A Virtual Environment Tool for Benchmarking Face Analysis Systems

Mauricio Correa[+,*], Javier Ruiz-del-Solar[+,*], Rodrigo Verschae[*]
[+]Department of Electrical Engineering, Universidad de Chile
[*]Advanced Mining Technology Center, Universidad de Chile
{macorrea, jruizd}@ing.uchile, rodrigo@verschae.org

Abstract. In this article, a virtual environment for realistic testing of face analysis systems under uncontrolled conditions is proposed. The key elements of this tool are a simulator, and real face and background images taken under real-world conditions with different acquisition conditions, such as indoor or outdoor illumination. Inside the virtual environment, an observing agent, the one with the ability to recognize and detect faces, can navigate and observe the face images, at different distances, and angles. During the face analysis process, the agent can actively change its viewpoint and relative distance to the faces in order to improve the recognition results. The virtual environment provides all behaviors to the agent (navigation, positioning, face's image composing under different angles, etc.), except the ones related with the analysis of faces (detection, recognition, pose estimation, etc.). In addition we describe different kinds of experiments that can be implemented for quantifying the face analysis capabilities of agents and provide usage example of the proposed tool in evaluating a face recognition system in a service robot.

Keywords: Face analysis**,** Face Recognition, Face Recognition Benchmark, Evaluation Methodologies, Virtual Simulation Environment, Simulator.

1 Introduction

Face analysis plays an important role in building HRI (Human-Robot Interaction). 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 security applications in dynamic environment, such as security cameras at airports. Evaluating face analysis systems for such environments and conditions is not straightforward, in particular in the cases when 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 evaluation and training methodologies. For instance, the very well known FERET database [9], has been very important in the development of face recognition algorithms for controlled environments in the last years. However, neither FERET nor other relatively new databases such as LFW [5], CAS-PEAL [4] and FRGC [10][8], among others [1][2][3], are 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 does not allow testing the use of those ideas. The use of a simulator could allow accomplishing this (viewpoint changes). However, a simulator is not able to generate faces and backgrounds that looks real/natural enough, which is a condition for the realistic testing of face recognition systems.

Nevertheless, the combined used of a simulation tool with real face and background images taken under real-world conditions could allow accomplishing the goal of providing a tool for testing face recognition systems under uncontrolled 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, a simulated virtual environment, a simulated agent moving in that environment, dynamic image's acquisition conditions, active vision mechanisms, predefined benchmark problems, and an evaluation methodology. The main goal of this paper is to provide such a virtual environment. In this environment: persons are located at different positions and orientations, where the face images are previously acquired under different pitch and yaw angles -- in-plane rotations can be simulated by software --, in indoor and outdoor variable lighting conditions. Inside this environment, an observing agent, the one with the ability to recognize faces, can navigate and observe the real face images (with real background information), at different distances, angles (yaw, pitch, and roll) and with indoor or outdoor illumination. During the recognition process the agent can actively change its viewpoint to improve the face recognition results. In addition, different kinds of agents and agents' trajectories can be simulated, such as an agent navigating in an scene with people looking in different directions (mimicking a home environment), an agent performing a circular scanning (such as in a security checkpoint), or a person approaching to a security camera.

The proposed virtual environment could be of high interest in the development and testing of applications related with the visual analysis of human faces. It allows comparing, quantifying, training and validating face analysis capabilities of agents, and in general intelligent machines, under exactly equal working conditions. In the current communication we focus on face recognition, although its use in others face analysis problems is straightforward. The simulator will be made available for academic use in the near future.

This article is organized as follows. In the following subsection, related work in face databases and evaluation methodologies is outlined. In Section 2, the proposed virtual environment is described, as well as the functionality that the agent should provide. An example of the application of this tool is presented in Section 3. Finally, we conclude in Section 4.

Related Work

The availability of standard databases, benchmarks, and evaluation methodologies is crucial for the appropriate comparison of face recognition systems. There is a large amount of face databases and associated evaluation methodologies that consider different number of persons, 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. Basic information about face databases can be found in [2][11].

