Fusion of domain-specific and trainable features for gender recognition from face images

Azzopardi, George; Greco, Antonio; Saggese, Alessia; Vento, Mario

Published in:
IEEE Access

DOI:
10.1109/ACCESS.2018.2823378

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 4, 2018, accepted April 2, 2018, date of publication April 5, 2018, date of current version May 24, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2823378

Fusion of Domain-Specific and Trainable Features for Gender Recognition From Face Images

GEORGE AZZOPARDI[1], ANTONIO GRECO[2], ALESSIA SAGGESE[2], AND MARIO VENTO[2]
[1]Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, 9712 CP Groningen, The Netherlands
[2]Department of Information Engineering, Electrical Engineering and Applied Mathematics, University of Salerno, 84084 Fisciano, Italy

Corresponding author: Antonio Greco (agreco@unisa.it)

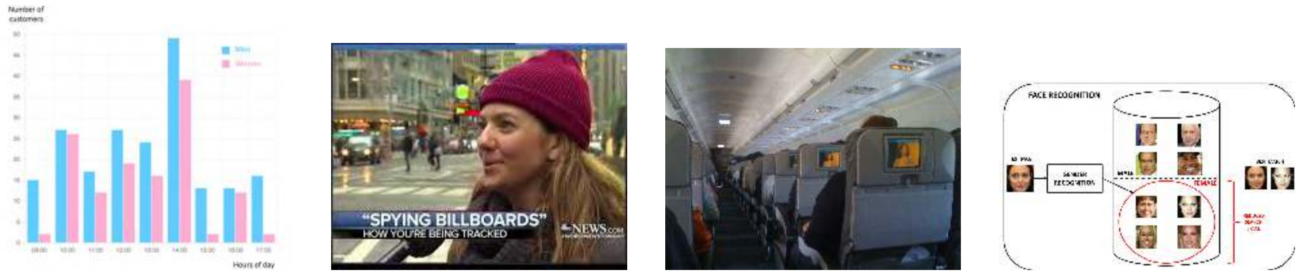
ABSTRACT The popularity and the appeal of systems which are able to automatically determine the gender from face images are growing rapidly. Such a great interest arises from the wide variety of applications, especially in the fields of retail and video surveillance. In recent years, there have been several attempts to address this challenge, but a definitive solution has not yet been found. In this paper, we propose a novel approach that fuses domain-specific and trainable features to recognize the gender from face images. In particular, we use the SURF descriptors extracted from 51 facial landmarks related to eyes, nose, and mouth as domain-dependent features, and the COSFIRE filters as trainable features. The proposed approach turns out to be very robust with respect to the well-known face variations, including different poses, expressions, and illumination conditions. It achieves state-of-the-art recognition rates on the GENDER-FERET (94.7%) and on the labeled faces in the wild (99.4%) data sets, which are two of the most popular benchmarks for gender recognition. We further evaluated the method on a new data set acquired in real scenarios, the UNISA-Public, recently made publicly available. It consists of 206 training (144 male, 62 female) and 200 test (139 male, 61 female) images that are acquired with a real-time indoor camera capturing people in regular walking motion. Such experiment has the aim to assess the capability of the algorithm to deal with face images extracted from videos, which are definitely more challenging than the still images available in the standard data sets. Also for this data set, we achieved a high recognition rate of 91.5%, that confirms the generalization capabilities of the proposed approach. Of the two types of features, the trainable COSFIRE filters are the most effective and, given their trainable character, they can be applied in any visual pattern recognition problem.

INDEX TERMS COSFIRE, face, gender recognition, SURF, trainable features.

I. INTRODUCTION

The face is one of the most important parts of the human body, since it has some distinctive physical and expressive features which allow the identification of certain properties. For instance, by just looking at faces, humans recognize the gender and the ethnicity [1], [2], estimate the age [3], deduce the emotions and the state of mind [4], [5], determine if the person has a familiar face or is a stranger [6], and verify or recognize the identity of the individual [7]–[9]. Although all faces consist of the same parts in a specific spatial arrangement (the relative positions of the nose, eyes and others), primates have enviable abilities to use the subtle features and draw conclusions from faces in a remarkable seemingly and effortless operation. As a matter of fact, there is a neurophysiological evidence that the visual cortices of primates have single neurons that are selective to faces [10]. This fact demonstrates the importance that evolution gave to faces. In recent years the variety and appeal of faces motivated several researchers to work on the problem of automatic face analysis coming from images and videos. In particular, gender recognition is considered among the most common and challenging problems [11].

The algorithms for the automatic classification of gender have a lot of potential for commercial applications (see Figure 1). Indeed, information such as gender, age and race are desired features that managers are eager to have for sophisticated market analysis. That information helps them to acquire more insights about the needs of customers with respect to the possible products that they can offer. Typical examples in the retail are given by the smart billboards or user interfaces which are able to modify their visual display depending on the gender of the person interacting with them.



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

FIGURE 1. Examples of gender recognition applications. (a) A chart that shows the number of the daily visitors in a store, used to carry out market analysis (from *www.aitech.vision*). (b) A woman who talks about smart billboards. (c) Screens installed in front of the passengers in a plane, which can be equipped with smart cameras so as to show tailored promotional material. (d) The architecture of a system which performs a pre-classification of the gender in order to reduce the search space in large face databases.

Another field where gender recognition algorithms can play an important role is video-surveillance. Indeed, there is a great demand for applications that are able to perform face recognition of suspicious individuals by analyzing images captured by surveillance cameras. The main challenge of these systems is the computational time needed for searching a match between the input face image and the thousands of samples stored in a reference database. One way for reducing the search space is to first detect the gender [12] and possibly the age group [13] and/or ethnicity [14] of the given face and then compare it only with the images with the same properties in the database. Although gender recognition from face images may appear a simple task, it is important to note that even human beings may find it challenging in certain situations. In fact, the study [15] demonstrates that the performance of humans in such a task reaches an accuracy lower than 95%. In addition to the inherent difficulties of the topic, the automatic detection and analysis of a face is further affected by different problems. First of all, existing face detection algorithms achieve reasonable performance on frontal faces, but the accuracy gradually decreases when the face is tilted horizontally or vertically with respect to the camera. Certain combinations of facial features also affect gender classification, as shown in Figure 2. The most challenging aspects are surely related to the pose variations and the partial occlusions of the face, for example with scarves, hats and glasses. While the former can be solved to some extent by normalizing the pose of the face using alignment algorithms [16], the latter is harder to solve due to the variety of all the possible occlusions. Furthermore, it has been demonstrated that the performance of gender recognition algorithms is strongly affected by the age of the people, as well as by their race or expression [17]. For instance, the wrinkles formed on elderly women may make their faces similar to elderly men. Figure 2 shows also other challenges, such as variation in the illumination and contrast.

In the last years, a great deal of literature has been produced with methods attempting to solve gender recognition [11]. While significant progress has been observed, automatic systems have not yet reached the generalization capability needed to achieve good performance even in presence of variations in age, race, pose, illumination, and so on.

A. RELATED WORKS

Although it is not possible to define a taxonomy to partition the methods for gender recognition, we can roughly recognize two different classes, depending on the type of features used. Most of the approaches rely on *handcrafted* features, which require expert knowledge for manually designing domain-specific features. Other approaches are indeed *trainable*, in that distinctive features can be automatically learned from training data. The advantage of using handcrafted features is the possibility to exploit the domain knowledge to identify the elements that distinguish the faces of men from those of women, such as intensity, texture, shape and geometry. Indeed, trainable features may capture aspects of the face that a human could not notice. Moreover the procedure for the extraction of such features does not rely upon domain knowledge.

The approaches belonging to the first category use various types of features based on color [18]–[20], texture [21]–[23] and shape [24]–[26] information. Almost all of them share a similar architecture that consists of three steps: i) the detection and the cropping of the face using the well known Viola-Jones algorithm [27]; ii) the pre-processing of the image, in order to normalize the face in terms of dimension, pose and illumination; iii) the extraction of the features used to recognize the gender.

