Eye Tracking to Enhance Facial Recognition Algorithms

Balu Ramamurthy
School of Computing
Clemson University
bramamu@clemson.edu

Brian Lewis
School of Computing
Clemson University
bblewis@clemson.edu

Andrew T Duchowski
School of Computing
Clemson University
andrewd@cs.clemson.edu

ABSTRACT
In this paper, we are using eye tracking to prove that the areas on a face that people use for facial recognition are similar to the areas of the face used in face recognition algorithms. These areas have a substantial effect on the face recognition algorithms.

Author Keywords
Eye Traking, Face Recognition,LBP,Gabor,HOG

INTRODUCTION
With the rise in security concerns over the past decade, a facial recognition system, a computer application for automatically identifying or verifying a person from a known set of images, has become a popular research area in computer vision. Inspite of the extensive research effort that has gone into these algorithms, we have yet to see a system that can be deployed effectively in an unconstrained setting. The only system that does seem to work well in all face recognition challanges is the human visual system. [8]. Therefore it makes eminent sense in understanding how humans visual system works and translating them into machine based algorithms.

In the numerous attempts going on to understand the human visual system,there are studies on humans and computer based facial recognition algorithms [4] [1] [8]. These studies basically compare how face recognition algorithms are inpar with humans. The results of these studies shows us that facial recognition algorithms are currently in a stage of surpassing humans in certain criterias like illumination, rembering large number of faces, etc....

Though these studies compared the humans with the computer based facial recognition algorithms, there were only a little effort in trying to combine or make use of the human visual system in computer based facial recognition algorithms, i.e there were no studies which tried to find the places where the humans look at in a face for recognition and improvising those results into facial recognition algorithms.

We assume that the places where humans look at on a face (called an Area of Interest or AOI) must have some importance in facial recoginition. Using eye tracking technology, we find these area of interests in our experiment, and study the effect of these area of interests in the facial recognition algorithms.

Modern face recognition methods can be generally divided into two categories: holistic matching methods and local matching methods. The Holistic method uses the entire face structure as a feature for recognition, while local matching method uses local features of the face for recognition.Recently,the localmatching approaches have shown promising results in face recognition. [9]

In this study, we collect the eye tracking data of 20 participants in a facial recognition experiment and use that data to find the common area of interest in the faces. Then we incoprate these AOIs in two local matching algorithms to study their effects on the algorithms.

BACKGROUND
In this section,we describe in brief about the two algorithms that are chosen for comparison,which are Local Binary Patterns (LBP) and Histograms of Gradients (HOG).

LBP and HOG were chosen because both of these algorithms split the face into different regions, then they extract the features from these regions and combine them all into a single large feature vector for face recognition. These regions resemble to the area of interest in our study.

LBP is a texture measure of pixels that quantifies intensity patterns in the neighborhood of the pixel [3]. Facial recognition has used patches such as spots, edges, corners, and other types of distinct texture patterns. Each LBP pixel score is calculated by counting the intensity pattern changes in a defined neighborhood of the pixel, p, which has a defined radius of r. The input image is in gray scale, and divided into M blocks, each with N pixels. Each LBP vector for a pixel is generated and then placed into a histogram with b(t) bins where b(t) = p(p-1)+3. The ordered set T(I) = T(1), , T(N) where T(1),,T(M) are the texture histograms that correspond to the different blocks. T(I) was then converted into a vectorized form of T(i) of M x b(t) dimensions. City-block distance was used to compare the texture features for I1 and I2.

Histograms of oriented gradients (HOG) are another common tool in classification through the method of object detection. This technique uses a localized approach to object detection [4]. The basic thought behind HoG is that an object can be described in shape and appearance through the distribution of the intensity gradients. HOGs has several benefits over other descriptor methods such as its work on localized parts of the image. The HOG descriptor is particularly suited for human

CHI 2012, May 5–10, 2012, Austin, TX, USA.

(a) Original image (b) Normalized image

Figure 1: Image before and after performing CSU Normaliztion

detection in images. Gradient values are computed with a one-dimensional centered, point derivative discrete mask in the horizontal and vertical directions. The filter used is [-1,0,1].

A histogram is then generated from these gradient values. A histogram is generated for each cell of pixels. Every pixel in the cell gives a weight to the orientation-based histogram channel. The histogram channels can go from 0 to 180 or 360 degrees depending on the gradient being signed or unsigned. It was found that unsigned gradients with 9 histogram channels work the best in human detection.

