CSCanada
Advances in Natural Science
Vol. 9, No. 1, 2016, pp. 1-4
DOI:10.3968/8444

ISSN 1715-7862 [PRINT]
ISSN 1715-7870 [ONLINE]
www.cscanada.net
www.cscanada.org

An Effective L_0 - *SVM* Classifier For Face Recognition Based on Haar Features

WANG Yunpeng[a],*; XIA Xiaogang[a]

[a]School of Science, Xi'an University of Science and Technology, Xi'an, China.
*Corresponding author.

Received 24 November 2015; accepted 20 January 2016
Published online 26 February 2016

Abstract

Face recognition is an important research topic in pattern recognition, and in which, it is a striking direction that how to extract the useful features to express face. In this paper, we present a technique for face recognition by L_0 -*SVM* classifier based on Haar features. Firstly, a mass of Haar features are produced by different kinds of Haar template. Then basing on the Haar features and according to the DC algorithm, L_0-*SVM* classifier is constructed in order to enhance computational and time efficiency, as well as its validity is proved in theory. Finally, experimental results on databases show that the method can effectively improve the recognition rate of the face with a small scale of samples.

Key words: Face images; Haar features; L_0-*SVM* classifier; DC programming

Wang, Y. P., & Xia, X. G. (2016). An Effective L_0-*SVM* Classifier For Face Recognition Based on Haar Features. *Advances in Natural Science*, *9*(1), 1-4. Available from: http://www.cscanada.net/index.php/ans/article/view/8444
DOI: http://dx.doi.org/10.3968/8444

INTRODUCTION

Viola and Jones (2001) proposed a face detection method based on Haar features and Adaboost, which not only have a good effect on human face detection, but also greatly improves the detection speed. Among them, the Haar features are a representation method based on gray scale, which can reflect the transformation of gray level, and Haar features are more effective compared with the pixel method. Usually, there are three kinds of Haar feature template (Lienhart & Maydt, 2002) (as Figure 1): edge features, linear features and diagonal features. In which, characteristic values can be expressed as the pixels difference between in the black area and the white area.

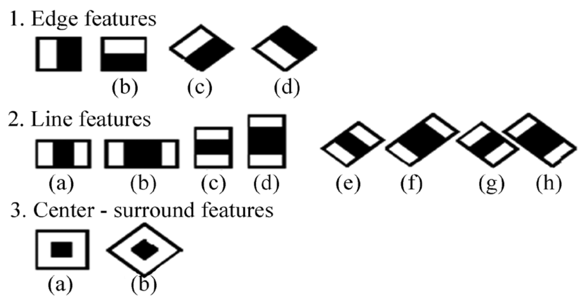

Figure 1
Three Kind of Features

Figure 2
Different Location and Dimensions

For each kind of feature, the rectangle feature can be produced at any location and any size (as Figure 2).

Where, Although belong to the same class of rectangular templates as well as the same starting position, they are different rectangular feature due to the different size. As a result, for a 32×32 pixel figure, the dimension of feature set composed of five features (as Figure 3) is as high as 135,168.

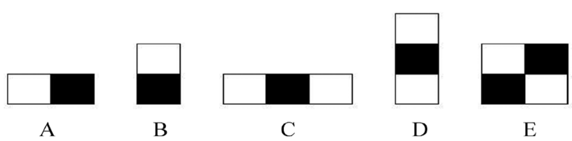

Figure 3
Five Basic Features

In this paper, a method of combines Haar features and L_0 - *SVM* are proposed for face recognition. When in solving operation, based on the theory of DC programming and DC algorithm, the appropriate approximation function of L_0 - *SVM* is constructed to improve the algorithm's operation efficiency. DC programming and DC algorithm have constituted the theoretical foundation of the nonconvex optimization, which has been widely developed through the efforts of Thi Hoai An Pham, Dinh Tao Le and other scholars. Thiao, Dinh and Le(2008) proposed a linear programming algorithm combined the DC programming to solve the optimization problem of the L_0 -relevant objective function. Le Hoai, Le Thi, Pham and Huynh (2013) to solve the block clustering problem by using DC programming and DC algorithm, and by which the computational complexity is reduced, however the traditional method had to solve a complex combinatorial optimization problem.

