KERNEL BASED SYNTHETIC DISCRIMINANT FUNCTION FOR OBJECT RECOGNITION

Kyu-Hwa Jeong, Puskal P. Pokharel, Jian-Wu Xu, Seungju Han, Jose C. Principe

Computational NeuroEngineering Laboratory
Department of Electrical and Computer Engineering
University of Florida, Gainesville, FL 32607 U.S.A.
E-mail: {khjeong, principe}@cnel.ufl.edu

ABSTRACT

In this paper a non-linear extension to the synthetic discriminant function (SDF) is proposed. The SDF is a well known 2-D correlation filter for object recognition. The proposed non-linear version of the SDF is derived from kernel-based learning. The kernel SDF is implemented in a nonlinear high dimensional space by using the kernel trick and it can improve the performance of the linear SDF by incorporating the image's class higher order moments. We show that this kernelized composite correlation filter has an intrinsic connection with the recently proposed correntropy function. We apply this kernel SDF to face recognition and simulations show that the kernel SDF significantly outperforms the traditional SDF as well as is robust in noisy data environments.

1. INTRODUCTION

Correlation filters have been applied successfully to target detection and recognition problems such as automatic target detection (ATR) [1] and face image recognition [2]. Object recognition is performed by cross-correlating an input image with a synthesized template (filter) and the correlation output is searched for the peak, which is used to determine whether the object of interest is present or not. It is well known that matched filters are the optimal linear filters for signal detection under linear channel and white noise conditions [3]. For image detection, matched spatial filters (MSF) are optimal in the sense that they provide the maximum output signal to noise ratio (SNR) for the detection of a known image in the presence of white noise, under the reasonable assumption of Gaussian statistics [4]. However, the performance of the MSF is very sensitive to even small changes in the reference image and the MSF cannot be used for multiclass pattern recognition since the MSF is only optimum for a single image. Therefore distortion invariant composite filters have been proposed in various papers [1].

The most well known of such composite correlation filters are the synthetic discriminant function (SDF) [5] and its

variations. In the conventional SDF approach, the filter is matched to a composite image that is a linear combination of the training image vectors such that the cross correlation output at the origin has the same value with all training images. The hope is that this composite image will correlate equally well not only with the training images but also with other distorted versions of that training images, even with test images in the same class. The shortcomings of the conventional SDF are that the SDF does not consider any input noise and it has a poor rejecting ability for out of class images since it controls only a single point in the output correlation plane. Minimum variance SDF (MVSDF) filter has been proposed in [6] taking into consideration additive input noise. The MVSDF minimizes the output variance due to zero-mean input noise while satisfying the same linear constraints as the SDF. One of major difficulty in MVSDF is that we often do not know the noise covariance exactly; even when we do know it, we need its inversion and it may be computationally impossible in practice. Another correlation filter that produces a sharp correlation peak is the minimum average correlation energy (MACE) filter [7]. The MACE minimizes the average correlation energy of the output over the training images subject to the same linear constraints as the SDF filters.

Most of these are linear correlation filters. A nonlinear extension to MACE filter has been proposed in [8]. Recently, kernel based learning algorithms have been exploited due to the fact that linear algorithms can be easily extended to nonlinear versions by the kernel method [9]. The kernel matched spatial filter (KMSF) has been proposed for hyperspectral target detection in [10]. A new generalized correlation function, called correntropy, defined in a nonlinear *reproducing kernel Hilbert space* (RKHS) has been proposed in [11] and its application to the matched filter has been presented in [12].

In this paper we propose a new kernel SDF using a Mercer kernel. In section 2, the conventional SDF filter is reviewed briefly. In section 3, we propose a new kernel SDF and show the connection with the correntropy. In section 4, we present simulation results for face recognition and section 5 summarizes and points out some further research.

2. SYNTHETIC DISCRIMINANT FUNCTION

We consider a 2-dimensional image data as a $d \times 1$ column vector $\mathbf{x}$, where d is the number of pixel in the image. This 1-dimensional image data can be obtained by lexicographic ordering the image rows. The SDF filter is matched to a composite image $\mathbf{h}$, where $\mathbf{h}$ is a linear combination of the training image vectors

$$\mathbf{h} = \sum_{i=1}^{N} a_i \mathbf{x}_i, \tag{1}$$

where N is the number of training images and the coefficients a_i are chosen to satisfy the following constraints

$$\mathbf{h}^T \mathbf{x}_j = u_j, \qquad j = 1, 2, \cdots, N, \tag{2}$$

where T denotes the transpose and u_j is a desired cross correlation output peak value. In vector form, we define the training image data matrix $\mathbf{X}$ as

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N], \tag{3}$$

where the size of matrix $\mathbf{X}$ is $d \times N$. Then the SDF is the solution to the following optimization problem

$$\min \mathbf{h}^T \mathbf{h}, \quad \text{subject to } \mathbf{X}^T \mathbf{h} = \mathbf{u}. \tag{4}$$

It is assumed that $N < d$ and so the problem formulation is a quadratic optimization subject to an under-determined system of linear constraints. The optimal solution is

$$\mathbf{h} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{u}. \tag{5}$$

Once $\mathbf{h}$ is determined, we apply an appropriate threshold to the output of the cross correlation, which is the inner product of the test input image and the filter $\mathbf{h}$ and decide on the class of the test image.

