Simultaneous Face Detection and Recognition Using Viola-Jones Algorithm and Artificial Neural Networks for Identity Verification

Conference Paper · April 2013

DOI: 10.1109/TENCONSpring.2014.6863118

6 authors, including:

Argel Bandala

De La Salle University

Elmer P. Dadios

De La Salle University

Simultaneous Face Detection and Recognition using Viola-Jones Algorithm and Artificial Neural Networks for Identity Verification

Ma. Christina D. Fernandez*, Kristina Joyce E. Gob*, Aubrey Rose M. Leonidas*, Ron Jason J. Ravara*, Argel A. Bandala**, Elmer P. Dadios**

*Department of Computer Science
Polytechnic University of the Philippines
Manila, Philippines

** Gokongwei College of Engineering
De La Salle University
Manila, Philippines

***Abstract* — The study presented aims to design and develop a face recognition system. The system utilized Viola Jones Algorithm in detecting faces from a given image. Also the system used Artificial Neural Networks in recognizing faces detected from the input. Upon experimentation the system generated can recognize human faces with accuracy of 87.05%. The system performs at its best if the person is around 150cm away from the camera with an accuracy rate of 87.59%. Also, the best amount of lighting for the recognition system is at 480 lumens with an accuracy rate of 88.64%. Lastly, the system also performs at its best if the person is directly facing the camera or at 0 degrees with respect to the camera.**

Index Terms – ***Face Detection, Face Recognition, Artificial Neural Networks, Image Processing***

I. INTRODUCTION

Nowadays, machine face recognition technologies still encounter a lot of challenges that it needs to overcome before it can be widely accepted and deployed. The aim of recognition technologies is to mimic or even exceed the ability of the human mind in recognizing faces and/or other objects [1].

In light of this, researchers have endeavored during the past years to produce and use various algorithms to address several problems that arose with the development and implementation of face recognition systems [2]. However, only a few of these systems really have the ability to recognize faces the way the human mind does.

Another problem that surfaced with the implementation of face recognition systems is the fact that the use of facial databases invades the privacy of people and, if not properly secured, may also leak certain information about these individuals to unauthorized people [3]. Aside from these, researchers have also been dealing with the problem of varying illumination between trained and actual images. Certain studies show that the difference in illumination affects the efficacy of face recognition systems [4]. Also, the angle of rotation of the human face is also one factor that still requires further study since most face recognition systems limit the angle of rotation of their subjects' faces to either just facing the camera directly or having only a slight difference from it.

This study aims to design a system that would be capable to address these problems through the use of an algorithm that may recognize human faces in real time.

II. BLOCK DIAGRAM OF THE SYSTEM

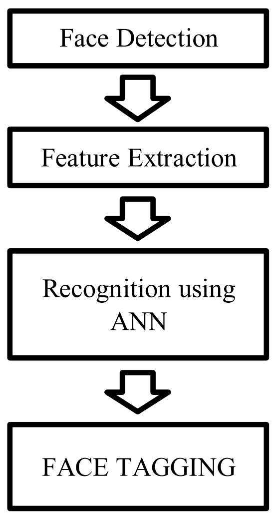


Figure 1 The General Block Diagram of the System

Figure 1 illustrates the block diagram used in the system. As can be seen there are four blocks utilized by the system. The first block, the face detection block, categorizes the variation of colors in the input image and derives the area occupied by the faces in this image. When the positions of the faces were detected, certain features will be extracted from them which will be done in the second block. The extracted features will be represented in numerical form that will serve as the input of the third block. The output of the third block is also a numeric value. Because of this a last block is used as an interpreter between the user and the system to determine the identity of the person detected by the system.

A. Face Detection Block

This block is the one that processes the input image. The input image is an image frame grabbed from the video stream transmitted by the camera. This block will serve as the starting point of the system. This block feeds the succeeding block with an image which is derived from the original image. The algorithm used for this part is the Viola-Jones Algorithm for Face Detection. This algorithm works by looking for Haar features. These features are basically black and white rectangles that the algorithm hunts for in an image. The algorithm adds the number of pixels of the rectangles to a box depending on whether it is contained in a threshold or not.

This step processes the image in such a way wherein it first tries to determine whether a human face is present in the image or not. If not, it then takes the succeeding image frame from the video and conducts the same process repeatedly until a face is detected. Once a face is detected, the system crops the image, removing the background and leaving only the face region.

B. Feature Extraction Block

This block begins after the face detection block. It accepts, as input, the processed image produced by the previous block. This block is a crucial part of any face recognition system. One of the important factors for this part is the measurement of certain biometric facial features which can be considered as unique for every person.

In this study, there are 7 facial features to be extracted and these are the skin color, color of the eye, the distance between the two eyes, the width of the nose, the height and width of the lips, and the distance between the nose and the lips.

