

Adaptive HOG-LBP Based Learning for Palm Tracking

Jinbing Gao

Department of Computer Science and Technology
Beihang University
Beijing, China
jarvis.gao@gmail.com

Qinghua Cao

Department of Computer Science and Technology
Beihang University
Beijing, China
caoqinghua@buaa.edu.cn

Abstract—Hand tracking is a challenging research direction in computer vision field. Although gesture tracking algorithm has been widely applied to the system of human-computer interaction, it is difficult to meet the robustness and the real-time requirements because of the hand lacks sufficiently rich texture information for discrimination. In this paper, we propose a new method named Adaptive HOG-LBP features to track the palm in the unfettered color images by fusing the HOG features and LBP features. The fused features could give more texture information of the edge features and sub-structure of the palm, and are less sensitive to light variations and background clusters. Macro structural features of the palm contain the maximum amount of information that can be used for discrimination, so we can use the fusion of HOG features and LBP features to process palm detection. The fusion features are inputted into linear SVM classifier learning. Experimental results show that in our own established palm dataset, performance has been improved significantly. Particularly worthy of note is that in our own established database, most of gesture pictures are of low quality, and the palm library contains a wealth of inter-plane and out-of-plane rotated pictures.

Keywords-palm tracking; adaptive HOG-LBP; SVM classifier

I. INTRODUCTION

As a natural and intuitive human-computer interaction approach, vision based hand recognition has been widely used in the actual life. Because the gesture itself has diversity, ambiguity, and the time and space differences, combined that the hand are complex deformable as well as visual discomfort qualitative, computer vision based gesture recognition is a challenging research topic. There are many related literatures for hand gesture recognition using different approaches. 3D model based methods [2, 5] may accurate description of the movement of the hand and its shape, but most of them are too computational expensive to be used. Shape based gesture recognition [6] could achieve better results for rigid objects, but would be not suitable for the articulated objects. Skin color modeling [8, 9] can effectively detect color information, but it's too sensitive to light and shadow to be used for reliable hand tracking. Especially when the background appears skin color, the recognition result is vulnerable to be interfered. Haar-like features [7] and boosting algorithm has recently been widely

adapted to gesture recognition, but most of its performance depends on the training process.

In this paper, we studied Histograms of Oriented Gradients [1] and Local Binary Patterns [12], and give a new fusion method named Adaptive HOG-LBP fusion for increasing the accuracy of hand detection. The fused features are putted into SVM classifier for learning. HOG features operate in the local cell unit of the image, so it can be well applied to the object detection for that it is less sensitive to the geometric deformation and the optical deformation of the object in the image. But HOG performs poorly when the background is cluttered with noisy edges. Local Binary Pattern is complementary in this aspect. LBP features have quick calculation speed, good robustness to light. But because the size of LBP window is fixed and has nothing to do with the image, LBP may have errors in the texture characteristics extraction, so it is difficult to satisfy the requirements of different roughness and scale texture. There are two traditional features fusion methods. One is co-training method [4], this method generally need to use two or more classifiers to complete the identification of the target. Although co-training can achieve better detection results, it requires complex calculations for the need to train at least two classifiers. Another method is merging characteristic matrix directly [3], this method is simple, but it can't be a good combination of multiple features characteristics. To overcome the disadvantages of the above two methods, we proposed the new adaptive HOG-LBP fusion algorithm. The algorithm fuses HOG features and LBP features, and reduces the amount of calculation in the case of guaranteed performance. Experimental results show that the method in this paper can achieve good performance in hand recognition and tracking. It also can be adapted to low quality color images and be able to achieve real-time processing performance on the PC platform with an average frame rate about 30FPS.

The rest of the paper is organized as follows: Section 2 describes some related research. Section 3 is devoted to a detailed description of the method proposed in this paper. The experimental results are given in Section 4. Finally, we conclude the full text and give future research directions.



Figure 1. Naked palm detection on PC with a webcam. The images show that our method can achieve good tracking performance when the palm moves before the webcam.

