Engineering Applications of Artificial Intelligence 64 (2017) 340–351

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Robust multimodal multivariate ear recognition using kernel based simultaneous sparse representation

Sayan Banerjee, Amitava Chatterjee *

Jadavpur University, Electrical Engg Deptt, Kolkata, PIN –700032, India

ARTICLE INFO

Keywords:
Multimodal multivariate ear recognition
Joint sparse representation
Joint dynamic
Sparse representation
Kernel robust simultaneous sparse coding

ABSTRACT

In this paper, we propose a novel multivariate multimodal ear recognition method which exploits correlation between left and right ear modality of an individual for his/her identification using joint sparse representation and its variant, joint dynamic sparse representation based classification approach. To make the problem much more robust against outliers that might be resulted from illumination variation or noises due to inaccurate measurements or from partial occlusion due to hair or ornaments — especially for female subjects, we employ a novel weighted multivariate regression scheme under joint sparse as well as joint dynamic sparse penalization. That particular scheme learns a set of weights iteratively for each and every residual corresponding to each observation and subsequently, during the time of classification, gives lesser weight to elements detected as outliers such that they are not able to participate for query set representation. To further improve accuracy of the system, the proposed method is kernelized to tackle non-linearity infusion made by pose variations and occlusions. In the end, extensive experimentations are carried out over a novel database developed in our laboratory to compare performance of the proposed method to several competitive, state-of-the-art methods in order to check suitability of the proposed classification method for various real life applications.

1. Introduction

Biometrics based human identification is receiving extensive amount of interest among researchers for the past few decades for its large-scale applications in security system designing, video based surveillance, human–computer interaction, robotics and more (Jain et al., 2004). Among different biometric modalities such as face, iris, fingerprint, hand shape, gait etc., ear is gaining rapid attention as it has certain advantages (Islam et al., 2008; Yuan et al., 2011; Shekhar et al., 2013) over others, particularly for its robustness with respect to facial expression change and age (Islam et al., 2008; Yuan et al., 2011; Shekhar et al., 2013). But in current scenarios where more and more attempt is being taken to make biometric systems much more user friendly and flexible, unimodal systems are not being able to perform well enough. Therefore in recent years, different multimodal biometric systems with several features are coming into sight as they have sufficient ability to alleviate limitations of unimodal systems through integration of multiple source of information corresponding to multiple modalities.

In this paper we propose a completely novel ear based multimodal multivariate biometrics system which identifies a particular person using both of his ears. Earlier, ear had been fused with several different types of biometric traits such as face (Xu and Mu, 2007a; Yuan et al., 2007; Xu and Mu, 2007b; Chang et al., 2003; Wang et al., 2007; Pan et al., 2008; Yazdanpanah et al., 2010; Abate et al., 2007), finger knuckle (Tharwat et al., 2012), gait (Yazdanpanah et al., 2010), palm-print (Faez et al., 2008), iris (Nadheen and Poomina, 2013), fingerprint (Abate et al., 2007) but no one did ever try to propose a multimodal ear based biometric authentication system based upon only left and right ear of an individual. Motivation behind this proposal primarily lies over three facts. First of all, ears of an individual carry high structural and textual resemblance which may subsequently help to increase robustness of the employed classifier. Secondly images corresponding to left and right ear can be acquired simultaneously using only a single type of sensor therefore data acquisition time and corresponding cost become less. And lastly, multiple measurements or observations of different modalities of an individual may help to boost system's accuracy by advocating spatio-temporal information carried with those observations.

