Landmarks-SIFT Face Representation for Gender Classification

Yomna Safaa El-Din¹, Mohamed N. Moustafa², and Hani Mahdi¹

¹ Department of Computer and Systems Engineering,
Ain Shams University,
Cairo, Egypt

{yomna.safaa-eldin,hani.mahdi}@eng.asu.edu.eg

² Department of Computer Science and Engineering,
American University in Cairo,
New Cairo, Egypt

moustafa@ieee.org, m.moustafa@aucegypt.edu

Abstract. Existing methods for gender classification from facial images mostly rely on either shape or texture cues. This paper presents a novel face representation that combines both shape and texture information for gender classification. We propose extracting the Scale Invariant Feature Transform (SIFT) descriptors at specific facial landmarks positions, hence encoding both the face shape and local-texture information. Moreover, we propose a decision-level fusion framework combining this Landmarks-SIFT with Local Binary Patterns (LBP) descriptor extracted for the whole face image. LBP is known of being tolerant against uncontrolled image capturing conditions. Competitive correct classification rates for both controlled (97% for FERET) and uncontrolled (95% and 94% for LFW and KinFace) benchmark datasets were achieved using our proposed decision-level fusion.

Keywords: gender classification, SIFT, facial landmarks, LBP, fusion.

1 Introduction

Gender classification is useful for many applications such as those which require interaction between humans and computers, surveillance purposes from videos, and collecting demographic information for gender-based advertising. Correct gender recognition can also help narrow the search space for face recognition and verification [1]. For such reasons, automatic gender classification has been widely investigated in literature, using different features and classification approaches.

The extraction of features from the image is the most crucial step of the classification process, and many techniques have been proposed to extract different types of features from the images. These features can be categorized mainly into appearance-based, and shape-based features. Appearance-based approaches used the preprocessed texture of the face, while shape-based methods rely on features extracted from the face shape such as geometric relations between facial

landmarks. Studies have shown that the combination of several classifiers leads to the improvement of the overall classification accuracy [2,3].

In this paper, we propose a new face-descriptor based on SIFT features. The proposed representation implicitly combines both shape and texture information as it extracts the SIFT features at specific facial landmarks instead of using an interest point detector as in the original SIFT algorithm. We test the proposed face representation with an SVM classifier to classify gender from facial images.

Moreover, we present a fusion mechanism than combines the decisions of our proposed classifier with another depending on the appearance-based Local Binary Patterns, which are local-texture descriptors that extract information from blocks on the whole face. We also utilize this fusion scheme in combining several classifiers each trained using the same features-type yet extracted from different facial regions.

The robustness of our proposal is tested on three publicly available datasets: the FERET [5], the Labeled Faces in the Wild (LFW) [6] databases and the newly formed UB KinFace database [7]. Experimental results show that our proposed face-descriptor achieves state-of-the-art performance which is even boosted by merging it with LBP features at the decision-level.

The rest of the paper is organized as follows; Section 2 reviews previous related work, Section 3 describes our proposed face representation, and Section 4 introduces our proposed method for the fusion of these descriptors. Section 5 states the databases we used, then explains and discusses the experiments and results. We conclude our work in the final section.

2 Related Work

Various gender classification methods are reported in the literature, from which we review here only the related studies.

Lian and Lu [8] used LBP with SVM for multi-view gender classification, while some researches have combined it with other information; Ylioins et al. [9] and Alexandre [10]. Shan [11] investigated gender classification on real-life faces of LFW, where boosted Local Binary Patterns (boostedLBP) were used with SVM. They used LBP to describe faces, then used Adaboost to select the discriminative LBP features, followed by an SVM for classification.

A shape-based approach was presented by Cao et al. [12] that investigates whether topological information extracted from facial landmarks can be used to efficiently perform gender classification. They used pairwise distances and angles between each two landmarks.

Mäkinen and Raisamo [2] presented an overview on the topic of gender classification from face images, and introduced combined results of some classifiers. While Li et al. [3] combined information of hair and clothing with extracted facial features for gender classification.

Wang et al. proposed the use of dense-SIFT in [13,14], where SIFT descriptors are extracted at regular image grid points and concatenated into a vector to represent the face.

