Unconstrained gender classification by multi-resolution LPQ and SIFT

Huu-Tuan Nguyen and Trinh Thi Ngoc Huong
Faculty of Information Technology
Vietnam Maritime University
484 Lach tray, Ngo Quyen, Hai Phong
Email: huu-tuan.nguyen@vimaru.edu.vn, huongtn@vimaru.edu.vn

***Abstract*—One of the most critical tasks in building a gender classification is how to describe the human face as a highly discriminative feature vector. To this end, in this paper we introduce a new handcrafted feature extraction method for unconstrained gender classification problem. From one input face image, we generate its smaller version and apply two LPQ operators on both of them. We then combine the obtained LPQ features with the SIFT feature extracted from the given image to constitute a global facial description. In the classification stage, the binary SVM classifier is used for determining the gender of the test images. To evaluate the recognition performance of the proposed methods, we carry out experiments upon two widely used unconstrained face databases Adience and LFW. The results show that our approach attains good classification rates (96.51% and 80.5% on LFW and Adience databases, respectively) and can be comparable with state-of-the-art systems.**

I. Introduction

From a human face image, computer vision systems can extract a lot of helpful clues, e.g. identity, expression, age, ethnic and gender, that help us to understand deeper about related individual. Gender, among others, is one of the most important attributes as it can be used in many real life applications such as targeting advertising, human-machine interaction, security and demographic research. Thus, a plethora of approaches have been proposed to build automatic gender classification systems [1], [2].

Given a face image, a gender classification system will output the corresponding gender based on the similarity of the image with other images within a gallery (or reference) set. This task can be done at a glance by human beings but it is a very challenging quest for machines to proceed. Similar to a human vision system, there are multiple stages in a gender classification framework. The first stage is face detection whose objective is to locate exactly the locations of face regions in an input image. After the detection step, an alignment technique is used to align the detected face images so that they have the same facial fiducial landmarks and are in frontal direction. The next stage is to apply a feature extraction method to extract the most discriminative features from the aligned images and store the results as high dimensional feature vectors. Then the feature vectors are fed into the classification stage to estimate the gender of the test images by a classifier.

The face detection stage is usually performed by Haar-like features [3] as in [4]. In the meantime, there are several choices for face alignment: one can use a simple rotate algorithm based on eyes' coordinates or follows an automated alignment method such as a simple approximation that used a 3D single, unmodified surface for face alignment and achieved very good results in face recognition and gender classification in [5]. According to our best of knowledge, the biggest variation is lied on the feature extraction stage where many methods were introduced. Interesting results were reported by a multi-view system [6] in which LBP (Local Binary Patterns) [7] and SVM were used for extracting facial features and classification, respectively. The modified LBP [8] and SVM were also combined and obtained quite high results [9]. A multi-scale fusion approach was presented by Alexandre [10] with the usage of both shape and texture features extracted on multi-scale images. In [11], LBP was fused with DCT and offered a recognition rate of 98.16% on FERET database [12]. While pre-cited systems carried out experiments upon constrained face database, the learned LBP system in [13] conducted tests on unconstrained database LFW [14] and generated a performance of 94.81%. Also examining the recognition performance on real life face images, in [15] LBP and FLBP (Four-patch Local Binary Patterns) [16] were used and provided excellent accuracy on Adience database whose images were gathered under unconstrained conditions. Another gender classification system, which had amazing results, was proposed by Levi et al. [17]. The authors utilized a simple convolution neural network upon intensity images and showed the best recognition rate on Adience database.

The goal of this paper is building a gender classification framework to deal with unconstrained face images. Towards this end, we adopt and fuse two powerful feature extraction methods, the Local Phase Quantization (LPQ) [18] and SIFT [19], for facial images' descriptions. For an input image, we first generate a smaller version image and the LPQ features are extracted from that two intensity images (the original input one and its resized copy). Simultaneously, the SIFT method is applied upon the given image to produce SIFT features. Lastly, multi-resolution LPQ and SIFT features are fused to form a final facial representation. In this work, the binary SVM classifier is selected for classification stage. To assess the accuracy performance of the proposed framework, we follow the standard protocols and perform experiment on two unconstrained and public face databases Label Face in the Wild (LFW) [14] and Adience [15]. The comparisons with other state-of-the-art systems show that our approach can work well under real life conditions and offers excellent recognition accuracies on both data sets.

