PCA-SVM 을 이용한 Human Detection 을 위한 HOG-Family 특징 비교

Evaluation of HOG-Family Features for Human Detection using PCA-SVM

Nurul Arif Setiawan*, and Chil Woo Lee**

Abstract Support Vector Machine (SVM) is one of powerful learning machine and has been applied to varying task with generally acceptable performance. The success of SVM for classification tasks in one domain is affected by features which represent the instance of specific class. Given the representative and discriminative features, SVM learning will give good generalization and consequently we can obtain good classifier. In this paper, we will assess the problem of feature choices for human detection tasks and measure the performance of each feature. Here we will consider HOG-family feature. As a natural extension of SVM, we combine SVM with Principal Component Analysis (PCA) to reduce dimension of features while retaining most of discriminative feature vectors.

핵심어: SVM, PCA, Human detection, dimensionality reduction, HOG.

1. Introduction

Human detection is one important research in Human-Computer Interaction. Given the ability to detect human, then a computer can do situational and environment analysis and/or providing human with services. Human detection is considered difficult since human appearance is highly varied. The complexity of the problem is added with uncontrolled dynamic environment.

There are two important components of learning-based detector (where the art of human detector lies): features and learning machine algorithm. It is important that features can robustly represent the objects that we want to find and those features can be separated from non-objects optimally and good machine learning must offer good generalization and robust discrimination of classes.

Support Vector Machine (SVM) is one of powerful learning machine and has been applied to human detection task with acceptable performance [1]. The success of SVM for classification tasks in one domain is affected by observation modeled as features which represent the instance of specific class. Given the discriminative features, then SVM learning will give good generalization and consequently we can obtain good classifier. However, finding good features to represent human with all possible pose and environment variation is considered difficult. Several human detection systems using redundant features and then utilizing Adaboost algorithm to learn important features [11]. Recently, Histogram of Oriented Gradient (HOG) descriptor has been used successfully applied for human detection with Support Vector Machine (SVM) as the learning machine [1]. As shown in [2], HOG is one robust feature for human detection with several strong points. HOG encode object's shape by using gradient structure, it captures objects spatial information by grid quantization, and the local normalization makes it illumination invariance.

There are two main issues that we will discuss in this paper. First is evaluation of several variant of HOG which we will call HOG-family. These include non-overlapped, dense [1], spatial pyramidal [3] and multiscale image pyramid of histogram. Second, is evaluation of SVM combined with PCA. The combination is expected will mainly affect on the

*주저자 : 전남대학교 전자컴퓨터공학과 석사 e-mail: arif@image.chonnam.ac.kr

**교신저자 : 전남대학교 전자컴퓨터공학과교수; e-mail: leecw@chonnam.ac.kr

speed of SVM process as the feature dimension is reduced by PCA. With this we want to reduce processing time without losing the original performance.

This paper is organized as follows. In section 2, we will describe the HOG-family features. SVM and combination of PCA for feature reduction with SVM will be discussed in section 3. In section 4 we will describe the data sets and performance measure and give the performance comparison and analysis. Section 5 will conclude this paper and outline future direction of this research.

2. HOG-family Features

Utilization of orientation histogram as shape encoding descriptors has been used in [5] for hand gesture recognition. Later it was developed into robust local feature descriptors known as SIFT [6]. In SIFT, the features are computed at a sparse set of scale-invariant key points, rotated to align their dominant orientations and used individually. The HOG as proposed by [1,2], are computed in dense grids at a single scale without dominant orientation alignment. The grid position of the block implicitly encodes spatial position relative to the detection window in the final feature vector. In addition to dense HOG [1], we will evaluate non-overlapped HOG as the basic HOG variant. We will evaluate spatial pyramid HOG and multiscale image pyramid HOG. Each of the features encoded different shape information from gradient image.

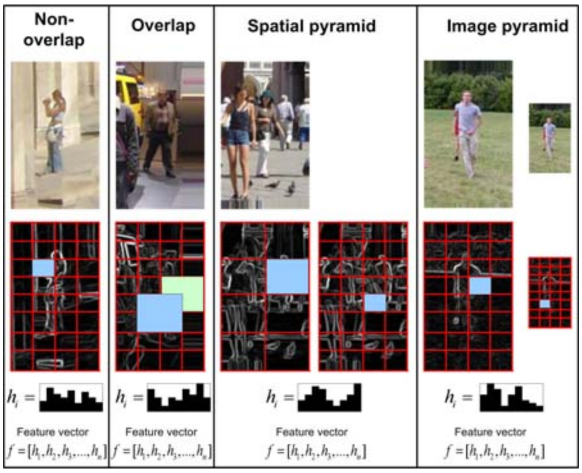


Figure 1. HOG Family features for human detection.

