Expert Systems with Applications 40 (2013) 5077–5084

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Feature selection for face recognition based on multi-objective evolutionary wrappers

Leandro D. Vignolo [a,*], Diego H. Milone [a], Jacob Scharcanski [b]

[a] Grupo de Investigación en Señales e Inteligencia Computacional, Departamento de Informática, Facultad de Ingeniería y Ciencias Hídricas, Universidad Nacional del Litoral, CONICET, Argentina

[b] Instituto de Informatica and Dept. de Engenharia Eletrica, Universidade Federal do Rio Grande do Sul, Caixa Postal 15064, 91501 970 Porto Alegre, RS, Brazil

ARTICLE INFO

Keywords:
Wrappers
Multi-objective genetic algorithms
Feature selection
Face recognition

ABSTRACT

Feature selection is a key issue in pattern recognition, specially when prior knowledge of the most discriminant features is not available. Moreover, in order to perform the classification task with reduced complexity and acceptable performance, usually features that are irrelevant, redundant, or noisy are excluded from the problem representation. This work presents a multi-objective wrapper, based on genetic algorithms, to select the most relevant set of features for face recognition tasks. The proposed strategy explores the space of multiple feasible selections in order to minimize the cardinality of the feature subset, and at the same time to maximize its discriminative capacity. Experimental results show that, in comparison with other state-of-the-art approaches, the proposed approach allows to improve the classification performance, while reducing the representation dimensionality.

1. Introduction

Face recognition has received significant attention due to its promising applications in security systems and human-computer interaction, which has motivated important new developments in research areas such as image processing and artificial intelligence. In general, the methodologies are developed for face images acquired under controlled conditions, but in practical situations, face recognition systems usually must also deal with changing conditions like variations in pose, expression and illumination, which introduce intra-class variability in the extracted features with respect to the training data (Li and Jain, 2011; Milborrow and Nicolls, 2008; Wen, 2012). In a face recognition problem, a given face image is classified into K previously known face classes. This is usually done using a model trained with the feature vectors extracted from a database of face images (Cevikalp and Triggs, 2010; Oh et al., 2013).

Two main approaches exist in face recognition, those which are based on holistic methods and the others based on analytic techniques (Kong et al., 2005). Holistic methods, such as eigenfaces (Turk and Pentland, 1991), use global characteristics of the face images. On the other hand, analytic techniques, like the Active Shape Models (ASM) (Cootes et al., 1995; Wang et al., 2013), extract face features related to the eyes, the nose, the mouth, etc.

In facial modeling with ASM, a number of points (i.e. image locations) are selected from an input image, but only some of these points are useful for characterizing the face, since the others have small contributions to discrimination, or are noisy. As the training of ASM converges towards salient edges, if these edges are distorted by noise or some other artifact, like local illumination variation, erroneous feature matchings might arise (Behaine and Scharcanski, 2012). Despite recent improvements made to ASM techniques, the matching errors may be undesirably high at some face locations (Hill et al., 1996; Kim et al., 2007). Even after some new implementations that improve the landmark location accuracy, the detection of facial features with varying pose and illumination is still challenging (Milborrow and Nicolls, 2008; Zheng et al., 2008). Usually, once a set of face image locations (i.e. points) is selected by the ASM method, a number of features describing each face location is extracted. Then, the resulting feature vectors representing the faces are usually of high dimensionality, which makes the classification task more difficult (Bishop, 2007). Also, large feature sets are prone to overfitting and, hence, to achieve poor generalization performance (Handl and Knowles, 2006).

In (Behaine and Scharcanski, 2012), the authors proposed to improve the ASM performance in face recognition by weighting the facial features according to a method based on adjusted mutual information. As the authors shown, this criterion allowed the selection of the most relevant landmark points, in order to improve the face classification results. However, the flexibility provided by the full set of features obtained by the ASM approach has not yet been fully explored by means of feature selection techniques, in order to

* Corresponding author. Tel.: +54 342 4575233x191; fax: +54 342 4575224.
E-mail addresses: ldvignolo@fich.unl.edu.ar (L.D. Vignolo), d.milone@ieee.org (D.H. Milone), jacobs@inf.ufrgs.br (J. Scharcanski).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.03.032

reduce the dimensionality of the representation while improving the face classification results. On the other hand, significant progresses have been made with the application of different artificial intelligence techniques for feature selection. In particular, many works rely on evolutionary algorithms for feature subset optimization (Chatterjee and Bhattacherjee, 2011; Hsu et al., 2011; Li et al., 2010; Pedrycz and Ahmad, 2012), and for the search of optimal representations (Charbuillet et al., 2009; Vignolo et al., 2011a,b; Vignolo et al., 2013). In (Vignolo et al., 2012) a genetic wrapper was proposed for the selection of the most relevant features for improving the accuracy of face recognition. Nevertheless, this wrapper was focused on classification accuracy improvement, which limits the proposed method since it overlooks other important issues in face classification (e.g. feature space dimensionality and class overlap).

In order to guide the search within the space of feasible face classification solutions, here we propose the use of a Multi-Objective Genetic Algorithm (Coello Coello et al., 2007). This method allows to overcome the above mentioned limitations by maximizing the face classification accuracy, while minimizing the number of features and the mutual information. Two different strategies for the representation of the candidate solutions are proposed and compared, and the generalization performance of the feature subset selection is assessed using an independent data set.

