Face Description for Perceptual User Interfaces

M. Castrillón-Santana, J. Lorenzo-Navarro, D. Hernández-Sosa, and J. Isern-González

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

Abstract We investigate mechanisms which can endow the computer with the ability of describing a human face by means of computer vision techniques. This is a necessary requirement in order to develop HCI approaches which make the user feel himself/herself perceived. This paper describes our experiences considering gender, race and the presence of moustache and glasses. This is accomplished comparing, on a set of 6000 facial images, two different face representation approaches: Principal Components Analysis (PCA) and Gabor filters. The results achieved using a Support Vector Machine (SVM) based classifier are promising and particularly better for the second representation approach.

1 Introduction

Nowadays human faces are everywhere. Not only we are exposed to facial patterns due to the fact of living in community, but also faces are present on magazine covers, commercials, news, ads, etc. Faces provide a channel which fills a main part of the non verbal communication held during human encounters [1]. They convey both dynamic information and signals of great interest for social interaction such as gender, age and more. The interpretation of those signals is a basic ability to be included in any Vision Based Interface [27] which makes use of Computer Vision technology to perceive the user in a Human Computer Interaction (HCI) context.

Many facial analysis papers have particularly focused on the face recognition and verification problems. A well known corpus used to evaluate recognition techniques is the FERET database [22] and more recently the Face Recognition Vendor Test. Verification approaches have their own framework, the BANCA protocol [2]. However, other facial descriptors which are particularly useful to describe unknown individuals, or to realize changes in human appearance during social interaction, have not excited the interest of the researchers similarly. Certainly, gender classification and facial expression recognition are exceptions [19,21]; but other descriptors such as race, glasses, moustaches, beards, hair color, hair style, eyes color, etc., have not been widely considered.

Recent developments suggest that these descriptors can be of interest for automatic face processing. Indeed, local context is taking more importance in the literature for detection and recognition [26]. Some authors have evidenced that the local context is used differently by individuals, e.g. people born and living in Europe would pay more attention to hair and its color while people born and living in Japan would not consider the hair as an identification cue [6,24].

In this paper our main effort is to provide results in the direction of face description by means of additional semantic labels. These labels or descriptors can be used during HCI in order to endow the computer with abilities that can make the user feel himself/herself perceived. For this aim, we compare two well known state of the art approaches for face representation: 1) Principal Components Analysis (PCA), and 2) Gabor filters. Both representations are used by a Support Vector Machine (SVM) based classifier to provide an automatic suggestion.

An introduction to the techniques employed for face representation is given in Section 2. Section 3 presents and discusses the experimental results achieved. Some conclusions are summarized in Section 4.

2 Facial Description

In this work, we have paid attention only to inner facial features, trying to cover a small subset of the semantic facial descriptors which can be extracted from a single face image: gender, race, and the presence of moustache and glasses.

Different recent works have tackled the problem of gender recognition. A recent approach based on perfectly aligned images outperforms humans in low resolution images [19]. In [18] a Gabor wavelet representation on selected points is used with good results in gender and race classification. In relation with the others descriptors, there are different references [15,29] which try to detect the presence of glasses in a face, but we have none tackling the presence of moustache.

2.1 Face Representation

Principal Components Analysis (PCA). PCA decomposition is a well known technique used to reduce data redundancy. This representation schema chooses the dimension reduction that maximizes the scatter of the projected samples. PCA has been used extensively for face representation since the work described in [17], due to the fact that it provides a reduced representation without a significant lost of information. As seen in Figure 1, once an image, I, is projected, different coefficients represent the image in this space of reduced dimensionality, $I_{PCA} = \{v_1, v_2, v_3, ..., v_n\}$. The original image can be recovered by means of a linear combination, defined by these coefficients, of the different eigenvectors plus the average image.

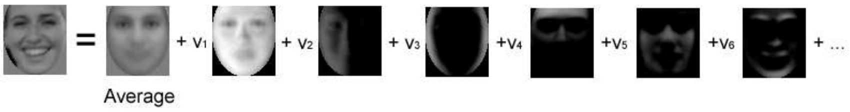

Figure 1. PCA decomposition of an image into the average image and the linear combination of different eigenvectors, known as eigenfaces.

