Gender recognition from face images with trainable COSFIRE filters

George Azzopardi
University of Malta
george.azzopardi@um.edu.mt

Antonio Greco, Mario Vento
University of Salerno
agreco@unisa.it, mvento@unisa.it

Abstract

Gender recognition from face images is an important application in the fields of security, retail advertising and marketing. We propose a novel descriptor based on COSFIRE filters for gender recognition. A COSFIRE filter is trainable, in that its selectivity is determined in an automatic configuration process that analyses a given prototype pattern of interest. We demonstrate the effectiveness of the proposed approach on a new dataset called GENDER-FERET with 474 training and 472 test samples and achieve an accuracy rate of 93.7%. It also outperforms an approach that relies on handcrafted features and an ensemble of classifiers. Furthermore, we perform another experiment by using the images of the Labeled Faces in the Wild (LFW) dataset to train our classifier and the test images of the GENDER-FERET dataset for evaluation. This experiment demonstrates the generalization ability of the proposed approach and it also outperforms two commercial libraries, namely Face++ and Luxand.

Keywords. Gender recognition, COSFIRE, trainable filters, faces

1. Introduction

In recent years the recognition of the gender from face images has attracted interest in both fundamental and applied research. From the fundamental point of view it is very intriguing to understand how for human beings gender recognition is an effortless operation which is done very rapidly, but for a computer vision algorithm the task could be very challenging. The difficulties emerge from the possible variations of a face captured by a camera [1], which depend on the image acquisition process (pose of the face, image illumination and contrast, background), the intrinsic differences between people's faces (expression, age, race), as well as the occlusions (sunglasses, scarves, hats). From the applied research point of view, there is a commercial interest to have systems that can automatically recognize the gender from face images. Examples include surveillance systems that can assist to restrict areas to one gender only, faster processing in biometrics systems that rely on face recognition, custom user interfaces depending on the gender of the person interacting with them, smart billboards designed to attract the attention of male or female audience, and systems for the collection of data in support of market analysis.

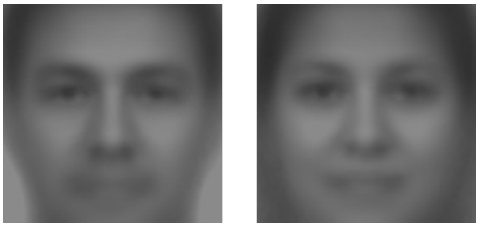


Figure 1. The average face of (a) men and (b) women computed from a subset of the FERET dataset [2].

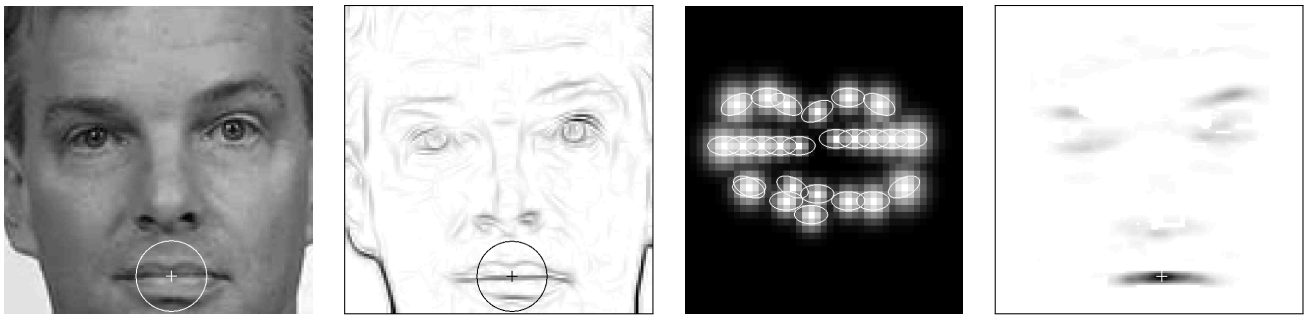
In Fig. 1 we illustrate the average face images of men and women generated from a subset of the FERET dataset [2]. From these images, one may observe differences in the intensity distribution especially in the hair and eyes regions. Based on these observations, many researchers use the pixel intensity values of the faces to train a binary classifier for gender recognition [3, 4, 5].