The FERET database [9] and its associated evaluation methodology is a standard choice for evaluating face recognition algorithms under controlled conditions. Other popular databases used with the same purpose are Yale Face Database [12] and BioID [13]. Other databases such the AR Face Database [14] and the University of Notre Dame Biometrics Database [15] 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 [16] and PIE [17] are the most utilized databases to test the performance of algorithms under variable illumination conditions. Yale Face contains 5,760 single light source images of 10 subjects, each seen under 576 viewing conditions (9 poses x 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 [5] consists of 13,233 images faces of 5,749 different persons, 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, due to LFW images are obtained from news, which in general are taken by professional photographers, they are obtained under good illumination conditions, and mostly in indoors.

FRGC ver2.0 database [8] consists of 50,000 face images divided into training and validation sets. The validation set consists of data from 4,003 subject sessions. A subject session consists of controlled and uncontrolled images. The uncontrolled images were taken in varying illumination conditions in indoors and outdoors. Each set of uncontrolled images contains two expressions, smiling and neutral.

2 Proposed Simulation and Testing Tool

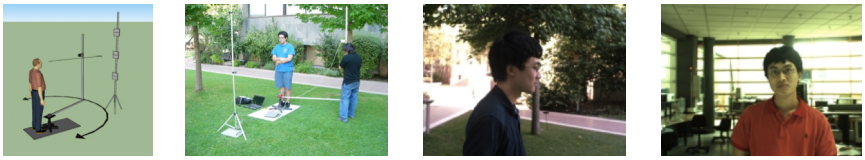
The proposed simulator allows that an observing agent navigate inside a virtual environment (VE), and observe a set of persons. The faces of each of these persons are previously scanned under different yaw and pitch angles, and under different indoor and outdoor illumination conditions. This allows that every time that the agent observes a person's face at a given distance and viewpoint inside the VE, the corresponding images/observations can be composed using real faces and background images, instead of being generated by the simulator.

Image Acquisition and Database Construction

Real face images are acquired at different yaw and pitch angles using a CCD camera mounted in a rotating structure (see Fig. 1a). The person under scan is in a still position, while the camera, placed at the same height than the person's face and at a fixed distance of 140 cm, rotates in the axial plane (the camera height is adjustable). An encoder placed in the rotation axis calculates the face's yaw angle. There are not restrictions on the person's facial expression. The system is able to acquire images with a resolution of 1°. However, in this first version, images are taken every 2°. The

scanning process takes 25 seconds, and we use a 1280 x 960 pixels CCD camera (DFK 41BU02 model). In the frontal image, the face's size is about 200x250 pixels.

Variations in pitch are obtained by repeating the described process with the different required pitch angles. In each case, the camera height is maintained, but the person looks at a different reference points in the vertical axis, which are located at 160 cm in front of the person (see Fig. 1a). In our experience, pitch angles of -15°, 0°, and 15° give account of typical human face variations. In addition, background images for each location, camera-height, and yaw-pitch angle combination are taken with the acquisition device, in order to be able to compose the final real images to be shown to the agent. In Fig. 1(c-d) are shown some images taken with the device.



(a) (b) (c) Yaw: 90°, Pitch: -15 Outdoor (d) Yaw: 0° Pitch: 15, Indoor

Figure 1. (a) Diagram of the acquisition system. (b) The system operating in outdoors. (c)-(d) Examples of images taken using the device in outdoors and indoors.

It is important to remark that the acquisition device is portable (it does not require any special installation), and therefore it can be used at different places. Thus, the whole acquisition process can be carried out at different locations (street environment, laboratory environment, mall environment, etc.). In our case we use at least two different locations for each person, one indoor (laboratory with windows), and one outdoors (gardens inside our school's campus).

With the acquired images a face database containing 50 persons is built. For each person, 726 registered face images (121x3x2) are stored. The yaw angle range is -120° to 120°, with a resolution of 2°, which gives 121 images. For each different yaw, 3 different pitch angles are considered (-15°, 0°, and 15°). For each yaw-pitch combination, indoor and outdoor images are taken. In addition, background images corresponding to the different yaw-pitch angles, place and camera-height combinations are also stored in the database.

Virtual Environment and Agent Navigation

By means of the acquired image DB, a virtual environment is implemented (See Fig 2). A virtual environment is defined by *N* subjects on a Global Map. An observing agent with the ability to analyze (e.g. detect and recognize) faces has the possibility of navigating and making observations inside this scenario. In the virtual scenario, the simulator can generate faces at different distances, angles and light conditions (indoors or outdoors). During the face recognition process, the agent can move in the map to change its point of view and distance to the subject using methods of active vision in order to modify its observations and improve its face recognition results.

