Human Visual System Based Framework For Gender Recognition

Cherinet G. Zewdie[1,2] and Hubert Konik[1,2]
[1] *Université de Lyon, CNRS, 37 Rue du Repos, 69007 Lyon, France*
[2] *Université Jean Monnet, Laboratoire Hubert Curien, UMR5516, 42000 Saint-Etienne, France*

Keywords: Human Visual System, Gender Recognition, Salient Region, HVS Inspired Gender Recognition, Local Binary Pattern.

Abstract: A face reveals a great deal of information to a perceiver including gender. Humans use specific information (cue) from a face to recognize gender. The focus of this paper is to find out this cue when the Human Visual System (HVS) decodes gender of a face. The result can be used by a Computer Vision community to develop HVS inspired framework for gender recognition. We carried out a Pyscho-visual experiment to find which face region is most correlated with gender. Eye movements of 15 observers were recorded using an eye tracker when they performed gender recognition task under controlled and free viewing condition. Analysis of the eye movement shows that the eye region is the most correlated with gender recognition. We also proposed a HVS inspired automatic gender recognition framework based on the Psycho-visual experiment. The proposed framework is tested on FERET database and is shown to achieve a high recognition accuracy.

1 INTRODUCTION

A human face reveals a great deal of information to a perceiver. These information include Gender, Ethnicity, Facial expression etc. Gender is an important demographic attribute of people.

The Human Visual System (HVS) has an amazing ability to decode gender of a face across different cultures and ethnicities with few exceptions (Bruce et al., 1993) in a very short time. Only specific face region or cue is used by HVS to derive the gender information from a face (Bruce et al., 1993)(Ng et al., 2012). That is, only salient face region correlated with gender recognition is used, where salient means the most important.

The ability to recognize gender using computer vision has many applications in surveillance, human-computer interaction, content-based indexing and retrieval, biometrics, targeted advertizing etc. In human-computer interaction, it helps to design humanoid robots with knowledge to address humans appropriately (as Mr or Ms). In content-based indexing and retrieval, the ability to recognize gender helps in annotating the gender of the people in an image for the millions of images on the internet. In biometrics application such as face recognition, the ability to recognize gender can cut the time required for searching by half by training a separate classifier for male and female. In targeted advertizing, a computer vision powered billboards can display advertisements relevant to the person looking at the billboard based on gender (Ng et al., 2012).

Different methods for automatic gender recognition have been proposed (Alexandre, 2010)(Gutta et al., 1998)(Andreu and Mollineda, 2008)(Kawano et al., 2004)(Buchala et al., 2004)(Moghaddam and Yang, 2002)(Makinen and Raisamo, 2008). However, none of the proposed methods try to mimic the way HVS performs gender recognition. These methods rely on the power of image processing by spending computational time on whole face image to extract geometric based information using measurements of facial landmark or appearance based information using some operations or transformations applied on pixels of the face image with the idea that such extracted information is different for female and male.

Luis A. Alexandre (Alexandre, 2010) proposed a multi scale approach (considered as one of the state of the art results) to solve the problem of gender recognition. He proposed extracting features from different face image resolutions and then a classifier (SVM) is trained for each feature type. The decision from each classifier is fused using majority vote to obtain the final classification result. For each resolution, he extracted intensity, shape and texture features. Histograms of edge directions were used as shape fea-

G. Zewdie C. and Konik H..
Human Visual System Based Framework For Gender Recognition.
DOI: 10.5220/0005176102540260
In *Proceedings of the International Conference on Agents and Artificial Intelligence* (ICAART-2015), pages 254-260
ISBN: 978-989-758-074-1

tures, pixel values as intensity and uniform LBP as texture features. He tested this approach on a pre-processed FERET database (geometric normalization such as the same eye position, face rotation, and placing mouth in a fixed position) and reported only one miss classified image out of 107 test images (accuracy of 99.07 %). The high accuracy reported here is as a result of a very complex approach that may not be suitable for real time applications.