All the above mentioned methods, except [2], show very good classification accuracies on only one database, while our proposed method showed robustness to varying image conditions as it achieves best results on both the FERET and the LFW databases.

3 Face Representation

In this section, we introduce our proposed face descriptor based on SIFT features, and we briefly explain the LBP local-feature extractor.

3.1 SIFT-Based Descriptor

SIFT features were proposed by Lowe [15], and have proven to be very useful for object recognition and matching applications. These features are invariant to image scaling, translation and rotation, and partially invariant to illumination changes and affine projection. The advantage of such invariant features, is that the preprocessing stage of accurate face alignment is not required.

The SIFT descriptor is a local feature as it describes the region surrounding a keypoint, in a specific scale and orientation. We briefly describe the details of the SIFT object recognition algorithm in the following subsection, then we explain our proposed use of SIFT for gender recognition.

Scale Invariant Feature Transform. Lowe's [15] SIFT algorithm for object recognition starts with building a scale space of the image by convolving it with a variable-scale Gaussian kernel and then calculates the difference of Gaussian (DoG) between each successive two layers in the scale space. The extremas of the DoG are then selected as candidate interest points, from which elements with low contrast and edge responses are excluded. The remaining elements become keypoints and each is assigned an orientation using the precomputed gradients.

After the detection of keypoints, each is assigned a descriptor that summarize information on local image gradient flows. The idea of the SIFT descriptor is similar to that of the HOG descriptor where a histogram of gradient orientations is computed for an interest region around the keypoint, using N_p bins for each spatial direction and N_o bins for the orientation. In this paper, we used the original SIFT descriptor with 4 spatial bins and 8 orientation bins; that is 4×4 subregions, each with 8 angle bins covering 45° each, generating a descriptor with dimension $= N_p^2 N_o = 128$

Finally, the object recognition step involves matching of all keypoints, extracted from an image, to object prototype keypoints, recorded in a corresponding database.

Proposed Landmarks-SIFT. Unfortunately, the use of an interest point detector in the original SIFT-based object recognition algorithm is generally unsuccessful in face or gender recognition. This is because only a few keypoints per face can be detected due to the general small resolution and low quality of face images in real life image capture scenarios.

Another drawback of using the original SIFT algorithm for gender recognition, is the final step, in which keypoints matching is performed. This matching process detects the presence or absence of a certain object, which is infeasible for the two-class classification task of gender recognition. Besides, the detected keypoints for each face do not necessarily represent the same facial features as their number and location vary from face to face. Hence, the sparse SIFT features cannot be used with standard machine learning method tools like SVM.

Our first contribution in this paper is a face representation based on extracting SIFT descriptors at specific predefined facial landmarks '*Landmarks-SIFT*', not on a constant grid as the dense-SIFT of [13,14]. To the best of our knowledge, this technique has not been used for gender classification before.

Active Shape Model (ASM) is used to automatically locate 76 facial landmarks, as shown in Figure 1, and at the location of each landmark, a 128-dimensional SIFT descriptor is obtained from 16×16 pixels around the landmark. The descriptors are then concatenated into a single vector and used as input to an SVM classifier, which is a classic pattern classification algorithm instead of the keypoints' matching process. We adopted SVM with Radial Basis Function (SVM+RBF).

Fig. 1. Feature extraction process for Landmarks-SIFT, LBP and dense-SIFT

3.2 Local Binary Patterns

Local Binary Patterns (LBPs) [4] are features that are calculated from pixel intensities in a pixel neighborhood. It is a very simple but efficient algorithm for local texture information extraction and it is also stable under illumination changes and rotation.

The original local binary pattern operator labels the pixels of an image by thresholding the 3×3 or more neighborhood of each pixel with the center value and concatenating the results to form one binary number.

Uniform patterns LBP is an extension to the original LBP operator where all the uniform patterns have separate bins, while the rest of the patterns are collected in one bin. The uniform patterns are those which have at most 2 transitions from 0 to 1 and vice versa.

In this research, the operator $LBP_{8,r}^{u^2}$ is used to translate the pixels values into binary numbers, using eight neighbors around each pixel at the radius of $r = 3$. The image is then partitioned into 9×9 non-overlapping blocks, as shown in Figure 1, and the histogram of each block is calculated and concatenated to form one features vector.