978-1-5090-2100-0/16/$31.00 ©2016 IEEE

The organization of this work is structured as follows. The multi-resolution LPQ and SIFT approach for feature extraction and the corresponding gender classification system are presented in Section 2. In Section 3, we show the comparisons of experimental results upon LFW and Adience data sets with those of other works. Finally, some conclusions and future works are described in Section 4.

II. Multi-resolution LPQ and SIFT features

A. LPQ

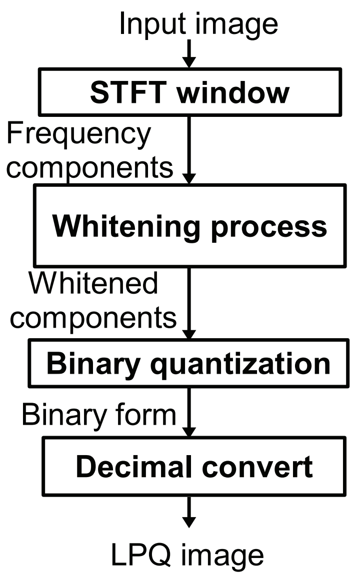

Fig. 1: Step in LPQ operator

The LPQ method, whose steps are illustrated in Fig. 1, is based on the blur-invariant property from the phase component of an image in the frequency representation. To generate an LPQ image, a Short Term Fourier Transform (STFT) is performed on a window patch at every image pixel by using four scalar low frequencies. This is achieved efficiently through multiple 1D convolutions of the input image with kernels derived from a, which is the highest frequency value that ensures the blur invariant property. Let $M \times M$ is the size of the window patch, the value of a is computed as:

$$a = \frac{8}{M-1}. \tag{1}$$

Then a whitening transformation process is applied with four real and four imaginary components to make LPQ features more efficient. In this step, a parameter, denoted as ρ, is used when computing the transformation matrices. After that, the whitened components are binary quantized to produce binary strings for all pixels in the input image. These binary forms are next converted to decimal values and results in as a gray-scale image. For detailed equations in above-mentioned steps, one can refer to [18] and the references therein.

The LPQ features are extracted from the LPQ image via two steps. In the first step, the LPQ image is sub-divided into non-overlapped rectangular sub-regions. The histogram sequences of those sub-regions are computed and are concatenated to constitute the LPQ feature vector for one input image in the next step.

B. SIFT

Introduced by Lowe [19], SIFT features are not only scale and rotation invariant but also are strong to illumination variations. SIFT method consists of several steps: scale-space extrema detection, keypoint localization, orientation assignment and keypoint descriptor. In the first step, the goal of searching all scales and locations are accomplished by using Difference-of-Gaussian filter to detect potential points that can be scale and rotation invariant. The keypoints are chosen, at the second step, based on their stability measured by model that is fit for determining scale and location. The third step assigns orientations to every keypoint based on directions of image gradient. In the last step, each keypoint descriptor is computed as an orientation histogram of its neighborhood.

C. Facial representation by multi-resolution LPQ and SIFT

The feature extraction method introduced in this work is constructed by using multiple LPQ operators and SIFT to extract phase based and scale-invariant features from face images. LPQ is proven to be an elite method in the face recognition literature whilst SIFT is widely known as a local feature representation. All the steps of our proposition is demonstrated in Fig. 2. Firstly, a LPQ descriptor is used on the input image to capture original phased based features. In the meantime, a smaller version of the input image is created and another LPQ operator is applied upon that copy for extracting additional phase based texture features. The two LPQ operators are different by their parameters, the windows' sizes, ρ and the grids for sub-dividing LPQ images, and the resulting features are complementary to make the final description more powerful. We do not slavishly believe that just one LPQ descriptor is sufficient for face representation when the gender classification system has to cope with unconstrained factors such as variations of lighting conditions, head pose, age, expressions, occlusions and race. Hence, by using LPQ operators on multi-resolution images, we expect that the final description contains more meaningful characteristics than when only one LPQ descriptor is applied. One raised question is that whether more LPQ features on other copies derived from the input image are better? We tried such direction but there were two issues: the accuracy performance was not improved while the computational workload was heavier since the size of the feature vector was larger. Thus, empirically, two LPQ operators are chosen.