Other methods try to divide the face into regions with the hypothesis that processing only such information could discriminate between female and male. Y.Andreu et al (Andreu and Mollineda, 2008) proposed a method that exploits the role of facial parts in gender recognition. Given the image of a face, a number of subimages containing mouth, eye, nose, chin; and an internal face (made up of eye, nose, mouth and chin), external face (made up of hair, ears and contour) and the full face are extracted and converted to appearance based data vectors. Classification using SVM, KNN, and Quadratic Bayes Normal Classifier (QDC) tested on FERET (Phillips et al., 1998) (2147 images) and XM2VTS (Andreu and Mollineda, 2008) (1378 images) showed that individual parts include enough information to be able to discriminate between genders with accuracy of above 80% and the joint contribution of the parts (internal face) produced accuracy of over 95%. The motivation to evaluate the contribution face parts was to deal with situations where face images are partially occluded.

However, unlike these methods, we propose to solve the problem of gender recognition by simulating the HVS. We argue that the task of gender recognition can be done in a more conducive manner, if only face region correlated with gender recognition (salient region) is processed as it happens in HVS.

We conducted a Pyscho-visual experiment to find facial region (salient region) associated with gender recognition. An eye-tracking device which records fixations and saccades has been used for our experiment. Fixations are used to describe the visual attention when it is directed towards a salient region for the assigned task and the eye gathers most of the information during fixations. Saccades are eye movements between fixations. We also propose a human visual system based novel framework for gender recognition. The proposed framework creates a new feature space by extracting Uniform Local Binary Pattern (ULBP) (Ojala et al., 2002) features from only identified salient region of the face. The result from our experiment can also be used by the computer vision community to develop robust algorithms by identifying salient region containing discriminative information for gender recognition.

Our contribution in this study is three fold

1. Using a pyscho-visual experiment, we have statistically and using gaze maps identified the face region (salient region) correlated with gender (both male and female) according to human vision.
2. We used a novel approach of conducting an experiment using eye tracker to find gender cue from a face and the reported result is validated in the domain of human behavior study and cognitive science.
3. We proposed a novel HVS based framework for automatic gender recognition which achieves high recognition accuracy by only processing the salient region.

The rest of the paper is organized as follows: in section 2, we will present the psycho-visual experiment we carried out to record eye movements (localize salient region). Section 3 discusses the result of the analysis of the experiment. In section 4, we will present the proposed automatic gender recognition framework. Results and Discussion of the automatic recognition are presented in section 5. Conclusion and Future works are given in section 6.

2 PYSCHO-VISUAL EXPERIMENT

A gender recognition task is assigned to the observers. We then recorded their eye movements under controlled and free viewing condition. The recorded eye movements is analyzed to find which face region is salient for the displayed gender (stimulus).

2.1 Participants and Stimuli

Fifteen students and teachers from the University of Jean Monnet volunteered for the experiment. The subjects were between 20 and 34 years of age. They all had normal or corrected to normal vision. They were given a short briefing about the experiment and the apparatus before the start of the experiment.

The stimuli were created by choosing 20 gray scale face images (10 female and 10 male) from FERET dataset (Phillips et al., 1998) with a criterion of maximizing diversity. The criterion includes different ethnicities, faces with special characteristics (such as glass, beard), different ages (young and old).

2.2 Apparatus

A video based eye tracker called EyeLink II from SR research is used to record the eye movements of the human observers as they performed the gender categorization task. The EyeLink II is equipped with three miniature infrared cameras with one mounted on a lightweight headband for head motion compensation and the other two mounted on arms attached to headband for tracking both eyes.

The stimuli were presented on a 19 inch CRT monitor with a resolution of 1024 x 768 and a refresh rate of 85 Hz. The viewing distance was 70 cm resulting in a visual angle of 5.76^0 x 6.42^0.

2.3 Procedure

We preformed the experiment in a dark and quiet room to avoid any situation that would distract the observers. There was nothing in front of the observers field of view except the stimulus. The experiment was designed in such a way that it won't be tiresome for the observers or loose interest in the experiment.

