Gender Recognition using Hog with Maximized Inter-Class Difference

M. E. Yildirim[1], O. F. Ince[2], Y. B. Salman[3], J. Kwan Song[2], J. Sik Park[2] and B. Woo Yoon[2]
[1]*Department of Electrical Engineering, Bahcesehir University, Istanbul, Turkey*
[2]*Department of Electronics Engineering, Kyungsung University, Busan, South Korea*
[3]*Department of Software Engineering, Bahcesehir University, Istanbul, Turkey*

Keywords: Gender Recognition, Random Forest, Histogram of Oriented Gradients, Inter-Class Difference, Adaboost.

Abstract: Several methods and features have been proposed for gender recognition problem. Histogram of oriented gradients (Hog) is a widely used feature in image processing. This study proposes a gender recognition method using full body features. Human body from side and front view were represented by Hog. Using all bins in the histogram requires longer time for training. In order to decrease the computation time, descriptor size should be decreased. Inter-class difference was obtained as a vector and sorted in a descending order. The bins with the largest value were selected among this vector. Random forest and Adaboost methods were used for the recognition. As a result of both tests, the classifier using first 100 bins with maximum difference gives the optimum performance in terms of accuracy rate and computation time. Although Adaboost performed faster, the accuracy of random forest is higher in full body gender recognition.

1 INTRODUCTION

Object recognition is an important and interesting subject of computer vision. An object recognition system finds objects in the real world from an image, using object models which are known a priori. A human can easily distinguish between a male and female (Bruce, 1993). However, recognizing a gender by a smart system is a challenging task. It is an essential field that can be applied for surveillance systems, medical purposes, content based indexing, biometrics, demographic collection, targeted advertising and human computer interaction.

Most of the existing approaches for gender recognition rely only on facial features (Alexandre, 2010; Wu et al., 2011; Wu et al., 2010; Makinen and Raisamo, 2008). These studies were generally applied to standard databases having high resolution aligned frontal faces. However, people can appear in different scales and viewpoints in real-world images (Khan et al., 2014). In real time conditions, where videos are taken by a Closed Circuit Television(CCTV) system, capturing face with enough details to extract features on it can`t be satisfying. CCTV cameras operating for security are mostly located in quite far distance from people.

Recently, Zhang et al., (2013) proposed two pose-normalized descriptors based on deformable part models for attribute description. Ng et al., (2013) obtained 80.4% accuracy rate by presenting a gender recognition system based on convolutional neural network (CNN) on color images, which automatically learns the most informative features during training. Bourdev et al., (2011) achieved approximately 82.4% accuracy by employing poselets that represent small parts of the body under a specific pose, to recognize several attributes including gender.

Significant aspects of recognition are feature selection and extraction. Performance and accuracy of the system can be increased by using proper features which should conform to several criterias such as uniqueness, performance, collectability, acceptability and circumvention (Gou et al., 2012; Hossain and Chetty, 2011; Yildirim et al., 2014).

The recognition system should have high accuracy rate as well as low processing time and computation load. The accuracy rate can be low due to real world conditions such as poses, clothing style and color, occlusion and shadows. There are only a

Yildirim, M., Ince, O., Salman, Y., Song, J., Park, J. and Yoon, B.
Gender Recognition using Hog with Maximized Inter-Class Difference.
DOI: 10.5220/0005715401060109
In *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2016) - Volume 3: VISAPP*, pages 108-111
ISBN: 978-989-758-175-5

few studies found in literature on body based gender recognition due to such conditions (Ng et al., 2012).

This study proposes a gender recognition method using full body features. Dataset for this work has been created from the CCTV videos taken at random times of the day at certain cameras. People in the vidoes are extracted and represented by Hog.

A hog descriptor has large number of bins. Thus, can cause much computation time for training and testing. One method to increase the processing speed is decreasing the descriptor size. The important point is that, accuracy should stay sufficient while the computation speed is increased. In this sense, we aimed to select the bins that gives the most distinguishing information on classes.

This paper is organized as follows. In section 2, we briefly mention about Hog feature. In section 3, random forest method is explained. In section 4, gender recognition is given with the description of datasets and experiments.

2 HISTOGRAM OF ORIENTED GRADIENTS

Dalal and Triggs (2005) proposed Hog algorithm. The basic idea is that local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients or edge directions, even without precise knowledge of the corresponding gradient or edge positions (Dalal and Triggs, 2005).

The computation procedure is as follows: to obtain a gradient image, a mask with [-1, 0, 1], without smoothing is applied to the image. For the computation of the Hog descriptor the gradient image is divided into 16x16 pixel non-overlapping blocks of four 8x8 pixel cells. After calculating the gradients, they are mapped into 9 bins within a range of $0^\circ - 180^\circ$.

