UNDERSTANDING CRITICAL FACTORS IN GENDER RECOGNITION

E. Grosso, A. Lagorio, L. Pulina, and M. Tistarelli

March 4, 2012

University of Sassari

Computer Vision Laboratory

Technical Report No. CVL -2012-001

UNIVERSITY
of
SASSARI

UNDERSTANDING CRITICAL FACTORS IN GENDER RECOGNITION

E. Grosso, A. Lagorio, L. Pulina, and M. Tistarelli

University of Sassari
Computer Vision Laboratory
Porto Conte Ricerche - Loc. Tramariglio
07041 Alghero (SS)
www.uniss.it

March 4, 2012

Technical Report No. CVL -2012-001

Abstract

Gender classification is a task of paramount importance for face recognition researchers, and it can be useful in a large set of applications. In this paper we investigate the gender classification problem from a non-conventional perspective. In particular, we are interested in understanding which factors critically affect the accuracy of available technologies, better explaining differences between face and gender recognition.

To this purpose we propose a novel challenging protocol over the dimensions of the Face Recognition Grand Challenge version 2.0 database (FRGC2.0) and we evaluate our protocol with respect to several classification algorithms, and processing different types of features, like Gabor and LBP. Our results show that gender classification is unexpectedly independent from factors like the race of the subject, face expressions, and variations of illumination conditions.

1 Introduction

Gender classification is a well-established problem in the field of automatic face recognition, and – as reported, e.g., in [15] – it is a task of paramount importance for face recognition researchers. A successful gender classification can boost a large number of advanced applications like search engines, surveillance systems and interfaces, and can help to steer gender-specific services. In the scientific literature, e.g., in [10], gender classification is often reported as one of the most challenging problems related to face recognition. In the last two decades, the computer vision scientific community has proposed several approaches. Starting from the seminal work of Golomb, Lawrence, and Sejnowski [6], key contributions are due to Cottrell and Metcalfe [5] – who proposed a multi-layer neural network approach – and Brunelli and Poggio [3], who detailed in the early 90s a system based on HyperBF networks. More recently, Moghaddam and his co-authors [12] proposed a methodology based on Support Vector Machine (SVM) with Radial Basis Function kernels. Mäkinen and Raisamo [11] surveyed several methodologies based on Multilayer Neural Network, SVM and Discrete AdaBoost. Lapedriza and colleagues [9] investigated the usage of boosting classifiers, like AdaBoost and JointBoosting. Finally, Shobeirinejad and Gao [17] presented a technique in which a histogram intersection is used as a measure of similarity for classification.

Most of the contributions listed above, in particular recent ones, agree on a generic processing scheme composed of a preliminary feature extraction step, followed by a classification algorithm. This scheme proved to be effective in face recognition and the extension to the gender recognition problem has been quite straightforward and equally effective. In fact, these two steps are semantically very different: feature extraction has to do with image signals which are considered relevant for the problem (for instance the skin color could be extremely relevant for race detection) whilst classification has to do

This research has received funding from Autonomous Region of Sardinia (Italy), L.R. 07/08/2007, n. 7, under grant agreement CP 2_442, "Adaptive Biometric Systems: Methodologies, Models, and Algorithms".

with the optimal partition of the feature space, possibly taking into account existing constraints.

In this paper we investigate the gender classification problem by an extended empirical analysis on the Face Recognition Grand Challenge version 2.0 (FRGC2.0) dataset [14]. To this extent, a first contribution concerns the proposal of a challenging experimental protocol for the gender recognition problem. Inspired by the above feature extraction-classification dicothomy and from experiments detailed in [14], we describe a procedure based on the exploitation of the dimensions offered by the data collected in FRGC2.0 – i.e., subject, face expression, race, environmental conditions. The proposed protocol is for general purpose, and it can be easily extended to other datasets and to different features and classifiers.