The virtual environment offers all the functions that the agent could have in a real scenario (navigation, positioning), and generates images observed by the agent at a given time, images that can contain faces at different angles, etc.

Given (X_A, Y_A, θ_A), where (X_A, Y_A) is the current position of the agent and (θ_A) the current orientation of the agent in the Global Map, for navigating in the

environment, the agent can position itself using the following functions provided by the navigation module of the virtual scenario:

- $MoveAgent(\Delta x, \Delta y)$ commands the agent to move in the Global Map relative to its current position. The final pose of the agent would be $(X_A + \Delta x, Y_A + \Delta y, \theta_A)$.
- $TurnAgent(\Delta\theta)$ commands the agent to turn in the Global Map relative to its current orientation. The final pose of the agent would be $(X_A, Y_A, \theta_A + \Delta\theta)$.
- $SetAgentPosition(x, y, \theta)$ allows the agent to set its position in the Global Map. The final pose of the agent would be (x, y, θ).

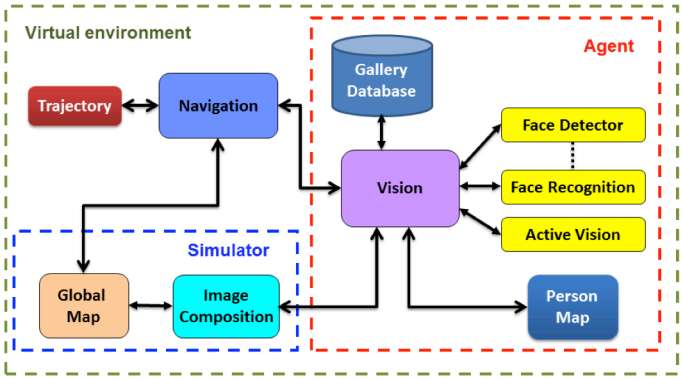


Figure 2. Diagram of the virtual environment

Given the relative position and orientation between the agent and the observed subject (information stored in the Global Map), the simulator generates the corresponding image as seen from the agent's point of view. The image generation processes more than a rendering process is an image composition process, in which real face and background images, acquired using the device described previously, are used to compose the scene as observed by the agent. This is implemented by the *Image Composition* module. This module reads from the Global Map the position and pose of the agent, as well as the position and pose of the subjects and generates the observation for the agent. Subject's out-of-plane rotations are restricted to the available face images in the sagittal and lateral planes, while there are no restrictions for in-plane rotations. If desired, the simulator can also generate noisy observations, considering the uncertainness in the movement of the agent. Every time the agent changes its pose, the simulator generates/composes the corresponding images if is required by the vision module. For instance, Figure 3 shows a given sequence of agent poses, and the corresponding images composed by the simulator.

The agent, at any given, can navigate to the next point on a defined trajectory (if it is using one) or use active vision to modify its observations and improve face recognition results. Different trajectories are implemented, and described below.

The virtual environment provides to the agent a *Person Map* to store information of the different individuals detected within the environment, thought the agent can use its own implementation. The main idea of this module is to store the position of subjects, and to use the information when a person has been detected and included in the Person Map. Given a Global Map and a detected face in a generated image, the distance from the person to the agent is estimated. Using the agent's current position and this estimation, the position of the subject on the global map is estimated. Then,

each detected person is stored at $(X_i, Y_i, \theta_i, E_i)$ with, (X_i, Y_i) position on the map, θ_i the pose and E_i is the estimation error of the position.

In addition, the virtual environment provides a *Gallery* module that the agent can use to store face information of the subjects seen so far (and the corresponding ID in the Person Map) in order to perform the recognition. This gallery can be build online or offline depending on the experiment being performed.

The *Trajectory* module allows to define different kinds of movements the agent can perform in the virtual environment, allowing to define different kinds of scenarios. It is implemented has a sequence of points (X_i, Y_i) that the agent has to flow or that the agent can use as reference.

(a) $x = 100, y = -60, \theta = -30$ (b) $x = 60, y = 65, \theta = 48$ (c) Agent in different positions and observed face

Figure 3. Example of agent's positioning (c), and observations composed by the simulator. The agent is located in (a) and moves to position (b); both generated images are shown.

Details about the *Trajectory* module are given below.