4 Decision-Level Fusion

In this section, we propose a fusion method to combine the decisions of several classifiers into one final decision in order to achieve higher gender recognition accuracy. Since our first classifier depends on features extracted from specific face regions, its performance can be boosted through combining it with another classifier that extracts information from the whole face area. In general, the fusion process is performed either before classification (features-level) or after classification (decision-level). We adopt the decision-level fusion mechanism with the general architecture depicted in Figure 2.

Fig. 2. Decision-level fusion mechanism. Each Expert returns an output o with score s

We have N individually trained single-feature classifiers, a combiner is trained using a different subset of images named ‘*Fusion-Training*’ set. Each expert n is used to classify the images in this set and return a set of classification results associated with confidence values (scores).

First step, before training the combiner, involves the calculation of a ‘*likelihood*’ function for the scores of each expert individually. This is done by dividing the range of score values returned by an expert into I intervals and calculating a ‘*likelihood*’ value for each interval separately. The ‘*likelihood*’ value of a score interval is calculated as the difference between the percentage of correctly classified images, having a score value within this interval, and those that were wrong.

We shift all calculated ‘*likelihood*’ values to be above zero and finally normalize to $[0 - 1]$. For our machine, we choose $I = 20$.

Next, for training the combiner, each image is represented by an N -dimensional vector to train the combiner using SVM. Each feature in that vector represents the ‘*likelihood*’ value of the confidence score returned by expert n , signed with the classification decision; positive for male and negative for female.

5 Experiments and Results

5.1 Datasets

To evaluate our proposed method, we used three publicly available face image databases, two of which are benchmark datasets: the FERET [5] and the Labeled Faces in the Wild [6]. The third set: UB KinFace database [7], has not been used before for the purpose of gender classification except for [17], it has been mainly used for the purpose of kinship verification and recognition.

The FERET dataset contains studio-setting constrained images, while LFW and KinFace offer a collection of faces captured from the web, representing a variation of expressions and lighting conditions. From the FERET database we used one image per subject from the frontal ‘*fa*’ gallery. For the LFW we selected the frontal faces and formed a set containing 4500 males and 2340 females, having at most two images of the same subject. The UB KinFace dataset contains 600 images of 400 people, the images are divided into 200 groups, each is composed of child, young parent and old parent images.

For each dataset, we created three subsets: ‘*Training*’ set (used to train the classifiers), ‘*Fusion-Training*’ set (used to train the combiner in decision-level fusion), and, finally, ‘*Testing*’ set (used to evaluate the classification performance).

In our experiments, we adopted a 5-folds cross validation test scheme, where the images are divided into five folds, keeping the same ratio between male and female faces; duplicate images of the same subject (in case of LFW and KinFace) are placed in the same fold. One fold is used for ‘*Testing*’, two for ‘*Training*’ and the remaining two for ‘*Fusion-Training*’; this process is repeated five times and the average is reported. The number of faces used for each database is shown in Table 1.

Table 1. Number of faces used for each database (Male/Female)

	FERET	LFW	KinFace
All	600/405	4500/2340	440/160
1 Fold	120/81	900/468	88/32

For implementation, images preprocessing, training and testing, we used MATLAB. Colored images are transformed to gray-scale, and then the lighting was enhanced using MATLAB’s built-in function *imadjust* which increases the contrast by remapping the intensity values to fill the entire range of $[0 - 255]$.

5.2 Region-Based Landmarks-SIFT

In a first experiment, we study the contribution of each group of facial landmarks on the gender classification performance. We split the 76 landmarks into 6 groups according to the facial regions they describe. The 15 landmarks on the face-contour are divided into two groups: ‘*Contour-L*’ and ‘*Contour-R*’ with the chin point repeated in each group. The points located on the left eye and the left eyebrow are collected in a fourth group of 15 points, ‘*Eye-L*’, and the same for the right eye, ‘*Eye-R*’. Twelve points are grouped in the ‘*Nose*’ group while the remaining 19 landmarks belong to the final ‘*Mouth*’ group.

(a) Gender classification results of facial-landmarks groups.