The SIFT features are extracted from the original input image and are accumulated with above LPQ ones to make the global feature vector. In this work, orientation histograms of SIFT are not built with automatically detected keypoints. Rather than that, we split the input image into smaller regions by evenly divided horizontal and vertical lines and specify these lines' intersection points as keypoints in original SIFT method. The root of this approach is that the lengths of SIFT vectors of different face images are different since the images' numbers of keypoints are not identical. In Fig. 2, the white dots are crossing points upon which we extract the SIFT feature

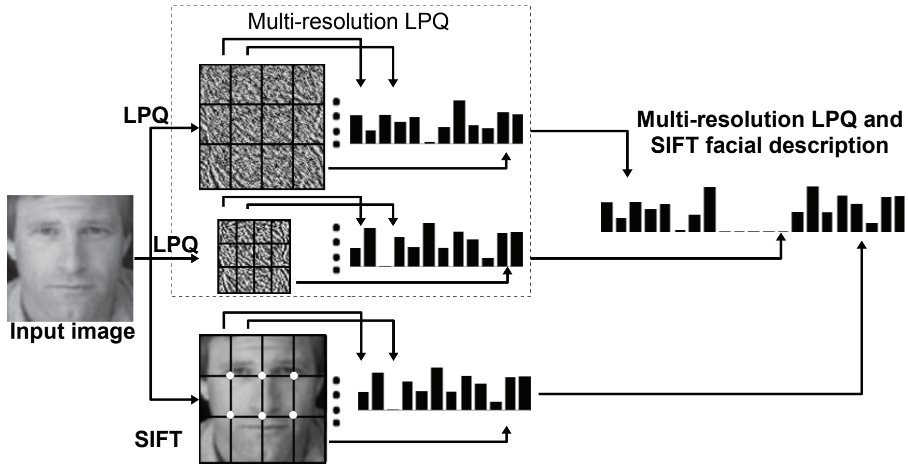

Fig. 2: Steps in using multi-resolution LPQ and SIFT for feature extraction

for our method. We adopt SIFT features to fuse with multi-resolution LPQ features due to several interesting properties they have:

1) As both SIFT and LPQ features are histogram based, we can expect that the fusion of them at feature level can bring higher discrimination when applying in gender classification problem.
2) LPQ feature is per se blur tolerant and has local phase information while SIFT feature is based on local gradient orientation and invariant to scale and rotation. So when one global description is built by combining them together, it encompasses many powerful characteristics from two kind of features.

In this work, the biggest grid sizes for the two LPQ descriptions (the one applies on the original input image of 128×128 resolution and the other for the smaller resized copy of 64×64 resolution) are 11×11 and 7×7, respectively. Additionally, the SIFT features are calculated from $7 \times 7 = 49$ intersection points. Consequently, the final feature vector's size is at most $(11 \times 11 + 7 \times 7) \times 256 + 49 \times 128 = 49280$ dimensions.

D. Gender classification by multi-resolution LPQ and SIFT

Equipped with the feature extraction method described in previous Section, we form a gender classification framework (illustrated in Fig. 3) by plugging the binary SVM classifier into the classification stage. The input images are first aligned based on the their eyes locations (LFW database) or simply cropped the center regions (Adience database). No preprocessing is applied and the resulted images are fed directly into the feature extraction stage where the feature vectors are extracted for all training and probe images. Next, the SVM trainer uses training vector to generate the decision plane for determining the gender of the test images afterward. Related to SVM tasks, the C++ interface of the library LibSVM 3.21 [20] is adopted to implement our system since it offers high speed and accuracy. Plus, the LPQ and SIFT methods are implemented using C++ language and OpenCV (http://opencv.org) library.

