Learning Symmetry Features for Face Detection Based on Sparse Group Lasso

Qi Li, Zhenan Sun, Ran He, and Tieniu Tan

Center for Research on Intelligent Perception and Computing,
National Laboratory of Pattern Recognition, Institute of Automation,
Chinese Academy of Sciences, Beijing, China
{qli,znsun,rhe,tnt}@nlpr.ia.ac.cn

Abstract. Face detection is of fundamental importance in face recognition, facial expression recognition and other face biometrics related applications. The core problem of face detection is to select a subset of features from massive local appearance descriptors such as Haar features and LBP. This paper proposes a two stage feature selection method for face detection. Firstly, feature representation of the symmetric characteristics of face pattern is formulated as a structured sparsity problem and sparse group lasso is used to select the most effective local features for face detection. Secondly, minimal redundancy maximal relevance is used to remove the redundant features in group sparsity learning. Experimental results demonstrate that the proposed feature selection method has better generalization ability than Adaboost and Lasso based feature selection methods for face detection problems.

Keywords: Face detection, sparse group lasso, minimal redundancy maximal relevance.

1 Introduction

Face detection is a key problem and a necessary step to many facial analysis algorithms, eg, face recognition, facial expression analysis, head pose estimation. How to efficiently compute and express the difference between faces and nonfaces is still a challenging task. Feature selection and appropriate classifier are needed to solve this problem. While the classification step is widely explored and quite standard, the feature selection process needs to be further researched for face detection. In this paper, we focus on feature selection method in face detection which is also a fundamental and important problem in pattern recognition and computer vision [1,6].

In the past decades, hundreds of approaches to face detection have been proposed. One of the most successful appearance-based methods is proposed by Viola and Jones [2]. The success of this method comes from a powerful feature selection method based on a well-known cascaded Boosting framework. Since then a large number of methods have been proposed following the general face detection architecture. Recently, Destrero et al. [1] proposed a sparsity enforcing

Z. Sun et al. (Eds.): CCBR 2013, LNCS 8232, pp. 162–169, 2013.

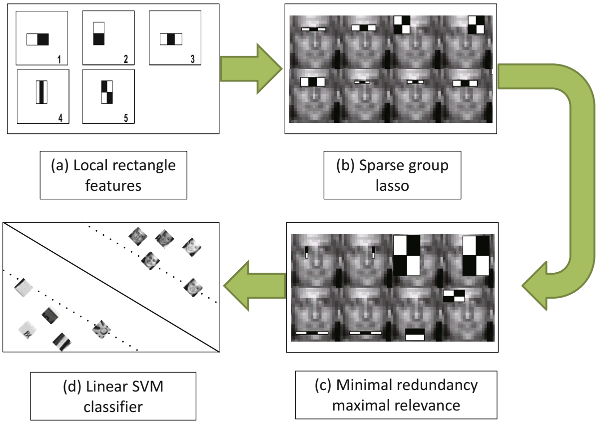


Fig. 1. Flowchart of the proposed framework. (a) Five different types of rectangle features. (b) Symmetry features selection via sparse group lasso. (c) Reducing the redundant features via minimal redundancy maximal relevance. (d) The final classification results using a linear SVM classifier.

method for learning face features. Lasso regression model was adopted to produce a sparse solution of a linear model which can be seen as a feature selection process. The sparsity based feature selection method is proved to be more effective than Viola and Jones feature selection method especially for the training set of limited size. However, there are still two unsolved problems in sparsity enforcing method: the first one is how to select discriminating features while preserving the internal symmetry structure of faces, and the second one is how to reduce the redundant features when using the sparsity enforcing method for feature selection.

To solve the first problem, we consider using sparse group lasso to select symmetry features which play important roles in object detection, recognition and matching [3,8]. Compared with lasso for feature selection, sparse group lasso not only has the property of performing feature selection but preserving the internal symmetry characteristics of faces. To solve the second problem, we choose minimal-redundancy-maximal-relevance (mRMR) to reduce the redundant features selected by the sparse group lasso algorithm. Figure 1 shows the proposed two-stage framework.

The contributions of this paper are summarized as two points. First, we combine knowledge-based methods and appearance-based methods for face detection. As to knowledge-based method, we make use of symmetry characteristics of faces. As to appearance-based method, we use sparsity enforcing method and mRMR to choose the features which are meaningful and representative. Second, the proposed framework has a better generalization ability than other feature selection methods. Besides, the number of training examples we used is usually less than Viola and Jones face detection method. So it can be easily applied to other types of less common objects.

2 The Proposed Framework for Face Detection

In this section, we discuss symmetry features and propose a two stage feature selection framework for face detection. In the first stage, we use sparse group lasso to select groups of symmetry rectangle features. In the second stage, the mRMR method is used to further reduce the selected symmetry features.

