World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:3, No:12, 2009

View-Point Insensitive Human Pose Recognition using Neural Network and CUDA

Sanghyeok Oh, Keechul Jung

***Abstract*—**Although lots of research work has been done for human pose recognition, the view-point of cameras is still critical problem of overall recognition system. In this paper, view-point insensitive human pose recognition is proposed. The aims of the proposed system are view-point insensitivity and real-time processing. Recognition system consists of feature extraction module, neural network and real-time feed forward calculation. First, histogram-based method is used to extract feature from silhouette image and it is suitable for represent the shape of human pose. To reduce the dimension of feature vector, Principle Component Analysis(PCA) is used. Second, real-time processing is implemented by using Compute Unified Device Architecture(CUDA) and this architecture improves the speed of feed-forward calculation of neural network. We demonstrate the effectiveness of our approach with experiments on real environment.

***Keywords*—**computer vision, neural network, pose recognition, view-point insensitive, PCA, CUDA.

I. Introduction

POSTURE recognition has received a significant amount of attention given its importance for human-computer interfaces, teleconferencing, surveillance, safety control, animation, and several other applications [12].

Image-based human pose analysis has been a hot trend in computer vision research domain, but still remain difficult problems. First of problems, to reduce the processing time we use both logical and technical methods. Logically, Principle Component Analysis (PCA) is used to reduce the dimension of feature vector. And technically, Compute Unified Device Architecture (CUDA) is used to calculate feed forward process of neural network in real-time.

It is important to obtain view-independence for vision-based pose recognition in many movement-based human computer interaction (MB-HCI) applications. On one hand, view-invariant pose recognition would be straightforward when body kinematics such as joint angles can be reliably recovered from the input images [13].

This paper presents real-time pose recognition system by using histogram-based feature, neural network, PCA and CUDA.

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

A. Data Driven

For view-point insensitivity, recognition system should have many features from various view-points. Because of the system use neural network, it is needed to get enough training data. Actually there exists infinite camera position, and it is impossible to get images from every positions. This is the reason of the system use only 5 cameras. The database consists of features which are extracted from these cameras.

B. Neural Network

To have view-point insensitivity, neural network is used. Recognition at various view-points is difficult because there are lots of possibilities. Because of 2D image has only shape information, it is quiet difficult to recognize the pose at each position. By using feature database neural network can recognize human pose at various view-point.

C. Real-time Processing

Real-time processing is possible through CUDA and PCA. CUDA is GPGPU SDK from NVIDIA and this architecture is good for calculation of a lot of data at once. Because of our recognition system using 5 cameras as input, parallel processing is needed to calculate many features at once. Feed forward calculation is implemented by CUDA and this module is helpful to achieve real-time processing.

PCA is used to overcome curse of dimensionality. After extract feature vector by using histogram based method, dimension of vector is too high to use and PCA is suitable way to reduce the dimension.

The paper organized as follows. Recognition system overview is introduced in section 2. At first, feature vector extraction is provided with histogram-based feature. After this, neural network training is introduced. In the end, experimental results are presented.

II. Human Pose Recognition System Overview

The system architecture of proposed system consists of four parts which are feature database generation, histogram-based feature extraction, PCA and neural network. The overview of recognition system is illustrated in Fig. 1.

S. Oh is with the Department of Media, Graduate School of Soongsil University, Seoul, KOREA (e-mail:hyeoki@gmail.com).

K. Jung is with the Department of Media, Graduate School of Soongsil University, Seoul, KOREA (corresponding author to provide phone: 82-8-812-7520; fax: 82-2-822-3622; e-mail: jungkeechul@gmail.com).

