 16 silhouette images are generated per pose by capture module.

Fig. 2 Silhouette Images Captured by Pose Capture Module

B. Contour Shape Feature

2D shape feature representative of multiple views are extracted from multiple views silhouette images. We use a simple approach to extract the 2D shape feature from the contour. Let, the silhouette contour of the posture as $S = \{p_1, p_2, \ldots, p_n\}$, is extracted using boundary following algorithm and store the contour points in clockwise order. The starting point is the first point $p_1 = (px_1, py_1)$. The center point $c_{x,y}$, of the posture contour is computed(see Eq. 1)

$$c_{x,y} = \left(\frac{\sum_{i=1}^{n} px_i}{n}, \frac{\sum_{i=1}^{n} py_i}{n} \right) \quad (1)$$

In equation (1), n is the total number of points extracted from the silhouette contour. Since we are using multiple view positions, each view of the same pose is varied. In addition, human shape and size are various. Therefore, the silhouettes generated from multiple cameras are various in sizes. A fixed length of contour points for each image is required for normalization. We have fixed the size of the contour points as f points.

$$\hat{p}[i] = p[i * \tfrac{n}{f}], \forall i \in [1 \ldots f] \quad (2)$$

Then, with the fixed length of silhouette contour points, the gradient between the center point $c_{x,y}$ and each contour point $\hat{p}[i] = (\hat{p}x_i, \hat{p}y_i), \forall i \in [1 \cdots f]$ is computed. The accumulation of gradient is allocated in 12 bins shape descriptor, $B = \{b_1, b_2, \ldots, b_{12}\}$ as shown in Fig. 2. We created 12 bins with 30 degree width for each bin. B is the 12 dimensions features that use to represent the 2D shape of each silhouette image.

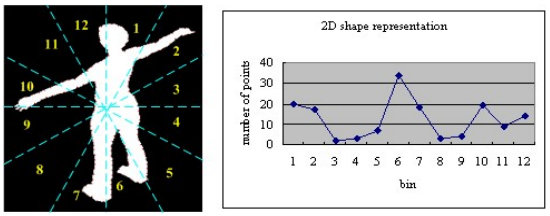

Fig. 3 Bins distribution template

IV. Neural-net Training

Neural network algorithm was introduced for recognition [3], [4]. In the case of using neural network, manipulate number of nodes of input and output layer is needed. So, 12 nodes of input layer is used because of contour shape feature module generate vector of 12 degree. We use 9 human models for that reason, and output layer has 9 nodes. One hidden layer is used and we consider various nodes in hidden layer to test recognition rate.

As you can see in Fig. 4, we make 9 models, 8 poses and 9 similar sets. Fig. 3 is a set of every pose of model No. 1. This is a set of every model and similar poses. For example, case of pose 1 and camera 3, there are 9 similar set of each pose, and 8 models. So, we have 72 similar images of pose 1 and camera 3. Illustrated in Fig. 5, feature pattern from pose 3 and nine models are almost same. But there exist ambiguous images at camera 1. Pose 1, 2 and pose 4, 5 are almost same because contour shape feature method cannot distribute these images, so we do not use images captured by left and right side cameras.

