Pattern Recognition 41 (2008) 3842 – 3859

Contents lists available at ScienceDirect

Pattern Recognition

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

Exploiting Voronoi diagram properties in face segmentation and feature extraction

Abbas Cheddad[a,*], Dzulkifli Mohamad[b], Azizah Abd Manaf[c]

[a]School of Computing and Intelligent Systems, Faculty of Computing and Engineering, University of Ulster, Northern Ireland BT48 7JL, UK
[b]Faculty of Computer Science and Information System, University of Technology Malaysia (UTM), Johor, Malaysia
[c]Faculty of Computer Science, Malaysian Military Academy (ATMA), Kem Sungai Besi, Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 28 January 2008
Received in revised form 15 April 2008
Accepted 9 June 2008

Keywords:
Biometric
Voronoi diagram
Delaunay triangulations
Distance transformation
Features extraction

ABSTRACT

Segmentation of human faces from still images is a research field of rapidly increasing interest. Although the field encounters several challenges, this paper seeks to present a novel face segmentation and facial feature extraction algorithm for gray intensity images (each containing a single face object). Face location and extraction must first be performed to obtain the approximate, if not exact, representation of a given face in an image. The proposed approach is based on the *Voronoi diagram (VD)*, a well-known technique in computational geometry, which generates clusters of intensity values using information from the vertices of the external boundary of *Delaunay triangulation* (*DT*). In this way, it is possible to produce segmented image regions. A greedy search algorithm looks for a particular face candidate by focusing its action in elliptical-like regions. VD is presently employed in many fields, but researchers primarily focus on its use in skeletonization and for generating Euclidean distances; this work exploits the triangulations (i.e., Delaunay) generated by the VD for use in this field. A *distance transformation* is applied to segment face features. We used the BioID face database to test our algorithm. We obtained promising results: 95.14% of faces were correctly segmented; 90.2% of eyes were detected and a 98.03% detection rate was obtained for mouth and nose.

1. Introduction

Computer science is becoming an important central discipline for a variety of scientific fields, leading it to become an increasingly multidisciplinary research area. The capability of taking advantage of advances in computer science to model more "*human-like*" behaviour has granted added value to the field and thereby introduced new research areas. However, the *human visual system* (*HVS*) remains the most complete and efficient-yet-complicated vision system that has ever been seen. In fact, researchers in the area of computer vision are trying to emulate the essential functionalities of the HVS. The fact that a face cannot be described in words has sparked the conceptualization of new technology based on *machine vision*. *Face recognition technology* (*FRT*), an example of that effort, is the study of the process of identifying a face in a given image pattern, segmenting it from the rest of the image (background), and then proceeding to a recognition phase. Indeed, the fact that faces can vary in terms of position, size, skin intensity values, lighting effects, facial expressions, the presence or absence of hair or glasses, and occlusion (faces with hidden part(s)) makes this a challenging exercise. All of these factors contribute to our inability to easily predict the appearances of faces.

Unlike fingerprints, a process such as face recognition could—at least in principle—be used to recognize people "passively," that is, without their knowledge or cooperation. Ref. [1] is a very interesting article which discusses face recognition from an ethical perspective.

The success of a recognition technique depends to a certain extent on two things: the first is minimization of the computational burden and the second is maximization of the speed of the process that leads to an accurate result. One method used by some face recognition researchers is to manually crop the face area from every relevant image [2] . For practical applications, however, this method would not be helpful.

The face recognition process usually requires three steps [3,4]. The first step involves locating the face area, a process known as *face localization*. The second step involves extracting facial features; this step is very critical because the next and final step depends solely on its outcome. The final step is classification of the face image based on the features vector. However, the first and second steps may overlap,

* Corresponding author. Tel.: +44 7907570340.
E-mail addresses: cheddad@gmail.com, cheddad-a@ulster.ac.uk (A. Cheddad), dzulkifli@utm.my (D. Mohamad), azizah07@citycampus.utm.my (A.A. Manaf).

0031-3203/$ - see front matter
doi:10.1016/j.patcog.2008.06.007

and in certain cases face features are extracted first in order to locate the face region. A process in which a face is first detected by a coarse location of facial features and then fine features are subsequently extracted is known as a bottom-up process [5].

A number of algorithms have been proposed to extract facial features based on different approaches. Ref. [6] used a specific 3∗3 mask for edge detection to crop faces from the background. The author's mask operations on the image utilize a closed contour. Images were all shot in a black background and a black cloth was draped around the subject's neck. Therefore, in this case, clipping the faces was a trivial task, and this method is considered one of the more ancient, direct techniques for face segmentation.

Refs. [7,8] used snakes or active contours, and represent a popular method of object boundary detection. In this work, a controlled continuity spline function is used to transform the shape of the curve so as to minimize the energy function relative to the initial state of the curve (energy minimization). The final curve will then mimic the external boundary of the object. The main difficulties in this method are the computational burden, which is quite high, the sensitivity to noise, the fact that a good initial point is very hard to estimate and that the method will always converge to a solution, whether or not this solution is the desired one or not.

