

Hand Movement Tracking and Recognizing Hand Gestures

Kwang-Chae Park¹ and Ceol-Soo Bae^{2*}

¹Electronic Engineering, Chosun University

²Information and Communication Engineering, Kwandong University

핸드 제스처를 인식하는 손동작 추적

박광채¹, 배철수^{2*}

¹조선대학교 전자공학과

²관동대학교 정보통신공학과

Abstract This paper introduces an Augmented Reality system recognizing hand gestures and shows results of the evaluation. The system's user can interact with artificial objects and manipulate their position and motions simply by his hand gestures. Hand gesture recognition is based on Histograms of Oriented Gradients (HOG). Salient features of human hand appearance are detected by HOG blocks. Blocks of different sizes are tested to define the most suitable configuration. To select the most informative blocks for classification multiclass AdaBoostSVM algorithm is applied. Evaluated recognition rate of the algorithm is 94.0%.

요 약 본 논문은 핸드 제스처에 의해 증강현실 내의 가상 객체 제어기술로, HOG기반의 핸드 제스처 인식을 제안하고 있다. 인식을 위한 특징점들은 HOG블럭들에 의하여 결정되며, 크기가 다른 여러 블럭들을 시험하여 가장 적절한 블럭구성을 결정하며, AdaBoostSVM기법을 사용하여 분류 목적에 가장 적절한 블럭들을 추출한다. 실험 결과 핸드 제스처 인식률은 94% 이었다.

Key Words : Augmented Reality, Hand Gesture Recognition, Histograms of Oriented Gradients

1. Introduction

Augmented Reality (AR) is the overlay of virtual computer graphics images on the real world, and has many potential applications in industrial and academic research. Why is Augmented Reality an interesting topic? Augmented Reality enhances a user's perception of and interaction with the real world. The virtual objects display information that the user cannot directly detect with his own senses. Thus AR enables a person to interact with the real world in ways never before possible. Some applications in such fields of AR like annotation and visualization, modeling, medicine, education etc need to be provided with a robust visual human understanding

system. It is a key for a machine to interact intelligently and effortlessly with a human-inhabited environment. One of the most natural and comfortable ways of Human-Computer interaction is using gestures. Because of many potentially important applications, gesture recognition is currently one of the most active application domains in computer vision. In this paper we propose an Augmented Reality system with a high speed hand gesture-based interface which allows user manipulate position and actions of virtual objects without any additional devices but simply by a hand gesticulation. This paper is organized as follows. Sections 2,3,4 and 5 give a description of the algorithm. In section 6 experimental results are demonstrated and in section 7 we conclude.

In this monograph is supported research funds of Chosun University in 2009

*Corresponding Author : Ceol-Soo Bae(Kwandong Univ.)

Tel: +82-11-369-0879 email: baecs@kd.ac.kr

Received July 23, 2013

Revised August 6, 2013

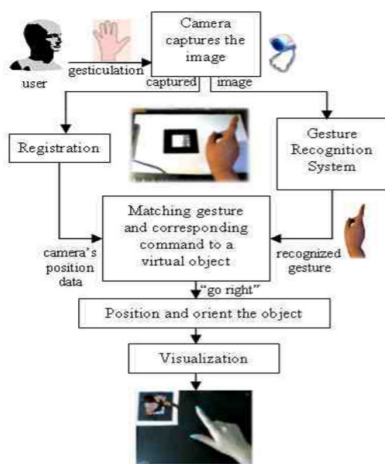
Accepted August 7, 2013

2. Work description

The idea of the algorithm is based on detecting separate parts of an object's appearance. When these parts are presented in a geometrically plausible configuration the object is detected and it is classified in one of the predefined object classes. The input image is divided into blocks which are characterized using Histograms of Oriented Gradient (HOG)[1]. For the classification purpose, we apply cascade of HOG which is based on the multiclass AdaBoost classifier[2].

