Engineering Applications of Artificial Intelligence 55 (2016) 37–46

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Image set based ear recognition using novel dictionary learning and classification scheme

Sayan Banerjee, Amitava Chatterjee*

Electrical Engineering Department, Jadavpur University, Kolkata 700032, India

ARTICLE INFO

Article history:
Received 4 September 2015
Received in revised form
29 February 2016
Accepted 9 May 2016

Keywords:
Ear recognition
Force field transformation
Metaface learning
Reduced time metaface learning
l_2-norm regularized kernel affine hull based collaborative representation based classification (L2RKAH-CRC)

ABSTRACT

In this work ear recognition of a moving person with the help of a single fixed-in-position video camera is investigated, a novel problem undertaken to the best of our knowledge and belief. The challenges associated with this work are that during data capturing process each moving subject appears in unconstrained views together with random noise, motion blur and different level of illumination. Secondly, the collection of large number of redundant samples for each subject makes the training database bulky, which eventually increases classification time. So, in this paper, we introduce a novel, reduced time, metaface dictionary learning approach which by employing Frobenius norm based regularization reduces redundancy of training database with much lesser time as compared to other available database learning methods. In general, ear samples with random poses, severe motion blur and different levels of illumination loose significant class specific information and hence inflict severe nonlinearity to the system. The simple and well accounted solution for the above problem is kernel framework which makes samples of different classes linearly separable by elevating them to higher dimensions. In our proposed solution, we have used novel l_2-norm regularized affine hull based kernel collaborative representation based classification scheme, which represents each query set as an affine hull and then collaboratively represents this hull over the linear span of gallery sets of all classes in the high dimensional space. Finally, the query set is assigned to that particular class which gives least representation or residual error among all the available classes. Results of extensive experimentations carried out over an indigenously developed database in our laboratory (named ERVIDJU) aptly demonstrate that our proposed method is a valid biometric identifier and it can produce superior performance compared to several other contemporary algorithms, developed for similar purposes.

1. Introduction

In the last few year, ear has been becoming as a very interesting means for biometric authentication and several fascinating research and advancements have been made in that regard. The main reasons behind its popularity over other biometric modalities are truly considerable and therefore we illustrate some of its features next. First, shape of the ear remains almost unaffected with changes in facial expression. Second, ear extraction from the redundant background becomes quite easier as it remains fixed over human side face lifelong. Furthermore, it has a surprising quality to retain its shape intact throughout its entire lifetime (Yuan et al., 2011; Islam et al., 2008; Hurley et al., 2007; Ross and Abaza, 2011) and the possibility of capturing ear image from a distance also validates its stand as an effective contributor for video-based surveillance systems (Ross and Abaza, 2011).

On the other hand, person authentication from a video is also a current area of research due to its wide application of access control, video surveillance and so on. But here we would like to mention thattill now video based face recognition, using image set-based formulation, has been only exploited and investigated so far (Wang et al., 2008; Huang et al., 2013; Shakhnarovich et al., 2002; Zhu et al., 2014; Cevikalp and Triggs, 2010; Yang et al., 2013; Lee et al., 2003; Arandjelovic et al., 2005; Hu et al., 2004). Therefore, examining all those advantages of ear over face mentioned, we propose a novel biometric system where right side ear of different moving person are considered for their authentication. As one can find that those ear image samples contain less-informative background (i.e. presence of hairs and some portions of throats) which cannot be effectively eliminated even after employing the best possible manual tuning or cropping. Hence, the utilization of a feature extraction module as a pre-processing step becomes necessary. In this paper, we apply a recently developed feature extraction scheme that is force field transform for this purpose. Force field transform was originally proposed by Hurley et al. (2002a,

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

http://dx.doi.org/10.1016/j.engappai.2016.05.005

2002b, 2000, 2005) and was primarily developed for the purpose of ear feature extraction where a force field is created around each pixel with the collective effort of surrounding pixels (Hurley et al., 2002a, 2002b, 2005, 2000; Banerjee and Chatterjee, 2015). The transform ultimately replaces intensity of each pixel by that force field. This force field is a vector quantity so pixels having identical neighbors will be represented by null (Hurley et al., 2002a, 2002b, 2000, 2005; Banerjee and Chatterjee, 2015). So the effect of an approximately smooth background around any ear sample can be significantly eliminated by applying aforesaid transform.

Again, one can see that there is another problem with the system. As each subject was observed for a definite time during data acquisition process, hence an ample number of samples are available for each subject in the training dataset, where many of those images bear similar properties. Moreover, the associated problem with the redundancy grows in leaps in bounds with the improvement in the acquisition quality of the video camera (i.e. with possibility of capturing higher number of frames per second). Therefore, the problem can potentially be very hazardous with excessive growth in the dimension of the data. To overcome this problem, a novel, dictionary-learning algorithm is proposed which effectively picks up significant samples from that oversized dataset and learns a more compact dictionary. A Frobenius norm based regularization enforced upon representation matrix makes the process less time consuming as compared to other state-of-the-art learning methods and therefore facilitates its use where continuous dynamic learning is required.

