

Cone-Restricted Subspace Methods

Takumi Kobayashi and Nobuyuki Otsu
National Institute of Advanced Industrial Science and Technology
1-1-1 Umezono, Tsukuba, Japan

Abstract

In pattern recognition, feature vectors are occasionally subject to non-negative constraints. This characteristic can be expressed by a cone in feature vector space. In this paper, we propose cone-restricted subspace methods. The proposed methods admit the scaling and additivity of vectors as well as ordinary subspace methods; in addition, vectors can be strictly classified at the boundary of the cone. Some experimental results for face and person detection demonstrate the effectiveness of the proposed methods.

1. Introduction

In pattern recognition, it is important to appropriately classify the feature vectors extracted from object patterns. The classifier is required to allow for the fundamental structure of sample feature vectors while remaining robust to irrelevant variations. These irrelevant variations may be different for different recognition tasks, but a classifier can be constructed by assuming appropriate sample distributions.

Subspace methods assuming linear variation of samples have been developed over many years [9, 11, 3]. The sample feature vectors, e.g., image vectors in an illumination cone [4] and the features with the additivity property [6, 5], vary linearly. In face recognition [3] and abnormality detection [7] using CHLAC features [5], subspace methods have worked well. Subspace methods, however, are not necessarily preferable in terms of approximating the distribution of samples since the distribution cannot be precisely approximated in a linear subspace. As shown in Fig. 1(a), sample vectors seldom cover the whole subspace and are usually contained in restricted regions, which increases false positive errors in the subspace method. Moreover, the dimensionality of the subspace is critical. Although the variance of the distribution is different across each dimension of the subspace, in subspace methods, the differences are not

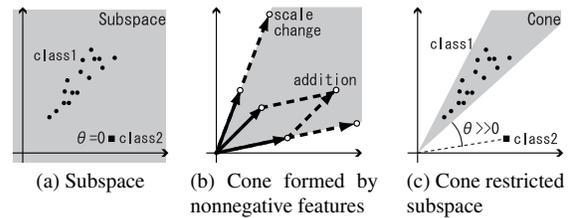


Figure 1. Subspace and cone

taken into account in the actual dimensions employed. Thus, the number of dimensions greatly affects the classification performance.

Usually, extracted features, such as the physical quantities of image vectors and histograms of SIFT [6] and CHLAC [5], have non-negative values. As shown in Fig. 1(b), non-negative feature vectors are subject to the following variations: scaling, e.g., by changes of pixel values arising from illumination changes, and additivity, e.g., arising from multiple objects in the image. These scaling and additivity are continuously brought about, and as a consequence, the sample feature vectors form a cone structure. It should be noted that, in this paper, samples are assumed to be unimodal.

We propose novel subspace methods which utilize a cone: *cone-restricted subspace methods* to deal with the problems of subspace methods discussed above. The proposed methods approximate the sample distribution by a cone pointed from the origin as in Fig. 1(c) by focusing on the property of non-negative feature vectors. They can strictly classify the vectors at the boundary of the cone, while admitting scaling and additivity of vectors as well as ordinary subspace methods.

2. Cone Restricted Subspace Method

Samples in feature vector space span a convex cone defined as follows:

$$C : \left\{ \mathbf{x} \mid \mathbf{x} = \sum_{i=1}^N \alpha_i \boldsymbol{\xi}_i = \boldsymbol{\Xi} \boldsymbol{\alpha}, \alpha_i \geq 0 \right\}, \quad (1)$$

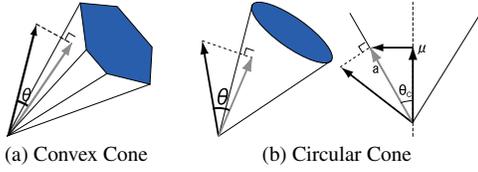


Figure 2. Cones

where N is the number of sample vectors $\xi_i \in \mathbf{R}^d$, and the α_i are non-negative coefficients. The cone C is obtained from the subspace ($\{x | x = \sum_{i=1}^N \alpha_i \xi_i\}$) by restriction to *non-negative* coefficients. The cone is, of course, as robust as the subspace, with respect to scaling and (non-negative) additivity of sample vectors. The sample distribution can be approximated by the cone more precisely as shown in Fig. 1(c), and the discrimination power should be improved.