3. KERNEL BASED METHOD

3.1. Kernel method

The basic idea of kernel algorithm is to transform the data $\mathbf{x}_i$ from the input space to a high dimensional feature space of vectors $\Phi(\mathbf{x}_i)$, where the inner products can be computed using a positive definite kernel function satisfying Mercer's condition [9],

$$k(\mathbf{x}, \mathbf{y}) = < \Phi(\mathbf{x}), \Phi(\mathbf{y}) > . \tag{6}$$

This simple and elegant idea allows us to obtain nonlinear versions of any linear algorithm expressed in terms of inner products, without even knowing the exact mapping function Φ. A particularly interesting characteristic of the feature space is that it is *reproducing kernel Hilbert space* (RKHS).

In this paper, we use the Gaussian kernel, which is the most widely used Mercer kernel,

$$k(\mathbf{x} - \mathbf{y}) = \frac{1}{\sqrt{2\pi}\sigma} exp - (\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}). \tag{7}$$

3.2. Kernel SDF

Based on the kernel methodology, the previous optimization problem for the SDF can be solved in a higher dimensional kernel feature space by transforming each element of the matrix of exemplars $\mathbf{X}$ to $\Phi(\mathbf{X}_{ij})$ and $\mathbf{h}$ to $\Phi(\mathbf{h})$, thus forming a higher dimensional matrix $\Phi(\mathbf{X})$ whose ijth feature vector is $\Phi(\mathbf{X}_{ij})$.

Then we can extend the SDF optimization problem to the nonlinear feature space by

$$\min \Phi^T(\mathbf{h})\Phi(\mathbf{h}), \quad \text{subject to } \Phi^T(\mathbf{X})\Phi(\mathbf{h}) = \mathbf{u}. \tag{8}$$

where the dimensions of the transformed $\Phi(\mathbf{X})$ and $\Phi(\mathbf{h})$ are $\infty \times N$ and $\infty \times 1$, respectively for the Gaussian kernel. Then the solution in kernel space becomes

$$\Phi(\mathbf{h}) = \Phi(\mathbf{X})(\Phi^T(\mathbf{X})\Phi(\mathbf{X}))^{-1}\mathbf{u}. \tag{9}$$

We denote $\mathbf{K}_{XX} = \Phi^T(\mathbf{X})\Phi(\mathbf{X})$, which is a $N \times N$ full rank matrix whose ijth element is given by

$$(\mathbf{K}_{XX})_{ij} = \sum_{k=1}^{d} k(x_{ki}, x_{kj}), \; i, j = 1, 2, \cdots, N. \tag{10}$$

Although $\Phi(\mathbf{h})$ is a infinite dimensional vector, the output of this filter is going to be an $N \times 1$, which can be easily computed using these kernels.

Let $\mathbf{Z}$ be the matrix of vector images for testing and its number of testing images are L. We denote $\mathbf{K}_{ZX} = \Phi^T(\mathbf{Z})\Phi(\mathbf{X})$, which is $L \times N$ matrix whose each element is given by

$$(\mathbf{K}_{ZX})_{ij} = \sum_{k=1}^{d} k(z_{ki}, x_{kj}), \; i = 1, 2, \cdots, L, j = 1, 2, \cdots, N. \tag{11}$$

Then the $L \times 1$ output vector of the kernel SDF is given by

$$\mathbf{y} = \Phi^T(\mathbf{Z})\Phi(\mathbf{h}) = \mathbf{K}_{ZX}\mathbf{K}_{XX}^{-1}\mathbf{u}. \tag{12}$$

By applying an appropriate threshold to the output in (12), we can detect and recognize the testing data without generating the composite filter in a feature space.

Here, we can see that this kernel SDF is related to correntropy. Correntropy, as proposed in [11], is a positive definite function that generalizes the correlation function to nonlinear (non Gaussian) manifolds. The correntropy of the random process $x(t)$ at instances t_1 and t_2 is defined as

$$V(t_1, t_2) = E[k(x_{t_1}, x_{t_2})], \tag{13}$$

where E is the expectation operator and k is a kernel function that obeys the Mercer's conditions. The values in (10) are the cross correntropy between ith training image and jth training image at the zero lag and the values in (11) are the cross correntropy between ith testing image and jth training image at the zero lag. Then we can say that the output of the kernel SDF is related to the zero-lag correntropy between test images and training images. Therefore, the kernel SDF generalizes the correlation at the zero lag in the nonlinear feature space.