Gabor filters. The linear receptive field (RF) for simple cell responses in the primary visual cortex (V1) can be modeled by two-dimensional Gabor filters [16]. Most of these cells are combined in pairs, one cell of each pair has even symmetry and the other one has odd symmetry [23]. These considerations allow for a definition of a biologically motivated filter [10]:

$$p_k(x) = \frac{k^2}{\sigma^2}\exp(-\frac{k^2}{2\sigma^2}x^2)(\exp(ikx) - \exp(-\frac{\sigma^2}{2})) \tag{1}$$

where k defines the frequency, orientation and location of the filter. These filters have been used in recent years as independent components to represent natural images [20].

The convolution of an image with these filters provides –for each pixel– a vector whose dimension depends on the number of orientations and scales used. In this paper we used 4 different orientations and 4 different scales, as shown in Figure 2. Thus, for each pixel 32 values are obtained (4 scales, 4 orientations and 2 symmetries). Therefore, the convolution yields a representation of higher dimensionality than the original image.

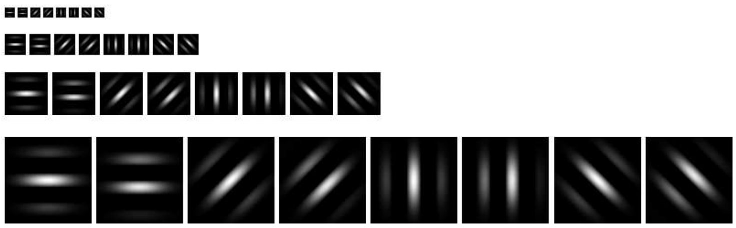

Figure 2. The bank of filters applied sized 11×11, 23×23, 47×47, 95×95.

Because of the resulting high dimensionality, different authors have considered ways to reduce the number of points, orientations, or scales used. In [7] a genetic algorithm accomplishes this reduction for texture classification. For faces, a weighted grid is employed for recognition configuring a representation that is more robust to pose changes [12]. For facial expression recognition the selection of the Gabor filters giving all the possibilities in terms of position, orientation and scale, is performed by an Adaboost approach followed by a SVM-based classifier. This approach yields a system that can provide real-time classification performance [3]. A similar approach for face recognition is described in [30] but in this case, it uses the intra-face and extra-face difference spaces for classification instead of a SVM-based classifier.

In our work we have considered applying the Gabor filters to a smaller number of image points thereby decreasing redundancy and providing faster classification. In order to have some preliminary results, our selection approach is quite simple in contrast to the Adaboost approach used in [3] and [30]. This simple approach performs different scans as described in Figure 3. On each scan the previously stored Gabor points (if any) are combined with each pixel. The location of the pixel that provides the best performance is then added to the list of points only if it improves the previous best rate.

```
change ← true
max_performance ← 0
Resets Gabor points list
while change do
  change ← false
  for each pixel do
    Compute Gabor representation using the Gabor points list and the current pixel location
    Get new_performance for the set
    if new_performance > max_performance then
      change ← true
      max_performance ← new_performance
      Stores current pixel location in a temporary location
    end if
  end for
  if change then
    Adds best pixel location to the Gabor list
  end if
end while
```

Figure 3. Gabor points selection algorithm

3 Experiments

3.1 Datasets and libraries

The dataset contains 6000 face images taken randomly from internet and selected samples from facial databases such as the BIOID [11]. They have been annotated by hand to get their eye positions and labelled according to the different semantic descriptors considered: female/male, clear/dark, glasses/no glasses, moustache/no moustache. These images have been normalized according to eye positions obtaining 59×65 pixels images. Table 1 summarizes the composition of the test and training sets used for the experimental setup.

Every face analyzed is transformed to both face representation spaces described above. For each representation and descriptor a classifier is computed based on the widely used and powerful Support Vector Machine (SVM) approach [28].

Different libraries have been used for these experiments. The OpenCV [14] library provides tools for PCA computation and projection. The Gabor filters have been computed adapting for OpenCV the David Bolme implementation [5]. Finally the SVM classifier was implemented making use of the available LIBSVM [9].