2.4 Eye Movement Recording

The eye position of the observers was tracked at 500 Hz with an average noise of 0.010. Before the recording process started, we performed calibration and validation. Calibration is used to collect fixations on target points, in order to map raw eye data to gaze position. A nine point calibration is chosen for this experiment. For observers, the calibration step is fixating on a nine sequentially and randomly displayed points on different locations of the screen. The validation step is used to find the gaze accuracy of the calibration. A threshold error of 1^0 has been selected as the greatest divergence that could be accepted.

The head mounted eye tracker compensates for head movements so that observers have the flexibility to perform experiments in a free viewing condition. The data collected from the experiment is also not affected by the head movements as a result of this compensation.

3 RESULTS AND DISCUSSIONS

The primary data that we obtain using the eye tracker is the fixations and saccades of observers collected while performing the gender categorization task. Two types of analysis can be done on the collected fixations once eye blinks are separated and removed. They are gaze map construction and statistical analysis of the fixations. However, since our stimuli are static, it is very important that we know in advance the duration of the fixations that we analyze. Each of the 20 stimulus was presented to each observer for 3000 ms. But we analyzed the fixations collected during the first 800 ms. This is because, for a defined gender recognition task, when an observer is presented with a stimulus, the observer recognizes gender around $\approx 200 - 250ms$ after stimulus onset (Mouchetant-Rostaing et al., 2000). However, the categorization task sometimes happen around $\approx 250 - 700ms$. To include all possibilities, fixations during the first 800 ms were analyzed. We have also removed the first fixation as this is related to the cross sign that precedes every stimulus presentation so that all observers start the task from the same center face location.

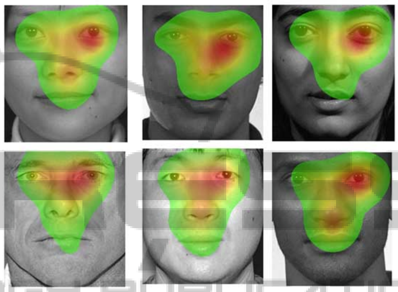

Figure 1: Gaze maps for gender recognition.

3.1 Gaze Map Construction

The simple, however, important output that we can obtain from the recorded fixation points is a gaze map. Figure 1 gives the first impression that gazes from all observers for both female and male stimuli are attracted to eyes, nose and mouth face regions. The second important fact we draw from the gaze map is that the eyes region is the most salient. This is because the superimposed color blobs is warmer for the eyes region compared to the mouth and the nose regions. That means, most of the fixations from the fifteen observers were attracted to the eyes region.

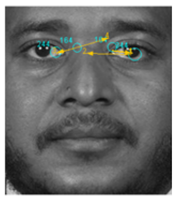

Figure 2: Saccades analysis for a face with special characteristic.

3.2 Statistical Analysis of Fixations

The conclusion that we made from the gaze map analysis is that the eyes region is the most salient for gender recognition task. We need to confirm this conclusion through statistical analysis of the collected fixations for different face regions. To confirm this, we computed and statistically analyzed the average percentage of trial time observers have fixated in the different face regions (eyes, nose and mouth) during the first 800*ms* for each stimulus presentation.

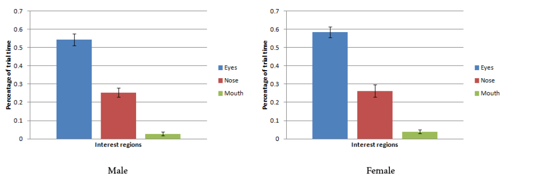

Figure 3: Statistical analysis of fixations for both female and male stimuli.