The HOG-family features are shown in Fig. 1 and can be described as follows:

1. Non-overlapped. This is the simplest HOG representation. Histogram of each cell is normalized with the block norm.
2. Overlapped (Dense). This is close to [1] implementation and it uses overlapped blocks as the basic descriptor.
3. Spatial pyramid. The pyramid HOG is build by using different cells number of each levels and then concatenate histogram form each level to create the feature.
4. Multiscale. The image pyramid structure is created from the image and histogram is calculated at each level. The feature vector is the concatenation of histogram from each level.

2.1 HOG Calculation

Let I denote an image of width m and height n and $I(x,y)$ denote the pixel intensity in position (x,y), then the HOG descriptor can be computed by the following:

1. Preprocessing image by using gamma correction.
2. From image I, we compute image gradient in x and y direction by 1-D centered mask [-1, 0, 1]

   where $g_x(x,y)$ and $g_y(x,y)$ denotes the x and y

   $$g_x(x,y) = I(x+1,y) - I(x-1,y)$$
   $$g_y(x,y) = I(x+1,y) - I(x-1,y) \quad (1)$$

   components of the image gradient, respectively.
3. The magnitude $m(x,y)$ and the orientation $\Theta(x,y)$ are computed by

   $$m(x,y) = \sqrt{g_x(x,y)^2 + g_y(x,y)^2}$$
   $$\theta(x,y) = \tan^{-1}(g_y(x,y)/g_x(x,y)) \quad (2)$$

   The orientation is ranged from [0-2π].
4. We divide the image into s_wxs_h non-overlapping cells. For each cells, we quantize the orientation $\Theta(x,y)$ for all pixels into s_b orientation bins weighted by its magnitude $m(x,y)$.
5. The feature is normalized by the sum of blocks. In all HOG variant, we use a magnitude of a block of 2x2 cells to normalize each of cells. By blocks normalization we capture the information in the surroundings cells also included.
6. The histogram from each cell is formed into one feature vector with configuration for each HOG variant is shown in Fig. 1.

3. SVM Classifier

SVM are a set of related supervised learning methods used for classification and regression. They belong to a family of generalized linear classifiers. A special property of SVM is that they simultaneously minimize the empirical classification error and maximize the geometric margin; hence they are also known as maximum margin classifiers [7,8].

Support vector machines use kernel to map input vectors to a higher dimensional space where a maximal separating hyperplane is constructed. Two parallel hyperplanes are constructed on each side of the hyperplane that separates the data. The separating hyperplane is the hyperplane that maximizes the distance between the two parallel hyperplanes. The parallel hyperplane on each side (positives and negatives) is represented by sets of trained features data points known as support vectors (SVs). For details treatment on SVM we suggest the reader to refer to [8,9]. In this evaluation, we use SVM-Light library implementation [10].

In classification problem, given a trained SVM model, a new observed data points can be evaluated by [9]:

$$y(\mathbf{x}) = \sum_{n=1}^{N} a_n t_n k(\mathbf{x}, \mathbf{x}_n) + b \quad (3)$$

where parameters $\mathbf{x}_n$ is the SV, t_n is sign of SV where $t_n = \{1, -1\}$, a_n is multipliers come from Lagrange constraint where a_n is non zero for SVs, b is the bias parameters, N is number of SVs, and $\mathbf{x}$ is the new observed data and $k(\mathbf{x}, \mathbf{x}_n)$ is the kernel function.

3.1 SVM with PCA

Speeding up SVM has been an interest to bring SVM for practical and real-time problem. Based on Eq. 3, there are two options to speed up SVM: Reducing the number of SVs N or reducing the size of features vector $\mathbf{x}$. The motivation to use PCA is to preprocess features by reducing its size while maximizing feature variance before we use SVM. SVM involves inner product spanned by the feature vectors. This is represented by kernel function in Eq. 3. The longer the features, more time is needed. Thus PCA is used to speed-up SVM process by reducing features while simultaneously retaining representative parts. This also reduces memory requirement for SVM training. However, projected features by PCA will change the result of SVM since basically SVM just sees another features set. It has been shown in [12] that SVM is invariant under PCA transform thus the properties of SVM are retained.