The success of joint sparse representation model for multimodal multivariate biometrics recognition as demonstrated in Shekhar et al. (2013) swayed us to apply the same for our case. The concerned model actually engages same set of training samples across different observations and

* Corresponding author.
E-mail address: cha_ami@yahoo.co.in (A. Chatterjee).

http://dx.doi.org/10.1016/j.engappai.2017.06.011
Received 28 May 2016; Received in revised form 6 March 2017; Accepted 16 June 2017
0952-1976/

as well as modalities for representation of corresponding query samples and therefore, as a result, associated coefficient vectors or rather the coefficient matrix become row-sparse with minimum possible sparsity level to ensure that a small fraction of training samples are participated for every query sample's representation. Till now, JSR has been applied to solve various pattern recognition problems (Chen et al., 2013; Ramezani et al., 2012; Wang et al., 2013; Hu et al., 2014; Panagakis et al., 2014; Zhang et al., 2015; Li et al., 2015; Zhang et al., 2013; Chen et al., 2011; Qi et al., 2013). In Yuan et al. (2012) Yuan et al. introduced multiple feature fusion based visual object recognition using joint sparse representation method. They had also fused different complementary descriptors encoded in form of kernel matrices for object recognition. Yuan et al. (2012). Later Zhang et al. in their paper (Zhang et al., 2012a) had employed JSR for automatic multi-view target recognition using Synthetic Aperture Radar and had obtained satisfactory results. This comparatively a new sparse strategy had also been implemented for different biomedical applications such as recognition of brain activity pattern associated to perceptual and cognitive components (Ramezani et al., 2012), Alzheimer's disease classification using MRI and CET scan (Wang et al., 2013), etc. Very recently it has been applied for multiple object tracking application too (Hu et al., 2014), and for music genre classification (Panagakis et al., 2014). Where in Panagakis et al. (2014), the author along with joint sparse representation charged low rank penalization over the representation matrix so that through the joint effort of these two representations the system can become least susceptible to noise presents in test samples (Panagakis et al., 2014). JSR has been appearing as an effective tool for hyper spectral images classification too where significant researches has been carrying out worldwide (Zhang et al., 2015; Li et al., 2015; Zhang et al., 2013; Chen et al., 2011; Qi et al., 2013).

A general drawback of JSR method is that when test samples become too unconstrained posed then the basic condition for the joint sparse representation which invokes them to share the same set of atoms is become too optimistic. In that context Zhang et al. in Jhang et al. (2012), proposed a relaxed joint sparse penalization method called joint dynamic sparse representation for multi-view, unconstrained posed face recognition where relaxation has been provided at atom level such that atoms are selected individually based upon their degree of coherence to corresponding residuals. The same author had also successfully implemented this method for transients acoustic signal classification in a multiple sensors based measuring environment (Zhang et al., 2012b). Where they further applied proximity prior based joint recovery scheme for signal classification which exploits similarity among signals in true measurement space rather than the sparse measurement space (Zhang et al., 2012b).

Drawing significant observations from Jhang et al. (2012) and Zhang et al. (2012b) we further apply that relaxed joint dynamic sparse representation method in our experiment. But regarding that pose variation problem, it should be carefully noted that whether that variable pose is actually infusing sufficient amount of non-linearity into the system or not. Because, if that happens then it would be even difficult for the joint dynamic sparse representation method to select appropriate atoms for representation of query samples and consequently classification accuracy will be hampered. One obvious solution for the aforementioned problem is to project them into a high dimensional kernel metric space and then involve appropriate kernel for getting subsequent solution (Zhang et al., 0000). Under that consideration we apply KPCA based dimension reduction technique as illustrated in Zhang et al. (0000) to implicitly solve that non-linearity problem and to maintain the system characteristic unperturbed (underdetermined). Furthermore to make the system less sensitive to outliers, we propose a novel robust simultaneous sparse coding approach which is basically a generalized version of the following methods (Yang et al., 2013, 2012, 2011) and has been primarily developed for multivariate regression problem.

The rest of the paper is organized as follows. In Section 2 joint sparse representation based multivariate multimodal biometrics classification problem is described and corresponding solution is proposed using simultaneous orthogonal matching pursuit algorithm. In Section 3 Joint dynamic sparse representation based classification approach and corresponding solution strategy is illustrated. In Section 4, novel robust simultaneous sparse coded joint dynamic sparse representation and joint sparse representation methods are outlined with detailed mathematical formulations. The kernel based solution of the proposed method is further given in Section 5. In Section 6 image acquisition and MULTI-EAR-JU database creation method is described. In Section 7 results of extensive experimentations carried out over MULTI-EAR-JU are reported and in Section 8 a precise conclusion is presented.

