MEG: Multi-Expert Gender classification from face images in a demographics-balanced dataset

Modesto Castrillón-Santana[1], Maria De Marsico[2], Michele Nappi[3], and Daniel Riccio[4]

[1]Universidad de Las Palmas de Gran Canaria, Spain. Email: mcastrillon@siani.es
[2]Sapienza University of Rome, Italy. Email: demarsico@di.uniroma1.it
[3]University of Salerno, Fisciano (SA), Italy. Email: mnappi@unisa.it
[4]University of Naples Federico II, Italy, Email: daniel.riccio@unina.it

Abstract. In this paper we focus on gender classification from face images, which is still a challenging task in unrestricted scenarios. This task can be useful in a number of ways, e.g., as a preliminary step in biometric identity recognition supported by demographic information. We compare a feature based approach with two score based ones. In the former, we stack a number of feature vectors obtained by different operators, and train a SVM based on them. In the latter, we separately compute the individual scores from the same operators, then either we feed them to a SVM, or exploit likelihood ratio based on a pairwise comparison of their answers. Experiments use EGA database, which presents a good balance with respect to demographic features of stored face images. As expected, feature level fusion achieves an often better classification performance but it is also quite computationally expensive. Our contribution has a threefold value: 1) the proposed score level fusion approaches, though less demanding, achieve results which are rather similar or slightly better than feature level fusion, especially when a particular set of experts are fused; since experts are trained individually, it is not required to evaluate a complex multi-feature distribution and the training process is more efficient; 2) the number of uncertain cases significantly decreases; 3) the operators used are not computationally expensive in themselves.

1 Introduction

Demographic data is widely used in marketing for customer profiling, to implement both marketing strategies and personalized recommendation systems. The most commonly studied demographics for that purpose are ethnicity, age and gender. However, the same information can be useful in forensic and security-related applications too. The work in [1] discusses how the preliminary determination of those demographics, though being considered soft biometrics, can significantly improve identity recognition performance by strong traits, e.g., face. In this paper, we focus on the gender classification (GC) problem from face images.

Within the context of the Computer Vision literature related to automatic GC, the face appearance attracts the main efforts, as reflected by most recent works [2]. However, additional elements of interest have been employed to tackle the problem, such as the face local context, the whole human body, the hair, or the clothes [3]. It is to notice that the influence of demographics on human appearance cannot be sharply identified. For instance, age and ethnicity can also affect the difference between male and female appearance. Therefore, despite the specific demographics under investigation, the benchmark dataset should be fairly balanced with respect to each factor [4]. This especially holds if a training phase is required. This motivated our choice of EGA (Ethnicity, Gender and Age) dataset [5] where this aspect is especially cared of. Its images are annotated with corresponding demographics information. In the present study, such annotations represent the ground-truth for assessing demographic classification performance.

In the present work, we propose to address the problem of GC from face images by a multi-expert approach. We chose a number of local operators able to capture different aspects of images, and able to possibly provide gender-discriminative information. Then, we investigated the most appropriate way to combine them in order to obtain gender-discriminative information. We tested expert combination using either feature-level or score level fusion. The former is expected to provide more accurate classification, yet at the expense of extra computational resources and of a most demanding training process. In score-level fusion approach, experts are trained individually, and this makes the training process much more feasible and parallelizable, since it is not required to evaluate a complex multi-feature distribution. If this allows to achieve only slightly worse results, we can accept this as a good compromise between accuracy and cost.

It is to underline that we did not plan to demonstrate the performance of either new operators or new fusion strategies. Our contribution has a threefold value. 1) The achieved performance demonstrates that, when suitably applied, score fusion can provide results that are comparable to those obtained by feature fusion. However, in the former case we have a vector with size equal to the number of experts, while in the latter the size is the sum of sizes provided by the single experts, unless a further expensive step of feature selection/learning is performed. 2) Even when the accuracy is quite similar, the number of uncertain cases significantly decreases, i.e., we have less situations that possibly require manual intervention. Therefore, the obtained system is overall more efficient. 3) These outcomes hold notwithstanding the use of quite light/popular operators, and we consider this a further added value.

