

Learning Sample Subspace with Application to Face Detection

Jianzhong Fang and Guoping Qiu
School of Computer Science, The University of Nottingham
{jzf|qiu}@cs.nott.ac.uk

Abstract

In this paper, we present a novel maximum correlation sample subspace method and apply it to human face detection [1] in still images. The algorithm starts by projecting all the training samples onto each sample and selects the sample with the largest accumulated projection as the first subspace base vector. After a base vector is selected, all other samples are made orthogonal to the current base vector and which is in turn used to form the training samples for learning the next base vector. Each subspace base is created by a one-pass process and therefore the method is computationally very efficient. These bases form a transform and we use it to derive discriminative features for face detection by training a support vector machine classifier. We perform testing on both CMU and MIT face detection image data sets. Extensive experiments demonstrate that our results are comparable to those published in state of the art literature.

1. Introduction

In this paper, we present a novel adaptive learning approach to selecting class specific sample subspace and use human face detection to illustrate the usefulness and effectiveness of the new method. In the following section, we present the development of a learning procedure to construct a set of orthogonal base from training samples. In section 3, we apply the new set of orthogonal base to construct feature vectors for (face/nonface) pattern representation. In section 4, we use the new feature vectors for human face detection and present experiment results. We give concluding remarks in section 5.

2. Maximum Correlation Sample Subspace

There are several techniques available for dimension reduction in face detection. Principal component analysis or PCA [6], is probably one of the most well-known. PCA derives a transform from the eigenvectors corresponding to the largest eigenvalues of the covariance matrix of the training samples. The objective of PCA is to seek a representation, which is optimal in terms of minimal mean square error between the representation and the original data. Although PCA is optimal for dimension reduction

(in mean square error sense), it is not a transform for developing discriminative features optimal for classification. In face detection and other pattern recognition tasks, feature representation vectors not only have to be compact but also discriminative.

There are several reasons that make PCA not the best transform for developing feature vectors for pattern recognition. It is well known that PCA is optimal when the data is under the unimodal Gaussian assumption, and this assumption is almost always not true in real world applications. In many applications, such as face detection, training sample size is often small. The eigenvectors point to directions that have the largest variance on “average”. This “average” may smear the representation due to both limited sample size and non-Gaussian nature of the data. This is reminiscent to linear filtering in digital signal/image processing which has the tendency of removing Gaussian noise but destroying (non-Gaussian) signal features as well. One of the alternatives to linear filtering is nonlinear filtering. One type of nonlinear filtering, often known as rank order filtering, ranks the data according to a certain criterion and take one of the data samples of a certain rank as output (e.g., median filtering takes the sample in the middle as output)

We have developed a method for learning a transform in which, the transform bases are selected from the training samples based on a maximal correlation criterion. It is related to PCA in the sense that the base vectors are orthonormal, however, they are not the eigenvectors of the covariance matrix. The transform bases are directly selected from the training samples. The relation between this new transform and PCA is analogy to that between linear and rank order filtering in digital signal processing.

Let $\mathcal{S} = \{X_1, X_2, \dots, X_N\}$ be the set of training samples, where X_i is an M dimensional vector. We use the following algorithm to construct K orthonormal base vectors

Maximum Correlation Sample Subspace Algorithm (MCSSA)

Step 1: Set $k = 1$, $E(0) = 0$, T = a preset value, and compute

$$E_0 = \frac{1}{N} \sum_{i=1}^N \langle X_i, X_i \rangle^2$$

Step 2: Calculate

$$A_j = \sum_{i=1}^N \langle X_i, X_j \rangle^2, \forall j \quad (1)$$

Step 3: Set

$$e(k) = \frac{X_n}{\|X_n\|} \text{ if } A_n > A_j, \forall j \text{ and } j \neq n$$

Step 4: Calculate

$$E(k) = E(k-1) + \frac{1}{N} \sum_{i=1}^N \langle X_i, e(k) \rangle^2 \quad (2)$$

Step 5: If $\frac{E(k)}{E_0} \geq T$, then stop, otherwise, for $\forall i$

do:

$$X_i = X_i - \langle X_i, e(k) \rangle e(k) \quad (3)$$

Step 6: Set $k = k + 1$, if $k > K$, stop, else go to **Step 2**

The algorithm starts by projecting all the training samples onto each sample and select the sample with the largest accumulated projection as the first base vector. After a base vector is selected, all other samples are made orthogonal to this base by (3), and which is in turn used to form the training samples for the construction of next base vector. The algorithm produces a set of K base vectors. The number of base vectors (K) is determined by the preset value of T . The larger T is, more bases will be created. It is easy to show that the base vectors in the set are orthonormal vectors.

