

Kernel Feature Extraction Approach for Color Image Recognition

Xiaoyuan Jing^{1,2}, Kun Li¹*, Songsong Wu¹, Yongfang Yao¹, Chao Wang¹

¹College of Automation, Nanjing University of Posts and Telecommunications, Nanjing, China

²State Key Laboratory of Software Engineering, Wuhan University, Wuhan, China

* Email:lk674372895@126.com

Abstract—Color Image Recognition is one of the most important fields in Pattern Recognition. Both Multi-set canonical correlation analysis and Kernel method are important techniques in the field of color image recognition. In this paper, we combine the two methods and propose one novel color image recognition approach: color image kernel canonical correlation analysis (CIKCCA). Color image kernel canonical correlation analysis is based on the theory of multi-set canonical correlation analysis and extracts canonical correlation features among the color image components. Then fuse the features of the color image components in the feature level, which are used for classification and recognition. Experimental results on the FRGC-v2 public color image databases demonstrate that the proposed approach acquire better recognition performance than other color recognition methods.

Keywords—color image; feature extraction; kernel method; canonical correlation analysis; discriminant analysis

I. INTRODUCTION

Recent year traditional gray image recognition has been unable to meet the demands of people, color image recognition is attracting a large number of researchers [1].

After the study of the neural network learning, kernel method [2] has become one of the most significant ways in the pattern recognition and machine learning field. It is based on statistical learning theory and support vector machine (SVM) and utilizes the nonlinear kernel function to reduce the computational complexity, which gained a wide range of application. LDA is a traditional statistical identification method, and at the same time, its widely using proved its efficiency in face recognition. However, kernel methods show better classification results than LDA and other traditional statistical methods. These years, some scholars put forward some typical kernel methods, which got better results than KDA [3, 4], KPCA[5] and KCCA [6]. Based on the analysis of these methods, we can find that kernel methods can greatly reduce the time complexity of the algorithm, and maximize the identification features in the high dimension space, therefore, kernel methods have been widely used.

CCA (Canonical Correlation Analysis) [7] is proposed by H.Hotelling in 1936, it has become one of the most widely used methods in color image recognition. It maximize the correlation by project the two groups of low dimensional variables to a lower dimensional space. Sun at al. [8] using CCA to improve the performance of multi-label classifier based on MMSE estimator, i.e, under a moderate high dimensional data structure condition, CCA can solve a minimum mean square error problem by constructing a class dependent instruction matrix. Based on the equivalent, CCA

can effectively solve the problem of minimum mean square error. Melzer at al. [9, 11] put forward kernel canonical correlation analysis algorithm for appearance model. Kernel CCA is a kind of CCA nonlinear extension, the input data is nonlinear transformed through the kernel function. Similar to traditional nonlinear method, kernel CCA is particularly suitable for the analysis of two related data set. Compare with other nonlinear feature fusion algorithms, such as kernel PCA, kernel CCA has an obvious advantage in the prediction of object position from a lot of pictures. Jing at al. [10] proposed a kind of color image canonical correlation analysis algorithm for color face image feature extraction and recognition. Although the traditional CCA can only solve the problem of correlation between the two groups of variables, and cannot be directly used to extracting features in color images that contain three groups of variables, this paper utilize CCA to solve the feature extraction involved three groups of variables, and apply preprocessed algorithm in the three color components of the color image to make the effectiveness of the algorithm verified in the database.

The rest of the paper is organized as follow: section 2 will give the principle of kernel discriminant analysis. In section 3, the novel color image kernel canonical correction analysis will be described in detail. Experiment results will be given in section 4. At last, section 5 will conclude the whole paper.

II. THE PRINCIPLE OF KERNEL DISCRIMINANT ANALYSIS

The kernel discriminant analysis is aimed to transform the original data from a low dimensional space into a nonlinear high dimensional kernel space. Assume that the original input sample sets $X = (x_1, x_2, x_3, \dots, x_N)$. For a nonlinear mapping: $\Phi: R^d \rightarrow K$, after mapping, the original samples can be mapped into the high dimensional kernel space K . In the kernel space K , it is devoted to find a linear transformation which can both maximize the between-class scatter matrix and minimize the within-class scatter matrix. Then we obtain a set of mapped samples: $X^\theta = (\theta(x_1), \theta(x_2), \theta(x_3), \dots, \theta(x_N))$, where N represents the number of samples per class. According to reproducing kernel theory, if kernel function $k(x, y)$ satisfies Mercer condition [9,11], the complex inner product calculation in feature space can be replaced.

