Measurement 48 (2014) 128–135

Contents lists available at ScienceDirect

Measurement

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

Kernel common discriminant-based multimodal image sensor data classification

Shu-Po Bu [a], Jiaqing Qiao [b], Jun-Bao Li [b,*]

[a] Department of Electronic and Communication Engineering, Suzhou Institute of Industrial Technology, Suzhou 215104, China
[b] Department of Automatic Test and Control, Harbin Institute of Technology, China

ARTICLE INFO

Article history:
Received 8 July 2013
Received in revised form 9 September 2013
Accepted 21 October 2013
Available online 7 November 2013

Keywords:
Image classification
Feature extraction
Kernel learning

ABSTRACT

Multimodal images (for example, optical image, MR, mammography) are widely used in many practical areas, for example, face recognition, image retrieval, and medical assisted diagnosis. In this paper, we proposed a novel image recognition method of kernel common discriminant based image classification. Firstly, we analyze the limitations of the traditional discriminative common vector (DCV) on the nonlinear feature extraction for image owing to the variations in illuminations. In order to overcome this limitation, we extend DCV with kernel trick with the space isomorphic mapping view in the kernel feature space and develop a two-phase algorithm of KPCA + DCV. The experiments are implemented on WDBC, ORL, YALE, MIAS databases to testify the performance of proposed method.

1. Introduction

Feature extraction and recognition is the most important research topic in pattern recognition area, such as image recognition. Recognition performance of the practical image recognition system is largely influenced by the variations in illumination conditions, viewing directions or poses, facial expression, aging, and disguises. On the feature extraction, the current methods are classified into signal processing and machine learning methods. On the signal processing methods, Gabor wavelets are widely used to extract the facial features for recognition. Gabor wavelets capture the properties of spatial localization, orientation selectivity, and spatial frequency selectivity to cope with the variations in illumination and facial expressions because its kernels are similar to the two-dimensional receptive field profiles of the mammalian cortical simple cells. Gabor feature vector is therefore derived from the Gabor wavelet representation of image images for image recognition. Gabor methods were reported to perform excellently on feature extraction for image recognition in the previous works [1–5]. On the feature extraction based on machine learning, dimensionality reduction is a popular method and widely used in image recognition. Among them, the most popular method are Principal Component Analysis (PCA) [7,8], Independent Component Analysis (ICA) [9,10], Linear Discriminant Analysis (LDA) [6]. Moreover, in recent years, kernel-based nonlinear feature extraction techniques have attracted much attention in the areas of pattern recognition and machine learning [11–13]. Some algorithms using the kernel trick are developed in recent years, such as kernel principal component analysis (KPCA) [14], kernel discriminant analysis (KDA) [15,17] and support vector machine (SVM) [16]. Researchers have developed a series of KDA algorithms [12,18–22]. PCA seeks a linear optimal transformation matrix to minimize the mean squared error criterion, and PCA is generalized to form the nonlinear curves such as principal curves [23] and its extension such as principal images [24]. Principal curves and principal surimages are the nonlinear generalizations of principal components and subspaces respectively. It has turned out that discretized principal curves are essentially equivalent to self-organizing maps (SOM) [25,26]. As an extended SOM, the

* Corresponding author. Tel.: +86 452 86402998.
E-mail address: junbaolihit@gmail.com (J.-B. Li).

0263-2241/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.measurement.2013.10.033

visualization-induced SOM (ViSOM) algorithm directly preserves the distance information on the map along with the topology [27]. ViSOM represents a discrete principal curve or surimage and produces a smooth and graded mesh in the data space and captures the nonlinear manifold of data [28]. Other nonlinear manifold algorithms, such as Isomap [29] and Locally Linear Embedding (LLE) [30], were proposed in recent years. Moreover, Isomap and LLE were well studied on map of the training data but rarely on the test data, while Locality Preserving Projection [31] may be simply applied to any new data point and locate it in the reduced representation space. LPP is still linear and the class information is not used, in order to solve its limitation, the improved Class-wised Locality Preserving Projection is proposed in our previous work [11], and other recent works includes ultrasound image [34], texture analysis [35] and face data [36].

In this paper, we proposed a novel image recognition method of kernel *common discriminant based classification*. Firstly, we analyze the limitations of the traditional discriminative common vector (DCV) [32] on the nonlinear feature extraction for face recognition owing to the variations in poses, illuminations and expressions. In order to overcome this limitation, we extend DCV with kernel trick with the space isomorphic mapping view in the kernel feature space and develop a two-phase algorithm of KPCA + DCV, because the space isomorphic mapping theory is feasible to analyze the data distribution in the mapping space for the classification. With this theory, the distribution is preserved with KDA from the high-dimensional input space to the low-dimensional feature space. It is easy to understand to analyze and improve the performance of KDA for classification. The experiments are implemented on WDBC, ORL, YALE, MIAS databases to testify the performance of proposed method.