2.1 Symmetry Features

We use the rectangle features proposed by [2], which have a strong discriminating power and can be efficiently computed by integral images. Figure 1(a) shows five different kinds of used rectangle features, which are computed over different locations, sizes and aspect ratios. For each image patch will generate tens of thousands of features. So it is necessary to select a small set of compact and meaningful features.

We propose a new concept called symmetry features which reflect the mirror characteristics of faces. In a cropped face image, we use the middle column of that image as the symmetry axis. Then each feature at the right side of the symmetry axis corresponds to the same size, same aspect ratio feature at the left side of the symmetry axis. We put these two features as one group. If the center of a feature just locates at the symmetry axis, then it doesn't have a mirror feature. We put this single feature as one group. So each group of features at most has two features. Those groups of features are called symmetry features.

2.2 Feature Selection Using Sparse Group Lasso

As mentioned in Section 1, if we use the l_1 based methods for feature selection directly, we will lose symmetry characteristics of faces. Sparse group lasso proposed by [4] has a nice property of selecting features at the group and individual predictor levels. The standard form of sparse group lasso is as follows:

$$\beta = \arg\min_{\beta \in R^P} \left(\begin{array}{c} \left\| Y - \sum_{l=1}^{L} X_l \beta_l \right\|_2^2 \\ + \lambda_1 \sum_{l=1}^{L} \sqrt{\rho_l} \|\beta_l\|_2 + \lambda_2 \|\beta\|_1 \end{array} \right) \quad (1)$$

where $X \in R^{n \times p}$ represents the training set of n elements with a dictionary of p features, $Y \in R^{n \times 1}$ is the output labels, $\beta \in R^{p \times 1}$ is the parameter vector. X is divided into L non-overlapping groups of features $(X_1, ..., X_L)$. The element of $X_l (l = 1, ..., L)$ is composed by symmetry features. $\beta = (\beta_1, \beta_2, ..., \beta_l)$ is the group parameter vector. $\sqrt{\rho_l}$ terms accounts for the varying group sizes. λ_1 controls the sparsity of features within a group and λ_2 controls the sparsity of the selected groups features. Depending on λ_1 and λ_2, the sparse group lasso yields sparsity at both the individual and group feature levels.

2.3 Reducing Redundant Features

More variables than needed are usually selected by the sparse group lasso when using cross validation to yield parameters λ_1 and λ_2. Compared with [2], a large number of Haar-like rectangle features will be selected by sparse group lasso. Although those features have high correlation with the detection results, lots of the same sizes and aspect ratios features almost in the same position are chosen twice or even more. In order to overcome the drawback of sparse group lasso, we choose mRMR [5] in the second stage. The features selected by mRMR can capture the face information in a broader scope by reducing the mutual redundant feature set.

Given two variables x and y, their correlation coefficient is defined as:

$$\rho(x;y) = \frac{cov(x,y)}{\sigma_x \sigma_y} = \frac{E[(x-\mu_x)(y-\mu_y)]}{\sigma_x \sigma_y} \tag{2}$$

Minimal redundancy is defined as:

$$\min R(S), R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} \rho(x_i, x_j) \tag{3}$$

Maximal relevance is defined as:

$$\max D(S,c), D = \frac{1}{|S|} \sum_{x_i \in S} \rho(x_i; c) \tag{4}$$

where S is the set of features, c is the corresponding target class of feature set S, $\rho(x_i, x_j)$ is the correlation between feature i and j, $\rho(x_i, c)$ is the correlation between feature i and c. Eq. (3) aims to select the subset of features such that the correlation between themselves are minimal. Eq. (4) aims to ensure that the selected subset features have the discriminating power when they represent different classes.

The criterion combining Eq. (3) and Eq. (4) is called minimal-redundancy-maximal-relevance (mRMR). Optimization both of them requires combining them into a single criterion function as follows:

$$max(D(S,c)/R(S)) \tag{5}$$

Incremental search methods proposed by [5] can be used to solve the optimization problem (5). Although we enforce group penalties in Eq. (1), those features selected by the first stage is not necessarily the symmetry features. First we analyze the features selected by sparse group lasso. If the groups of features are composed by two rectangle features, one of them who represents the entire group is sent to the second stage. The final output of rectangle features contains the whole groups of rectangle features that mRMR selected.

3 Experiments

The Jensen database [7] is used to evaluate different feature selection methods for face detection. All of the positive examples of Jensen database are taken from FERET database [9] and LFW database [10]. 5,000 training images and 5,000 testing images are selected sequentially where the positive and negative examples are evenly distributed between training and testing images. We resize all of the images to a base resolution of 19 by 19. Five different kinds of local rectangle features are used to generate about 64,000 rectangle features for each of the training and testing images.