Figure 3 confirms the conclusion made from the gaze map shown in figure 1 about the saliency of eye region for gender recognition task. That is observers use most of the information from the eye region to categorize a stimulus as female or male. It can be easily observed from figure 3 that the average percentage of trial time spent is around 54.3%, 25.4% and 2.8% in the eyes, nose and mouth regions, respectively for male face; and 58.6%, 26.3% and 3.9%, respectively in the eyes, nose and mouth regions for female face. For both genders, the amount of time spent in the eye region is more than twice the amount spent in the nose region. The error bar represents the standard error (SE) of the mean. The saliency of the eye region is further confirmed by the analysis of saccades shown in figure 2. In the figure, we see a face stimulus with a beard. Beard is a characteristic associated with male face. Even in the presence of this characteristic, observers have still searched for gender information from the eye region. All the eye movements between fixation (saccades) were in the eye region. The discovery of the eye region as carrying information for gender recognition is consistent with the result by Brown et al (BrownU and Perrett, 1993) where sixteen male faces were averaged to create a male prototype and sixteen female faces where averaged to create a female prototype. Individual face regions such as eyes, brows, nose, mouth and chin were shown to subjects for gender recognition. The authors reported that all regions except the nose carried some information about gender.

4 PROPOSED AUTOMATIC GENDER RECOGNITION FRAMEWORK

Using the conclusion from section 3, we are going to test the idea that algorithmically gender can be recognized by processing only salient region. The proposed

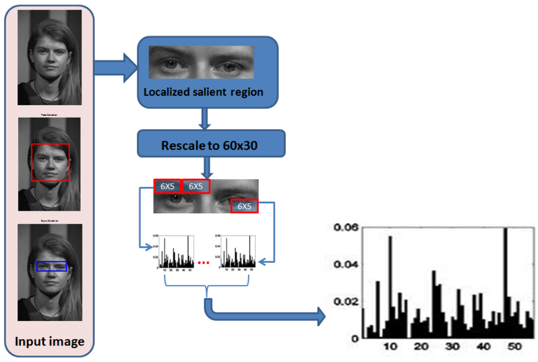

Figure 4: Human Visual System based feature extraction framework from a face for gender recognition

framework is shown in figure 4.

4.1 Localize Salient Region

As shown in figure 4, for a given face, we first detect the interest region. That means we try to detect the eye region since the eye region is the most salient for gender recognition. There are different detection algorithms. We chose to use the detection algorithm proposed by Viola and Jones (Viola and Jones, 2004) since the algorithm has a minimum false positives compared to other detection algorithms. Once we detect the eye region, we localize that region as the region from where we extract the features.

4.2 Feature Extraction

Once the eye region is localized, the next task would be to represent it using feature vectors(descriptors). There are different feature extraction methods. However, the type of features they extract (intensity, shape, texture etc) and their discriminative power is different depending on the problem at hand. The Uniform Local Binary Pattern (ULBP), which is the extension of the Local Binary Patterns (Ojala et al., 2002) is shown in different studies to have a high discriminative power for gender recognition (Ng et al., 2012). Based on their track record, we used ULBP to extract texture features from the salient eye region.

When describing a face with LBP patterns, it is important to consider local features instead of global features. For this reason, we computed the ULBP using a small local window and concatenated the features together. The other reason for doing this is that since the salient region we are considering is only the eyes region and considering the fact that there are similarities between female face eye region and male face eye region, the discriminative power of features could be improved if we capture the information locally. This approach is used by different researchers (Alexandre, 2010) even if the features were not extracted from a salient region but the whole face region.

4.3 Classifiers

Once the images are represented with features with a high degree of discriminative power, the next step will be to feed these feature vectors to a machine learning algorithm that learns a model for each gender. Gender recognition is a binary classification. A face image can only be a male or female. So one class is positive and the other is negative. If a given face image is not classified as a male (which in our study is a positive class) then it will automatically be classified as a female. There are a number of machine learning algorithms to solve the gender recognition problem. We tested SVM, KNN, C4.5 Decision tree and ensemble methods such as Bagging, Random Forest and Adaboost.

5 RESULTS FOR AUTOMATIC GENDER RECOGNITION

5.1 Database Selection

The challenging problem when solving gender recognition is the selection of database to test on since there is no specific database (containing multiple ethnicities) made for gender recognition. Limited by this problem, just like other researchers, we also selected images from a facial recognition database called FERET (Phillips et al., 1998) to test our algorithm. In total, we selected 207 images from FERET database to test our algorithm. 127 of the total is used for training and 80 is used for testing. The distribution for male and female is shown in table 1. No images of the same person is repeated in the database.