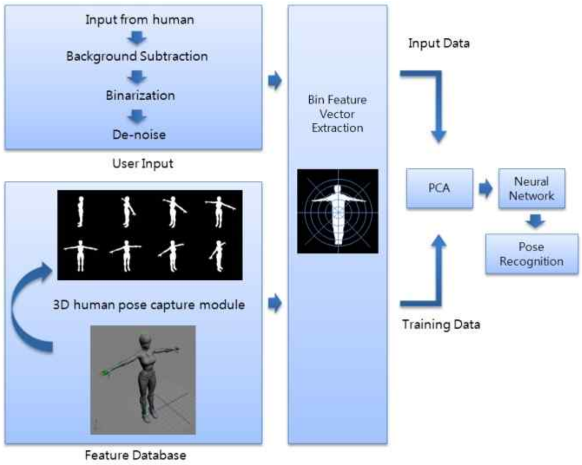

Fig. 1 Human Pose Recognition System

Fig. 2 Recognition Poses

To construct feature database, human pose images are captured by 5 cameras. These cameras are placed at around of human and used to take a photo when human takes a pose. It is converted from photo to silhouette image manually. Histogram-based feature is extracted and used as input of PCA. After reducing dimension of vector, neural network is trained. This is training task of recognition system, and input process is almost same. To use real-time camera input, system uses background subtraction, binarization and de-noising. After these processes silhouette images are used as input of pose recognition process.

III. Feature Vector Extraction

In this paper, it is proposed that human pose recognition by histogram-based feature which is extracted from 2D silhouette image.

A. 2D Silhouette Image Generation

Database is constructed by using 12 poses and 9 persons. 5 cameras are used to take a photo and total picture is 540.

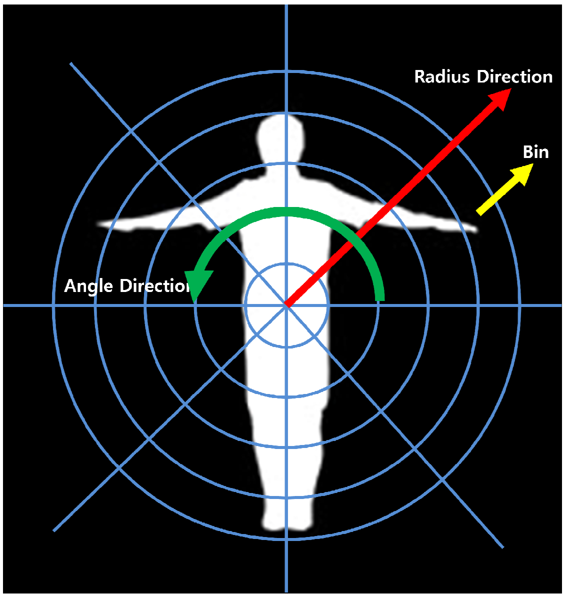

Fig. 3 Histogram-based feature from silhouette image

B. Histogram-based feature from silhouette image

In our approach, a feature with a position invariant property is desired. Position and rotation invariant feature for 3D object matching is proposed [14]. This feature is easily extracted from silhouette image. We use their approach and named the extracting histogram-based feature, Cylindrical Histogram Feature. Extracting procedures are listed below and an outline illustration is presented in Fig. 3.

1) Silhouette image is generated from 2D picture, and a central axis is set to the center of gravity.
2) Based on the central axis, 2D space is divided into r_r, radius divisions, and r_θ angle divisions. The total resolution r_q will be $r_r r_\theta$, and a corresponding 1D array of the same size is prepared. When the index values of radius and angle divisions are respectively defined as i_r, i_θ, index I of the array is calculated from (1).

$$I = i_r r_\theta + i_\theta \quad (1)$$

3) When a pixel ν is included in the region of bin b_I, the distance from the central axis d_υ is added to the corresponding array element. After this operation is applied to every pixel in the effective range, each array element is normalized by dividing it with a d_υ total. The result array is defined as a feature vector q(*t*)(see Eq. (2)). This weighting operation, has an effect to weaken the influence of pixels near the central axis, eventually increasing robustness. As a result of normalization, the final inference result will not depend on the resolution of pixel space or the subject's position. Our method can manipulate number of divisions. These properties offer high reusability of the training model.

$$q(t) = \frac{\Sigma_{\upsilon \in b_I} d_\upsilon}{\Sigma_\upsilon d_\upsilon} \quad (2)$$

IV. Neural Network 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. Because of the system using PCA to reduce dimension of feature vector, total input nodes are up to total selected eigen values. And nodes of output layer is 12, because of the feature database has 12 poses. One hidden layer is used and we consider various nodes in hidden layer to test recognition rate.