2 The Pool of Experts

In this work, we consider the following set of local descriptors, that have already been applied in different scenarios of facial analysis: 1) Local Binary Patterns (LBP) [6]; 2) Local Gradient Patterns (LGP) [7]; 3) Local Ternary Patterns (LTP) [8]; 4) Local Derivative Patterns (LDP) [9]; 5) Weber Local Descriptor (WLD) [10]; and 6) Local Phase Quantization (LPQ) [11].

LBP. Since the work by Ahonen et al. [12], LBP is used as descriptor for facial images. In the original definition each pixel in turn is the center of a 3×3 window and is encoded by comparing its value with each of the neighboring ones. A pixel in the neighborhood is assigned a 1 if its value is greater than the value of the central pixel (local threshold value for a kind of window binarization), and a 0 otherwise. The code for the central pixel is produced by concatenating the 1s and 0s of the neighborhood into a binary number. A histogram of the resulting codes is used to represent the texture. LBP are therefore computed easily, and have proven their capacity of discrimination in different real world texture classification problems, while exhibiting a notorious robustness to monotonic gray-scale changes. Their definition has been extended to arbitrary circular neighborhoods of radius R with P neighbors. To achieve higher robustness, the image is divided into a predefined grid. The final feature vector is obtained by concatenating the histograms of the single cells of the grid, following a Bag of Words scheme.

LGP. LBP technique has inspired a number of variations based on similar measurements of characteristics in a pixel neighborhood of possibly varying size. LGP operator uses the gradient values of the eight neighbors of a given central pixel. These are computed as the absolute value of intensity difference between each pixel and its neighboring one. Gradient values substitute pixel values, and their average is assigned to the given central pixel and is used as the threshold value for LGP encoding, which is performed as for LBP. Also the LGP operator is extended in a way similar to LBP to use different sizes of neighborhoods.

LTP. LTP extend LBP to 3-valued codes. Gray levels in a intensity range of width $\pm t$ around g_c are quantized to zero, those above this are quantized to $+1$ and those below it to -1. The binary LBP code is replaced by a ternary LTP code. Since t is a user-specified threshold, LTP codes can be made resistant to noise, but no longer invariant to gray level transformations as LBP codes are. When using LTP for matching one could compute 3^n valued codes. However, an alternative coding scheme suggested by the authors splits each ternary pattern into its positive and negative parts (Upper Pattern and Lower Pattern respectively). These are treated as two separate channels of LBP descriptors, for which separate histograms and similarity metrics are computed, and finally combined.

LDP. LBP can be considered to represent a kind of first-order circular derivative pattern of images, i.e., a micropattern generated by the concatenation of the binary gradient directions. LDP increases the detail of coded information since it represents a high-order local pattern descriptor, by encoding directional pattern features based on local derivative variations. The n^{th} order LDP encode the $(n-1)^{th}$ order local derivative direction variations. While basic LBP encode the relationship between the central point and its neighbors in a 3×3 window, the LDP templates are more complex and extract high-order local information by encoding various distinctive spatial relationships contained in a given local region. Given an image I, the first-order derivatives are denoted as I'_α where α=0°, 45°, 90° and 135°. Given g_c a point in I, and g_p, $p = 0, \ldots, P-1$ its neighbors, the four first-order derivatives at g_c can be written as:

$$LDP^1(g_c) = \left\{ \begin{array}{l} I'_{0^\circ}(g_c) = I(g_c) - I(g_3), I'_{45^\circ}(g_c) = I(g_c) - I(g_2), \\ I'_{90^\circ}(g_c) = I(g_c) - I(g_1), I'_{135^\circ}(g_c) = I(g_c) - I(g_0) \end{array} \right\} \qquad (1)$$

The second-order directional LDP, LDP^2_α, in direction α is defined as:

$$LDP^2_\alpha(g_c) = \left\{ \begin{array}{l} f(I'_\alpha(g_c), I'_\alpha(g_0)), f(I'_\alpha(g_c), I'_\alpha(g_1)), \ldots \\ \ldots, f(I'_\alpha(g_c), I'_\alpha(g_7) \end{array} \right\} \tag{2}$$

where $f(.,.)$ is a binary function which determines the type of local pattern transition, and encodes the co-occurrence of two derivative directions at different neighboring pixels as:

$$f(I'_\alpha(g_c), I'_\alpha(g_p)) = \begin{cases} 0, if I'_\alpha(g_c) \cdot I'_\alpha(g_p) > 0 \\ 1, if I'_\alpha(g_c) \cdot I'_\alpha(g_p) \leq 0 \end{cases} \tag{3}$$

