Neurocomputing 72 (2008) 102–111

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Enhancing the correntropy MACE filter with random projections ☆

Kyu-Hwa Jeong [a,*], Jose C. Principe [b]

[a] Intel Mask Operation, 2200 Mission College Blud, Santa Clara, CA 95052, USA
[b] Computational NeuroEngineering Laboratory, Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA

ARTICLE INFO

Available online 17 September 2008

Keywords:
Minimum average correlation energy
Pattern recognition
Correntropy
Random projection
Dimensionality reduction
Face recognition

ABSTRACT

The minimum average correlation energy (MACE) filter is a well known correlation filter for object recognition. Recently, a nonlinear extension to the MACE filter using the correntropy function in feature space has been introduced. Correntropy is a positive definite function that generalizes the concept of correlation by utilizing higher order moment information of signal structure. Since the MACE is a spatial matched filter for an image class, the correntropy MACE (CMACE) can potentially improve its performance. Both the MACE and CMACE are basically memory-based algorithms and due to the high dimensionality of the image data, the computational cost of the CMACE filter is one of the critical issues in practical applications. We propose to use a dimensionality reduction method based on random projections (RP), which has emerged as a powerful method for dimensionality reduction in machine learning. We apply the CMACE filter with random projection (CMACE-RP) to face recognition and show that it indeed outperforms the traditional linear MACE in both generalization and rejection abilities with small degradation in performance, but great savings in storage and computational complexity.

1. Introduction

Correlation filters have been applied successfully to target detection and recognition problems such as automatic target recognition (ATR) [41] and face recognition [34,44,43]. The advantages of the correlation filter approach are simplicity and rotation and shift invariance properties [42]. 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 [37]. 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 [39]. 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 [41].

The most well known of such composite correlation filters are the synthetic discriminant function (SDF) [20] 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 (false) images since it controls only a single point in the output correlation plane. The minimum variance SDF (MVSDF) filter has been proposed in [40] to take 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 the major difficulties in MVSDF is that we often do not know the noise covariance exactly. Another correlation filter that is widely used is the minimum average correlation energy (MACE) filter [29]. The MACE minimizes the average correlation energy of the output over the training images to produce a sharp correlation peak subject to the same linear constraints as the MVSDF and SDF filters. In practice, the MACE filter performs better than the MVSDF with respect to rejecting out-of-class input images. The MACE filter, however, has been

☆ This work was supported in part by the National Science Foundation under Grant ECS-0601271.
* Corresponding author.
E-mail addresses: kyu-hwa.jeong@intel.com, khjeong@cnel.ufl.edu (K.-H. Jeong), principe@cnel.ufl.edu (J.C. Principe).

0925-2312/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2007.12.045

shown to have poor generalization properties, that is, images in the recognition class but not in the training exemplar set are not recognized well. Therefore, the optimal trade-off filters have been proposed by [32] to combine the properties of various SDFs. Most of these are linear correlation filters. A nonlinear extension to the MACE filter using information theoretic learning (ITL) criterion has also been proposed in [13].

Recently, kernel-based learning algorithms have been heavily researched due to the fact that linear algorithms can be easily extended to nonlinear versions by the kernel method [35]. The kernel matched spatial filter (KMSF) has been proposed for hyperspectral target detection in [26] and the kernel SDF has been proposed for face recognition [23]. More recently, a new generalized correlation function, called correntropy has been introduced by our group [33]. Correntropy is a positive definite function, which measures a nonlinear similarity between random variables (or stochastic processes) and it involves high-order statistics of input signals, therefore it could be a promising candidate for machine learning and signal processing. Correntropy defines a new reproducing kernel Hilbert space (VRKHS)[1] of the same dimension as the input, which simplifies the formulation of analytic solutions in the VRKHS. Applications to the matched filter [30], chaotic time series prediction [31], face recognition [24] and synthetic aperture radar (SAR) automatic target recognition [21,22] have been presented in the literature. In [24,22] we exploit the linear structure of the VRKHS to formulate the MACE filter with similar equations as in the input space, which we called the correntropy MACE (CMACE). However, due to the nonlinear relation between the input space and this feature space, the CMACE corresponds to a nonlinear filter in the input space and has been shown to possess better generalization and rejecting performance than the conventional MACE.