Fig. 1. Sample images: (a)person A (b) person B (c) person C with additive Gaussian noise (SNR=10db).

4. SIMULATIONS

In this section, we show the performance of the proposed kernel based SDF filter for face image recognition. In the simulations, we used the facial expression database collected at the Advanced Multimedia Processing Lab at the Electrical and Computer Engineering Department of CMU [13]. The database consists of 13 subjects, whose facial images were captured with 75 varying expressions. The size of each image is 64×64. Sample images are depicted in Fig. 1. The sample images with additive Gaussian noise with 10 dB SNR are shown in Fig. 1(c). Only three images per person were utilized for training one SDF per person. In order to evaluate the performance of the SDF in this data set, we examined $975(13\times75)$ correlation outputs. From these results and the ones reported in [2] as well as due to paper size limitations, we picked and report the results on only two person's who produced the worst performance with the conventional SDF method. We test with all the images of each person's data set resulting in 150 outputs for each class. The simulation results have been obtained by averaging (Monte-Carlo approach) over 100 randomly chosen training sets to minimize the problem of performance differences due to splitting the relatively small database in training and testing sets.

Fig. 2 shows the average output peak values for image recognition when we use only $N = 3$ images as training. The desired output peak value should be close to one when the test image belongs to the training image class. Fig. 2 (Top) shows that the correlation output peak values of the conventional SDF in both true and false classes not only overlap but are also close to one. As a result the system will have great difficulty to recognize these two individuals because they can be interpreted as belonging to the same class. Fig. 2 (Bottom) shows the output values of kernel SDF and we can see that the two images can be recognized well even with a small number of training images. Fig. 3 shows the ROC curves with different number of training images (N). In the kernel SDF with $N = 3$, the probability of detection with zero false alarm rate is 1. However, the conventional SDF needs at least 25 images
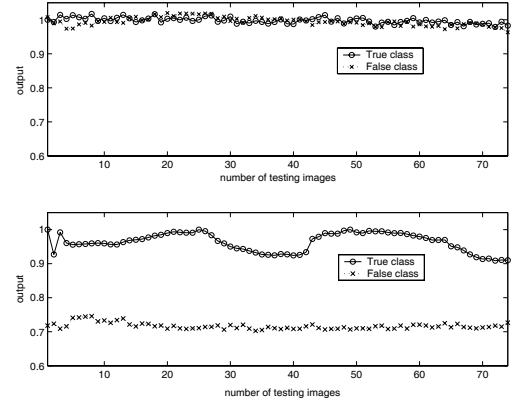

Fig. 2. The output peak values when only 3 images are used for training (N=3), (Top): Conventional SDF, (Bottom): Kernel SDF.
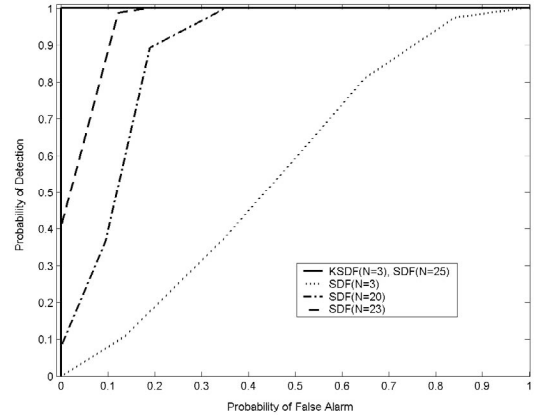

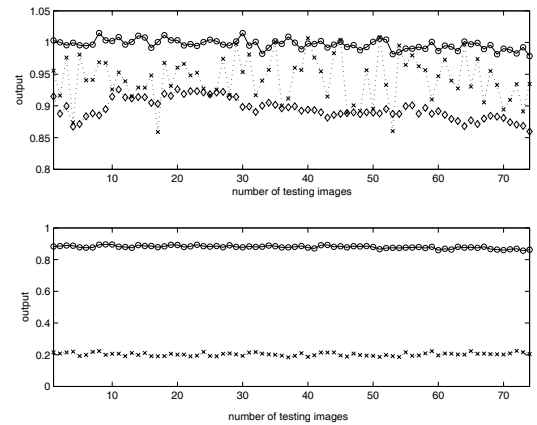**Fig. 3**. The comparison of ROC curves with different number of training images (N).