Finally, the second order Local Derivative Pattern $LDP^2(I)$ is the concatenation of the codes according to the different directions. Higher order derivatives are computed in a similar way. For more details, see [9].
WLD. Differences of pixel intensity within a local neighborhood are also encoded by WLD. It is inspired by Weber's Law, stating that human perception of a pattern depends both on the change of a stimulus and also on its original intensity. Therefore, WLD consists of two components: differential excitation and orientation. The former one is a function of the ratio between the relative intensity differences of a current pixel against its neighbors, and the intensity of the current pixel itself. The orientation component is the gradient orientation of the current pixel, and computed as in [13]. For a given image, both components make up a concatenated WLD histogram. For further details see [10].
LPQ. The codes produced by the LPQ operator are insensitive to centrally symmetric blur (e.g., due to motion, or out of focus). Even the LPQ operator is computed locally at every pixel location and the resulting codes is summarized in a histogram. The method is based on the blur invariance property of the Fourier phase spectrum. It uses the local phase information extracted using the short-term Fourier transform (STFT) computed over a $M \times M$ neighborhood at each pixel position x. For details see [11].

Feature vectors vs feature images

Given the image I, the application of the local operator O (LBP, LDP, LGP, etc.) will produce either a feature vector V or a feature image F. In the first case, the operator returns either a single histogram from the frequency of the extracted codes, or the concatenation of histograms corresponding to the cells of a square grid. In our system, all operators but LDP divide the 64×100 original image in a 5×5 grid, therefore producing a vector of size 6400. When a feature image F is returned instead, each pixel in the original image is substituted by the corresponding code assigned to it by O. All operators considered assign a 8 bit code to each pixel in I, which can be interpreted as a gray level. Figure 1 shows an example of the feature images corresponding to an image, with LTP Upper and Lower patterns fused together.

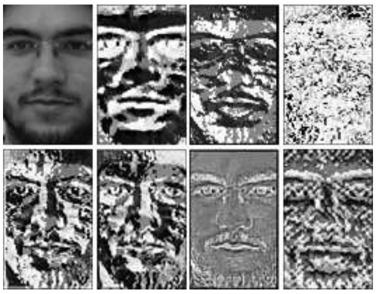

Fig. 1. From top left, original and coded images (LBP, LGP, LTP, LDP, WLD, LPQ)

2.1 Score generation by Likelihood Ratio (LR)

The Likelihood Ratio (LR) was introduced in biometrics to evaluate the membership of a submitted probe either to the class of genuine users (those enrolled in the system), or to that of impostors (unregistered users). In accordance with the Neyman-Pearson lemma, the authors of [14] experimentally assess that the LR test represents the optimal test to assign the score vector X to either genuine or impostor users, if, when False Acceptance Rate (FAR) is fixed at Ψ, we find a constant η which maximizes Genuine Acceptance Rate (GAR). However, optimality is constrained by the precision of genuine and impostor score distributions estimates. In this work, given a face image in input, the LR is used to produce a gender-discriminative (*male/female*) score; comparing that score with a threshold properly fixed in advance, the system decides if the input face belongs either to the class *male* or *female*. A training phase is needed to estimate $f_{male}(x)$ and $f_{female}(x)$ distributions, and test performances depends on the quality of such training.