Some face analysis modules (face detection, face recognition, face alignment, face gender classification, etc.) and the active vision module are available in the virtual environment, but can be redefined by the user to evaluate their own modules. For example, if the user wants to evaluate a face recognition module, he can make use of face detector that uses the ground truth (perfect face detection), one that uses a Viola&Jones detector [19] or one implemented by the user himself.

Trajectory Generation

The virtual environment provides three different kinds of trajectories: constrained navigation, predefined navigation and free navigation. Constrained navigation is used to simulate agents that cannot actively move in the environment. Predefined navigation simulates agents that can move freely, but that have a predefined route that could be followed or not. Free navigation allows the agent to move freely and it does not provide any predefined route to the agent.

- Constrained navigation: the agent must follow a predefined constrained trajectory. Five kinds of these trajectories are provided (see Figure 4 (a-e)):
  a) Frontal: the agent approaches the subject having a frontal view of him.
  b) Site-to-side: the agent moves perpendicular to an imaginary line coming from the observed person.
  c) Circular: the agent moves around a person at a fixed distance.
  d) Strafe: This movement is a mix of Frontal and Side-to-side. The agent moves with respect to an imaginary line that is not perpendicular to the frontal view of the subject. The movement does not maintains a fixed distance as the agent is approachs the subject.
  e) Random: the agent places randomly in front of the subject, positioning inside

a rectangle in front of the subject.

- Predefined navigation with active vision: the agent follows a predefined trajectory, but it can position itself freely at any position in the map (Fig. 4 (f)).
- Free navigation: the agent can move freely in the map.

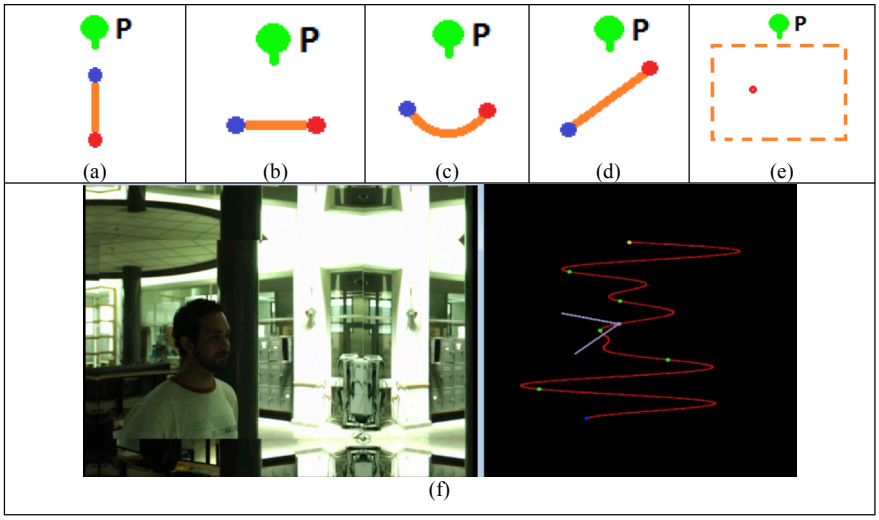


Fig. 4. Examples of the implemented trajectories. See main text for details.

Using these trajectories it is possible to simulate different kinds of scenarios, such as a service robot navigating in a house, a static security camera (with moving subjects), or a scanning device performing a circular moving around the subject. In addition we have included the option of simulating occlusions in the maps.

3 Testing Methodology

In order to recognize faces properly, the agent needs to have the following modules:

1. *Face Detection*: The agent detects a face (i.e. the face region) in a given image.
2. *Face Pose Estimation*: The agent estimates the face's angular pose in the lateral, sagittal and coronal plane.
3. *Active Vision*: Using information about the detected face and its pose, and information observed in the input images, the agent can take actions in order to change its viewpoint for improving face's perception.
4. *Face Recognition*: The identity of the person contained in the face image is determined. The module can include abilities such as face alignment or illumination compensation.
5. Other modules could be also used. For example a gender classifier could be also evaluated, thought this functionality has not been implemented yet (the database has the required information). Age and race classification in not available and would require capturing a new database.

The virtual environment already provides some well-known state of the art methods that can be used for some of these modules. Among others, it provides

interfaces for the OpenCV Viola&Jones face detector, an LBP-based face recognition, as well as perfect face and eye detection using the ground truth. Face pose estimation is also provided using the ground truth.