II. RELATED WORK

HOG features are first proposed by Navneet Dalal and Bill Triggs [1] to be used as the feature set of the pedestrian detection. HOG features not only have the geometric invariant characteristic, but also are able to directly adapt to the changes of light, so HOG is also widely used in object detection. In the field of hand recognition, face recognition, pedestrian detection, vehicle detection, the use of the HOG features exhibits excellent performance. HOG first calculate the value of the gradient of the image, and then do the gradient histogram statistics on different levels. So HOG is insensitive to direction, scale and light. LBP features [4] have first been found to be a powerful feature for texture classification; it has further been determined that when LBP is combined with the HOG classifier, it improves the detection performance considerably on some datasets. As HOG features [15] can give a good description of the shape and LBP can well reflect the texture information, many researchers fused them together for object detection. Wang Xiaoyu et al combined HOG and LBP to a global-part detector to detect human with partial occlusion handling [3]. Zhang et al proposed a boosted Local Structured HOG-LBP based object detector following the bottom-up paradigm after studied the description of object structure from both feature level and topology level [16]. Guo use LBP and HOG

methods to evaluate the aging effect on gender characterization with age variation [17]. Zhang et al proposed a novel feature called aHOG which combined with LBP can do better palm information mining [11]. These studies about the HOG and LBP prove that the fusion of the two class features can achieve better performance in object recognition, but most of the existing fusion methods just directly combine the two features together and do not fully use the two features. So a simple and effective fusion method is needed.

III. ADAPTIVE HOG-LBP FUSION FOR HAND DETECTION

Inspired by the work of Wang Xiaoyu et al in Human Detection [3] and Zhang's work in Hand Posture Recognition [11], we proposed the novel feature called Adaptive HOG-LBP feature for palm tracking in this paper. Two different types of features, HOG and LBP are fused to train a classifier for hand posture recognition. We believe that the description ability of LBP and HOG to the object is different. So, we need to determine the strength of HOG and LBP description ability on the objects or the parts of the objects. LBP and HOG are both based on the histogram statistics, and HOG features vote in cell structure and merge to block structure for normalization. So we can do LBP features statistics in the cell structure at the same time. And for the same block structure, we use SVM to learn these two features to compute confidence of HOG and LBP. After that, we do the weighted fusion based on the confidence of these two features in the block structure and normalize the fused features. Finally, we calculate the Adaptive HOG-LBP features which can be well applicable to the case of the sub-structure of the object are not similar. In the palm, because of the different sub-structure of the palm portion and finger part, the description ability of HOG and LBP are not same to partial palm. So our method can be used in the detection of the palm. The feature fusion can make full use the advantages of the two features.

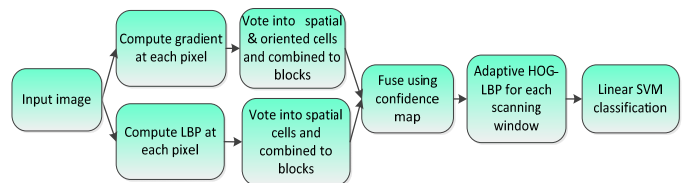


Figure 2. The framework of Adaptive HOG-LBP detector.

A. Confidence Map Computing

In order to ensure that the fused features are adaptive and the features fusion can make full use of the advantages of the two features, we need to determine the fusion proportion according to description ability of the features of objects. At this point, we need to train two classifications. In order to better reflect the local features of gestures, we divide the hand gesture into smaller modules, and then calculate the confidence map of HOG and LBP by using labeled positive samples and negative samples. This allows the features fusion to better reflect the edge features of the palm. The

following detailed describe the calculation process of the degree of confidence of these two characteristics.

a) Compute gradients at each pixel of the input image and vote them into spatial and oriented cells. We choose 1-D centred $[-1,0,1]$ mask to calculate separate gradients for each colour channel, and take the one with the largest norm as the pixel's gradient vector. And then we vote the gradient magnitude to the rectangular cells with orientation bins spaced over $0^\circ \sim 180^\circ$.

b) Compute LBP at each pixel and vote them into spatial cells. The cells of LBP are the same local spatial regions as the cells of HOG. We compute LBP over all three colour channels of RGB color space independently, and then concatenating the results together[13]. The principle of standard LBP calculation is showed in Figure 3 In each colour channel, we use $LBP_{(8,1)}^{U_2}$ norm to calculate the features.