Classification is based on the angle θ between the input vector y and the orthogonally projected vector onto the convex cone C (Fig. 2(a)), which is defined as

$$\begin{aligned} \theta &= \arcsin(\min_{x \in C} \|y - x\| / \|y\|) \\ &= \arcsin(\sqrt{\min_{\alpha \geq 0} \|y - \Xi \alpha\|^2} / \|y\|) \end{aligned} \quad (2)$$

where $0 \leq \theta \leq \pi/2$.¹ This is calculated by using the non-negative least square method (NNLS) [2]. However, the computational cost increases rapidly as the number of samples, N , increases. In order to reduce the computational cost, we propose three kinds of cone-restricted subspace method: The first method is based on the strict convex cone of samples by finding its essential basis vectors, the second is to approximate the convex cone by a covering convex cone, and the third is to approximate it by a circular cone.

2.1. Strict Convex Cone

From the sample vectors, we select a small number of essential basis vectors supporting the convex cone spanned by samples. Given sample vectors $\mathbf{X} = [x_1, \dots, x_N]$, *convex-redundant* samples are defined as $x_t = \mathbf{X}_{\tilde{t}} \alpha_{\tilde{t}}$ ($\alpha_{\tilde{t}} \geq 0$), where \tilde{t} indicates the subset excluding the t -th sample. It is easily proven that the convex cone (Eq.(1)) is invariant even if the convex-redundant samples are eliminated. Thus, by eliminating such samples, we can obtain the *convex-basis* samples. In order to find and eliminate the convex-redundant samples, the leave-one-out scheme is applied: the sample vector x_t is eliminated if the angle to the convex cone spanned by $\mathbf{X}_{\tilde{t}}$, which is calculated by Eq.(2) with

¹In the case of non-negative feature vectors and many other cases, any angle satisfies the constraint.

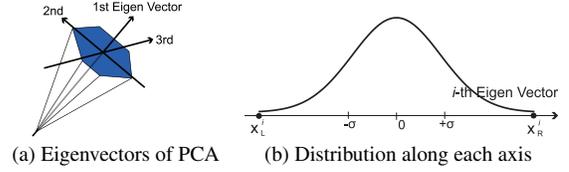


Figure 3. Covering convex cone

$y \leftarrow x_t, \Xi \leftarrow \mathbf{X}_{\tilde{t}}$, is below a predefined threshold (close to 0). This results in a reduced number of vectors and a drastic reduction in the computational cost of Eq.(2).

2.2. Covering Convex Cone

The convex cone of sample vectors can be approximated by a covering convex cone spanned by a small number of basis vectors. Consider the convex hull of the projections of samples onto the unit (hyper)sphere. The convex hull can be approximated by a covering convex hull constructed using PCA as follows. First, sample vectors are projected onto the sphere by normalization ($z_t = x_t / \|x_t\| \in \mathbf{R}^d$) and then PCA is applied to the auto-correlation matrix of z_t . The eigenvector of the first (biggest) eigenvalue is the direction vector of the cone and the other eigenvectors are the principal axes of the distribution on the sphere (Fig. 3(a)).² By determining two points, denoted by x_L, x_R , which are the edges of the distribution on each principal axis (Fig. 3(b)), the convex hull composed of these edge points approximately covers the convex hull of samples on the sphere. Thus, for the i (≥ 2)-th eigenvector, two basis vectors are defined as

$$\begin{aligned} \xi_{2i-3} &= e_1 + x_L^i e_i = e_1 - k \sqrt{\lambda_i} e_i \\ \xi_{2i-2} &= e_1 + x_R^i e_i = e_1 + k \sqrt{\lambda_i} e_i, \end{aligned} \quad (3)$$

where e_i are eigenvectors, λ_i are eigenvalues corresponding to the variances of the distribution, and k is the scaling parameter. When applying the above PCA, we determine the reduced dimensionality r ($< d$) according to the contributing rate $\eta_i = \sum_{j=2}^i \lambda_j / \sum_{j=2}^d \lambda_j$, and so obtain $2(r-1)$ basis vectors.

2.3. Circular Cone

The convex cone of samples can be approximated by a circular cone, defined as follows. Given projection onto an unit sphere, as shown in Fig. 4(a), a circular cone is defined by the circle which is the intersection of the plane and the sphere:

$$\{x \mid \mu^T x / \|x\| \geq b\}, \quad (4)$$

²This holds in the case that feature vectors lie in a restricted region, in particular, a non-negative region.

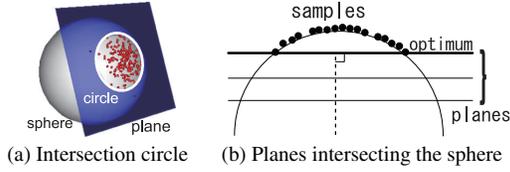


Figure 4. Circular cone

where μ is the direction vector of the convex cone of samples and $\mu^T x = b$ is the plane intersecting the sphere. The circular cone has a simple formulation that enables us to easily calculate angles, as we describe later. The circular cone is an approximate representation of the convex cone of samples since, on the sphere, the convex hull composed of an increased number of the samples approaches a circle. The plane ($\mu^T x = b$) is obtained from sample vectors as follows.