These are then detected, extracted, and measured from the person's processed face image. These measurements are then passed through processes which will produce a representation of these characteristics in numerical vector form.

C. Recognition Using ANN Model Block

In this block, the data taken from the images are simulated through an ANN which has been previously trained using a set of 150 training images for every person to be recognized by the system. The input to this part will be a vector array with 14 values which was produced by the previous block.

D. Face Tagging Block

The result from the simulation in the previous block is then used by the recognition system to tag the appropriate name to the person whose image the camera has captured. This data, however, is in binary form so this block is also responsible in evaluating the expression into a certain value and matching it to a person's name in the name list. However, if the interpreted value is not one of the values listed in the roster, then the name returned will be automatically predefined as "Unknown".

After that, the name of the person is then attached to his/her image captured by the camera. Then, depending on whether the person is authorized or not, a warning message may be issued.

III. SYSTEM FLOWCHART

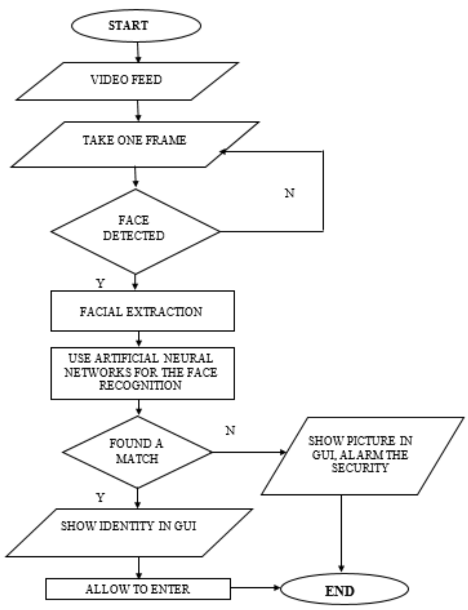
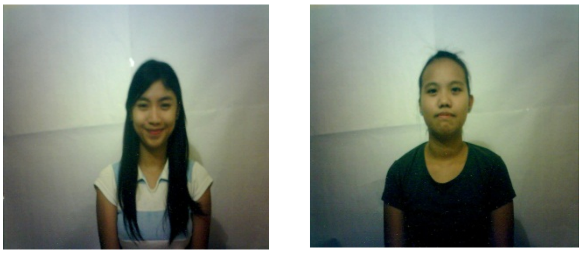


Figure 2 Flow Chart of the Recognition System

The steps followed by the system in implementing the process blocks discussed in Figure 1 are represented here in detail by the flowchart in Figure 2. As shown in Figure 2, the system begins the face recognition process by grabbing an image from the video stream provided by the camera.

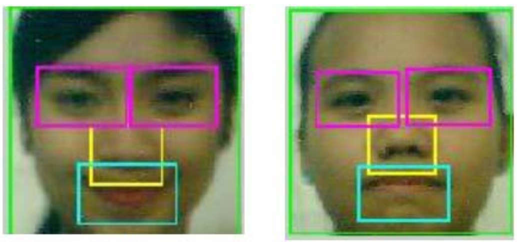


(A) (B)

Figure 3 Sample Input Image

Figure 3(a) and 3(b) shows sample image frames from the video feed. The input for this process, which comes directly from the camera, is a continuous stream of images transmitted to the system, and the system grabs a frame at a time for processing. The images in figure 2 are the type of images expected by the system.

As discussed above, the system determines through a set of processes if a person's face is present in the image. In the case that no faces are found from the image, the system will automatically repeat the previous sequence until a face is detected.



(A) (B)

Figure 4 Extracted Face Features

Figure 4 shows the cropped face image with the facial regions of interest to the researchers are enclosed in a box. The cropped image is just a small part of the original image that contains a person's face. As observed, there are only three regions of the human face that are considered in the study. These are the eyes, nose, and mouth regions.

From the isolated regions, the researchers drew various features such as the color of the skin which was calculated from the average skin color from the border of the yellow box.

Another feature of interest extracted from the image is the color of the lips, calculated from the average color of the lip portion isolated from the cyan box. Color of the eyes was drawn by taking the average color of the middle of the eye region. Eye color was taken as two separate values: one for the left and one for the right eye. These values are bounded by the violet box in the upper portion of the face. Another feature that the system considers is the width and height of some of the facial parts considered in the study.

For instance, in the cyan box that encloses the lips, the number of pixels is summed up across and from top to bottom of the lip portion in order to produce the height and width of the lips. Another feature drawn is the distance between the two purple boxes which was taken to measure the distance between the two eyes.

One more feature drawn from the image is the width of the nose which was taken the same way as the width of the lips. Lastly, the distance between the yellow and cyan box is also measured to represent the distance between the nose and the lips. These values are then compiled into a vector array to serve as input for the ANN used in the study.