2. Joint sparse representation based multimodal biometrics classification scheme

Consider, $\mathbf{D}^h = \left[\mathbf{D}_1^h, \mathbf{D}_2^h, \dots, \mathbf{D}_c^h\right] \in \Re^{m\times n}$, as the dictionary of n number of training images, reshaped into m dimensional vectors referred as atoms, acquired across c different classes under hth modality (it should be noted that for our experiment total number of modalities (nom)) is 2 where $h = 1$ signifies left ear and $h = 2$ signifies right ear. $\boldsymbol{D}_i^h \in \Re^{m\times n_i}$ denotes the dictionary associated to ith class. Consider $\boldsymbol{Y} = \left[\boldsymbol{Y}^1, \boldsymbol{Y}^2, \dots, \boldsymbol{Y}^{nom}\right]$ as the joint query set where each query set $\boldsymbol{Y}^h = \left[\boldsymbol{y}_1^h, \boldsymbol{y}_2^h, \dots, \boldsymbol{y}_p^h\right] \in \Re^{m\times p}$ is holding p observations of the test subject. In this context it should be mentioned that all of the observations of a particular query subject however acquired across different modalities but they belong to the same class therefore ideally they should share the same class-level sparsity pattern (Shekhar et al., 2013; Zhang et al., 2012a). Furthermore, due to presence of considerable amount of correlations, they should share the same atom-level sparsity pattern (Shekhar et al., 2013; Zhang et al., 2012a) too as shown in Fig. 1. But depending upon the problem formulation, degrees of contribution of the fixed set of participating atoms for the reconstruction of different observations across different modalities may be distinct. Therefore it is reasonable to enforce joint sparse penalization (Shekhar et al., 2013; Yuan et al., 2012; Zhang et al., 2012a; Chen et al., 2013) over the set of coefficient matrices $\boldsymbol{X} = \left[\boldsymbol{X}^1, \boldsymbol{X}^2, \dots, \boldsymbol{X}^{nom}\right]$ such that they along with the set of multimodal dictionaries can efficiently represent the joint query set.

The joint sparse representation model, like (Zhang et al., 2012a; Chen et al., 2011), tries to recover $\hat{\boldsymbol{X}}$, the set of desired coefficient matrices, through the following constrained optimization problem

$$\hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \sum_{h=1}^{nom} \left\| \boldsymbol{Y}^h - \boldsymbol{D}^h \boldsymbol{X}^h \right\|_F^2$$
$$\text{such that } \|\boldsymbol{X}\|_{l_0/l_2} \le S \tag{1}$$

where $\|\boldsymbol{X}\|_{l_0/l_2}$ is called l_0/l_2 mixed-norm of the representation set $\boldsymbol{X}$ (Zhang et al., 2012a) which can be calculated by first applying l_2-norm on each row of $\boldsymbol{X}$ and then determining the number of non-zero elements of the resultant vector (Zhang et al., 2012a). S indicates the number of non-zero rows that are present in $\hat{\boldsymbol{X}}$ which further suggests that maximum of S number of atoms across each modality and under each observation should simultaneously take part for reconstruction of $\boldsymbol{Y}$.

Now we can see that the problem in (1) is NP-hard hence it should be solved by some greedy algorithms (Chen et al., 2011).

In our work we have adopted Simultaneous Orthogonal Matching Pursuit (SOMP) method like (Chen et al., 2011) for approximately solving Eq. (1) due to its accuracy over other similar greedy methods and it is for the first time, to the best of our knowledge, SOMP is applied to solve multimodal biometrics recognition problem.

In SOMP, atoms across different modalities are chosen based on their correlations with the corresponding residual and it should be noted that only one atom is chosen at a time (Chen et al., 2011). The SOMP method terminates its operation when the average residual value