Pattern Recognition 52 (2016) 160–173

Contents lists available at ScienceDirect

Pattern Recognition

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

A realistic virtual environment for evaluating face analysis systems under dynamic conditions

Mauricio Correa [a], Javier Ruiz-del-Solar [a,b,*], Rodrigo Verschae [a,1]

[a] Advanced Mining Technology Center, Universidad de Chile, Chile
[b] Department of Electrical Engineering, Universidad de Chile, Av. Tupper 2007, 837-0451 Santiago, Chile

ARTICLE INFO

Article history:
Received 14 November 2014
Received in revised form
7 September 2015
Accepted 12 November 2015
Available online 22 November 2015

Keywords:
Face analysis
Face recognition
Face
recognition benchmark
Evaluation methodologies
Virtual simulation environment
Simulator

ABSTRACT

This paper proposes a new tool for the evaluation of face analysis systems under dynamic experimental conditions. The tool primarily consists of a virtual environment where a virtual agent (e.g., a simulated robot) carries out a face analysis process (e.g. face detection and recognition). This virtual agent can navigate in the virtual environment, where one or more subjects are present, and it can observe the subjects' faces from different distances and angles (yaw, pitch, and roll), and under different illumination conditions (indoor or outdoor). The current view of the agent, i.e. the image that the agent observes, is generated by composing real face and background images acquired prior to their usage in the virtual environment. In the virtual environment, different kinds of agents and agents' trajectories can be simulated, such as an agent navigating in a scene with people looking in different directions (mimicking a home-like environment), an agent performing a circular scanning (such as in a security checkpoint), or a camera-based surveillance system observing a person. In addition, during the recognition process the agent can actively change its viewpoint seeking to improve the recognition results. The proposed tool provides to the developer all functionalities needed to build the evaluation scenario: a set of real face images with real background information, a virtual agent with navigation capabilities, a scenario configuration (number, position and pose of the subjects to be observed), an agent trajectory definition, the generation of the simulated agent's view-dependent images, some basic active vision mechanisms, and the ground truth data (e.g. face id and pose for every observation), allowing the evaluation of face analysis methods under realistic conditions. Three usage examples are presented: the study of the robustness of face detection and face recognition methods under pose variations, and the evaluation of an integrated face analysis system to be used by a service robot. The proposed methodology may be of interest for researchers and developers of face analysis methods, in particular in the robotic and biometrics communities.

1. Introduction

Face analysis[2] plays an important role in building computer vision systems, HRI (Human–Robot Interaction) systems, and in general in any system that uses vision to interact naturally with humans or to process information of humans in a given scene. Human detection and human identification based on face information are key abilities of intelligent machines whose purpose is to interact with humans. Face analysis is also very important in surveillance applications in dynamic environment, such as security cameras at airports, and is also being included in consumer electronics, such as face detection and smile detection in cameras. Evaluating face analysis systems for such environments and conditions is not straightforward, in particular, in the cases where the recognition system uses active vision mechanisms to change its viewpoint or position in the scene.

A very important aspect in the development of face analysis methodologies is the use of suitable databases, and reproducible testing and training methodologies. For instance, the well-known FERET database [1], has been very important in the development of face recognition algorithms for controlled environments. However, neither FERET nor other relatively new databases such as LFW [2], CAS-PEAL [3] and FRGC [4,5], among others [6–8], are

* Corresponding author at: Department of Electrical Engineering, Universidad de Chile, Av. Tupper 2007, 837-0451 Santiago, Chile. Tel.: +56 2 2977 1000/ +56 2 2978 4207; fax: +56 2 6720162.
E-mail addresses: macorrea@ing.uchile.cl (M. Correa), jruizd@ing.uchile.cl (J. Ruiz-del-Solar), rodrigo@verschae.org (R. Verschae).

[1] Rodrigo Verschae is now with the Graduate School of Informatics, Kyoto University, Japan.

[2] Face analysis is understood as any computational procedure related to the analysis of faces, with the most common procedures being detection, recognition, alignment and expression recognition, but many other exist, such as super-resolution, gender classification, smile detection, pose estimation, blink detection, kindness recognition, etc.

http://dx.doi.org/10.1016/j.patcog.2015.11.008
0031-3203/

able to provide real-world testing conditions for evaluating face recognition systems that include the use of innovative mechanisms such as spatiotemporal context and active vision, which are required in applications that consider the dynamic interaction with humans in the real world. Even the use of video face databases (e.g. [9–12]) does not allow testing the use of those ideas, because the video sequence is taken using pre-defined viewpoints. The use of a virtual face simulator could allow accomplishing the changes in viewpoints. However, such a simulator would not be able to generate faces and backgrounds that look real/natural enough, which is an important requirement for the realistic testing of face recognition systems.

Nevertheless, the combined used of a simulation tool with real face images and background images taken under real-world conditions could allow accomplishing the goal of providing a tool for testing face recognition systems under uncontrolled, dynamic conditions. In this case, more than providing a database and a testing procedure, the idea would be to supply a virtual environment that offers a database of real face images and real background images, a simulated virtual environment, a virtual agent moving in that environment, active vision mechanisms for the virtual agent, predefined benchmark problems, ground truth data, and an evaluation methodology.