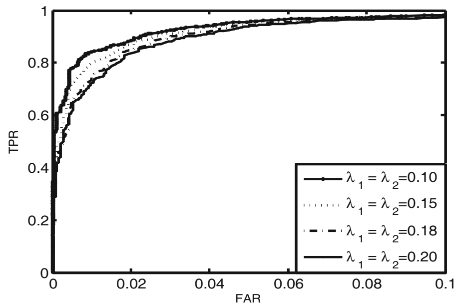
3.1 Settings

Before using sparse group lasso to select symmetry features, we use l_2 norm to normalize the dataset so that each column of the training and testing dataset has unit l_2 norm. A subset of features are selected according to our proposed two stage framework. These features are used to represent the whole dataset. Considering the computational advantage, a linear SVM classifier is used. The linear SVM model is obtained based on the training set and 5-fold cross validation is used to tune the parameters. Finally we analyze the generalization ability of our method over the testing set. ROC curves are used to evaluate the performance of different feature selection methods for face detection.

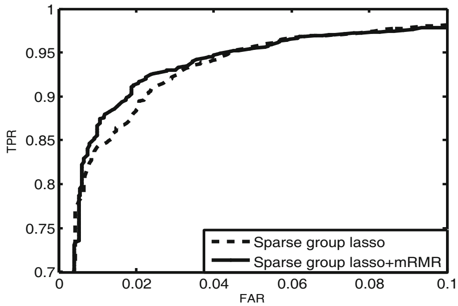
The First Stage. How to determine the values of λ_1 and λ_2 is an important problem when sparse group lasso is used to select the symmetry features. In practice, the parameters λ_1 and λ_2 are set to an equal and small value because of two reasons. (1) If λ_1 and λ_2 are set to be large values, the large penalty terms will make the number of selected symmetry features relatively small. These symmetry features cannot be used for classification because of the low performance. (2) A large number of symmetry features are available to preserve the internal symmetry characteristics of faces if identical value is given to λ_1 and λ_2. These large number of symmetry features not only have meaningful representation but also can achieve a high level classification performance.

Figure 2(a) shows ROC curves with different values of λ_1, λ_2. We choose $\lambda_1 = 0.10, \lambda_2 = 0.10$ as the final parameters in the first stage of our method. This procedure leaves us with a set of 1087 groups of features (totally 2054 features). A large number of symmetry features appear around eyes, noses and mouths. Although the size of these representative symmetry features is much less than that of the original set, it is still higher than we actually need.

The Second Stage. After the selection procedure of the first stage, mRMR is used to further choose a small subset of features. At each step, mRMR choose one group of features which maximize the optimization problem (5). Finally we choose 39 groups of features (totally 67 features) which achieve the best detection accuracy after the second stage. Figure 2(b) compares the performance of our method with and without the second stage. We can see that the procedure



(a)



(b)

Fig. 2. Different ROC curves with the horizontal line representing the false accept rate and the vertical line representing the true positive rate. (a) ROC curves with different λ_1, λ_2 using sparse group lasso method in the first stage, (b) ROC curves of our method with and without the second stage.

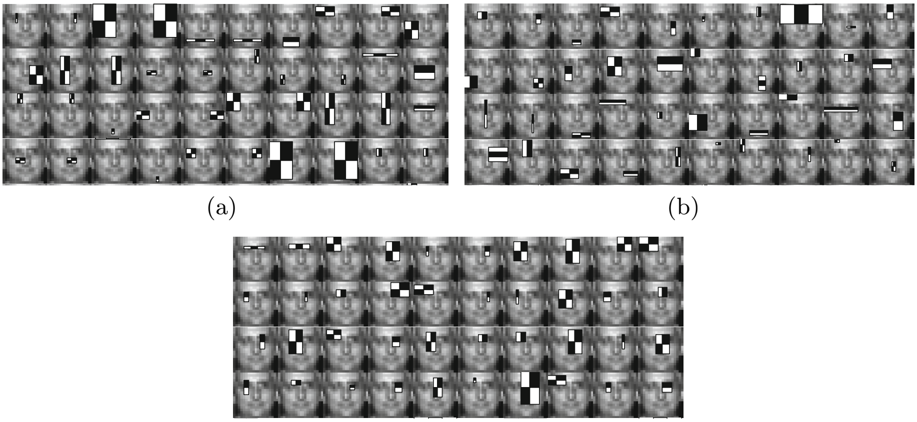
of the second stage not only reduces the number of features significantly, but actually has a small gain (about 5%) at low false accept rate in terms of the detection result. This observation indicates that mRMR can select a small subset of compact and meaningful features. Figure 3(a) shows the top 40 features selected by our method. We can see that the order of the the features is changed after the second stage, and the symmetry features selected by our method are salient features appearing at different locations and sizes of faces.