Our second, and more relevant contribution, concerns the application of the proposed protocol to a significant set of features and classifiers, proving that gender classification should be treated as a problem very different from face classification. In our experiments, 1-Nearest-Neighbour [1], Aggregation Pheromone density based pattern Classification (APC) [8], and Support Vector Machines [4], are used as classifiers. Feature extraction is based on Gabor features – see, e.g., [16] –, Local Binary Patterns (LBP) [13] and raw pixel values with histogram equalization.

Noticeably external factors critically affecting the accuracy of face recognition like race, expressions and environmental conditions, are in the case of gender almost irrelevant. We also report interesting insights related to the feature extraction step. Particularly, Gabor features turn to be an effective choice in uncontrolled environment, while, in the case of controlled environment raw pixel values perform equally well.

The paper is structured as follows. In Section 2 we introduce the notation used and we give a brief description of the FRGC2.0 dataset. We also briefly introduce both the classification algorithms and the feature extraction methods herewith employed. In Section 3 we describe our experimental setup, detailing the experimental protocol. Section 4 shows the results of the experimental protocol applied to a selected set of features and classifiers. Finally, in Section 5 conclusions are drawn.

2 Data, Algorithms, and Features

2.1 The FRGC2.0 dataset

Our empirical evaluation is based on 2D images comprised in the Face Recognition Grand Challenge dataset, version 2.0. The dataset is composed of more than 50,000 images, see Figure 1 for some samples. As reported in [14], the dataset is composed of both a training and a validation set, denoted in the following as Γ and Σ, respectively.

We model both Γ and Σ as sets composed of the total amount of the subjects (males and females) involved in the images collections. Therefore, we can look at the FRGC2.0 training set as a set $\Gamma = \{\underline{\gamma}_1, \ldots, \underline{\gamma}_n\}$, with $n = 291$ (the total amount of involved subjects), in which each $\underline{\gamma}_j$ denotes the pool of images related to the subject j. Each image $\gamma_{jk} \subset \underline{\gamma}_j$ is characterized by a tuple of three elements $< C, E, R >$, where:

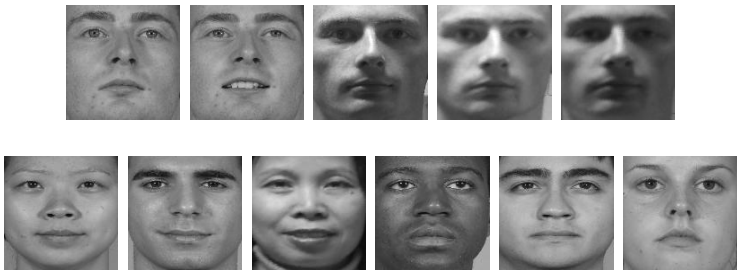

Figure 1: Images samples from the FRGC2.0 database. Neutral, smiling and different light conditions images from the same person are depicted in the first row. In the second row, they are depicted images related to the different races, namely Asian, Asian Middle Eastern, Asian Southern, Black or African American, Hispanic, and White.

- $C = \{c2l, c3l, u\}$ denotes the types of control, i.e., controlled images with two or three studio lights ($c2l$ and $c3l$, respectively), and uncontrolled images (denoted as u).
- $E = \{BlankStare, Happiness\}$ denotes the facial expressions in the dataset images, i.e., neutral and smiling, respectively.
- $R = \{A, AME, AS, BAA, H, U, W\}$ denotes the race of the subject, where, following the categorization of FRGC2.0, A stands for "Asian", AME for "Asian Middle Eastern", AS for "Asian-Southern", BAA for "Black or African American", H for "Hispanic", while U stands for "Unknown", and, finally, W denotes "White" race.

The same is done for the test set $\Sigma = \{\underline{\sigma}_1, \ldots, \underline{\sigma}_n\}$, with $n = 466$.

