Expert Systems With Applications 56 (2016) 282–290

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Subject adaptation using selective style transfer mapping for detection of facial action units

Amin Mohammadian[a], Hassan Aghaeinia[b], Farzad Towhidkhah[a,*], Seyyede zohreh seyyedsalehi[a]

[a] Dept. of Biomedical Engineering, Amirkabir University of Technology, Tehran 15914, Iran
[b] Dept. of Electrical Engineering, Amirkabir University of Technology, Tehran 15914, Iran

ARTICLE INFO

Article history:
Received 22 December 2014
Revised 21 February 2016
Accepted 10 March 2016
Available online 17 March 2016

Keywords:
Action unit detection
Style transfer
Subject adaptation

ABSTRACT

A key assumption in some learning methods in intelligent systems is that the probability distribution of both test and training data is the same. However, the identical distribution assumption does not hold true for person-independent facial expression recognition. This is because the appearance of an expression may significantly vary for different people. Therefore, domain adaptation methods have been proposed to bring the performance of a person-independent system closer to a person-dependent one. Mismatched conditions between training data and new subject data vary based on the individual. Selective style transfer mapping (SSTM) is an instance-transfer method that will not require re-training classifier and can be classifier independent. This method is proposed to increase the generalization ability of action unit detection through the selection of style transfer mapping type (linear or nonlinear) for different persons. We also propose a rapid SSTM that uses the neutral vectors from a particular person (a small amount of data) to improve action unit detection. Rapid SSTM is also first method for style transfer using a SSTM framework without the need of full labelled target data. The F1 score of selective style transfer mapping for action unit detection in the UNBC-McMaster database is 81.15, which is a significant ($P < 0.05$) improvement over the style transfer mapping, which is 60.30. The results also show that our approach can effectively perform the task of action unit detection with better generalization of the subjects in the other database. The training database was CK+, but the test database was UNBC-McMaster.

1. Introduction

Based on the fact that the human face is sensitive to emotional states, automatic recognition of facial expressions plays an important role in the study of psychological phenomena and the development of nonverbal communication (Cohn, 2010). The analysis of facial expressions is a visual pattern recognition problem in which the facial *Action Units* (AUs), or emotional expressions have to be recognized.

The *Facial Action Coding System* (FACS) is the most well-known and comprehensive method for the recognition of facial gestures (Cohn, Zlochower, Lien, & Kanade, 1999). It is a system designed to describe changes in facial expression in terms of visually observable movements of the facial muscles. 44 different AUs have been defined that are anatomically related to the contraction of either a specific facial muscle or a set of facial muscles. These AUs (independent of any interpretation) can be used as the basis for any higher order decision making process, including the recognition of basic emotions (Senechal, Bailly, & Prevost, 2014), cognitive states (e.g. interest), and psychological states such as pain. Although the FACS provides a good foundation for the AU-coding of facial images by human observers, automatic AU recognition remains difficult (Pantic & Rothkrantz, 2004).

The facial AU detection rate is sensitive to certain imaging conditions such as lighting and viewpoint, as well as subject variability (Gong, McKenna, & Psarrou, 2000). Subject variability is an essential factor in the performance of a facial expression recognition system (Cohen, Sebe, Garg, Chen, & Huang, 2003; Maronidis, Bolis, Tefas, & Pitas, 2011). The difference between the probability distributions of the training and test data is considered the basic assumption in person-independent recognition systems (Zhang & Liu, 2013). Additionally, the training sample size and the variety of samples play an important role in the system's learning and generalization capabilities. One solution for improving the system's reliability is to use larger and more diverse training data, but image data collection is a time-consuming and expensive task.