Table 1: Distribution of images as training and test set selected from FERET database.

	Original size	**Total number**	**Training**		**Test**	
			Female	*male*	*Female*	*Male*
FERET	512X768	207	60	67	40	40

5.2 Results

The classifiers listed above were used for learning and classifying the face images as either gender. The results from these classifiers are shown in table 2. The displayed accuracy value in the table is the accuracy of each classifier on the unseen data (on the 80 test images). A 5-fold cross validation is used to choose parameters for the classifiers and kernels used. In addition to measuring the accuracy in terms of correctly classified test images, we have also used ROC curve (value) to evaluate the performance of our algorithm.

Table 2: Classification results of the proposed framework for different classifiers.

	ACCURACY %	Alexander et al (Alexandre, 2010) State of the art
SVM + RBF + Cross validation (Weka)	100	99.01
KNN +Cross validation	100	-
C4.5 decision tree	100	-
Bagging	100	-
Random Forest	100	-
Adaboost (Tree)	100	-

The result of the analysis of ROC curve for SVM, KNN, C4.5 Decision tree, Bagging, Random Forest and Adaboost is shown in figure 5.

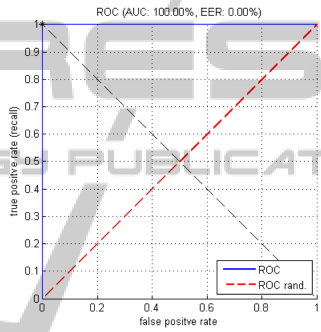

Figure 5: Receiver Operating Characteristics (ROC) for SVM, C4.5, KNN, Bagging, Random Forest, Adaboost.

All the accuracy results presented in table 2 were obtained on the 80 test images. Each of the classifiers shown in the table classified the given test images into their respective classes. We believe that different factors have contributed to the reported high accuracy of the result. The first reason is the processing of only salient region of the face correlated with gender recognition. This is because features with no discriminative power complicates the model that the classifier learns and thus reduce the predicting capability of the learned model. However, exploring the HVS to find the face information used by human when recognizing gender and using only those discriminative information with computer vision algorithms to recognize gender would produce an impressive result as shown in the table. The second reason is the type of features (texture) used and how the features are obtained. Texture features are shown in different studies to have produced a good result for gender recognition problem. Even if we claim that the salient region we used has a discriminative power, it may not produce the result shown in the table if the information from this region is not captured properly. Since the eye salient region for both male and female bears resemblance, capturing the features at a global level doesn't help in discriminating between the eye region of male face and female face. For this reason, we captured the information at a local level by using a small window. This helps us to capture the small spatial information of the eye region for both gender. The window covers the full image size by sliding horizontally and vertically. The result of the other window sizes we tried is shown in table 3. The third reason is the choice of classifiers. SVM, KNN and C4.5 are shown to produce a good result for a gender recognition problem in the survey made by (Ng et al., 2012). It is also no surprise, in addition to the quality features we had, that we obtained excellent result with Bagging, Random Forest and Adaboost since they combine results from a number of week classifiers. The number of weak classifiers (trees) used are 800, 500 and 90 for Bagging, Radom Forest and Adaboost, respectively.

Table 3: Accuracy(%) of different window size for different classifiers.

Window size	Bagging	Random Forest	Adaboost	KNN	SVM
6x5	100	100	100	100	100
10x10	73.75	71.25	75	63.75	73.75
10x12	73.75	73.75	78.75	60	75.25
15x20	63.75	67.5	72.5	56.25	68.5
30x60	65	66.25	62.5	65	59.25

We have also compared our result with the method considered as one of the state of the art results proposed at (Alexandre, 2010). As we repeatedly explained before, it is difficult to compare results from gender recognition algorithms since authors use different databases to test their algorithm. However, comparisons are usually made in the literature between algorithms tested on images from the same database (not necessarily the same images). For this, since we selected the images from a FERET (Phillips et al., 1998) dataset, we compared our result with (Alexandre, 2010) since this method is also tested on FERET as shown in table 2. One miss classified image is reported in (Alexandre, 2010) from 107 images. However, the high recognition accuracy is obtained owing to the very complex approach used. Considering the complexity of the proposed multiscale decision fusion approach, we believe that our method is simple and achieves a very good accuracy.