2.2 Algorithms

In this work, we model the gender classification problem as a binary pattern classification one. In binary classification problems, a set of patterns is given, i.e., input vectors $X = \{\underline{x}_1, \ldots \underline{x}_k\}$ with $\underline{x}_i \in \mathbb{R}^n$, and a corresponding set of labels, i.e., output values $Y \in \{0, 1\}$ – in our case, male and female. We think the labels as generated by some unknown function $f : \mathbb{R}^n \to \{0, 1\}$ applied to the patterns, i.e., $f(\underline{x}_i) = y_i$ for $i \in \{1, \ldots, k\}$ and $y_i \in \{0, 1\}$. The task of a binary classifier c is to extrapolate f given X and Y, i.e., construct c from X and Y so that when given some $\underline{x}^\star \in X$, $c(\underline{x}^\star)$ will equal $f(\underline{x}^\star)$; such task can be achieved *training an inductive model* of c.

In the following, we briefly review the classifiers that we use in our empirical analysis.

- **1-nearest-neighbor** (1-NN): It is a classifier yielding the label of the training instance which is closer to the given test instance, whereby closeness is evaluated using some proximity measure, e.g. Euclidean distance; we use the method described in [1] to store the training instances for fast look-up.

- **Aggregation Pheromone density based pattern Classification** (APC): It is a pattern classification algorithm modeled on the ants colony behaviour and distributed adaptive organization in nature. Each data pattern is considered as an ant, and the training patterns (ants) form several groups or colonies depending on the number of classes present in the data set. A new test pattern (ant) will move along the direction where average aggregation pheromone density (at the location of the new ant) formed due to each colony of ants is higher and hence eventually it will join that colony. The reader is referred to [8] for further details.

- **Support Vector Machine** (SVM): It is a supervised learning algorithm used for both classification and regression tasks. Roughly speaking, the basic training principle of SVMs is finding an optimal linear hyperplane such that the expected classification error for (unseen) test patterns is minimized. The reader is referred to [4] for further details.

2.3 Features

We compute different input vectors X of both Γ and Σ using the raw values of the pixels and extracting Gabor and LBP features.

As a basic feature we use the raw pixel values (PV in the following) of the image converting the image matrix in a mono-dimensional vector. To compensate for illumination changes, histogram equalization is first applied; in order to reduce the dimension, images are therefore scaled to 64x64 pixels.

Concerning the Gabor features, first image dimension is scaled with respect to the position of the eyes. Next, a bank of Gabor filters with 5 scales and 8 orientations is applied to the preprocessed image on the 64 nodes of a uniform 8×8 grid superimposed to the image, obtaining – for each preprocessed image – a feature vector composed of 2560 elements. In order to do that, we used a piece of software built on top of the Feature Extraction Library (FELib) [18].

Concerning the Local Binary Pattern operator (LBP), it was originally designed for texture description but recently it has been successfully applied to face description and gender classification. The original LBP operator assigns a label to every pixel of an image by thresholding the 3x3 neighborhood of each pixel with the center pixel value and converting the result in a binary number. In order to describe the textures at different scales, the LBP operator has been extended to use neighborhoods at different distances from the considered pixel. The operator was denoted as $LBP_{P,R}$, where P is the number of sampling points on a circle of radius R. An interesting extension of LBP takes in account the bitwise transitions of the obtained binary pattern [13]. According with the results obtained by Ahonen et al. in [2] the $LBP^{u2}_{8,2}$ operator was selected in order to obtain a good trade-off between description performance and feature vector length. The considered image is divided in a 7x7 windows and the $LBP^{u2}_{8,2}$ is applied to each window. The histograms are computed independently within each window and then they are concatenated. The resulting histogram has size m x n where m is the number of windows (49 in this case) and n is the length of a single $LBP^{u2}_{8,2}$ histogram (10 in this case) so the total histogram size is 490.

	#	**F**	**M**
Γ_t	1027	42.65%	57.35%
Σ_v	1292	43.19%	56.61%
Σ_a	1262	45.01%	54.99%
Σ_b	1958	50.56%	49.45%
Σ_c	1958	50.56%	49.45%

Table 1: Synopsis of training, validation and test sets. The table is structured as follows. The first column shows the name of set (in the case of test sets, we report groups only), and it is followed by three columns. The first column ("#") reports the total amount of images in the set, while the remaining two ("F" and "M") report the percentage of the images in the set, labeled as female and male, respectively.

