Expert Systems With Applications 89 (2017) 129–137

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Multi-Objective Differential Evolution for feature selection in Facial Expression Recognition systems

Uroš Mlakar*, Iztok Fister, Janez Brest, Božidar Potočnik

University of Maribor, Faculty of Electrical Engineering and Computer Science, Koroška cesta 46, SI-2000, Slovenia

ARTICLE INFO

Article history:
Received 31 January 2017
Revised 23 June 2017
Accepted 23 July 2017
Available online 24 July 2017

Keywords:
Facial Expression Recognition
Feature selection
Differential evolution
Feature vector differences
Multi-objective optimization

ABSTRACT

This paper proposes an efficient feature selection system applied to a Facial Expression Recognition (FER) system. This system, capable of recognizing seven prototypical emotions including neutral expression, is based on a histogram of oriented gradient descriptor (HOG) and difference feature vectors. The emotion feature selection was carried out by using an appropriately modified multi-objective differential evolution algorithm. The number of used features was minimized, while the emotion recognition accuracy of the support vector machine classifiers was maximized simultaneously. 'The emotion-specific features' and 'the more discriminative features over all emotions' selection strategies were developed, whereby the latter strategy proved to be more efficient using the Friedman statistical test. This person-independent FER system with proposed feature selection was validated on three commonly used evaluation databases, where the mean emotion recognition rate was 98.37% on the Cohn Kanade database, 92.75% on the JAFFE database, and 84.07% on the MMI database, while the number of used features lowered up to 89% with respect to the original difference feature vector length. Compared to the state-of-the-art, the proposed FER method offers good results, while also greatly lowering the number of used features, which, in return, minimizes the computational cost of training the classifiers. The optimization proposed in this paper can be generalized easily to a feature selection for an arbitrary multi-objective, as well as many-objective, problem.

1. Introduction

Research on Facial Expression Recognition (FER) is focused either on recognizing emotions indirectly through the facial action coding system (Ekman & Rosenberg, 1997), or on trying to identify emotions directly from facial images. The following basic emotional states are usually being recognized (Mlakar & Potočnik, 2015): Anger, disgust, fear, surprise, sadness, happiness, and sometimes also the neutral emotion.

A typical FER system consists of: (i) Face acquisition module, (ii) Feature extraction and selection module, and (iii) Feature classification module (Mlakar & Potočnik, 2015). According to our experiences, feature extraction and selection plays a key role in the emotion recognition process. By using inappropriate features, even state-of-the-art classifiers would demonstrate low efficiency (Lajevardi & Hussain, 2012).

The two most common facial feature extraction methods are: (i) Geometric feature-based methods and (ii) Appearance-based methods (Tian, Kanade, & Cohn, 2005). Geometric feature-based methods construct a feature vector as a collection of shapes and locations of important facial features (e.g., eyes, mouth, and nose), while the appearance-based methods extract features by special filters, which are applied either to the whole face or to some specific facial regions.

The obtained feature vector is a higher-level representation of the facial image and is, typically, high dimensional. A large amount of data makes the training of a classifier hard and ineffective. To overcome this problem, a feature selection should be performed in order to select an optimal subset of features that contain sufficient information for correct classification. Two methods are found in Lajevardi and Hussain (2012) for feature selection, namely filter-based and wrapper-based. The former uses a proxy measure to rank feature subsets. Such measure is simple and usually computed fast, while still holding sufficient information for the classification (Guyon & Elisseeff, 2003). On the other hand, the wrapper-based method utilizes a predictive model to score/rank the subsets. Each subset is used to train a model, which is then evaluated (e.g., on

* Corresponding author.
E-mail addresses: uros.mlakar@um.si (U. Mlakar), iztok.fister@um.si (I. Fister), janez.brest@um.si (J. Brest), bozidar.potocnik@um.si (B. Potočnik).

http://dx.doi.org/10.1016/j.eswa.2017.07.037
0957-4174/

a hold-out set). This selection technique, although computationally more intensive, is usually more efficient.

Exhaustive selection methods are extremely intensive computationally, especially if dealing with high dimensional features. Therefore, many researches have focused on heuristic algorithms such as Evolutionary Algorithms (EAs), that try to find the near-optimal solutions in real-time. It has been pointed out that any non-exhaustive selection method does not guarantee to find the optimal feature subset, but rather provides a satisfactory local optimum (Peng, Long, & Ding, 2005). Plenty of papers related to feature selection for FER in combination with EAs have been published to date, e.g., Yu and Bhanu (2006), Zavaschi, Britto, Oliveira, and Koerich (2013), Olague, Hammoud, Trujillo, Hernández, and Romero (2009), Soyel, Tekguc, and Demirel (2011) and Lajevardi and Hussain (2012).

Cited methods solve the feature selection as single-objective problem, where the selected features depend heavily on a classifier's accuracy. Additionally, these methods converged towards a local optimum or to a sub-par result, respectively. A step forward were methods that treat the feature selection as a multi-objective problem, where a vector was optimized of more conflicting objectives (Deb, 2001). Naturally, the feature selection can be represented as the Multi-Objective Optimization Problem (MOOP), with the goal to maximize the classifier's performance and minimize the number of features simultaneously.

