ICIC Express Letters
Part B: Applications
Volume **7**, Number **3**, March **2016**

 ISSN 2185-2766
pp. **607–612**

ONE NOVEL PHOG DESCRIPTOR FOR HANDWRITTEN DIGITS RECOGNITION

Hongyan Zhang[1,*] and Genping Lei[2]

[1]College of Information Science and Technology
Zhengzhou Normal University
No. 6, Yingcai Street, North Daxue City, Zhengzhou 450044, P. R. China
[*]Corresponding author: zhy86866@163.com

[2]Information Engineering Department
Henan Mechanical and Electrical Vocational College
Zhengzhou 451191, P. R. China

Received September 2015; accepted December 2015

Abstract. *Handwritten digits recognition is a current hot issue and also a challenging problem. Extracting robust features from handwritten digits is a core problem. In this paper, we used the PHOG (Pyramid Histogram of Oriented Gradient) method to well describe the characteristics of handwritten digits and get our PHOG descriptors. HOG can well describe the local appearance and shape of handwritten digits by computing the gradient or edge direction density distribution. It combines the histograms of oriented gradient of local area (such as 10×10 pixels block) in handwritten digits image to form the feature. In order to better represent the granularity of block, we construct a three-tier pyramid HOG with the different size of blocks to form the descriptor for each handwritten digit image. Finally, in order to reduce the dimensionality as well as drop the redundant information of descriptors, we adopt the classical linear dimensionality reduction algorithm PCA (Principal Component Analysis) to learn a subspace and get our novel PHOG descriptor. We conduct experiment on the MNIST handwritten digits database with 2,000 training samples and 2,000 testing samples, and the experiment results have shown the good accuracy and effectiveness of our descriptor.*

Keywords: Handwritten digits recognition, Pyramid histogram of oriented gradient, Three-tier pyramid HOG, Principal component analysis

1. **Introduction.** Handwritten digits recognition, a branch of image recognition is the common issue in the field of pattern recognition and computer vision. It plays an important role in postal mail sorting, bank check amounts processing and other forms of data processing. However, when it comes to the practical application, handwritten digits recognition is still a challenging problem. The recognition accuracy and recognition speed both are the primary problem to be solved.

The difficulties in handwritten digits recognition are mainly the following two points.

1) The glyph of digit has little difference, and possesses simple strokes. These intrinsic properties make it difficult to distinguish handwritten digits.

2) Although only 10 kinds of digits exist, the handwriting ways vary widely and have regional characteristics, which leads to poor commonality of recognition algorithms.

Different algorithms have been proposed to acquire corresponding solutions to these problem. [1] used the separability measure to get the corresponding kernel parameter and a fast method for handwritten digit recognition was proposed. [2] adopted the artificial neural networks algorithm (RBMs) and achieved an excellent effect in handwritten digit recognition. [3] extracted the gist feature for handwritten digits and adopted random oblique decision trees for classification. However, all these algorithms cannot well describe the local appearance and shape of handwritten digits images, while our proposed algorithm does describe the local target by gradient or the edge direction density distribution well in handwritten digits images. Different from the above algorithms, we used the PHOG method to well describe the characteristics of handwritten digits and adopt the classical linear dimensionality reduction algorithm PCA to learn a subspace. Finally, we got our robust PHOG descriptors. Our proposed algorithm shows good performance both on recognition rate and recognition speed.

The content of this paper is organized as follows. In Section 2, histogram of oriented gradient and the classical linear dimensionality reduction algorithm PCA will be introduced. Our novel PHOG descriptor for handwritten digits recognition is proposed in Section 3. In Section 4, experiment setup and results are described followed by the conclusion in Section 5.

2. Related Work.

2.1. Histogram of oriented gradient (HOG).

HOG feature is a kind of object detection feature descriptor in the field of computer vision. HOG can well describe the local appearance and shape of handwritten digits by computing the gradient or edge direction density distribution. It combines the histograms of oriented gradient of local area (such as 10×10 pixels block) to form the feature, which also possesses a certain degree of invariance against local geometric and photometric transformations. Navneet Dalal and Bill Triggs put forward the HOG feature for pedestrian detection, which achieved great success.

In general, the main steps of HOG feature extraction are as follows.

1) Standardize gamma space and color space.

2) Compute the gradient of original image. Figure 3 reveals the gradient of handwritten digits images.