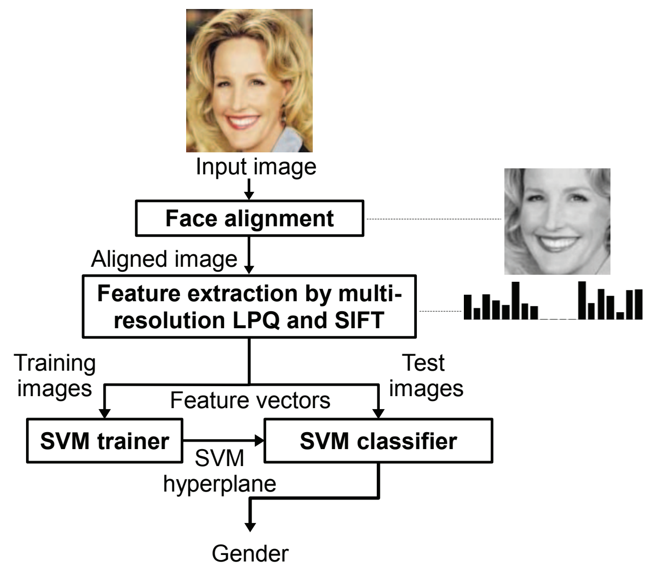

Fig. 3: Gender classification framework by LPQ and SIFT.

Before being used by the SVM classifier, the feature vectors are transformed into a latent space with the usage of a kernel function. Via empirical experiments with different available functions, the polynomial kernel is eventually selected due to the highest classification rates it achieves. Let X and Y be the two feature matrices, then the polynomial kernel matrix between them is calculated as

$$K(X,Y) = (X^T * Y + c)^d, \tag{2}$$

where d is fixed as 1 and c is set to 0.

III. Experimental results

Since the objective of this paper is to build a gender classification that can deal with unconstrained conditions, we choose two public databases, LFW and Adience, which have images acquired in outdoor environment and in real life situations, to evaluate the performance of our framework. All the tests are done following the standard protocols. The results are compared with those of other existing systems that use the same evaluation methods.

A. Results on LFW database

Fig. 4: Sample cropped images from LFW database.

The Labeled Faces in the Wild (LFW) [14] (Fig. 4 show some cropped samples) is one of the most widely used databases for studying the unconstrained facial analysis problems, including the gender classification one. It contains 13,233 real life pictures of 5,749 people collected on the Internet. Variations of illumination, facial expressions, head pose, occlusions and qualities are main properties of this database. In this work, we use the benchmark protocol proposed in [21] to perform a 5-folds experiment with LFW to evaluate the performance of our system. For more details, one can refer to the online file at the URL http://fipa.cs.kit.edu/downloads/LFW-gender-folds.dat.

For face cropping and alignment, we utilize the manual eyes' coordinates available at http://lda.tsu.tula.ru/FD/lfw_eyes.zip to rotate, crop and then resize down to 128×128 resolution to proceed the experiments.

The comparisons between our approach's results and those of other systems are presented in table I. It can be seen from table I that our proposed framework outperforms all other state-of-the-art results. This superiority is even more impressive when considering that most of other systems (except the one in [23]) used much less images than ours. With the average classification rate reaches to 96.51%, we conclude that the combination of multi-resolution LPQ and SIFT can work well with unconstrained images captured from real life scenarios under challenging conditions, such as variations of head pose, age, expressions, occlusions and low quality test images.

B. Results on Adience database

Adience [15] (whose some samples are shown in Fig. 5) is a recent public database designed for age and gender estimation systems' evaluations. It has 26,580 images (of 2,284 individuals) collected from user-uploaded galleries of Flickr and these images were captured by smart-phones' cameras. Thus, the human faces within the database are influenced by various unconstrained conditions and under different real life scenarios. Hence, Adience images are divergent in illumination conditions, pose, age, resolution and are challenging to any facial analysis system. It is worth mentioning that tests on Adience are more difficult than those upon LFW database.

