Computer Vision and Image Understanding 147 (2016) 69–76

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Capturing global spatial patterns for distinguishing posed and spontaneous expressions

Shangfei Wang[a,*], Chongliang Wu[a], Qiang Ji[b]

[a] Key Lab of Computing and Communication Software of Anhui Province, School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui 230027, PR China
[b] Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy 12180, NY, USA

ARTICLE INFO

Article history:
Received 28 February 2015
Accepted 18 August 2015

Keywords:
Global spatial patterns
Posed and spontaneous expression distinction
Privileged information

ABSTRACT

In this paper, we introduce methods to differentiate posed expressions from spontaneous ones by capturing global spatial patterns embedded in posed and spontaneous expressions, and by incorporating gender and expression categories as privileged information during spatial pattern modeling. Specifically, we construct multiple restricted Boltzmann machines (RBMs) with continuous visible units to model spatial patterns from facial geometric features given expression-related factors, i.e., gender and expression categories. During testing, only facial geometric features are provided, and the samples are classified into posed or spontaneous expressions according to the RBM with the largest likelihood. Furthermore, we propose efficient inference algorithm by extending annealing importance sampling to RBM with continuous visible units for calculating partition function of RBMs. Experimental results on benchmark databases demonstrate the effectiveness of the proposed approach in modelling global spatial patterns as well as its superior posed and spontaneous expression distinction performance over existing approaches.

1. Introduction

Spontaneous expressions reveal one's real emotions, while posed expressions may disguise one's inner feelings. Automatically distinguishing between spontaneous and posed expressions can benefit many real life scenes. For example, service robots can make human–robot interaction more realistic by perceiving users' true feelings. Doctors can be more certain during diagnosis by knowing patients' genuine feelings. Detectives may detect a lie by differentiating posed expression from spontaneous ones.

Behavior research indicates that posed and spontaneous expressions are different from each other in both temporal and spatial patterns. Temporal patterns involve the speed, amplitude, trajectory and total duration of onset and offset. For example, Ekman et al. [1,2] revealed that the trajectory appears often smoother for spontaneous expressions than for posed ones, and the total duration is usually longer, and onset is more abrupt for posed expressions than spontaneous expressions in most cases. Spatial patterns mainly consist of the movement of facial muscles. Ekman et al. [1] found that both zygomatic major and orbicularis oculi are contracted during spontaneous smiles, while only zygomatic major is contracted for posed smiles, as shown in Fig. 1. Furthermore, the contraction of zygomatic major is more likely to occur asymmetrically for posed smiles than spontaneous ones [3]. Recently, some works reveal contradictory findings in spatial patterns. For example, Krumhuber et al. [4] questioned the differences of orbicularis oculi muscle movements between posed and spontaneous smile. Schmidt *et al* [5] suggested that asymmetry of facial movements may play a much smaller role in distinguishing posed and spontaneous smile. But they observed other differences between posed and spontaneous smile, such as smile intensity [4], amplitude, maximum speed, and duration [5]. Despite lack of a consensus on the differences between posed and spontaneous expression, we believe there indeed exist differences in spatial and temporal facial patterns between posed and spontaneous facial expressions as demonstrated by existing research. And, the goal of this research is to automatically capture the differences and to leverage them for distinguishing posed and spontaneous facial expressions.

Inspired by the observations from nonverbal behavior research, researchers have begun to pay attention to posed and spontaneous expression distinction. The main components of posed and spontaneous expression distinction consists of feature extraction and classification. Although various features are proposed to describe temporal patterns embedded in spontaneous and posed expressions, and many classifiers are adopted, most studies only focus on one kind of expression, such as smile or pain, and little research explicitly models

* Corresponding author.
E-mail addresses: sfwang@ustc.edu.cn (S. Wang), clwzkd@mail.ustc.edu.cn (C. Wu), qji@ecse.rpi.edu (Q. Ji).

http://dx.doi.org/10.1016/j.cviu.2015.08.007
1077-3142/

(a) Posed smile (b) Spontaneous smile

Fig. 1. Posed and spontaneous smile, frames in (a) are from posed smile, and frames in (b) are from spontaneous smile. Both the zygomatic major (facial mouth area) and the orbicularis oculi (eyes area) are contracted during spontaneous smiles (the first frame of (b)), while only the zygomatic major is contracted for posed smiles (the first frame of (a)). The contraction of zygomatic major is more likely occur asymmetrically for posed smiles than spontaneous ones (second frame in (a) and (b)).

spatial patterns embedded in posed and spontaneous expression respectively. Furthermore, little research incorporates expression-related factors, such as gender, age, for posed and spontaneous expression classification. Thus, in this paper, we propose restricted Boltzmann machine (RBM) to explicitly capture the high-order spatial patterns embedded in posed and spontaneous expressions from facial geometric features, and incorporate gender and expression categories as privileged information during spatial pattern modeling. Specifically, we construct multiple RBMs with continuous visible units to model high order spatial patterns embedded in posed and spontaneous expressions given expression-related factors. During training, contrastive divergence (CD) [6] is adopted to learn the parameters of RBMs. During testing, only facial geometric features are provided, and the samples are classified into posed or spontaneous expressions according to the RBM with the largest likelihood. Furthermore, to calculate the partition function of RBMs, we extended annealing importance sampling (AIS) [7] to RBM with continuous visible units case.

