REAL-TIME HAND DETECTION BASED ON MULTI-STAGE HOG-SVM CLASSIFIER

Jiang Guo[1,a] *Jun Cheng*[12,b] *Jianxin Pang*[1,c] *Yu Guo*[1,d]

[1] Guangdong Provincial Key Laboratory of Robotics and Intelligent System,
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
[2] The Chinese University of Hong Kong
{[a]jiang.guo, [b]jun.cheng, [c]jx.pang, [d]yu.guo}@siat.ac.cn

ABSTRACT

In this paper, we propose a real-time hand detection method with multi-stage HOG-SVM classifier. Unlike traditional methods based on learning which make decomposition of feature vector or combination of different types of features or classifiers, upon the division of background into several categories, we propose a multi-stage classifier which combines several SVM classifies each of which is trained to distinguish corresponding divisions of background and target. Furthermore, in order to improve speed performance, skin color in-formation and integral histogram are also applied. Experiment results demonstrate that the proposed algorithm works well under multiple challenging backgrounds in real-time speed (16 frames per second).

Index Terms— human-computer interaction, hand detection, HOG, SVM classifier, integral image

1. INTRODUCTION

Human-Computer Interaction (HCI) has been a hot topic in recent years. Recognition of body language (such as gesture, posture and motion etc.) is an important and challenging job for HCI, among which, hand detection is a fundamental task.

The methods of object detection generally can be categorized into non-learning based methods and learning based methods. Generally speaking, for non-learning based hand detection, the first step is skin color segmentation [1, 2, 3, 4]. After skin segmentation, edge information and template are utilized in [4]; [1] detects the region of the forearm, enabling location of the hand; 1-D self-organizing mapping (SOM) combining motion information in HSV color space is used for hand region segmentation in [2]. Overall, strong assumptions are required in non-learning based methods, such as clear and complete edge are required [4], hands must be moved [2] or uniform illumination above forearm that are not occluded by sleeve [1]. Thus, non-learning based hand detection method could not handle complex background and complicated illumination. Detection methods based on machine learning are proven to be more robust against illumination and background changes than non-learning based methods. A robust, feature-separable and strong classification-capable classifier is essential to learning based method. Methods [3] based on machine learning usually extract novel Haar feature for hand detection.

Dalal and Triggs proposed Histograms of Oriented Gradient (HOG) descriptors for pedestrian detection [5], exerting a great influence. In the same year, Fatih Poriki proposed integral histogram, which greatly improves the speed of HOG feature extraction [6]. HOG is extensively discussed in related literature featuring improvements in various aspects. For example, HOG and LBP (Local Binary Patterns) are combined for pedestrian detection with partial occlusion in [7]. Linear interpolation is introduced to quantize the gradient orientation of each pixel in [8]. [9] verified that classification effect of RBF-SVM (Radial Basis Function-Support Vector Machine) is significantly better than the effect of Linear-SVM, but more time consuming. The usual way is to construct more complex cascade classifiers or make a combination of different types of classifiers. [10, 11] combine adaboost and SVM in different ways improving the algorithm speed and accuracy effectively. Combination of multiple SVM classifier are carried out in various ways in [12, 13], such as binary Tree Architecture [12] or SVM network [13].

HOG feature is utilized in many kinds of object detection, such as pedestrian detection [5, 7, 8, 9, 11], traffic sign detection [12, 13], vehicle detection [10] and hand detection [14]. Verified by above experiments, HOG feature has a strong advantage in describing the complex non-rigid object.

In this paper, a novel multi-stage classifier is proposed which applies optimized HOG features to hand detection for the first time. The remainder of paper is organized as follows. In the next section, we describe in detail the improved HOG feature. In section 3, the architecture of our proposed cascade classifier is presented. Section 4 gives comparison of experimental results and discussion. Finally, some conclusions remarks are provided in section 5.