To obtain PCA from our data sets, we use all both positives and negatives training images since PCA do not handle labeled data. If we only used positives images then we will obtain positive data' s principal component only while in testing stage, we do not know whether the new observed data is positives or negatives.

4. Performance Evaluation

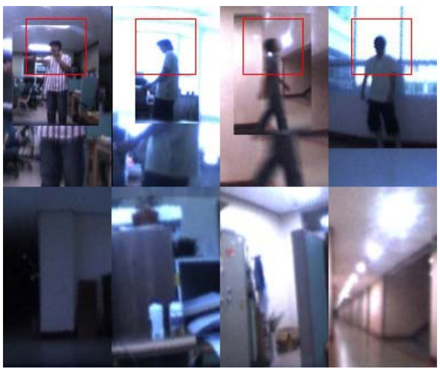
4.1 Evaluation Measure

To quantify the HOG-family performance on binary classifier SVM we use Precision Recall (PR) curve. We use this measure because we are interested in knowing how many of the objects it detects, and how often the detections it makes are false. Another reason is that most of the majority of data is negatives, small number of true positives will gives small effect to e.g. DET or ROC curve shape, and thus it is difficult to observe the differences in performance.

PR curve measure the proportion of recall defined as true positive against the proportion of precision which is defined as number of true detected divided by total true hypothesis.

4.2 Data Sets

The main purpose of this evaluation is to choose a suitable implementation of HOG for human detection system in omnidirectional camera mounted on a mobile robot [4]. Thus we have created new data sets to represent the application domain of our system. The data sets are taken by our omnidirectional camera system so the training images are representative for our purpose. Because the omnidirectional image is converted into panoramic, sampling step degrades the



quality of edges and the image contains many noise and artifact.

Figure 2. ImageLab person data sets (top) positive normalized - the rectangle show 48x48 upper body part - and (bottom) negative images.

The training data sets contains 249 96x160 normalized positives images of person's body as seen in panoramic images in various poses in different illumination condition. We use 98 negatives images which then we sampled for training the SVM classifier. Although the dataset has full body size, we only focuses on upper body 48x48 detector since we want the detector to detect human in radius 1−4 m around the robot. If human is too close then we cannot detect the lower part of the body. The test data sets consist of 100 positives normalized person images and 85 negatives images.

We only use one run of training without re−training by augmenting images from false positives obtained with initial training set as used in [2]. The test phase also performed on normalized positives images. Although this might seems restrictive, we can assume given the generalization of the SVM and good choice of training data sets, if a feature perform well in this data sets then it will perform comparably well in multiscale detection schema as used in practice [2,4].

The evaluation will be performed under some fixed HOG parameters. For our data set (detection of 48x48 upper body), each cells size is 12x12 pixels, blocks is defined by 2x2 cells with stride is 12 pixels in both direction (for dense type). We use 249 positives and 953 negatives images for training. The testing data sets contain 100 positives and 7332 negatives.

To evaluate performances of PCA−SVM we use two projection sizes: one−third and one−sixth of original feature size. This is shown in Table 1. For the SVM kernel we use RBF kernel.

Table 1. Full size and PCA projected size of the HOG-family

HOG−type	Full size	1/3	1/6
Non−overlap	144	48	24
Dense	324	108	54
Spatial pyramid	189	63	31
Multiscale	288	96	48

4.3 Evaluation Result

The performances of HOG−family in PR curves can be seen in Fig. 3. For full size feature, at 0.8 precision, multiscale−HOG which we propose has better recall compared to other HOG variant including dense−HOG from [1,2].

We also can see that PCA reduces performance significantly in 1/6 feature size and for 1/3 feature size, dense−HOG and multiscale−HOG have comparable performance compared to full feature size. To make this clear, Fig. 4 shows the comparison of full feature size and 1/3 PCA projected feature for dense HOG and multiscale HOG.