3.2 Comparison with Other Methods

We also compare our method with other feature selection methods, such as Adaboost and conventional Lasso method without considering the symmetry characteristics of faces. Different features are selected by the three methods respectively and then these selected features are used to represent the whole training dataset to train different linear SVM models. Then we evaluate the effectiveness of the selected features over the testing set by the ROC curves.

The same number of features (totally 67 features) are selected using Adaboost algorithm [2] and the top 40 features are shown in Figure 3(b). Conventional Lasso method without considering the symmetry characteristics of faces is also compared with our method. Different number of features can be selected by tuning the parameter λ. Notice that when tuning the parameter λ, the number of selected features is not consistent. In order to have a fair comparison, a little higher number of features (103 features) are selected during the process. Figure 3(c) shows the top 40 features selected by conventional Lasso method. From Figure 3(c) we can see that most of the selected features seem to appear around the salient parts of faces.

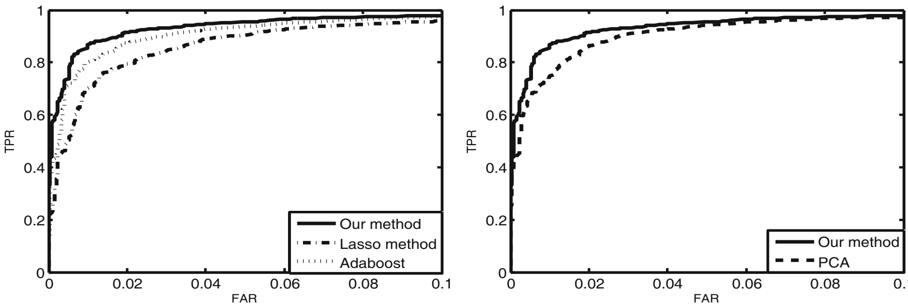
The comparison of the proposed method with Adaboost and conventional Lasso method is shown in Figure 4(a). From Figure 4(a), we observe that the true positive rate (TPR) of all methods increases quickly when false accept rate (FAR) is smaller than 0.01, and TPR of all methods tends to be similar



(a) (b)

(c)

Fig. 3. Top 40 rectangle features selected by different methods: (a) our method, (b) Adabooost method, (c) conventional Lasso method



(a) (b)

Fig. 4. Comparison of different feature selection methods for face detection: (a) ROC curves of our method, Adaboost method and conventional Lasso method with the same linear SVM classifier, (b) ROC curves of our method and PCA with the same linear SVM classifier

when FAR is larger than 0.06. We also observe that our method outperforms the other two methods. The improvement of our method against Adaboost and conventional Lasso method is nearly 5% and 10% at a relatively low FAR.

Similar to [1], we also compare our method with the classic dimensionality reduction method PCA. The projection matrix which contains the eigenvectors of the training dataset is used as the projection matrix to project the training and testing dataset to a 67 dimensional matrix. Then the training dataset is used to train a linear SVM classifier and the final classification result over testing dataset is showed in Figure 4(b). We can see that TPR of our method is about 5% higher than PCA at a relatively low FAR.

4 Conclusions

In this paper we have presented a novel two stage framework to learn symmetry features for face detection. Sparse group lasso and mRMR are used to reduce the redundant features while at the same time preserving the symmetry characteristics of faces. Experimental results have shown that our method outperforms other traditional feature selection methods under the same conditions. As parts of our future work, we will further research on implementing a robust face detection system.

Acknowledgments. This work is funded by the National Natural Science Foundation of China (Grant No. 61273272,61103155), the International S&T Cooperation Program of China (Grant No. 2010DFB14110) and the Instrument Developing Project of the Chinese Academy of Sciences (Grant No. YZ201266).

References

1. Destrero, A., De Mol, C., Odone, F., Verri, A.: A regularized framework for feature selection in face detection and authentication. International Journal of Computer Vision 83, 164–177 (2009)
2. Viola, P., Jones, M.J.: Robust real-time face detection. International Journal of Computer Vision 57, 137–154 (2004)
3. Levi, K., Weiss, Y.: Learning object detection from a small number of examples: the importance of good features. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 53–60 (2004)
4. Friedman, J., Hastie, T., Tibshirani, R.: A note on the group lasso and a sparse group lasso, arXiv:1001.0736, 164–177 (2010)
5. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence 27, 1226–1238 (2005)
6. He, R., Tan, T., Wang, L., Zheng, W.S.: l2, 1 Regularized correntropy for robust feature selection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2504–2511 (2012)
7. Jensen, O.H.: Implementing the Viola-Jones face detection algorithm. PHD thesis, Technical University of Denmark, Denmark (2008)
8. Hauagge, D.C., Snavely, N.: Image matching using local symmetry features. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 206–213 (2012)
9. Phillips, P.J.: Moon: The Facial Recognition Technology (FERET) Database, `http://www.itl.nist.gov/isd/humanid/feret/feret_master.html`
10. Huang, G.B., Mattar, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In: Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition (2008)