Signal Processing: *Image Communication* 36 (2015) 179–189

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

A face detection and location method based on Feature Binding

Jing Jin [a], Bin Xu [a], Xiaoliang Liu [a], Yuanqing Wang [a,*], Liqun Cao [b], Lei Han [b], Biye Zhou [b], Minggao Li [b]

[a] *Nanjing University, School of Electronic Science and Engineering, No. 163, Xianlin Dadao, Nanjing 210046, China*
[b] *Navy General Hospital, Beijing 100088, China*

ARTICLE INFO

Article history:
Received 23 December 2014
Received in revised form
13 May 2015
Accepted 17 June 2015
Available online 7 July 2015

Keywords:
Feature Binding
Face detection
Face location
Pattern recognition
Human–machine interaction

ABSTRACT

A face detection and location method based on Feature Binding (FB) is proposed in this paper. The features used for face detection and location are classified and bound into groups. The information of each group is extracted separately during face detection. Through the combination with the constraint relationship, the precise location of the face in the image could be identified by confidence coefficients of all groups. Experimental results show that this proposed method can improve the accuracy rate obviously and has good detection effect on obscured faces. Besides, FB can be good to adapt to varieties of features.

1. Introduction

With the development of image processing and pattern recognition, the demands for intelligent processing have been growing. As a hotspot in the field of pattern recognition, face detection has been widely used in applications, such as authentication, attendance system and electronic passport. Besides, it has also been applied in a new naked-eye auto-stereoscopic display which requires accurate human face and eye locations [1]. Therefore, the accuracy of face detection and location is particularly important.

The primary task of face detection is to identify whether there is a certain face in a given image or image sequence, while that of face location is to calculate the face details including the position, size, quantity and spatial distribution. However, it is challenging to detect a face from an image with complex background because of the varying characteristics in faces such as scales, positions, orientations, and postures, as well as different facial expressions and light conditions.

With the increasing knowledge of human face, different kinds of face detection algorithms have been proposed. Conventional methods of face detection could be classified into four categories: a knowledge-based method, a feature invariant method, a template matching method and a statistical-based method. Knowledge-based method is a top-down method. It encodes human faces by regular database, which is formed by typical faces, and develops a series of criteria according to the relationships among the facial features. Faces can be detected when these criteria are met in the test region. Yang proposed a mosaic diagram in 1994 [2]. Afterwards, modified mosaic diagram [3] and a classification method [4] are proposed respectively. However, these methods strongly rely on the priori knowledge and have low recognition rate for changing faces. Different from the knowledge-based method, the feature invariant method is a down–up method, which includes a morphological method [5], a random labeled graph method [6] and a wavelet decomposition method [7]. The objective of these methods is to find invariable structural features under the variation of postures, perspectives or illumination conditions. Hence, the face could be detected and located according to these features.

* Corresponding author. Tel.: +86 13913977866.
E-mail address: yqwang@nju.edu.cn (Y. Wang).

http://dx.doi.org/10.1016/j.image.2015.06.010
0923-5965/

Nonetheless, these features can be seriously damaged due to illumination, noise or shadows. It would be difficult to carry out these methods when the boundaries of the features are weakened or the edges of the shadow are intensified. Template matching method is a method which firstly stores several standard face modes to describe the whole face and facial features respectively, and then calculates the relationship between the input image and stored modes [8]. Haar feature [9] is one of the most common used feature which uses gray and gradient information of image. The classic deformable templates are Active Shape Models (ASM) [10,11] and Active Appearance Models (AAM) [12]. Considering the deflection of face, Active Performance Model (APM) which is based on appearance [13] and Smooth Statistical Shape Model (SSSM) [14] are developed. Nevertheless, template matching methods involve a large amount of calculation and have low recognition rate for the faces with scale and rotation transformations. Taking full advantage of statistical analysis and machine learning, statistical-based method obtains statistical features of positive and negative samples from training images so that the classifiers could be built to detect faces. Statistical-based method is related to many kinds of methods, such as subspace method [15], Neural Network (NN) [16,17], Support Vector Machine (SVM) [18], Hidden Markov Model (HMM) [19], and a Boosting method [20]. Based on these methods, face detection systems could get good performance in terms of accuracy and detection speed, as well as good effect in circumstances with multi-posture faces. However, there are many requirements for training samples. For example, the accuracy of these methods depends on each training feature. At the same time, the long training time is also an influence factor.