For example, in [18] Moghaddam and Yang propose to use raw information (the pixel intensity values of face images) to form vectors and use them to train a SVM classifier with an RBF or a polynomial kernel. Their main drawback is that they are not invariant to translation. If the same face is shifted by just a few pixels, the resulting feature vector may be completely different. Another problem is the dimensionality of the obtained feature vector, which increases with the resolution of the face image. In order to address this problem, Yang *et al.* [19] compared the performance of various dimensionality reduction techniques applied on the pixel intensity values, such as principal component analysis (PCA), 2D principal component analysis (2DPCA),

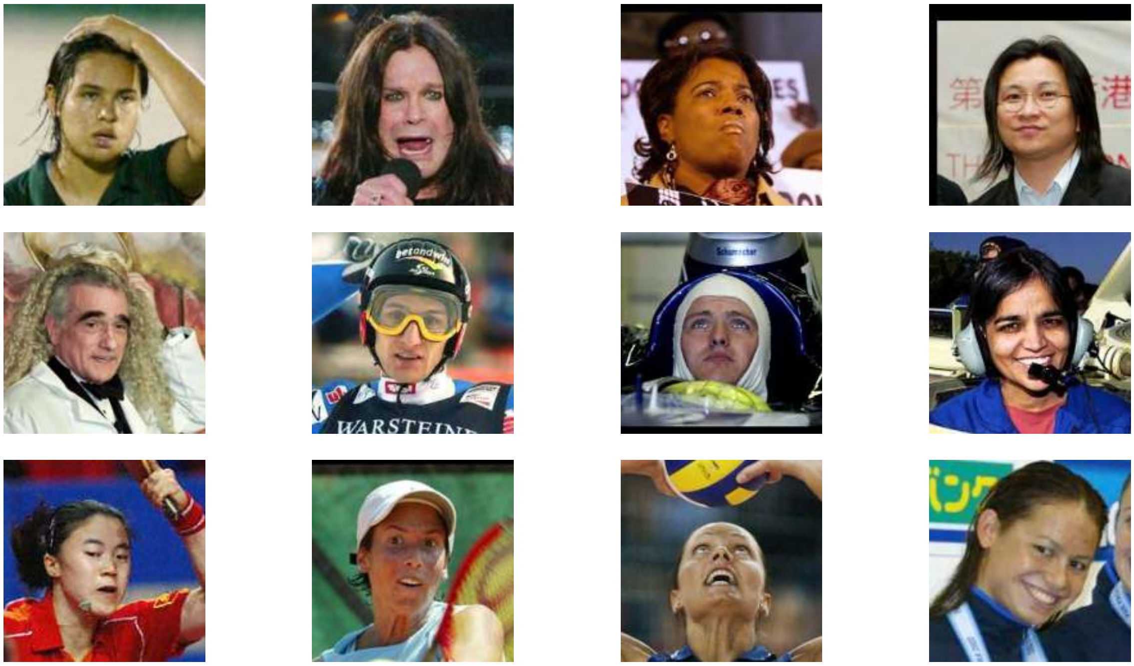


FIGURE 2. **Typical problems of gender recognition systems: (first row) different expressions, ages or races; (second row) occlusions with wig, ski mask, balaclava, and microphone; (third row) different poses and illumination conditions.**

independent component analysis (ICA) and linear discriminant analysis (LDA). In [20] Baluja and Rowley use the relationship between the intensities of the pixels in face images. They consider ten types of a pixel comparison operator, which provide a binary decision, as weak classifiers to learn a model using the Adaboost method [28]. Lian and Lu [21] extract and concatenate local binary pattern (LBP) histograms [29], from different regions of the face, in a single feature vector, and trained a SVM classifier for gender recognition. The rationale of this approach is that a texture descriptor could be able to capture the differences between the smoother skin of a woman and the rougher skin of a man, especially in presence of beard. Eidinger *et al.* [30] and Azarmehr *et al.* [31] use a pair of different LBP variants, FBLBP and MSLBP respectively, for the automatic recognition of gender and age. Dago-Casas *et al.* [32] extract Gabor wavelets from a sparse uniform grid of face points and compute a face descriptor combining them with LBP histograms. Since not all the regions of the face are significant in terms of texture, in [22] and [23] other researchers propose to use Adaboost to carry out feature selection and to use only the LBP histograms from the most discriminant parts of the faces. In [24] Singh *et al.* propose the use of histograms of gradients (HOG) [33] to represent the shape of a face and use it as a descriptor for gender recognition. In [25] Guo *et al.* demonstrate that the performance of a gender classifier based on HOG features is affected by age. This idea is further investigated in [26], where the authors find dependencies among facial demographic attributes, especially between gender, age and pose facial attributes. Other researchers also try to combine several typologies of color, shape and texture features, in order to improve the performance of their gender classifiers [34]–[36]. The rationale behind those approaches is that color, texture and shape features can be complementary, as demonstrated in [34], in the sense that they capture different aspects of human faces and can improve gender recognition when used together.

Other domain specific approaches rely on the extraction of handcrafted features from specific points, known as fiducial points [37]. Brunelli and Poggio [38] propose a face descriptor that computes 18 fiducial distances between points representing the locations of the eyes, nose, chin, mouth corners and others. el-Din *et al.* [39] extract SIFT [40] descriptors from these so-called facial landmarks and used them to form a long feature vector.

As for the second category, the deep learning-based methods [41]–[44], which gained popularity in recent years, are the most common trainable approaches. Levi and Hassner [41] perform automatic age and gender classification using deep-convolutional neural networks (CNN) [45]. van de Wolfshaar *et al.* [42] train a dropout-SVM using the deep features selected by a CNN. Ranjan *et al.* [43] propose a multi-task learning framework, called HyperFace, for simultaneous face detection, landmark localization, pose estimation and gender recognition using CNNs. Jia and Cristianini [44] design a method to generate a classifier using a mixture of face datasets composed of four million images and about 60, 000 features.

Another trainable approach recently applied to gender recognition is based on COSFIRE filters [46]. The main

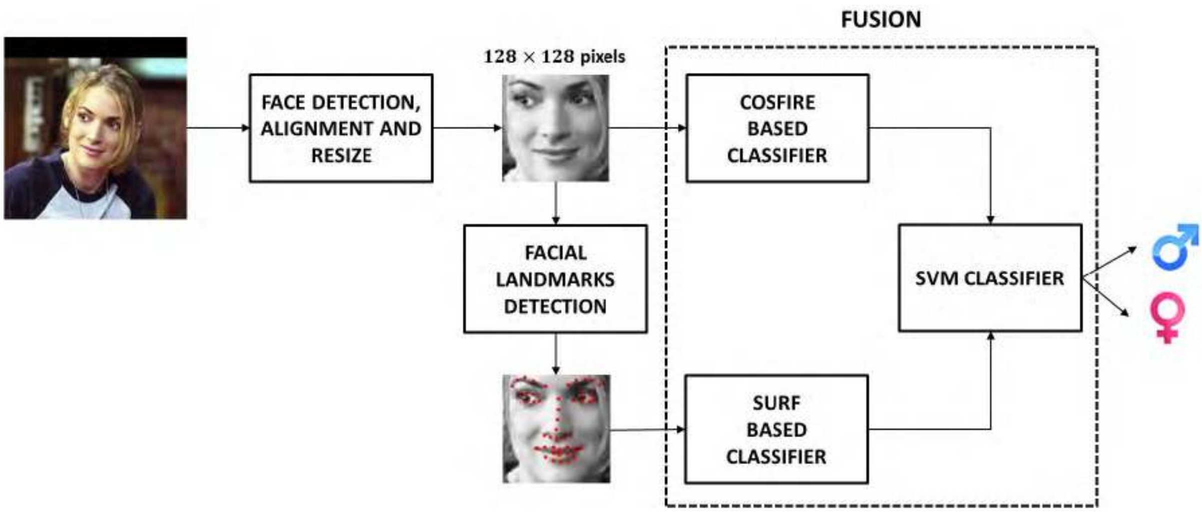


FIGURE 3. Architecture of the proposed method.

advantage of such filters is that, unlike deep learning architectures, only a small set of training images are required for the configuration step. COSFIRE filters, essentially shape-selective filters, are trainable in the sense that their shape selectivity is determined in an automatic configuration procedure using prototype patterns of interest. In [47] 180 COSFIRE filters were configured by using a pseudo-random procedure to select the prototype patterns of interest. For a given image, a face descriptor was then obtained by taking the maximum responses of the resulting 180 COSFIRE filters in a spatial pyramid with three levels [48]. The method was found to be very effective on frontal faces and its generalization capability was demonstrated by carrying out a cross-dataset experiment.