Mercer condition: For any $\phi(x) \neq 0$, and $\int \phi(x)^2 dx < \infty$, $k(x, y)$ satisfy:

$\iint \phi(x)k(x,y)\phi(y)dxdy > 0$, $k(x,y)$ is called as kernel function.

Kernel function $k(x,y)$ is defined as follows:

$$K = (X^\Phi)^T T X^\Phi. \tag{1}$$

Assume that S_w^θ and S_b^θ are the within-class scatter and between-class scatter matrices, which are defined as follows:

$$S_w^\theta = \frac{1}{N} X_\theta T_w X_\theta^T, \tag{2}$$

$$S_b^\theta = \frac{1}{N} X_\theta \Omega T_b \Omega^T X_\theta^T, \tag{3}$$

where $X^\theta = (\theta(x_1), \theta(x_2), \theta(x_3), \dots, \theta(x_N))$.

III. COLOR IMAGE KERNEL CANONICAL CORRELATION ANALYSIS

A. Introduction of the color image kernel canonical correlation analysis

Assume there are N pairs of data in the kernel space $(x_i^\Phi, y_i^\Phi, z_i^\Phi)$, $i = 1, \dots, N$, $x_i^\Phi \in R^p$, $y_i^\Phi \in R^q$, $z_i^\Phi \in R^s$. They respectively belong to three sample set X^Φ, Y^Φ, Z^Φ . We will find three projection directions $\varphi_x, \varphi_y, \varphi_z$, to maximize the correlation coefficients between the three functions $\varphi_x^T(x_i - \bar{x}), \varphi_y^T(y_i - \bar{y}), \varphi_z^T(z_i - \bar{z})$, and we can get the canonical correlation characteristics between the three groups of data set. Their correlation coefficients r_1, r_2, r_3 are defined as follows:

$$r_1 = \frac{\varphi_x^T X^\Phi (Y^\Phi)^T \varphi_y}{\sqrt{\varphi_x^T X^\Phi (X^\Phi)^T \varphi_x} \sqrt{\varphi_y^T Y^\Phi (Y^\Phi)^T \varphi_y}}, \tag{4}$$

$$r_2 = \frac{\varphi_y^T Y^\Phi (Z^\Phi)^T \varphi_z}{\sqrt{\varphi_y^T Y^\Phi (Y^\Phi)^T \varphi_y} \sqrt{\varphi_z^T Z^\Phi (Z^\Phi)^T \varphi_z}}, \tag{5}$$

$$r_3 = \frac{\varphi_z^T Z^\Phi (X^\Phi)^T \varphi_x}{\sqrt{\varphi_z^T Z^\Phi (Z^\Phi)^T \varphi_z} \sqrt{\varphi_x^T X^\Phi (X^\Phi)^T \varphi_x}}, \tag{6}$$

where $X^\Phi = [x_1^\Phi - \bar{x}^\Phi, \dots, x_N^\Phi - \bar{x}^\Phi]$, $Y^\Phi = [y_1^\Phi - \bar{y}^\Phi, \dots, y_N^\Phi - \bar{y}^\Phi]$ and $Z^\Phi = [z_1^\Phi - \bar{z}^\Phi, \dots, z_N^\Phi - \bar{z}^\Phi]$. $\bar{x}^\Phi, \bar{y}^\Phi, \bar{z}^\Phi$ is mean vector of X^Φ, Y^Φ, Z^Φ respectively. In order to gain the maximum of the correlation coefficients r_1, r_2, r_3 , we transform (4),(5),(6) into the following optimization problem:

$$\begin{aligned} & \max \varphi_x^T X^\Phi (Y^\Phi)^T \varphi_y + \varphi_y^T Y^\Phi (Z^\Phi)^T \varphi_z + \varphi_z^T Z^\Phi (X^\Phi)^T \varphi_x, \\ & s.t. \quad \varphi_x^T X^\Phi (X^\Phi)^T \varphi_x = 1, \\ & \quad \varphi_y^T Y^\Phi (Y^\Phi)^T \varphi_y = 1, \\ & \quad \varphi_z^T Z^\Phi (Z^\Phi)^T \varphi_z = 1. \end{aligned}$$

