Pattern Recognition 60 (2016) 890–900

Contents lists available at ScienceDirect

Pattern Recognition

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

Facial action unit recognition under incomplete data based on multi-label learning with missing labels

Yongqiang Li [a,c], Baoyuan Wu [b,*], Bernard Ghanem [b], Yongping Zhao [a], Hongxun Yao [c], Qiang Ji [d]

[a] School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin 15001, China
[b] The Visual Computing Center, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia
[c] School of Computer Science and Technology, Harbin Institute of Technology, Harbin 15001, China
[d] Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

ARTICLE INFO

Article history:
Received 6 April 2016
Received in revised form
10 June 2016
Accepted 4 July 2016
Available online 7 July 2016

Keywords:
Face action unit recognition
Incomplete data
Multi-label learning

ABSTRACT

Facial action unit (AU) recognition has been applied in a wild range of fields, and has attracted great attention in the past two decades. Most existing works on AU recognition assumed that the complete label assignment for each training image is available, which is often not the case in practice. Labeling AU is expensive and time consuming process. Moreover, due to the AU ambiguity and subjective difference, some AUs are difficult to label reliably and confidently. Many AU recognition works try to train the classifier for each AU independently, which is of high computation cost and ignores the dependency among different AUs. In this work, we formulate AU recognition under incomplete data as a multi-label learning with missing labels (MLML) problem. Most existing MLML methods usually employ the same features for all classes. However, we find this setting is unreasonable in AU recognition, as the occurrence of different AUs produce changes of skin surface displacement or face appearance in different face regions. If using the shared features for all AUs, much noise will be involved due to the occurrence of other AUs. Consequently, the changes of the specific AUs cannot be clearly highlighted, leading to the performance degradation. Instead, we propose to extract the most discriminative features for each AU individually, which are learned by the supervised learning method. The learned features are further embedded into the instance-level label smoothness term of our model, which also includes the label consistency and the class-level label smoothness. Both a global solution using st-cut and an approximated solution using conjugate gradient (CG) descent are provided. Experiments on both posed and spontaneous facial expression databases demonstrate the superiority of the proposed method in comparison with several state-of-the-art works.

1. Introduction

Facial expression is one of the most natural nonverbal communication media that individuals use to regulate interactions with each other. Expressions can express the emotions, clarify and emphasize what is being said, and signal comprehension, disagreement and intentions [7]. Machine understanding of facial expressions will provide powerful information to describe the emotional states and psychological patterns of individuals. Due to the huge potential in many applications, including social robotics, affective online tutoring environment, intelligent Human–Computer interaction (HCI), etc., automatic facial expression recognition has recently gained great attention and become a hot topic [2,3].

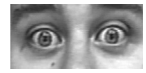
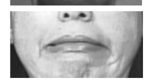
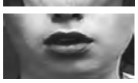
One of the most widely studied expression descriptors is the six basic expressions named anger, fear, disgust, happiness, sadness and surprise, which are universal and unrelated with race and culture [4]. However, these six basic expressions only represent a small set of human facial expressions. In fact, human emotion is composed of thousands of expressions, though most of them differ in subtle changes of a few facial features. Facial Action Coding System (FACS) developed by Ekman [5] has been demonstrated as a powerful means for representing and characterizing a large number of facial expressions through the combination of only a small set of action units (AUs). According to FACS, each AU is related to the contraction of a specific set of facial muscles. FACS defines 32 AUs, i.e., 9 AUs in the upper face, 18 in the lower face and 5 AUs that cannot be partitioned as belonging to either the upper or the lower face [7]. The readers are referred to **[5]** for detailed definition and explanation of all AUs. We list the 16 AUs we recognize in this work in Table 1.[1] The aim of AU recognition is

* Corresponding author.
E-mail address: wubaoyuan1987@gmail.com (B. Wu).

[1] Readers are referred to: http://www.cs.cmu.edu/face/facs.htm, where the pictures, facial muscles and descriptions of all AUs are listed.

http://dx.doi.org/10.1016/j.patcog.2016.07.009
0031-3203/

Table 1
A list of several frequent AUs, their interpretations, corresponding face regions and facial muscles. (adapted from [5]).