Going back to the considerations about the pros and cons of using domain-specific or trainable features, in this work we propose to combine the domain-free and trainable COSFIRE filters, configured using aligned images, with the handcrafted SURF features [49] extracted from 51 facial landmarks related to eyes, nose and mouth. Hereinafter we refer to these methods as *COSFIRE-based* and *SURF-based*, respectively. We expect that such features are complementary since they capture, in principle, different aspects of the human face. The COSFIRE-based classifier should be able to find the differences in the shape of male and female faces, while the SURF-based classifier should rely on the descriptors extracted from the facial landmarks to discriminate the local differences between the faces of men and women. For these reasons the proposed method should increase the robustness to the aforementioned face variations. Moreover, the fusion of trainable and handcrafted features should ensure a better tradeoff between generalization and specificity. To the best of our knowledge the proposed approach is the first one that combines domain-independent features with domain-specific ones.

The rest of the paper is organized as follows. In Section II, we provide a description of the multi-expert architecture together with a description of COSFIRE- and SURF-based features. In Section III we report the results achieved by the proposed approach on two benchmark datasets and on a new dataset collected in the University of Salerno. Finally, in Section IV we provide a discussion of some aspects of our method before drawing conclusions in Section V.

II. THE PROPOSED METHOD

Figure 3 illustrates the architecture of the proposed system. Note that we do not perform any pre-processing to the given images. Preliminarily, we apply the Viola-Jones algorithm [27] on the input image to detect the faces occurring in it. Then we use a face alignment algorithm to normalize the pose and resize the resulting cropped faces to 128 × 128 pixels. Such resolution allows to maximize the accuracy of the subsequent steps, namely the COSFIRE filters configuration and the facial landmarks extraction. The domain-independent COSFIRE-based method is applied directly to the obtained face image. Indeed, for the SURF-based method we detect 51 facial landmarks belonging to the eyes, the nose and the mouth and extract the SURF descriptors at the keypoints indicating the facial landmarks. Finally, we fuse the outputs of the SURF-based and COSFIRE-based classifiers in another SVM classifier.

In the next sections we provide more details concerning each of the above mentioned components.

A. FACE DETECTION AND ALIGNMENT

For a given image, we use the Viola-Jones algorithm [27] to detect all faces in different scales without applying preliminary pre-processing filters. Subsequently, we use the method proposed in [50] to detect a set of 51 facial landmarks. The original algorithm is able to find 68 fiducial points, but we noticed that sometimes the bounding box of the face, namely the output of the Viola-Jones algorithm, excludes any of the 17 points which belong to the face contour, so we decided

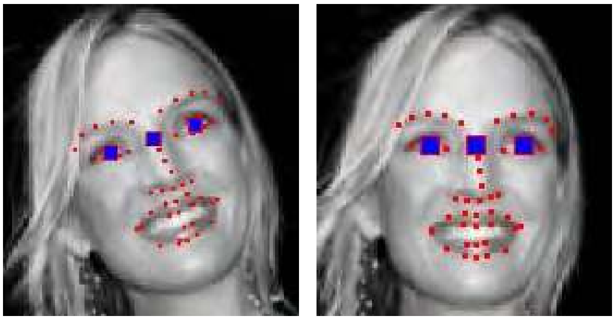


FIGURE 4. Representation of the proposed face alignment algorithm. The 51 red dots indicate the positions of the facial landmarks. The three blue markers, from left to right, indicate the left eye center, the center of the line that connects the two eyes, and the right eye center.

to exclude them. Then we compute the average location of each of the two sets of eye-related landmarks. This allows to determine the orientation of the line which connects these two points, namely the orientation of the face, and use that angle to horizontally align the face image. We rotate the image around the center of the line that connects the two eyes. Figure 4 depicts an example of a face image before and after the alignment. In practice, in order to avoid having a black background in the rotated image, we first crop the face image by using a bounding box that is twice as large as the one determined by the Viola-Jones algorithm. Then we rotate the image and use the Viola-Jones bounding box to crop the face in the rotated image. Finally, we resize each horizontally aligned face image to a fixed size of 128 × 128 pixels.

B. COSFIRE-BASED CLASSIFIER

Trainable COSFIRE filters have been demonstrated to be effective in various computer vision applications, including object localization and recognition [46], [51], [52], vessel-like segmentation [53], [54], and contour detection [55], [56]. A COSFIRE filter is trainable, in that its selectivity is determined in a one-step configuration process that automatically analyzes a given prototype pattern of interest. The resulting non-linear COSFIRE filter can then be applied to images in order to localize patterns that, to a certain extent, are similar to the prototype. Below we describe briefly the required processing to configure and apply COSFIRE filters, and subsequently use their responses to form feature vectors.

1) CONFIGURATION OF A COSFIRE FILTER

The idea of a COSFIRE filter is to combine the responses of some low-level detectors that are selective for simple features in order to determine the selectivity for a more complex feature or a shape. In [46], for instance, it was shown that by combining the responses of orientation-selective Gabor filters at certain positions, one could configure a COSFIRE filter that is selective for a complex shape, such as a traffic sign. By simply changing the input low-level detectors from Gabor filters to difference-of-Gaussians, one could achieve very effective contour [55], [56] and vessel [53], [54] detectors. These two types of COSFIRE filters are essentially shape detectors and do not take colour into account. Recently, a new type of COSFIRE filters were proposed which take input from color blob detectors and have been found to be more effective than Gabor-based COSFIRE filters in object recognition datasets where colour plays an important role [52].

In this work we use the original Gabor-based type of COSFIRE filters [46] as we do not consider colour to be a distinctive feature for gender recognition. In an automatic configuration process we first apply a bank of Gabor filters with eight orientations and five scales and superimpose their responses. Then we consider a number of concentric circles around a point of interest and determine the positions along these circles at which we obtain local maxima Gabor responses. For each such a point we form a tuple with four parameters $(\lambda, \theta, \rho, \phi)$, where λ nd θ denote the scale and orientation, respectively, of the Gabor filter that achieves the maximum response at that position, while ρ and ϕ, respectively, denote the distance and polar angle with regards to the point of interest. Finally, we denote by $S_f = \{(\lambda_i, \theta_i, \rho_i, \phi_i) \mid i \in \ 1 \ldots n\}$ a set that contains the 4-tuples that represent all n points at which we achieve local maximum Gabor responses.

The center point used in a given prototype is the position at which the resulting COSFIRE filter will obtain the maximum response. It can either be specified manually or selected automatically. For the application at hand, we choose such locations randomly in the training face images and use their surroundings as local prototype patterns to configure COSFIRE filters. In this way, a COSFIRE filter is selective for a small part of a face. Figure 5 shows the configuration procedure of two COSFIRE filters by using parts of the eyebrows as prototype patterns selected from a male and a female face images.