1. APPROXIMATION ALGORITHM OF L_0 - *SVM*

DC programming and DC algorithm (DCA) is one of the effective methods that used to solve the non-convex optimization problem. In 1985, the DC framework and DCA is introduced (Le Thi & Pham, 2005) as an optimization method, as well as extensively developed under the efforts of (Le Thi & Pham,1997) .

DC programming and DCA constitutes the backbone of the non-convex programming and global optimization. In general, a non-convex objective function $F(x)$ and its DC decomposition can be defined as:

$$\alpha = \inf \{F(x) = G(x) - H(x) \mid x \in R^n\} . \tag{3}$$

Where $F(x)$ is called DC function, and $G(x)$, $H(x)$ are the appropriate sub-continuous convex function on R^n, called the DC components of $F(x)$.Under the lack of the global optimum, he DC programming often can obtain the approximate optimal solution by using the local optimal condition.

Given the primal DC program:

$$\begin{aligned} \min \quad & g(x) - h(x) \\ \text{s.t.} \quad & x \in X \end{aligned} \tag{4}$$

And the dual:

$$\begin{aligned} \min \quad & h^*(y) - g^*(y) \\ s.t. \quad & y \in Y \end{aligned} \tag{5}$$

Where Y is the dual space of X; $g^*(y)$ and $h^*(y)$ respectively denote the conjugate functions of $g(x)$ and $h(x)$. Here we are interested in the link between global optimality conditions and local conditions.

DCA (Le Thi & Pham, 2005), which based on the local optimality conditions and DC programming, have been introduced and improved under the joint efforts of Le Thi Hoai An and Pham Dinh Tao since 1993. Although it can't guarantee the global solution for general DC programming, the optimization problem often converges to a global one when providing a suitable starting point. Through the development of the past twenty years, DCA was successfully applied in non-convex optimization field.

In this paper, we address the problem of image recognition by combining Haar feature and L_0 - *SVM*, which solving process is completed by approximation approaches. Our methods are based on DC programming and DCA, an efficient approach to promote the good approximation algorithm.

2. DC PROGRAMMING AND DCA FOR SOLVING THE PROBLEM

It is worth mentioning that a DC function f has many DC decompositions and each DC decomposition corresponds to a kind of DCA. So the key to the problem is how to develop an efficient algorithm based on the generic DCA scheme, which can be divided in two stages: the search for an appropriate DC decomposition and a good initial point. In the paper, concave approximation are given, so we first introduce an approximation algorithms of L_0 - *SVM*.

Concave approximation (Conc-SVM)

If $x \in R^n$ and $\alpha > 0$, concave function η can be denoted by

$$\eta(x) = \begin{cases} 1 - \varepsilon^{-\alpha x} & \text{if } x \ge 0 \\ 1 - \varepsilon^{\alpha x} & \text{if } x \le 0 \end{cases} . \tag{6}$$

Then zero-norm $||w||_0$can be approximated by

$$\|w\|_0 \approx \sum_{i=1}^{n} \eta(w_i) . \tag{7}$$

We get an equivalent form of objective Function (2)

$$\begin{aligned} min \quad & C\sum_{i=1}^{n}\xi_i + \sum_{i=1}^{n}\eta(w_i) \\ s.t. \quad & \begin{cases} y_i(w^T x_i + b) \ge 1 - \xi_i , \quad i = 1,...,n, \\ \xi_i \ge 0. \end{cases} \end{aligned} \tag{8}$$

The DC decomposition of $\eta(x)$ is: $\eta(x) = g(x) - h(x)$

Where $g(x) = \begin{cases} \alpha x & \text{if } x \ge 0 \\ -\alpha x & \text{if } x \le 0 \end{cases}$ and

$$h(x) = \begin{cases} \alpha x - 1 + \varepsilon^{-\alpha x} & \text{if } x \ge 0 \\ -\alpha x - 1 + \varepsilon^{-\alpha x} & \text{if } x \le 0 \end{cases} .$$