3. Algorithm

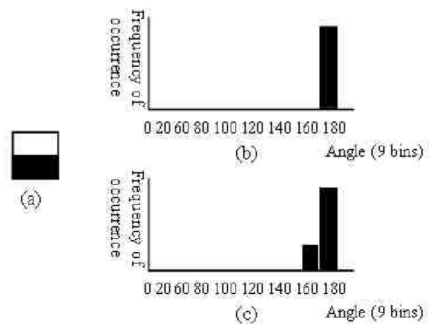
Video frames are processed in a real-time. For registration a marker card is placed in the camera's view to calculate the real camera position and orientation relative to the cards. If a human hand appears in a camera's view, system detects it, analyzes hand's motions and performs according reaction. This section describes algorithm of hand gesture recognition which is used in the proposed AR system's interface. The method is based on Dalal-Triggs method for human detection. Here we give a brief description of the algorithm. First, this approach detects separate parts of hand appearance. Then if these parts are presented in a geometrically plausible configuration hand is detected and hand's posture is classified to one of predefined gesture classes.



[Fig. 1] The proposed system's scheme

4. Histogram of oriented gradients

The HOG representation has several advantages. It captures edge or gradient structure that is very characteristic of local shape and it does so in a spatial areas. The orientation bins are evenly spaced over $0^\circ - 180^\circ$. Using bins of range $0^\circ - 360^\circ$ does not give any advantages because gradient values of range $0^\circ - 180^\circ$ and gradient values of range $180^\circ - 360^\circ$ differ only in a sign. The vote is a function of the gradient magnitude at the pixel. As it was shown in [1] using the magnitude itself gives the best results.



[Fig. 2] Simple illustration of HOG: (a) - an image of a horizontal edge, (b) - raw image (HOG has counts at only one orientation value), (c) - HOG for blurred image

5. Features capturing

The method used in the proposed work is relied on the dense set of blocks all over the entire detection window. Each detection window is divided into plenty of blocks. Blocks are set in a sliding fashion, so blocks can overlap with each other. Each block is divided into 2×2 sub-regions. To characterize each pixel, its gradient orientation was discretized (including its magnitude) into 9 histogram bins. Each sub-region thus is characterized through the 9-bin Histogram of Oriented Gradients and each block consists of a concatenated vector of all its sub-regions' HOGs. Therefore, each block is represented by a 36-dimensional feature vector. Experiments with different blocks size were done. Tested blocks have sizes from 16×16 pixels to 64×64 pixels (sub-regions 8×8 pixels to 32×32 pixels correspondently).

6. Classifier

AdaBoost is an algorithm for constructing a "strong" classifier as linear combination of "simple" "weak" classifiers $h_1(x)$:

$$f(x) = \sum_{i=1}^T \alpha_i h_i(x) \quad (\text{where } \alpha_i \in R) \quad (1)$$

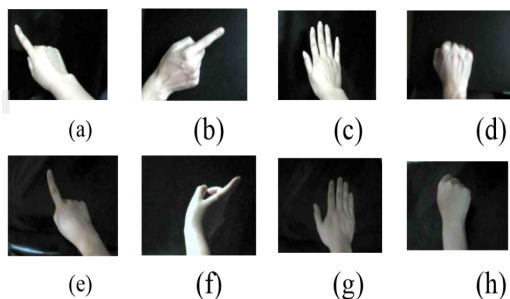
This algorithm has several advantages:

- Very simple to implement
- Does feature selection resulting in relatively simple classifier
- Fairly good generalization

In this algorithm linear SVM (Support Vector Machines)[3] is used as a "week" classifier. Since we require the system to recognize several gestures we need an algorithm of a multiclass classification. Multiclass AdaBoost is based on a singleclass AdaBoost and works in a similar way.

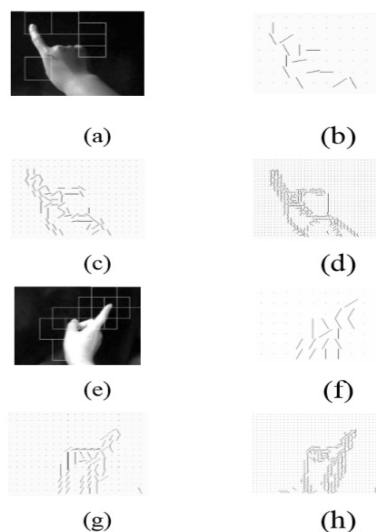
7. Experimental results

To train multiclass AdaBoostSVM classifier a training vocabulary of hand gestures was built. Figure 3 shows several samples of the training subset for four gestures. To make system robust to different lightning conditions the classifier was trained on images captured under several kinds of illumination.