2) RESPONSE OF A GABOR-BASED COSFIRE FILTER

The response of a Gabor-based COSFIRE filter is computed in four simple steps, namely filter-blur-shift-multiply. In the first step we determine the unique pairs of the parameters (λ, θ) from the set S_f and apply Gabor linear filtering in the Fourier domain with those parameter values. Secondly, in order to allow for some tolerance with respect to the preferred positions, for the $i-$th tuple we blur the corresponding Gabor response map with a Gaussian function whose standard deviation σ_i is a linear function of the distance ρ_i. In practice, we use the linear function $\sigma_i = \sigma_0 + \alpha\rho_i$, with σ_0 and α set to the default values ($\sigma_0 = 0.67$, $\alpha = 0.1$) proposed in [46]. Thirdly, we shift each blurred Gabor response by the polar vector $(\rho_i, -\phi_i)$, so that all afferent Gabor responses meet at the support center of the concerned COSFIRE filter. Finally, we use the geometric mean function, essentially multiplication, to combine all blurred and shifted Gabor responses and come to a scalar value in each position of a given image. Figure 5(d) and Figure 5(h) show the response maps of the two COSFIRE filters applied to the two images from which we use some local patterns for their

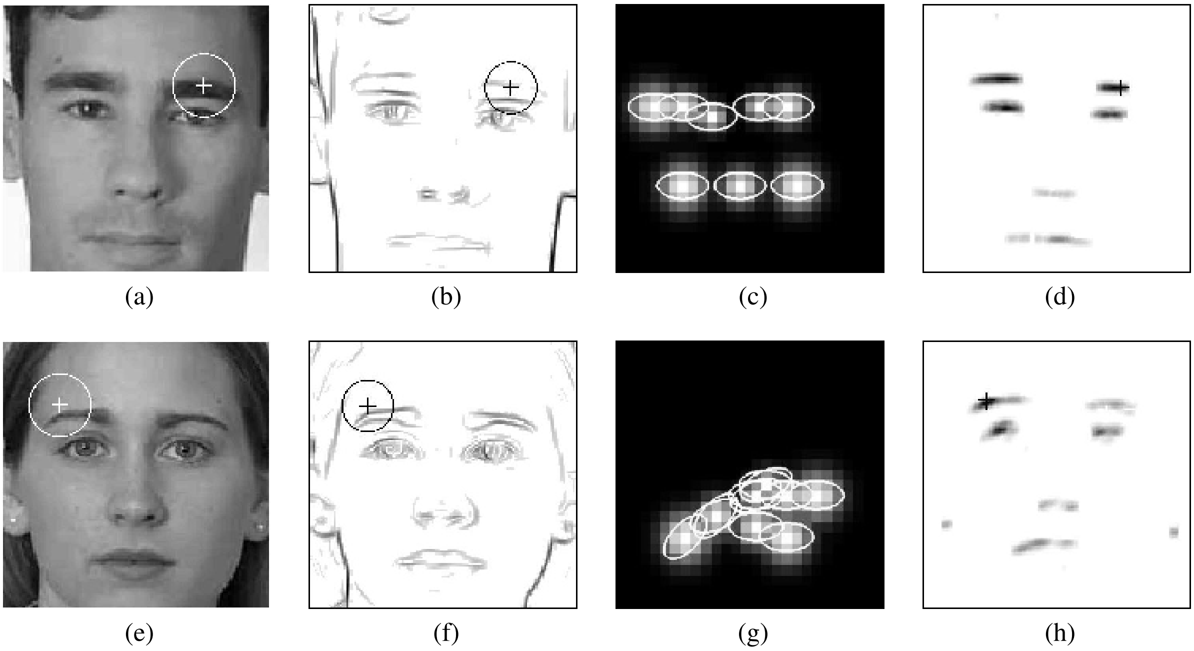


FIGURE 5. Configuration of two COSFIRE filters using (a-d) a training male face image and (e-h) a training female face image, both of size 128 × 128 pixels. (a,e) The encircled regions indicate the prototype patterns of interest which are used to configure the two COSFIRE filters. (b,f) The superposition of inverted response maps of a bank of Gabor filters with 16 orientations ($\theta = \{0, \pi/8, \ldots 15\pi/8\}$) and a single scale ($\lambda = 4$). (c,g) The structures of the COSFIRE filters that are configured to be selective for the prototype patterns indicated in (a) and (e). (d,h) The inverted response maps of the concerned COSFIRE filters to the input face images in (a) and (e). The darker the pixel the higher the response.

configuration. They respond in locations where the local patterns are very similar to the eyebrow prototype parts.

In [46], it was also demonstrated how tolerance to rotation, scale and reflection could be achieved by the manipulation of parameter values. These invariances are, however, not necessary for this application.

3) FORMING A FEATURE DESCRIPTOR AND LEARNING A CLASSIFICATION MODEL

By using k local patterns that we randomly select from the training face images, we configure k COSFIRE filters that are selective for different parts of male and female faces. For a given face image we then apply the collection of k COSFIRE filters and use a spatial pyramid of three levels from which we take the COSFIRE filter responses. In level zero, where there is only one tile, we take the global maximum responses of all COSFIRE filters across the entire image. In level one and level two, we divide each COSFIRE response map, respectively, into $(2 \times 2 =)$ 4 and $(4 \times 4 =)$ 16 tiles and take the maximum response in each tile. For k COSFIRE filters and a spatial pyramid of $(1 + 4 + 16 =)$ 21 tiles we describe a face image with a $21k-$element feature vector. We normalize to unit length the set of k COSFIRE filter maximum responses per tile. Figure 6 depicts the spatial pyramids of the two above configured COSFIRE filters obtained from a test female face image and a bar graph with the values of the resulting $(21 \times 2 =)$ 42-elements descriptor.

We use the $21k-$element feature vectors of all training images to train an SVM classification model with the following chi-squared kernel:

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

where x_i and y_j are the feature vectors of the $i-$th and the $j-$th training images, while the parameter ϵ[1] represents a very small value and it is used to avoid numerical errors. This COSFIRE-based descriptor that we propose is inspired by the concept of population coding from neuroscience [57] as well as from the spatial pyramid matching approach [48].

C. SURF-BASED CLASSIFIER

For the hand-crafted and domain specific approach we compute the 128-element SURF descriptors of the inner 51 facial landmarks, namely the points related to eyes, nose and mouth. Figures 3 and 4 show graphic representations of the fiducial points of two faces. Then we normalize to unit length each descriptor. This approach results in a $(51 \times 128 =)$ 6528-element feature vector for each face image.

Finally, we use these vectors to train an SVM classifier with a linear kernel.

D. ENSEMBLE OF SVM CLASSIFICATION MODELS

We propose to combine the decisions made by the COSFIRE- and SURF-based classifiers using a stacked classification scheme. We use the output scores of the COSFIRE- and SURF-based SVM classifiers achieved from the training