3 Experimental Setup

Aim of this section is to develop an experimental protocol useful to analyze the sensitiveness of classification algorithms with respect to several image characteristics – i.e., expression and race of the subject, illumination variations – in the gender classification problem.

In order to do that, the FRGC2.0 dataset was used in our experiments. Considering 2D images, it is composed of about 40,000 images related to 466 subjects of different races. Images were taken at different illumination conditions, and subjects had different expressions. Considering the total amount of images, 43.93% depicts female subjects, while the remaining depicts male subjects.

In order to accomplish our goals, our target is to compute classification models trained on data having specific values of C, E, and R. To do that, we train the algorithms described in Section 2.2 selecting controlled images – two studio lights – related to "White" subjects (the most recurrent in the FRGC2.0) having a neutral expression. In other words, we train the classifiers on a set Γ_t in which, for each subject j, $|\underline{\gamma_j}|$ is equal to the total amount of images $\gamma_{jk} \subset \underline{\gamma_j}$ such that $(C = c2l) \wedge (E = BlankStare) \wedge (R = W)$.

In machine learning literature, it is well-established that classifiers performance could vary with respect to different parameter tunings. In order to have a fair comparison among classifiers, we proceed as follows. First, we split the FRGC2.0 validation set Σ in two parts. The first one is composed of the images related to the 291 subjects also occurring in Γ. This partition – Σ_v in the following – is used for parameter tuning purpose. We compute Σ_v with the same criteria of Γ:

Σ_v : for each subject j, $\sigma_{jk} \subset \underline{\sigma_j}$ such that $(C = c2l) \wedge (E = BlankStare) \wedge (R = W)$

Concerning the test set, we consider the partition of Σ composed of the images related to the 175 subjects *not occurring* in Γ. We compute 9 different test sets that can be organized in 3 groups:

- Σ_a: for each subject j, $\forall \sigma_{jk} \subset \underline{\sigma_j} \in \Sigma$ such that $\underline{\gamma_j} \notin \Gamma$ and $(E = BlankStare) \wedge (R = W)$. The rational of such group is to have the same facial expression and

Classifier	PV		Gabor		LBP	
	Acc.	**Par.**	**Acc.**	**Par.**	**Acc.**	**Par.**
`1-NN`	97.54%	–	98.99%	–	96.59%	–
`APC`	95.90%	$\delta = 2$	99.07%	$\delta = 0.1$	94.27%	$\delta = 50$
`SVM`	98.80%	$c = 2$, $g = 0$	99.46%	$c = 2$, $g = 2$	97.29%	$c = 16$, $g =$ 2e-06

Table 2: Parameter optimization for the considered algorithms. The table is organized as follows. The first column shows the name of the algorithms, and it is followed by three groups of columns, reporting the results of the optimization considering PV, Gabor and LBP features. Each group of columns is composed of two subcolumns, reporting the accuracy (column "Acc.") of the computed model, and the related parameters (column "Par.").

race of the training set, in order to have a baseline for our comparisons. This group is composed of three test sets, i.e., $\Sigma_{a,c2l}$, $\Sigma_{a,c3l}$, $\Sigma_{a,u}$, representing test sets in which σ_{jk} has a value of C equal to $c2l$, $c3l$, and u, respectively.

- Σ_b: for each subject j, $\forall \sigma_{jk} \subset \underline{\sigma_j} \in \Sigma$ such that $\underline{\gamma_j} \notin \Gamma$ and $(E = BlankStare)$. In this group are involved images that are not constrained to a particular value of R. The rational of such group is to compare the accuracy of the classifiers with respect to the race of the involved subjects. Also this group if composed of three test sets, i.e., $\Sigma_{b,c2l}$, $\Sigma_{b,c3l}$, $\Sigma_{b,u}$.

