Applied Computing and Informatics (2014) **10**, 1–13

King Saud University
Applied Computing and Informatics
www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

Effective and precise face detection based on color and depth data

CrossMark

Loris Nanni [a,*], Alessandra Lumini [b], Fabio Dominio [a], Pietro Zanuttigh [a]

[a] *DEI, University of Padova, via Gradenigo, 6, 35131 Padova, Italy*
[b] *DISI, University of Bologna, via Venezia, 52, 47521 Cesena, Italy*

Received 3 February 2014; revised 31 March 2014; accepted 7 April 2014
Available online 21 April 2014

KEYWORDS

Face detection;
Skin detection;
Depth map;
Viola–jones detector

Abstract In this work an effective face detector based on the well-known Viola–Jones algorithm is proposed. A common issue in face detection is that for maximizing the face detection rate a low threshold is used for classifying as face an input image, but at the same time using a low threshold drastically increases the number of false positives. In this paper several criteria are proposed for reducing false positives: (i) a skin detection step is used to reject a candidate face region that does not contain the skin color, (ii) the size of the candidate face region is calculated according to the depth data, removing the too small or the too large faces, (iii) images of flat objects (e.g. candidate face found in a wall) or uneven objects (e.g. candidate face found in the leaves of a tree) are removed using the depth map and a segmentation approach based both on color and depth data.

The above criteria permit to drastically reduce the number of false positives without decreasing the detection rate. The proposed approach has been validated on three datasets composed of 180 samples including both 2D and depth images. The face position inside samples has been manually labeled for testing.

* Corresponding author. Tel.: +39 3493511673.
E-mail address: nanni@dei.unipd.it (L. Nanni).

Peer review under responsibility of King Saud University.

ELSEVIER Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.aci.2014.04.001

A Matlab version of the system for face detection and the full testing dataset will be freely available from http://www.dei.unipd.it/node/2357.

1. Introduction

Face detection has attracted the attention of many research groups due to its widespread application in many fields as surveillance and security systems, as human–computer interface, face tagging, behavioral analysis, content-based image and video indexing, and many others (Zeng et al., 2009). Face detection is the first crucial step for facial analysis algorithms (i.e. face recognition/verification, head tracking, and facial expression recognition) whose goal is to determine whether or not faces are present in an image and eventually return their location and extent (i.e. a bounding box). It is a more challenging problem than face localization in which a single face is assumed to be inside an image.

Most of the literature in this field deals with frontal face detection from two-dimensional (2D) images: the problem is often formulated as a two-class pattern recognition problem aimed at classifying each sub-window of a given size of the input image as either containing or not containing a face (Jin et al., 2004). Then the classification is performed by common technologies for 2D facial recognition such as Eigenface, Fisherface, waveletface, PCA (Principal Component Analysis), LDA (Linear Discriminant Analysis), Haar wavelet transform, and so on. The Viola–Jones detector (Viola and Jones, 2001) is probably the most famous approach for frontal 2D detection: it involves exhaustively searching an entire image for faces, with multiple scales explored at each pixel using Haar-like rectangle features boosting classification. Two different face detection strategies based on slightly modified Viola–Jones are proposed in Anisetti (2009). In Küblbeck and Ernst, 2006 boosting has also been used in conjunction with Modified census transform (MCT), to improve illumination invariance. In (Huang et al., 2007) a method able to detect faces with arbitrary rotation-in-plane and rotation–off-plane angles in still images or video sequences is proposed. In (Jianxin et al., 2008) is designed a classifier that explicitly addresses the difficulties caused by the asymmetric learning goal (the minority class is the face class).

Despite the success of these and of several other methods, designed to provide accurate detection performance under variable conditions (Zhang and Zhang, 2010), most of difficulties in precise face detection still arises in the presence of illumination changes and occlusions. One possible way to improve algorithms for face detection is to incorporate models of the image processing that efficiently integrates multiple cues, such as stereo disparity, texture and motion.

For example, Microsoft Kinect is a depth sensing device that couples the 2D RGB image with a depth map (RGB-D) which can be used to determine the depth of every object in the scene. Each pixel in Kinect's depth map has a value