[1] In practice we use the Matlab in-built function *eps*

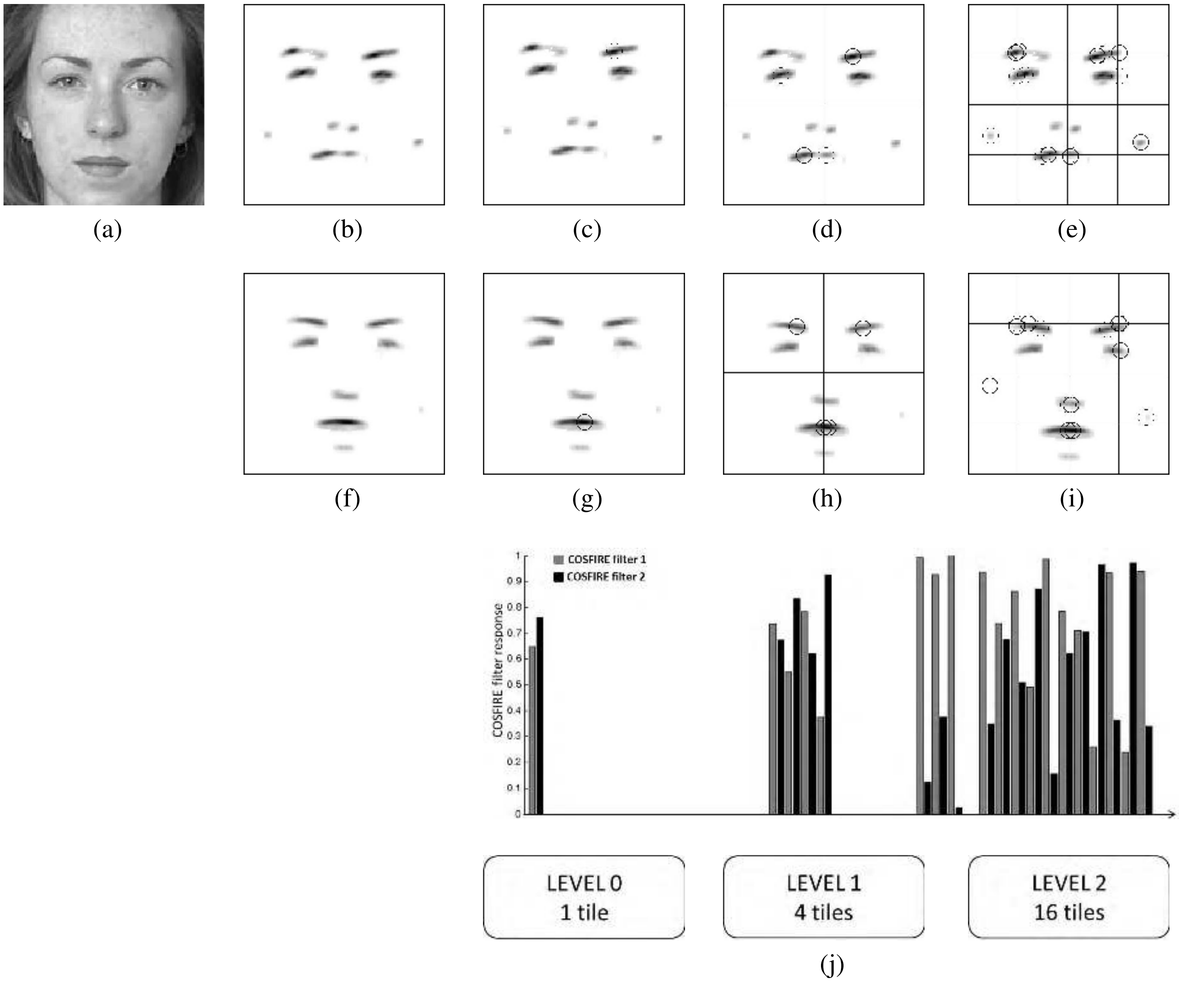


FIGURE 6. **Application of the two COSFIRE filters to a test face image. (b,f) Consideration of three-level spatial pyramids to the response maps of the COSFIRE filters. (c,g) In level zero we consider only one tile, which has the same size of the given image. (d,h) In level one we consider four tiles in a 2×2 spatial arrangement. (e,i) In level two we consider 16 tiles in a 4×4 grid. For each of the 21 tiles the circle indicates the location of the maximum response. (j) The resulting face descriptor, which consists of ($21 \times 2 =$) 42 values. The responses are normalized to unit length for each tile.**

images as feature vectors to learn another SVM with a linear kernel. The final layer determines the classification of a given test image.

III. EXPERIMENTAL RESULTS

In this section we provide details on the datasets used together with the adopted experimental setup for the evaluation of the proposed method, followed by the results.

A. DATASETS

We use three different datasets, namely GENDER-FERET [34], LFW [58] and UNISA-Public [59].

While in the last years several datasets have been proposed for benchmarking gender recognition algorithms, to the best of our knowledge there is no dataset which has become a standard de facto. Also, often there is not a standard way for evaluating algorithms on the existing datasets, as the partitioning of training and test sets is not clearly defined. This makes it hard to compare the performance of a new algorithm with existing ones.

In order to aim for some standardization, we use the GENDER-FERET subset that we created from the well known FERET [60] dataset and already published in [34]. The GENDER-FERET dataset has a balanced number of male and female face images and it is pre-partitioned into 474 training (237 males and 237 females) and 472 test (236 males and 236 females) images. This dataset consists of frontal faces acquired in controlled conditions with different illumination, background, age, expression and race. Moreover, it contains only one face for each person, which can be present either in the training or in the test set, but not in both. Figure 7 shows some examples of face images from the GENDER-FERET dataset.

In order to test the proposed method on faces with more different poses and in order to evaluate the impact of face alignment technique, we also use the Labeled Faces in the Wild (LFW) dataset [58]. LFW contains more than 13,000 images of 5,749 subjects designed to study the problem of face recognition in uncontrolled conditions. The images show famous people busy in different activities, such as recording

FIGURE 7. Examples of GENDER-FERET images.

FIGURE 8. Examples of LFW faces (first row) and the corresponding aligned images (second row).

an interview, playing sports, doing a fashion show and others. In our experiments we aligned the LFW face images using the algorithm described in Section II-A. Figure 8 shows four original LFW images and the corresponding aligned faces.

As recommended in [23], [32], and [36], we performed a 5-fold cross validation and computed the average accuracy of the proposed method.

The LFW images are unconstrained pictures and are taken using professional cameras. For this reason, the dataset consists of high resolution faces without motion blur, which is in contrast to real video based systems that come with typically lower resolution and with motion blur. To the best of our knowledge there is not a publicly available dataset of face images captured by a video camera under realistic conditions. The lack of such a dataset restricts researchers from an elaborative evaluation of their algorithms. In order to test the proposed method in scenarios of everyday life, we used a new dataset, namely *UNISA-Public*, that is publicly available for benchmarking purposes. The dataset contains 406 faces of 58 different persons (42m, 16f) captured while walking towards a fixed camera. We detected the faces using the Viola-Jones algorithm and cropped the faces with bounding boxes that are twice as large as the detected ones. The expansion of the bounding boxes allow us to perform face alignment without adding black backgrounds to the rotated images. The UNISA-Public dataset exhibits variations not covered by the other datasets. Indeed, the illumination conditions are quite controlled since the original videos were acquired in an indoor environment close to the entrance door of a building. The persons are, however, unaware of the camera and they approach it without a precise rule, so their faces are detected with different poses and expressions. Sometimes the individuals also make sudden movements that cause motion blur. Furthermore, the face images have different sizes depending on the distance of the detected persons from the camera. Examples of images from the UNISA-Public dataset are shown in Figure 9. For our experiments we randomly partitioned the dataset in two subsets, the training and the test set. In both sets we collected the face images of 29 persons, namely 21 men and 8 women, so as to preserve the a priori distribution of males and females of the whole dataset.

B. EXPERIMENTS

In this section, we evaluate the effectiveness of our method on the three aforementioned datasets. First, we report the results of the COSFIRE- and the SURF-based methods separately. Finally, we show the results achieved by considering the two heterogeneous classifiers as base classifiers and use their predicted values as input of a meta classifier.

1) RESULTS WITH COSFIRE-BASED CLASSIFIER

We performed the experiments with the COSFIRE-based method following the same procedure reported in [47]. Further to the findings in [47] we configured 180 COSFIRE filters for both the GENDER-FERET and the LFW data sets. In practice, we configured 90 COSFIRE filters from randomly selected male face training images and 90 from randomly selected female face training images. Since the UNISA-Public dataset has less than 90 face training images from the minority female class we could not configure 90 COSFIRE filters per class. Having only 62 female training images we decided to configure 60 COSFIRE filters for each gender class, amounting to a total of 120 COSFIRE filters for the UNISA-Public dataset. Then, for each image, we designed a pseudo-random procedure to select a region of 19×19 pixels, as we did in [47], and used it as a prototype pattern of interest to configure a COSFIRE filter. The pattern was considered valid if the corresponding COSFIRE filter resulted in at least five defining features (i.e. five tuples). Otherwise we chose a new prototype and repeated the procedure until that condition was respected. For the configuration of the COSFIRE filters we used the same parameters as reported in [47]: $\rho = \{0, 3, 6, 9\}$, $t_1 = 0.1$, $t_2 = 0.75$, $\sigma_0 = 0.67$, $\alpha = 0.1$.

Table 1 shows the results of the COSFIRE-based method on the three benchmark datasets. It consistently outperforms the domain-specific method that uses the SURF descriptor for the fiducial landmarks. In particular, it achieves an accuracy of 94.1% on the GENDER-FERET dataset, an accuracy of 89.9% on the UNISA-Public dataset and a remarkable accuracy of 99.3% on the LFW dataset.

2) SURF-BASED RESULTS

The SURF-based algorithm does not require the configuration of any parameters. Table 1 shows that this approach, like the COSFIRE-based, achieves a high accuracy rate on the LFW dataset (96.1%), moderate performance on the GENDER-FERET dataset (89.2%) and a rather low accuracy on the UNISA-Public dataset (82.9%).