Due to the system need to do several processes, it is hard to get recognition result in real-time. Therefore processes are divided into two parts which are CPU managed and GPU manage. Image capture and feature extract module are CPU managed. Because of feed forward process needs high cost of CPU time, it can be expressed by matrix product easily and this property is suitable for GPU implementation. There are 5 cameras and system needs to process in real-time. Each camera capture 30 frames per second and total frame is 150.

A. Feed forward calculation using CUDA

Normally, matrix production needs lots of calculation and it spends lots of CPU time. This is the reason of using GPU and this architecture takes part of overall process to help CPU processing. Because of the system using one hidden layer it needs only two times of matrix product. First of all feature vector is inserted by input of process and multiplied by weight matrix between input and hidden layer. Result is calculated by multiplication of second weight matrix. Sigmoid function also implemented by CUDA.

V. Experimental Result

We performed experiments on image-based human pose recognition. Experiments were performed on 3 occasions, 1) without PCA and CUDA : high dimensional vectors are used without parallel processing 2) using PCA : use dimensional reduction algorithm. 3) using CUDA : use GPU processing. And we manipulate the total nodes of hidden layer to find the best recognition rate. The main idea of these experiments is to confirm the effectiveness of PCA and CUDA. The advantage of PCA, which is mainly speed, was expected to be confirmed between experiments 1) and 2). Advantage of CUDA was expected to be confirmed between 2) and 3). PCA is helpful to speed up of overall process and CUDA is good to implement real-time processing.

A. Recognition Pose

12 poses are used to train neural network and recognize. 5 cameras are used to capture the images with 9 persons. Total image is 540, 2/3 is used as train data of neural network. 1/3 is used as test data to get recognition rate. 5 cameras are settled on the front of human. And the angle interval of each camera is 24°.

B. Histogram-based Feature

Resolution of histogram-based feature depends on how many angle and distance divider are used. Although 4 angle divider separate image into 8 spaces, it is not suitable spaces to extract feature from human shape. Also less angle divider could not divide silhouette image effectively. Therefore 5 or more angle divider is used in our experiments.

C. PCA and CUDA

Although low number of bin has dimensional advantage, it causes lower recognition rate. On the other hand, higher number of bin cause higher recognition rate, but it carries curse of dimensionality. To achieve both higher recognition rate and dimensional advantage, PCA is used. Dimension of feature vector which is extracted by using 3 distance level and 4 angle divider is 24. Because of using PCA, it can be reduced enormously and result vector has almost same property of original vector. In proposed system, PCA is helpful to speed up the recognition process.

TABLE I
Recognition System Test without PCA

Type	***Distance Level***	***Angle Divider***	***Dimension of Input vector***	***Recognition Rate***	***Time(ms)***	
					CPU	***CUDA***
1	3	4	24	86.1%	5.1	1.3
2	3	5	30	95.0%	5.2	1.3
3	3	6	36	94.0%	5.4	1.3
4	4	4	32	87.0%	5.3	1.5
5	4	5	40	93.0%	5.7	1.5
6	4	6	48	90.0%	5.9	1.6
7	5	4	40	86.5%	5.9	1.4
8	5	5	50	94.4%	5.8	1.5

TABLE Ⅱ
Recognition System Test Using PCA

Type	***Distance Level***	***Angle Divider***	***Dimension of Input vector***	***Recognition Rate***	***Time(ms)***	
					CPU	***CUDA***
1	3	4	8	85.0%	1.8	1.3
2	3	5	8	92.2%	1.8	1.3
3	3	6	9	98.9%	1.9	1.3
4	4	4	8	85.7%	1.9	1.3
5	4	5	10	96.1%	1.9	1.5
6	4	6	10	91.1%	1.8	1.3
7	5	4	8	84.6	1.8	1.4
8	5	5	9	93.3%	1.8	1.3

CUDA, GPU SDK of NVIDIA, is powerful to process feed forward calculation. To get result, only two times of matrix product and one times of sigmoid processing are needed.