(7)

Theorem 1: Function (7) is equivalent to solve the following three characteristic equations:

$$\begin{cases} X^\Phi (Y^\Phi)^T (Y^\Phi (Y^\Phi)^T)^{-1} Y^\Phi (Z^\Phi)^T (Z^\Phi (Z^\Phi)^T)^{-1} Z^\Phi (X^\Phi)^T \varphi_x \\ = \lambda X^\Phi (X^\Phi)^T \varphi_x \\ Y^\Phi (Z^\Phi)^T (Z^\Phi (Z^\Phi)^T)^{-1} Z^\Phi (X^\Phi)^T (X^\Phi (X^\Phi)^T)^{-1} X^\Phi (Y^\Phi)^T \varphi_y \\ = \lambda Y^\Phi (Y^\Phi)^T \varphi_y \\ Z^\Phi (X^\Phi)^T (X^\Phi (X^\Phi)^T)^{-1} X^\Phi (Y^\Phi)^T (Y^\Phi (Y^\Phi)^T)^{-1} Y^\Phi (Z^\Phi)^T \varphi_z \\ = \lambda Z^\Phi (Z^\Phi)^T \varphi_z \end{cases}, \tag{8}$$

Proof:

By using the Lagrange multiplier method, function (8) can be:

$$\begin{aligned} L = & \varphi_x^T X^\Phi (Y^\Phi)^T \varphi_y + \varphi_y^T Y^\Phi (Z^\Phi)^T \varphi_z + \varphi_z^T Z^\Phi (X^\Phi)^T \varphi_x \\ & + \lambda_x (1 - \varphi_x^T X^\Phi (X^\Phi)^T \varphi_x) + \lambda_y (1 - \varphi_y^T Y^\Phi (Y^\Phi)^T \varphi_y) + \\ & \lambda_z (1 - \varphi_z^T Z^\Phi (Z^\Phi)^T \varphi_z) \end{aligned}, \tag{9}$$

$\lambda_x, \lambda_y, \lambda_z$ is the Lagrange multiplier.

Take the derivative of $\varphi_x, \varphi_y, \varphi_z$ respectively, and set their partial derivative equal zero, that is:

$$\frac{\partial(L)}{\partial(\varphi_x)} = 0, \quad \frac{\partial(L)}{\partial(\varphi_y)} = 0, \quad \frac{\partial(L)}{\partial(\varphi_z)} = 0, \tag{10}$$

After the premultiplication $\varphi_x^T, \varphi_y^T, \varphi_z^T$ to the partial derived formula the following will be:

$$\varphi_x^T X^\Phi (Y^\Phi)^T \varphi_y + \varphi_x^T X^\Phi (Z^\Phi)^T \varphi_z = 2\lambda_x, \tag{11}$$

$$\varphi_y^T Y^\Phi (X^\Phi)^T \varphi_x + \varphi_y^T Y^\Phi (Z^\Phi)^T \varphi_z = 2\lambda_y, \tag{12}$$

$$\varphi_z^T Z^\Phi (Y^\Phi)^T \varphi_y + \varphi_z^T Z^\Phi (X^\Phi)^T \varphi_x = 2\lambda_z, \tag{13}$$

Put (11), (12), (13) into (9), and set $\frac{\partial(L_1)}{\partial(\varphi_x)}, \frac{\partial(L_2)}{\partial(\varphi_y)}, \frac{\partial(L_3)}{\partial(\varphi_z)}$ equal zero, the following will be:

$$Y^\Phi (Z^\Phi)^T \varphi_z - 2\lambda_y Y^\Phi (Y^\Phi)^T \varphi_y = 0, \tag{14}$$

$$Z^\Phi (X^\Phi)^T \varphi_x - 2\lambda_z Z^\Phi (Z^\Phi)^T \varphi_z = 0, \tag{15}$$

$$X^\Phi (Y^\Phi)^T \varphi_y - 2\lambda_x X^\Phi (X^\Phi)^T \varphi_x = 0, \tag{16}$$

Through the deduction, the characteristic equations can be got as the Theorem1.