(b) Fused gender classification results of facial-landmarks groups.

Fig. 3. Gender classification results of facial-landmarks groups and their fusion

Figure 3(a) shows the classification results obtained by each group of landmarks separately, along with the results obtained using all the 76 landmarks, which corresponds to a feature-level fusion of the 6 groups. It can be noticed that the eyes region achieves the highest classification rates of the individual groups, yet its performance does not exceed 90% correct rate, hence the need of combining the information extracted from all the regions. In Figure 3(b), we compare the results obtained by fusing the 6 classifiers using our decision-level fusion mechanism versus the voting decision-level fusion and also against feature-level fusion.

5.3 Fusion Results

In this subsection, we present the results obtained using our Landmarks-SIFT face representation and its fusion with LBP. We also compare with the dense-SIFT descriptor. For Landmarks-SIFT, we used Stasm [16] which is an extended version of ASM to automatically locate the 76 landmarks on the face image. While for LBP and dense-SIFT, the image is preprocessed and the face area is extracted as a square with dimensions relative to the inter-eyes distance. The cropped face is scaled to be of size 144×144 and then divided into 9×9 blocks

each of size 16×16 pixels. The operator $LBP_{8,3}^{u^2}$ is used to extract and form the LBP features vector of the image.

Table 2 presents the average 5-folds cross-validation results for each dataset and compares the results of the our proposed Landmarks-SIFT face representation with LBP and dense-SIFT. The table also shows the results of our proposed decision-level fusion mechanism and compares it with feature-level fusion. From Table 2, it can be noticed that our proposed fusion scheme improves accuracy by up to 2% over the best single-feature classifier.

Table 2. Gender classification results of the single-feature classifiers and their fusion

	FERET	LFW	KinFace
Single-Feature Classifiers :-			
LBP	94.43%	93.10%	90.05%
Dense-SIFT [$9 \times 9 = 81$ Keypoints]	94.63%	92.85%	90.41%
Landmarks-SIFT			
-All 76 Keypoints	95.22%	93.06%	93.10%
-Decision-level Fusion of 6 Groups	96.42%	93.40%	93.10%
Landmarks-SIFT + LBP :-			
-Feature-level Fusion:	96.81%	93.97%	94.28%
-Decision-level Fusion:			
*[All 76 Landmarks]+[LBP]	96.62%	94.33%	94.28%
* $6 \times$ [Landmarks-Group]+[LBP]	97.11%	95.00%	94.45%

5.4 Comparison with Previous Work

Table 3 compares our proposed approach with previously published classification results on all mentioned datasets. For LFW, we compared with Shan's [11] work and showed comparative results. For the KinFace dataset, we compared with our previous proposal in [17] where we used simple features extracted the face image and combined them at the decision-level achieving 90% accuracy.

Table 3. Comparison with previously published results

	FERET	FERET_MR	LFW	KinFace
Proposed Decision-level Fusion of Landmarks-SIFT+LBP:	97.11%	97.00%	95.00%	94.45%
	95.80%	92.86%	94.81%	90.00%
Best Published Results:	Li et al. [3]	Mäkinen and Raisamo [2]	Shan [11]	Y. S. El-Din et al. [17]

For FERET, Li et al. [3] selected 782 images from the dataset with equal numbers of males and females, and used one quarter for testing and the rest for training. They reported best average results of 95.8%, using fuzzy combination. Our work outperforms the reported methods on the FERET dataset with

average 97.1%. We have also reported results on a subset of FERET that was used by Mäkinen and Raisamo in [2], which we refer to as FERET_MR. The subset contains 760 face images divided equally between male and female from which they used 80% for training and 20% for testing. They combined six gender classification methods using four types of combination methods, and achieved 92.86% accuracy. We report results of our method on their set, splitting the 80% that are used for training equally between the ‘*Training*’ and ‘*Fusion-Training*’ subsets.

6 Conclusion and Future Work

In this paper we proposed a new SIFT-based representation for face images that combines both texture and shape information. We compared our proposed Landmarks-SIFT on the task of gender classification, with a local texture descriptor, LBP, that extracts local-information from blocks covering the whole face image. Competitive results were reported on three publicly available datasets: FERET, LFW and the KinFace databases.