The organization of this paper is as follows. First a brief introduction to the use of ASM for face modeling is given in Section 2, and next our multi-objective wrapper for the selection of features for face classification is presented in Section 3. Section 4 describes our experiments and discuss the results obtained for face classification. Finally, our conclusions and ideas for future work are presented in Section 5.

2. Active shape models for facial recognition applications

The ASM approach is used to represent shapes and their expected ways of deforming as learned from a training set. For this, it uses flexible point distribution models (PDM), based on the positioning of selected points in the face image examples (Hill et al., 1996). This PDM iteratively deforms to fit the shape of an object, constrained to vary in the way learned from a set of training examples. When applied to face recognition, the ASM is trained on a set of sample faces, and N points are used to represent the shape of each face within its class (see Fig. 1(a)).

Nevertheless, matching errors may arise in the location of the PDM points, often called *landmarks*, in a face image (see Fig. 1(b)) (Behaine and Scharcanski, 2012). Then, considering a training image set with K face classes, each class $k = 1, \ldots, K$ is represented by N landmark points $S_{k,\epsilon} = \{p_i(x_i + \epsilon_{x_i}, y_i + \epsilon_{y_i})^k\}$, where $i = 1, \ldots, N$, (x_i, y_i) are the coordinates of the landmark point p_i and $(\epsilon_{x_i}, \epsilon_{y_i})$ are the respective location errors. Every relevant facial characteristic (e.g. eye centers, mouth contours, etc.) is represented by a set of landmarks p_i, and the particularities of each point in the image are described by Q features (e.g. chrominance, texture, etc.). The features at landmark p_i will be denoted $\{F_{j,i}\}$, with $j = 1, \ldots, Q$.

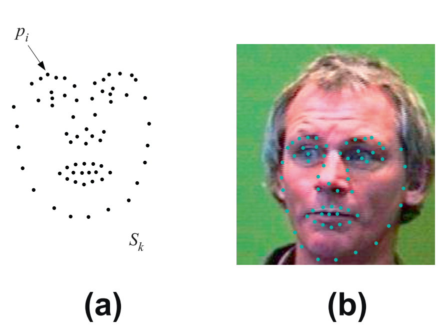

Fig. 1. Illustration of the landmark points used to model a face (a) and their location on an image (b) (Behaine & Scharcanski, 2012).

In order to describe each one of the N landmark points p_i, the mean $\mu_{F_{j,i}}$ and the variance $\sigma^2_{F_{j,i}}$ of the measurements of each feature j taken within a defined neighborhood of that point are commonly used (Behaine and Scharcanski, 2012). These are computed for all features $F^m_{j,i}$, with $m = 1, \ldots, M$, where M is the number of training image samples,

$$\mu_{F_{j,i}} = \frac{1}{w^2} \sum_{r=1}^{w} \sum_{q=1}^{w} \mu_{j,i}(r,q), \tag{1}$$

$$\sigma^2_{F_{j,i}} = \max_{r,q \in W} \left\{ \sigma^2_{j,i}(r,q) \right\}, \tag{2}$$

where (r,q) are the pixel coordinates within the window W (of size $w \times w$), centered at the landmark point p_i (Behaine and Scharcanski, 2012), $\mu_{j,i}(r,q) = \frac{1}{M}\sum_{m=1}^{M} F^m_{j,i}(r,q)$ and $\sigma^2_{j,i}(r,q) = \frac{1}{M}\sum_{m=1}^{M} \left(F^m_{j,i}(r,q) - \mu_{j,i}(r,q)\right)^2$.

To consider the feature variability within the $w \times w$ neighborhood of landmark p_i, the maximum window variance was used in (2). The window size was set to $w = 2\max\{\sigma_\epsilon\}$, where σ_ϵ is the standard deviation of landmark location errors, measured during ASM training. The probability density of location errors at each landmark point is assumed to be approximately Gaussian (Shi et al., 2006).

In this work, the face detector proposed by Demirel and Anbarjafari (2009) is used, and the process applied to the database of face images in order to obtain the ASM-based set of features is described in detail in Behaine and Scharcanski (2012), Vignolo et al. (2012).

3. Multi-objective wrapper for face feature selection

Genetic algorithms (GAs) are meta-heuristic optimization methods, inspired on the process of natural evolution, that are capable of finding global optima in complex search spaces (Youssef et al., 2001). These optimization algorithms need to evaluate a problem-dependent objective function to guide the search. However, in most real-world problems we may be interested in satisfying more than one objective, and the optimization of one objective may conflict with the other objectives. In general, the solution of a multi-objective optimization problem is not a single point, but a set of points known as the Pareto optimal front (Kim and Liou, 2012).

Different modifications to the traditional GAs were proposed in order to tackle multi-objective problems (Fonseca and Fleming, 1993). One generic approach is to combine the individual objective functions into a single aggregative function, or to consider all but one objective as constraints. Another generic approach is to determine a Pareto optimal, or nondominated set of solutions. This means, a set of solutions for which none of the objective values can be improved without detriment in some of the other objective functions. This approach takes advantage of the population-based nature of GAs, which allows the generation of several elements of the Pareto set in a single run (Coello Coello et al., 2007).

Particularly, the Multi-Objective Genetic Algorithm (MOGA) is a variation of the classical GA, in which the rank of an individual is the number of chromosomes in the population by which it is dominated (Fonseca and Fleming, 1993). This technique addresses the search toward the true Pareto front, while maintaining diversity in the population (Konak et al., 2006). A problem that arises in Pareto based multi-objective evolutionary algorithms is the difficulty