Both the conventional MACE and the CMACE are memory-based algorithms, therefore, in practice, the drawback of this class of algorithms is both the storage requirements and the high computational demand. The output of the CMACE filter is obtained by computing the product of two matrices defined by the image size and the number of training images, and each elements of the matrix requires $O(d^2)$ computations, where d is the number of image pixels. This quickly becomes too complex in practical settings even for relatively small images. The fast CMACE filter using the fast Gauss transform (FGT) has been proposed in [21] to increase the computational speed of the CMACE filter, but the storage is still high. When the number of training images is N, the total computation complexity of one test output of the CMACE is $O(d^2N(N+1))$ and this can be reduced to $O(pcdN(N+1))$, where p is the order of the Hermite approximation and c is the number of clusters utilized in the FGT $(p, c \ll d)$.

In this paper we use a dimensionality reduction pre-processor based on random projections (RP) to decrease the storage and meet more readily available computational resources. RP has emerged as a powerful method for dimensionality reduction in machine learning and image compression [14,5], outperforming principal component analysis (PCA) in many applications. A theoretical underpinning to compressive sampling is currently being explored [10].

The organization of this paper is as follows. In Section 2, the MACE filter is reviewed briefly. In Section 3, we review also the CMACE filter based on correntropy and present the dimensionality reduction method with RP in Section 4. In Section 5, we present simulation results for face recognition and Section 6 summarizes and points out some further research.

[1] In this paper, we call the RKHS induced by correntropy VRKHS.

2. MACE filter

In this section, we summarize the MACE filer formulation in [29]. We consider a two-dimensional ith image as a $d \times 1$ column vector $\mathbf{x}_i$, where d is the number of pixel. This one-dimensional discrete sequence can be obtained by lexicographically reordering the image. The conventional MACE filter is formulated in frequency domain. The discrete Fourier transform (DFT) of the sequence $\mathbf{x}_i$ is denoted by $\mathbf{X}_i$ and we define the training image data matrix $\mathbf{X}$ as

$$\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N], \tag{1}$$

where the size of $\mathbf{X}$ is $d \times N$ and N is the number of training data. Let the vector $\mathbf{h}$ be the filter in space domain and its Fourier transform vector be $\mathbf{H}$. We are interested in the correlation of the input image and the filter. The correlation of the ith image sequence $x_i(n)$ with filter sequence $h(n)$ can be written as

$$g_i(n) = x_i(n) \otimes h(n). \tag{2}$$

By Parseval's theorem, the correlation energy of the ith image can be written as a quadratic form

$$E_i = \mathbf{H}^{\mathrm{H}}\mathbf{D}_i\mathbf{H}, \tag{3}$$

where $\mathbf{D}_i$ is a diagonal matrix of size $d \times d$ whose diagonal elements are the magnitude squared of the associated element of $\mathbf{X}_i$, that is, the power spectrum of $x_i(n)$ and the superscript H denotes the Hermitian transpose. The objective of the MACE filter is to minimize the average correlation energy over all signals while simultaneously satisfying intensity constraint at the origin for each signal. The value of the correlation at the origin can be written as

$$g_i(0) = \mathbf{X}_i^{\mathrm{H}}\mathbf{H} = c_i \tag{4}$$

for all $i = 1, 2, \ldots, N$ training images, where c_i is the user specified output correlation value at the origin for the ith image. Then the average energy over all training images is expressed as

$$E_{\mathrm{avg}} = \mathbf{H}^{\mathrm{H}}\mathbf{D}\mathbf{H}, \tag{5}$$

where

$$\mathbf{D} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{D}_i. \tag{6}$$

The MACE design problem is to minimize E_{avg} while satisfying the constraint, $\mathbf{X}^{\mathrm{H}}\mathbf{H} = \mathbf{c}$, where $\mathbf{c} = [c_1, c_2, \ldots, c_N]$ is an N-dimensional vector. This optimization problem can be solved using Lagrange multipliers, and the solution is

$$\mathbf{H} = \mathbf{D}^{-1}\mathbf{X}(\mathbf{X}^{\mathrm{H}}\mathbf{D}^{-1}\mathbf{X})^{-1}\mathbf{c}. \tag{7}$$

It is clear that $\mathbf{h}$ can be obtained from $\mathbf{H}$ by an inverse DFT. Once $\mathbf{h}$ is determined, we apply an appropriate threshold to the output correlation plane and decide on the class of the test image.

3. Nonlinear version of the MACE in RKHS using correntropy

3.1. Correntropy function in RKHS

Cross correntropy or simply correntropy [33] is a generalized similarity measure between two arbitrary vector random variables X and Y defined as

$$V(X, Y) = E[\kappa_\sigma(X - Y)], \tag{8}$$

where E is the mathematical expectation and κ_σ is the Gaussian kernel given by

$$\kappa_\sigma(X - Y) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{\|X - Y\|^2}{2\sigma^2}\right\}, \tag{9}$$