Further differences can be observed in terms of texture. This could be due to the softer facial features of women and more pronounced eyebrows, while men have a rougher skin especially in the presence of beard. The most popular texture descriptors for face images are the histograms of local binary patterns (LBP) [6, 7, 8].

One may also observe a variation in the shape of the face. The face of a woman is generally more rounded, while the face of a man is more elliptical. In [9], the authors exploited this aspect and proposed the use of histogram of gradients (HOG) descriptor [10] for the recognition of gender. In other works, shape-based features have been combined with other types of features in order to have a more robust classifier [11, 12, 13, 14].

Finally, there are also many subtle differences in the geometry of the faces. The average face of a man has closer eyes, a thinner nose and a narrower mouth. These obser-

978-1-5090-3811-4/16/$31.00



(a) (b) (c) (d)

Figure 2. (a) A training face image of size 128×128 pixels. The encircled region indicates a prototype pattern of interest which is used to configure a COSFIRE filter. The plus marker indicates the center of the prototype. (b) The superimposed (inverted) response maps of a bank of Gabor filters with 16 orientations ($\theta = \{0, \pi/8, \dots 15\pi/8\}$ and a single scale ($\lambda = 4$). (c) The structure of a COSFIRE filter that is configured to be selective for the prototype pattern shown in (a). (d) The (inverted) response map of the concerned COSFIRE filter to the input image in (a). The darker the pixel the higher the response.

vations triggered the investigation of what are known as facial fiducial distances, which are essentially the distances between certain facial landmarks (e.g. nose, eyes contour, eyebrows) [15]. The fiducial points may be detected using active shape model [16] or deep learning techniques [17, 18].

We propose to use trainable COSFIRE (Combination of Shifted Filter Responses) filters [19, 20] for gender recognition from face images. COSFIRE filters have already been found to be highly effective in different computer vision tasks, including contour detection [21, 22], retinal vessel segmentation [23], object localization and recognition [24, 25], and handwritten digit classification [26]. COSFIRE filters are trainable shape detectors. The term trainable refers to the ability of determining their selectivity in an automatic configuration process that analyses a given prototype pattern of interest in terms of its dominant orientations and their mutual spatial arrangement. Our hypothesis is that by configuring multiple COSFIRE filters that are selective for different parts of the faces we can capture the subtle differences that distinguish the faces of men and women.

The remaining part of the paper is organized as follows. In Section 2 we describe how we form COSFIRE-based descriptors. In Section 3 we evaluate their performance on a subset of FERET and compare them with an approach that relies on handcrafted features. We discuss certain aspects of the proposed approach in Section 4 and finally we draw conclusions in Section 5.

2. Method

In the following we give an overview of the trainable COSFIRE approach and show how we use it to form face descriptors. For further technical details on COSFIRE filters we refer the reader to [19].

2.1. COSFIRE filter configuration

The selectivity of a COSFIRE filter is determined in an automatic configuration process that analyses the shape properties of a given prototype pattern of interest. This procedure consists of the following steps. First, it applies a bank of Gabor filters of different orientations and scales to the given prototype image. Second, it considers a set of concentric circles around the prototype center and chooses the local maximum Gabor responses along these circles. The number of circles and their radii values are given by the user. For each local maximum point i the configuration procedure determines four parameter values; the scale λ_i and the orientation θ_i of the Gabor filter that achieves the maximum response at that position, along with the polar coordinates (ρ_i, ϕ_i) with respect to the prototype center. Finally, it groups the parameter values of all points in a set of 4-tuples:

$$S_f = \{(\lambda_i, \theta_i, \rho_i, \phi_i) \mid i = 1 \dots n\} \quad (1)$$

where f denotes the given prototype pattern and n represents the number of local maximum points.

In Fig. 2a we show an image of a face. We use the encircled region as a protoype to configure a COSFIRE filter to be selective for the same and similar patterns. In Fig. 2b we show the superimposed response maps of a bank of Gabor filters which is used in the configuration stage and in Fig. 2c we illustrate the structure of the resulting COSFIRE filter. The ellipses represent the properties of the determined contour parts. Their sizes and orientations indicate the parameters λ_i and θ_i of the concerned Gabor filters. We explain the function of the white blobs in the next section.

