Signal Processing: *Image Communication* 43 (2016) 54–67

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Optimization-based methodology for training set selection to synthesize composite correlation filters for face recognition

Everardo Santiago-Ramirez [a], Jose A. Gonzalez-Fraga [a,b,*], Everardo Gutierrez [a], Omar Alvarez-Xochihua [a]

[a] *Facultad de Ciencias, Universidad Autónoma de Baja California, Carretera Transpeninsular Ensenada-Tijuana, Núm. 3917, Colonia Playitas, Ensenada, Baja California, C.P. 22860, México*
[b] *Instituto Politécnico Nacional–CITEDI, Ave. del Parque Núm. 1310, Mesa de Otay, Tijuana, Baja California, C.P. 22510, México*

ARTICLE INFO

Article history:
Received 21 June 2015
Received in revised form
13 December 2015
Accepted 10 February 2016
Available online 19 February 2016

Keywords:
Training set selection
Face recognition
Composite correlation filter
Optimization algorithm
Pattern recognition

ABSTRACT

Face recognition has been addressed with pattern recognition techniques such as composite correlation filters. These filters are synthesized from training sets which are representative of facial classes. For this reason, the filter performance depends greatly on the appropriate selection of the training set. This set can be selected either by a filter designer or by a conventional method. This paper presents an optimization-based methodology for the automatic selection of the training set. Given an optimization algorithm, the proposed methodology uses its main mechanics to iteratively examine a given set of available images in order to find the best subset for the training set. To this end, three objective functions are proposed as optimization criteria for training set selection. The proposed methodology was evaluated by undertaking face recognition under variable illumination and facial expressions. Four optimization algorithms and three composite correlation filters were used to test the proposed methodology. The Maximum Average Correlation Height filter designed by Grey Wolf Optimizer obtained the best performance under homogeneous illumination and facial expressions, while the Unconstrained Nonlinear Composite Filter designed by either Grey Wolf Optimizer or (1+1)-Evolution Strategy obtained the best performance under variable illumination. The proposed methodology selects training sets for the synthesis of composite filters with competitive results comparable to the results reported in the face recognition literature.

1. Introduction

Face recognition has a wide range of applications, such as face ID, access control, security, surveillance, smart cards, law enforcement, face databases, multimedia management, and human–computer interaction [1]. Face recognition has several advantages over other biometric technologies: it is natural, nonintrusive, easy to use, and can capture faces both covertly and at a distance [1]. However, the development of face recognition technology remains a challenging task due to the fact that facial appearance alters due to motion, illumination, and environment [2]. Different approaches have been proposed for performing this task [2,3], which can be categorized into feature-based methods and template-matching methods. In feature-based methods the input image is processed to extract relevant features such as lines, edges, curves, facial features and geometrical constraints [4], which are then

* Corresponding author. Tel.: +52 646 1744560x113.
E-mail addresses: everardo.santiagoramirez@uabc.edu.mx (E. Santiago-Ramirez), angel_fraga@uabc.edu.mx (J.A. Gonzalez-Fraga), everardo.gutierrez@uabc.edu.mx (E. Gutierrez), aomar@uabc.edu.mx (O. Alvarez-Xochihua).

http://dx.doi.org/10.1016/j.image.2016.02.002
0923-5965/

compared to stored facial features to carry out the recognition. In template matching methods, several standard face patterns are stored in order to describe the face as a whole, with the recognition carried out by matching the input image with stored patterns.

The goal of feature extraction is to produce a few descriptors to capture the essence of an input image [5,6]. However, this reduction of the information involves two major problems. Firstly, it is not easy to identify which features could be best for the recognition. Secondly, with a few number of features it is difficult to discriminate two similar faces which belong to two different people. These problems could be addressed using the image as a template to provide as much information to classifiers for a robust recognition. However, it has recently been proposed in [7,8] methods that combine both feature and template methods on getting a major robustness.