Inductive transfer learning has been used to establish a person-specific model for facial expression recognition. In inductive transfer learning only a small amount of labeled data in the target

* Corresponding author. Tel.: +98-21-6454-2363; fax: +98-21-66468186.
E-mail addresses: a.mohammadian@aut.ac.ir (A. Mohammadian), Aghaeini@aut.ac.ir (H. Aghaeinia), Towhidkhah@aut.ac.ir (F. Towhidkhah), Z.seyyedsalehi@aut.ac.ir (S.z. seyyedsalehi).

http://dx.doi.org/10.1016/j.eswa.2016.03.023
0957-4174/

domain is available. Learning a classifier solely from the labeled target data may suffer serious overfitting (Chen, Liu, Tu, & Aragones, 2013). Selective transfer machine, as a transductive learning method, has been introduced to personalize a generic classifier of facial AUs by attenuating person-specific biases (Chu, De la Torre, & Cohn, 2013). Moreover, other techniques have been proposed to use primary information from a new person to adapt the system to that new person or to enrich the overall training set (Cavalin, Sabourin, Suen, & Britto, 2009). The task-dependent multi-task multiple kernel learning (TD-MTMKL) method has been used to distinguish between neutral facial expressions and the presence of AUs. When the system detects a set of AUs that usually occur with specific emotions, it is essential for the system not only to determine the commonalities across multiple AU detection tasks but also to adapt to the differences in tasks (Zhang & Mahoor, 2016). Maximum Likelihood Linear Regression (MLLR) conducts an affine transformation from the adaptation data to maximize the likelihood of it. The parameters of the hidden Markov model (HMM) are then re-estimated by the transformation with the regression matrix (Leggetter & Woodland, 1995). Cross-domain adaptation techniques require prior knowledge of new domains, or subjects, to be applied to them (Bruzzone & Marconcini, 2010). For two-class classification, a scenario of transfer learning has been considered, where only data samples from one of the two classes (e.g., the negative class) are available in the target domain (Chen & Liu, 2014). A classifier for a new target individual is inferred with unlabeled data. Based on a regression framework, this method was proposed to learn a mapping between a marginal distribution of the samples associated with a target person and the parameters of her/his personalized classifier. This distribution is represented by the set of parameters of the classifier in the source subject and by all of the unlabeled samples in the target subject (Sangineto, Zen, Ricci, & Sebe, 2014; Zen, Sangineto, Ricci, & Sebe, 2014).

By contrast, a novel framework has been proposed for writer adaptation (Zhang & Liu, 2013). The Style Transfer Mapping (STM) projects the new samples by mapping them without reweighting classifiers or the classifier parameters. This mapping projects the data of different writers onto a style-free space where the independent classifier can achieve significantly higher accuracy. The STM needs various samples of the new person to increase their generalization ability. Existing approaches of transfer learning is classified in instance-transfer, features-transfer and parameter-transfer approaches (Pan & Yang, 2010). STM is based on the features-transfer approach and is not a classifier-dependent technique, and hence can be used with any classifier.

Based on the dynamic modelling of facial expressions, an extension of the STM method was proposed to increase the generalizability of the recognition system. This method preserves the dynamic characteristics while modifying the diversity generated from the information on the new person and reducing the effect of inter-personal variation (subject style) via transfer mapping, estimated by using virtual samples (Mohammadian, Aghaeinia, & Towhidkhah, 2015). Regularized latent task structure proposed for learning person-specific facial AU detection models. This method was able to produce subject-specific AU detection models even without any training data for the target task by exploiting annotated data of the same subject but for a different (Almaev, Martinez, & Valstar, 2015). A Confident Preserving Machine (CPM) was proposed that follows an easy-toward classification strategy. During testing, CPM then learns a person-specific classifier using "virtual labels" provided by confident classifiers. This step is achieved using a quasi-semi-supervised (QSS) approach. Hard samples are typically close to the decision boundary, and the QSS approach disambiguates them using spatio-temporal constraints (Zeng, Chu, De la Torre, Cohn, & Xiong, 2015).