(A) (B)

Figure 5 Annotated Video Output of the System

Figure 5 shows the output of the face tagging block of the system which is the cropped face image of the person annotated by his/her name. The result of the ANN is a binary vector representing a name from the roster. However, if the interpreted value of the ANN is not interpretable to a name from the roster, then "Unknown" is the string tagged as the person's name automatically.

IV. ARCHITECTURE OF THE NEURAL NETWORK

Figure 6 illustrates the neural network used in the study. Fourteen (14) values representing the features extracted from the image are the input used. The network is composed of neurons which are used to correctly identify a person's face.

In figure 7 the first layer or cluster of neurons is used to represent the input values to the network. The second layer, or hidden layer, is the one that deals with the interface between the input and output neurons. In this study, 20 hidden layers are used.

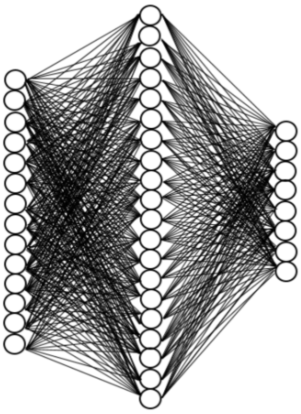


Figure 6 Architecture of the ANN

The third and final layer is the output layer. As can be observed in Figure 6, 4 output neurons were utilized. The output of the network is represented as a binary number. Only 4 output neurons were used since the experiment only accommodated 7 persons.

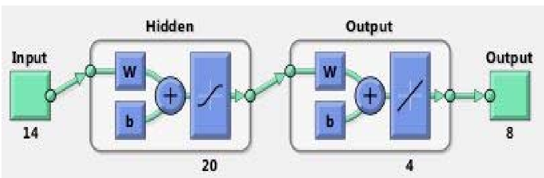


Figure 7 Architecture of the Neurons

A neuron is composed of a weight, a bias and an activation function, shown in figure 7. The values of these parts are calculated in the training stage of the study and are done prior to the implementation of the system [5] [6].

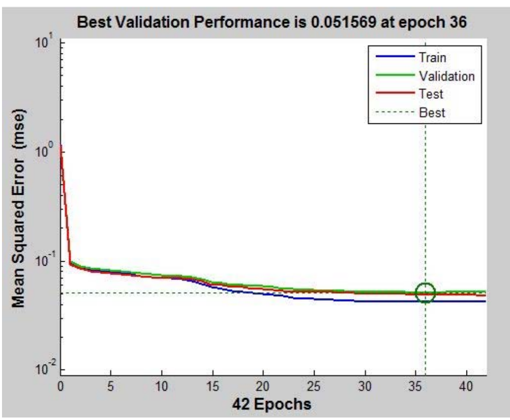


Figure 8 Performance of the Neural Network

Figure 8 shows the performance of the neural network. As can be observed the best performance of the system was seen in the 42nd iteration of the training. At this stage the system delivers a mean square error of 0.051569. This shows that the system is performing with an accuracy rate of 87.05%.

V. EXPERIMENTATION OF THE SYSTEM

After building the system, tests were conducted to determine the overall accuracy and precision of the system. These tests are also used to determine the effect of certain parameters to the system's accuracy.

The experiments were conducted in a closed room where the amount of lighting can be controlled. In this room, 10 images of each person used as subjects were taken per group of parameters (e.g. 100 cm. for distance, 230 lumens for luminance, and 0° for the orientation). Each image was then tested 30 times against the system in order to determine the precision of the system through the use of the standard deviation shown in equation 1.

$$\sigma = \sqrt{\frac{\Sigma(\bar{x} - x)^2}{n - 1}} \quad (1)$$

Where: x is the mean of the accuracy of the system,
x is the individual accuracy of the system per image,
and n is the number of tests.

For the accuracy of the system, a set of computations were used. Since the result of the ANN is a binary vector, the absolute error is determined by taking the sum of all the differences of the expected output and the measured output.

$$AE = \sum_{t0}^{tn} |Estimate - Actual| \quad (2)$$

The relative absolute error is computed per trial to determine the absolute error in terms of percentage (%) shown in equation 3.

$$RAE = \frac{AE}{n} \text{ x } 100 \quad (3)$$

And, finally, to compute for the accuracy of the system, the relative absolute error is subtracted from 100%.

$$\text{Accuracy} = 100 - \text{RAE} \quad (4)$$

Then, the mean of the accuracies were computed to produce the overall accuracy of the system.