Usage Example: evaluating a face recognition algorithm.

As an example, we evaluate a face recognition algorithm used by an agent corresponding to a robot moving in an environment with 10 persons. The face recognition method to be evaluated corresponds to a local-matching face recognition that is well suited for robot application because of its processing speed. It is based on histograms of LBP (Local Binary Patterns) features [18]. Following [7], histogram intersection (HI) is used as similarity measure. The face images are scaled to 81x150 pixels and divided into 40 regions to compute the LBP histograms.

The number of persons (*N*) in the first experiments was set to 10 in order to reduce the computational cost, and also considering that this number is enough to evaluate the modules; recall that the same subject may be observed in many frames. In the last two experiments, 20 persons were considered. The experiments are respited 10 times considering a different subjects of the 50 subjects of the DB.

The height of the agent is fixed and equal to the base height of the persons (160 cm). The height of subjects follows a uniform distribution in [136, 184] cm, i.e. a 15% variation around the agent's height.

With respect to how the gallery is constructed, we consider two testing modes:

- *Offline Gallery Mode*: The VE generates a face gallery before the recognition process starts. The gallery contains one image of each person to be recognized. The gallery's images are frontal pictures (no rotations in any plane), taken under indoor illumination conditions. This is the standard operation mode.

- *Online Gallery Mode*: There is no gallery. The agent needs to cross two times the virtual scenario. In the first round, it should create the gallery online. In the second round, the subjects change position in the scene, and the gallery is used for recognition. In both rounds, the agent sees the person's faces at variable distance and angles, in indoor or outdoor illumination conditions. The subjects pose and the illumination conditions are randomly chosen in all cases.

We consider the case of predefined trajectories. With respect to how the agent moves, we consider two cases: with and without the use of active vision. In both cases we consider the use of online and offline gallery DB, which gives 8 cases in total.

In all experiments, the map settings are the same in order to compare the different combinations. Each experiment is repeated 10 times, considering different trajectories (see Fig 4(f) for an example) and different subjects selected randomly out of the 50 subjects in the database. Table I shows the obtained results. The first, second and third columns show which face detector was used, whether the active vision module was used and how the gallery database was build. The fourth column displays the number of subjects added to the gallery database (normally when the gallery is build online, there are some false detections); the fifth column shows the rate of number of subjects correctly detected while traversing the environment; the sixth column shows the face recognition rate out of all subjects in the scene; and the seventh column shows the face recognition rate considering only the detected subjects.

From the first two rows of the table we can that the Person Map module is used, the recognition rate improves from 78.41% to 86.77%. This is because the same subject is

not added more than one time to the map, and at the same time there are less false detections. In all the other results (following row in the table) the Person Map is used. From the table it can also be observed that using the active vision module allows, among other things, to build a better gallery database (when the gallery database is built online), and to improve the recognition rate. When the database is built offline, using active vision improves the recognition rate from 86.77% to 92.92% on average.

Table 1. Evaluation of a Face Recognition system based on LBP features [18]

(*) It does not use Person Map. (+) 20 subjects are present in the scene; otherwise there are 10 subjects.						
Face Detection	Active Vision	Gallery Database	Persons added to the gallery	Persons correctly detected [%]	Recognition [%] (out of all subjects in the scene)	Recognition [%] (out of the detected subjects)
Viola&Jones	No	Offline	10.0	84.0%	----	78.4% (*)
Viola&Jones	No	Offline	10.0	84.0%	73.0%	86.8%
Viola&Jones	No	Online	14.8	84.0%	59.0%	70.2%
Viola&Jones	Yes	Offline	10.0	84.0%	78.0%	92.9%
Viola&Jones	Yes	Online	12.2	84.0%	73.0%	86.9%
Ground Truth	Yes	Offline	10.0	100.0%	92.0%	92.0%
Ground Truth	Yes	Online	10.0	100.0%	90.0%	90.0%
Viola&Jones	No	Offline	20.0 (+)	91.8%	83.0%	90.5%
Viola&Jones	Yes	Offline	20.0 (+)	91.8%	87.5%	95.7%

4 Conclusions and Projections

A virtual environment for testing face detection and recognition systems under uncontrolled conditions is proposed. The testing tool combines the use of a simulator with real face and background images taken under real-world conditions. Inside the virtual environment, an agent navigates and observes real face images, at different distances and angles, and with indoor or outdoor illumination. During the face detection and recognition process, the agent can actively change its viewpoint and relative distance to the faces in order to improve the recognition results. This tool allows evaluating multiple face analysis methods.