- Σ_c: for each subject j, $\forall \sigma_{jk} \subset \underline{\sigma_j} \in \Sigma$ such that $\underline{\gamma_j} \notin \Gamma$ and $(E = Happiness)$. The rational of such group is to compare the accuracy of the classifiers with respect to a different face expression. Also this group is not constrained by a particular value of R. It is composed of three test sets: $\Sigma_{c,c2l}$, $\Sigma_{c,c3l}$, $\Sigma_{c,u}$.

Cardinalities and label distributions of the sets are reported in Table 1.

4 Empirical Results

Aim of our first experiment is to train gender classification model with 1-NN, APC, and SVM. In order to do that, we perform a parameter grid search involving both APC and SVM, and we proceed as follows:

- Concerning APC, we explore the parameter δ related to the pheromone intensity of described in [8].

- Concerning SVM, we consider a C-SVC with a Radial Basis Function (RBF) kernel. In particular, we explore the parameter space related to both cost c and the parameter g of the kernel RBF.

We test the obtained models on Σ_v, and the results of these experiments related to the best configuration found – in terms of accuracy – are depicted in Table 2. For all the experiments in the following, when we refer to APC and SVM, we will intend that such algorithms are tuned by using the parameters shown in Table 2.

Test set	**1-NN**	**APC**	**SVM**
$\Sigma_{a,c2l}$	83.60%	84.86%	95.80%
$\Sigma_{a,c3l}$	81.70%	84.15%	95.01%
$\Sigma_{a,u}$	64.10%	67.99%	76.39%
$\Sigma_{b,c2l}$	81.31%	82.38%	91.37%
$\Sigma_{b,c3l}$	81.10%	82.69%	91.78%
$\Sigma_{b,u}$	63.89%	66.70%	75.49%
$\Sigma_{c,c2l}$	80.03%	80.54%	88.07%
$\Sigma_{c,c3l}$	81.00%	81.15%	89.07%
$\Sigma_{c,u}$	65.83%	68.08%	78.19%

Table 3: Evaluation results using PV features. The table is composed of four columns. The first one ("Test set") denotes the test set on which classifiers has been evaluated. The three following columns report the accuracy performance (in percentage) related to 1-NN, APC, and SVM (columns "1-NN", "APC", and "SVM", respectively).

In our next experiment, we evaluate the performance of 1-NN, APC, and SVM trained on Γ_t using PV feature, and tested on the test sets described in Section 3. Table 3 shows the results of such experiment.

Looking at the table, we can see that SVM outperforms the other classifiers, reporting an accuracy greater than 90% on all test sets having $C = c2l$ and $C = c3l$ in both groups Σ_a and Σ_b. SVM is also the most performing classifier for $C = u$. Looking at the table in detail, concerning $\Sigma_{a,c2l}$, we can see that SVM accuracy is more than 10% greater than both 1-NN and APC accuracies. We can consider the results related to $\Sigma_{a,c2l}$ as a reference for the experiments in the following, because it is composed of images having the same value of C, E, and R used to compute Γ_t. Considering now the results related to $\Sigma_{a,c3l}$, we report that the accuracy of all classifiers is very close to the one reported for $\Sigma_{a,c2l}$. As a consequence, we can conjecture that all considered classifiers are robust with respect to controlled illumination variations. Looking now at the results in uncontrolled environment – $\Sigma_{a,u}$ –, we report a leak of classifiers performance. SVM is yet the best classifier, but its accuracy is about 20% less than the one reported for both $\Sigma_{a,c2l}$ and $\Sigma_{a,c3l}$.

Considering now the results related to Σ_b group, we can see the same pattern described for Σ_a: SVM outperforms the other classifiers, and there is a lack of performance for $C = u$. Also the accuracy results for all classifiers are very close to the ones reported for Σ_a. However, as discussed in Section 3, images comprised in such test sets are not constrained from a particular value of R. As a consequence, we can conjecture that classifiers performance are not affected by this element. We can reach analogous conclusions looking at the results related to the group Σ_c, in which also E is not constrained.