The rest of the paper is organized as follows: Section 2 presents an overview of the related works on posed and spontaneous distinction. The detailed introduction of our method is given in Section 3. Section 4 discusses the experimental results. Finally, the paper is concluded in Section 5.

2. Related work

Current research of posed and spontaneous expression differentiation mainly consists of two steps: feature extraction and classification. For feature extraction, most research proposes features specially designed for differentiating posed expressions from spontaneous ones. For example, Cohn and Schmidt [8] proposed temporal features, i.e., duration, amplitude, and the ratio of amplitude to duration. Valstar [9] defined several mid-level feature, including intensity, speed, duration, symmetry, trajectory and the occurrence order of brow actions, from the displacements of facial fiducial points. Dibeklioglu et al. [10] extracted distance and angular features to discriminate the movements of eyelids. They [11] further extracted amplitude, duration, speed, and acceleration to describe dynamics of eyelid, cheek, and lip corner movements. Seckington [12] defined six features including morphology, apex overlap, symmetry, total duration, speed of onset and speed of offset, to represent temporal dynamics, which is essential for distinguishing between posed and spontaneous smiles. In addition to defining posed vs. spontaneous expression specified features, some research adopts commonly used features for expression recognition. For example, Littlewort et al. [13] fed the extracted Gabor wavelet features into support vector machine (SVM) to recognize 20 facial action units as the middle-level features for posed and spontaneous pain classification. Pfister et al. [14] proposed a spatio-temporal local texture features, CLBP-TOP. Zhang et al. [15] used scale-invariant feature transform (SIFT) appearance features and facial animation parameters (FAP) geometric features.

After feature extraction, classifiers should be trained. Cohn and Schmidt [8] adopted a linear discriminant classifier for posed and spontaneous smile recognition. Littlewort *et al* [13] employed SVM, Adaboost, and linear discriminant analysis to classify posed and spontaneous pain from recognized 20 facial action units. Valstar [9] adopted gentle Boost and relevance vector machines to distinguish posed vs. spontaneous brow actions. Dibeklioglu et al. [10] used continuous HMM, k-NN and naive Bayes classifiers to differentiate spontaneous smiles from posed ones. They [11] also employed individual SVM classifiers for different facial regions, and fuse them to classify genuine and posed smiles. Seckington [12] proposed to use dynamic Bayesian networks (BN) to model the temporal dynamics to distinguish between posed and spontaneous expressions. Zhang et al. [15] adopted minimal redundancy maximal relevance for feature selection, and SVM as classifier for discrimination between posed and spontaneous versions of six basic emotions. Although various approaches have been developed for posed and spontaneous expression differentiation, there still exist several limitations. First, most computer vision works only focus on one specific expression, such as smile. To the best of our knowledge, only two works [14,15] considered all six basic expressions (i.e., happiness, disgust, fear, surprise, sadness and anger) for posed and spontaneous expressions recognition. Zhang et al. [15] investigated the performance of a machine vision system for posed and spontaneous expressions recognition of six basic expression on USTC-NVIE database. Pfister et al. [14] proposed a generic facial expression recognition framework to differentiate posed from spontaneous expressions from both visible and infrared images on SPOS database.

Furthermore, most current works applied different classifiers for posed and spontaneous expression recognition, without capturing the spatial patterns embedded in posed and spontaneous expressions explicitly. We call them feature-driven method. Only recently, Wang et al. [16] proposed multiple BN to capture posed and spontaneous spatial facial patterns respectively given gender and expression categories. We call it a model-based method. Their recognition results on the USTC-NVIE and SPOS databases outperform those of the state of the art. However, due to the first-order Markov assumption of BN, their model can only capture the local dependencies among geometric features instead of the global and high-order relations among them. Furthermore, finding the optimal structure of a large geometric feature network for posed and spontaneous expression recognition is difficult. Compared with BN, RBM can model higher-order dependencies among random variables by introducing a layer of latent units [17]. It has been widely used to model complex joint distributions over structured variables such as image pixels. Thus, in this paper, we propose to use RBM to explicitly model complex joint distributions over feature points , i.e., spatial patterns, embedded in posed and spontaneous expressions respectively.

In addition, little work incorporates expression-related factors, such as gender, age and expression categories, for posed and spontaneous expression distinction, although researches indicate that different gender have different facial expression manifestation, face structures develop with ages, expression manifestation varies with ages, and different expressions usually evokes different spatial patterns [18,19]. Recently, Dibeklioglu et al. [11] analyzed effect of age and gender on posed and spontaneous expression distinguishing by using age or gender as one feature. Wang et al. [16] employed gender and expression categories as privileged information to help classify posed and spontaneous expressions. Compared with these two works, the former requires expression-related factors during both training and testing, while the later requires expression-related factors only for training. It means expression-related factors should be predicted during testing in the former. Such sequential approach may propagate the error of expression-related factor recognition