Correlation filter-based face recognition [9–11,5,12,13] belongs to the template matching category [5,14,15]. A correlation filter is a spatial-frequency array that is specifically designed from a set of training images which are representatives of a particular class [5,14]. This correlation filter is compared to an input image by obtaining a cross-correlation, usually in the frequency domain for computational efficiency. The correlation filters have caused interest in researchers due to several attractive properties: shift invariant [5], robustness to noise [13], graceful degradation of the response to occlusions [14], ability to simultaneously localize and classify objects [15], it can be efficiently implemented in the frequency domain via Fourier transform [15,16], and can be trained either on-line or off-line [10]. Although correlation filters generally are designed with images directly, these can be extended to feature spaces such as is proposed in [7]. This extension, besides the properties above mentioned, significantly reduce the storage requirement. For this reason, the LBPI images [8] were used to capture the best features while preserving both intrinsic appearances and spatial relations (see Section 5.1). Advantages of the correlation is that it can be implemented digitally and optically using the Fourier transform [5]. In [5,13] there can be found enough theory for digital implementation, while in [17] the latest developments in optical correlation are described.

Correlation filter performance depends greatly on the training set. For this reason, it is important to choose a training set that describes the general expected distortions of the object [18,15] and it must be small for computational conveniences [18]. Moreover, it is important to take into account that using too few or too many images may lower filter performance [15]. One approach for training set selection consists of distributing N images uniformly over the range of expected distortions [18]. This approach may be exemplified as follows [18]: if in-plane rotation is the expected distortion, face images obtained by rotating a single image by increments of 6° can be used to obtain 60 training face images. These 60 training face images are used for the synthesis of an in-rotation invariant composite filter. A similar approach is proposed in [19] for out-plane rotation distortion in 3D objects. However, there are distortions where it is impossible to obtain a uniformly sampled set, for example due to the local shadows presented in face images captured under unconstrained environments. In this case, a more appropriate approach is to capture face images with local shadows in different regions of the face. Thus, a correlation threshold c can be used to select the most representative images as it follows [18,15]. First, an image is selected from a given set of available images and used to train the filter using that single image alone, after which the filter is correlated to all the available images. The image with which the filter performed the worst is then selected as an additional training image. The training set now has two images, which are used to train a new filter. The new filter is again correlated to all the available images, and once again, the image with which the filter performed the worst is added to the training set. This process is continued until all remaining images achieve a similarity score greater than or equal to τ_c. In [20] a method is proposed for designing composite correlation filters from a randomly selected initial training set. Then, more images are added to the training set using the same process as the previous approach. Although these approaches can be used for selecting training sets, filter performance may decrease with additional images before τ_c is reached. This indicates that the approach is not capable of capturing the most representative images from the available image set. The training set selection problem consists of selecting the best subset from the available images set. Therefore, this problem can be posed as an optimization problem, as proposed in [21], which outlines the combination of two optimization algorithms for the synthesis of composite correlation filters. Firstly, population management is performed by the Strength Pareto Evolutionary Algorithm 2 (SPEA2) [22]. Secondly, individuals in the population are represented as string of variable length. Genetic operators are then applied to individuals using the Speciation Adaption Genetic Algorithm (SAGA) [23]. This optimization-based algorithm designs only one specific type of composite correlation filter, named MOSDF (Multi-Objective Synthetic Discriminant Function).

The algorithms proposed in the literature are capable of selecting training sets for designing correlation filters with limited performance. Moreover, as previously mentioned, they are not capable of capturing the most descriptive training images. Furthermore, the proposed algorithms have not been systematically evaluated for face recognition use in a biometric context. Therefore, it is necessary to develop an optimized methodology for selecting the most descriptive training images for the synthesis of composite correlation filters as it is reported in the literature.

This paper presents an optimization-based methodology for the automatic selection of training sets. Given an optimization algorithm, the proposed methodology uses its main mechanics to iteratively examine a given image set to find the best subset for the training set. To this end, this study proposes three objective functions which can be used as optimization criteria for training set selection. The main components of the proposed methodology are individual representation, population management, operator, and objective functions. First, the individuals are represented as a binary position vector. Second, the population management consists of a strategy for selecting the best individual from an individual parent and their descendants. Third, an operator is applied to individuals to