3) Construct histogram of oriented gradient for each 'cell'.

4) Combine the 'cells' into the larger 'block' and implement normalization for the block.

5) Collect the histograms of oriented gradient of local area into the final descriptor.

The main steps of HOG feature extraction are shown in Figure 1.

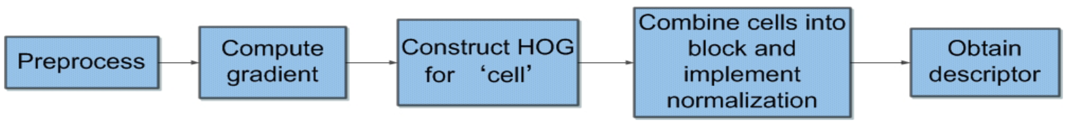

Figure 1. The main steps of HOG feature extraction

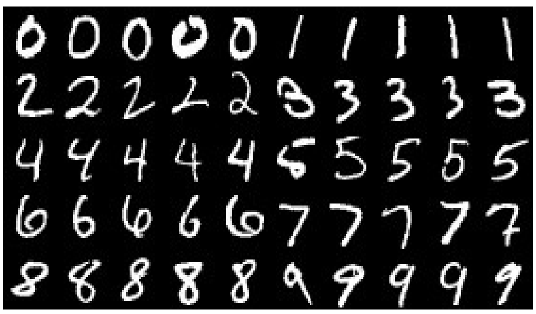

Figure 2. The original handwritten digits images

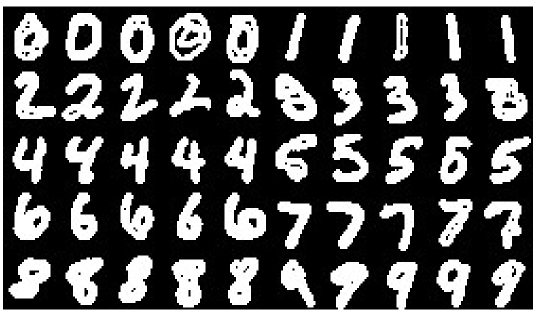

Figure 3. The gradient of handwritten digits images

2.2. **The classical linear dimensionality reduction algorithm PCA.** High-dimensional data is common in practical applications, such as text, images and videos whose dimensionalities are thousands or ten thousands. However, it is reasonable to suspect that the high-dimensional real data of nature probably resides on a manifold with a lower dimensionality. That is, a small number of features (parameters) can well characterize the underlying structure of real data in some cases. This leads to represent the data in a lower dimensional space, where the intrinsic structure of data can be well visualized [5,6].

PCA (Principal Component Analysis) is one of the most typical dimensionality reduction methods, also called '$Eigenface$'. The PCA aims to preserve the global Euclidean structure by exploring a set of basis functions which are mutually orthogonal to pursue the directions of maximum variance. When the data is embedded in subspace which is linear or almost linear, PCA can be guaranteed to learn a subspace.

Let $X = x_1, x_2, \ldots, x_i, \ldots, x_N$ denote the samples set with N samples. Then, we can capture the subspace whose basis functions correspond to the directions of maximum variance [7],

$$y_i = W^T x_i, \quad i = 1, 2, \ldots, N \tag{1}$$

where, W denotes the projection matrix which maps the original n-dimensional space onto the d-dimensional linear subspace ($d << n$), $W = a_1, a_2, \ldots, a_i, \ldots, a_d$, ($\lambda_1 > \lambda_2 > \ldots > \lambda_d$). We can obtain the W by solving the following eigen decomposition

$$Qa_i = \lambda_i a_i \tag{2}$$

where $Q = XX^T$ is the covariance matrix.

3. **Our Novel PHOG Descriptor for Handwritten Digits Recognition.** In this section, we will describe our novel PHOG descriptor for handwritten digits recognition in detail which is different from the HOG in some aspects. Firstly, we construct a three-tier pyramid HOG to form the descriptor for each handwritten digits image. Then, in order to reduce the dimensionality as well as drop the redundant information of descriptors, we adopt the classical linear dimensionality reduction algorithm PCA to learn a linear subspace for the descriptors.

Local appearance and shape of handwritten digits images can be well described by HOG. It combines the histograms of oriented gradient of local area (such as 10×10 pixels block) in handwritten digits image to form the feature, which also possesses a certain degree of invariance against local geometric and photometric transformations. In order to better represent the granularity of block, we construct a three-tier pyramid HOG with the different size of blocks to form the descriptor for each handwritten digits image.

