Pattern Recognition Letters 29 (2008) 1537–1543

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Face recognition using HOG–EBGM ☆

Alberto Albiol *, David Monzo, Antoine Martin, Jorge Sastre, Antonio Albiol

I-TEAM, Universidad Politecnica de Valencia, Spain

ARTICLE INFO

Article history:
Received 25 May 2007
Received in revised form 10 January 2008
Available online 7 April 2008

Communicated by H.H.S. Ip

Keywords:
Face recognition
EBGM
SIFT
HOG
Local image descriptors

ABSTRACT

This paper presents a new face recognition algorithm based on the well-known EBGM which replaces Gabor features by HOG descriptors. The recognition results show a better performance of our approach compared to other face recognition approaches using public available databases. This better performance is explained by the properties of HOG descriptors which are more robust to changes in illumination, rotation and small displacements, and to the higher accuracy of the face graphs obtained compared to classical Gabor–EBGM ones.

1. Introduction

Face recognition is a research field that has attracted much attention in the past years due to its many applications in public surveillance, video annotation, multimedia and others.

Many algorithms have been proposed for face recognition, we recommend (Zhao et al., 2003) for a complete survey on the topic. Face recognition techniques can be broadly classified into holistic and feature-based. Holistic methods, such as PCA (Turk and Pentland, 1991) or LDA (Belhumeur et al., 1996), project input faces onto a dimensional reduced space where recognition is carried out. These type of methods can be deemed as classical today. However, they are still very popular due to their simplicity and good performance. The greatest problem of eigen-based methods it that in some sense they assume that faces are rigid objects which can be reconstructed with linear combinations of a set of eigenfaces or fisherfaces, which is not true. Also when the illumination conditions change the performance of these methods degrades rapidly. To overcome these difficulties many variations of the original methods have been proposed (del Solar and Navarrete, 2005).

Feature-based methods try to recognize faces using its facial components: eyes, nose, mouth, etc. Active Appearance Model (Cotes et al., 2001) and Elastic Bunch Graph Matching (EBGM) (Wiskott et al., 1996) fall into this category. In EBGM faces are represented as graphs with nodes at facial landmarks (such as eyes, tip of the nose, etc.). Each node contains a set of Gabor wavelet coefficients, known as a jet. To increase robustness of EBGM to changes in expression and illumination, new approaches have been proposed. For example, in (Shin et al., 2007) a graph matching approach which replaces the original Gabor features is presented. Following this line of research, in this paper we present a EBGM algorithm in which Gabor features have been replaced by Histograms of Oriented Gradients (HOG) (Bicego et al., 2006) descriptors, which are inherited from the Scale Invariant Feature Transform (SIFT) proposed by Lowe (2004). SIFT has emerged as a cutting edge technology for extracting distinctive features from images, to be used in algorithms for tasks like matching different views of an object or scene. SIFT achieves invariance to scale changes by extracting keypoints at the local extrema of the scale-space representation of the image, then each keypoint is represented using histograms of image gradients, in the sequel HOG descriptor. HOG descriptors have also been proposed for pedestrian detection (Bay et al., 2006; Mikolajczyk and Schmid, 2005; Dalal and Triggs, 2005). In these approaches objects are assumed to be at a fixed scale and are divided into small connected regions at fixed positions. Then, for each region a HOG descriptor is obtained and the combination of these descriptors is used to represent the object.

SIFT has also been recently proposed for face recognition (Bicego et al., 2006), however this approach totally differs from ours. In the Bicego's algorithm keypoints are located at the local extrema of the scale-space as in the original Lowe's approach (Lowe, 2004). The main problem of this approach is that there is no control on the number, position and scale of the keypoints. However, in our algorithm the keypoints represent specific facial landmarks which are

☆ This work has been supported by the Technical University of Valencia: Programa de apoyo a la investigación y desarrollo PAID-06-06 and the CDTI Hesperia project.

* Corresponding author. Tel.: +34 96 387 73 09.
E-mail address: alalbiol@dcom.upv.es (A. Albiol).

0167-8655/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2008.03.017

detected first as explained below. Once facial landmarks are detected we use HOG descriptors to represent them.

The rest of the paper is organized as follows. First, Section 2 presents in detail how HOG descriptors are built. Next, Section 3 describes our EBGM algorithm that uses HOG descriptors. Sections 4 and 5 show the experimental setup and recognition results. Finally some conclusions and future research are drawn in Section 6.

Henceforth we will use the term Gabor–EBGM, or simply EBGM, when referring to the original EBGM algorithm presented in (Wiskott et al., 1996) while we will use the term HOG–EBGM when referring to the algorithm presented in this paper.

2. HOG descriptors

As mentioned previously SIFT has emerged as one of the most used detection/description schemes for its ability to handle image transformations like scale changes (zoom), image rotation, and illumination. The major steps of the SIFT algorithm are:

(1) Scale-space extrema detection.
(2) Orientation assignment.
(3) Keypoint descriptor.