2. HOG FEATURE EXTRACTION AND IMPROVING

Dalal and Triggs proposed HOG feature [5] which used in multiple occasions of object detection successfully. But the

The study has been financially supported by: CAS and Locality Cooperation Projects (ZNGZ-2011-012), Guangdong Innovative Research Team Program (No.201001D0104648280), Guangdong-Hong Kong Technology Cooperation Funding (2011A091200001), Guangdong-CAS Strategic Cooperation Program (2012B090400044)

speed of HOG feature extraction is not fast, and in sliding window detection, there are lots of double counting within the same area. Fatih Porikli proposed integral histogram for fast histogram extraction [6] which can also be utilized in HOG feature, namely integral HOG. As such, in this work, we will use the integral HOG for fast HOG feature extraction. Equation (1) and (2) are used for horizontal gradient $g_x(x, y)$ and vertical gradient $g_y(x, y)$ computation of each pixel $I(x, y)$ in image I. Equation (3) and (4) are used for gradient magnitude $m(x, y)$ and gradient orientation $O(x, y)$ computation:

$$g_x(x, y) = [-1, 0, 1] * I(x, y) \tag{1}$$

$$g_y(x, y) = [-1, 0, 1]^T * I(x, y) \tag{2}$$

$$m(x, y) = \sqrt{g_x(x, y)^2 + g_y(x, y)^2} \tag{3}$$

$$O(x, y) = \arctan\left(\frac{g_y(x, y)}{g_x(x, y)}\right) \tag{4}$$

In this paper, $O(x, y)$ is discretized into 6 bins (6 possible directions uniformly distributed in [0,180). In order to get fast histogram calculation in image, integral histogram technique [6] is utilized. The integral histogram can be defined in Eq. (5):

$$h_k(x, y) = \sum_{i=0, j=0 \& O(i,j) \in bin_k}^{x,y} m(i, j) \tag{5}$$

where $h_k(x, y)$ represents the value of pixel in integral histogram image H_k corresponding to k-th bin. The procedure of calculating integral HOG feature is illustrated in Fig.1. We can get any sub-histogram in an image using only three float operation in each channel of histogram H_k.

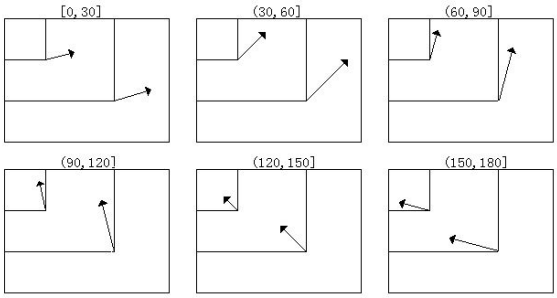


Fig. 1. gradient orientation is uniformly quantized into 6 bins , constructing integral HOG image for each bin.

In order to deal with the problem of ignorance in quantizing the gradient orientation and interpolation in spatial position, linear interpolation is utilized to quantify the gradient orientation of each pixel [8].

In this paper, samples are normalized to 32×40 pixels, each sample is divided into cells of 8×8 pixels and each group of 2×2 cells is integrated into a block. For each pixel $I(x, y)$, the gradient magnitude $m(x, y)$ and orientation $O(x, y)$ is computed in these cells. Then a local orientation histogram of gradients is formed and divides the gradient angle into 6 bins. Each block is thus represented by a 24-D feature vector normalized to an $L2$ unit length, each sample is represented by 3×4 blocks with 1/2 block overlapping, generating a 288-D feature vector as illustrated in Fig.2.

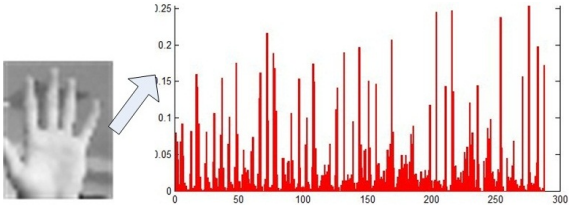


Fig. 2. For a sample of hand, extract enhanced HOG feature.

3. MULTI-STAGE SVM CLASSIFIER CONSTRUCTING

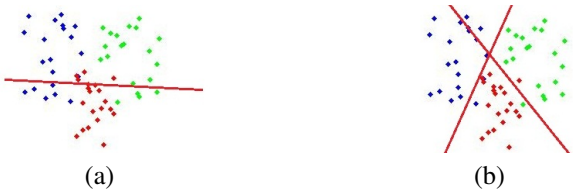
Conventional algorithms treat background simply as one indiscrete category. Actually, the classification judgment interface was directly affected by the distribution of background. Thus, Appropriate division is conducive to classification. A novel multi-stage classifier is proposed in this paper. The flow chart of our multi-stage classifier is demonstrated in Fig.4.