Formula (9) makes three input data set canonical correlated. We call (9) as canonical correlation analysis on three data sets. Moreover, if we want to make kernel canonical correlation analysis on the three color component of the color image, we have to use kernel function instead of dot product operation, so we need:

Theorem 2: The desegregation $\varphi_x^*, \varphi_y^*, \varphi_z^*$ to (7) is $\varphi_x^* \in span(X^\Phi), \varphi_y^* \in span(Y^\Phi)$ and $\varphi_z^* \in span(Z^\Phi)$.

Proof:

Take the following factor decomposition $X^\Phi (X^\Phi)^T$:

$$X^\Phi (X^\Phi)^T = E \Lambda E^T = \sum_{i=1}^p \lambda_{xi} e_i e_i^T, \quad (17)$$

Assume $X^\Phi (X^\Phi)^T$ is singular, the first equation of (8) is expressed as a new equation:

$$\begin{aligned} (X^\Phi (X^\Phi)^T)^{-1} X^\Phi (Y^\Phi)^T (Y^\Phi (Y^\Phi)^T)^{-1} Y^\Phi (Z^\Phi)^T \\ (Z^\Phi (Z^\Phi)^T)^{-1} Z^\Phi (X^\Phi)^T \phi_x = \lambda \phi_x \end{aligned}, \quad (18)$$

To all solutions ϕ_x^* of (18), each ϕ_x^* is a feature vectors' linear combination of $X^\Phi (X^\Phi)^T$. To (17), each column vector of E are in the range of X^Φ , and this process is effective to all columns' linear combination, then we will get $\phi_x^* \in span(X^\Phi)$, also we can attain $\phi_y^* \in span(Y^\Phi)$ and $\phi_z^* \in span(Z^\Phi)$ in the same way.

According to the theorem 2, for $\phi_x^*, \phi_y^*, \phi_z^*$ in (18), there are $\phi_x, \phi_y, \phi_z \in R^N$, $\phi_x^* = X^\Phi \phi_x$, $\phi_y^* = Y^\Phi \phi_y$, $\phi_z^* = Z^\Phi \phi_z$. And (4) can be transformed as follow:

$$\begin{aligned} \max \phi_x^T (X^\Phi)^T X^\Phi (Y^\Phi)^T Y^\Phi \phi_y + \phi_y^T (Y^\Phi)^T Y^\Phi (Z^\Phi)^T Z^\Phi \phi_z \\ + \phi_z^T (Z^\Phi)^T Z^\Phi (X^\Phi)^T X^\Phi \phi_x, \\ \text{s.t. } \phi_x^T (X^\Phi)^T X^\Phi (X^\Phi)^T X^\Phi \phi_x = 1, \\ \phi_y^T (Y^\Phi)^T Y^\Phi (Y^\Phi)^T Y^\Phi \phi_y = 1, \\ \phi_z^T (Z^\Phi)^T Z^\Phi (Z^\Phi)^T Z^\Phi \phi_z = 1. \end{aligned}, \quad (19)$$

Formula (19) shows that we can use the dot product form to represent (8), then we can utilize kernel method to substitute the dot product operation. Define kernel matrix K_x, K_y, K_z as follows:

$$K_{xij} = e^{-\|x_i^\Phi - x_j^\Phi\|/\sigma}, \quad (20)$$

$$K_{yij} = e^{-\|y_i^\Phi - y_j^\Phi\|/\sigma}, \quad (21)$$

$$K_{zij} = e^{-\|z_i^\Phi - z_j^\Phi\|/\sigma}, \quad (22)$$

Where $K_x, K_y, K_z \in R^{N \times N}$, σ is a parameter. Then $X^\Phi (X^\Phi)^T, Y^\Phi (Y^\Phi)^T, Z^\Phi (Z^\Phi)^T$ in (19) is replaced by K_x, K_y, K_z , respectively. We may get the kernel canonical correlation analysis form of three variables of the color image:

$$\begin{aligned} \max \phi_x^T K_x K_y \phi_y' + \phi_y^T K_y K_z \phi_z' + \phi_z^T K_z K_x \phi_x', \\ \text{s.t. } \phi_x^T K_x^2 \phi_x' = 1, \\ \phi_y^T K_y^2 \phi_y' = 1, \\ \phi_z^T K_z^2 \phi_z' = 1. \end{aligned} \quad (23)$$