Due to the disadvantages of the methods mentioned above, it could be an imperative requirement to find a "biological visual" recognition method. In the theory of visual perception, there is a "binding problem" that concerns the way in which people select and integrate the separate features of objects in the correct combinations [21]. Therefore, the concept of Feature Binding (FB) in the field of pattern recognition is proposed by combining with the theory of visual sensing. It is a method based on feature subspace. The features on a human face are grouped according to their locations, and the cluster of all the features in each group is regarded as a "feature set". During face detection, each feature set is detected respectively at first. The presence of a certain human face could be confirmed once the amount of constrained feature sets have reached the threshold, and the face can be located according to the locations of these feature sets. The results show that the accuracy of face location algorithm with FB could be improved more obviously compared with the traditional method. Meanwhile, the detection process can be accelerated and protection from the external interference better with this method. Besides, it has also shown good detection results for partial obscured faces in practical applications, because it only needs to reach the thresholds of countable feature sets rather than all of them. In addition, FB is an adaptive and robust method which can be applied to a variety of features rather than a particular one. Moreover, in circumstances with many kinds of features mentioned above, such as Haar feature and ASM feature, FB can be utilized to obtain better detection results.

The paper is organized as follows. In Section 2, the concept of Feature Binding is introduced. Section 3 describes a case used for Haar feature in which the advantage of the proposed approach would be demonstrated. Experiments and another case which is used for ASM are shown in Section 4, and Section 5 provides some concluding remarks.

2. Concept of Feature Binding

2.1. Feature Binding in biological vision

In biological perception theory, "feature" means an irreducible attribute of an object such as its color, orientation, form, and motion, and "binding" is the dynamic linkage of multiple features leading to the perception of a given object as a coherent and unified whole [22].

Recent findings have suggested that a primate brain codes perceived events in a distributed fashion, which are integrated into object files – episodic bindings of object-related information. Hommel put forward that the brain addresses these problems by creating multi-layered networks of bindings-"event files". These bindings produce systematic but often surprising and counter-intuitive interactions between perception and action planning [23] as well as their impairments.

Researchers in neurophysiology field have discovered that vision cells of visual pathway at all levels generally have receptive field property [24]. According to the property of reception field, cells on the visual cortex can be divided into simple cells, complex cells and hypercomplex cells. Simple cells, which are also called as direction selectivity cells, are suitable for the detection of contrasty straight edge. Complex cells have larger receptive field than simple cells. Besides, they have certain directions and shift invariance properties which are good for invariant features detection. There are some requirements for the length of the strip stimulation of hypercomplex cells, so that the optimal stimulus which can cause a strong reaction is the endpoint or the inflection point with a certain directivity. Hubel and Wiesel put forward the famous Receptive Field Level hypothesis through the research on visual cortex cells, which assumes that the receptive field of senior neuron is converged orderly by many lower neurons [25,26].

2.2. Feature Binding in pattern recognition

In the theory of computer vision, it is a common method to use local area as the processing element. Local areas have been divided artificially and regularly in previous algorithms, which lead to dispersive information and inconspicuous regional characteristics. In order to get a relatively large difference between each two local areas and obtain similar features in each area at the same time, the image shall be divided according to its own feature attributes. In the field of pattern recognition, this problem can be solved through feature subspace. With this method, test image is projected to varied feature spaces and the optimal features in these spaces best capable of distinguishing samples are selected. Besides, these features could constitute a feature subspace with certain rules [27,28].

From the perspective of information theory, the image can be divided into redundant part and mutational part according to its composition. People are often more