

An Analysis of Automatic Gender Classification

Modesto Castrillón-Santana¹ and Quoc C. Vuong²

¹ IUSIANI

Edificio Central del Parque Científico-Tecnológico
Universidad de Las Palmas de Gran Canaria
35017 Las Palmas - Spain
`mcastrillon@mcastrillon.ulpgc.es`

² Division of Psychology

Henry Wellcome Building for Neuroecology
Newcastle University
Framlington Place
Newcastle upon Tyne
NE2 4HH - UK
`q.c.vuong@ncl.ac.uk`

Abstract. Different researches suggest that inner facial features are not the only discriminative features for tasks such as person identification or gender classification. Indeed, they have shown an influence of features which are part of the local face context, such as hair, on these tasks. However, object-centered approaches which ignore local context dominate the research in computational vision based facial analysis. In this paper, we performed an analysis to study which areas and which resolutions are diagnostic for the gender classification problem. We first demonstrate the importance of contextual features in human observers for gender classification using a psychophysical "bubbles" technique. The success rate achieved without internal facial information convinced us to analyze the performance of an appearance-based representation which takes into account facial areas and resolutions that integrate inner features and local context.

Keywords: Gender classification, local context, PCA, SVM

1 Introduction

Humans make use of different static and dynamic features to successfully identify and interact with other people. Gender is one of them.

Gender recognition is an often-studied topic in both the psychology and, since the first work in the early 90s [1, 2], the computer vision community. Recently, very high performance was obtained using only facial information [3, 4]. To the best of our knowledge, in computer vision this task has been based almost exclusively on internal facial features, except in some recent work which analyzed the discriminative capability of external facial features [5] such as hair [6].

This latter finding is consistent with the human recognition system, which makes use of external and other features for gender recognition, such as gait,

hair, clothes, voice, etc. [7, 8]. To illustrate this, if we look at different paintings, e.g., August Macke’s *Hat Shop*, or Rene Magritte’s *Son of Man*, it is observed that even when internal facial features are not distinguishable at all, different contextual details depicted by the painter allow us to confidently infer the gender of the person depicted. This observation is also supported by psychophysical results, which suggests that local context is a robust cue to gender for low resolution or noisy images [8, 9]. This paper investigates the extent to which the facial area and resolution chosen for gender classification influence in the performance of an automatic gender classifier.



Fig. 1. Image samples (62×54) pixels.

2 Evidence of Local Context Influence in Human Perception

In the psychology literature, the local context around the face has been established as a source of information to characterize an individual. The local context is defined as the information surrounding the internal features of the face, including hair and possibly the environment surrounding the head (i.e., local background) [10].

According to Torralba [11], object-centered approaches dominate the research in computational vision based facial analysis. Most systems have been designed for a resolution range in which the inner face elements are clearly distinguishable. This focus restricts their reliability due to the fact that these systems are easily fooled in situations with poor viewing conditions in terms of pose, resolution, illumination, etc. Such conditions have been systematically studied in psychophysical experiments [10], and the results suggest that humans make use of the local context as the level of detail decreases to achieve greater robustness. However, Sinha’s previous work also suggests that the role of local context is not exclusively utilized at low resolution. Indeed, under some circumstances, the local context provides enough discriminative features which indicates that internal facial details are not the only sources of gender information that are taken into consideration [8, 9].

To investigate the influence of local context for gender recognition, we first analyzed which contextual features are used by human observers for gender classification. For that purpose, we used a psychophysical ”bubbles” technique [12]



Fig. 2. Two different images samples after applying randomly distributed bubble. The inner facial features are always masked but the eye locations are indicated to observers by two red dots. Figure 1 shows some samples without any mask applied but down-sampled to 62×54 pixels.

to isolate those regions used by humans when only local contextual information around the head and shoulders is available in the image. The technique was used in [12] to identify internal facial features that provided diagnostic information for gender classification. With high-resolution face images, the gender was correctly determined using just the eyes and mouth.

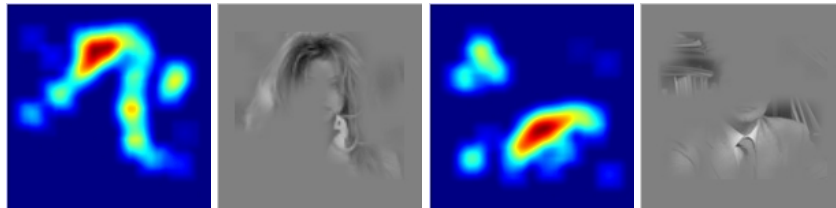


Fig. 3. The left pair shows the diagnostic image and the superimposition of the image with a female sample. The diagnostic image shows diagnostic regions (and their strength) that lead to accurate classification by humans. The right pair shows the results for males.