It is important to locally normalize the gradient strengths by grouping the cells into spatially-connected blocks. The HoG is the vector of the components of the normalized cell histograms from every block. These blocks often overlap which means some cells contribute more than once to the final descriptor.

METHOD

Participants

A sample population of 20 college students and faculty from Clemson University participated in this study. Participants were screened for vision defects so that any noise due to these problems was eliminated. Only Indian male and female participants were used in the study inorder to avoid the race effects on facial recognition.[4] [2]

Apparatus

The eye-tracker used in this study is a Tobii ET-17500 (Tobii Technology, Inc., Falls Church, VA) embedded in a TFT 17 monitor with a resolution of 1280x1024. This is a table-mounted (desktop based) binocular eye tracker. Near infrared light-emitting diodes are used to capture the reflection patterns on the corneas of the participants eyes. Subjects were asked to sit approximately 50 cm away from the display screen which provided the stimulus. Eye position data was sampled at 50 Hz, with a position accuracy of 0.5. The

Figure 2: Identification experiment stimuli : A pair of images of two different people, level easy

display, study parameters and data collection were managed by the ClearView 2.7.1 eye tracking software on a Windows XP computer. [Tobii, 2003]. The fixation filter was set within the Clearview software at 30 pixels radius and 100ms fixation duration.r was used to track the eye movement of participants. All experiments were created and run in the Tobii Studio program.

Stimulus

Face stimuli were chosen from the large grayscale Facial Recognition Technology (FERET) dataset [5] [6]. The FERET dataset contains 11338 images of 994 subjects (591 males and 403 females). The dataset contains faces with various expressions,poses and illumination.The FERET dataset was divided into various smaller subsets for different kind of experiments. Among these subsets we use the fafb subset which contains only images with slight change in facial expression (which are easy to identify),fafc subset which contains only faces which vary in illumination (which are difficult to identify).

To make the task as challenging as possible, we narrowed the available pairs to include only Caucasian males and females. Restriction to this relatively homogenous set of faces eliminates the posibility that algorithms or humans can base identify comparisons on surface facial characteristics associated with race or age. For the same reason, the face pairs presented to participants were also matched by same sex.

We then normalize the chosen images and also get rid of the extra texture such as background,clothes,etc ... as shown in figure 1. This helps in narrowing down the attention of the participants to the face instead of the backgroud, hair style, clothing, etc ... In this study we use the normalization algorithm provided by the CSU Face Identification Evaluation System [7] because of its popularity and use among the biometric research group.

Procedure

Figure 3: Verification experiment stimuli : A 2 * 3 grid of people, of difficulty level hard

Upon arrival, the participants were informed about the purpose of the study, i.e. to use eye movements to try and determine how humans try to identify and recognize faces using an eye tracker. Next each subject was asked to read the information consent. After which the participant was asked a few demographic questions (i.e. age, gender profession, vision) and answers were be recorded on a task record form. Participants were then asked to sit approximately 50 cm away from the display and their eyes were calibrated with the eye tracker using a five point grid included with Tobii Studio. All the subjects participated in two different experiments.

1.Verification, in this experiment, a pair of normalized faces next to each other are shown to the participant for 5 seconds in which time they were asked to verify whether the faces belonged to same person or of different person.The probe(test) image was displayed on the left and the target image on the right as shown in figure 2.The participants were asked to rate the image pairs using a 1-4 scale,

1. sure they are same
2. think they are same
3. think they are different
4. sure they are different

For the purpose of conducting ingerential statistical analyses on the behavioral data, participant responses were transformed into ”same” or ”different” judgments for individual pair of faces,so that it would be easy to elimnate the data in which the answer is wrong.Responses 1 and 2 were deemed ”same” and responses 3 and 4 were deemed ”different”.

2.Identification, in this experiment the participants were shown face of two persons right after each other and then were asked to identify them from a set of faces aligned in a 2*3 grid format as shown in figure 3 by identifying the face by their block numbers.

Figure 4: 7*7 blocks of region in the Verification Experiment.

Figure 5: 7*7 blocks of region in the Identification Experiment for the image identified.

Both of the experiment had 3 sets of stimulis. The first stimuli was for training the participant on the experiment and is ignored in the observation. Using a latin square,each participant viewed the next two set of stimuli in different orders of difficulty in which one of them is a easy set while the other was a hard one.

Result

Amongest the 19 images used as stimuli,the regular and training images, of which later were used to give the participant a feel of the experiment and the former one used with unlimited time contained equal number of fixations on all blocks were excluded,leaving with a total of 12 images

Face were divided into 7*7 blocks to imitate the experimental setup of the face recognition algorithms.The eye tracking experiment was used to analyse the percentage of fixations for each block.These blocks were created for all the match and non match images and were scaled accordingly to match the image’s scale as shown in the figures 4 and 5.