Descriptor	Training set Female	Training set Male	Test set Female	Test set Male
Gender	1223	1523	835	2246
Descriptor	**Training set Clear**	**Training set Dark**	**Test set Clear**	**Test set Dark**
Race	574	316	4811	306
Descriptor	**Training set No**	**Training set Yes**	**Test set No**	**Test set Yes**
Glasses presence	912	692	4042	356
Moustache presence	710	480	4389	426

Table 1. Training and test sets containing 59×65 pixels images. We have tried to build balanced (in number) training sets. For some descriptors one class has not so many samples, for that reason the training set is reduced and therefore the test set has much more samples of the typical class in the dataset: clear skin, no glasses, no moustache.

3.2 PCA + SVM

The PCA space was computed using 4000 samples of the face dataset requiring 12 hours in a PIV 2.2 Ghz. The paper described in [8] analyzed the performance of different classifiers based on a PCA+SVM approach modifying the number of eigenfaces used for classification. The authors concluded that 70 coefficients provide a good trade-off between correct recognition rate and training processing time, see Figure 4.

That said, it can be considered the fact, as already referred by different authors [4], that some eigenfeatures selected have no interest for the problem analyzed. Thus, the first eigenfaces contain generally information related with illumination that is not useful here, or for example some eigenfeatures contain information that may not be discriminant for the glasses presence problem. Other authors have considered a more precise selection of them, for example in [25] the authors do not just take the first n eigenfeatures but select them by means of a genetic algorithm. Instead of this, we have considered the use of a Gabor filters based representation whose results are presented in the next section.

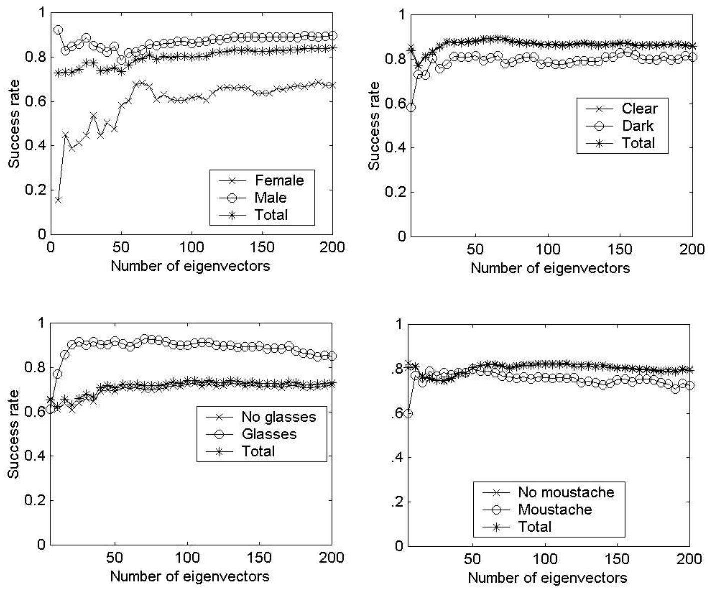

Figure 4. PCA+SVM performance. Top left) gender, top right) race, bottom left) glasses presence, bottom right) moustache presence.

3.3 Gabor filters + SVM

The Gabor location selection approach described above, see Figure 3, was applied to each training set in Table 1 to choose the best configuration based on the Gabor banks. For any of the problems after one scan, the overall hit resulted in at least 99% for the training set. Due to this fact, in the experiments presented here we have selected new points to add based on the performance of the test set (however no image from the test set was used to train the classifier) in order to achieve a longer performance evolution. The final performance and locations achieved for the different problems are depicted in Figures 5 and 6 respectively.

As suggested in Figure 6, the selected points are located in image regions related to the feature being analyzed. In the case of gender, it seems to be a concentration close to the mouth and eyes, which fits with psychophysical results achieved in [13]. The points to check the moustache and glasses presence are selected close to the possible location of those elements in the face (even when all the image pixels were considered). However, we have not yet a clear explanation for the Gabor points selected for race classification.