As illustrated in Fig.3, red points represent positive samples, blue and green points represent two category backgrounds. In Fig.3(a), the background is considered as a whole part. Therefore, only single linear category interface is formed which will cause more error. In Fig.3(b), background is divided into two parts, and two linear categories interface are formed. The classification results represent that these two category interfaces are more in line with the real data distribution.

Too simple classifier will cause bad effect on classification results because the category interface could not reflect the distribution of data. Thus, in this paper, background is divided into five categories appropriately.

The first step of our multi-stage classifier is skin segmentation. The RGB color space should be transferred to HSV color space where H subspace is utilized for segmentation. For a detection window, the number of pixels whose H value is greater than 90 is calculated. If the number is greater than half of total number of pixels in the detection window, the detection window passes this level classifier. To avoid double counting, the integral image is also utilized here. After this step, at least 50% of the detection window is filtered.

Backgrounds are divided into 5 categories including face, arm, simple background, complex background and fist. Each type of background constructs a classifier with hand respec-



(a) (b)

Fig. 3. (a) treat two classes of backgrounds as a whole part, (b) treat two classes of background respectively.

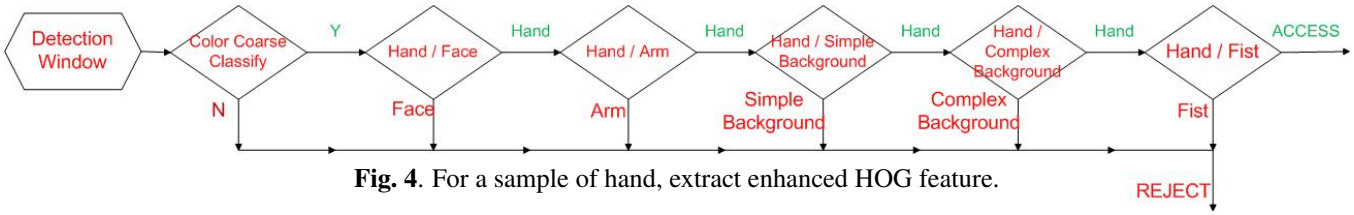


Fig. 4. For a sample of hand, extract enhanced HOG feature.

tively. The detection window is considered to be hand only if it passes all 5 SVM classifiers, as illustrated in Fig.4.

Each classifier is guaranteed a high hit rate (HR), with-out requiring very low false positive rate (FPR). From the point of view of probability and statistics, if HR and FPR for each classifier are H_i and F_i respectively, then the HR and FPR of the multi-stages classifier is:

$$HR = \prod_{i=1}^{5} H_i \quad and \quad FPR = \prod_{i=1}^{5} F_i \tag{6}$$

Quite low FPR with our multi-stage classifier can be achieved excluding the necessity of very low rate of F_i.

The speed of our multi-stage classifier S_{stage} (ms/frame) is affected by the weak classifier sort order and can be calculated by:

$$S_{stage} = \sum_{i=1}^{5} p_i s_i \tag{7}$$

$$s_1 \leqslant S_{stage} \leqslant \sum_{i=1}^{5} s_i \tag{8}$$

Where s_i represents the speed of each weak classifier in multi-stages classifier. p_i represents the possibility of through all the previous weak classifier of stage i.

The equation (7) and inequality in (8) indicate that the sooner the detection window is rejected, the faster the S_{stage} could be arrived. Because face and arm could easily pass the skin region segmentation, hand/face and hand/arm SVM classifiers take precedence in the detection flow.

The speed of SVM classifier S_{SVM} and multi-stage classifier S_{stage} is determined by:

$$S_{SVM} \propto N_{\dim} N_{SV} \tag{9}$$

$$S_{stage} \propto \sum_{i=1}^{m} N_{\dim} N_{SV_i} \tag{10}$$

Where N_{SV} represents the number of support vectors, N_{dim} represents the dimension of feature vector. The speed of multi-stage classifier is determined by the front m weak classifiers which the detection window reaches.