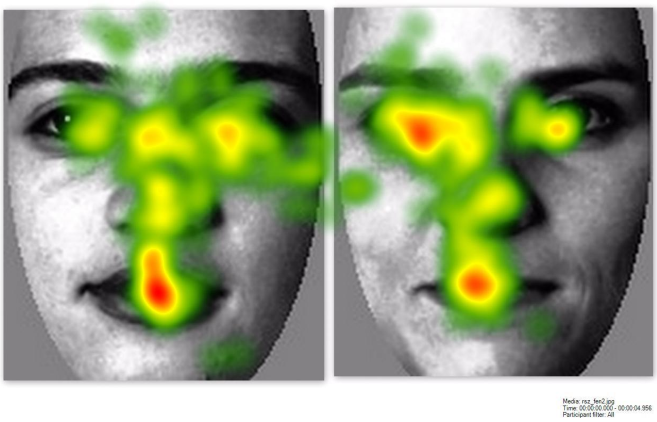

Figure 6: Heat map for a verification experiment image.

Experiment Group	LBP	HOG
All	75.35	65.27

Table 1: Experiment I results

Heat maps of the images show that the regions of the eyes, nose, and lips are the most important locations in the face which help in identifying the person. It can also be infered from the heat maps that the regions are not consistent across all the face i.e. some faces have more distinctive eyes than the others,for some its lips and nose and for some its eyebrows as well. For example in the figure 8,the correct female is the second person who has her eye closed in the image. Participant who used nose/lips for identifying people werenot able to identify her,whereas the people who looked at eyebrows were able to identify the person correctly which is visible from the heat map.

The participants were divided into three groups; male, female, and one group including everyone. Each group gave us a different percentage fixations for blocks which can be seen in the figure. The mean of the percentage fixations were calculated for all the group as shown in the table.Our eyetracking results show that female are more good at recognizing people than their counterparts.The weightage of the blocks were decided manually by looking at the distribution of the fixation across the number of blocks.

This process of assigining the weightage is a trial and error one. In this paper,we have conducted 4 experiments with different weights for each group and also for each of the algorithm i.e both LBP and HOG. For Experiment I, the blocks were given a equal weightage of 1 for all of them in order to get their normal recognition accuracy as shown in the table 1.

In experiment II, weights were given from range 0 to 3 with each weight containing equal number of blocks. This experiment was carried out in order to give the weights equal probability. The results of experiment 2 were below the normal recognition accuracy rate(2) ,this might be due to equal distribution of weights across the blocks.

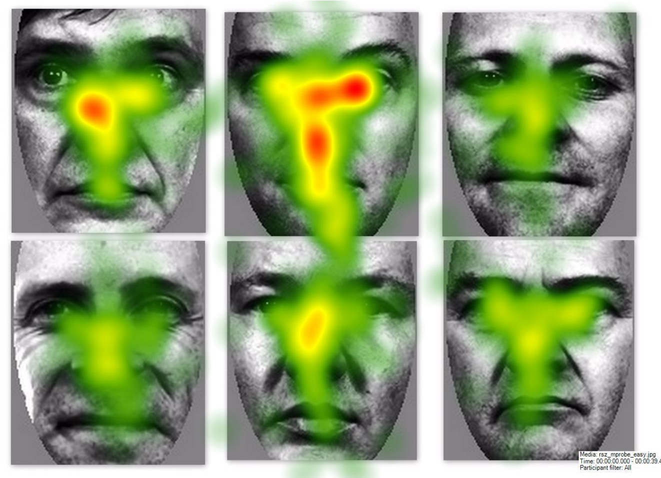

Figure 7: Heat map for a Identification experiment image.

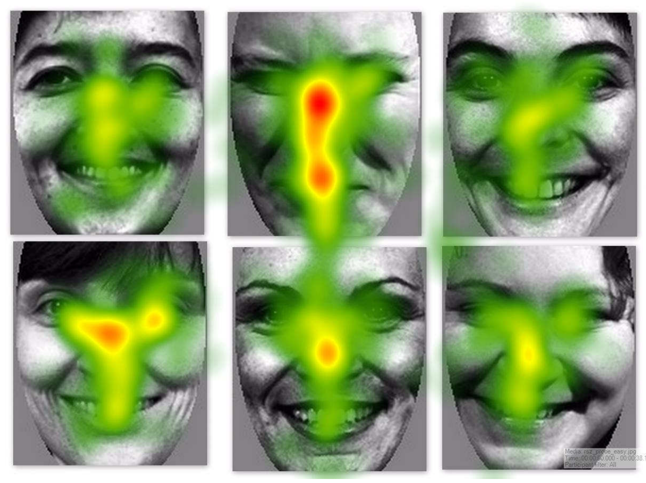

Figure 8: Heat map for a Identification experiment image.