The specific steps of constructing a three-tier pyramid HOG are as follows.

1) Enlarge handwritten digits images. The size of original handwritten digits image on MNIST database is 28×28. In order to obtain the efficient HOG descriptor, we resize the image into 56×56.

2) Compute the gradient information for each handwritten digits image. In order to capture the contour of the digit, we simply use a template $[-1\ \ 0\ \ 1]$ as gradient operator to perform a convolution operation with the original image at each pixel. Then we will get the gradient in x and y directions respectively, the gradient magnitude and gradient orientation.

$$G_x(x, y) = Digit(x + 1, y) - Digit(x - 1, y) = Digit(x, y) * [-1 \quad 0 \quad 1] \tag{3}$$

$$G_y(x, y) = Digit(x, y + 1) - Digit(x, y - 1) = Digit(x, y) * [-1 \quad 0 \quad 1] \tag{4}$$

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \tag{5}$$

3) Use Canny operator to detect the edge of the handwritten digits. Figure 7's 'Detect edge by Canny' shows the edge extracted by this operator. Firstly, smooth the handwritten digits image with the Gaussian filter. Then, compute the gradient magnitude and orientation. Thirdly, carry out the non-maxima suppression on gradient magnitude. Finally, employ the double-threshold algorithm to detect and connect edge and let E denote the edge points set.

4) Adopt the 8 connecting area method to label the connecting area of the edge points matrix. Then, we can compute the gradient magnitude at each edge point by Equation (6) and which bins it belongs to at each edge point (see Figure 8 and Figure 9). And 8 connecting area is shown in Figure 6. What the 8 connecting area is; Figure 4 denotes the sample of connecting area and '1' denotes to be connected while '0' not connected. Figure 5 shows what the 4 connecting area (only 4 directions: up, down, left and right) is. Figure 6 is the 8 connecting area (8 directions: up, down, left, right and the 4 diagonal directions.)

$$G_{Canny}(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \tag{6}$$

where $(x, y) \in E$, E denote the edge points set.

$$Angle(x, y) = \arctan\left(G_y(x, y)/G_x(x, y)\right) \tag{7}$$

1	0	1	0	0	0	0	0	0	1
1	1	0	0	0	0	0	0	1	1
0	1	0	0	0	0	1	1	0	0
0	1	1	0	0	0	0	0	0	0

FIGURE 4. The sample of connecting area

1		2							4
1	1							4	4
	1					3	3		
	1	1							

FIGURE 5. 4 connecting area

1		1							2
1	1							2	2
	1					2	2		
	1	1							

FIGURE 6. 8 connecting area

5) Construct a three-tier pyramid HOG with the different size of blocks for each handwritten digits image. In order to better represent the granularity of block, we construct a three-tier pyramid HOG with the different size of blocks. Then, employ the L2hys method to normalize each histogram of oriented gradient of every 'block'. The three-tier pyramid HOG is demonstrated in Figure 7. The parameters in this paper we choose are: bin is 10, and angle is 360.

For each block, we will get a histogram of oriented gradient with 10 bins, and meanwhile, we use the gradient magnitude as the vote weight for each point which belongs to the corresponding 'bin'. See Equation (8) to understand this.

$$h_{block}(b) = \sharp\left\{G_{Canny}(x, y) \,\middle|\, \frac{Angle(x, y)}{nAngle} \in bin(b)\right\} \tag{8}$$

Among them, $b = 1, 2, \ldots, 10$; $nAngle = 36°$

6) Collect the histograms of oriented gradient over all the 'blocks' and use PCA to learn a linear subspace for the descriptors. Finally, we will obtain our novel PHOG descriptor for handwritten digits recognition.