Compared with PCA, the bases selected by the new algorithm are more robust against outliers, just as rank order filtering is more robust against outliers than linear filtering. The algorithm is adaptive and suitable for online training. In the next section, we will use the algorithm to develop transform base vectors for human face representation.

3. Human Face Representation in Maximum Correlation Sample Subspace

The maximum correlation sample subspace developed in the last section can have various applications. In this section we apply it to represent face subspace. In this case, all the samples in the training set $\mathcal{S} = \{X_1, X_2, \dots, X_N\}$ are M dimensional face image patterns. Examples of the maximum correlation sample subspace base vectors are shown in Fig. 1.

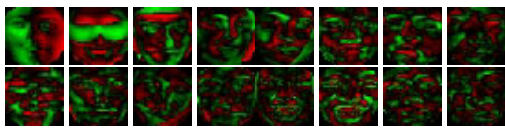


Fig. 1. Example of a set of 16 face subspace bases obtained from a set of over 3000 face image patterns.

After the creation of the orthonormal base set Φ_K , the original pattern can be projected onto the subspace as

$$Y_i = \Phi_K^T X_i \quad (4)$$

where Y_i is a K ($K \ll M$) dimensional projection vector of the face pattern X_i . The dimensionality of the projection vectors is much lower than that of the original data and can be used for recognition applications. In the next section, we use such a representation scheme to train a support vector machine for human face detection.

4. Experimental Results

To illustrate the effectiveness of the maximum correlation sample subspace in representing patterns for classification, we apply it to face detection. We first collected face image sample from various sources. In total, there are 3061 face samples forming the training set. In order to train the classifier, non-face samples are collected from Corel data set. The non-face categories include scenery, texture, surface, fur, desert, underwater shimmering, art and skin patches. There are a total of 3239 initial non-face samples. All the face samples and non-face samples are normalized to 32×32 pixels in size.

In order to achieve good contrast and normalize the amount of energy for images taken under different conditions, local histogram equalization is performed on the 32×32 pixel normalized image. We also apply a spade-shaped mask to eliminate the background noise, and finally a 934 dimensional vector is retrieved for the face or non-face image representation.

A set of orthonormal bases is developed using the maximum correlation sample subspace algorithm (MCSSA) from the face image set only. In our particular data set, 168 bases are generated for $T = 91\%$. A support vector machine classifier [4] is trained to play the role of decision-maker. We employ exhaustive search method by moving rectangular windows of different sizes across the testing images. We use 20 different window sizes in search of different sized faces, the smallest face size is defined as 20×24 pixels and the biggest face is defined as 149×182 pixels, each scale is about 10% larger than the previous one.

To evaluate the performance of our method, we perform testing on 3 publicly available data sets that are widely used in the face detection community. Set A and C is provided by CMU Set B is provided by MIT, all the data were obtained from the CMU website [5]. Table 1 lists the correct detection and false alarm rates of each set and combined sets at a particular threshold.

Data Set	Faces (Images)	Faces Correctly Detected	Correct Detection Rate	False Positive Rate
Set A (CMU)	169(42)	149	88.1%	1.20E-5
Set B (MIT)	155(23)	133	85.8%	1.23E-5
Set C (CMU)	183(65)	166	90.7%	1.29E-5
Set A, B and C	507(130)	448	88.3%	1.26E-5

Table 1. Testing results performed on open data sets [5] over 130 images with 507 faces for a particular threshold.

In total, our detector evaluated 41,679,874 patterns on Set A, Set B and Set C. The false positive rate is very small compared with evaluated patterns. The full characteristics of a detector is represented by the receiver operating curve (ROC). The ROC curve of our detector is shown in Fig. 2.