[Fig. 3] Subset of gestures vocabulary. Sample of training set for gestures: (a) "left", (b) "right", (c) "stop", (d) "go"; (e),(f),(g),(h) gestures for the same commands under different lightning conditions

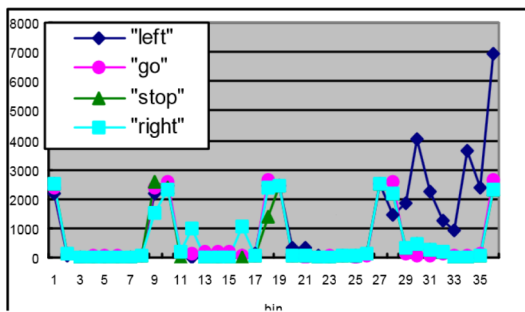
As it was shown in [4] using blocks of variable size allows capturing semantic parts in an object much more sensitively than using blocks of a fixed size. However, experiments show that using variable sizes of blocks slows down the performance while recognition rate does not significantly increase. It can be explained by the fact that for the AR system range of possible hands scales is not so wide (camera is fixed on a user's head). In practice using blocks of size 64×64 pixels (sub-regions of size 32×32 pixels) for 320×240 detection window is preferable. Totally 266 overlapping blocks are produced per frame. Figure 4 shows the most distinctive blocks selected on the example of two kinds of gestures. As it can be seen AdaBoost algorithm selects the most informative blocks which explicitly describe corresponding gesture.



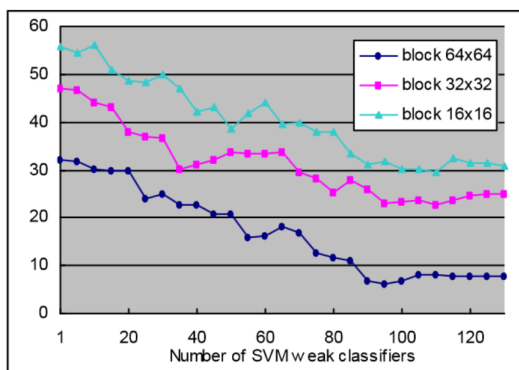
[Fig. 4] Constructed blocks for gestures: (a) "left", (b) "right". Corresponding gradients are calculated for different sizes of blocks: (b), (f) 64×64 ; (c), (g) 32×32 ; (d), (h) 16×16 .

In Fig. 5 feature vector representing one of the most distinctive blocks selected for "left" gesture recognition is shown. As it can be seen the block graph for this gesture is quite different from the graphs for the blocks of the same position for remain gestures. And that is why this block can efficiently distinguish "left" gesture from any others. In order to investigate the influence of number of "weak" classifiers in AdaBoostSVM and size of blocks on

the proposed system, we performed experiments with the number of SVM's varying from 1 to 130 and block sizes equal to 64×64, 32×32 and 16×16 pixels. Figure 6 shows the comparison results. It can be seen that the blocks' size has significant effect on the final generalization of the proposed system. The best performance show blocks with size equal to 64×64 pixels. Using these blocks recognition error 1.8 times less than that of 32×32 blocks and 2.3 times less than of 16×16 blocks. This means that large blocks are much more informative than small blocks. Moreover using large blocks decreases the block population which does contribute significantly to the performance of our system regarding to calculation time. Also as it can be seen the lowest error rate (over 6%) can be achieved when AdaBoostSVM includes 95 SVM "weak" classifiers.



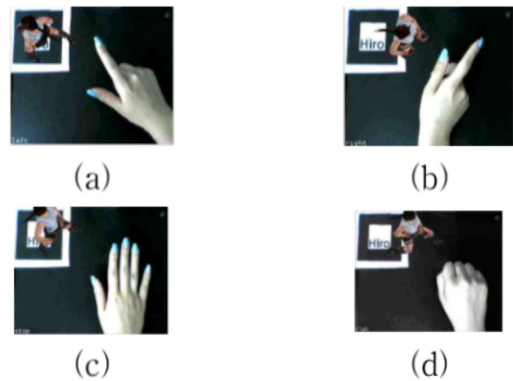
[Fig. 5] Vector corresponding to the best block for "left" gesture recognition classifier.



[Fig. 6] Comparison results of recognition error for different sizes of blocks

After gesture is recognized and registration is done system renders a virtual object using all obtained

information. A real-time video stream is augmented with an MD3 model. To demonstrate developed interface, in the example given below MD3 model's animation and position are controlled by user. For example, using gestures shown in Fig. 3(a) and 3(b) user makes the object go to the right and to the left respectively. Using gesture of the Fig. 3(c) and 3(d) user changes an animation state from idle to running.