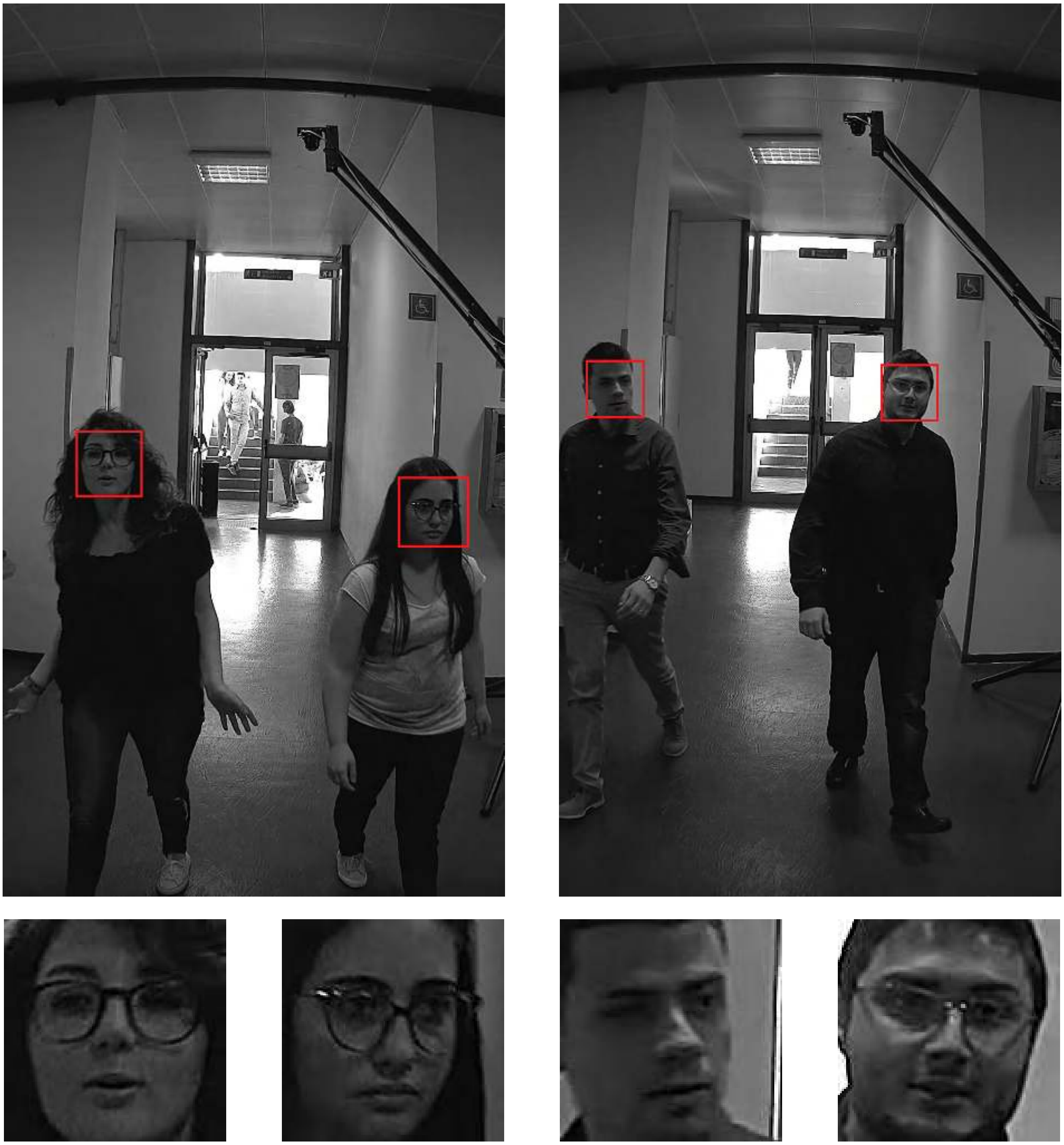


FIGURE 9. Examples of images taken from the UNISA-Public dataset. The images in the first row are the original frames captured by the camera. The face images in the second row are the automatically detected faces that we used as input to the proposed gender recognition algorithm.

3) FUSING COSFIRE- AND SURF-BASED CLASSIFIERS

In this section we analyse the complementarity of the two considered features, namely the trainable COSFIRE- and the domain-specific SURF-based approaches. In particular, Figure 10 gives an idea of the performance achievable by combining the decisions of the proposed classifiers over the three considered datasets. The ideal fusion technique would be the one that is able to take the right decision when at least one of the methods classifies the gender correctly. This happens in two situations: (1) both the decisions are correct (in blue in the figure); (2) only one of the two decisions is correct (in red in the figure). Such experiment proves the complementarity of the COSFIRE- and the SURF-based features, since the fusion may achieve more than 99% of accuracy on the GENDER-FERET and LFW datasets and almost 96% on the UNISA-Public.

In order to test the effectiveness of the proposed stacked classification scheme, we compare it with other two well

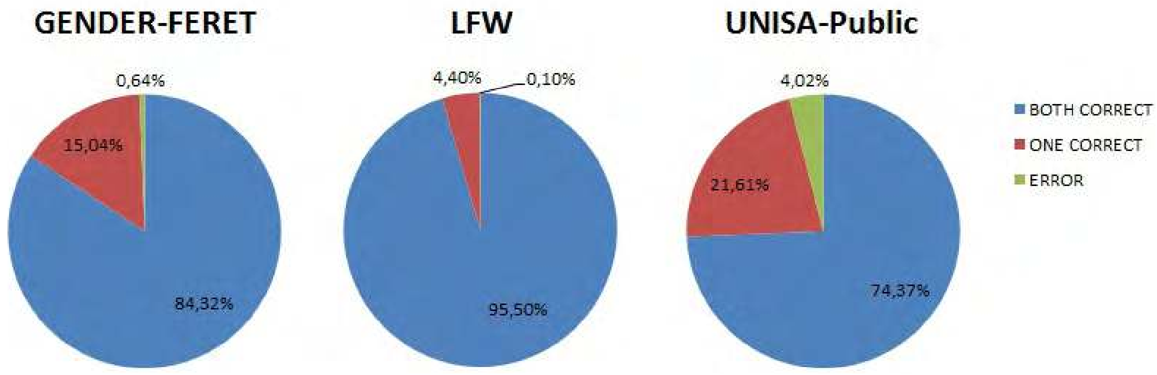


FIGURE 10. Accuracy achievable by fusing COSFIRE- and SURF-based classifiers.

TABLE 1. Results of the COSFIRE- and SURF-based methods on the GENDER-FERET, LFW and UNISA-Public datasets.

Dataset	Method	Accuracy (%)
GENDER-FERET	COSFIRE-based	**94.1**
	SURF-based	89.2
LFW	COSFIRE-based	**99.3**
	SURF-based	96.1
UNISA-Public	COSFIRE-based	**89.9**
	SURF-based	82.9

known fusion techniques, namely the majority voting and the weighted voting schemes. For the former, we trust the decision of the SVM classifier that achieves the highest absolute score. For the latter we consider the accuracies on the training sets achieved by the COSFIRE- and SURF-based classifiers as prior probabilities and use them for the final classification. Suppose $P_C(TR)$ and $P_S(TR)$ represent the prior probabilities of the COSFIRE- and SURF-based classifiers, then the weighted scores of both classifiers are computed by multiplying $P_C(TR)$ and $P_S(TR)$ with the corresponding absolute SVM scores. The classifier which yields the highest absolute weighted score is entrusted. Table 2 reports the results achieved on the three datasets using the three different fusion methods. The highest improvement is observed on the UNISA-Public dataset where the error rate is reduced by almost (1.6/10.1) 16%. This substantial improvement indicates that the stacked classification scheme is rather effective, even if it is not able to achieve the ideal performance depicted in Figure 10.

Notable is the fact that the stacked classification fusion method outperforms the other two methods in all three datasets. The majority voting and the weighted voting schemes obtain almost the same performance and, even, achieve lower accuracies results than that of the COSFIRE-based method on the LFW dataset.

4) COMPARISON WITH OTHER METHODS

In order to prove the effectiveness of the proposed approach with respect to other methodologies, we perform a comparative analysis. However, since the UNISA-Public is a new dataset, we can only use GENDER-FERET and LFW, that have already been used by the scientific community.

TABLE 2. Results with three fusion techniques on the GENDER-FERET, LFW and UNISA-Public benchmark datasets.