Fig. 3 shows pictorial examples of our testing results. These results are at least comparable to state of the art. It is worthy mentioning that our experiments used all 130 images which contained 507 faces whilst most others only included a subset of the testing data, e.g., [2], [3], reported results using 125 testing images contained fewer faces.

5. Concluding Remarks

In this paper, we have presented a new method, the maximum correlation sample subspace algorithm (MCSSA), for developing orthogonal linear bases and we have applied it for developing transforms for human face detection. The method is similar to PCA in spirit, however, it differs from PCA in a number of significant ways. Whilst the bases of PCA captures the averaged directions with the largest variance, the bases of the new MCSAA are derived directly from one of the samples. Therefore, when the samples are noisy (nonGaussian) or the training sample sizes are limited, MCSAA may be more advantageous for representing patterns for classification purposes. Whilst the bases of PCA are computed through manipulation of (very large) covariance matrix or through iterative learning, the bases of MCSAA

are obtained by a deterministic one-pass processes. The relationship between PCA and MCSAA can be understood as similar to that between the linear and rank order filtering in digital signal/image processing. To illustrate the effectiveness and usefulness of such a subspace, we have successfully applied it to face detection, which achieves results that are competitive to state of the art.

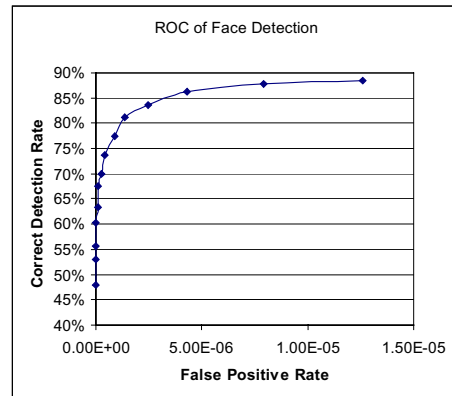


Fig. 2. ROC curve of MCSSA subspace face detector

References

1. M-H Yang, D. Kriegman and N. Ahuja. "Detecting Face in Images: A Survey". IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 24, no. 1, pp. 34-58, January, 2002.
2. Stan Z. Li, et al "Statistical Learning of Multi-View Face Detection". ECCV 2002, LNCS 2353, pp. 67-81, 2002
3. A. H. Schneiderman and T. Kanade. "Probabilistic Modeling of Local Appearance and Spatial Relationships for Object Recognition". CVPR '98. pp. 45-51.
4. E. Osuna, R. Freund, and F. Girosi. Training Support Vector Machines: An Application to Face Detection. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Pages 130-136, 1997.
5. URL <http://www.cs.cmu.edu/~har/faces.html>.
6. S. Haykin, Neural Networks: **A Comprehensive Foundation** (2nd Edition) , Englewood Cliffs, NJ: Prentice-Hall

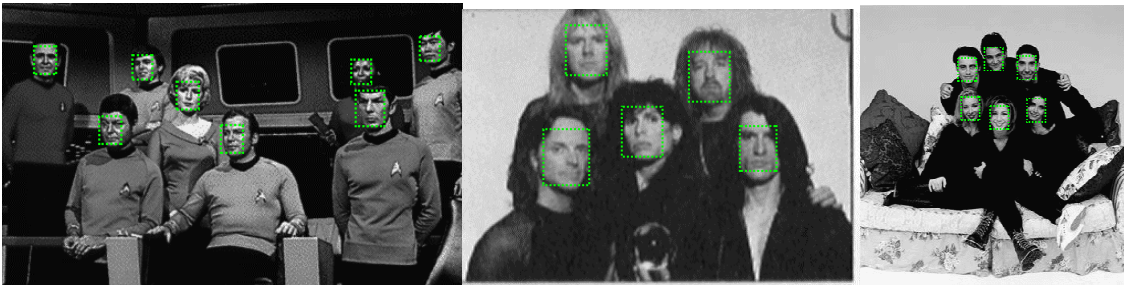




Fig. 3. Examples of detection results. Please note that if the image is printed in black and white, the detection windows may not be clearly visible, in that case either print in colour or view the electronic version