4. **Experiment Setup and Results.** In this section, we will design corresponding experiment on MNIST handwritten digits database to test and validate the performance of our algorithm mentioned in Section 3. MNIST handwritten digits database, a subset of NIST database, was built from the NIST's SD-3 and SD-1. Among them, SD-3 was collected among Census Bureau employees and SD-1 was collected among high-school students [8]. In this paper, we choose the former 400 images for each digit from 0 to 9 to construct our own sub MNIST database with 4,000 images of '*bmp*' format

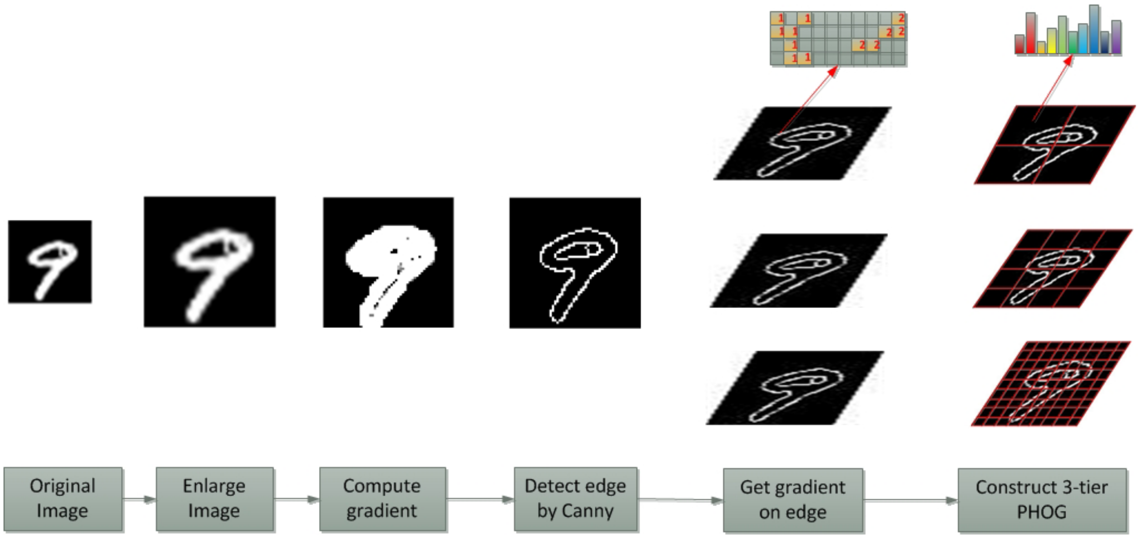

FIGURE 7. The construction process of a three-tier pyramid HOG with the different size of blocks: 'Get gradient on edge' in the fifth column computes the gradient on edge by the 8 connecting area method; 'Construct 3-tier PHOG': the i-th tier contains 2^i blocks and we will get a histogram with 10 bins for each block

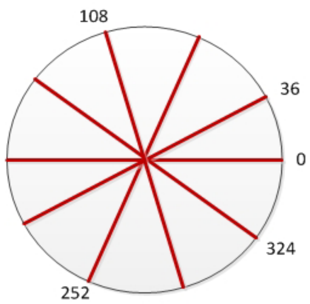

FIGURE 8. The bins of gradient orientation (each bin for 36°)

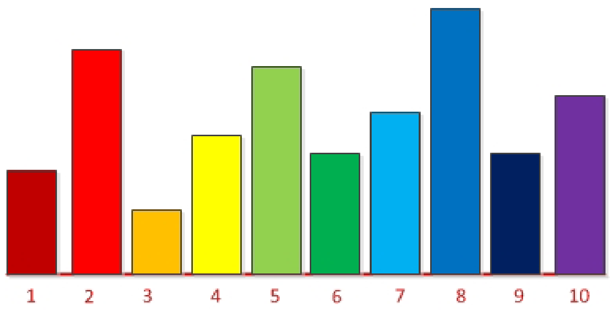

FIGURE 9. The histogram of oriented gradient of each block with 10 bins

in total. We design 5 sets of training data with 10%, 20%, 30%, 40%, 50% of our sub MNIST database respectively, and put the remaining data into test set, that is $T_r 10\%/T_e 90\%, T_r 20\%/T_e 80\%, \ldots, T_r 50\%/T_e 50\%$. We iterate 10 times for each group of training and testing data, and see the average value as the final recognition rate. The part of handwritten digits images in our sub MNIST database is shown in Figure 2.

The recognition rates of various algorithms are demonstrated in Table 1. Among them, '1-NN' denotes the 1-nearest neighbor algorithm and 'PHOG' does our proposed algorithm, where '1-NN' (when $K = 1$) is just the special case of 'KNN'.

