Kernel uncorrelated neighborhood discriminative embedding for feature extraction

Xuelian Yu and Xuegang Wang
University of Electronic Science and Technology
School of Electronic Engineering
Chengdu, 610054, China

Abstract. Feature extraction is a crucial step for pattern recognition. Recently, some manifold learning algorithms have drawn much attention. Although their properties of locality preserving are fairly significant, most manifold-based algorithms have limits to solve classification problems. First, they do not have good discriminant ability. Second, they fail to remove the redundancy among the extracted features. We present a new feature extraction method, called kernel uncorrelated neighborhood discriminative embedding (KUNDE), which integrates two abilities of manifold learning and pattern classification. The purpose of KUNDE is to preserve the within-class neighboring geometry while maximizing the between-class scatter. Optimizing an objective function in a kernel feature space, nonlinear features are extracted. Moreover, by putting a simple uncorrelated constraint on the computation of the basis vectors, the extracted features via KUNDE are statistically uncorrelated and thus contain minimum redundancy. Experimental results on radar target recognition indicate the promising performance of the proposed method. © *2007 Society of Photo-Optical Instrumentation Engineers.*
[DOI: 10.1117/1.2821866]

1 Introduction

In the past few years, some nonlinear manifold learning algorithms have been proposed to discover the nonlinear structure of the manifold by investigating the local geometry of samples, such as isomap,[1] locally linear embedding (LLE),[2,3] and Laplacian eigenmap.[4] These methods are defined only on training data, and the issue of how to map new test data remains difficult. Therefore, they cannot be applied directly to recognition problems.

Recently, some manifold-based linear algorithms, like neighborhood preserving embedding (NPE),[5] resolve the difficulty by finding a mapping on the whole data space, not just on training data. However, they have a common inherent limitation: they deemphasize discriminant information that is very important for the recognition task. In addition, the basis vectors of these methods are statistically correlated, and then the extracted features contain much redundancy. As a result, the overlapped information can distort the distribution of the features and even degrade the recognition performance.

In this work, a new feature extraction algorithm, called kernel uncorrelated neighborhood discriminative embedding (KUNDE) is proposed to address the problems mentioned earlier. On the one hand, the method explicitly considers both the within-class neighboring information and the between-class scatter information and emphasizes the discriminant information. On the other hand, the method obtains statistically uncorrelated features with minimum redundancy by putting a simple uncorrelated constraint on the computation of the basis vectors. Mapping the input data to some high-dimensional feature space using the kernel technique, nonlinear features are extracted.

2 Kernel Uncorrelated Neighborhood Discriminative Embedding

Given a data set $X=[x_1,\ldots,x_N]$ in R^D, suppose that each data point x_i belongs to one of the C classes $\omega_1,\ldots,\omega_C$, and each class contains n_c ($c=1,\ldots,C$) samples. The data is then mapped into an implicit high-dimensional feature space F by a nonlinear mapping function $\phi:x\in R^D \to \phi(x)\in F$. The problem is to find a transformation matrix V that maps these points to be new points $Y=[y_1,\ldots,y_N] =V^T\phi(X)$ in $R^d(d\ll D)$, where $y_i=V^T\phi(x_i)$.

The algorithm procedure of KUNDE is stated here:

1. Construct the kernel matrix $K=\phi(X)^T\phi(X)$ whose elements are $K_{ij}=k(x_i,x_j)$, where k is a kernel function that satisfies $k(x_i,x_j)=[\phi(x_i)\cdot\phi(x_j)]$.
2. Compute the weight matrix W by minimizing the reconstruction error $E(W)=\Sigma_i\|\phi(x_i)-\Sigma_j W_{ij}\phi(x_j)\|^2$, where $\Sigma_j W_{ij}=1$, and $W_{ij}\neq 0$ if $\phi(x_j)$ is one of the n identical-label nearest neighbors of $\phi(x_i)$; otherwise, $W_{ij}=0$. An efficient way to minimize the error can refer to Ref. 3.
3. Compute matrices M, L, and G as follows: $M=(I -W)(I-W)^T$, $L=I-E$, and $G=I-(1/N)ee^T$, where I is an identity matrix, $E_{ij}=1/n_c$ if x_i and x_j belong to the cth class; otherwise, $E_{ij}=0$, and $e=(1,\ldots,1)^T$.
4. Solve the generalized eigenvalue problem $K(M +L)Ka_i=\lambda_i KGKa_i$, with $\lambda_1<\cdots<\lambda_d$, and constitute the matrix $A=[a_1,\ldots,a_d]$.
5. For any data point x in R^D, the embedded feature in R^d is given by $y=V^T\phi(x)=A^T[k(x_1,x),\ldots, k(x_N,x)]^T$.

3 Theoretical Justification

In this section, we provide the theoretical analysis of the KUNDE algorithm. As mentioned earlier, the objective of KUNDE is to preserve the within-class neighboring geometry while maximizing the between-class scatter in the low-dimensional space.

First, we characterize the within-class geometry of each data point in the feature space F by linear coefficients that reconstruct the data point from its n identical-label nearest neighbors. To preserve the within-class neighboring relations, the basic idea is that the same weights that reconstruct the point $\phi(x_i)$ in F should also reconstruct its embedded counterpart y_i in R^d. Therefore, we should minimize the following cost function:

$$J_1(V) = \sum_i \left\| y_i - \sum_j W_{ij} y_j \right\|^2 = \| Y(I-W) \|^2,$$

$$= trace[Y(I-W)(I-W)^T Y^T],$$

$$= trace[V^T \phi(X) M \phi(X)^T V], \qquad (1)$$

where $M = (I-W)(I-W)^T$.

Second, since the purpose of KUNDE is to solve classification problems, we should make the embedded vectors from different classes far from each other. Here, we propose to maximize the between-class scatter:

$$J_2(V) = trace(V^T S_b^\phi V), \qquad (2)$$

where S_b^ϕ is the between-class scatter matrix in F. According to Ref. 6, the total scatter matrix S_t^ϕ, within-class scatter matrix S_w^ϕ, and between-class scatter matrix S_b^ϕ can be represented as follows:

$$S_t^\phi = (1/N) \sum_{i=1}^N (\phi(x_i) - m)(\phi(x_i) - m)^T,$$

$$= (1/N)\phi(X)(I - ee^T/N)\phi(X)^T = \phi(X) G \phi(X)^T,$$

$$S_w^\phi = \sum_{c=1}^C \sum_{x \in \omega_c} [\phi(x) - m_c][\phi(x) - m_c]^T,$$

$$= \phi(X)(I - E)\phi(X)^T = \phi(X) L \phi(X)^T,$$

$$S_b^\phi = S_t^\phi - S_w^\phi = \phi(X)(G - L)\phi(X)^T = \phi(X) B \phi(X)^T, \qquad (3)$$

where $G = I - (1/N)ee^T$, $L = I - E$, and $B = G - L$. Thus, Eq. (2) can be rewritten as:

$$J_2(V) = trace[V^T \phi(X) B \phi(X)^T V]. \qquad (4)$$

Combining Eqs. (1) and (4), we should minimizing the following objective function:

$$J(V) = \frac{trace[V^T \phi(X) M \phi(X)^T V]}{trace[V^T \phi(X) B \phi(X)^T V]}. \qquad (5)$$