Considering now Gabor features, we report the results of our experiment in Table 4. Looking at the table, we can see that also in this case SVM outperforms the other classifiers, reaching an accuracy greater than 90% on all test sets having $C = c2l$ and

Test set	**1-NN**	**APC**	**SVM**
$\Sigma_{a,c2l}$	82.88%	86.53%	91.92%
$\Sigma_{a,c3l}$	83.44%	87.48%	92.95%
$\Sigma_{a,u}$	64.66%	68.70%	78.68%
$\Sigma_{b,c2l}$	83.20%	86.57%	90.60%
$\Sigma_{b,c3l}$	83.71%	87.13%	91.62%
$\Sigma_{b,u}$	65.37%	68.18%	77.99%
$\Sigma_{c,c2l}$	74.44%	80.75%	90.14%
$\Sigma_{c,c3l}$	74.11%	81.61%	90.81%
$\Sigma_{c,u}$	63.48%	64.86%	71.12%

Table 4: Evaluation results using Gabor features. The table is organized as Table 4.

Test set	**1-NN**	**APC**	**SVM**
$\Sigma_{a,c2l}$	79.63%	81.93%	90.89%
$\Sigma_{a,c3l}$	81.30%	81.06%	91.36%
$\Sigma_{a,u}$	53.80%	51.34%	51.98%
$\Sigma_{b,c2l}$	82.79%	85.03%	91.73%
$\Sigma_{b,c3l}$	82.53%	83.76%	92.13%
$\Sigma_{b,u}$	49.08%	52.35%	54.44%
$\Sigma_{c,c2l}$	80.08%	81.97%	90.40%
$\Sigma_{c,c3l}$	77.43%	78.65%	89.32%
$\Sigma_{c,u}$	56.84%	54.70%	55.41%

Table 5: Evaluation results using LBP features. The table is organized as Table 3.

$C = c3l$. SVM is also the most performing classifier for $C = u$. Looking at the table in detail, concerning $\Sigma_{a,c2l}$, we can see that SVM accuracy is more than 5% greater than APC accuracy, and about 9% greater than 1-NN accuracy. Considering now the results related to $\Sigma_{a,c3l}$, we report that the accuracy of all classifiers is very close to the one reported for $\Sigma_{a,c2l}$. Looking now at the results in $\Sigma_{a,u}$, we report a leak of classifiers performance also in this case. SVM is yet the better classifier, but its accuracy is about 14% smaller than the one reported for both $\Sigma_{a,c2l}$ and $\Sigma_{a,c2l}$. We also report that performance of both 1-NN and APC decrease of about 20%. Considering now the results related to Σ_b group, we can see the same pattern described for Σ_a: SVM outperforms the other classifiers, and there is a lack of performance for $C = u$. Also the accuracy results for all classifiers are very close to the ones reported for Σ_a.

Our last experiment is analogous to the previous one, with the noticeable difference that we used LBP features instead of Gabor features. Table 5 shows the results of such experiment.

Looking at the table, we can see that SVM is the best classifier – in terms of accuracy – also in this case. Looking in detail the results related to Σ_a, Σ_b, and Σ_c, we can also report the same pattern found using both PV and Gabor features. This fact

confirm our conjecture that gender classification is independent from both values of E and R, and also from controlled illumination variations. However, differently from previous tables, LBP features are almost useless in the uncontrolled cases $\Sigma_{a,u}$, $\Sigma_{b,u}$, and $\Sigma_{c,u}$.

5 Conclusions

The paper investigates the gender classification problem trying to understand which factors critically affect the accuracy of available technologies. The proposed protocol exploits the dimensions of FRGC2.0 database analyzing the sensitivity of a two-steps feature extraction-classification approach with respect to three different classifiers and three orthogonal types of features.

The results of our empirical analysis can be summarized as follows:

- Gender classification is independent from the race of the subjects. Our results show that training an inductive model on a set of images composed of subject of only one race, the accuracy of the classifiers is about the same if in the test set we involve subjects of different races.