2.2. COSFIRE filter response

The response of a COSFIRE filter is computed by combining the responses of the involved Gabor filters indicated in the set S_f. For each tuple i in S_f a Gabor filter with

a scale λ_i and an orientation θ_i is applied. Then, the next step considers the respective Gabor responses at the locations indicated by the polar coordinates (ρ_i, ϕ_i) and applies a multi-variate function to them to obtain a COSFIRE response in every location (x, y) of an input image. For efficiency purposes, in practice the Gabor response maps are shifted by the corresponding distance parameter value ρ_i in the direction opposite to ϕ_i. In this way, all the concerned Gabor responses meet at the support center of the filter.

In order to allow for some tolerance with respect to the preferred positions, the Gabor response maps are also blurred by taking the maximum of their neighbouring responses weighted by Gaussian function maps. The standard deviation σ_i of such a Gaussian function depends linearly on the distance ρ_i from the support center: $\sigma_i = \sigma_0 + \alpha\rho_i$ where σ_0 and α are constants determined empirically on the training set. In Fig. 2c the white blobs indicate the Gaussian function maps that are used to blur the response maps of the corresponding Gabor filters. The standard deviations of the Gaussian functions increase with increasing distance from the support center of the COSFIRE filter.

Finally, the response of a COSFIRE filter r_{S_f} in a location (x, y) is achieved by combining the blurred and shifted Gabor filter responses $s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x, y)$ by geometric mean:

$$r_{S_f} = \left(\prod_{i=1}^{n} s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x, y) \right)^{\frac{1}{n}} \quad (2)$$

In Fig. 2d we illustrate the (inverted) response map of the configured COSFIRE filter to the image in Fig. 2a. For clarity purposes the zero values are rendered as white pixels and the non-zero values are rendered as shades of gray. The darker the pixel the higher the COSFIRE response. The maximum response is correctly obtained in the center of the prototype that was used to configure the concerned COSFIRE filter. The filter, however, achieves other responses (lower than the maximum) to patterns that are similar to the prototype. In general, the filter responds to features that consist of a horizontal edge surrounded by two curvatures pointing outwards.

2.3. Face descriptor

We form a descriptor for face images by using the maximum responses of a collection of COSFIRE filters that are selective for different parts of a face. In the example illustrated in Fig. 2 we demonstrate the configuration and application of one COSFIRE filter that is selective for the central region of the lips. Similarly, we may use other parts of the face to configure more COSFIRE filters. For a given test image we then apply all COSFIRE filters and consider a spatial pyramid of three levels. In level zero we consider only one tile, which is the same size of the given image, in level one we consider four tiles in a 2×2 spatial arrangement and in level two we consider 16 tiles in a 4×4 grid.

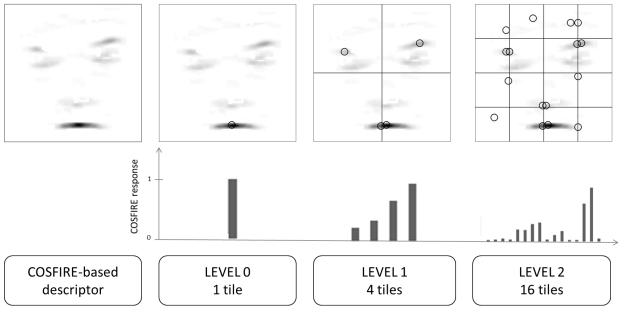


Figure 3. Example of the COSFIRE face descriptor using a single filter. The circles indicate the locations of the maximum filter responses in a three-level spatial pyramid, while the bar plots represent the values of the maximum responses.

For each of the 21 tiles we take the maximum value of every COSFIRE filter. This means that for k COSFIRE filters the descriptor results in a $21k$-element vector. We normalize to unit length the set of k COSFIRE filter responses in each tile. Fig. 3 shows the computation of the 21-element vector after the application of a single filter.

The proposed approach that use the responses of multiple COSFIRE filters for the description of a face is inspired by the hypothesis of population coding in neuroscience. Neurophysiologists believe that a shape is described by the collective response of a set of shape-selective neurons in visual cortex [27]. Further inspiration was obtained from the spatial pyramid matching approach with bags of features [28].

2.4. Classification model

We use the resulting descriptors from the images in a given training set to learn an SVM classification model with the following chi-squared kernel $K(x_i, y_i)$:

$$K(x_i, y_j) = \frac{(x_i - y_j)^2}{\frac{1}{2}(x_i + y_j) + \epsilon} \quad (3)$$

where x_i and y_j are the descriptors of training images i and j, and the parameter ϵ represents a very small value[1] in order to avoid division by zero errors. In practice, we use the libsvm library [29] with the above mentioned custom kernel and for the remaining parameters we use the default values.