From the classification point of view, it is evident that the algorithm involves a set-to-set based operation. Here, a query set is a collection of randomly posed test samples generated from a video sequence captured over a subject of interest and it is checked with every learned gallery set corresponding to each subject. An l_2-norm regularized kernel affine hull based collaborative representation based classifier (named L2RKAH-CRC)has been proposed that first deals with non-linearity and next it creates affine hull from the query set. Subsequently, it collaboratively represents this hull over each gallery set of every class and assigns it to that particular class label which produces minimum representation error. A weighted l_2 lllnorm based regularization over representation vectors of query as well as the whole gallery makes the classifier more stable and reflects factor of collaboration among intra class query samples in addition with inter class gallery samples to classify a particular query set.

The rest of the paper is organized as follows. In Section 2 a detailed literature survey is presented. Section 3 discusses the mathematical formulation of force field transform whereas, in Section 4, basic metaface learning along with our proposed learning approach has been presented in detail. Section 5 discusses the l_2 norm regularized kernel affine hull based collaborative representation basedclassifier. In Section 6 image acquisition and creation of the ERVIDJU database method is described. In Section 7 extensive performance evaluation carried out using our ERVIDJU database is reported. The conclusions are presented in Section 8.

2. Literature survey

2.1. Force field feature extraction technique

It was first proposed by Hurley et al. in the year of 1999 (Hurley et al., 2002a) where they had shown that an entire image under consideration could be transformed into a force field image by assuming that each pixel applies a force to every other pixel which is proportional to pixel"s intensity and inversely proportional to square of the distance. They had used small potential energy wells as ear features which are basically clustering or converging points of force field lines and suggest local energy peaks in the scalar potential energy surface (Hurley et al., 2002a, 2002b, 2000). They had demonstrated that the potential energy matrix is invertible and, therefore, the original image can be fully recovered from the potential energy surface. They had also shown that the relative positions of those wells are invariant to scaling and rotation of an ear and these wells are quite robust to random noise (Hurley et al., 2002a, 2002b, 2000). Hurley et al. (2000) had applied this particular feature extraction technique over human face samples for the first time and obtained satisfactory results. Hurley et al. (2002b) presented more detailed discussions and mathematical explanations of the corresponding feature extraction method. Later, Dong and Mu employed this force field transform technique in conjunction with Nonlinear Kernel Fisher Discriminant Analysis (NKFDA) for multi-pose ear feature extraction (Hurley et al., 2005). They actually took whole force field transformed ear as samples rather than potential wells for their work (Hurley et al., 2005).

Here it should be mentioned that we got on an average 7 features (irrespective of the number of randomly initialized force field lines) over different ear samples using the potential energy well concept as mentioned above. Moreover, ear samples for our problem were itself of small in size as it were captured from a distance. Here specifically we would like to mention, that our data are severely corrupted by motion blur, random noise and it has some illumination variations, which degraded the essential quality of those features. Furthermore, the number of features are insufficient. As a consequence, the system becomes more underdetermined and as a whole performance of the classifier is compromised. Henceforth, we considered whole force field transformed ear samples for experimentation.

2.2. Dictionary learning methods

Dictionary learning methods were formulated primarily with the aim of learning a dictionary from the original training samples such that it can properly represent the samples. General Sparse Representation based Classification (SRC) and Collaborative Representation based Classification (CRC) techniques create a dictionary from the original samples without any modification (Wright et al., 2009; Zhang et al., 2011, 2012b) hence performance of those algorithms directly depends upon availability of samples instead of the dictionary. In 2006, Aharon et al. proposed KSVD based dictionary learning algorithm which can create an overcomplete dictionary (Aharon et al., 2006) and therefore promotes application of sparse representation based classification framework over the problem. But due to unsupervised nature of the learning method, this particular algorithm suffers from reduced classification accuracy. Supervised dictionary learning methods (Zhang and Li, 2010; Jiang et al., 2013) on the other hand can take advantages of readymade class label information and therefore can make the learned dictionary along with the coding coefficients much more discriminative to improve further classification results. Exploiting the above fact, Zhang et.al. in the year of 2010 proposed discriminative KSVD based dictionary learning and utilized discriminative capability of the learned dictionary over face recognition problem (Zhang and Li, 2010). In the year of 2013, Jiang et al. proposed another variant of KSVD, named label consistent KSVD (Jiang et al., 2013) which infuses discriminative quality over general KSVD based learned dictionary effectively.

Shared dictionary learning under supervised learning method is comparatively a new learning approach which was developed formally to learn a dictionary from highly correlated training samples such that each learned atom with class specific information additionally have information of the rest of the classes to some extent. Due to that prescribed nature the learned dictionary becomes quite precise (Zhang and Li, 2010).

The above paragraph show cased class specific dictionary learning where class representation residual is exploited for