[The Plane for Circular Cone] In general, sample vectors are elliptically distributed, although we approximate the distribution by a circular cone, i.e., a circle on the sphere. For accurate approximation, whitening is required to convert the distribution into an isotropic one. The whitening is based on PCA applied to the samples on the sphere as in Sec.2.2. As shown in Fig. 3(a), the second or subsequent eigenvectors of PCA are the principal axes of the elliptical distribution on the sphere. Since we can arbitrarily determine the radius of the circle, samples are scaled along the third or subsequent eigenvectors, based on the standard deviation along the second eigenvector:

$$\hat{z}_t = \mathbf{S} \mathbf{E}^T z_t. \quad (5)$$

where $\mathbf{S} = \text{diag}(1, 1, \sqrt{\lambda_2/\lambda_3}, \dots, \sqrt{\lambda_2/\lambda_r})$, $\mathbf{E} = [e_1, \dots, e_r]$.

The optimum circular cone appropriately approximating the sample distribution is equivalent to the minimal circle containing samples on the sphere (Fig. 4(b)). The radius of the circle decreases as the distance ($b/\|\mu\|$) between the plane and the origin increases (Fig. 4(b)). Thus, the plane is obtained by

$$\min_{\mu, b} \frac{1}{2} \|\mu\|^2 - b \quad \text{s.t.} \quad \mu^T \frac{\hat{z}_i}{\|\hat{z}_i\|} \geq b. \quad (6)$$

The solutions are normalized; $\mu \leftarrow \mu/\|\mu\|$, $b \leftarrow b/\|\mu\|$. This is the same formulation as 1-class SVM [10] and can be solved by convex quadratic programming of SVM.

[Angle to Circular Cone] The angle to the circular cone is calculated more easily than that to the convex cone in Eq.(2). Firstly, an input vector y is projected into the subspace and scaled ($\hat{y} = \mathbf{S} \mathbf{E}^T y$). Secondly, we calculate the unit vector a , which is the projection of \hat{y} onto the circular cone. Suppose that the slice surface



Figure 5. Images in database

including \hat{y} , a and the direction vector μ is as shown in Fig. 2(b). Then a and the angle θ are calculated by

$$a = \cos \theta_C \mu + \sin \theta_C \frac{\hat{y} - (\hat{y}^T \mu) \mu}{\|\hat{y} - (\hat{y}^T \mu) \mu\|} \quad (7)$$

$$\theta = \arccos \frac{y^T \mathbf{E} \mathbf{S}^{-1} a}{\|y\| \|\mathbf{E} \mathbf{S}^{-1} a\|} = \arccos \frac{\hat{y}^T \mathbf{S}^{-2} a}{\|y\| \|\mathbf{S}^{-1} a\|},$$

where $\theta_C (= \arccos(b))$ is the spread angle of the circular cone. Due to the reduced dimensionality r of \hat{y} and a , the computational time of Eq.(7) is less than that of Eq.(5). Thus, the angle θ can be calculated as fast as subspace methods ($O(dr)$). Note that, for the circular cone, any angles ($0 \leq \theta \leq \pi$) can be calculated.

2.4. Summary

The proposed methods are summarized below.

Method	Approx. Accuracy	Comp. Cost	Angle Range
Strict Convex Cone	↑ high	↑	$0 \leq \theta \leq \frac{\pi}{2}$
Covering Convex Cone			
Circular Cone			↓ small

These methods have trade-offs in terms of approximation accuracy and computational cost.

3. Experimental Results

We applied the proposed methods to face and person detection using the MIT-CBCL Face Dataset [1] and the Pedestrian Dataset [8], respectively. In this experiment, only positive samples of target objects (face/person images) were used for constructing the cone, i.e., one-class learning. For face detection, we used 2429 face images (19×19) for training, and 472 face and 28121 face-free images for testing (Fig. 5(a)). We extracted SIFT-like features: histograms of 8-oriented gradients from 2×2 subregions (32 dimensions). For person detection, 400 person images (64×128) for training, and 524 person and 3000 person-free images for testing were used (Fig. 5(b)), and the SIFT-like image features from 4×8 subregions were extracted (256 dimensions).