In the current study, images containing aligned male and female face images were revealed through a mask of small randomly distributed Gaussian windows (“bubbles”). That is, the presence of a bubble over a region showed that region. To prevent human observers from making the classification on the basis of internal facial features, the bubbles were never placed on any internal facial region, as shown in Figure 2. Therefore, observers classified the gender based on information contained in the local context surrounding the face. Across observers, masks leading to correct responses are summed and normalized to reveal image regions that were diagnostic for gender classification. The resulting normalized images are referred to as diagnostic images.

The results (i.e. diagnostic images) for a dataset of 962 images (124×108 pixels) are shown in Figure 3. They indicated that the regions used by humans

to discriminate gender when internal facial features were not visible depended on the gender of the test face. For male faces, the neck provided the most diagnostic information (i.e., the tie or the shirt neck provided discriminative information). By comparison, for females faces, long hair was frequently used as a diagnostic cue. With these regions observers correctly classified the stimuli 70 – 75% on average, without seeing too much internal facial details.

3 Automatic Gender classification

The psychophysical results above suggested that the local contextual information around the head and shoulders (e.g., hair, neck tie) provided diagnostic gender information. As we mentioned above, however, automatic gender classification have been predominantly based on internal facial features. For example, the recent work described in [3] used perfectly aligned and cropped frontal faces represented using Principal Components Analysis (PCA). A Support Vector Machine (SVM) classifier [13] was subsequently used to classify the PCA-based representations. This classifier achieved high gender classification performance even at low resolution.

We adopted the same appearance-based representation space, i.e. PCA, and classification approach used by Moghaddam et al. [3]. However, instead of manually collecting a training set for our experiments, we moved to a less restricted scenario in which an automatic face detector [14] provided the input data source extracted from video streams to incrementally build up the training set. In our experiments, gender classification was performed each time a new individual was encountered, but the system used the information extracted to iteratively upgrade the gender classifier.

Our assumption is that the system can be improved by a supervisor which confirms or corrects its classifications. Labelling interactively for learning is not new [15]. The online training is therefore unconstrained in the sense that an individual can be presented in an arbitrary order in front of a camera and labelled by a supervisor.

4 Experiments

4.1 Representation

The initial representation space, based on PCA, has been set up using a dataset of 7000 images taken randomly from the Internet. These images were annotated by hand, and normalized according to eye position to obtain samples that were 59x65 pixels each.

We randomly selected half of this set for training, and the other half for testing. We then computed a collection of SVM-based classifiers [13] which takes into considering account different number of eigenvalues for the face representation. Based on the results shown in Figure 5, we employed 100 eigenvalues to represent faces in our online experiments.



Fig. 4. Samples of different facial areas used for automatic recognition. From left to right: *Cropped A* (without hair line), *Cropped A IN* (same but with Illumination Normalization [16]), *Uncropped* (including local context), *Cropped B* (including the hair line) and *Eyes*. The first two areas have a dimension of 59×65 pixels, the second two areas 75×90 pixels, and the last area has a dimension of 15×11 pixels.

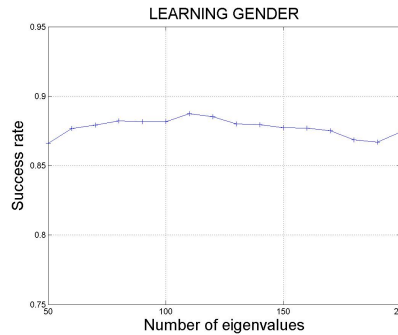


Fig. 5. Correct recognition rate achieved with different classifiers, according to the number of eigenvalues employed, computed offline with around 3400 training images.

To test successive meetings, we built up a database which made use of broadcast television. The database contains around 1130 different video streams with resolution equal to or greater than 320×240 pixels. The sequences correspond to approximately 850 different identities. Unfortunately we do not have permission from most of the sources to share the database.

For our face representation, we used incremental PCA (IPCA) approach described in [17, 18] instead of using a fixed PCA space. Thus, the learning procedure starts with an initial PCA representation space, and an initially empty training set. The system meets individuals in a random order, after each meeting, it determines a classification (once a sample is available for each class) for that meeting based on some significant exemplars extracted. Incorrectly classified exemplars are then employed to 1) increase the training set to update the gender classifier, and 2) to incrementally update the PCA representation space [17, 18].

