Pattern Recognition 43 (2010) 1039–1049

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.de/locate/pr

A data mining approach to face detection

Wen-Kwang Tsao, Anthony J.T. Lee *, Ying-Ho Liu, Ting-Wei Chang, Hsiu-Hui Lin

Department of Information Management, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan, ROC

ARTICLE INFO

Article history:
Received 28 May 2007
Received in revised form
25 February 2009
Accepted 3 September 2009

Keywords:
Face detection
Feature pattern
Maximal frequent pattern
Data mining
Support vector machine

ABSTRACT

In this paper, we propose a novel face detection method based on the MAFIA algorithm. Our proposed method consists of two phases, namely, training and detection. In the training phase, we first apply Sobel's edge detection operator, morphological operator, and thresholding to each training image, and transform it into an edge image. Next, we use the MAFIA algorithm to mine the maximal frequent patterns from those edge images and obtain the positive feature pattern. Similarly, we can obtain the negative feature pattern from the complements of edge images. Based on the feature patterns mined, we construct a face detector to prune non-face candidates. In the detection phase, we apply a sliding window to the testing image in different scales. For each sliding window, if the slide window passes the face detector, it is considered as a human face. The proposed method can automatically find the feature patterns that capture most of facial features. By using the feature patterns to construct a face detector, the proposed method is robust to races, illumination, and facial expressions. The experimental results show that the proposed method has outstanding performance in the MIT-CMU dataset and comparable performance in the BioID dataset in terms of false positive and detection rate.

1. Introduction

Face detection is a fundamental problem in many computer vision applications. It can be used to locate a face, and as a front–end for applications such as face recognition system, surveillance and security system, human computer interaction (HCI) system, etc. Face is a highly non-rigid object. The challenge of detecting human faces from an image mostly comes from the variation of human faces such as races, illumination, facial expressions, face scales, head poses (off-plane rotations), face tilting (in-plane rotations), occlusions, etc. Also, environment issues such as lighting conditions, image quality, and cluttered backgrounds may cause great difficulties.

Many face detection methods have been proposed. These methods can be classified into three categories [1]: knowledge-based and feature invariant methods, template matching methods, and appearance-based methods.

The knowledge-based and feature invariant methods locate the features of a face such as eyes, nose, mouth, or even skin color, and then group them together by considering their geometrical relationships [2–4]. Bhuiyan et al. [5] proposed a method to convert pixels in color images from the RGB model to the YIQ model to find skin colors and group them into skin regions. Chiang et al. [3] presented a model to convert pixels from the RGB model to the r–g plane and find skin and lip colors in an image between two parabolas on that plane. Once the skin regions are located in an image, facial components or features like eyes, eyebrows, nose or mouth can be detected within these skin regions. A skin region with proper facial components is reported as a face candidate. Lee et al. [6] used the directional template and blob map to locate major facial features, i.e., two eyes and a mouth. Usually, a verification process such as facial components' geometrical relations is needed to reduce false positives [7]. One major problem about knowledge-based and feature invariant methods is that it is hard to translate human knowledge into well-defined rules. If the rules are too strict, they may fail to detect human faces. On the other hand, if the rules are too general, they may result in too many false positives.

Template matching methods aim to find face features that exist even when the pose or lighting conditions are varying [1]. Usually, a standard face template (pattern) is manually predefined or constructed by a function. The existence of a face is determined based on the similarity or correlation between the input image and the standard face pattern in terms of eyes, nose, mouth, and face contour. Silhouettes have also been used as templates for locating face candidates [8]. Bhuiyan et al. [5] formed the template of a face image, which is obtained from averaging the gradation levels of pixels of face samples. The template face image is then shifted through the whole image to find the location with the most suitable matches. Jesorsky et al. [9] used a face

* Corresponding author.
E-mail addresses: d93725001@ntu.edu.tw (W.-K. Tsao), jtlee@ntu.edu.tw (A.J. Lee), d94725011@ntu.edu.tw (Y.-H. Liu), r94725006@ntu.edu.tw (T.-W. Chang), d8725003@ntu.edu.tw (H.-H. Lin).

0031-3203/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2009.09.005

template to roughly locate a face candidate. Then an eye template is exploited to verify the located face candidates. One disadvantage of template matching methods is that it is not easy to obtain a good template from training face images. Another disadvantage is that it is difficult to enumerate templates for different poses.