2. Method

In this section, we analyze the traditional DCV algorithm from a novel view of isomorphic mapping, and we apply the kernel trick to DCV for the nonlinear image classification. In the practical application, multimedia multi-sensor systems are used in various monitoring system. For example, buses, homes, shopping malls, schools, and so on. These systems are implemented accordingly in an ambient space. Multiple sensors, such as audio and video, are used for identification and ensure safety. The framework of the multimedia multi-sensor smart environment includes many different sensors used to collect the multimedia data. The data is pre-processed or processed with a media processor, and then the features of the sensor data are extracted for event detection. The application framework is shown in Fig. 1.

2.1. Theory derivation

The main idea of discriminant common vector (DCV) is to develop the common properties of all classes through eliminating the differences of the samples within the class. The within-class scatter matrix is created for the common vectors instead of using a given class's own scatter matrix. DCV obtains the common vectors with the subspace methods and the Gram–Schmidt orthogonalization procedure to propose discriminative common vectors. We present the novel formulation of common vector analysis from space isomorphic mapping view as follows. The Fisher discriminant analysis in Hilbert space H based on Fisher criterion function is defined as [32]

$$J(\varphi) = \frac{\varphi^T S_B \varphi}{\varphi^T S_W \varphi} \tag{1}$$

where $S_W = \sum_{i=1}^{C}\sum_{j=1}^{N}\left(x_i^j - m_i\right)\left(x_i^j - m_i\right)^T$ and $S_B = \sum_{i=1}^{C} N(m_i - m)(m_i - m)^T$ are both positive in the Hilbert space H. Sepecially, $\varphi^T S_W \varphi = 0$, the Fisher criterion is transformed as

$$J_b(\varphi) = \varphi^T S_B \varphi, \ (\|\varphi\| = 1) \tag{2}$$

The data become well separable under criterion (2) where $\varphi^T S_W \varphi = 0$, which are analyzed with space isomorphic mapping in the Hilbert space as follows [26]. Supposed a compact and self-adjoint operator on Hilbert space H, then its eigenvector forms an orthonormal basis for H, ($H = \Psi \oplus \Psi^{\perp}$). Then an arbitrary vector φ ($\varphi \in H$) is represented $\varphi = \phi + \zeta$ ($\phi \in \Psi$, $\zeta \in \Psi^{\perp}$), and the mapping $P{:}H \to \Psi$ where $\varphi = \phi + \zeta \to \phi$ and the orthogonal projection ϕ of φ onto H. According to $P{:}H \to \Psi$ with $\varphi = \phi + \zeta \to \phi$, the criterion in (2) is transformed as

$$J_b(\varphi) = J_b(\phi) \tag{3}$$

So, P is a linear operator from H onto its subspace Ψ. Supposed $\alpha_1, \alpha_2, \ldots, \alpha_m$ of S_W, $\Omega_w = span\{\alpha_1, \alpha_2, \ldots, \alpha_q\}$ and $\overline{\Omega}_w = span\{\alpha_{q+1}, \alpha_{q+2}, \ldots, \alpha_m\}$ are the range space and null space of S_W respectively, where $\mathbb{R}^m = \overline{\Omega}_w \oplus \Omega_w$, $q = rank(S_w)$. Since Ω_w and $\overline{\Omega}_w$ are isomorphic to $\mathbb{R}^q$ and $\mathbb{R}^p$ ($p = m - q$) respectively, $P_1 = (\alpha_1, \alpha_2, \ldots, \alpha_q)$ and $P_2 = (\alpha_{q+1}, \alpha_{q+2}, \ldots, \alpha_m)$, the corresponding isomorphic mapping is defined by

$$\varphi = P_2\theta \tag{4}$$

Then criterion in (3) is converted into

$$J_b(\theta) = \theta^T \widehat{S}_b \theta, \ (\|\theta\| = 1) \tag{5}$$

where $\widehat{S}_b = P_2^T S_b P_2$. The stationary points $\mu_1, \ldots, \mu_d$ ($d \leqslant c - 1$) of $J_b(\theta)$ are the orthonormal eigenvectors of $\widehat{S}_b$ corresponding to the d largest eigenvalues. The optimal irregular discriminant vectors with respect to $J_b(\varphi)$, $c\varphi_1, \varphi_2, \ldots, \varphi_d$ are acquired through $\varphi_i = P_2\mu_i (i = 1, \ldots, d)$. So the irregular discriminant feature vector y of the input vector x is obtained by

$$y = (\varphi_1, \varphi_2, \ldots, \varphi_d)^T x \tag{6}$$

Supposed that the stationary points $\mu_1, \ldots, \mu_d (d \leqslant c - 1)$ of $J_b(\theta)$ be the orthonormal eigenvectors of $\widehat{S}_b$ corresponding to the d largest eigenvalues, then

$$\widehat{S}_b \mu_i = \lambda \mu_i \quad i = 1, 2, \ldots, d \tag{7}$$

Then

$$P_2 \widehat{S}_b \mu_i = \lambda P_2 \mu_i \quad i = 1, 2, \ldots, d \tag{8}$$