The main goal of this paper is to provide such a virtual environment. In this environment, virtual subjects are located at different positions and with different orientations in a virtual map. Inside the virtual environment, a virtual agent (a virtual entity with the ability to detect, recognize and analyze faces) can navigate and observe face images from different distances and angles (yaw, pitch, and roll). The current view of the agent, i.e. the image that the agent observes, is generated by the virtual environment using real face images previously acquired in indoor and outdoor variable lighting conditions with several pitch and yaw angles (in-plane rotations can be simulated by software), as well as real background images. In the virtual environment, different kinds of agents and agents' trajectories can be simulated, such as an agent navigating in a scene with people looking in different directions (mimicking a home-like environment), an agent performing a circular scanning (such as in a security checkpoint), or a camera-based surveillance system observing a person. In addition, during the recognition process, the virtual agent can actively change its viewpoint seeking to improve the recognition results.

We believe that the proposed methodology and evaluation tool are of interest to researchers involved in development and testing of applications related with the visual analysis of human faces. Its use allows comparing, quantifying and validating face analysis capabilities of agents, and in general intelligent machines, under dynamic working conditions. One of its more relevant features is that it allows repeatability of the experiments. Therefore, it allows the comparison and evaluation of one or more algorithms without damaging the moving agent (e.g. the robot), and with short evaluation times. In the current work we focus on face recognition and detection, although the use of the tool is straightforward in other face analysis problems, such as pose estimation, gender classification, and age estimation.

It is worth mentioning that a special acquisition device was designed and built to acquire face and background images under different view angles, which are essential for the operation of the virtual environment. The simplicity and modularity of the device allows its rapid deployment and use in real-world locations such as streets, gardens, shopping malls, etc.

This article is organized as follows. First, related work on existing face analysis and evaluation methodologies is outlined (Section 2). Afterwards we describe the proposed virtual environment (Section 3), where we give a detailed description of its different modules. Later, we present some usage examples of the proposed system (Section 4), to finally conclude (Section 5).

2. Related work

The availability of standard databases, benchmarks, and evaluation methodologies is crucial for the appropriate development and comparison of face analysis systems. There is a large number of face databases and associated evaluation methodologies that consider different number of subjects, camera sensors, and image acquisition conditions, and that are suited to test different aspects of the face recognition problem such us illumination invariance, aging, expression invariance, etc. (e.g. the surveys and comparatives studies [13–17,46]). An overview and basic information about existing face databases can be found in [7,18]. Although some new databases (e.g. LFW database [2] and Photoface database [19]) are designed to include real-world images, most databases and evaluation protocols (including LFW database [2] and Photoface database [19]) are designed to test methods using images captured by static cameras. Also, similar methodologies are commonly used in face recognition infrared images [20,21].

Out of the existing databases for face recognition, probably the most well known is the FERET database [1] and its associated evaluation methodology, which has become the standard choice for evaluating face recognition algorithms under controlled conditions. Alternative popular databases used with the same purpose are the Yale Face Database [22] and BioID [23]. Other databases, such the AR Face Database [24], ORL database [47] and the University of Notre Dame Biometrics Database [25], include faces with different facial expressions, illumination conditions, and occlusions. However, from our point of view, all of them are far from considering real-world conditions.

The Yale Face Database B [26] and PIE [27] are the most utilized databases to test the performance of algorithms under variable illumination conditions. The Yale Face database contains 5 760 single light source images of 10 subjects, each seen under 576 viewing conditions (9 poses $\times$ 64 illumination conditions). For every subject in a particular pose, an image with ambient (background) illumination was also captured. PIE is a database containing 41,368 images of 68 people, each person under 13 different poses, 43 different illumination conditions, and with 4 different expressions. Both databases consider only indoor illumination.

The LFW database [2] consists of 13,233 face images of 5749 different subjects, obtained from news images by means of a face detector. There are no eyes/fiducial point annotations; the faces were just aligned using the output of the face detector. The images have a very large degree of variability in the face's pose, expression, age, race, and background. However, given that the LFW images are obtained from news, which in general are taken by professional photographers, the images are obtained under good illumination conditions, and mostly in indoors.

FRGC ver2.0 database [5] consists of 50,000 face images divided into training and validation sets of controlled and uncontrolled images. The uncontrolled images were taken under varying illumination conditions in indoors and outdoors. Each set of uncontrolled images contains two expressions, smiling and neutral.

The Photoface database [19] is a database of 3D faces, which consist of 3187 sessions of 453 subjects, captured in two recording periods of approximately six months each. The Photoface device was located in an unsupervised corridor allowing real-world and unconstrained capture. Each session comprises four differently lit colour photographs of the subject, from which surface normal and albedo estimations can be calculated. This allows for many testing scenarios and data fusion modalities. Eleven facial landmarks have been manually located on each session for alignment purposes. Additionally, metadata such as gender, facial hair, pose and expression is available.

The EURECOM Kinect Face Dataset [28] consists of multimodal facial images of 52 people (14 females, 38 males) acquired with a Kinect sensor. In each session images are collected according to different facial expressions, lighting and occlusion conditions: neutral,