Fig. 4. The output values of noisy test input images with additive Gaussian noise when 25 images are used for training (N=25),(Top): Conventional SDF, circle-true class with SNR=10dB, cross-false class with SNR=-2dB, diamond-false class with no noise, (Bottom): kernel SDF, circle-true class with SNR=10dB, cross-false class with SNR=-2dB.
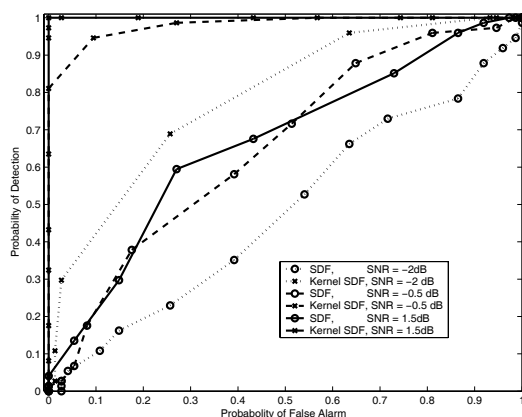
Fig. 5. The ROC curves of noisy test input images with different SNRs when 10 images are used for training (N=10).

for training in order to have the same detection performance as the kernel SDF.

One of major problems of the conventional SDF is that the performance can be easily degraded by additive noise in the test image since SDF does not have any special mechanism to consider input noise. Therefore, it has a poor rejecting ability for a false class image. Fig. 4 (Top) shows the noise effect on the conventional SDF. When the class images are seriously distorted by additive Gaussian noise with a very low SNR (-2dB), the correlation output peaks of some test images become great than 1, hence wrong recognition happens. The results in Fig. 4 (Bottom) are obtained by the kernel SDF. The kernel SDF shows a much better performance even in a very low SNR environment. The comparison of ROC curves between the kernel SDF and the conventional SDF in the case of noisy test input with different SNRs is shown in Fig. 5. We can see that the kernel SDF outperforms the SDF and achieves a robust pattern recognition performance in a very high noisy environment.

5. CONCLUSIONS

In this paper, we have proposed and evaluated a kernel based nonlinear SDF method for object recognition. We presented experimental results for face recognition. Using the kernel trick, the nonlinear version of SDF can be easily implemented in a higher dimensional feature space and this kernel SDF overcomes the two main shortcomings of the conventional SDF; one is poor rejecting performance and the other is the effect of the input noise. Simulation results show that the detection and recognition performance of the kernel SDF is much better than that of the SDF in particular with a small number of training data, which indicates that the filter generalizes better. The kernel SDF is also robust in a higher input noise data environment. Although we have focused this study on the SDF, we think that similar ideas could be applied to its

variants such as MVSDF and MACE filter by exploiting the correntropy function. Tests in more difficult datasets are also needed to fully characterize the performance of the kernel SDF.

6. REFERENCES

[1] B.V.K. Vijaya Kumar, "Tutorial survey of composite filter designs for optical correlators," vol. 31, pp. 4773–4801, 1992.

[2] M. Savvides, B.V.K. Vijaya Kumar, and P.K. Khosla, "Face verification using correlation filters," in *Proceedings of Third IEEE Automatic Identification Advanced Technologies*, Tarrytown,NY, 2002, pp. 56–61.

[3] G.L. Turin, "An introduction to matched filters," *IEEE Trans. Information Theory*, vol. 6, pp. 311–329, 1960.

[4] A. VanderLugt, "Signal detection by complex spatial filtering," *IEEE Trans. Information Theory*, , no. 10, pp. 139–145, 1964.

[5] C.F. Hester and D. Casasent, "Multivariant technique for multiclass pattern rcognition," *Appl.Opt*, vol. 19, pp. 1758–1761, 1980.

[6] B.V.K. Vijaya Kumar, "Minimum variance synthetic discriminant functions," *J.Opt.Soc.Am.A*, vol. 3, no. 10, pp. 1579–1584, 1986.

[7] A. Mahalanobis, B.V.K. Vijaya Kumar, and D. Casasent, "Minimum average correlation energy filters," *Appl.Opt*, vol. 26, no. 17, pp. 3633–3640, 1987.

[8] J.W. Fisher and J.C. Principe, "Recent advances to nonlinear minimum average correlation energy filter," *Optical Engineering*, October 1997.

[9] B. Scholkopf and A. J. Smola, *Learning with Kernels*, The MIT Press, 2002.

[10] H. Kwon and N. M. Nasrabadi, "Kernel matched signal detectors for hyperspectral tareget detection," in *Proceedings of ICASSP*, 2005, vol. 4, pp. 665–668.

[11] I. Santamaría, P.P. Pokharel, and J.C. Principe, "Generalized correlation function: Definition,properties and application to blind equalization," *to appear IEEE Trans. Signal Processing*.

[12] P.P. Pokharel, R. Agrawal, and J.C. Principe, "Correntropy based matched filtering," in *Proceedings of MLSP*, Sept. 2005, pp. 148–155.

[13] "http://www.amp.ece.cmu.edu : Advanced multimedia processing lab at electrical and computer eng., cmu," .