4.2 Results

As mentioned above, for our experiments, we used an automatic face detector [14] that is freely available for research purposes. To avoid artifacts produced by the order in which videos are presented or meetings held, the experimental procedure was repeated randomly several times for five different facial regions, see Figure 4. We present the results achieved averaged across these repetitions..

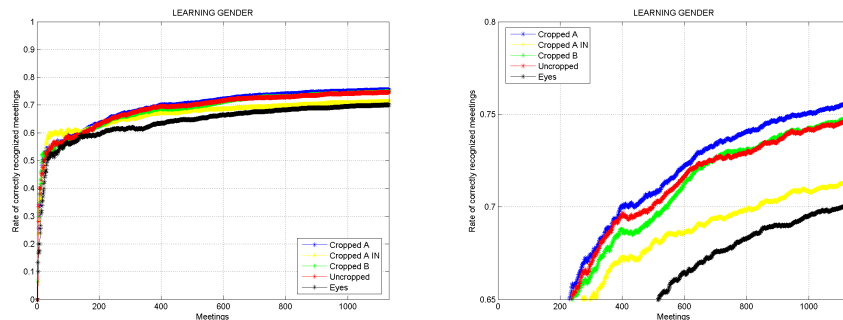


Fig. 6. Gender classification performance evolution (zoomed on the right). For each meeting, the cumulative rate of correct classification is indicated.

Figure 6 presents the results achieved using the different facial areas shown in Figure 4. The plot represents the cumulative success rate of classified meetings. Initially, the system classified gender poorly. However in the final stages, three curves reported a success rate greater than 75%, which indicates that the system was able to recognize correctly more than 847 sequences of held meetings at that point, and this tendency continued to grow. Two curves in Figure 6 clearly showed relative poor performance: *Eyes*, and *Cropped A IN*.

We were also interested in testing with an independent and larger test set of images of images not contained in the video streams. Figure 7 presents the results achieved using again the facial areas shown in Figure 4 for the test set used to compute the initial PCA (containing 7000 faces annotated by hand). The overall classification performance is slightly better, and again the curves for the *Eyes*, and *Cropped A IN* areas show worst results than the other areas.

Figure 8.left shows the number of samples automatically included in the training set by the different areas. The *Cropped B* appeared to build a training set that was slightly smaller than the *Cropped B* and *Uncropped* areas. The *Cropped B* does not include information related to hair, but does include the forehead which could, in some situations, be affected by hair.

Another question which arose during these experiments was the influence of spatial resolution. In [3], it was stated that the automatic system can easily

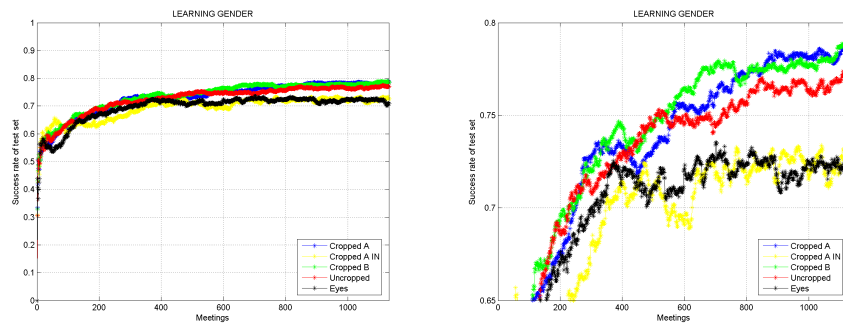


Fig. 7. Gender classification performance evolution (zoomed on the right). For each meeting, the cumulative rate of correct classification is indicated for the independent test set.

outperform a human at low resolution. Therefore, we tested four different resolutions, as shown in Figure 9. For resolutions lower than 18×22 , we observed a significant reduction in classification performance and an increase in the number of samples, as shown in Figure 8.right. Figures 10 and 11 depict the results for the three best curves (*Cropped A*, *Cropped B* and *Uncropped*) using different resolutions. Based on these results, we do not see a clear advantage in performance for any one of those areas. Overall, however, the lowest resolution revealed a decrease in performance and a corresponding increase in the size of the training set.

5 Conclusions and Future Work

In the present work, we first analyzed the influence of outer facial regions on human gender classification performance. The results indicated that humans do not only use inner facial regions to classify gender.

Motivated by the human data, we subsequently employed a state-of-the-art approach to investigate which inner and outer facial areas provided better information for an automatic gender classifier. The entire process is performed automatically from face detection to exemplars selection. Only a supervisor is required to confirm or correct a classification performed by the system. This is the key point for the learning system due to the fact that incorrectly classified patterns are added to the training set.