In contrast to template matching methods, appearance-based methods learn templates from training samples (faces and non-faces) instead of learning from experts; therefore the training process is usually time-consuming. However the detection speed of these methods is usually fast, sometimes even real-time. This category of methods includes statistical methods [10], neural networks [11–13], support vector machines (SVM) [14–16], and multiple classifier combination [17–19].

Shih and Liu [15] proposed the DFA-SVM method which integrates the discriminating feature analysis, distribution-based face class modeling, and SVM. Heisele et al. [16] built a 5-level hierarchy of SVM classifiers with lower level classifiers rejecting most parts of the background, and the upper level classifiers performed the detailed detection. Then, PCA is applied to the top-level classifier to choose relevant image features. With the combination of a hierarchy of SVM classifiers and feature reduction, the detection process is speeded up in comparison with similar classification methods [20].

Viola and Jones [17] proposed an object detection scheme, which is fast enough for real-time applications and can be effectively applied to face detection. This is accomplished by the integration of a new image representation called integral image, a learning algorithm based on AdaBoost [21], and a cascade scheme is to combine the boosted classifiers. Roth et al. [22] proposed a method, called SNoW, for detecting faces with different features, expressions, and poses under different lighting conditions. Nilsson et al. [23] proposed an extended version of the SNoW classifier by a split-up process. One of the drawbacks of the SNoW-based methods is that they require a large number of face and non-face patches. Cristinacce and Cootes [24] used a joint shape and texture appearance model to generate a set of region template detectors, where a face is detected by aggregating the responses of these detectors.

One problem of the appearance-based methods is that it needs a lot of positive and negative training examples, and the learning process is very time-consuming. Another problem is that the learned result is hard to interpret, thus is difficult to be adjusted to fit various applications.

Of the three categories of methods mentioned above, the disadvantages of these methods are (1) the knowledge and the relation of the features is hard to define, (2) the training process of the appearance-based methods is time-consuming, or (3) a good template is hard to obtain.

Therefore, in this paper, we propose a data mining approach to obtain the features patterns of human faces automatically and efficiently. We first apply Sobel's edge detection operator, morphological operator, and thresholding to each training image, and transform it into an edge image. Next, we use the MAFIA [25] algorithm to mine the maximal frequent patterns from those edge images and obtain the positive feature pattern. Similarly, we can obtain the negative feature pattern from the non-edge images, each of which is a complement of an edge image. Based on the positive and negative feature patterns mined, we construct a face detector containing three cascaded classifiers to prune non-face candidates. That is, we not only focus on the traditional facial feature patterns like eyes or mouth, which are called positive feature pattern, but also focus on the facial regions that do not contain any facial feature patterns like cheeks, which are called negative feature pattern.

Unlike knowledge-based and feature invariant methods, our proposed method can find the feature patterns automatically. Moreover, by using both the positive and negative feature patterns to construct a face detector, our proposed method is robust to races, illumination, and facial expressions.

The rest of the paper is organized as follows. The preliminary concept and problem definitions are described in Section 2. Our proposed method is presented in detail in Section 3. The performance analysis is shown in Section 4. Finally, the conclusions and future work are discussed in Section 5.

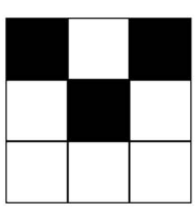

Fig. 1. An example image.

2. Preliminary concept and problem definition

Before describing our proposed method, we define some notations used later.

Window: A window is a fix-sized region in an image.

Item: An item is a pixel in a window. An item is denoted by its coordinate (x,y). That is, the item is located at the xth column and the yth row, where the column and row numbers are started from 0.

Pattern: A pattern in an image contains a set of items, which contains all items' coordinates. For example, a pattern in the image shown in Fig. 1 can be presented as {(0,0), (2,0), (1,1)}.

Super-pattern: If every item in pattern X can be found in pattern Y, we can say that Y is a super-pattern of X or X is a sub-pattern of Y.

Support: The support of a pattern is defined as the percentage of images containing the pattern in the database.

Frequent pattern (*FP*): A pattern is *frequent* if its support is not less than the user-specified minimum support threshold.

Maximal frequent pattern (*MFP*): A frequent pattern X is *maximal* if none of X's super-patterns are frequent.

3. Our proposed approach

Our proposed approach contains two phases: training phase and detection phase.

Fig. 2. Processed image: (a) input image; (b) after histogram equalization; (c) after horizontal edge detection; (d) filtered edge; (e) after dilation.