3. Evaluation

3.1. Dataset

To the best of our knowledge there is not yet a standard dataset for the evaluation of gender recognition algorithms.

[1] In Matlab we use the function eps

Most of the available datasets are designed for face recognition purposes, and hence they do not make available the gender labels. For this reason we decided to use a subset of the FERET dataset [2], which is publicly available with the name GENDER-FERET[2].

In Fig. 4 we show some examples of faces available in our new dataset, which consists of 946 frontal faces (473 m, 473 f). We randomly divided the dataset into a training set that consists of 237 men and 237 women, and a test set containing 236 men and 236 women. In both the training and the test sets there are faces with different expressions, illumination, skin colour, and backgrounds. The face of a person is represented either in the training set or in the test set but not in both.

3.2. Preprocessing

We applied the Viola-Jones algorithm [30] to every image in the dataset and resized the detected faces to a fixed size of 128×128 pixels.

3.3. Experiments with COSFIRE filters

In the following we evaluate the effectiveness of the proposed approach on the GENDER-FERET dataset. We performed a number of experiments by configuring and using increasing number of COSFIRE filters. In the first experiment we configured 10 filters with the following procedure. First, we randomly chose five training faces of men and five training faces of women. Then, for each randomly picked face we chose a random region of size 19×19 pixels and used it as a prototype to configure a COSFIRE filter. If the selected prototype resulted in a COSFIRE with less than 5 tuples we considered it as not enough salient and chose a new one. The filters were configured with the default parameters $t_1 = 0.1$, $t_2 = 0.75$, $\sigma_0 = 0.67$ and $\alpha = 0.1$ as proposed in [19]. We only mention that in the configuration of the filters we considered Gabor filter responses along three concentric circles and the center point: $\rho = \{0, 3, 6, 9\}$. The sizes of the prototype patterns together with the number and radii of the concentric circles were determined empirically on the training set.

Then we executed further experiments by incrementing the set of COSFIRE filters by 10 at a time up to 250. In Fig. 5 we plot the accuracy rate as a function of the number of filters used. For each set of COSFIRE filters we plot two values, one of which is the training accuracy rate that is achieved by 10-fold cross validation on the training set, and the other one is the accuracy rate obtained on the test set. With only 10 filters that result in a feature vector of $(21 \times 10 =)$ 210 elements we achieved 83.79% and 81.4% accuracy rates on the training and test sets, respectively.

[2] http://mivia.unisa.it/database/gender-feret.zip

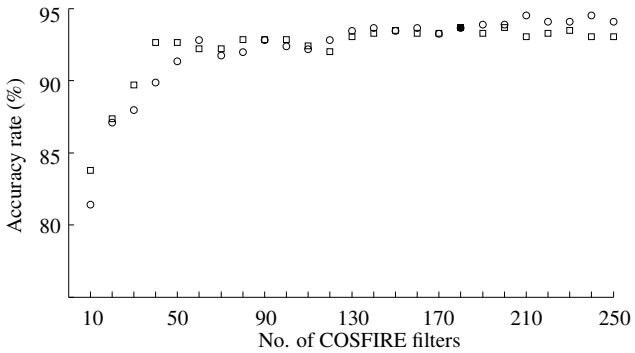


Figure 5. Experimental results in the form of accuracy rate as a function of the number of COSFIRE filters used. The square markers indicate the accuracy rate on the training set with a 10-fold cross validation while the circles indicate the accuracy rates on the test set. The solid square marker indicates the maximum accuracy rate on the training set, which is achieved with 180 filters.

The accuracy increased rapidly up to 60 filters and then increased slowly until it reached a plateau. The maximum accuracy rate of 93.68% on the training set was achieved with 180 COSFIRE filters. By using the same 180 filters we achieved 93.66% accuracy on the test set.

3.4. Comparison with handcrafted features

We compared the proposed *trainable* approach with an approach that relies on two *handcrafted* feature descriptors namely histogram of gradients (HOG) [10] and local binary patterns (LBP) [6], as well as raw pixel values. The selection of these types of features is motivated from the fact that they extract different information from a given image. The HOG descriptors extract information about the edges that essentially describe the shape, LBP descriptors extract information about the texture and the raw pixel values describe the intensity distribution.