Hog has become popular for various problems of pattern recognition. Collins et al. (2009) developed a descriptor named PixelHOG which was found to perform better than other descriptors based on Pyramid Hog (Bosch et al., 2007) and Pyramid of Words (Lazebnik et al., 2006). Bourdev et al., (2011) combined color histogram, Hog and skin features to represent poselets to gather gender information to make it robust against pose and occlusion.

3 CLASSIFICATION METHODS

3.1 Random Forest

A random forest multiclass classifier consists of a number of trees, with each tree using some form of randomization. The leaf nodes of all trees are labeled by estimations of the posterior distribution over the image classes. Each internal node contains a test that best splits the space of data to be classified. An image is classified by sending it down each tree and aggregating the reached leaf distributions. Randomness can be injected at two points during training: in subsampling the training data so that each tree is grown using a different subset; and in selecting the node tests.

The trees here are binary and are constructed in a top-down manner. The binary test at each node can be chosen in one of two ways: (i) randomly, for example data independent; or (ii) by a greedy algorithm which picks the test that best separates the given training examples. Best here is measured by the information gain

$$\Delta E = -\sum_i \frac{|Q_i|}{|Q|} E(Q_i) \tag{1}$$

caused by partitioning the set Q of examples into two subsets $Q_i$according the given test. Here $E(q)$is the entropy $-\sum_{j-1}^{N} p_j log_2(p_j)$ with p_j the proportion of examples in q belonging to class j, and $|\,.\,|$ the size of the set. The process of selecting a test is repeated for each nonterminal node, using only the training examples falling in that node. The recursion is stopped when the node receives too few examples or when it reaches a given depth (Bosch et al. , 2007).

A test image is passed down through each random tree until it reaches a leaf node. All the posterior probabilities are then averaged and the argmax is taken as the classification result of the image.

3.2 Adaboost

Boosting is a well known statistical method that uses the original distribution of positive and negative examples to compute simple rules also called weak classifiers and combines them to create a stronger classifier. AdaBoost is the most commonly used for binary classification, but it can also handle multiple classes with minor modifications (Freund and Schapire, 1996).

4 GENDER RECOGNITION

4.1 Dataset

In this study, we established and used our own database for gender recognition. The images were captured from the CCTV videos recorded around the campus of a university in Korea. In the captured images, only the pedestrians which are in the range of 15 meters from both sides and 7 meters ahead of the camera. The camera is stationary and standing at a height of 6 meters from the ground. There are 700 samples in each class for training. Each class includes 100 different people with 7 different poses for each person. Figure 1 shows samples from the pedestrian dataset, in which first two images are female and the other two are male samples.

Figure 1: Sample images from the used dataset.

4.2 Experiments and Results

The size of training images was 48x96 pixels. Each image was divided into 55 blocks with 50% overlap. Every block had 4 cells with 2x2 configuration and each cell is represented by 9 gradient bins. As a result, descriptor has 1980 bins per image in histogram.

The tests were conducted in WEKA 3.6 tool. Table 2 presents the accuracy rates for the given dataset. Adaboost and random forest were performed as classifiers to collect results by two types of tests: 5-fold and 10-fold cross validation. In an n-fold cross validation test, the dataset is divided into n subsets. One of these subsets is used for testing and remaining n-1 subsets are used for training. This procedure is repeated for n times.

1980 bins takes much computation time in both random forest (2.72 seconds) and adaboost (18.4 seconds) algorithms. Whereas using 100 bins takes 2.26 and 1.64 seconds in a 10-fold cross validation tests. That's why, we decreased the number of bins used for building the classifier. Table 1 shows the time spent for training for 5, 10, 100, 200 and 1980 bins with random forest and adaboost algorithms.

In order to find out the most appropriate number of bins to select, the procedure is as follows: we have extracted the histogram of each image in the class. The average of each bin is calculated over 700 samples for both classes. Afterwards, absolute difference between two classes is calculated. A predefined number of bins with highest difference value were chosen to represent the class for training and testing steps.

Table 1: Time for training with different methods.

Method	Test	5	10	100	200	1980
Random Forest	5-fold	1.25	1.45	2.23	2.64	2.69
	10- fold	1.36	1.52	2.26	2.65	2.72
Adaboost	5-fold	0.11	0.19	1.33	2.64	18.4
	10- fold	0.13	0.16	1.64	2.48	19.4

Table 2 illustrates the recognition rates for five different sets. The first four sets consist of the first 5, 10, 100 and 200 highest value bins of the inter-class difference vector respectively. The last set includes the full length vector with 1980 bins.

Table 2: Accuracy rates for gender recognition.

Method	Test	5	10	100	200	1980
Random Forest	5-fold	78.1	84.9	89.5	90.4	92.3
	10- fold	78	84.2	89.4	90.6	92.5
Adaboost	5-fold	75.6	76.8	81.7	82.1	85.6
	10- fold	75.1	76	81.3	81.5	84.7

For all cases, random forest reveals more accurate recognition than adaboost. Adaboost can reach its highest accuracy rate of 85.6% when it is using 1980 bins and it is consuming 18.4 seconds. Using 1980 bins gives the highest recognition rate but requires excess time for building the classifiers.