The first step is used by SIFT to achieve invariance to scale changes. This is done by extracting SIFT features only at the local extrema of the scale-space representation of the image. The next step aims to obtain image rotation invariance. To that end, at each extrema of the scale-space representation, SIFT finds the dominant orientation using image gradient information and then, all image gradients are made relative to this dominant direction.

While these two techniques have proved to be very useful for images that are arbitrarily scaled or rotated, the fact is that these normalization stages remove information which might be useful for recognition when images are not scaled or rotated. In this paper, we assume that the exact location of both eyes is known a priori. To detect the eyes precisely, we have developed an algorithm that uses a mixed approach of boosted classifiers (Viola and Jones, 2001) and again HOG descriptors. However this problem can be deemed as precise face localization and it is not treated here. Since the exact location of the eyes is used to normalize faces, we do not expect any changes in either scale or rotation. For this reason, we skip the two first steps of the SIFT algorithm and only adopt the last step from Lowe's approach, the keypoint descriptor. This keypoint descriptor is also called HOG in the literature.

The HOG descriptor is a local statistic of the orientations of the image gradients around a keypoint. More formally, each descriptor is a bundle of histograms composed of pixel orientations given by their gradients. The number of possible orientations (histogram bins) is referred to as N_o. Each histogram in the bundle describes a specific area around the keypoint. These areas correspond to the cells of a $N_p \times N_p$ squared grid centered on the keypoint (see Fig. 1). The original paper (Lowe, 2004) sets the parameters of the descriptor to $N_p = 4$ cells for each spatial direction and $N_o = 8$ bins for each histogram in the bundle resulting in a total of $N_p^2 N_o = 128$ elements in a HOG descriptor.

In our work, each spatial cell is a square of 5×5 pixels. This size is chosen accordingly to the distance between eyes of the normalized faces, which in our work is 40 pixels. The results presented in Section 5.1.1 will further justify this selection.

Similar to Lowe's original approach, the contribution of each pixel gradient to the histogram is weighted by the gradient modulus and a Gaussian window. The Gaussian window is centered at the keypoint coordinates and its standard deviation equals to half the extension of the spatial range, which is 10 pixels. Also the pixel contribution is distributed into adjacent spatial cells and orientations bins using trilinear interpolation. This is important to avoid all boundary effects in which the descriptor abruptly changes as a sample shifts smoothly from being within one cell to another or from one orientation to another. Gaussian windowing and trilinear interpolation also increases the robustness of the descriptor against small displacements of the keypoint location.

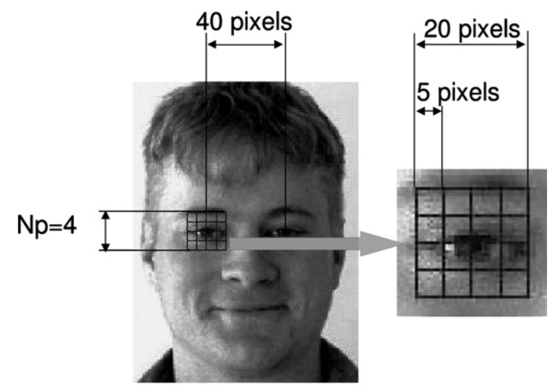

Fig. 1. Normalized face and the spatial cells of the right eye HOG descriptor.

Finally, HOG descriptors are normalized to increase invariance to illumination changes. As in the Lowe's algorithm, first the 128 elements vector are normalized to unit length. This normalization cancels changes in image contrast. Notice that we do not care about changes in brightness, a constant added to pixel values, because they are suppressed by image gradients. Finally, the descriptor is saturated so that no values over 0.2 are allowed and again re-normalized to unit length. This final step is done to reduce non-linear illumination changes.

3. Elastic bunch graph matching

The main idea of EBGM is that a novel face pattern can be recognized by first localizing a set of facial landmarks and then measuring the similarity between these landmarks and those extracted from a set of faces of each individual. Traditionally, EBGM algorithms use Gabor jets as features for both localization and matching of facial landmarks. In our approach, we replace Gabor coefficients by the HOG descriptors presented in Section 2.

Our implementation of EBGM is based on the algorithm developed by Wiskott et al. (1996) which was included by the Colorado State University (CSU) as a baseline algorithm for comparison of face recognition algorithms (Bolme et al., 2003). Basically, this HOG–EBGM algorithm can be decomposed into three steps:

(1) Image normalization.
(2) Creation of face graphs.
(3) Graph matching.

The objective of the image normalization step is to reduce variability produced by changes in illumination, scale and rotation.

The next step creates a face graph after the detection of facial landmarks. Of course, the success of the recognition algorithm depends on a good selection of facial landmarks. More precisely, facial landmarks need to be very distinctive between different people and also be relatively easy to detect in a fully automatic system. Our face graph follows the structure proposed in the CSU project (Bolme, 2003) with 25 facial landmarks which are shown in Fig. 2, the numbers indicate the search order.

It is important to mention, that not all facial areas contribute equally to face recognition. Several studies (Zhao et al., 2003) have shown that the area around eyes and nose are very important for