Dataset	Fusion technique	Accuracy (%)
GENDER-FERET	Majority voting	94.5
	Weighted voting	94.3
	Stacked classification	**94.7**
LFW	Majority voting	99.0
	Weighted voting	99.0
	Stacked classification	**99.4**
UNISA-Public	Majority voting	90.9
	Weighted voting	90.9
	Stacked classification	**91.5**

TABLE 3. Comparison of the results on the GENDER-FERET dataset.

Method	Description	Accuracy (%)
Azzopardi et al. [34]	RAW LBP HOG	92.6
Azzopardi et al. [47]	COSFIRE	93.7
Proposed	COSFIRE SURF	**94.7**

Table 3 shows the performance comparison with two methods on the GENDER-FERET dataset. Our approach outperforms the method proposed in [34], which exploits the use of pixel intensity values, texture and shape features. Moreover, the new algorithm improves the performance of our previous method, based only on COSFIRE-based features [47]. The improvement is attributable to two factors. Firstly, the face alignment algorithm contributes to an improved accuracy (94.1% vs 93.7%). Secondly, the fusion with the SURF-based method further improves the results.

In Table 4 we compare our results with existing approaches. The approach that we propose is more effective than the methods proposed by Dago-Casas *et al.* [32] and Shan [23], while it outperforms the approach proposed by Tapia and Perez [36]. As we can see from the table, all the methods use an SVM classifier. Thus, the main improvement is due to the combination of the chosen descriptor, which proved to be very representative for the problem at hand.

TABLE 4. Comparison of the results on the LFW dataset.

Method	Description	Accuracy (%)
Dago-Casas et al. [32]	Gabor	94.0
Shan et al. [23]	Boosted LBP	94.8
Tapia and Perez [36]	LBP	98.0
Proposed	COSFIRE SURF	**99.4**

IV. DISCUSSION

The experimental results confirm the effectiveness of the proposed approach for gender recognition from face images. We demonstrated that the fusion of domain-specific (SURF-based) and trainable (COSFIRE-based) features allows to find a good tradeoff between generalization and specificity. Indeed, the proposed method achieves a remarkable accuracy on three benchmark datasets, that consist of face images acquired in various controlled and uncontrolled scenarios. Such images have been analyzed without any preliminary pre-processing steps besides the face alignment, which turned out to be very important to achieve high accuracy rates. In this work we only used a simple alignment method which can only rotate the face to an upright position. In future we aim to investigate further sophisticated algorithms that are more robust to other transformations, such as skewness, and that can deal with occluded faces.

It is worth noting that the COSFIRE-based method achieves the best performance on the three considered datasets. Consequently, it provides the most significant contribution to the multi-expert classification. The proposed domain specific method, namely the SURF-based, does not generalize as much as the trainable filter approach. While the COSFIRE-based method is much more effective, the SURF-based method is more efficient. COSFIRE filters are, however, highly parallelizable and in future we aim to develop a parallel implementation of this approach which could run on modern GPUs.

The results suggest that the trainable COSFIRE-based and the domain-specific SURF-based approach are complementary to each other as the accuracy increases when combined together. The stacked classification scheme outperforms the other widely used combination techniques, namely the majority voting and the weighted voting rules. This evidence is not surprising because the classifier is able to learn a non linear combination. However, we demonstrated that the combination has the potential to achieve further improvements. As a future direction, we will investigate different fusion techniques that are able to maximize the performance. In this paper we focused on the investigation of robust features for gender recognition from face images, while the investigation of the most appropriate machine learning tool was beyond the scope of this work. In future we will investigate more sophisticated machine learning methods such as boosting for COSFIRE filter selection and random forest approaches.

V. CONCLUSIONS

We propose a novel method for recognizing the gender by analyzing face images using a fusion of trainable and domain-specific features. The experimentation, performed over three benchmark datasets, confirms that our method is able to deal with faces acquired in completely different environments. The COSFIRE- and the SURF-based descriptors are able to represent different characteristics of the human face, which allow to effectively discriminate the gender of a person.

The proposed method is effective and relatively simple to use. The code is publicly available and the algorithm can be easily evaluated using other face images for experimental purposes. Moreover, the trainable part of our approach, namely the COSFIRE-based method, can be applied to other visual pattern recognition problems as it does not require domain knowledge.

REFERENCES

[1] M. Ng, V. M. Ciaramitaro, S. Anstis, G. M. Boynton, and I. Fine, "Selectivity for the configural cues that identify the gender, ethnicity, and identity of faces in human cortex," *Proc. Nat. Acad. Sci. USA*, vol. 103, no. 51, pp. 19552–19557, 2006.

[2] S. Fu, H. He, and Z.-G. Hou, "Learning race from face: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 12, pp. 2483–2509, Dec. 2014.

[3] Y. Fu, G. Guo, and T. S. Huang, "Age synthesis and estimation via faces: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 1955–1976, Nov. 2010.

[4] M. E. Hasselmo, E. T. Rolls, and G. C. Baylis, "The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey," *Behavioural Brain Res.*, vol. 32, no. 3, pp. 203–218, 1989.

[5] B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey," *Pattern Recognit.*, vol. 36, pp. 259–275, Jan. 2003.

[6] E. T. Rolls, M. J. Tovee, D. G. Purcell, A. L. Stewart, and P. Azzopardi, "The responses of neurons in the temporal cortex of primates, and face identification and detection," *Experim. Brain Res.*, vol. 101, no. 3, pp. 473–484, 1994.

[7] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, 2003.

[8] C. A. Nelson, "The development and neural bases of face recognition," *Infant Child Develop.*, vol. 10, nos. 1–2, pp. 3–18, 2001.

[9] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1701–1708.

[10] J. V. Haxby, E. A. Hoffman, and M. I. Gobbini, "The distributed human neural system for face perception," *Trends Cognit. Sci.*, vol. 4, no. 6, pp. 223–233, 2000.

[11] C. B. Ng, Y. H. Tay, and B.-M. Goi, *Recognizing Human Gender in Computer Vision: A Survey*. Berlin, Germany: Springer, 2012, pp. 335–346.

[12] U. Park and A. K. Jain, "Face matching and retrieval using soft biometrics," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 3, pp. 406–415, Sep. 2010.

[13] B.-C. Chen, C.-S. Chen, and W. H. Hsu, "Cross-age reference coding for age-invariant face recognition and retrieval," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 768–783.

[14] D. Riccio, G. Tortora, M. De Marsico, and H. Wechsler, "EGA—Ethnicity, gender and age, a pre-annotated face database," in *Proc. IEEE Workshop Biometric Meas. Syst. Secur. Med. Appl. (BIOMS)*, Sep. 2012, pp. 1–8.

[15] V. Bruce *et al.*, "Sex discrimination: How do we tell the difference between male and female faces?" *Perception*, vol. 22, no. 2, pp. 131–152, 1993.