3.4 Discussion

Table 2 presents the best results achieved for each problem using both approaches, i.e. PCA+SVM and Gabor-based+SVM. The overall performance is

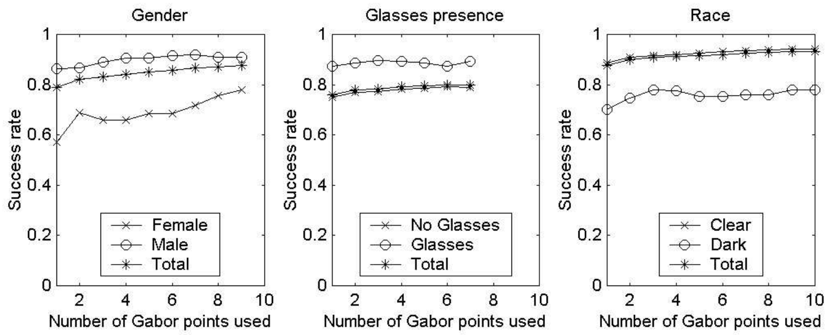

Figure 5. Gabor+SVM. Iterative improvement for the test set for gender, glasses presence and race.

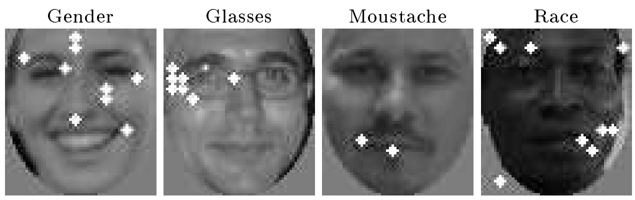

Figure 6. Gabor points location selected by the approach for the different descriptors.

clearly better using the second approach without showing a significant increase in computational cost. This result is due to the fact that only some selected points are used to compute the Gabor features.

However, the first approach makes possible a shorter training stage as the new classifier setup can be built very fast as only a single SVM classifier training is necessary (if the PCA space is considered fixed). On the other hand, the compilation of the Gabor points list which provides the best performance for a given set, requires an undefined number of scans on the image and the computation of multiple SVM classifiers during the selection process. This fact is a disadvantage of this approach if we consider the possibility of an interactive system which were able to learn incrementally, and therefore would need to retrain online.

4 Conclusions and Future Work

A dataset of still face images has been used to analyze the possibility of automatic suggestion of semantic descriptors given a human face image. In some cases these descriptors have not been considered previously by facial analysis literature. Two face representation approaches have been compared, getting better results for

	PCA + SVM			Gabor filters		
Descriptor	Recog. rate	N. eigenvalues	Proc. time	Recog. rate	N. filters	Proc. time
Gender	82.9%	160	5	87.6%	9	7
Race	89.1%	65	1	93.2%	10	5
Glasses presence	72%	70	1	79.8%	7	2
Moustache presence	81.8%	65	1	85.8%	2	1

Table 2. Summary comparing the best classifier for each approach. The processing time required for classification is indicated in milliseconds.

the one which is based on the biologically motivated Gabor filters. The resulting classifiers perform reliably in both cases for real time operation.

Future work should also extend the facial descriptors domain and take into account the possibility of providing a weighted output. For example the race descriptor should not have a binary output. Indeed a degree could be provided and/or some other classes could be added such as Asian, Indian, Hispanic, etc. Additionally a comparison must be performed with other approaches for Gabor banks location selection such as Adaboost and genetic algorithms. This can also be applied to the different eigenfeatures in the first approach, or to the number of Gabor filters, orientations, and scales in the second. A combination of both approaches can also be analyzed.

The face dataset must also be increased due to the fact that currently it does not contain a large number of samples for some of the descriptor classes. We think that a reason to this is that in our main source, i.e. internet, it is easier to gather images corresponding to young, caucassian and good looking (!) people. Therefore, the collection of new samples for non trendy features requires a longer search. Bigger databases would likely be needed to provide better performance, particularly for those descriptors which present a clear border among the different classes, e.g. glasses presence.

Acknowledgments

Work partially funded by research projects Univ. of Las Palmas de Gran Canaria UNI2003/06, UNI2004/10 and UNI2004/25, Canary Islands Autonomous Government PI2003/160 and PI2003/165 and the Spanish Ministry of Education and Science and FEDER funds (TIN2004-07087).