Now, we turn to the statistically uncorrelated constraint. Assuming that any two different components y_i and y_j ($j \neq i$) of the extracted feature $y = V^T x$ are uncorrelated, this means that:

$$E\{[y_i - E(y_i)][y_j - E(y_j)]\} = v_i^T S_t^\phi v_j = 0, \qquad (6)$$

where v_i and v_j are two different columns of the matrix V. Besides, v_i should be normalized. Without loss of generalization, let v_i satisfy:

$$v_i^T S_t^\phi v_i = 1. \qquad (7)$$

Then, from Eqs. (6) and (7), we get:

$$V^T S_t^\phi V = I. \qquad (8)$$

As a result, KUNDE can be formulated as the following constrained minimization problem:

$$\min_{V^T S_t^\phi V = I} \frac{trace[V^T \phi(X) M \phi(X)^T V]}{trace[V^T \phi(X) B \phi(X)^T V]}$$

$$= \min_{V^T \phi(X) G \phi(X)^T V = I} \frac{trace[V^T \phi(X) M \phi(X)^T V]}{trace[I - V^T \phi(X) L \phi(X)^T V]}. \qquad (9)$$

Further, Eq. (9) is equivalent to:

$$\min_{V^T \phi(X) G \phi(X)^T V = I} trace[V^T \phi(X)(M + L)\phi(X)^T V]. \qquad (10)$$

Since each column of V should lie in the span of all training samples in F, there exist coefficients $\alpha_j (j = 1, \ldots, N)$ such that $v = \sum_{j=1}^N \alpha_j \phi(x_j) = \phi(X)a$, where $a = [\alpha_1, \ldots, \alpha_N]^T$. Therefore, Eq. (10) becomes:

$$\min_{A^T KGKA = I} trace[A^T K(M + L)KA], \qquad (11)$$

where K is the kernel matrix, which is defined in Sec. 2.

Last, the constrained minimization problem is reduced to a generalized eigenvalue problem, as follows:

$$K(M + L)Ka = \lambda KGKa. \qquad (12)$$

The matrix A is determined by the d eigenvectors corresponding to the first d smallest eigenvalues of Eq. (12). Once A is obtained, for any data point x in the input space, the nonlinear feature is given as $y = V^T \phi(x) = A^T [k(x_1, x), \ldots, k(x_N, x)]^T$.

4 Experimental Results

In this section, experiments are performed on radar target recognition with measured and simulated range profiles, respectively. The performance of the KUNDE algorithm is compared with those of kernel NPE (KNPE),[5] KPCA,[7] and KFDA.[8] The Gaussian kernel $k(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / \sigma^2)$ is adopted, and the parameter σ is simply set to 1. Since we
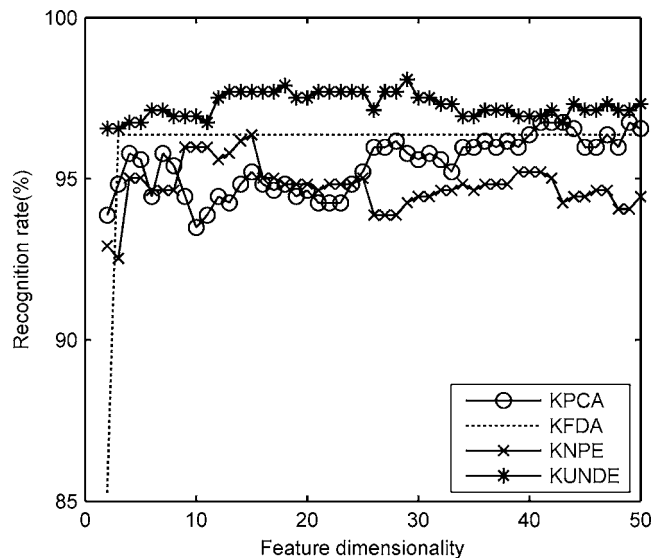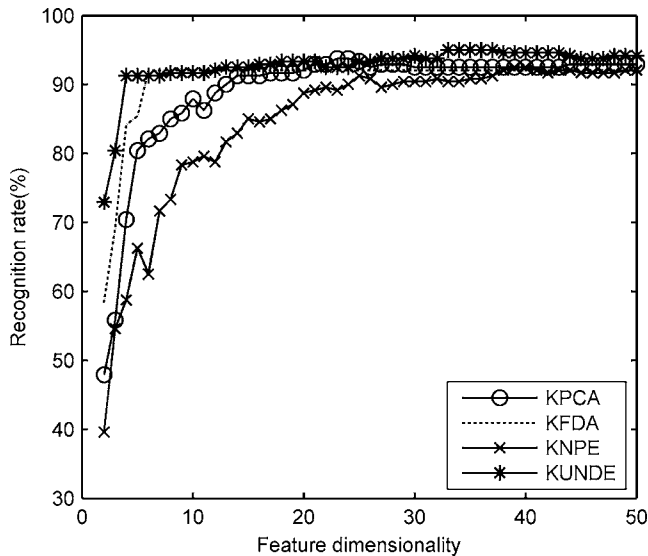