All the experts in the system that exploit LR for score generation execute the same operation pipeline, the only difference being the local operator O, among those discussed above, that each of them uses to extract relevant features from I and transform it into a feature image F. For each pixel (x, y) in the image F the training phase learns two probability distributions Pr_{male} and Pr_{female} using a training set of images of faces, whose gender male / female is obviously known. Of course, training and testing sets have no intersection. During matching, each pixel in the feature image F contributes to the calculation of the final score by voting by its own partial score $s(x, y)$. This is computed according to the learned distributions using the standard formula defined for LR:

$$s(x, y) = 2 \cdot \frac{log(f_{female}(F(x, y)))}{log(f_{male}(F(x, y)))} \tag{4}$$

The partial score produced by Eq. 4 generally has a negative value if the pixel votes for the class male and positive otherwise. The greater is the confidence of the pixel when voting for a class, the higher is the absolute value of the assigned partial score. There is an area of uncertainty in the interval around the 0, for which the partial score can be considered noise, rather than a really useful contribution for the calculation of the final score s. For this reason, we fix a threshold th_p for the partial score (here it has been experimentally set to $th = 1.3$) . The final score is calculated as:

$$s = \frac{1}{S} \sum \delta(s(x,y)) \cdot s(x,y) \tag{5}$$

where δ is the Dirac function returning 1 only if $|s(x,y)| \geq th_p$ and $S = \sum_{x,y} \delta(s(x,y))$. Similar considerations hold for the global score, if it is too close to the border between classes. Here, we deal with ambiguity of fused scores too. The first column of Table 1 summarizes the results for the single operators.

2.2 Score generation by Support Vector Machines (SVM)

The second column of Table 1 summarizes the accuracy for experts when using a SVM classifier. For each expert we evaluated both linear and RBF kernels. The Table includes the best performance achieved, that was provided by a linear kernel for LBP, LPQ, WLD and LTP, while a RBF was better for LDP and LGP. The trade-off between margin and error, i.e., parameter C, was always fixed at $C = 1$, while for the RBF kernel, the gamma value was fixed at $gamma = 0.07$. In both columns the best achieved values are in bold. Those results are reported considering a decision threshold of 0 (negative vs. positive values). A SVM-based classifier outputs a score that also indicates the sample proximity to the threshold, and thus might be further used to evaluate the individual classification quality or ambiguity.

Table 1. Gender classification accuracy of the single experts using either SVM on feature vectors or LR on feature images. As for SVM, the best performance was provided by a linear kernel for LBP, LPQ, WLD and LTP, and by RBF for LDP and LGP.

Protocol	LR Accuracy	SVM Accuracy (# features)
LBP	86.36%	91.36% (6400)
LGP	90.005%	88.18% (6400)
WLD	91.36%	92.27% (6400)
LPQ	85.00%	90.45% (6400)
LTP^{low}	90.45%	91.36% (6400)
LTP^{high}	89.55%	**93.18%** (6400)
LDP	**93.19%**	86.36% (16384)

3 Fusion Strategies

Fusion can occur either at feature, matching score, or decision level. The first one retains the most information, but it is usually computationally more demanding. The last one looses too much information before the final result. The best compromise is usually achieved at score level. In this work we investigate and compare multi-expert systems using three different fusion protocols.

F-SVM performs feature level fusion and is based on Support Vector Machines (SVM): a single linear SVM is trained on the compound feature vectors, which are obtained here by stacking the histograms produced by the above described methods; given the set of experts, $\varOmega = E_1, E_2, ..., E_n$, the protocol combines the whole set of feature vectors $\varPhi_{E_{i,k}} = f_1, f_2, ..., f_{m_{E_{i,k}}}$ extracted by individual experts E_i for a given image I_k in the new compound vector.

S-SVM is based on SVM too but uses score level fusion: more first stage SVMs are trained on the different features vectors, which are represented here by histograms produced by the above methods; the protocol then feeds the responses of individual experts E_i for a given image I_k to a second stage SVM classifier; given $\varOmega$, a set of experts, and their respective scores s_i, a new feature vector is composed as $\varSigma = s_1, s_2, ..., s_n$, and fed to a preliminary trained linear SVM.