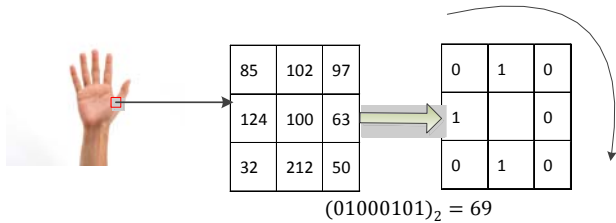


Figure 3. The standard LBP calculation principle

c) Compute the confidence of HOG and LBP in each block. We groupe cells into blocks to calculate the histogram of the two features independently. And then we use linear SVM classifier to calculate the confidence of HOG and LBP,

$$C = \frac{1}{1 + \exp(-d)}, d = W\Phi(x) + b \quad (1)$$

In expression (1), d denotes the distance between the feature vector of the candidate window and the separating hyperplane, we use sigmoid function to transform d to the confidence, W and b are the parameter of SVM classifier.

We use C_{HOG} to denote the confidence of HOG and use C_{LBP} to denote the confidence of LBP.

d) Build the confidence map for each detection window. Each detection window includes a plurality of blocks, so we can build the confidence map by merging the confidence of each block.

B. Fusion based on Confidence Map

After the calculation of the confidence map for each detection window, we then fuse HOG and LBP to a novel

feature called Adaptive HOG-LBP features. The fused features,

$$Feature(i) = C_{HOGi} \cdot V_{HOGi} + C_{LBPi} \cdot V_{LBPi}, \quad (2)$$

are calculated in each block. C_{HOGi} is the HOG confidence of block i , V_{HOGi} is the HOG feature vector of block i , C_{LBPi} is the LBP confidence of block i , V_{LBPi} is the LBP feature vector of block i . And then we use $L1$ -norm to normalize all the cells in one block. We group the normalized blocks into a detect window and merge the block vectors into one vector. Finally, we obtain the Adaptive HOG-LBP features vector in each detect window.

C. Adaptive HOG-LBP Detector for Palm Tracking

We use a sliding window method to detect the palm, and use a linear SVM classifier to classify. First of all, we take samples of the palm to establish a gesture dataset. Secondly, the confidence of HOG and LBP are calculated by using SVM and the features are fused based on confidence. Thirdly, we train a linear SVM classifier based on the fused features. And then we could use the trained classifier to detect hand gesture. The framework of the whole detector is showed in Figure 2. For each input image of 320×480 pixels, we use sliding window to extract the Adaptive HOG-LBP feature and put it into the SVM classifier. In order to ensure the accuracy of detection, feature extraction and classification go on in multi-scale [10]. In addition, we use the candidate window to accelerate the speed of detection, in order to achieve real-time tracking requirements.

IV. EXPERIMENT

In order to conduct the experiment, we built a palm dataset by ourselves. We capture 5378 palm images in 12 different occasions from 28 different persons as positive samples and randomly cut 6946 images without hand palm as negative samples. Besides, the samples are color pictures collected from motion hand gesture and always show with motion blur. Moreover in our work we do the classification without any pre-process such as noise filtering, brightness balance or sharpening. All these points show the challenging of our samples.

In the experiment, we extract traditional HOG-LBP and Adaptive HOG-LBP features separately in the training sets and test sets. As can be seen from the Figure 4 of recall-precision curves, our approach obtains the detection efficiency of 95.2% precision rate with the recall of 90%. Relative to the traditional HOG-LBP features, our method has significantly improved the performance. The average detection time is about 90ms when the palm appears in the detection window. In the worst case, the detector needs to detect the whole image with the time of about 120ms. So, our hand tracking system could satisfy the real-time requirements. In our work, in order to ensure the wide usability of the hand detection, we design a hand tracking

system consisting of an ordinary home web camera and a PC. The camera model is Logitech C270, supports 640×480 pixels resolution and 30FPS sampling rate. The PC has Dual-core CPU at 2.50 GHz and 2.00GB DDR2 RAM. The hand detector is showed in Figure 1.

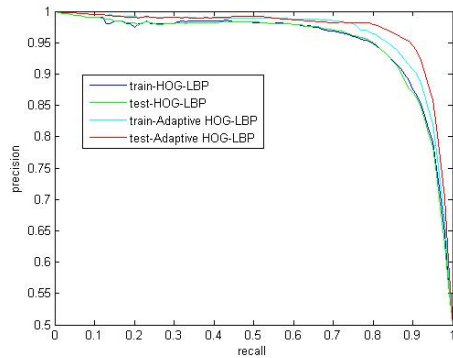


Figure 4. The recall-precision curve of the purposed method.