TABLE 1. The recognition rates of various algorithms on our sub MNIST database

Algorithm	$T_r 10\%/T_e 90\%$	$T_r 20\%/T_e 80\%$	$T_r 30\%/T_e 70\%$	$T_r 40\%/T_e 60\%$	$T_r 50\%/T_e 50\%$
1-NN	0.8133	0.8495	0.8704	0.8865	0.8943
PHOG	**0.8873**	**0.9081**	**0.9251**	**0.9334**	**0.9370**

Here, we will demonstrate the difference in computing time cost between two kinds of algorithms. There are 5 sets of training and testing data and we iterate only one time on the machine whose CPU is Inter(R) Core(TM) i5-3230 2.60GHz and the memory is 4.00GB, as is demonstrated in Table 2.

TABLE 2. The computing time cost (seconds) between two kinds of algorithms on our sub MNIST database

Algorithm	1	2	3	4	5
1-NN	14.578	31.028	40.192	46.823	49.312
PHOG	**7.609**	**19.495**	**27.068**	**40.195**	**40.872**

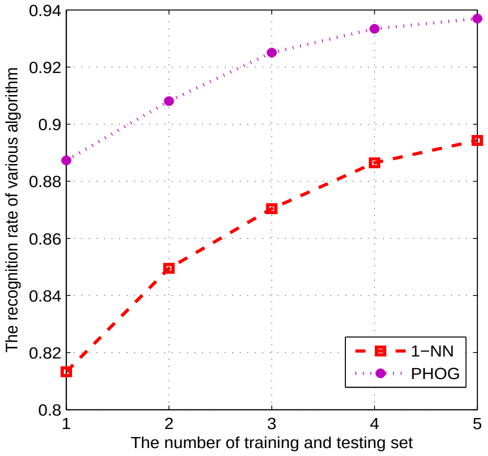

FIGURE 10. The recognition rate of various algorithms on our sub MNIST database

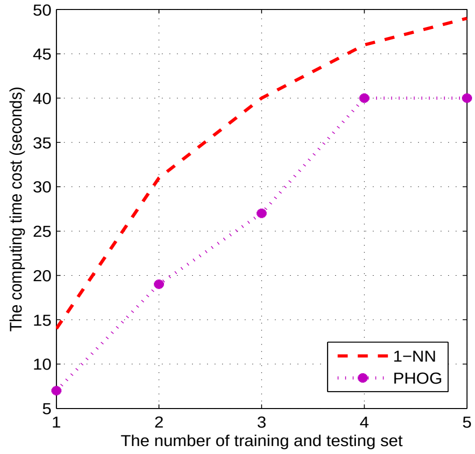

FIGURE 11. The computing time cost between two kinds of algorithms

From Figure 10 corresponding to Table 1 and Figure 11 corresponding to Table 2, we can safely draw the conclusion that our novel algorithm is better than the original algorithm both on accuracy and computing speed.

5. **Conclusion.** In this paper, we construct a three-tier pyramid HOG with the different size of blocks to form the descriptor for each handwritten digit image aiming to better represent the granularity of block. And, in order to reduce the dimensionality as well as drop the redundant information of descriptors, we adopt PCA dimensionality reduction algorithm to learn a subspace and finally get our novel PHOG descriptor. The experiment results have revealed that our novel PHOG descriptor performs well both in accuracy and efficiency.

REFERENCES

[1] Q. Li and L. Chen, A fast handwritten digit recognition algorithm based on improved SVM, *The 1st International Conference on Information Sciences, Machinery, Materials and Energy*, 2015.
[2] K. O'Shea, *Massively Deep Artificial Neural Networks for Handwritten Digit Recognition*, arXiv preprint arXiv:1507.05053, 2015.
[3] T. N. Do and N. K. Pham, Handwritten digit recognition using GIST descriptors and random oblique decision trees, *Some Current Advanced Researches on Information and Computer Science in Vietnam*, pp.1-15, 2015.
[4] N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol.1, pp.886-893, 2005.
[5] X. He, D. Cai, S. Yan et al., Neighborhood preserving embedding, *The 10th IEEE International Conference on Computer Vision*, vol.2, pp.1208-1213, 2005.
[6] D. Cai, X. He and J. Han, Semi-supervised discriminant analysis, *IEEE the 11th International Conference on Computer Vision*, pp.1-7, 2007.
[7] A. M. Martínez and A. C. Kak, PCA versus LDA, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.23, no.2, pp.228-233, 2001.
[8] *MNIST Handwritten Digits Database*, http://yann.lecun.com/exdb/mnist/.