And objective Function (8) become

$$\begin{aligned} \min_{(w,b,\xi)\in\Omega} \quad & C\sum_{i=1}^{n}\xi_i + \sum_{i=1}^{n} g(w_i) - \sum_{i=1}^{n} h(w_i) \\ s.t. \quad & \begin{cases} y_i(w^T x_i + b) \ge 1 - \xi_i , \quad i = 1,...,n, \\ \xi_i \ge 0. \end{cases} \end{aligned} \tag{9}$$

Or equivalently

$$\min_{(w,b,\xi)\in\Omega} \quad C\left(\sum_{i=1}^{n} \sup\left[1 - y_i(w^T x_i + b)\right]_+\right) + \sum_{i=1}^{n} g(w_i) - \sum_{i=1}^{n} h(w_i) \tag{10}$$

Let

$$G_1(w,b,\xi) = C\sum_{i=1}^{n}\xi_i + \sum_{i=1}^{n}|\alpha w_i|$$

$$H_1(w,b,\xi)=\sum_{i=1}^{n}\left(|\alpha w_i|-1+\varepsilon^{-\alpha w_i}\right).$$

Apparently, G_1 and H_1 are convex function which satisfying the DC decomposition. The DCA corresponding to (9) as follows:

Algorithm	**Conc-SVM**

Initialization: $X^0=(w^0,b^0,\xi^0)$, $k\leftarrow 0$.

Repeat

1\. Compute $Y^k=(\overline{w}^k,\overline{b}^k,\overline{\xi}^k)\in \nabla H_1(X^k)$ is calculated by

$$\overline{b}^k=\frac{\partial H_1}{\partial b^k}=0;\quad \overline{\xi}^k=\frac{\partial H_1}{\partial \xi_i^k}=0;$$

$$\overline{w}_i^k=\frac{\partial H_1}{\partial w_i^k}=|\alpha|-\alpha\varepsilon^{-\alpha w_i}.$$

2\. Compute

$$X^{k+1}(w^{k+1},b^{k+1},\xi^{k+1})\in \arg\min\left\{C\sum_{i=1}^{n}\xi_i+\sum_{i=1}^{n}|\alpha w_i|-\left\langle Y^k,X\right\rangle : X=(w,b,\xi)\in\Omega\right\}.$$

3\. $k=k+1$

Until $\left\|X^k-X^{k-1}\right\|\big/(1+\left\|X^k\right\|)<\varepsilon$.

3. NUMERICAL EXPERIMENTS

All the algorithms we proposed have been coded in opencv, vs 2005 and implemented on a *Intel*(*R*) Core(TM) i3-2120 CPU @3.30GHz, RAM 8GB. We evaluate the performance of various approaches on the face image dataset YALE and FERET. The YALE database, collected at Yale University in 2000, contains 10 subjects and 5,850 images under various view points and illumination conditions, each image pixel is 64×64. The FERET database, sponsored by the United States Department of Defense, consists of 14,051 eight-bit grayscale images of human heads with different pose, illumination and expression, each image pixel is 112×92, is one of the most widely used face databases.

We compared our method *H*+Conc-SVM (H+ denote Haar+) on the image database YALE and FERET with other frequently-used methods, such as Haar+kNN, Haar+Adaboost and *H*+L2-SVM. The conclusion we get as shown in Table 1.

Table 1
Running Result Under Various Algorithms

Algorithm	Database	No. of select features	Proportion	Recognition accuracy	Running time(s)
H+Conc-SVM	YALE	**64**	1.05%	**99.0%**	117
	FERET	**67**	1.113%	**99.4%**	102
H+L2-SVM	YALE	102	2.23%	97.7%	194
	FERET	99	1.96%	95.7%	346
Haar+kNN	YALE	132	4.13%	93.2%	**99**
	FERET	122	3.57%	92.9%	**123**
Haar+Adaboost	YALE	133	4.43%	92.8%	120
	FERET	109	3.40%	93.7%	332

Bold values in Table 1 correspond to the best results for each evaluating indicator. Recognition accuracy is obtained by the average value of 15 test results that each test has 40% random samples as the train sample and the rest as the test sample. As can be seen from Table 1, H+Conc-SVM achieved the highest recognition accuracy with a relatively small feature set, as high as 99.4% in the FERET database. Haar+kNN (Zhang, Berg, Maire & Malik, 2006) has a best in the running efficiency. Usually speaking, the experimental results of H+Conc-SVM are satisfactory, they have a good effect on the number of selected features, image recognition accuracy and running time.