S-LR uses score level fusion after the single experts have used LR to compute their individual scores using the feature images produced by the adopted operators; the S-LR protocol combines the responses of individual experts E_i for a given image I_k by examining them in pairs and selecting the best pair. More in detail, given $\varOmega$, a set of experts, each of which produces a score s_i, for each possible pair (E_i, E_j) with $i \neq j$, S-LR checks if both experts have voted for the same class (male, female), or $sign(s_i, s_j)$. If this is true, the pair of experts is assigned a value of $s_{i,j} = sign(s_i) \cdot \sqrt{s_i \cdot s_j}$, which represents the fused score. Otherwise, the protocol assigns the value $s_{i,j} = 0$ to the pair. At the end, the protocol S-LR selects the pair of experts that provides the maximum $s_{i,j}$ in absolute value, or $s_{global} = Max_{i,j}(|s_{i,j}|)$.

Ambiguous answers All protocols presented provide a score as the final result. The sign of this final score depends on the class (male, female), to which the input face was assigned by the system. Here, the males have a negative score, while females have a positive score. Some face images, produce a score which is very close to the value 0, which indicates a high degree of uncertainty of the response. It is possible to set a threshold th_s such that the response of the system is considered reliable if $abs(s_{global}) \geq th_s$ and ambiguous otherwise. The ambiguous answers are not necessarily discarded; they can be considered as particularly complex cases for the system, that need to be treated separately, for example with the interrogation of a further group of experts. In the experimental results, we present the curves that show the number of ambiguous answers and the performance of different indexes of accuracy adopted for the evaluation of the system versus variations of this threshold. It is worth underlining that strategies producing a similar genuine accept rate can differ by the number of ambiguous responses: of course, the lower this number, the better.

4 The Image dataset

EGA (Ethnicity, Gender and Age face database) has been designed and implemented to provide demographics balance among dataset images as well as flexibility even along time. It integrates into a single dataset face images from different databases. Many of such databases are available or will be available for research. Images are drawn from publicly available ones to create a more heterogeneous and representative dataset. In particular, in order to avoid copyright infringement, EGA has been conceived as a set of links to files previously processed by appropriate scripts, and of annotations, which are provided to organize images according to individual features such as ethnicity, gender and age. Each user can ask and obtain on her/his own the original datasets with the images needed to build EGA. The scripts will reorganize and rename all requested images, according to the structure that was devised for EGA. In this way, it is possible to easily reconstruct the whole dataset, but even to expand it, as new datasets become available and after they are annotated. At present, images are taken from CASIA-Face V5 [15], FEI [16], FERET [17], FRGC [18], JAFFE [19], and the Indian Face Database [20]. Not all images from the above databases are included in EGA, but subsets allowing an overall good balance in demographics percentages and a lower influence of factors different from demographics (e.g., pose, illumination and expression - PIE). EGA includes 469 subjects from five ethnicities: a) African-American (53), b) Asian (111), c) Caucasian (162), d) Indian (65), e) Latinos (68). For each of them, subjects are chosen aiming at the best possible balance to represent the two genders male and female. These two subgroups are further divided into three age ranges: a) young, b) adult and c) middle-aged, with adult being much more represented due to the composition of the original datasets. More details on EGA composition can be found in [5].

5 Experiments and Results

As described above, EGA includes 469 subjects. The training set for both LR and SVM was chosen to be balanced with respect to demographics, therefore we included the first half of male subjects and the first half of female subjects for each ethnicity, resulting in 124 male and 111 female subjects in the training set. All the remaining subjects were included in the test set. As for the latter, the first image for each subject was used for gallery, and the second for test.

Accuracy is defined as the number of correct classifications in relation to the total number of samples analyzed, $Acc = \frac{(TM+TF)}{M+F}$, where TM and TF stand for the number of correct male and female classifications, while M and F indicate the number of total male and female samples tested.

We remind that when exploiting feature level fusion, for all operators except LDP, histograms are extracted after dividing the image in a 5×5 grid and then chained in the final vector. As for LDP, the whole image is used. When exploiting score level fusion, feature images are used, with each pixel of the original image replaces by the code in $[0, 255]$ produced by the operator at hand. Table 2 shows the results, with best ones reported in bold. F-SVM and S-SVM reported similar results, but the first approach is particularly slow with limited possibilities of parallelization. It is also interesting to notice that the combination of operators providing the best results with all three fusion approaches

Table 2. Gender Classification accuracy of the fusion approaches.