According to the theorem 1, theorem 3 is promoted.

Theorem 3: The solution set of (19) is equivalent to the solution set of the following three characteristic equations:

$$\begin{cases} K_x K_y (K_y^2)^{-1} K_y K_z (K_z^2)^{-1} K_z K_x \phi_x' \\ = \lambda K_x^2 \phi_x' \\ K_y K_z (K_z^2)^{-1} K_z K_x (K_x^2)^{-1} K_x K_y \phi_y' \\ = \lambda K_y^2 \phi_y' \\ K_z K_x (K_x^2)^{-1} K_x K_y (K_y^2)^{-1} K_y K_z \phi_z' \\ = \lambda K_z^2 \phi_z' \end{cases}, \quad (24)$$

The proof is similar to Theorem 1.

CIKCCA is used in color image recognition, whose input data sets are a color image's preprocessed red, green, blue color component. That is the proposed CIKCCA. CIKCCA extracts canonical correlation characteristics of the three color components red, green and blue of the RGB color image, which proved to be effective in color image recognition.

B. Algorithm description

Assume X_R, X_G, X_B represent three color components R,G,B of a color image respectively. CIKCCA is realized as following steps:

Step 1: Calculate n_1 (number of the principal components) principal components $W_R = \{w_1^{pca}, w_2^{pca}, \dots, w_{n_1}^{pca}\}$ by using PCA on R color component.

Step 2: Calculate n_2 (number of the projection vectors) projection vectors $W_G = \{w_1^{lda}, w_2^{lda}, \dots, w_{n_2}^{lda}\}$ by using LDA on G color component.

Step 3: Calculate n_3 (number of the projection vectors) projection vectors $W_B = \{w_1^{msdd}, w_2^{msdd}, \dots, w_{n_3}^{msdd}\}$ by using MSDDA on B color component.

Step 4: Project R, G, B color components to Y_R, Y_G, Y_B space respectively by using W_R, W_G, W_B , that is $Y_R = W_R^T X_R, Y_G = W_G^T X_G, Y_B = W_B^T X_B$.

Step 5: Calculate K_R, K_G, K_B according to the (11),(12),(13).

Step 6: Calculate W_R, W_G, W_B by using CIKCCA K_R, K_G, K_B according to (24).

Step 7: Construct a new data set B

$$B = \begin{bmatrix} W_R^T K_R \\ W_G^T K_G \\ W_B^T K_B \end{bmatrix}, \quad (25)$$

Step 8: Apply the nearest neighbor classifier with the cosine distance to do classification.

IV. EXPERIMENTS

We will testify our algorithm on FRGC-V2 database. There are 12776 training images in FRGC-V2[12], including 16028 controlled images and 8014 uncontrolled images while controlled images have a better image

quality than uncontrolled images. Training sample set consist of both controlled images and uncontrolled images. Training set contains 222 people, each has 36-64 images. We choose 100 people, everyone's 24 images are used to construct the sample set we needed, and every image is preprocessed to 60*60. Fig.1 is an individual's demo images of FRGC-V2.

In order to verify the effectiveness of our method, we will compare the results of CIKCCA and the results of CICCA on FRGC-V2 database. In the experiment, we choose 6 samples among the 24 samples in each class as training samples, the remaining samples as test samples. Each method runs ten times randomly.

In the experiment of CICCA and CIKCCA, we exact the three color components' canonical correlation characteristics of a color image by utilizing PCA, LDA and MSDDA. In PCA, we choose c-1 principal components, in LDA, we choose N-1 discriminant vectors, in MSDDA, we choose N-c discriminant vectors, these components or vectors are used for dimension reduction, where N is the total number of training samples, c is the number of the class.