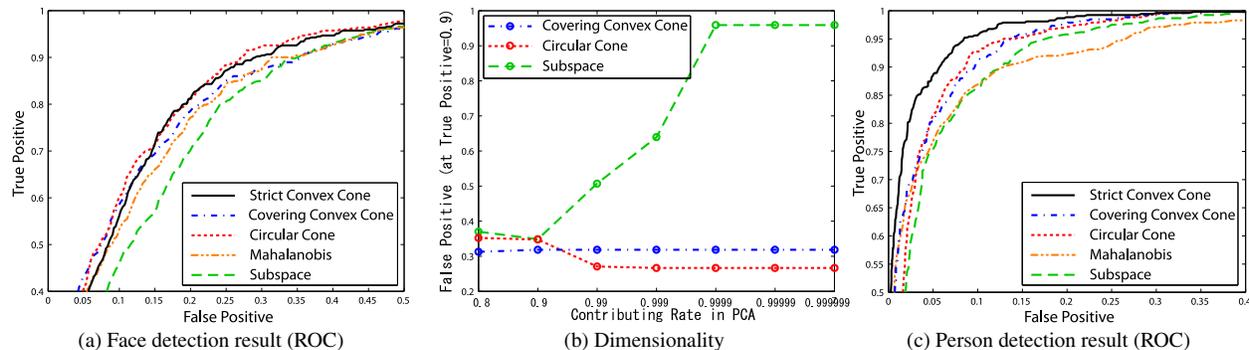


Figure 6. Experimental results.

The proposed methods were compared to methods using the Mahalanobis distance and the subspace of normalized vectors. The results of face detection are shown in Fig. 6(a). The circular cone and strict convex cone are superior to the others, and the covering convex cone has a similar performance to the Mahalanobis distance method, while the subspace method is least effective. The relationship between the contributing rate in PCA, which controls the dimensionality of the subspace, and the error rate is shown in Fig. 6(b). The performance of the subspace method is affected by the dimensionality and has a peak at the trade-off point. For the proposed methods, however, the error rate do not increase, even for high dimensionality. Since the proposed methods appropriately approximate the sample distribution by a cone, the effect of the redundant dimensionality in the subspace is reduced. Thus, we can simply employ higher dimensionality for the proposed methods. The results of person detection are shown in Fig. 6(c). The proposed methods are superior, and in particular, the strict convex cone is best. The figures of persons include more variance than those of faces, which results in a more complicated distribution. The subspace and normal distribution for the Mahalanobis distance is not able to capture such a complicated distribution. In summary, the effectiveness of approximation by cones has been demonstrated.

4. Conclusion

We have proposed three types of cone-restricted subspace method: the strict convex cone, the covering convex cone and the circular cone. They approximate the sample distributions by means of a cone pointed from the origin. They are able to strictly classify the vectors at the boundary of the cone, while admitting scaling and additivity of vectors as well as ordinary subspace methods. The experimental results for face and person detection have demonstrated the effectiveness of the proposed methods. The future work is to extend the pro-

posed methods to deal with multi-modal cones, while we assume unimodal cone in this paper.

Acknowledgement

This work was supported by KAKENHI (19700187). The authors thank Satoshi Itou for his advice on NNLS.

References

- [1] CBCL Face Database, www.ai.mit.edu/projects/cbcl.
- [2] R. Bro and S. Jong. A fast non-negativity-constrained least squares algorithm. *Journal of Chemometrics*, 11, 1997.
- [3] K. Fukui and O. Yamaguchi. Face recognition using multi-viewpoint patterns for robot vision. In *International Symposium of Robotics Research*, 2003.
- [4] A. Georghiades, D. Kriegman, and P. Belhumeur. Illumination cones for recognition under variable lighting: Faces. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1998.
- [5] T. Kobayashi and N. Otsu. A three-way auto-correlation based approach to human identification by gait. In *IEEE Workshop on Visual Surveillance*, 2006.
- [6] D. Lowe. Distinctive image features from scale invariant features. *International Journal of Computer Vision*, 60, 2004.
- [7] T. Nanri and N. Otsu. Unsupervised abnormality detection in video surveillance. In *IAPR Conference on Machine Vision Applications*, 2005.
- [8] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio. Pedestrian detection using wavelet templates. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1997.
- [9] E. Oya. *Pattern recognition and subspace method*. Sangyo Tosho Co., 1986.
- [10] B. Scholkopf, J. Platt, J. Shawe-Taylor, A. Smola, and R. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13, 2001.
- [11] O. Yamaguchi, K. Fukui, and K. Maeda. Face recognition using temporal image sequence. In *International Conference on Automatic Face and Gesture Recognition*, 1998.