[Fig. 7] AR system output video stream frames. MD3 model was given orders: (a) "go left", (b) "go right", (c) "stop", (d) "run"

The implemented hand gesture interface showed a high recognition rate while processing speed was 20 frames/sec which is good enough for the real-time system. In Fig. 7 several captured frames are demonstrated. The developed system shows well robustness and good controllability.

8. Conclusion

An Augmented Reality system recognizing hand gestures was developed. Results of the system evaluation were presented. Developed interface provides the AR system with the stable, fast, accurate and comfortable "human-virtual world" interaction mechanism. Experimentally it was found out that optimal size of blocks for feature vectors calculation is 64×64 pixels and optimal number of SVM "weak" classifiers included in multiclass AdaBoostSVM is 95. The implemented algorithm shows high processing speed and accuracy. An achieved gestures recognition rate is 94.0% which is quit well for a real-time system. Average processing speed is 20

frames/sec. Further algorithm can be improved by increasing number of recognizable gestures and reducing computation time.

References

- [1] Hiroyuki Arai, "Measurement of mobile antenna systems," Artech House, 2001.
- [2] Gary E. Evans, "Antenna measurement techniques," Artech House, 1990.
- [3] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1063 - 6919, 2005.
- [4] Robert E. Schapire and Yoram Singer, "Improved boosting algorithms using confidence-rated predictions," Machine Learning, vol. 37, no. 3, pp. 297 - 336, Dec 1999.
- [5] V. Vapnik, S. Golowich, and A. Smola. "Support vector method for function approximation, regression estimation, and signal processing", In M. Mozer, M. Jordan, and T. Petsche, editors, Proc. Advances in Neural Information Processing Systems 9, pp. 281 - 287, Cambridge, MA, 1997. MIT Press.
- [6] Zhu, Q.; Avidan, S.; Yeh, M-C; Cheng, K-W, "Fast Human Detection Using a Cascade of Histograms of Oriented Gradients", Proc. IEEE Computer Society Conference on Computer vision and Pattern Recognition, ISSN: 1063-6919, vol. 2, pp. 1491-1498, June 2006.
- [7] G. D. Abowd et. al., "Teaching and Learning as Multimedia Authoring: The Classroom 2000 Project," ACM Multimedia, pp. 187-198, 2000.
- [8] S. G. Deshpande & J.-N. Hwang, "A Real-time Interactive Virtual Classroom Multimedia Distance Learning System," IEEE Trans on Multimedia, vol. 3, no. 4, pp. 432-444, 2001.
- [9] D. Phung, S. Venkatesh & C. Dorai, "High Level Segmentation of Instructional Videos Based on Content Density," ACM Multimedia, 2002.
- [10] Q. Liu, Y. Rui, A. Gupta & J. J. Cadiz, "Automatic Camera Management for Lecture Room Environment," Int. Conf. on Human Fectirs in Computing Systems, 2001.
- [11] T. F. S. -Mahmood, "Indexing for topics in videos using foils," Int. Conf. CVPR, pp. 312-319, 2000.
- [12] J. Martin & J. B. Durand, "Automatic Gestures Recognition Using Hidden Markov Models," Int. Conf. Automatic Face and Gesture Recognition, 2000.
- [13] C. W. Ngo, T. C. Pong & T. S. Huang, "Detection of Slide Transition for Topic Indexing," Int. Conf. on Multimedia Expo, 2002.

Kwang-Chae Park

[Regular member]



- Feb. 1980 : Chosun Univ. Electronic Engineering, MS
- Feb. 1994 : Kwangwoon Univ. Electronics and Communications Engineering, PhD
- Nov. 2012 ~ current : Chosun Univ, Electronic information technical college, Dean
- current : Chosun Univ, Electronic Engineering, Professor

<Research Interests>

Data communication and Protocol, Signal process

Ceol-Soo Bae

[Regular member]



- Feb. 1981 : Myongji Univ. The Department of Electronic Engineering, MS
- Feb. 1988 : Myongji Univ. The Department of Electronic Engineering, PhD
- Feb. 1999 ~ Dec. 2001 : Kwandong Univ, An Engineering College, Dean
- Feb. 1981 ~ current : Kwandong Univ, Information and Communication Engineering, Professor

<Research Interests>

Image processing, Signal process