The resulting classifiers required a much smaller training set: less than 10% that of the classifiers used for offline learning. However, the performance of the offline classifier was still greater.

Our results indicated that the minimal resolution required for gender recognition should not be lower than 18×22 . However, we did not find a clear winner in

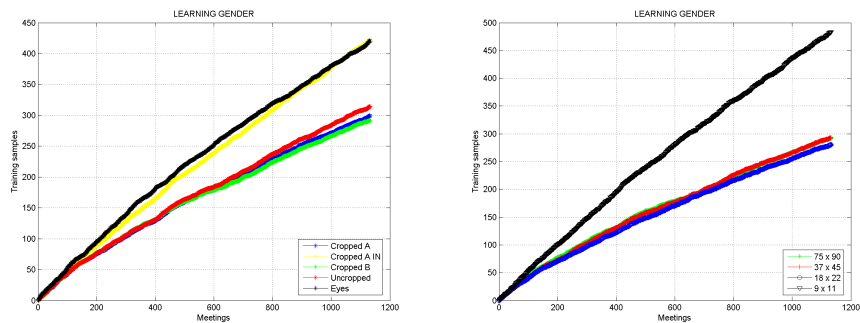


Fig. 8. Gender classification performance evolution. For each meeting, the cumulative size of training set is indicated. Left) comparison of facial areas used, Right) Comparison of resolutions used for *Cropped B*.



Fig. 9. *Cropped B* samples at three different resolution: a) 75×90 , b) 37×45 , c) 18×22 , and d) 9×11 pixels.

terms of the optimal area for gender classification (among *Cropped A*, *Cropped B* and *Uncropped*). According to this finding, for our configuration the local context, although useful for humans, does not seem to provide additional information for gender classification if internal features are present.

In the near future we plan to test the possibility of classifying without the availability of inner facial features. Additionally, we would like to test the possibility of combining the different classifiers to get a better metaclassifier for the gender classification problem.

Acknowledgments

Work partially funded by research projects Univ. of Las Palmas de Gran Canaria UNI2005/18 and the Spanish Ministry of Education and Science and FEDER funds (TIN2004-07087).

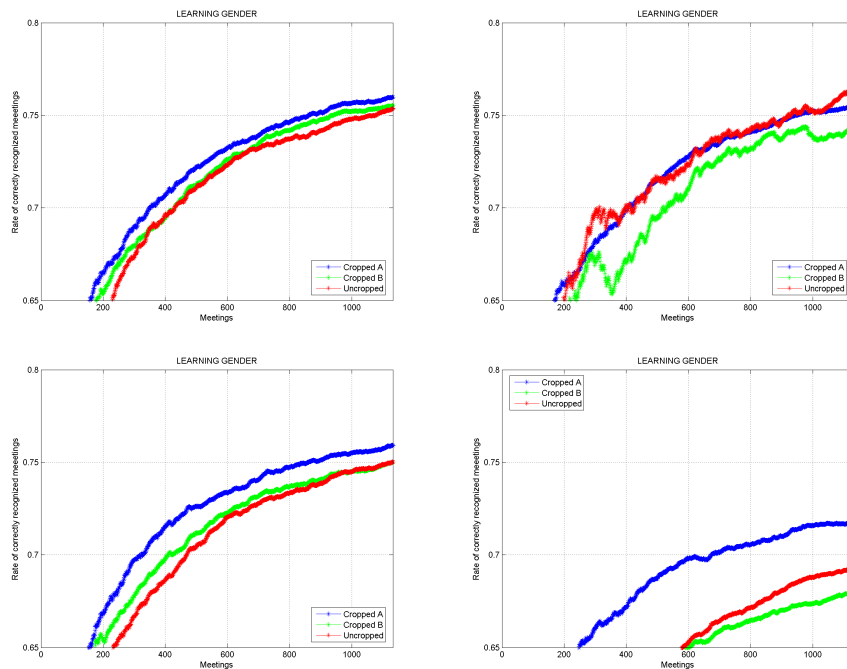


Fig. 10. Cropped A, B and Uncropped at different resolutions (full, half, 1/4 and 1/8) using the meetings results