Fusion protocol	ALL	LDP WLD	LTP^{low} LTP^{high}	LTP^{low} LTP^{high} WLD	LTP^{low} LTP^{high} WLD LBP	LTP^{low} LTP^{high} WLD LDP	LTP^{low} LTP^{high} WLD LBP LPQ	LTP^{low} LTP^{high} WLD LBP LPQ LDP
F-SVM	93.20%	92.27%	92.27%	93.63%	**94.09%**	**94.09%**	**94.09%**	93.18%
S-SVM	93.20%	92.27%	92.73%	**94.55%**	**94.55%**	**94.55%**	93.18%	93.18%
S-LR	91.16%	**92.08%**	90.08%	90.93%	90.443%	**92.08%**	89.96%	90.99%

is $\{LTP^{low}, LTP^{high}, WLD, LDP\}$, suggesting that the addition of LBP and LPQ only introduces redundant information if not noise. In any case, the accuracy achieved shows a rather limited improvement of feature or score level fusion compared to some single experts. Indeed, the best multi-expert fusion achieved 94.55 using S-SVM vs. 93.19 with LDP and LR-based classification. However, there is a positive effect too that is not shown by those results. As a matter of fact, it is important to give the appropriate focus to the ability of the compound system to reduce ambiguous cases.

In Figure 2 we show the accuracy improvement with respect of ambiguous responses (see Section 3), i.e., responses which are too close to the threshold dividing the two classes and therefore might produce errors.

In order to compute the improvement produced by setting apart ambiguous responses, we do not count them in the denominator of the expression for Accuracy. Those responses that are farther from the classification border present a better classification rate. Therefore, we can notice the compromise achieved: a slightly lower number of useful responses vs. an increased classification precision.

It is worth further underlining two observations: first, we should consider the curves obtained with respect to two thresholds, the one used for classification M/F and the one set for ambiguity. However, since the former distributions are normalized, the threshold is fixed and is 0 (negative vs. positive values) therefore this threshold is fixed and left implicit in the graphics; second, as mentioned before, discarding the ambiguous responses improves the accuracy of classification but has a cost, represented by the lower number of useful responses (accuracy is computed considering unambiguous responses only). The second observation implies that, since the accuracy is computed as a proportion with respect to the useful responses, the lower the number of ambiguous responses the better the system, even when producing the same accuracy value.

The advantage of the fusion approaches is evident in Figure 2, which shows that the multi-expert approach reduces the number of errors due to ambiguous cases. In order to maintain the figure readable, we only reported the curves corresponding to the best single classifier and to the fusion of all operators for the different protocols. SVM approaches report a better performance, behaving in a quite similar way for both F-SVM and S-SVM. However, the latter, even if it reported a slightly worse accuracy, would be preferred as it is more flexible in terms of usability and cost. We can notice that the curve for the single operator LDP starts better, but the fusion significantly improves the performance when a higher number of ambiguous responses is accepted, and this is a further advantage of the multi-expert approach.

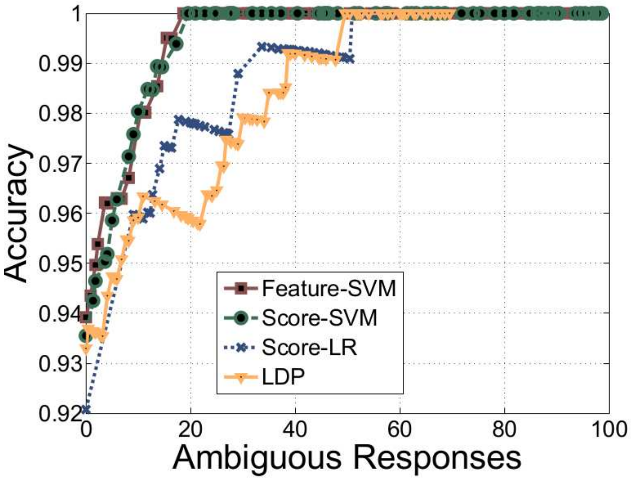

Fig. 2. Number of ambiguous cases vs. accuracy achieved for the remaining responses.