Fig. 1 Recognition rate versus reduced dimensionality on measured range profiles.

Fig. 2 Recognition rate versus reduced dimensionality on simulated range profiles.

focus only on feature extraction, as for classification, the nearest-neighbor classifier using a Euclidean metric is employed for the sake of simplicity.

4.1 Experiments on Measured Data

The measured data are from three flying airplanes, including An-26, Yark-42, and Cessna Citation S/II. Each airplane has 260 range profiles. In the experiments, each profile is preprocessed by energy normalization. Then, for each airplane, one third of all profiles are used for training and the rest for testing. Figure 1 shows the plots of recognition rates obtained by each method versus the reduced dimensionality. Note that the dimensionality of KFDA is at most $c-1$, where c is the number of classes.

From Fig. 1, it can be seen that KUNDE achieves the best recognition results at each feature dimensionality and KFDA performs second best, while KNPE and KPCA are relatively poor since neither of them considers discrimination information. This indicates that KUNDE has more discriminative power than the other three methods by incorporating the within-class neighboring information and the between-class scatter information and that the statistically uncorrelated features are very helpful for improving the recognition performance.

4.2 Experiments on Simulated Data

The simulated profiles are from six airplanes: Mirage, IDF, F16, J8II, SU27, and E2C, and each airplane has 60 pro-

files. Similarly, each profile is normalized to unit energy, and 20 profiles per target are used for training and the rest for testing. The recognition results are shown in Fig. 2. The results suggest that KUNDE outperforms again all the other methods. In addition, it can be seen that as the feature dimensionality increases, KUNDE has a higher recognition rate and KPCA and KNPE can also give satisfactory results.

5 Conclusions

A novel algorithm called kernel uncorrelated neighborhood discriminative embedding (KUNDE) is presented for pattern recognition. KUNDE has two prominent characteristics. First, it integrates the within-class neighboring information and the between-class scatter information, which enables its powerful discriminative ability. Second, it extracts statistically uncorrelated features with minimum redundancy by introducing an uncorrelated constraint, which is helpful to improve recognition performance. Experimental results on radar target recognition show that KUNDE achieves better recognition performance than all the other involved methods.

References

1. J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science* **290**(5500), 2319–2323 (2000).
2. S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science* **290**(5500), 2323–2326 (2000).
3. L. K. Saul and S. T. Roweis, "Think globally, fit locally: unsupervised learning of low dimensional manifolds," *J. Mach. Learn. Res.* **4**, 119–155 (2003).
4. M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.* **5**(6), 1373–1396 (2003).
5. X. F. He, D. Cai, S. C. Yan, and H. J. Zhang, "Neighborhood preserving embedding," in *Proc. 10th Int. Conf. Computer Vision* (2005).
6. X. F. He, S. C. Yan, Y. X. Hu, P. Niyogi, and H. J. Zhang, "Face recognition using Laplacian faces," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(3), 328–340 (2005).
7. B. Scholkopf, A. Smola, and K. R. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.* **10**, 1299–1319 (1998).
8. Q. S. Liu, R. Huang, H. Q. Lu, and S. D. Ma, "Face recognition using kernel-based fisher discriminant analysis," in *Proc. Int. Conf. Automatic Face and Gesture Recognition*, Washington (2002).