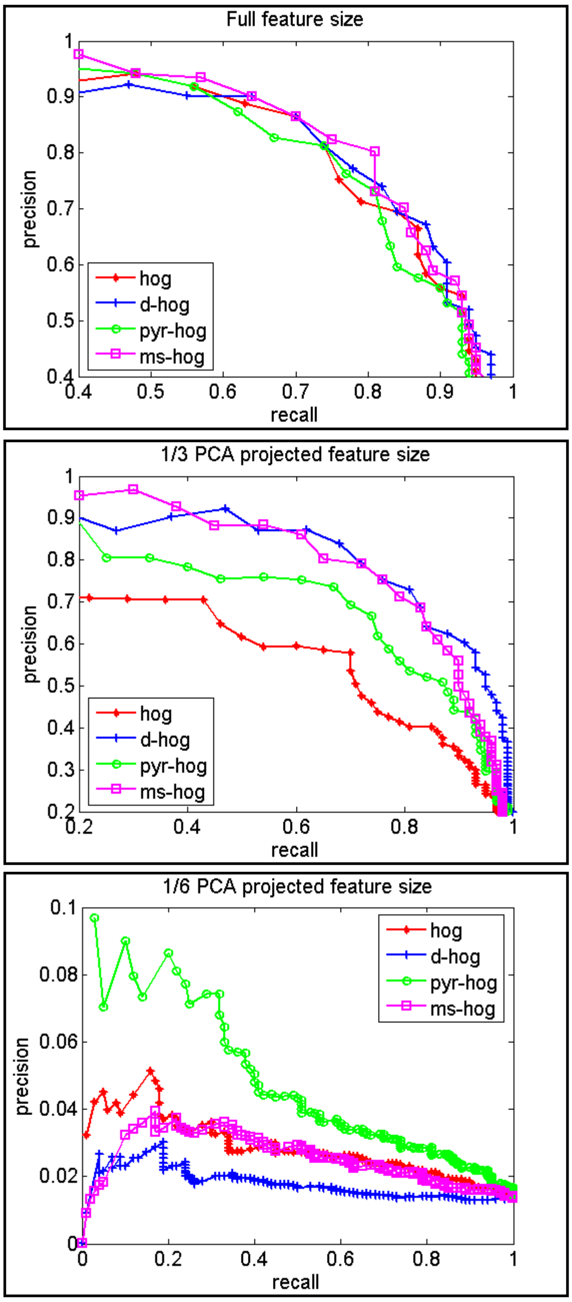
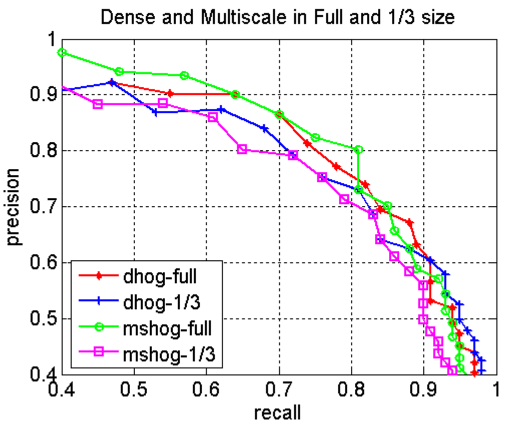


Figure 3. Performances on ImageLab data set with RBF kernel. (Top) full feature size (middle) 1/3 PCA projected feature and (bottom) 1/6 PCA projected feature

To show the advantage of PCA-SVM the processing time to extract feature and run svmlight classification binary for test data set is shown in Table 2. We only consider dense and multiscale since the performance of non-overlap and pyramid on PCA projected feature is much worse than original full size.

Figure 4. Performances of full feature size and 1/3 PCA projected feature size for dense and multiscale HOG.



Using PCA, at 0.8 precision, recall of multiscale-HOG is dropped nearly 10% while for dense-HOG the recall is dropped approximately 4%. From table 2, PCA can reduce processing time, however, at the expense of reduces performance as we can see in Fig. 3 and Fig. 4.

Table 2. Processing time of dense and multiscale HOG for all test data set.

HOG-type	Feature size	Feature extraction (ms)	SVM classification (s)
Dense - full	324	23911.69	1.75
Dense 1/3	108	18128.58	0.88
Multiscale - full	288	30830.73	1.68
Multiscale 1/3	96	26992.65	0.88

4.4 Discussion

We have shown the evaluation performance of four variant of HOG: non-overlap, overlap (dense), spatial pyramid and multiscale. The results that we obtain show several important points for analyzing feature or descriptors and understanding SVM property for classification task.