References

1. M. Argyle. *Bodily communication.* Methuen, 2nd edition, 1988.
2. E Bailly-Bailliere, S Bengio, F Bimbot, M Hamouz, J Kittler, J Mariethoz, J Matas, K Messer, V Popovici, F Poree, B Ruiz, and J-P Thiran. The banca database

and evaluation protocol. In J Kittler and M Nixon, editors, *Proc. Audio- and Video-Based Biometric Person Authentication*, pages 625–638, Berlin, June 2003. Springer.
3. M.S. Bartlett, Gwen Littlewort, Ian Fasel, and Javier R. Movellan. Real time face detection and facial expression recognition: Development and applications to human computer interaction. In *Computer Vision and Pattern Recognition*, 2003.
4. P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. on PAMI*, 19(7):711–720, 1997.
5. David Bolme. Elastic bunch graph matching. Master's thesis, Colorado State University, Computer Science Department, June 2003 2003.
6. Vicki Bruce and Andy Young. *The eye of the beholder*. Oxford University Press, 1998.
7. N. W. Campbell and B. T. Thomas. Automatic selection of gabor filters for pixel classification. In *Sixth International Conference on Image Processing and its Applications*, pages 761–765, July 1997.
8. M. Castrillón Santana, J. Lorenzo Navarro, D. Hernández Sosa, and Y. Rodríguez-Domínguez. An analysis of facial description in static images and video streams. In *2nd Iberian Conference on Pattern Recognition and Image Analysis*, Estoril, Portugal, June 2005.
9. Chi-Chung Chang and Chih-Jen Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.
10. J.G. Daugman. Complete discrete 2-d gabor transforms by neural networks for image analysis and compression. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 36(7), July 1988.
11. Robert W. Frischholz and Ulrich Dieckmann. Bioid: A multimodal biometric identification system. *IEEE Computer*, 33(2), February 2000.
12. B. Gokberk, L. Akarun, and E. Alpaydýn. Feature selection for pose invariant face recognition. In *International Conference on Pattern Recognition*, Barcelona (Spain), 2002.
13. F. Gosselin and P. G. Schyns. Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Research*, pages 2261–2271, 2001.
14. Intel. Intel Open Source Computer Vision Library, b4.0. www.intel.com/research/mrl/research/opencv, August 2004.
15. Zhong Jing and Robert Mariani. Glasses detection and extraction by deformable contour. In *International Conference on Pattern Recognition*, 2000.
16. J.P. Jones and L.A. Palmer. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophisiology*, 58(6):1233–1258, 1987.
17. Y. Kirby and L. Sirovich. Application of the karhunen-loève procedure for the characterization of human faces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(1), July 1990.
18. Michael J. Lyons, Julien Budyneck, and Shigery Akamatsu. Automatic classification of single facial images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(12):1357–1362, December 1999.
19. Baback Moghaddam and Ming-Hsuan Yang. Learning gender with support faces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(5):707–711, 2002.
20. Bruno A. Olshausen and David J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.

21. Maja Pantic and Leon J.M. Rothkrantz. Automatic analysis of facial expressions: The state of the art. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(12):1424–1445, December 2000.
22. P. Jonathon Phillips, Hyeonjoon Moon, Syed A. Rizvi, and Patrick J. Rauss. The feret evaluation methodology for face recognition algorithms. TR 6264, NISTIR, January 1999.
23. Daniel A. Pollen and Steven F. Ronner. Phase relationship between adjacent simple cells in the visual cortex. *Science*, 212:1409–1411, June 1981.
24. Pawan Sinha and T. Poggio Torralba. I think i know that face... *Nature*, 384(6608):384–404, 1996.
25. Zehang Sun, George Bebis, Xiaojing Yuan, and Sushil J. Louis. Genetic feature subset selection for gender classification: A comparison study. In *Sixth IEEE Workshop on Applications of Computer Vision*, December 2002.
26. Antonio Torralba. Contextual modulation of target saliency. *Advances in Neural Information Processing Systems*, 2001.
27. M. Turk. Computer vision in the interface. *Communications of the ACM*, 47(1):61–67, January 2004.
28. V. Vapnik. *The nature of statistical learning theory.* Springer, New York, 1995.
29. Bo Wu, Haizhou Ai, and Ran Liu. Glasses detection by boosting simple wavelet features. In *17th Int. Conf. on Pattern Recognition, Cambridge, UK*, pages 292–295, August 2004.
30. Peng Yang, Shiguang Shan, Wen Gao, Stan Z. Li, and Dong Zhang. Face recognition using ada-boosted gabor features. In *Proc. of the 6th International Conference on Automatic Face and Gesture Recognition*, 2004.