6 Conclusions

We analyzed the behavior of a number of local operators, namely LBP, LGP, LTP, LDP, WLD and LPQ, in addressing the problem of gender classification. Our proposal deals with fusing the results of such operators, in order to achieve an improvement of performance in terms of both accuracy and reliability (possible ambiguity of classification results). The benefits of adopting a fusion approach are confirmed by experimental results both in terms of accuracy and in terms of robustness to ambiguous samples. In particular, the identification of ambiguous cases allows to get better relative performance by discarding the corresponding responses. The cost to pay is a lower number of useful responses. However, a further advantage achieved by fusion is that such number increases (ambiguity decreases) with all presented fusion approaches. As a consequence, the same accuracy is achieved in conjunction with higher overall rate of classified images.

References

1. B. F. Klare, M. J. Burge, J. C. Klontz, R. W. Vorder Bruegge, and A. K. Jain. Face recognition performance: Role of demographic information. *IEEE Trans. on Information Forensics and Security*, 7(6):1789–1801, December 2012.

2. J. Bekios-Calfa, J. M. Buenaposada, and L. Baumela. Robust gender recognition by exploiting facial attributes dependencies. *Pattern Recognition Letters*, 36:228–234, 2014.
3. Choon Boon Ng, Yong Haur Tay, and Bok-Min Goi. Recognizing human gender in computer vision: A survey. In *Pacific Rim International Conference on Artificial Intelligence (PRICAI)*, pages 335–346, 2012.
4. A. Torralba and A. A. Efros. Unbiased look at dataset bias. In *Computer Vision and Pattern Recognition*, 2011.
5. D. Riccio, G. Tortora, M. De Marsico, and H Wechsler. EGA - Ethnicity, Gender and Age, a pre-annotated face database. In *2012 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BIOMS)*, pages 1–8, 2012.
6. Matti Pietikäinen, Abdenour Hadid, Guoying Zhao, and Timo Ahonen. *Computer Vision Using Local Binary Patterns.* Springer, 2011.
7. B. Jun and D. Kim. Robust face detection using local gradient patterns and evidence accumulation. *Pattern Recognition*, 45(9):3304–3316, 2012.
8. Xiaoyang Tan and Bill Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. In S.Kevin Zhou, Wenyi Zhao, Xiaoou Tang, and Shaogang Gong, editors, *Analysis and Modeling of Faces and Gestures, LNCS 4778*, pages 168–182. Springer Berlin Heidelberg, 2007.
9. Baochang Zhang, Yongsheng Gao, Sanqiang Zhao, and Jianzhuang Liu. Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor. *Image Processing, IEEE Trans. on*, 19(2):533–544, 2010.
10. Jie Chen, Shiguang Shan, Chu He, Guoying Zhao, M. Pietikainen, Xilin Chen, and Wen Gao. Wld: A robust local image descriptor. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(9):1705–1720, 2010.
11. V. Ojansivu and J. Heikkil. Blur insensitive texture classification using local phase quantization. In A. Elmoataz, O. Lezoray, F. Nouboud, and D. Mammass, editors, *Image and Signal Processing, LNCS 5099*, pages 236–243. Springer, 2008.
12. Timo Ahonen, Abdenour Hadid, and Matti Pietikäinen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12), December 2006.
13. D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
14. B. Ulery, A.R. Hicklin, C. Watson, W. Fellner, and P. Hallinan. Studies of biometric fusion. Technical Report IR 7346, NIST, 2006.
15. CASIA-FaceV5. `http://biometrics.idealtest.org/`.
16. The FEI face database. `http://www.fei.edu.br/~cet/facedatabase.html`.
17. P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss. The FERET Database and Evaluation Procedure for Face-Recognition Algorithms. *Image and Vision Computing J.*, 16(5):295–306, Sept 1988.
18. P. Phillips, P. 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 *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2005.
19. M. J. Lyons, S. Akamatsu, M. Kamachi, , and J. Gyoba. Coding facial expressions with gabor wavelets. In *Proceeding of the IEEE International Conference on Automatic Face and Gesture Recognition*, pages 200–205, 1998.
20. V. Jain and A. Mukherjee. The Indian Face Database. `http://vis-www.cs.umass.edu/~vidit/IndianFaceDatabase/`.