- Gender classification accuracy does not change in a noticeable way for controlled changes of illumination. We showed that, training classifiers on FRGC2.0 controlled images with two studio lights, and testing them on controlled images with three studio lights, the accuracy result is almost the same of the test performed on controlled images with two studio lights.

- Different face expressions do not influence in a noticeable way the gender classification accuracy applying SVM to Gabor and LBP features. This result is very clear applying SVM to both Gabor and LBP features. A limited degradation is reported for PV features, starting from a 96% results obtained for $\Sigma_{a,2cl}$. This fact is probably related to iconic information content of PV features, while both Gabor and LBP features are mainly related to the frequency image content.

As a final comment, our analysis confirms that race, expressions, and illumination conditions are for gender recognition almost irrelevant. Obviously, relaxing constraints concerning race and expression, the global accuracy of the recognition decreases. But this behaviour is independent both from classifiers and features.

Concerning classifiers, we report that SVMs always outperform other classifiers. Concerning features, our investigation confirms that Gabor features are an effective choice in the case of uncontrolled environment. Moreover, our analysis show that also trivial features as PV can be usefully adopted for gender classification.

As future work, we are planning to investigate additional dimensions of FRGC2.0, e.g. age, and to extend our analysis to other datasets, including masking and face occlusions. In addition, we plan to extend our analysis to additional feature representation and state-of-the-art gender classification methods, and carefully consider the statistical significance of the classifier results.

References

[1] D.W. Aha, D. Kibler, and M.K. Albert. Instance-based learning algorithms. *Machine learning*, 6(1):37–66, 1991.

[2] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12):2037–2041, 2006.

[3] R. Brunelli and T. Poggio. Hyperbf networks for gender classification. In *Proceedings of the DARPA Image Understanding Workshop*, volume 314. San Diego: CA, 1992.

[4] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[5] G.W. Cottrell and J. Metcalfe. Empath: Face, emotion, and gender recognition using holons. In *Proceedings of the 1990 conference on Advances in neural information processing systems 3*, pages 564–571. Morgan Kaufmann Publishers Inc., 1990.

[6] B.A. Golomb, D.T. Lawrence, and T.J. Sejnowski. Sexnet: A neural network identifies sex from human faces. *Advances in neural information processing systems*, 3:572–577, 1991.

[7] S. Gutta and H. Wechsler. Gender and ethnic classification of human faces using hybrid classifiers. In *Neural Networks, 1999. IJCNN'99. International Joint Conference on*, volume 6, pages 4084–4089. IEEE, 1999.

[8] A. Halder, A. Ghosh, and S. Ghosh. Aggregation pheromone density based pattern classification. *Fundamenta Informaticae*, 92(4):345–362, 2009.

[9] A. Lapedriza, M.J. Marin-Jimenez, and J. Vitria. Gender recognition in non controlled environments. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 3, pages 834–837. IEEE, 2006.

[10] H.C. Lian and B.L. Lu. Multi-view gender classification using local binary patterns and support vector machines. *Advances in Neural Networks-ISNN 2006*, pages 202–209, 2006.

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

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

[13] T. Ojala, M. Pietikainen, and T. Maenpaa. 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.

[14] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *Computer vision and pattern recognition, 2005. CVPR 2005. IEEE computer society conference on*, volume 1, pages 947–954. IEEE, 2005.

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

[16] L. Shen and L. Bai. A review on gabor wavelets for face recognition. *Pattern Analysis & Applications*, 9(2):273–292, 2006.

[17] A. Shobeirinejad and Y. Gao. Gender classification using interlaced derivative patterns. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 1509–1512. IEEE, 2010.

[18] J. Zhu, S.C.H. Hoi, M.R. Lyu, and S. Yan. Near-duplicate keyframe retrieval by nonrigid image matching. In *Proceeding of the 16th ACM international conference on Multimedia*, pages 41–50. ACM, 2008. `http://www.vision.ee.ethz.ch/ zhuji`.