6 CONCLUSION AND FUTURE WORK

This study presented the face region correlated with gender recognition for both female and male faces by studying the HVS. This conclusion is made by recording eye movements of 15 observers as they performed a gender recognition task on 20 images chosen from FERET database. The constructed gaze map and statistical analysis of the collected fixations show that the eye region is the most salient for gender recognition. The localized salient region can be used by the computer vision community to have an insight from where to extract discriminative descriptors as this is the most important step when developing robust and efficient automatic gender recognition algorithms. In addition the extraction of descriptors from only the salient region would result in a fast system as it reduces the computational complexity of the algorithms.

We also proposed a novel framework for automatic gender recognition based on the localized salient region. We have achieved a high recognition accuracy by processing only salient region of a face.

In this paper, we only considered frontal faces when we created stimuli for the experiment. In the future, we have plans to include different face orientations into the stimuli and see if the same conclusion can be made. We also intend to occlude the eye region and find out if secondary information can be used for gender recognition, (i.e if gender recognition is hierarchical). Including different image resolution in the stimuli is also one of the many works we planned in the future.

ACKNOWLEDGEMENTS

This paper is supported by the Erasmus Mundus Scholarship 2012-2014 sponsored by the European Union, and CNRS.

REFERENCES

Alexandre, L. A. (2010). Gender recognition: A multiscale decision fusion approach. *Pattern Recognition Letters*, 31(11):1422–1427.

Andreu, Y. and Mollineda, R. A. (2008). The role of face parts in gender recognition. In *Image Analysis and Recognition*, pages 945–954. Springer.

BrownU, E. and Perrett, D. (1993). What gives a face its gender. *Perception*, 22:829–840.

Bruce, V., Burton, A. M., Hanna, E., Healey, P., Mason, O., Coombes, A., Fright, R., and Linney, A. (1993). Sex discrimination: how do we tell the difference between male and female faces? *Perception*.

Buchala, S., Davey, N., Frank, R. J., Gale, T. M., Loomes, M. J., and Kanargard, W. (2004). Gender classification of face images: The role of global and feature-based information. In *Neural Information Processing*, pages 763–768. Springer.

Gutta, S., Wechsler, H., and Phillips, P. J. (1998). Gender and ethnic classification of face images. In *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, pages 194–199. IEEE.

Kawano, T., Kato, K., and Yamamoto, K. (2004). A comparison of the gender differentiation capability between facial parts. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 1, pages 350–353. IEEE.

Makinen, E. and Raisamo, R. (2008). Evaluation of gender classification methods with automatically detected and aligned faces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(3):541–547.

Moghaddam, B. and Yang, M.-H. (2002). Learning gender with support faces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(5):707–711.

Mouchetant-Rostaing, Y., Giard, M.-H., Bentin, S., Aguera, P.-E., and Pernier, J. (2000). Neurophysiological correlates of face gender processing in humans. *European Journal of Neuroscience*, 12(1):303–310.

Ng, C. B., Tay, Y. H., and Goi, B. M. (2012). Vision-based human gender recognition: a survey. *arXiv preprint arXiv:1204.1611*.

Ojala, T., Pietikainen, M., and Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987.

Phillips, P. J., Wechsler, H., Huang, J., and Rauss, P. J. (1998). The feret database and evaluation procedure for face-recognition algorithms. *Image and vision computing*, 16(5):295–306.

Viola, P. and Jones, M. J. (2004). Robust real-time face detection. *International journal of computer vision*, 57(2):137–154.