Neurocomputing 171 (2016) 515–523

Contents lists available at ScienceDirect

Neurocomputing

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

A multi-task model for simultaneous face identification and facial expression recognition

Hao Zheng [a,b,c,d], Xin Geng [a,*], Dacheng Tao [e], Zhong Jin [c]

[a] MOE Key Laboratory of Computer Network and Information Integration, Southeast University, Nanjing, China
[b] Key Laboratory of Trusted Cloud Computing and Big Data Analysis, Nanjing XiaoZhuang University, Nanjing, China
[c] Jiangsu Key Laboratory of Image and Video Understanding for Social Safety, Nanjing University of Science and Technology, Nanjing, China
[d] State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China
[e] Centre for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia

ARTICLE INFO

Article history:
Received 12 February 2015
Received in revised form
22 June 2015
Accepted 29 June 2015
Communicated by Qingshan Liu
Available online 20 July 2015

Keywords:
Face identification
Facial expression recognition
Multi-task learning

ABSTRACT

Regarded as two independent tasks, both face identification and facial expression recognition perform poorly given small size training sets. To address this problem, we propose a multi-task facial inference model (MT-FIM) for simultaneous face identification and facial expression recognition. In particular, face identification and facial expression recognition are learnt simultaneously by extracting and utilizing appropriate shared information across them in the framework of multi-task learning, in which the shared information refers to the parameter controlling the sparsity. MT-FIM simultaneously minimizes the within-class scatter and maximizes the distance between different classes to enable the robust performance of each individual task. We conduct comprehensive experiments on three face image databases. The experimental results show that our algorithm outperforms the state-of-the-art algorithms.

1. Introduction

Face recognition is still a very active research area in public security, human–computer interaction, financial security, etc. The two primary face recognition tasks are identification and verification. The goal of face identification task is to identify a person based on the face image, i.e., the captured face image needs to be matched to a gallery of known people. Therefore, addressing face identification task is the key issue for face recognition. Though typical applications of face recognition have been used in many years, many real world situations are still a challenge due to illuminations, pose, and occlusions [1]. At the same time, facial expression recognition is also an important topic due to its wide range of applications [2–5]. Most facial expression recognition methods aim to recognize a set of prototypic expressions (i.e. surprise, anger, joy, sadness, fear, and disgust). Though much progress has been made [6–10], the performance of facial expression recognition is still unsatisfactory due to the subtlety, complexity and variability of facial expressions.

For face recognition, many algorithms about classifiers have been proposed. The nearest- neighbor (NN) algorithm is simple and it is accurate and applicable to various problems. But the shortcoming of NN algorithm is that only one training sample is used to represent the test face image, so the nearest feature line classifier [11] was proposed through using two training samples for each class to represent the test face image. Then the nearest feature plane classifier [12] was proposed through using three samples to represent the test image. Later, for representing the test image by all the training samples of each class, the local subspace classifier [13] and the nearest subspace classifier [14,15] were proposed. Because samples from a specific object class are known to lie on a linear subspace, a linear regression classification (LRC) [16] algorithm was proposed by formulating the pattern recognition problem in terms of linear regression. Another well-known classifier is support vector machine (SVM) classifier [17], which is solidly based on the theory of structural risk minimization in statistical learning. It is well known that the SVM can efficiently perform a non-linear classification and map the inputs to a high-dimensional feature space, then find a large margin hyperplane between the two classes which can be solved through the quadratic programming algorithm. However, SVM cannot be applied when the vectors defining out samples have missing entries. It can be seen when occlusions are present for face recognition. Fortunately sparse representation based classification (SRC) was reported by Wright et al. [18] for robust face recognition. In Wright et al.'s pioneer work, the training face images are used as the dictionary of representative samples, and an input test image is coded as a sparse linear combination of these sample images via

* Corresponding author.
E-mail address: xgeng@seu.edu.cn (X. Geng).

http://dx.doi.org/10.1016/j.neucom.2015.06.079
0925-2312/

l_1-norm minimization. The experimental results in [18,19] of SRC were exciting in FR, which could lead to high classification accuracy, especially well handling the problems of face occlusion and corruption. Furthermore, many extended methods were proposed [20–22], e.g. Gabor SRC [20], Heteroscedastic SRC [21], and SRC for continuous occlusion [22].

For facial expression recognition, many previous works are based on statistical learning. Ekman [23] developed the Facial Action Coding System (FACS) in which the movements on the facial expression are described by action units. Many extension methods [24,25] were also proposed, and various classifiers were applied for facial expression recognition. In [26], after the facial features are extracted and represented, an artificial neural network (ANN) was employed to recognize the action units. In order to recognize the naturalistic affective expressions, Meng and Bianchi-Berthouze [27] adopted Hidden Markov Models (HMM) to model this spontaneous process and finalize the classification process. Then Cohen et al. [28] proposed a new architecture of HMM for automatically segmenting and recognizing human facial expressions. A SVM classifier [29] that incorporates statistical information of the classes under examination was also proposed and used to recognize either the six basic facial expressions or a set of Facial Action Units. In [30], popular machine learning methods including SVM were compared for facial expression recognition and SVM was proved to be the most effective classifier.

