Pattern Recognition Letters 53 (2015) 69–76

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

An adaptive bimodal recognition framework using sparse coding for face and ear ☆

Zengxi Huang [a], Yiguang Liu [b,*], Xuwei Li [b], Jie Li [b]

[a] School of Mathematics and Computer Engineering, Xihua University, Chengdu 610039, PR China
[b] Vision and Image Processing Laboratory, College of Computer Science, Sichuan University, Chengdu 610065, PR China

ARTICLE INFO

Article history:
Received 2 January 2014
Available online 22 October 2014

Keywords:
Multimodal recognition
Sparse coding
Biometric quality assessment
Adaptive feature fusion

ABSTRACT

In this paper, we propose an adaptive face and ear based bimodal recognition framework using sparse coding, namely ABSRC, which can effectively reduce the adverse effect of degraded modality. A unified and reliable biometric quality measure based on sparse coding is presented for both face and ear, which relies on the collaborative representation by all classes. For adaptive feature fusion, a flexible piecewise function is carefully designed to select feature weights based on their qualities. ABSRC utilizes a two-phase sparse coding strategy. At first, face and ear features are separately coded on their associated dictionaries for individual quality assessments. Secondly, the weighted features are concatenated to form a unique feature vector, which is then coded and classified in multimodal feature space. Experiments demonstrate that ABSRC achieves quite encouraging robustness against image degeneration, and outperforms many up-to-date methods. Very impressively, even when query sample of one modality is extremely degraded by random pixel corruption, illumination variation, etc., ABSRC can still get performance comparable to the unimodal recognition based on the other modality.

1. Introduction

Face recognition (FR) has received a great amount of attention and been improved significantly [1,2]. However, variations like pose, expression, illumination and occlusion are still challenges to FR, which is hence not yet as accurate and flexible as desired. On the other hand, the ear is considered as a unique characteristic of an individual and can be used for biometric recognition, which is non-intrusive as the face. Although occlusion due to hair or jewelry could pose difficulty to ear recognition (ER), the ear has several appealing advantages over the face: it has a stable structure with rich information, nearly unaffected by aging and facial expressions [3–8]. Besides, many feature extraction and classification techniques originally developed for the face are applicable to ER. Therefore, a multimodal biometric system based on face and ear is feasible and can address several limitations of face and ear unimodal systems [1, 4–7].

However, it must be noticed that image degeneration of either face or ear could possibly degrade the multimodal recognition performance. Particularly, when one modality confronts severe data degeneration, multimodal system may perform worse than the unimodal system using the other modality. Given the relative independence between the face and the ear, the key to robust multimodal biometric is in effective biometric quality-based adaptive fusion, which is necessary to assign lower weight to the less reliable modality while assign higher weight to the good one [1,9]. Biometric quality-based fusion involves biometric quality assessment and dynamic weight selection.

Nevertheless, biometric quality assessment remains an open issue to most biometric systems [10,11]. To our knowledge, there was not yet an effective and unified biometric quality measure suitable for both face and ear until in [6] we found that sparse coding error can be a favorable quality measure. Thereby, we have developed a quality-based multimodal approach called multimodal sparse representation-based classification (SRC) with feature weighting (MSRCW) [6]. Its promising robustness to data degeneration has been confirmed. However, MSRCW directly uses the sparse coding coefficients of multimodal feature to estimate the face and ear image qualities such that the quality assessment may not be accurate. Besides, MSRCW's logistic function-based feature weight selection is not flexible enough to enable it to make a good balance between utilizing discriminative information and reducing adverse effect of the less reliable modality. Aiming to alleviate these limitations, in this paper we attempt to develop a more general multimodal recognition framework called adaptive bimodal SRC (ABSRC). ABSRC is different from MSRCW in three aspects. First, ABSRC evaluates face and ear qualities based on their individual sparse coding results, which has

☆ This paper has been recommended for acceptance by Ajay Kumar.

* Corresponding author. Tel.: +86 028 85469688; fax: +86 028 85469688.
E-mail address: lygpapers@aliyun.com (Yiguang Liu).

http://dx.doi.org/10.1016/j.patrec.2014.10.009
0167-8655/

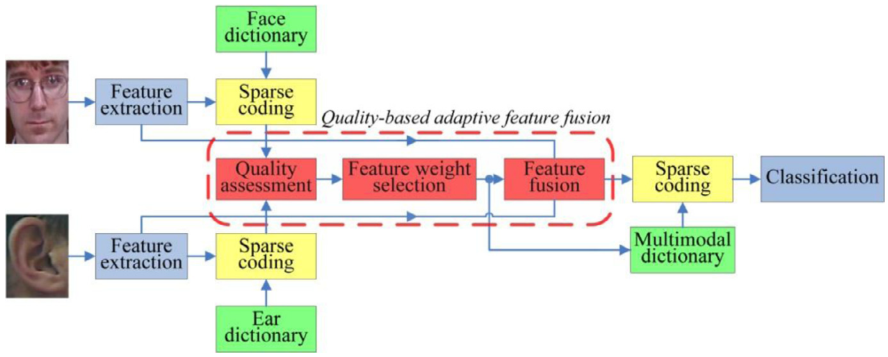

Fig. 1. The block diagram of ABSRC.

been experimentally confirmed to be more precise. Second, a piecewise feature weight function is elaborately designed for better coping with data degeneration. Last, from a utility perspective, the biometric quality measure is newly defined only using a certain part of coding residual because corruptions like isolated/impulsive noises and small-region occlusion often cause large coding residual but lead to a relatively small impact on biometric recognition. Furthermore, we provide more illustrations about the sparsity-based quality assessment and more experiments and discussions.