Recently, more important multi-objective methods were reported (Soyel et al., 2011; Zavaschi et al., 2013). A short analysis revealed that these, based mainly on a Genetic Algorithm (GA) and on a multi-objective GA, exhibited a modest emotion recognition accuracy even on simple data sets, e.g., the best results were around 90% on the Cohn Kanade (CK) public database. Additionally, these methods focused merely on the selection of discriminative features for each distinctive facial expression, and never tried to determine a single face area that would distinguish between all expressions simultaneously, in order to construct a classifier with a better generalization ability.

In this study, we addressed both shortcomings of the reviewed literature, namely: (i) Relatively low emotion recognition accuracy of multi-objective methods and (ii) Inability of the classifier to generalize to unknown subjects due to improper selection of feature subsets (i.e., due to improper selection of face areas). This research is based on our previous state-of-the-art method for the automated recognition of prototypical emotional states (Mlakar & Potočnik, 2015). That method, denoted as the baseline method in the sequel, is supplemented by a feature selection using the Multi-Objective Differential Evolution (DEMO) algorithm introduced in Robič and Filipič (2005). The baseline method employed the Histogram of Oriented Gradients (HOG) descriptor and recognized six emotional states on a level of feature vector differences (Mlakar & Potočnik, 2015). The difference vector was obtained as follows. First, the HOG descriptor was applied separately on the input emotional image (i.e., an image with expressed emotional state in its peak) and on the neutral image of the same person. Afterwards, both obtained vectors were simply subtracted to form the final feature vector, which was then classified by Support Vector Machines (SVM) (Mlakar & Potočnik, 2015). The main weakness of the described methodology is that the whole feature vector was always employed, whereas all features had an equal weight in the training of the classification model. Consequently, the emotion recognition was slower and the recognition accuracy was slightly lower compared to the most efficient state-of-the-art methods.

In this paper, the basic idea from our previous work (Mlakar & Potočnik, 2015) is extended by the feature selection algorithm based on multi-objective optimization. A subset of discriminative features was determined by the DEMO algorithm. Its choice was justified by the known fact that the DEMO algorithm has many advantages with respect to the multi-objective GA algorithm when solving MOOPs (Tušar & Filipič, 2007). Namely, it requires less control parameters, and demonstrates higher performance, computational speed, and robustness. Two conflicting objectives were tackled in our MOOP algorithm: (i) The maximization of the classifier's accuracy and (ii) The minimization of selected feature subset length. Our feature selection method is wrapper-based, whereas the classifier is a simple linear SVM. In addition, we adapted the original method in such a way that, besides the existing six emotions it is also possible to recognize the seventh, i.e. the neutral expression.

Key novelties of this research are: (i) A wrapper-based selection of important features from facial images utilizing a Multi-Objective Differential Evolution Algorithm, (ii) A comparative study of two different feature selection strategies aimed for the FER, namely the emotion specific feature selection and a selection of more discriminative features, valid for all emotions being recognized, (iii) Utilizing the feature vector differences within a feature selection strategy for facial expression recognition, and (iv) A new facial emotion recognition method as a whole.

Our research also has several contributions to the field of expert and intelligent systems:

1. To the best of our knowledge, the original DEMO algorithm has not been used as a feature selector for the facial expression recognition problem up to now. Therefore, our study is the first research that demonstrated an efficient utilization of the DEMO algorithm for the FER. The DEMO algorithm was also modified appropriately to support two different feature selection strategies (i.e., a class specific feature selection, and a selection of more discriminative features over all classes).
2. The optimization proposed in this paper can be generalized easily to a feature selection for an arbitrary NE-class ($NE \geq 2$) classification problem, where either class specific features or the more discriminative features over all classes are determined.
3. The feature vectors we were dealing with within the DEMO algorithm were constructed by the HOG texture descriptor. This descriptor is parametric as well. Consequently, another hard optimization problem should be solved in order to obtain the best setting of HOG descriptor parameters. In this study, we avoided this difficult problem elegantly by preparing in advance the five sets of more perspective parameters for the HOG descriptor (we determined these sets by the real-value coded-GA in our previous study (Mlakar & Potočnik, 2015)), whereat these sets were, afterwards, applied for the feature vector construction. Indirectly, these sets were also utilized within the DEMO algorithm. In this way, we demonstrated in practice how to carry out double optimization (i.e., tuning of the method's parameters and feature selection) with an acceptable computation cost.
4. Last but not least, our proposed FER method may serve as a framework for solving hard optimization problems in different domains, such as, for an example, for numerical function optimization (Abbass, Sarker, & Newton, 2001), business optimization (Babu & Gujarathil, 2007), power optimization (Basu, 2011) etc.

2. Feature extraction

The feature extraction step was borrowed from our previous work (Mlakar & Potočnik, 2015), whereby only small modifications were needed in this study. Namely, we omitted approximating a face by an ellipse. The baseline algorithm otherwise employed the well-established HOG (Dalal & Triggs, 2005) texture descriptor for extracting features from facial images. This extraction was conducted as follows. First, the image was divided into small regions