Optimum results are obtained in terms of computation time and accuracy rate with use of 100 bins for both methods. Using 200 bins results in lower accuracy rate when compared to 1980 bins test. However, it consumes almost same time for training. In contrast, 5-bins set takes the smallest computation time but gives lowest accuracy.

For the appointed purpose, the optimum selection is 100 bins. It gives a high accuracy rate of

89.5% for random forest and 81.7% for adaboost at a comparatively low processing time when compared to 200 bins and 1980 bins versions.

5 CONCLUSION

In this paper, a new study is given for gender recognition problem in public areas where facial features can`t be extracted. Inter-class difference vector is maximized and selected number of bins with highest value of this vector are used to build the classifiers with both random forest and adaboost algorithms. Five different sets are used and for our purpose 100 bins set gives the most satisfying results. It gives us 89.5% and 81.7% recognition rates random forest and adaboost respectively.

As a further study, we will apply the same method as multi-feature model with adding colour information or modified features.

ACKNOWLEDGEMENTS

This research was supported by Basic Science Research Program through the National Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (2012M3C1A1048865) and Busan Brain 21 funded by Busan City.

REFERENCES

L. A. Alexandre, 2010. Gender recognition: A multiscale decision fusion approach, *Pattern Recognition Letters*, Vol.31,No.11, pp.1422–1427.

J. Wu, W. Smith, and E. Hancock, 2011. Gender discriminating models from facial surface normals, *PR*, Vol.44, No. 12, pp.2871–2886.

J. Wu, W. Smith, and E. Hancock, 2010. Facial gender classification using shape-from-shading, *IVC*, Vol. 28, No. 6, pp.1039–1048.

E. Makinen and R. Raisamo,2008. Evaluation of gender classification methods with automatically detected and aligned faces, *PAMI*, Vol.30,No.3,pp.541–547.

F. S. Khan, J. van de Weijer, R. M. Anwer, M. Felsberg, and C. Gatta, 2014. Semantic pyramids for gender and action recognition, *IEEE Trans. Image Process.* , Vol.23, No.8, pp.3633–3645.

N. Zhang, R. Farrell, F. Iandola, and T. Darrell, 2013. Deformable part descriptors for fine-grained recognition and attribute prediction*, Proc. IEEE Int. Conf. Comput. Vis*. pp.729–736.

C. B. Ng, Y. H. Tay, and B.-M. Goi, 2013. A convolutional neural network for pedestrian gender recognition, *in Int. Symposium on Neural Networks (ISNN'13), LNCS*, Vol. 7951.

L. D. Bourdev, S. Maji, and J. Malik, 2011. Describing people: A poselet-based approach to attribute classification, in *IEEE Int. Conf. on Comp. Vis*. (ICCV).

J. Gou, L.Gao, P. Hou, and C. Hu, 2012. Gender recognition based on multiple scale textural feature, *5th Int. Congress on Image and Signal Process.,* Sichuan, China.

S. M. E. Hossain and G. Chetty, 2011. Next generation biometric identity verification based on face- gait biometrics*, Int. Conf. on Biomedical Eng. and Tech*. , Kuala Lumpur, Malaysia.

Mustafa E. Yildirim, J. S. Park, J. Song, and B. W. Yoon, 2014. Gender Classification Based on Binary Haar Cascade, *International Journal of Computer and Comm. Eng.* , Vol. 3, No.2.

C. B. Ng, Y. H. Tay, and B. M. Goi, 2012. Vision-based human gender recognition: A survey, in Proc. Pacific Rim Int. Conf. Artif. Intell.: Trends Artif. Intell., pp. 335–346.

N. Dalal and B. Triggs, 2005. Histograms of oriented gradients for human detection, Proc. of the IEEE Computer Society Conf. on Comp. Vis. and Patt. Rec., Vol.1, pp.886-893.

A. Bosch, A. Zisserman, 2007. Representing shape with spatial pyramid kernel, *in Proc. of the 6th ACM Int. Conf. on Image and video retriveal*. pp. 401-408.

S. Lazebnik, C. Schmid, and J. Ponce, 2006. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories*, in IEEE Comp. Society, Conf. on Comp. Vis. and PR*. Vol. 2, pp. 2169-2178.

A. Bosch, A. Zisserman, and X. Munoz, 2007. Image Classification Using Random Forests and Ferns, Proc. 11th IEEE Int. Conf. Comp. Vis. Conf.

M. Collins, J. Zhang, and P. Miller, 2009. Full body image feature representations for gender profiling, in IEEE 12th Int. Conf. On Comp. Vis. Workshops, pp.1235-1242.

Y. Freund and R. Schapire, 1996. Experiments with a new boosting algorithm, *Int. Conf. Machine Learning*.