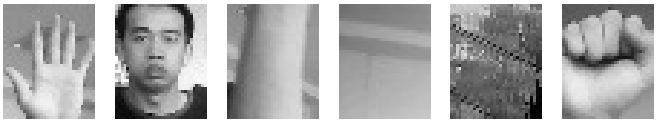
If the background is not divided, the number of support vector is large, therefore, the speed of SVM classifier is slow. If 5 types of backgrounds construct SVM classifier with hand respectively, the amount of support vectors is getting more in our multi-stage classifier, but support vectors of each weak classifier is getting much less. So with a appropriate sort order, the number of support vectors used in the stage classifier will be much less than the amount number in stage classifier. Experiments will demonstrate advantage of our multi-stage classifier in speed.

4. EXPERIMENT AND DISCUSSION

For the training of classifiers, numbers of positive and negative samples are collected. The number of each types of samples collected is shown in Table 1, and samples collected are shown in Fig. 5.

Table 1. Number of each types of samples.

Type	Hand	Face	Arm	S-BG	C-BG	Fist
Number	1967	1064	896	1232	720	1464



(a) (b) (c) (d) (e) (f)

Fig. 5. (a), (b), (c), (d), (e), (f) are hand, face, arm, simple background, complex background and fist respectively.

Samples are normalized to gray image with size of 32×40. RBF-SVM classifier supported by OpenCV lib is utilized to train and test. In order to compare experimental results, 2-class RBF-SVM, 6-class RBF-SVM and Haar-adaboost algorithm are also trained for comparison. 5 fold cross validation is utilized for experiment test. Fig.6 is the experiment ROC curve. The result is represented by false positive rate (FPR) and hit rate (HR) which are defined below:

$$\text{FPR} = \frac{number\ of\ false\ positive}{true\ positive + false\ positive}$$

$$\text{HR} = \frac{number\ of\ true\ positive}{true\ positive + false\ positive}$$

For the comparison of detection speed of each classifier, 8082 320×240 images captured with different backgrounds and illumination are tested. The magnification of sliding detection window is 1.2, 32×40 is the smallest detection window, 80×100 is the biggest detection window. The computer with 3.2 GHz CPU and 2 GB memory is utilized here. Table. 2 is the average detection speed of each classifier.

Table 2. Speed comparison of each classifier.(ms/frame)

	2-class classifier	6-class classifier	Our stage classifier	Haar-adaboost
Speed	148	194	65	58

It can be seen in Fig.6 that our multi-stage classifier outperforms the other three methods. The FPR is much less than any other methods at the same HR. The HR of multi-stages classifier is almost the same as that of the 6-class classifier. It

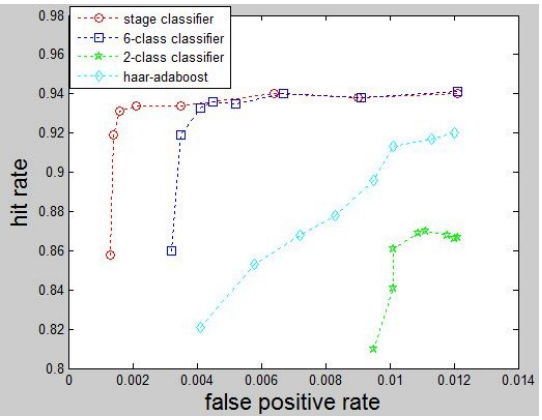


Fig. 6. The comparison of each classifier.

indicate that the most hand region will pass the all stages in our stages classifier not only several stages and most negative region will be rejected by our multi-stage classifier as analysis in section 3.

As Table 2 shows, the detection speed of our multi-stage classifier is close to Haar-adaboost, and much faster than 2-class HOG-SVM classifier and 6-class HOG-SVM classifier. Its mainly due to the well sorted order of our multi-stage classifier, most of detection window are rejected at first two stages.

The experiment results demonstrate that our multi-stage HOG-SVM classifier achieved expected performance as analysis above. Our proposed algorithm achieves high speed detection with high HR and low FPR in real time. Some experiment results under different background and illumination are shown in Fig.7.