Fig. 5: Sample cropped images from Adience database.

The Adience gender benchmark offers a full experiment to use all face images and a restricted test in which only frontal images are exploited. In this paper, we adopt the frontal-image experiment. In total, we use 13560 images. These images are divided into five folds and a 5-fold cross validation is carried out. The classification rate of each fold is calculated with the training set formed by images from the other 4 folds. The final accuracy performance is the average of 5 folds' results.

The original images of the database are first re-sampled down to 250×250 pixels and then the center regions of size 128×128 are cropped for performing the evaluations.

From results in table II, one can observe that even though our result is not the highest, it is still very excellent as only the best system [17] in the literature has better recognition rate while it surpasses other leading-edge approaches. The superiority of the system in [17] is understandable since it followed deep learning, a very power full direction, for gender prediction. However, our framework has some advantages over the deep learning approach: it is simpler and very fast. The total time for feature extraction and training is about 270 seconds (when running on a Core i5-3320m laptop) while the time for predicting one test image is negligible. In the mean time, system in [17] requires four hours to train each network (on a powerful GPU machine) and takes about 200ms to classify one probe image.

Despite a very good overall classification rate of 80.5%, this result on Adience benchmark is far behind the ones upon the LFW database in previous Section. The causes of this situation are enumerated below.

1) The variations of images in Adience database are larger than the ones in LFW database. For example, faces in LFW are of adult people while those in Adience are of all ages, from very little children to elderly individuals.
2) The gender estimation problem is more difficult when dealing with children faces, even by human beings.
3) A large number of images in Adience are misaligned. Thus, the recognition stage is heavily affected.

TABLE I: Rank-1 classification rate (%) on LFW database [14].

Approach	Mean	Std. Dev	No. of images
Best from [21]	94.44	N/A[1]	7443
Best from [13]	94.81	N/A[1]	7443
Best from [22]	94.01	N/A[1]	13088
Best from [11]	95.60	0.45	7443
Best from [23]	95.76	1.21	13233
Our framework	**96.51**	**0.67**	**13233**

[1] N/A: Not available result.

TABLE II: Rank-1 classification rate (%) on Adience database [15].

Approach	Mean	Std. Dev	Notes
Best of [15]	77.8	1.3	Only frontal images for training
Best of [5]	79.3	0.0	3D frontalize
Best of [24]	79.9	N/A[1]	Automatic detection and alignment
Our framwork	**80.5**	**2.6**	**No alignment technique**
Best of [17]	86.8	1.4	Deep learning

[1] N/A: Not available result.

Although the margins between our results and those of other approaches in tables I and II are not significant, the superiority of the proposed system is clear when we look at it in more detail. For example on LFW database, while the recognition rate of system in [23] is about 0.16% (95.76% versus 95.60%) higher than the one in [11], our framework yields an improvement of 0.75% in compared with it (96.51% versus 95.76%). Besides, this is an evidence for the challenges in unconstrained gender classification problem since the improvements (between all systems, including ours) are quite moderate (less than 1% in many cases).

IV. Conclusion

In this paper, we fuse multi-resolution LPQ and SIFT features for facial representations and use them for unconstrained gender recognition problem. Our gender classification framework is then formed by adopting binary SVM classifier for classification stage. For accuracy performance evaluation, we perform experiments upon two public face databases that have images acquired under real life conditions. Comparisons between our results and those of other existing systems show that the proposed approach can cope with unconstrained conditions, e.g. variations of illumination, age and expression, and attains good recognition performance.

For future work, we would like to improve the performance (computational speed and accuracy) by exploiting a learning method to reduce the length of feature vectors down to reasonable value and carry out experiments with automatically detected face images. Also, investigating other multi-resolution approaches for gender classification is a sound direction to pursuit.

References

[1] E. Makinen and R. Raisamo, "Evaluation of Gender Classification Methods with Automatically Detected and Aligned Faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 3, pp. 541–547, Mar. 2008.