Experiments against various variations show that ABSRC is obviously superior to MSRCW, multimodal SRC (MSRC) [6], and a multimodal extension of WGSR (MWGSR) [12]. ABSRC's advantage over MSRCW is especially significant in illumination experiments. Benefiting from the adaptive feature fusion, ABSRC achieves very promising robustness against data degeneration. Even when query image of one modality is extremely degraded by random pixel corruption,[1] illumination variation or face disguise, ABSRC can still obtain performance comparable to the unimodal recognition using the other modality.

The rest of this paper is organized as follows. We briefly review the related work in next section. In Section 3, our multimodal biometric method is described in details. Section 4 conducts experiments to evaluate the proposed ABSRC. Finally, conclusions are summarized in Section 5.

2. Prior work

In multimodal biometric systems, evidences can be fused at sensor, feature, match score, and decision levels [1]. Compared with other levels, fusion at feature level can exploit the most discriminative information while eliminate the redundant/adverse information from the original biometric data. Feature level fusion is particularly popular in multimodal biometric involving face and ear. Xu and Mu [13] reported a system integrating ear and profile face at feature level, which integrates the principle component analysis-based (PCA) features of face and ear by means of canonical correlation analysis (CCA). They achieved an accuracy of 97.37% on a subset of USTB ear database with 38 subjects. Further, they developed a method using kernel CCA-based (KCCA) feature fusion, which got an improved accuracy of 98.68% on the same dataset [14]. Abate et al. [4] applied iterated function systems (IFS) transformation to characterize the self-similarity of a face and/or ear image, then concatenated their features, described by a list of centroids, to form an overall feature vector. MSRC [6] combines the PCA features of face and ear with serial concatenation and employs SRC for multimodal classification, which was reported with much better performance than Xu's CCA-based methods.

The aforementioned methods are evidently better than the unimodal recognition using either face or ear alone in their reports. However, due to the usage of static fusion rules, they are sensitive to face and ear image degeneration, which has been clearly revealed in [6]. On the contrary, Huang et al. [6] proposed biometric quality-based feature fusion methods based on SRC [15] and RSC [16] models, namely MSRCW and MRSCW, respectively. They are quite robust to the image degeneration of individual modality, and meanwhile MRSCW outperforms MSRCW owing to its outlier (e.g., noise, corruption, occlusion etc.) detection skill RSC. However, besides sharing MSRCW's limitations uncovered above, MRSCW is computationally expensive because it has to re-calculate the multimodal feature dictionary and perform sparse coding in each iteration. Hence, many more robust but relatively time-consuming feature extraction algorithms like SIFT [17] and Log-Gabor filters [18,19] could not be adopted. Furthermore, MRSCW needs to know preliminarily the percentage of corrupted pixels in query image, which seems hard to be achieved in reality.

3. Adaptive bimodal SRC

SRC firstly encodes the query data with a training sample dictionary and then classifies it to the class which yields the least square coding error. Let $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_c]$ be the matrix formed by training samples of c classes, where $\mathbf{A}_i$ is the subset of class i. Let $\mathbf{y}$ be a query sample. In SRC [15], first $\mathbf{y}$ is sparsely coded on $\mathbf{A}$ via l_1-norm minimization

$$\hat{\alpha} = \arg\min \|\alpha\|_1 \ \text{s.t.} \ \|\mathbf{y} - \mathbf{A}\alpha\|_2 < \varepsilon, \tag{1}$$

where $\varepsilon > 0$ is a constant. Then classification is made by $g(\mathbf{y}) = \arg\min_i\{\gamma_i\}$, where $\gamma_i = \|\mathbf{y} - \mathbf{A}_i\hat{\alpha}_i\|_2$, $\hat{\alpha} = [\hat{\alpha}_1; \hat{\alpha}_2; \cdots; \hat{\alpha}_c]$, and $\hat{\alpha}_i$ is the coefficient vector associated with class i.

The proposed ABSRC employs a two-phase sparse coding strategy including separate sparse coding and joint sparse coding. As the block diagram plotted in Fig. 1, firstly, the face and ear features are separately coded on their corresponding dictionaries. Then the feature weights for fusion are calculated dynamically by using an elaborately designed piecewise function based on the sparsity-based biometric qualities. Secondly, ABSRC serially concatenates the weighted features to form a unique multimodal feature vector, which is then classified in the multimodal feature space by using SRC.

3.1. Feature extraction

This paper mainly focuses on biometric quality assessment and adaptive feature fusion of face and ear, also concerns about the generalization ability of the proposed adaptive bimodal framework in dealing with data degeneration. Hence, a simple, time-saving and general feature extraction algorithm like PCA [20], whose applications in FR and ER are known as "Eigenfaces" and "Eigenears" [7], is utilized, consistent with MSRCW [6]. Given c classes, let $\mathbf{A}^f = [\mathbf{A}^f_1, \mathbf{A}^f_2, \cdots, \mathbf{A}^f_c]$ and $\mathbf{A}^e = [\mathbf{A}^e_1, \mathbf{A}^e_2, \cdots, \mathbf{A}^e_c]$ separately denote the face and ear sample sets

[1] Random pixel corruption [15]: Replaces a certain percentage of image pixels by uniformly distributed random values within [0, 255].