Table I shows the result of recognition system without PCA. And Table II shows the result of recognition system using PCA. The results show view-point insensitivity and possibility of real-time processing. Without PCA maximum recognition rate is 95.0% and 98.9% is result of using PCA. This result shows the system can recognize human poses without fixed view-point. Experimental type 1 and 4 shows low recognition rate because it uses only 4 angle divider. Only 4 angle divider is not effective for recognition. Other experimental type which uses 5 or more angle divider shows higher recognition rate than type 1 and 4. Comparison between results of 1) and 2) indicate the effectiveness of PCA. It upgrades not only the recognition rate but also processing time. Case of using CUDA, processing time of feed forward calculation is only 1/5 of CPU. Processing time of CPU is increasing in proportion to dimension of input vector. On the other hand, processing time of CUDA is increasing slowly because parallel processing is used.

VI. Conclusion

The main contribution of our system is view-point insensitivity and speed that ensures real-time processing. Neural network trained by features extracted from 5 cameras can recognize human poses which captured from various view point. And histogram-based feature extraction which uses circular bins shows robustness towards silhouette human poses. But dimension of result vector is too high to process. PCA, dimensional reduction algorithm, is proper way to reduce the dimension and there exist a little informational loss. Result of PCA affects the whole process to speed up. Additional to this CUDA shows possibilities of parallel processing of neural network and it completes complex calculation in real-time. Future tasks are improvements in precision, representation of silhouette human pose, and robustness towards lots of poses.

Acknowledgment

This research was supported by the National IT Industry Promotion Agency (NIPA) under the program of Software Engineering Technologies Development and Experts Education.

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 with semi-global features", 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] A. Agarwal, B. Triggs, “Human Pose from Silhouettes by Relevance Vector Regression”, CVPR, vol2., pp882-888, 2004.
[6] R. Rosales, S. Sclaroff, “Learning and synthesizing human body motion and posture”, IEEE, 2000.
[7] F. Lb, R. Nevatia, “Single View Human Action Recognition using Key Pose Matching and Viterbi Path Searching”, CVPR, 2007.
[8] H. Yu, G. Sun, W. Song, X. Li, “Human Motion Recognition Based on Neural Network”, IEEE, Vol. 2, pp. 982, 2005.
[9] M. Voit, K. Nicket, R. Stiefelhagen, “Multi-view Head Pose Estimation using Neural Networks”, Computer and Robot Vision, pp. 347-352, 2005.
[10] B. Oh, "Face recognition by using neural network classifiers based on PCA and LDA" IEEE International Conference on Systems, Man and Cybernetics, vol. 2, pp. 1699-1703, Oct. 2005.
[11] A. Park, K. Jung, "Neural Networks Implementation using CUDA and OpenMP", DICTA 2008 (Digital Image Computing: Techniques and Applications) December 1st, 2008, pp. 155-161.
[12] Van den Bergh, M. Koller-Meier, E. Van Gool, "Fast body posture estimation using volumetric features". In IEEE visual motion computing, 2008
[13] B. Peng, G. Qian, Y. Ma, “View-Invariant Pose Recognition Using Multilinear Analysis and the Universum”, ISVC 2008, Part II, LNCS 5359, pp.581-591, 2008.
[14] M. Kortgen, G.-J. Park, M. Novotni, and R. Klein, “3D Shape Matching with 3D Shape Contexts,” in Proc. WSCG, 2003