V. CONCLUSION

In this paper, we propose Adaptive HOG-LBP features by fusing HOG and LBP for palm tracking. Compared to the traditional HOG-LBP features, our method could make full use the advantages of these two features and obtain better performance. In addition, we use this new feature to train a linear SVM classifier for real-time palm tracking. The experiment results show the feasibility of the proposed method.

However, our hand detector is not universal for characteristic human motion detection, which is the next problem to be tackled in our future work.

REFERENCES

- [1] N. Dalal and B. Triggs: "Histograms of oriented gradients for human detection". In CVPR 2005, volume 1, pages 886–893, 2005.
- [2] R. Rosales, V. Athitsos, L. Sigal, and S. Sclaroff. "3D Hand Pose Reconstruction using Specialized Mappings". In Proc. Intl. Conf. Computer Vision, 2001.
- [3] Xiaoyu Wang, Tony X. Han and Shuicheng Yan. "An HOG-LBP human detector with partial occlusion handling". In CVPR 2009, pages 32-39.
- [4] Ojala, T., Pietikainen, M. and Maenpaa, T.. "Multiresolution gray-scale and rotation invariant texture classification with Local Binary Patterns". In PAMI 2002 Volume 24, pages 971-987.
- [5] H. Fillbrandt, S. Akyol, K. F. Kraiss, "Extraction of 3D Hand Shape and Posture from Image Sequences for Sign Language Recognition". In IEEE International Workshop on Analysis and Modeling of Faces and Gestures, vol. 17, October 2003, pp. 181-186.
- [6] Y. Hamada, N. Shimada and Y. Shirai, "Hand Shape Estimation under Complex Backgrounds for Sign Language Recognition", in Proc. of 6th Int. Conf. on Automatic Face and Gesture Recognition, 2004.
- [7] Q. Chen, N. Georganas, and E. Petriu. "Real-time vision based hand gesture recognition using haar-like features". Instrumentation and Measurement Technology Conference Proceedings, 2007, pages 1–6.
- [8] Michal Kawulok. "Dynamic Skin Detection in Color Images for Sign Language recognition", In Image and Signal Processing - 3rd International Conferences, 2008.
- [9] Zhigeng Pan, Yang Li, Mingmin Zhang, Chao Sun, Kangde Guo, Xing Tang, Zhou, S.Z. "A Real-time Multi-cue Hand Tracking Algorithm Based on Computer Vision". In Virtual Reality Conference (VR) 2010, Pages 219 – 222.
- [10] Yikai Fang, Kongqiao Wang, Jian Cheng, Hanqing Lu. "A Real-Time Hand Gesture Recognition Method". In International Conference on Multimedia and Expo 2007, pages 995 – 998.
- [11] Shuai Zhang, Xiang Chen, Kongqiao Wang, Jiangwei Li, Yanwei Pang, He Yan. "Active Histogram of Oriented Gradient Based Learning for Palm Tracking". Advances in Intelligent and Soft Computing, 2012, Volume 133/2012, 683-690, DOI: 10.1007/978-3-642-27552-4_91.
- [12] T. Ahonen, A. Hadid, and M. Pietikinen. "Face description with local binary patterns: Application to face recognition". IEEE Trans. Pattern Anal. Mach. Intell., 28(12):2037–2041, 2006.
- [13] Chao Zhu, Charles-Edmond Bichot, Liming Chen. "Multi-scale Color Local Binary Patterns for Visual Object Classes Recognition". International Conference on Pattern Recognition (ICPR), IEEE ed. Istanbul, Turkey. pp. 3065-3068, 2010.
- [14] Yikai Fang, Jian Cheng, Jinqiao Wang, Kongqiao Wang, Jing Liu, Hanqing Lu. "Hand Posture Recognition with Co-Training". 19th International Conference on Pattern Recognition (ICPR), December 2008.
- [15] Xiujuan Chai, Yikai Fang, Kongqiao Wang. Robust hand gesture analysis and application in gallery browsing. IEEE International Conference on Multimedia and Expo (ICME), pp938-941. 2009.
- [16] Junge Zhang, Kaiqi Huang, Yinan Yu, Tieniu Tan. "Boosted local structured HOG-LBP for object localization". CVPR'11 Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition Pages 1393-1400.
- [17] Guodong Guo, Dyer C.R., Yun Fu, Huang T.S.. "Is gender recognition affected by age?". IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops), 2009.