[2] ——, "An experimental comparison of gender classification methods," *Pattern Recognition Letters*, vol. 29, no. 10, pp. 1544–1556, Jul. 2008.

[3] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[4] G. Shakhnarovich, P. A. Viola, and B. Moghaddam, "A unified learning framework for real time face detection and classification," in *FG, 2002. Proceedings. Fifth IEEE International Conference on*. IEEE, 2002, pp. 14–21.

[5] T. Hassner, S. Harel, E. Paz, and R. Enbar, "Effective face frontalization in unconstrained images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4295–4304.

[6] 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*, pp. 202–209, 2006.

[7] T. Ahonen, A. Hadid, and M. Pietikinen, "Face recognition with local binary patterns," *ECCV*, pp. 469–481, 2004.

[8] B. Froba and A. Ernst, "Face detection with the modified census transform," in *6th IEEE International Conference on FG, 2004. Proceedings*, May 2004, pp. 91–96.

[9] R. Verschae, J. Ruiz-del Solar, and M. Correa, "Gender classification of faces using adaboost," *Progress in Pattern Recognition, Image Analysis and Applications*, pp. 68–78, 2006.

[10] L. A. Alexandre, "Gender recognition: A multiscale decision fusion approach," *Pattern Recognition Letters*, vol. 31, no. 11, pp. 1422–1427, Aug. 2010.

[11] A. M. Mirza, M. Hussain, H. Almuzaini, G. Muhammad, H. Aboalsamh, and G. Bebis, "Gender Recognition Using Fusion of Local and Global Facial Features," in *Advances in Visual Computing*. Springer, 2013, pp. 493–502.

[12] P. Phillips, H. Moon, S. Rizvi, and P. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, 2000.

[13] C. Shan, "Learning local binary patterns for gender classification on real-world face images," *Pattern Recognition Letters*, vol. 33, no. 4, pp. 431–437, Mar. 2012.

[14] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments," in *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*, 2008.

[15] E. Eidinger, R. Enbar, and T. Hassner, "Age and Gender Estimation of Unfiltered Faces," *IEEE TIFS*, vol. 9, no. 12, pp. 2170–2179, Dec. 2014.

[16] L. Wolf, T. Hassner, and Y. Taigman, "Descriptor based methods in the wild," in *Workshop on Faces in'Real-Life'Images: Detection, Alignment, and Recognition*, 2008, 00201. [Online]. Available: http://hal.inria.fr/inria-00326729/

[17] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in *Proceedings of the IEEE CVPR Workshops*, 2015, pp. 34–42.

[18] T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkila, "Recognition of blurred faces using local phase quantization," in *ICPR*, 2008, pp. 1–4.

[19] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004, 23852.

[20] C.-C. Chang and C.-J. Lin, "LIBSVM: A Library for Support Vector Machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, May 2011.

[21] C. Shan, "Gender Classification on Real-Life Faces," in *Advanced Concepts for Intelligent Vision Systems*, ser. Lecture Notes in Computer Science, J. Blanc-Talon, D. Bone, W. Philips, D. Popescu, and P. Scheunders, Eds. Springer Berlin Heidelberg, Dec. 2010, no. 6475, pp. 323–331.

[22] P. Dago-Casas, D. Gonzlez-Jimnez, L. L. Yu, and J. L. Alba-Castro, "Single-and cross-database benchmarks for gender classification under unconstrained settings," in *Computer vision workshops (ICCV Workshops), 2011 IEEE international conference on*. IEEE, 2011, pp. 2152–2159.

[23] H. Liu, Y. Gao, and C. Wang, "Gender identification in unconstrained scenarios using self-similarity of gradients features," in *IEEE ICIP 2014*. IEEE, 2014, pp. 5911–5915.

[24] R. Azarmehr, R. Laganiere, W.-S. Lee, C. Xu, and D. Laroche, "Real-time embedded age and gender classification in unconstrained video," in *Proc. of the IEEE Conf. on CVPR Workshops*, 2015, pp. 57–65.