Table 1 shows comparison of the randomly experiment results between CICCA and CIKCCA, each class has 6 training samples and run 10 times randomly. The experiment result shows that CIKCCA increase the recognition rate by 1.02% compared to CICCA. Through nonlinear transformation, CIKCCA gets a better result than CICCA, which fully proved the superiority of the kernel methods.



Figure 1. Demo images of one individual on FRGC-V2 database

The experiment results are shown as follow:

TABLE I. EXPERIMENT RESULTS OF CICCA AND CIKCCA ON FGRC-V2

Methods	CICCA	CIKCCA
Random Times		
1	88.01	88.02
2	88.52	90.73
3	85.93	87.01
4	79.89	82.14
5	89.21	90.33
6	87.03	87.78
7	88.89	90.54
8	85.32	86.65
9	89.03	89.01
10	83.32	86.04
Average	86.89	88.09

V. CONCLUSION

Color image recognition methods and kernel methods are both significant hot spots in pattern recognition and machine learning. Based on the above two methods, this paper proposed a novel color image recognition method: CIKCCA — color image kernel canonical correlation analysis method. First of all, the three groups of color components are projected into a high dimensional kernel space by utilizing kernel function, and then canonical correlation characteristics are extracted in the high dimensional kernel space and fused in the feature level. At last, the fused vectors are used for classification and recognition. The experiment results shows that our proposed approach has a better result than classical canonical correlation analysis, validating the advantage of kernel method.

ACKNOWLEDGMENT

The work described in this paper was fully supported by the NSFC under Project No. 61073113 and Project No. 61272273, the New Century Excellent Talents of Education Ministry under Project No.NCET-09-0162, the Doctoral Foundation of Education Ministry under Project No. 20093223110001.

REFERENCES

- [1] K. Etemad and R. Chellapa, "Discriminant analysis for recognition of human face images," *Journal of Optical Society of America A*, vol. 14, No. 8, pp. 1724-1733, 1997.
- [2] H. Cevikalp, M. Neamtu, M. Wilkes, A. Barkana, "Discriminative Common Vectors for Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.27, No.1, pp. 4-13, 2005.
- [3] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, and K.R. Müller, "Fisher Discriminant Analysis with Kernels," *Proceeding of IEEE Neural Networks for Signal Processing Workshop*, pp. 41-48, 1999.
- [4] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Computation*, 12, pp. 2385-2404, 2000.
- [5] Yang, M.H., "Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods," *Proc. Fifth IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp. 215-220, 2002.
- [6] P.L. Lai and C. Fyfe, "Kernel and Nonlinear Canonical Correlation Analysis," *Int. J. Neural System*, vol. 10, No. 5, pp. 365-377, 2000.
- [7] H. Hotelling, "Relations between Two Sets of Variables," *Biometrika*, vol. 28, No. 1, pp. 312-377, 1936.
- [8] L. Sun, S.W. Ji, J.P. Ye, "Canonical Correlation Analysis for Multilabel Classification: A Least-Squares Formulation, Extensions, and Analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 33, No. 1, 194-200, 2011.
- [9] T. Melzer, M. Reiter, H. Bischof, "Appearance models based on kernel canonical correlation analysis," *Pattern recognition*, vol.36, No.1, pp. 1961-1971, 2003.
- [10] X.Y. Jing, S. Li, C. Lan, D. Zhang, "Color image canonical correlation analysis for face feature extraction and recognition," *Signal Processing*, vol.91, No.1, pp. 2132-2140, 2011.
- [11] J. Mercer, "Functions of Positive and Negative Type, and their Connection with the Theory of Integral Equations," *Philosophical Transactions of the Royal Society of London. Series A. Containing Papers of a Mathematical or Physical Character*, vol.209, No.1, pp. 415-446, 1909.
- [12] P. J. Phillips, P. J. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, "Overview of The Face Recognition Grand Challenge," *IEEE Conf. Computer Vision and Pattern Recognition*, vol.1, No.1, pp. 947-954, 2005,