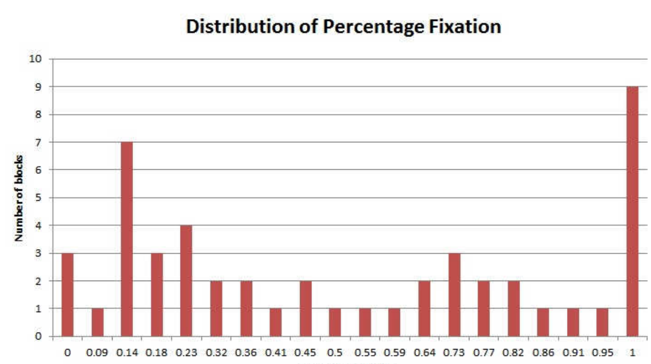

Figure 9: Percentage fixation distribution.

Experiment Group	LBP	HOG
Male	71.04	61.58
Female	72.88	63.34
All	73.12	63.17

Table 2: Experiment II results

Experiment Group	LBP	HOG
Male	73.44	61.86
Female	75.73	63.12
All	77.57	62.09

Table 3: Experiment III results

In experiment III, the weights were given from range 0 to 4. Blocks with percentage fixation less than 0.1 getting a 0, blocks with in fixation range of 0.1 to 0.5 getting a weight of 1 , 0.6 to 0.9 a weight of 2, and 0.9 to 1 a weight of 4. The weightage distribution for the blocks for experiment 2 and 3 can be found in the figure 10. The result of the experiment are tabulated in table 3.

In experiment IV, only the blocks with 100% percent fixation were used to understand their importance. All other blocks were given a weightage of 0 and block with 100% fixation was given a weightage of 1. The result of the experiment are shown in table 4.

Conclusion

From the results, one can notice that even though there is no significant increase in recognition accuracy of the algorithm performance, the block for which the percentage fixation is 100% constitutes to 50% of the recognition accuracy which substantiates our assumption that eyetracking can be used to improve the face recognition algorithm in this means. The Experiment IV and III which were a trial and error gives us hope that on further research in this area can lead us to a real breakthrough in future. We think that the scaling of the blocks for the images which were done manually would have been prone to errors. Avoiding this might help to boost up the results in furture. The number of participants also plays a vital part in all of this. The number of participants might also have a effect on the experiment and we also didn't have a solid paper to compare our findings or to follow a particular procedure in assigining the weights to the blocks. Trying to achieve the above said in near future, we thank the Clemson students who participated in our experiment.

Experiment Group	LBP	HOG
Male	63.23	55.23
Female	65.24	57.54
All	66.86	58.99

Table 4: Experiment IV results

REFERENCES

1. Hancock, P. J., Bruce, V., and Burton, M. A. A comparison of two computer-based face identification systems with human perceptions of faces. *Vision Research 38*, 15-16 (1998), 2277 – 2288.
2. Josephson, S., and Holmes, M. E. Cross-race recognition deficit and visual attention: do they all look (at faces) alike? In *Proceedings of the 2008 symposium on Eye tracking research & applications*, ETRA '08, ACM (New York, NY, USA, 2008), 157–164.
3. Ojala, T., Pietikinen, M., and Maenpaa, T. Multi-resolution grayscale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence 24*, 7 (2002), 971–987.
4. O'Toole, A., Phillips, P., Jiang, F., Ayyad, J., Penard, N., and Abdi, H. Face recognition algorithms surpass humans matching faces over changes in illumination. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 29*, 9 (sept. 2007), 1642 –1646.
5. Phillips, P., Moon, H., Rizvi, S., and Rauss, P. The feret evaluation methodology for face-recognition algorithms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 22*, 10 (oct 2000), 1090 – 1104.
6. Phillips, P., Wechsler, H., Huang, J., and Rauss, P. The feret database and evaluation procedure for face recognition algorithms. *Image and Vision Computing J 16*, 5 (1998), 295 – 306.
7. S., B. D., Beveridge, J. R., Teixeira, M., and Draper, B. A. The csu face identification evaluation system: its purpose, features, and structure. In *Proceedings of the 3rd international conference on Computer vision systems*, ICVS'03, Springer-Verlag (Berlin, Heidelberg, 2003), 304–313.
8. Sinha, P., Balas, B., Ostrovsky, Y., and Russell, R. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE 94*, 11 (nov. 2006), 1948 –1962.
9. Zou, J., Ji, Q., and Nagy, G. A comparative study of local matching approach for face recognition. *Image Processing, IEEE Transactions on 16*, 10 (oct. 2007), 2617 –2628.