The applicability of the proposed tool is validated with an example; a face recognition method, namely histograms of LBP features [18], is evaluated in an agent corresponding to a robot moving in an environment with 10 subjects. In this example the face recognition performance is evaluated in two cases: when the gallery is build online and offline. Also, the use of active vision mechanisms is favorably compared to the use of a static trajectory that the agent cannot modify. The results validate our hypothesis that the use of active vision mechanisms improves largely the whole face recognition process, especially in the case of large out-of-plane face rotations.

This tool is useful for testing face analysis systems, in particular for comparing different face detection and recognition systems under similar conditions (e.g. using a similar active vision approach). In addition other face analysis subsystems can be also evaluated. For example a gender classifier, a face pose estimator, or an eye detector could be evaluated (our DB has the required information). In the future, age and race classification could be also implemented, but it would require extending the database.

Acknowledgments. This work was financed by the FONDECYT Postdoctoral Program Grant N. 3120218 and by the FONDECYT program Grant N. 1090250.

References

[1] A. F. Abate, M. Nappi, D. Riccio, G. Sabatino, 2D and 3D face recognition: A survey, *Pattern Recognition Letters*, Vol. 28, pp. 1885-1906, 2007.

[2] *Face Recognition Home Page* (Available on June 4th, 2012): http://www.face-rec.org/databases/

[3] BeFIT - Benchmarking Facial Image Analysis Technologies home page (Avalilable on July 5th, 2012): http://fipa.cs.kit.edu/412.php

[4] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao, The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations, Trans. Sys. Man Cyber. Part A, Jan 2008, Vol 38, N. 1, pp 149-161.

[5] *Labeled Faces in the Wild Database* (Available on June 5th, 2012): http://vis-www.cs.umass.edu/lfw/index.html

[6] G.B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*. University of Massachusetts, Amherst, Technical Report 07-49, Oct. 2007.

[7] J. Ruiz-del-Solar, R. Verschae, and M. Correa (2009). Recognition of Faces in Unconstrained Environments: A Comparative Study. *EURASIP Journal on Advances in Signal Processing*, Vol. 2009, Article ID 184617, 19 pages.

[8] Face Recognition Grand Challenge, Official website public site (Available on June 30th, 2010): http://www.frvt.org/FRGC/

[9] P. J. Phillips, H. Wechsler, J. Huang and P. Rauss, The FERET database and evaluation procedure for face recognition algorithms, Image and Vision Computing J., Vol. 16, no. 5, pp. 295-306, 1998.

[10] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, Overview of the Face Recognition Grand Challenge, Proc. of the IEEE Conf. Computer Vision and Pattern Recognition – CVPR 2005, Vol. 1, p. 947-954.

[11] R. Gross, Face Databases, in Handbook of Face Recognition, S. Li and A.K Jain (Eds.), Springer-Verlag, pp. 301-327, 2005.

[12] Yale University Face Image Database public site (Available on June 5th, 2012): http://cvc.yale.edu/projects/yalefaces/yalefaces.html

[13] BioID Face Database public site (Available on June 5th, 2012): http://www.humanscan.de/support/downloads/facedb.php

[14] AR Face Database public site (Available on June 30th, 2010): http://cobweb.ecn.purdue.edu/~aleix/aleix_face_DB.html

[15] P.J. Flynn, K.W. Bowyer, and P.J. Phillips (2003). Assessment of time dependency in face recognition: An initial study, *Audio and Video-Based Biometric Person Authentication*, pp. 44-51.

[16] Yale Face Database B. public site (Available on June 30th, 2010): http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html

[17] PIE Database. Basic information in (Available on June 30th, 2010): http://www.ri.cmu.edu/projects/project_418.html

[18] T. Ahonen, A. Hadid, and M. Pietikainen, Face Description with Local Binary Patterns: Application to Face Recognition, *IEEE Trans. on Patt. Analysis and Machine Intell.*, Vol. 28, No. 12, pp. 2037-2041, Dec. 2006.

[19] P. Viola and M. Jones, Robust Real-Time Face Detection, Int. J. Comput. Vision, Vol 57, N. 2, May 2004, pp. 137-154.