[16] E. Murphy-Chutorian and M. M. Trivedi, "Head pose estimation in computer vision: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 4, pp. 607–626, Apr. 2009.
[17] C. BenAbdelkader and P. Griffin, "A local region-based approach to gender classification from face images," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)- Workshops*, Sep. 2005, p. 52. [Online]. Available: http://dx.doi.org/10.1109/CVPR.2005.388
[18] B. Moghaddam and M.-H. Yang, "Learning gender with support faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 707–711, May 2002.
[19] J. Yang, D. Zhang, A. F. Frangi, and J.-Y. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, Jan. 2004.
[20] S. Baluja and H. A. Rowley, "Boosting sex identification performance," *Int. J. Comput. Vis.*, vol. 71, no. 1, pp. 111–119, Jan. 2007.
[21] H.-C. Lian and B.-L. Lu, "Multi-view gender classification using local binary patterns and support vector machines," in *Proc. Int. Symp. Neural Netw.*, 2006, pp. 202–209.
[22] Z. Yang and H. Ai, "Demographic classification with local binary patterns," in *Proc. Int. Conf. Biometrics*, 2007, pp. 464–473.
[23] C. Shan, "Learning local binary patterns for gender classification on real-world face images," *Pattern Recognit. Lett.*, vol. 33, no. 4, pp. 431–437, Mar. 2012.
[24] V. Singh, V. Shokeen, and M. B. Singh, "Comparison of feature extraction algorithms for gender classification from face images," *Int. J. Eng. Res. Technol.*, vol. 2, no. 5, pp. 1313–1318, May 2013.
[25] G. Guo, C. R. Dyer, Y. Fu, and T. S. Huang, "Is gender recognition affected by age?" in *Proc. IEEE 12th Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Sep./Oct. 2009, pp. 2032–2039.
[26] J. Bekios-Calfa, J. M. Buenaposada, and L. Baumela, "Robust gender recognition by exploiting facial attributes dependencies," *Pattern Recognit. Lett.*, vol. 36, pp. 228–234, Jan. 2014.
[27] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.
[28] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of online learning and an application to boosting," in *Proc. Eur. Conf. Comput. Learn. Theory*, 1995, pp. 23–37.
[29] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
[30] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 12, pp. 2170–2179, Dec. 2014.
[31] R. Azarmehr, R. Laganière, W.-S. Lee, C. Xu, and D. Laroche, "Real-time embedded age and gender classification in unconstrained video," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2015, pp. 57–65.
[32] P. Dago-Casas, D. González-Jiménez, L. L. Yu, and J. L. Alba-Castro, "Single- and cross- database benchmarks for gender classification under unconstrained settings," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Nov. 2011, pp. 2152–2159.
[33] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, vol. 1, no. 1, pp. 886–893.
[34] G. Azzopardi, A. Greco, and M. Vento, "Gender recognition from face images using a fusion of SVM classifiers," in *Proc. 13th Int. Conf. Image Anal. Recognit. (ICIAR)*, Jul. 2016, pp. 533–538.
[35] L. A. Alexandre, "Gender recognition: A multiscale decision fusion approach," *Pattern Recognit. Lett.*, vol. 31, no. 11, pp. 1422–1427, 2010.
[36] 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," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 3, pp. 488–499, Mar. 2013.
[37] N. Wang, X. Gao, D. Tao, and X. Li. (2014). "Facial feature point detection: A comprehensive survey." [Online]. Available: https://arxiv.org/abs/1410.1037
[38] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 10, pp. 1042–1052, Oct. 1993.
[39] Y. S. El-Din, M. N. Moustafa, and H. Mahdi, "Landmarks-SIFT face representation for gender classification," in *Proc. Int. Conf. Image Anal. Process.*, 2013, pp. 329–338.
[40] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[41] G. Levi and T. Hassncer, "Age and gender classification using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2015, pp. 34–42.
[42] J. van de Wolfshaar, M. F. Karaaba, and M. A. Wiering, "Deep convolutional neural networks and support vector machines for gender recognition," in *Proc. IEEE Symp. Ser. Comput. Intell.*, Dec. 2015, pp. 188–195.
[43] R. Ranjan, V. M. Patel, and R. Chellappa. (2016). "HyperFace: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition." [Online]. Available: https://arxiv.org/abs/1603.01249
[44] S. Jia and N. Cristianini, "Learning to classify gender from four million images," *Pattern Recognit. Lett.*, vol. 58, pp. 35–41, Jun. 2015.
[45] Y. LeCun *et al.*, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1989.
[46] G. Azzopardi and N. Azzopardi, "Trainable COSFIRE filters for keypoint detection and pattern recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 490–503, Feb. 2013.
[47] G. Azzopardi, A. Greco, and M. Vento, "Gender recognition from face images with trainable COSFIRE filters," in *Proc. 13th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Aug. 2016, pp. 235–241.
[48] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2006, pp. 2169–2178. [Online]. Available: http://dx.doi.org/10.1109/CVPR.2006.68
[49] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 404–417.
[50] M. Uricár, V. Franc, and V. Hlavác, "Facial landmark tracking by tree-based deformable part model based detector," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Dec. 2015, pp. 10–17.
[51] G. Azzopardi and N. Petkov, "Ventral-stream-like shape representation: From pixel intensity values to trainable object-selective COSFIRE models," *Frontiers Comput. Neurosci.*, vol. 8, p. 80, Jul. 2014. [Online]. Available: http://www.biomedsearch.com/nih/Ventral-stream-like-shape-representation/25126068.html
[52] B. Gecer, G. Azzopardi, and N. Petkov, "Color-blob-based COSFIRE filters for object recognition," *Image Vis. Comput.*, vol. 57, pp. 165–174, Jan. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0262885616301895
[53] G. Azzopardi, N. Strisciuglio, M. Vento, and N. Petkov, "Trainable COSFIRE filters for vessel delineation with application to retinal images," *Med. Image Anal.*, vol. 19, no. 1, p. 46–57, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1361841514001364
[54] N. Strisciuglio, G. Azzopardi, M. Vento, and N. Petkov, "Supervised vessel delineation in retinal fundus images with the automatic selection of *B*-COSFIRE filters," *Mach. Vis. Appl.*, vol. 27, no. 8, pp. 1137–1149, 2016. [Online]. Available: http://dx.doi.org/10.1007/s00138-016-0781-7
[55] G. Azzopardi and N. Petkov, "A CORF computational model of a simple cell that relies on LGN input outperforms the Gabor function model," *Biol. Cybern.*, vol. 106, no. 3, pp. 177–189, Mar. 2012. [Online]. Available: http://dx.doi.org/10.1007/s00422-012-0486-6, doi: 10.1007/s00422-012-0486-6.
[56] G. Azzopardi, A. Rodríguez-Sánchez, J. Piater, and N. Petkov, "A push-pull CORF model of a simple cell with antiphase inhibition improves SNR and contour detection," *PLoS ONE*, vol. 9, no. 7, p. e98424, 2014. [Online]. Available: http://dx.doi.org/10.1371%2Fjournal.pone.0098424
[57] A. Pasupathy and C. E. Connor, "Population coding of shape in area V4," *Nature Neurosci.*, vol. 5, no. 12, pp. 1332–1338, Dec. 2002.
[58] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Univ. Massachusetts, Amherst, Amherst, MA, USA, Tech. Rep. 07-49, 2007.
[59] G. Azzopardi, A. Greco, A. Saggese, and M. Vento, "Fast gender recognition in videos using a novel descriptor based on the gradient magnitudes of facial landmarks," in *Proc. 14th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Aug./Sep. 2017, pp. 1–6.
[60] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.

GEORGE AZZOPARDI received the B.Sc. degree in computer science from the Goldsmiths, University of London, the M.Sc. degree in computer science from the Queen Mary University of London, and the Ph.D. degree *(cum laude)* in computer science from the University of Groningen, The Netherlands, in 2013. In 2015, he joined the University of Malta as a Lecturer. He is currently with the University of Groningen. He has published articles in several journals. His current research interests include pattern recognition, brain-inspired computer vision, and machine learning. He is a Co-Editor of the CAIP2015 conference proceedings and the First International Workshop on Computational Models of the Visual Cortex.

ANTONIO GRECO received the Ph.D. degree in computer science and computer engineering from the University of Salerno in 2018. His Ph.D. thesis was on the analysis of faces from images and videos in real time. His research activity is focused on computer vision and machine learning techniques for retail and video surveillance applications.

ALESSIA SAGGESE received the Ph.D. degree in computer engineering from the University of Salerno, Italy, and the University of Caen Normandy, France, in 2014. She is currently an Assistant Professor with the University of Salerno. Her research interests include basic methodologies and applications in computer vision and pattern recognition. She has been a member of the International Association for Pattern Recognition Technical Committee 15 on Graph-Based Representations in Pattern Recognition since 2012.

MARIO VENTO received the Ph.D. degree in computer engineering from the University of Naples Federico II in 1989. He is currently a Full Professor of computer science and artificial intelligence with the University of Salerno, Fisciano, Italy, where he is also a Coordinator of the Artificial Vision Laboratory and the Dean of the Department of Information and Electrical Engineering and Applied Mathematics. His research interests include real-time video analysis and interpretation for video surveillance applications, classification techniques, either statistical, syntactic, and structural, exact and inexact graph matching, multiexpert classification, and learning methodologies for structural descriptions.

Prof. Vento is a Fellow Scientist of the International Association Pattern Recognition (IAPR). He served as the Chairman for the IAPR Technical Committee 15 on Graph-Based Representation in Pattern Recognition from 2002 to 2006.