First is how consistent the feature in representing the object, in our case human body shape. Spatial-HOG is based on assumption that the pyramid can capture spatial layout by capturing histogram is several spatial scale. Non-overlap HOG also capture spatial information by using blocks normalization and it is much simpler (in term of feature size) with performance is better in several test situation. Multiscale-HOG try to capture the information at two scale and with feature size is smaller than dense-HOG, we have shown that it has comparable performance with dense-HOG.

The second is insight into how SVM treats the feature. By the kernel trick, SVM map feature vectors into higher dimension by only retaining its dot product. Thus, any spatial advantage of feature vectors in unobservable since SVM produce hyperplane from points in kernel space. Understanding which component of features favorable in classification or retain the spatial configuration can be advantageous in human detection task. We are planning to investigate this in the future.

Third is performance of PCA-SVM for human detection is affected by the feature chosen and the data sets. It also depends on size of projection that we choose. Generally, PCA worsen the performance so using PCA might not give any gain even if we can reduce processing time given a challenging data set. If the performance penalty is acceptable then using PCA is preferable. We infer that there is an optimal projection size where PCA can give comparable performance to original size.

Finally we note that PCA-SVM used in current research is not a principled or integrated approach to reduce feature size because SVM change the feature points is kernel space. That is, no guarantee the new point is same to original point. This makes the performance penalty is unpredictable. We will explore Joint classifier and feature optimization (JCFO) [13] which seeks sparsity in its use of both basis functions and features.

5. Conclusion

This paper has presented an extensive evaluation of HOG-family features for human detection task. We have assessed the performance of each feature with SVMs classifier and show the effect of applying PCA for reducing features. Our novel multiscale image HOG shows best performance compared to well known dense HOG type in RBF kernel. More integrated combination of reconstructive capability of PCA into SVM machine and multiscale detection will be carried out in the future.

References

This work was supported in part by MIC & IITA through IT Leading R&D Support Project.

References

[1] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: International Conference on Computer Vision & Pattern Recognition. Volume 2. (2005) 886 893.

[2] Dalal, N.: Finding people in images and videos. PhD thesis, Institut National Polytechnique de Grenoble (2006).

[3] Bosch, A., Zisserman, A., Munoz, X.: Representing shape with a spatial pyramid kernel. In: CIVR' 07: Proceedings of the 6th ACM international conference on Image and video retrieval. (2007) 401 408

[4] Arif, S.N.: Development of omnidirectional human detection module for mobile robot. Master' sthesis, Chonnam National University (2008)

[5] Freeman, W.T., Roth, M.: Orientation histogram for hand gesture recognition. In: International Workshop on Automatic Face and Gesture Recognition. (1995) 296 301

[6] Lowe, D.G.: Distinctive image features from scale invariant key points. International Journal of Computer Vision 60(2) (2004) 91 110

[7] Boser, B.E., Guyon, I.M., Vapnik, V.N.: A training algorithm for optimal margin classifiers. In: COLT ' 92: Proceedings of the fifth annual workshop on Computational learning theory. (1992) 144 152

[8] Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery 2(2) (1998) 121 167

[9] Bishop, C.M.: Pattern Recognition and Machine Learning (Information Science and Statistics). Springer (2006)

[10] Joachims, T.: Making large-scale SVM learning practical. In Schlkopf, B., Burges, C., Smola, A., eds.: Advances in Kernel Methods - Support Vector Learning, MIT Press (1999)

[11] O. Tuzel, F. Porikli, and P. Meer. Human detection via classification on Riemannian manifolds. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, pages 1 8, 2007.

[12] H. Lei, V. Govindaraju. Speeding Up Multi-class SVM by PCA and Feature Selection. Feature Selection in Data Mining (FSDM05), The 5th SIAM International Conference on Data Mining Workshop, California, USA, 2005.

[13] Krishnapuram, B.; Harternink, A.J.; Carin, L.; Figueiredo, M.A.T., "A Bayesian approach to joint feature selection and classifier design," IEEE Transactions on Pattern Analysis and Machine Intelligence , vol.26, no.9, pp. 1105-1111, Sept. 2004