Moreover, we presented a mechanism for combining the proposed Landmarks-SIFT with LBP at the decision-level to fuse the local and global features, which successfully boosted the gender classification performance by up to 2% from the best single-feature classifier. We also used this fusion technique to merge decisions of six classifiers each trained on SIFT-descriptors extracted at a subset of the facial landmarks, and showed that the fusion of decisions of several individual classifiers achieves better performance than using all the features with a single classifier.

High gender classification rates were reported in comparison to previously published results on the three datasets, where 97.2% was achieved for FERET, 95.03% for LFW and 94.28% for KinFace.

For future work, we will try to apply dimensionality reduction techniques to the SIFT feature vector in addition to using other features extracted from the face image including SURF, FREAK or boosted LBP. We would also like try more fusion methods, and also to experiment with SVMACs proposed by Ji and Lu [18].

References

1. Ross Beveridge, J., Givens, G.H., Jonathon Phillips, P., Draper, B.A.: Factors that influence algorithm performance in the face recognition grand challenge. In: *Computer Vision and Image Understanding*, pp. 750–762 (2009)
2. Makinen, E., Raisamo, R.: An experimental comparison of gender classification methods. *Pattern Recognition Letters* 29, 1544–1556 (2008)
3. Li, B., Lian, X.-C., Lu, B.-L.: Gender classification by combining clothing, hair and facial component classifiers. *Neurocomputing* 76(1), 18–27 (2012)
4. Ojala, T., Pietikainen, M.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(7), 971–987 (2002)

5. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The FERET Evaluation Methodology for Face-Recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 1090–1104 (2000)
6. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Tech. Rep. 07-49, University of Massachusetts, Amherst (October 2007)
7. Xia, S., Shao, M., Luo, J., Fu, Y.: Understanding kin relationships in a photo. *IEEE Transactions on Multimedia* (2012)
8. Lian, H.-C., Lu, B.-L.: Multi-view gender classification using local binary patterns and support vector machines. In: Wang, J., Yi, Z., Žurada, J.M., Lu, B.-L., Yin, H. (eds.) *ISNN 2006*. LNCS, vol. 3972, pp. 202–209. Springer, Heidelberg (2006)
9. Ylioinas, J., Hadid, A., Pietikäinen, M.: Combining contrast information and local binary patterns for gender classification. In: Heyden, A., Kahl, F. (eds.) *SCIA 2011*. LNCS, vol. 6688, pp. 676–686. Springer, Heidelberg (2011)
10. Alexandre, L.A.: Gender recognition: A multiscale decision fusion approach. *Pattern Recogn. Lett.* 31(11), 1422–1427 (2010)
11. Shan, C.: Learning local binary patterns for gender classification on real-world face images. *Pattern Recognition Letters* 33(1), 431–437 (2012)
12. Piccirilli, M., Adjeroh, D., Bourlai, T., Ross, A., Cao, D., Chen, C.: Can facial metrology predict gender? In: *Proc. of International Joint Conference on Biometrics (IJCB)* (2011)
13. Wang, J.G., Li, J., Yau, W.Y., Sung, E.: Boosting dense sift descriptors and shape contexts of face images for gender recognition. In: *Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 96–102 (2010)
14. Wang, J.G., Li, J., Yau, W.Y., Sung, E.: Dense sift and gabor descriptors-based face representation with applications to gender recognition. In: *Proceedings of the 2010 11th International Conference on Control Automation Robotics and Vision (ICARCV)*, pp. 1860–1864 (2010)
15. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), 91–110 (2004)
16. Milborrow, S., Nicolls, F.: Locating facial features with an extended active shape model. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008, Part IV*. LNCS, vol. 5305, pp. 504–513. Springer, Heidelberg (2008), <http://www.milbo.users.sonic.net/stasm>
17. El-Din, Y.S., Moustafa, M.N., Mahdi, H.: A Mixture of Two Gender Classification Experts. In: *25th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pp. 245–251. IEEE Computer Society, Brazil (2012)
18. Zheng, J., Lu, B.-L.: A support vector machine classifier with automatic confidence and its application to gender classification. *Neurocomputing* 74(11), 1926–1935 (2011)