Colour (RGB) transformation is another popular method which was recently invoked. It is a very fast algorithm. In Ref. [2] , an RGB colour matrix was nonlinearly transformed to YC_bC_r (intensity, chromatic blue and chromatic red) space. Ref. [9] followed a similar procedure, while Ref. [10] chose the following system to convert from (RGB) to (Y, C_b, C_r):

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} (0.299)(0.587)(0.114) \\ (-0.169)(-0.331)(0.500) \\ (0.500)(-0.419)(-0.081) \end{bmatrix} * \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1}$$

Their algorithm resulted in 85.92% correct face detection. Ref. [11] describes a method for extracting the skin area from an image using normalized colour information. Once the flesh region is extracted, its colour distribution is compared with a manually cropped and constructed model. Ref. [12] obtained a 90% correct detection rate using an identical method. Ref. [13] created a special hardware system; called the triple-band system, which included a near-IR (infra-red) illumination generator. Their idea is based on the premise human skin has a special reflection in the near-IR illumination. In other works (e.g., Ref. [14]) face segmentation has been done using another colour space transformation, namely HSV (hue, saturation and value) and shape information. First, skin-like regions are segmented based on the hue and saturation component information, and then a fine search in each of the regions is performed to detect elliptical shapes. This method normally gives large false alarms, which requires further processing either by considering other information or by making additional use of other techniques. Moreover, generating a skin-colour model is not a trivial task. For example, Ref. [15] used 43 million skin pixels from 900 images to train the skin-colour model and Ref. [16] manually segmented a set of images containing skin regions to generate a skin model.

Template matching technique was among the first pioneered algorithms. It was refined to develop a *deformable template matching* which was implemented to compensate for some of the drawbacks of former methods [17,18]. In their paper [17], Wang and Tan used six templates. Two eye templates and one mouth template were used to verify a face and locate its main features; afterwards, two cheek templates and one chin template were employed to extract the face contour. The performance claimed to be favourable; however, it could not detect faces with shadow, rotation or bad lighting conditions. One thing that they presented as an advantage was that they were able to choose a relatively big step in feature matching so as to reduce the computation cost. While it holds true that the said *big step* reduces the computation burden, it can, however, result in the loss of other significant information. The interesting point here is that the authors admitted this indirectly in their conclusion: "We suspect that one reason for this (failure) is that our template does not include enough information to distinguish faces in very complex background". Most importantly the algorithm gave false positives when rotations occurred, and when there were unwanted objects whose shapes were similar to ellipses. Creating a good template model is not an easy task, however, because it does not (in fact it cannot) take into account the variable appearance of faces. It suffers most when dealing with unknown face size and rotation. Resizing the model and rotating it moreover increases the processing time.

The first advanced neural network approach to report results on a large complex dataset was presented in Ref. [19]. Their system incorporated face knowledge in a "retinally" connected neural network. The neural network was designed to look at windows of 20 ∗ 20 pixels (i.e. 400 input units). This group introduced a hidden layer with 26 units: 4 units to look at 10 ∗ 10 pixel sub-regions, 16 to look at 5 ∗ 5 sub-regions, and 6 look at 20 ∗ 5 pixel overlapping horizontal stripes. The input window was pre-processed for lighting correction (a best fit linear function was subtracted) and histogram equalization. Neural network methods require a lot of training samples in order to increase their efficiency levels. The complexity of the network is considered to be a disadvantage because one cannot know whether the network has "cheated" or not. It is almost impossible to find out how the network determines its answers. Hence, this is also known as a black box model.

Ref. [20] uses the genetic algorithm (GA) to select a pair of eyes from among various possible blocks. The fitness value for each candidate (face) is calculated by projecting it onto the eigenfaces space. Even though GA is accurate, like the "snakes" process it suffers from being a time-consuming technique, because it fires its chromosomes to each and every location.

More relevant to our proposed work is Ref. [5], which suggests a symmetry-based method for face boundary extraction from a binarized facial image. Basically, the work details construction of Delaunay triangulations (DT) (which are the dual of the Voronoi diagram (VD)) from points of an edged image. The property of each triangle was examined geometrically, and the so-called *J-triangle* (junction triangle) was identified. These types of triangles act as linkers to repair the broken edges. The purpose of this is to prevent the face boundary from being merged with the background. This method may experience a heavy computational load if the image size increases; it will also be extremely difficult to deal with images with a complex background. The proposed method has limitations in cases of rotation or if the face of interest is wearing glasses, as stated in the paper. It is also sensitive to noise and the presence of a beard. The demonstrated success rate for detection of facial features upon segmentation is 89%. In contrast, as we will show later, our proposed method employs the VD and its properties on a select few unique feature points derived from the image histogram rather than on points associated with edges in the special domain. Therefore, the computational process for our method is much less than that of the method proposed in Ref. [5].

For an additional detailed survey of the different techniques, we direct the reader to the literature [21], as well as to Ref. [22], in which the recognition phase is discussed, for a more in-depth study.

1.1. Image content segmentation

Ref. [23] describes an interesting survey on colour image segmentation and presents tables comparing the advantages and