3.4.1 Raw pixels

First, we rescaled the intensity values in the range $[0, 1]$ by dividing by 255. Then we reshaped the face images of size 128×128 pixels to a vector of 16384 values.

3.4.2 LBP features

The LBP-based descriptor compares every pixel to its eight neighbours. This resulted in a binary string of eight bits which we converted to a scalar decimal value. Since we used eight neighbours the decimal values had a range of $[0, 255]$. We used a spatial grid of 3×3 and for each tile we generated an L2-normalized histogram of 256 bins. Finally, we concatenated all the histograms in a feature vector with $(256 \times 9 =)$ 2304 elements.

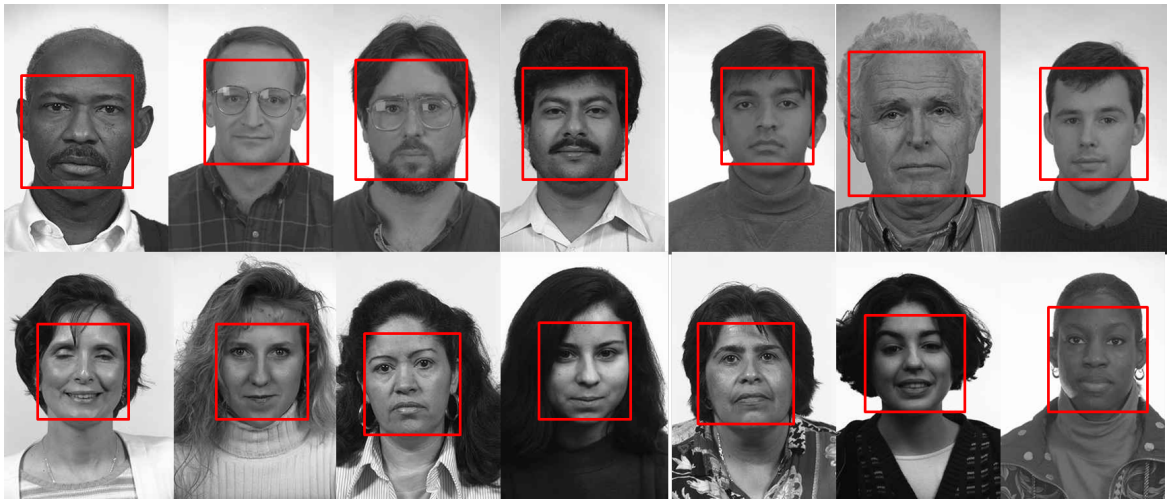


Figure 4. Examples of face images in the GENDER-FERET dataset. The square boxes indicate the faces detections by Viola-Jones [30].

3.4.3 HOG features

For the HOG-based descriptor, we first divided a face image in 49 blocks of 32×32 pixels that overlap by 50%. Then we divided each block in 4 non-overlapping tiles, and for each tile we generated an L2-normalized histogram of orientations with nine bins. We clipped the normalized histograms at 0.2 and normalized again. The result is a feature vector of $(49 \times 4 \times 9 =)$ 1764 elements.

3.4.4 Experiments

For the descriptor that is based on raw pixels we learned an SVM with a linear kernel. For the HOG- and LBP-based descriptors that generated histograms of features, we learned an SVM with a histogram intersection kernel for each of them. We evaluated all possible combinations of these three types of features by fusing the results of the corresponding SVM classifiers. Fusion was achieved by summing up the corresponding output probabilities of the classifiers. If the total male probability was larger than the total female probability then the image was classified as a man, otherwise it was classified as a woman.

Table 1 reports the accuracy rates that we achieved on the test set for different combinations of features.

3.5. Testing generalization capability

We performed another experiment to test the generalization capability of the proposed COSFIRE-based descriptors. We applied the same 180 COSFIRE filters to the training images of the Labeled Faces in the Wild (LFW) dataset [31] and learned an SVM classification model with the chi-squared kernel given in Eq. 3. The LFW dataset is designed for studying the problem of unconstrained face recognition. The face images in that dataset present challenging variations in pose, lighting, race, occlusions, and background. Moreover, the LFW dataset is imbalanced, it consists of 7508 images of men and 2296 images of women.