References

1. Abdi, H., Valentin, D., Edelman, B.G., O'Toole, A.J.: More about the difference between men and women: evidence from linear neural network and the principal component approach. *Perception* **24** (1995)
2. Brunelli, R., Poggio, T.: Hyperbf networks for gender classification. In: Proceedings of the DARPA Image Understanding Workshop. (1992) 311–314
3. Moghaddam, B., Yang, M.H.: Learning gender with support faces. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **24**(5) (2002) 707–711
4. Wu, J., Smith, W.A.P., Hancock, E.R.: Learning mixture models for gender classification based on facial surface normals. In: 3rd Iberian Conference on Pattern Recognition and Image Analysis, Girona, Spain (2007) 39–46
5. Lapedriza, A., Masip, D., Vitria, J.: Are external face features useful for automatic face classification? In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops. Volume 3. (2005) 151–157
6. Yacoob, Y., Davis, L.S.: Detection and analysis of hair. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **28**(7) (2006) 1164–1169
7. Bruce, V., Young, A.: *The eye of the beholder*. Oxford University Press (1998)
8. Jarudi, I., Sinha, P.: Relative roles of internal and external features in face recognition. Technical Report memo 225, CBCL (2005)

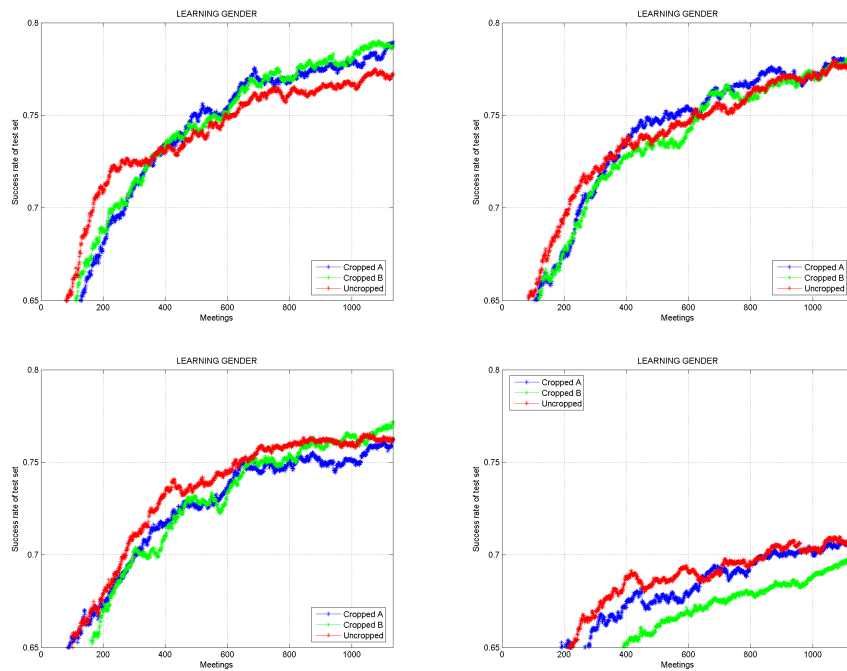


Fig. 11. Cropped A, B and Uncropped at different resolutions (full, half, 1/4 and 1/8) using the independent test set results

9. Sinha, P., Poggio, T.: I think I know that face ... *Nature* **384**(6608) (1996) 384–404
10. Sinha, P., Torralba, A.: Detecting faces in impoverished images. AI memo 2001-028, CBCL memo 208, Massachusetts Institute of Technology (2001)
11. Torralba, A.: Contextual priming for object detection. *International Journal of Computer Vision* **53**(2) (2003) 169–191
12. Gosselin, F., Schyns, P.G.: Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Research* (2001) 2261–2271
13. Vapnik, V.: *The nature of statistical learning theory*. Springer, New York (1995)
14. Castrillón, M., Déniz, O., Hernández, M., Guerra, C.: ENCARA2: Real-time detection of multiple faces at different resolutions in video streams. *Journal of Visual Communication and Image Representation* (2007) 130–140
15. Wersing, H., Kirstein, S., Goetting, M., Brandl, H., Dunn1, M., Mikhailova, I., Goerick, C., Steil, J., Ritter, H., Kierner, E.: Online learning of objects and faces in an integrated biologically motivated architecture. In: *ICVS*. (2007)
16. Xie, X., Lam, K.M.: An efficient illumination normalization method for face recognition. *Pattern Recognition Letters* **27**(6) (2006) 609–617
17. Artac, M., Jogan, M., Leonardis, A.: Incremental PCA for on-line visual learning and recognition. In: *Proceedings 16th International Conference on Pattern Recognition*. (2002) 781–784
18. Hall, P., Marshall, D., Martin, R.: Incremental eigenanalysis for classification. In: *British Machine Vision Conference*. Volume 1. (1998) 286–295