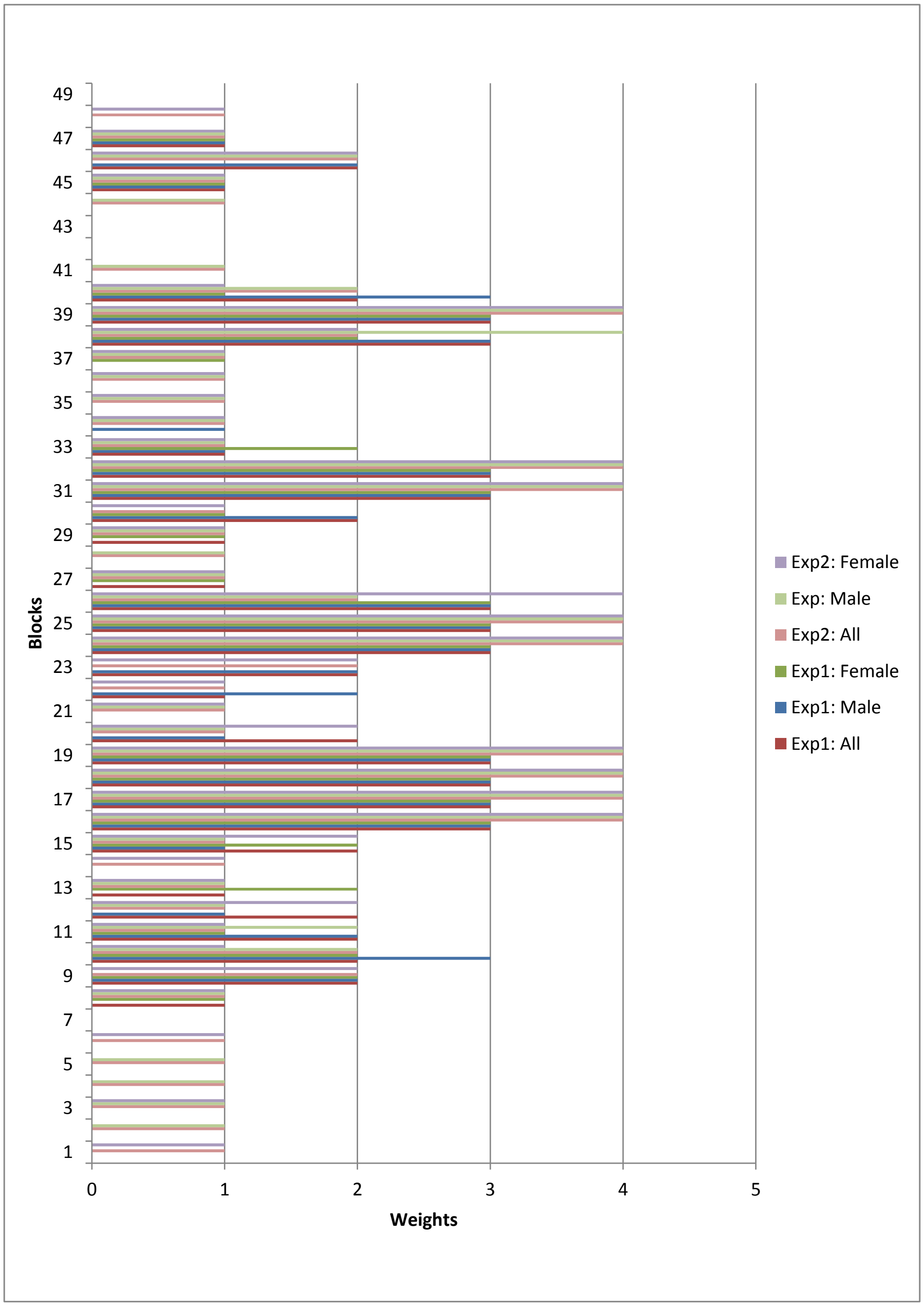

Figure 10: Block weightage for experiment II and III