We then applied the resulting classification model to the test images of the GENDER-FERET dataset, and achieved an accuracy rate of 90%, Table 1. This result gives a good indication of the generalization capability of our method. As a matter of fact, it is slightly higher than what the commercial libraries Face++[32] and Luxand [33] achieve.

3.6. Discussion

The proposed approach with trainable COSFIRE filters outperformed the combined approach of handcrafted features and raw pixels. Our approach achieved an accuracy rate of 93.7% and the best accuracy rate achieved with the other features was 92.6%. The three types of features that were used in the latter approach are complementary to each other as the accuracy increased substantially when combined together.

The COSFIRE-based descriptor that we propose is much more versatile than the handcrafted features. It is based on the configuration of COSFIRE filters with randomly selected local patterns from training images. They do not require domain knowledge and they only expect as input the size of the local patterns used for configuration, something which can be determined empirically. This characteristic makes the proposed COSFIRE-based approach suitable to other computer vision applications.

In Table 1 we report the results obtained with different configurations together with the results of two commercial libraries, namely Luxand [33] and Face++ [32]. These two libraries provide pre-trained classifiers which we used to evaluate the performance on the GENDER-FERET test set.

Table 1. Experimental results. The first column indicates the name of the dataset that was used for training: GF stands for GENDER-FERET and LFW stands for Labeled Faces in the Wild. The headings of the middle four columns indicate the features used to learn SVM classification models with the indicated linear, histogram intersection (H.Int) and chi-squared (χ^2) kernels. The check marks indicate which types of features are used to obtain the corresponding accuracy.

Training Dataset	**Raw Linear**	**LBP H.Int**	**HOG H.Int**	**COSFIRE χ^2**	**Acc %**
GF	✓				88.3
GF		✓			85.2
GF			✓		90.0
GF	✓	✓			90.3
GF	✓		✓		91.9
GF		✓	✓		91.5
GF	✓	✓	✓		92.6
GF				✓	**93.7**
LFW				✓	**90.0**
Mixture	Face++ [32]				89.6
Mixture	Luxand [33]				89.2

The accuracy rates that they achieve are lower than that of our method. We must point out, however, that Luxand and Face++ were trained with a set of images that is different than the one that we used for our approach. In order to simulate their scenario, we performed another experiment where we used the training images of the LFW dataset and the test images of the GENDER-FERET dataset. Also in this experiment, our method showed better effectiveness than the Face++ and Luxand libraries.

The comparison with Face++ and Luxand is interesting because these two libraries use a geometric approach for the detection of facial landmarks and gender recognition. The library Face++ detects the gender by evaluating the position of the fiducial points, identified using a multi-layer convolutional neural network [18]. Luxand FaceSDK can automatically identify a subject's gender based on a still image or motion stream. The SDK uses the coordinates of 66 facial feature points including eyes, eye contours, eyebrows, lip contours and nose tip [33].

There are various directions for future work. One direction is to evaluate the performance of the proposed method on a larger dataset that provides variations also in pose. Another direction is to use a keypoint detector technique, such as Harris affine detector [34] or fiducial points [35] and use the corresponding local patterns as prototypes to configure COSFIRE filters. This approach would provide more informative and possible more distinctive prototype patterns in comparison to the random region detector approach that we use in this work. Moreover, it would be interesting to investigate various functions to transform a COSFIRE filter response map into a descriptor. Here we kept it simple and only used the maximum values in a spatial pyramid.

4. Conclusion

The proposed method that is based on the trainable COSFIRE filters and combined with an SVM of a chi-squared kernel is highly effective for gender recognition from face images. It outperforms an ensemble of three classifiers that rely on the HOG and LBP handcrafted features along with the raw pixel values.

The approach that we propose does not rely upon domain knowledge and thus it is suitable for various image classification tasks.

References

[1] C. B. Ng, Y. H. Tay, and B. M. Goi. A review of facial gender recognition. *Pattern Analysis and Applications*, 18(4):739–755, 2015.

[2] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(10):1090–1104, 2000.

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

[4] S. Baluja and H. A. Rowley. Boosting sex identification performance. *International Journal of computer vision*, 71(1):111–119, 2007.

[5] J. Yang, D. Zhang, A. F. Frangi, and J. Y. Yang. Two-dimensional pca: a new approach to appearance-based face representation and recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(1):131–137, 2004.