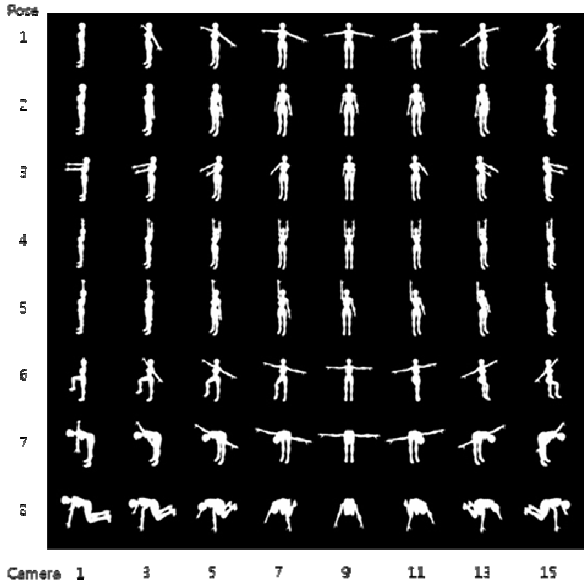

Fig. 4 Sequence of silhouette image of every pose of model No.1

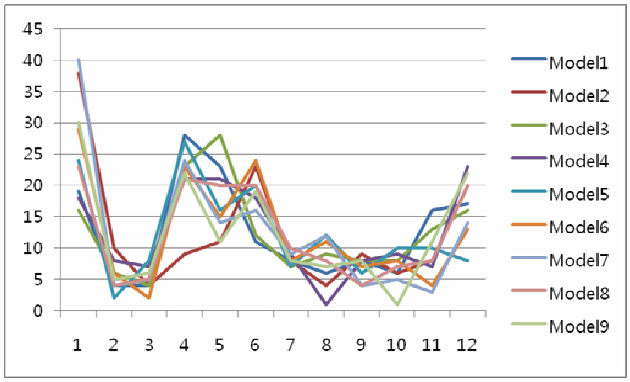

Fig. 5 Feature patterns of pose No. 1 of nine models

V. Experimental Result

We have explained our view-point insensitive pose recognition system using neural network. To construct feature database we generate 10,368 features by using 9 3D models, 8 poses, 9 similar sets and 16 images. Due to the ambiguous images, we do not use those feature to training or test data. Actually we use 12 images. Silhouette image is captured by pose capture module as shown in Fig. 2. Fig. 5 illustrate that feature of each pose has similar pattern. To training the neural network we use 2/3 features of all and 1/3 features are used as test data. We test neural network with various number of node of hidden layer. Test result is illustrated as shown in Table I and Table II.

TABLE I
Test result of neural network with non-training data

Training Data	Node	TestData	Correct	Error	Precision
5184	30	2592	1756	836	67.75
5184	70	2592	1924	668	74.23
5184	100	2592	1965	627	75.81
5184	130	2592	2037	555	78.59
5184	150	2592	2018	574	77.85
5184	170	2592	2004	588	77.31

TABLE II
Test result of neural network with training data

Training Data	Node	TestData	Correct	Error	Precision
5184	30	5184	3572	1612	68.90
5184	70	5184	4026	1158	77.66
5184	100	5184	2484	2700	47.92
5184	130	5184	4436	748	85.57
5184	150	5184	4459	725	86.01
5184	170	5184	4493	691	86.67

Table I represent the test results of neural network with non-training data. We have generated 7,776 features and 2/3(5,184) is used as training data, 1/3(2,592) is used as test data. As you can see, average precision is 75.3% at non-training data, and 81.24% at training data. We noticed that neural network with less nodes than 100 do not control the variation of feature data. Especially with 30 nodes has very poor recognition rate. Another problem that causes the poor recognition rate is due to some images are too ambiguous to recognize.

VI. Conclusion

In this paper, we have presented efficient view-point freedom human pose recognition system using neural network. We use 3D human model to capture silhouette images automatically. We use 16 cameras, placed at around of 3D human model, to generate multiple images. Then contour shape feature which has vector with 12 degree is extracted by

silhouette image. And neural network has trained by 2/3 features of all and it has been tested by non-training data. The average recognition rate is 74.5%.

For future work, we will add camera input system to test practical human pose recognition. And a lot of motion data is accessible at internet. By using these data, we can generate many poses without human labor. This will helpful of upgrade of our system.

REFERENCES

[1] Y. Sagawa, M. Shimosaka, T. Mori and T. Sato, "Fast Online Human Pose Estimation via 3D Voxel Data", Intelligent Robots and Systems, pp. 1034-1040, 2007.
[2] Catherine A., Xingtai Q., Arash M., Maurice., "A novel approach for recognition of human actions", Machine Vision and Applications, 2008, pp. 27-34, 2008.
[3] M. Voit, K. Nickel, R. Stiefelhagen, "Neural Network-Based Head Pose Estimation and Multi-view Fusion", LNCS 4122, pp. 291-298, 2007.
[4] C. Yuan, H. Niemann, "Neural networks for the recognition and pose estimation of 3D objects from a single 2D perspective view", Image and Vision Computing 19, pp. 585-592, 2001.
[5] URL downloads 3D models : http://www.turbosquid.com
[6] A. Agarwal, B. Triggs, "Human Pose from Silhouettes by Relevance Vector Regression", CVPR, vol2., pp882-888, 2004.
[7] R. Rosales, S. Sclaroff, "Learning and synthesizing human body motion and posture", IEEE, 2000.
[8] F. Lb, R. Nevatia, "Single View Human Action Recognition using Key Pose Matching and Viterbi Path Searching", CVPR, 2007.
[9] H. Yu, G. Sun, W. Song, X. Li, "Human Motion Recognition Based on Neural Network", IEEE, Vol. 2, pp. 982, 2005.
[10] M. Voit, K. Nicket, R. Stiefelhagen, "Multi-view Head Pose Estimation using Neural Networks", Computer and Robot Vision, pp. 347-352, 2005.