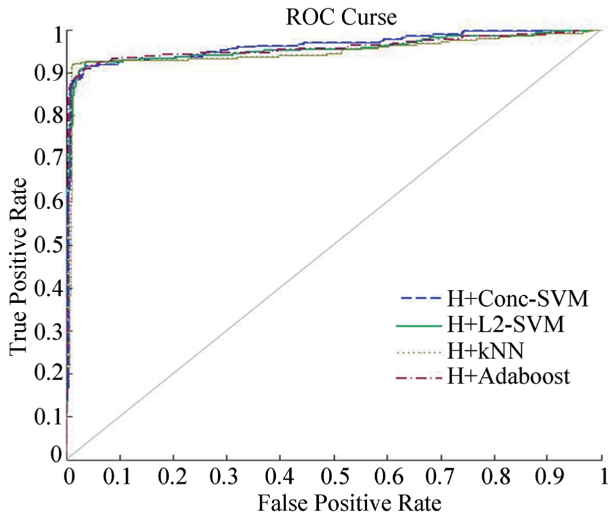

Figure 4
Recognition Accuracy of the Various Algorithms

As Figure 4, the ROC Curse reflects the accuracy of the four algorithms with different image database. The algorithm based on Haar and SVM can get higher recognition accuracy and higher than 95%. In addition, H+Conc-SVM achieved better classification results compared to the others.

CONCLUSION AND FUTURE WORK

In this work, we introduced a new L_0 - *SVM* classifier for face recognition base on Haar features. In order to improve the efficiency and save the running time of L_0 - *SVM*, we construct the approximation function of L_0 - *SVM*. Subsequently, the non-convex optimization problem is decomposed into a subtraction between two convex optimization, which is solved by DCA. We compared the performance of our approach with the Haar+kNN, Haar+Adaboost and H+L2-SVM through simulation study on face image database. Empirical results show our approach achieve better classification results and higher recognition accuracy compared to the others. We believe our approach is promising for face image recognition, especially with applications to the high-dimensional small-sample datasets.

There are several future directions for this study. Firstly, we would like to extend our method to handle more types of images, especially the abnormal face that face is obscured, wearing sunglasses, wearing a face mask and face with a lower resolution. Secondly, we will extend the work to feature selection in multi-class classification tasks.

REFERENCES

Le Hoai, M., Le Thi, H. A., Pham Dinh, T., & Huynh, V. N. (2013). Block clustering based on difference of convex functions (DC) programming and DC algorithms. *Neural Computation,* 259-278.

Le Thi, H. A., & Pham Dinh, T. (1997). Solving a class of linearly constrained indefinite quadratic problems by DC algorithms. *Journal of Global Optimization,* 2776-2807.

Le Thi, H. A., & Pham Dinh, T. (2005). The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems. *Annals of Operations Research,* 23-46.

Lienhart, R., & Maydt, J. (2002). An extended set of Haar-like features for rapid object detection. *The IEEE International Conference on Image Processing, 1,* 900-903.

Thiao, M., Pham Dinh, T., & Le Thi, H. A. (2008). DC programming approach for a class of nonconvex programs involving l_0-norm. *Modelling, Computation and Optimization in Information Systems and Management Sciences, 14,* 358-367.

Viola, P., & Jones, M. (2001). *Rapid object detection using a boosted cascade of a simple features* (pp.511-518). In proceeding of International Conference on Computer vision and Pattern Recognition.

Zhang, H., A., Berg, M., Maire, & Malik, J. (2006). SVM-KNN: Discriminative nearest neighbor classification for visual category recognition. *Computer Vision and Pattern Recognition,* 2126-2136.