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

[7] Z. Yang and H. Ai. Demographic classification with local binary patterns. In *Advances in Biometrics*, pages 464–473. Springer, 2007.

[8] C. Shan. Learning local binary patterns for gender classification on real-world face images. *Pattern Recognition Letters*, 33(4):431–437, 2012.

[9] V. Singh, V. Shokeen, and M. B. Singh. Comparison of feature extraction algorithms for gender classification from face images. In *International Journal of Engineering Research and Technology*, volume 2. ESRSA Publications, 2013.

[10] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

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

[12] J. E. Tapia and C. A. Perez. Gender classification based on fusion of different spatial scale features selected by mutual information from histogram of lbp, intensity, and shape. *Information Forensics and Security, IEEE Transactions on*, 8(3):488–499, 2013.

[13] J. Bekios-Calfa, J. M. Buenaposada, and L. Baumela. Robust gender recognition by exploiting facial attributes dependencies. *Pattern Recognition Letters*, 36:228–234, 2014.

[14] G. Azzopardi, A. Greco, and M. Vento. Gender recognition from face images using a fusion of svm classifiers. In *International Conference Image Analysis and Recognition*, pages 533–538. Springer, 2016.

[15] R. Brunelli and T. Poggio. Face recognition: Features versus templates. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (10):1042–1052, 1993.

[16] S. Milborrow and F. Nicolls. Locating facial features with an extended active shape model. In *Computer Vision–ECCV 2008*, pages 504–513. Springer, 2008.

[17] Y. Sun, X. Wang, and X. Tang. Deep convolutional network cascade for facial point detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3476–3483, 2013.

[18] E. Zhou, H. Fan, Z. Cao, Y. Jiang, and Q. Yin. Extensive facial landmark localization with coarse-to-fine convolutional network cascade. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 386–391, 2013.

[19] G. Azzopardi and N. Petkov. Trainable COSFIRE filters for keypoint detection and pattern recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2):490–503, Feb 2013.

[20] G. Azzopardi, L. Fernandez Robles, E. Alegre, and N. Petkov. Increased generalization capability of trainable cosfire filters with application to machine vision. In *23rd International Conference on Pattern Recognition (ICPR)*, 2016, in print.

[21] G. Azzopardi and N. Petkov. A CORF computational model of a simple cell that relies on lgn input outperforms the gabor function model. *Biological Cybernetics*, 106:177–189, 2012. 10.1007/s00422-012-0486-6.

[22] G. Azzopardi, A. Rodriguez-Snchez, J. Piater, and N. Petkov. A push-pull CORF model of a simple cell with antiphase inhibition improves SNR and contour detection. *PLoS ONE*, 9(7):e98424, 07 2014.

[23] G. Azzopardi, N. Strisciuglio, M. Vento, and N. Petkov. Trainable COSFIRE filters for vessel delineation with application to retinal images. *Medical Image Analysis*, 19(1):4657, 2014.

[24] G. Azzopardi and N. Petkov. Automatic detection of vascular bifurcations in segmented retinal images using trainable COSFIRE filters. *Pattern Recognition Letters*, 34:922–933, 2013.

[25] G. Azzopardi and N. Petkov. Ventral-stream-like shape representation: from pixel intensity values to trainable object-selective COSFIRE models. *Frontiers in computational neuroscience*, 8, 2014.

[26] G. Azzopardi and N. Petkov. A shape descriptor based on trainable COSFIRE filters for the recognition of handwritten digits. In *Computer Analysis of Images and Patterns (CAIP 2013) Lecture Notes in Computer Science*, pages 9–16. Springer, 2013.

[27] A. Pasupathy and C. E. Connor. Population coding of shape in area V4. *Nature Neuroscience*, 5(12):1332–1338, DEC 2002.

[28] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*, CVPR '06, pages 2169–2178, Washington, DC, USA, 2006. IEEE Computer Society.

[29] C. C. Chang and C. J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

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

[31] LFW. Labeled faces in the wild. Available: `http://vis-www.cs.umass.edu/lfw/`, 2007.

[32] Face++. Leading face recognition on cloud. Available: `http://www.faceplusplus.com/`, 2014.

[33] Luxand. Facial feature detection technologies. Available: `https://www.luxand.com/`, 2015.

[34] K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. In *Computer VisionECCV 2002*, pages 128–142. Springer, 2002.

[35] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1701–1708, 2014.