AUs	Picture	Interpretation	Facial Muscles	AUs	Picture	Interpretation	Facial Muscles
AU1		Inner Brow Raiser	Frontalis, pars medialis	AU2		Outer Brow Raiser	Frontalis, pars lateralis
AU4		Brow Lowerer	Corrugator supercilii, Depressor supercilii	AU5		Upper Lid Raiser	Levator palpebrae superioris
AU6		Cheek Raiser	Orbicularis oculi, pars orbitalis	AU7		Lid Tightener	Orbicularis oculi, pars palpebralis
AU9		Nose Wrinkler	Levator labii superioris alaquae nasi	AU12		Lip Corner Puller	Zygomaticus Major
AU14		Dimpler	Buccinator	AU15		Lip Corner Depressor	Depressor anguli oris
AU17		Chin Raiser	Mentalis	AU20		Lip Stretcher	Risorius platysma
AU23		Lip Tightener	Orbicularis oris	AU24		Lip Pressor	Orbicularis oris
AU25		Lip Part	Depressor labii inferioris or relaxation of Mentalis, or Orbicularis oris	AU27		Mouth Stretch	Pterygoids, Digastric

to recognize all present AUs from expressional images, and then to describe all possible facial expressions. More formally, an expressional image is denoted as an instance x, and m candidate AUs are represented as $\{\mathbf{c_1}, \mathbf{c_2}, \ldots, \mathbf{c_m}\}$. The label vector of x is denoted as $z \in \{-1, +1\}^m$, where a positive value indicates the presence of the corresponding class (AU), while a negative value means absence. Consequently, the AU recognition can be seen as the prediction of the complete label vector z of x, which is naturally formulated as a multi-label learning problem.

The majority of previous multi-label learning methods assume that each training instance is associated with a complete label assignment. However, in AU recognition, it is often difficult to obtain a complete label assignment for each training sample. For example, AU is typically manually labeled by trained human experts, which is expensive and time consuming. Furthermore, because of the ambiguity nature of AUs as well as the subjective difference, some AUs are difficult to label reliably and confidently. Hence a more realistic scenario is that we have to learn the model from incomplete data, i.e., a part of labels for some training instances are missing. To explicitly accommodate the missing labels, we utilize the definition of an incomplete label vector $y \in \{-1, 0, +1\}^m$ where a 0 indicates the missing label.

Wu et al. [8,9] proposed a multi-label learning with missing labels strategy for AU recognition, and achieved improved performance compared to several state-of-the-art methods. They utilize both the instance-level and class-level smoothness to build a unified graph, such that the label information can be propagated from the provided labels to missing labels. However, similar as most multi-label learning methods [12,13,43–44], studies [8,9] compute a shared instance-level similarity for all AU classes between two instances, based on the whole features of images. According to FACS, different AUs are caused by different sets of facial muscles, and hence produce changes of skin surface displacement or face appearance in different face regions. For example, as shown in Table 1, the contraction of muscle group *Occipito Frontalis* will produce AU1 while the contraction of muscle group *Mentalis* will cause AU15. These two AUs cause feature changes in different regions. Hence, features selected for AU1 are not discriminative for AU15, and vice verse (as demonstrated in Fig. 1). Thus, we believe that the shared instance-level similarity for all AU classes is unable to distinguish the subtle changes of different AUs.

In this work, we propose to firstly learn the discriminative features for each AU class by supervised learning. As a result, the feature noise from the occurrence of other AUs could be filtered. The class-specific instance-level similarity among instances is then computed based on the learned features, which is further incorporated into the MLML model. The proposed model investigates both the discriminative information from features and the label dependency information from labels in a principled manner. We also provide two efficient solutions, including the exact solution based on ST-CUT method [56] and the approximate solution based on conjugate gradient method [55]. Sufficient evaluations on both posed and spontaneous expression databases demonstrate the superiority of the proposed method compared to state-of-the-art.

2. Related works

Because of great potential application value in a wide range of different fields, there has been extensive research on facial expression analysis over the past decades. Typical strategy for facial actions recognition includes two phases: image feature extraction and machine analysis of facial actions. For image feature extraction, some works try to extract features based on the location of facial salient points [26,27], and the shapes of the facial components [28,29,1], which are usually referred as geometric features. Face surface and skin texture changes such as wrinkles, bulges, and furrows are employed as appearance features. Such kind of features includes Gabor feature [30], Haar feature [31,32], LBP feature [33,34], etc. Detailed survey on AU recognition is referred to [18].

Machine analysis methods can be grouped into separated-AU recognition methods and joint-AU recognition methods.

Separated-AU recognition methods usually try to learn discriminative classifier for each AU individually [31,35,30,26,27], which ignore the correlation dependency among AUs. In contrast,