However, in real-world applications for face identification and facial expression recognition, the number of training samples from each class is usually limited. Unfortunately, the performances of face identification and facial expression recognition by the above methods usually degrade significantly with the decrease of training samples. Therefore the small sample size problem becomes one of the most prominent issues in face identification and facial expression recognition. In fact, the goal of most face identification methods is to find a similarity measure invariant to illumination changes, pose, and facial expressions, so that images of faces can be recognized in spite of existence of these variation. While the goal of expression recognition is to find a model for non-rigid patterns facial expression, so that facial expression can be classified in spite of a wide range of variation. Therefore, information about facial appearance and expression patterns can be jointed to carry face identification and facial expression at the same time [31,32]. Both face identification task and facial expression recognition task aim to learn universal model which make the tasks are robust for various situation. In general, multi-task learning can be adopted to join face identification and facial expression recognition by extracting the appropriate shared information. Especially multi-task learning has been shown, both empirically and theoretically, to be able to significantly improve the performance of learning each task separately. In this paper, motivated by the above idea, we propose a multi-task facial inference model (MT-FIM) for simultaneous face identification and facial expression recognition. In MT-FIM, the classifiers of face identification and facial expression recognition are learned at the same time by extracting and utilizing appropriate shared information across them. More importantly, face identification and facial expression recognition can benefit from each other, resulting in better performance. In addition, we minimize the within-class distance and maximize the distance between different classes. Because MT-FIM is convex, optimization of MT-FIM is employed. In conclusion, the contributions of our work are three-fold.

1. We propose a MT-FIM method for face identification and facial expression recognition, which is a multi-task method that joints two related tasks. To the best of our knowledge, this is the first work to simultaneously learning the face identification and facial expression recognition.
2. Since MT-FIM adopts the multi-task and increases the samples size for each recognition task, the sample size problem can be solved efficiently. MT-FIM achieves better recognition results than the state-of-the-art methods.
3. In MT-FIM, we introduce the within-class covariance which makes the same class samples easier to cluster. MT-FIM strikes a balance between minimize the within-class distance and maximize the distance between different classes.

The rest of this paper is organized as follows. Section 2.1 reviews the technology of multi-task learning. Section 2.2 presents our multi-task facial inference model. Section 2.3 describes optimization process of MT-FIM, and then complexity of MT-FIM is discussed in Section 2.4. Finally, Section 3 conducts experiments to validate the proposed model and Section 4 concludes the paper.

2. Multi-task facial inference model (MT-FIM)

In this section, firstly we discuss the related multi-task learning. Then we present the proposed MT-FIM algorithm. Finally optimization of the algorithm is stated.

2.1. Multi-task learning

Multi-task learning has continuously received increasing attention in computer vision and image recognition. It aims to simultaneously learn multiple related tasks and utilizing the shared information among related tasks for improving performance of each task. In the past years many multi-task learning methods have been studied, existing methods can be broadly classified into two categories: implicit structure sharing and explicit parameter sharing. Methods under implicit structure sharing implicitly capture some common structures; for example, the methods in [33,34] constrain all tasks to share a common low rank subspace and the methods in [35,36] constraint all tasks to share a common set of features. While methods under explicit parameter sharing explicitly share some common parameters; examples include hidden units in neural networks [37,38], prior in hierarchical Bayesian models [39,40], parameters of Gaussian process [41], classification weight [42], feature mapping matrix [43], and similarity metric [44,45]. In two categories, methods under explicit parameter sharing were often adopted, and multi-task sparse learning was employed by learning multiple classifiers from different tasks to share similar parameter sparsity patterns. For increasing the training samples size, multi-task learning can effectively solve the problem of the small sample size. Especially the Lasso regularized methods [46,47] are widely used to image recognition for the simplicity and effectiveness. The Lasso method in MTL is a penalized least square method imposing a l_1-norm penalty on the regression coefficients and the parameter controlling the sparsity is shared by all related tasks. A general model is illustrated as below:

$$\min_{W} \sum_{i=1}^{t} \Gamma(W) + \rho \|W\|_1, \tag{1}$$

where W is classifier to be estimated from the training samples, $\Gamma(W)$ is the empirical loss on the training set, $\|W\|_1$ is the regularization term that encodes task, and t is the number of tasks. In the model, ρ is the regularization parameter for controlling sparsity.

2.2. Multi-task facial inference model (MT-FIM)

But in the multi-task model within-class discriminative information is not considered, so the multi-task learning sometimes