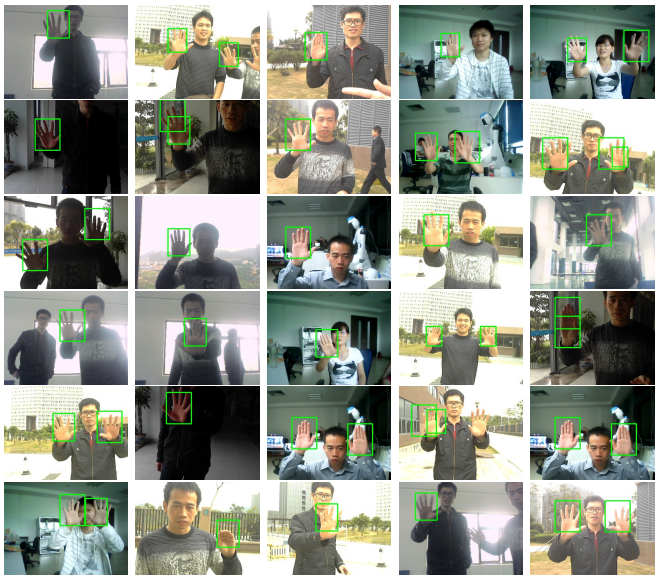


Fig. 7. The detection result of our multi-stages HOG-SVM.

5. CONCLUSION AND FUTURE WORK

In this paper, a multi-stage HOG-SVM hand detection algorithm is proposed. Through the HOG feature optimizing, multi-stage HOG-SVM classifier constructing with appropriate sort order, the hand detection algorithm works well in real time.

6. REFERENCES

[1] Dganit Maimon and Yehezkel Yeshurun, "Hand detection by direct convexity estimation," *Advanced Studies in Biometrics*, , no. 1, pp. 105–113, 2005.

[2] X Wu, L Xu, B Zhang, and Q Ge, "Hand detection based on self-organizing map and motion information," in *International Conference on Neural Networks and Signal Processing, 2003.*, 2003, pp. 253–256 Vol.1.

[3] M. Kolsch and M Turk, "Robust hand detection," in *Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004.*, 2004, pp. 614–619.

[4] Enver Sangineto and Marco Cupelli, "Real-time viewpoint-invariant hand localization with cluttered backgrounds," *Image and Vision Computing*, vol. 30, no. 1, pp. 26–37, Jan. 2012.

[5] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, vol. 1, pp. 886–893.

[6] F. Porikli, "Integral histogram: a fast way to extract histograms in Cartesian spaces," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, pp. 829–836 vol. 1.

[7] Xiaoyu Wang, Tony X. Han, and Shuicheng Yan, "An HOG-LBP human detector with partial occlusion handling," in *2009 IEEE 12th International Conference on Computer Vision*, Sept. 2009, pp. 32–39.

[8] J Shen, C Sun, W Yang, and Z Sun, "Fast human detection based on enhanced variable size HOG features," *Advances in Neural NetworksISNN 2011*, 2011.

[9] S. Paisitkriangkrai, C. Shen, and J. Zhang, "Performance evaluation of local features in human classification and detection," *IET Computer Vision*, vol. 2, no. 4, pp. 236, 2008.

[10] Xianbin Cao, Changxia Wu, Pingkun Yan, and Xuelong Li, "Linear SVM classification using boosting HOG features for vehicle detection in low-altitude airborne videos," in *2011 18th IEEE International Conference on Image Processing*, Sept. 2011, pp. 2421–2424.

[11] Qixiang Ye, Jianbin Jiao, and Baochang Zhang, "Fast pedestrian detection with multi-scale orientation features and two-stage classifiers," in *2010 IEEE International Conference on Image Processing*, Sept. 2010, pp. 881–884.

[12] Wei Liu, Jin Lv, Haihua Gao, Bobo Duan, Huai Yuan, and Hong Zhao, "An efficient real-time speed limit signs recognition based on rotation invariant feature," in *2011 IEEE Intelligent Vehicles Symposium (IV)*, June 2011, pp. 1000–1005.

[13] Fabio Boi and Lorenzo Gagliardini, "A Support Vector Machines network for traffic sign recognition," in *The 2011 International Joint Conference on Neural Networks*, July 2011, pp. 2210–2216.

[14] Liyuan Li, Qianli Xu, and Yeow Kee Tan, "Attention-based addressee selection for service and social robots to interact with multiple persons," in *Proceedings of the Workshop at SIGGRAPH Asia on - WASA '12*, New York, New York, USA, 2012, p. 131, ACM Press.