We addressed the issue of person-independent facial AU detection by adapting the system to a new person. To address the person-independent facial AU detection, we proposed a method based on the style transfer framework. The following requirements are necessary to conduct an effective style transfer method: (1) a transfer mapping must be able to completely transfer the new samples to the targets; (2) transfer, non-transfer, and its type must be selectable for different conditions; and (3) there must be enough adaptation data to estimate the transfer mapping. We proposed a powerful and flexible transformation as the mapping of the test feature to the space of the training features to meet these requirements. This transformation is not constrained to be linear. To summarize, the novel contributions of this paper are as follows:

- Different mappings were considered within the selective style transfer framework in order to increase the generalization ability for AU detection in matched and mismatched conditions between training data and new subject data.
- In developing the linear style transfer mapping found in (Zhang & Liu, 2013) a nonlinear transfer mapping has been proposed.
- We suggest a new source point set in style transfer framework that uses the neutral vectors from a specific individual (a small amount of data) and virtual synthesis model in order to improve AU detection.

The rest of this paper is organized as follows: Section 2 describes our proposed method for resolving the person-independent problem. Section 3 defines the baseline system for the facial AU detection. The experimental results and discussion are presented in Sections 4 and 5 and the conclusions are set out in Section 6.

2. Selective style transfer mapping

Suppose we are given labeled training samples $\{\boldsymbol{x}_i^{tr}, \boldsymbol{y}_i^{tr}\}_{i=1}^{l_{tr}}$, where $\boldsymbol{x}_i^{tr} \in R^D$ (D denotes the input dimensionality) is i*th* training input point drawn independently from a probability distribution with density $p_{tr}(\boldsymbol{x})$ and $\boldsymbol{y}_i^{tr} \in \{1, \ldots, K\}$ (K denotes the number of classes) is a training label following a conditional probability distribution with density $p(\boldsymbol{y}|\boldsymbol{x} = \boldsymbol{x}_i^{tr})$. In addition to the labeled training samples, suppose we are given unlabeled test input points $\{\boldsymbol{x}_i^{te}\}_{i=1}^{l_{te}}$, where $\boldsymbol{x}_i^{te}$ ($\in R^D$) is i*th* test input point drawn independently from a probability distribution with density $p_{te}(\boldsymbol{x})$. Note that $p_{tr}(\boldsymbol{x}) \neq p_{te}(\boldsymbol{x})$ in general, and thus the input distributions are different between the training and test phases. Our goal is to learn a transfer mapping that bring $p_{te}(\boldsymbol{x})$ closer to $p_{tr}(\boldsymbol{x})$ to predicts a class label y^{te} for a test input point $\boldsymbol{x}^{te}$.

We write an observation vector in the style s and content c as $\boldsymbol{x}^{sc}$. Content (facial expressions) is a representation of the intrinsic face configuration through the action as a function of time and is invariant to the person. Style is a time-invariant person parameter (which is used to identify the person). Facial expression style is also exist in the neutral vector of each person. The distinctive manner of expression (subject style) is the cause of inter-personal variations, which play an important role in the performance of facial expression recognition (Maronidis, et al., 2011; Seung Ho, Kostas Plataniotis, & Yong Man, 2014). We assume some model $\boldsymbol{x}^{sc} = f(\boldsymbol{a}^s; \boldsymbol{b}^c; \boldsymbol{W})$, where $\boldsymbol{a}^s$ is vector describing style s that is time invariant and $\boldsymbol{b}^c$ is vector describing content c, and $\boldsymbol{W}$ is a set of parameters which expresses the interaction between the two factors. We assume f is a asymmetric bilinear model, given as (1) in which the terms w_{iju} vary with style (Tenenbaum & Freeman, 2000).

$$\boldsymbol{x}_u^{sc}(t) = \sum_j a_{ju}^s b_{tj}^c \tag{1}$$

where $a_{ju}^s = \sum_i a_i^s w_{iju}^s$ and i, j and u denote the components of style, content, and observation vectors, respectively. t is time. In