V. REPRESENTATION OF DATA

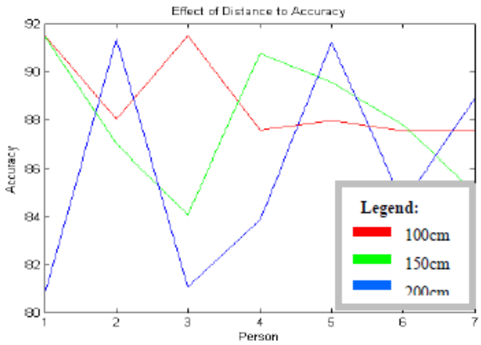


Figure 9 Effect of Distance to Accuracy

Figure 9 shows the graphical representation of the accuracy means of the system for the 100 cm, 150 cm, and the 200 cm tests. The graph shows that the system most accurately recognized a person when the person is 100 cm. away from the camera with an average accuracy rate of 88.8%. This is closely followed by the 150 cm. mark with an accuracy rate of 87.96%. Lastly, the overall average accuracy for the system when the person is 200 cm. away is 85.96%. This shows that the system can perform most accurately when the person is at a closer distance from the camera and least accurate when the person is too far from the camera.

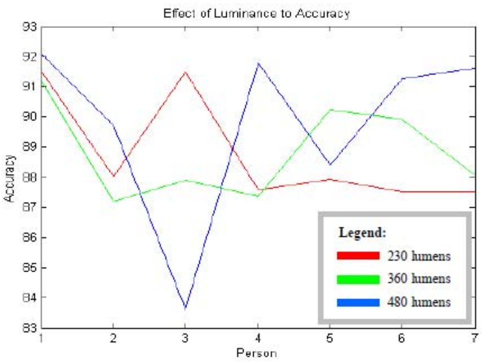


Figure 10 Effect of Luminance to Accuracy

Figure 10 illustrates the effect of the amount of lighting available to the accuracy of the system. The graph shows that the system performs at its best when the lighting is bright enough, in this case, at 480 lumens. The average accuracy rate calculated for this amount of lighting is 88.8%. The graph also illustrates the effect of dim lighting to the accuracy of the system through the 230 lumens mark which garnered an accuracy rate of 88.84% accuracy.

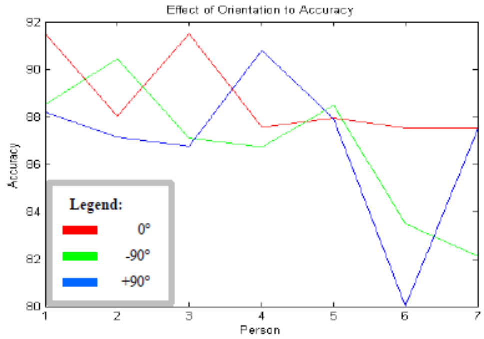


Figure 11 Effect of Orientation to Accuracy

The last parameter tested was the orientation. Figure 11 illustrates the accuracy rates for the different orientations used for the testing. The graph shows that persons facing the camera are the ones that the system can best recognize with an average accuracy rate of 88.8%. Also, it shows that the other two orientation values, -90 and +90 degrees, produced close accuracy rates of 86.69% and 86.91%, respectively.

VI. CONCLUSION

Through thorough analysis of data from the implementation/experimentation of the study entitled 'Multiple Face Detection and Recognition through Viola-Jones and Artificial Neural Networks,' the proponents have arrived with certain conclusions. The system can recognize human faces properly and accurately with an accuracy rate of 88.06%. The system performs at its best if the person is around 100cm away from the camera with an accuracy rate of 88.8%. Also, the best amount of lighting for the recognition system is at 480 lumens with an accuracy rate of 89.79%. Lastly, the system also performs at its best if the person is directly facing the camera or at 0 degrees with respect to the camera with an accuracy rate of 88.8%.

REFERENCES

[1] Benjamin Balas, Yuri Ostrovsky, and Richard Russell Pawan Sinha, "Face Recognition by Humans:Nineteen Results All Computer," March 2006.

[2] W. Zhao, R. Chellap, A. Rosenfeld, and P.J. Phillips, "Face Recognition: A Literature Survey," *ACM Computing Surveys (CSUR)*, vol. 35, no. 4, pp. 399-458, December 2000.

[3] "Facing Facts: Best Practices for Common Uses of Facial Recognition Technologies," Federal Trade Commission (FTC), Staff Report 2012.

[4] Jeffrey Ho and David Kriegman, "On the Effect of Illumination and Face Recognition," *Face Processing: Advanced Modeling and Methods*, 2005.

[5] Jerome Paul N Cruz et al., "Object Recognition and Detection by Shape and Color Pattern Recognition Utilizing Artificial Neural Networks," in *International Conference of Information and Communication Technology (ICoICT)*, Bandung, Indonesia, 2013, pp. 140 - 144.

[6] Leo S Bartolome, Argel A Bandala, Cesar Llorente, and Elmer P Dadios , "Vehicle Parking Inventory System Utilizing Image Recognition Through Artificial Neural Networks," in *IEEE Region 10 Conference*, Cebu, Philippines, 2012, pp. 1 – 5.