https://doi.org/10.33778/kcsa.2019.19.2.091

Local Prominent Directional Pattern을 이용한 얼굴 사진과 스케치 영상 성별인식 방법

Farkhod Makhmudkhujaev*, 채 옥 삼**

요 약

본 논문에서는 성별 인식을 위해 얼굴 영상을 효과적으로 기술하는 새로운 지역 패턴 방법 Local Prominent Directional Pattern (LPDP)를 제안한다. 제안된 LPDP 방법은 성별 인식에 중요한 얼굴 모양을 명확하게 구분하기 위해 주변 패턴이 누적된 히스토그램을 통계적으로 분석하고 패턴 변화가 크게 발생하는 픽셀을 부호화 한다. 통계적인 정보를 사용하는 얼굴 모양 구분에 중요한 뚜렷한 에지 방향 패턴 영역을 구분하는 중요한 정보를 제공 할 수 있다. 이는 뚜렷한 에지 방향 패턴이 나타나는 영역의 주변도 유사한 에지 방향 패턴이 나타내기 때문에 통계적으로 특정 방향이 히스토그램에 많이 누적될 수 있기 때문이다. 또한 통계적인 방법은 주변 영역의 정보를 많이 수용하기 때문에 잡음으로 발생하는 에지 방향 변화 오류에 강력한 장점이 있다. 제안된 방법은 기존 방법들 보다 더 강력한 성별인식에 중요한 얼굴 모양 구분 능력을 보여주면서 국소적으로 발생하는 잡음에 견고함을 보여준다. 우리는 제안된 방법의 성능을 평가하기 위해 밝기, 표정, 연령, 머리 포즈가 변화하는 성별 인식 데이터 셋에 다양한 실험을 실험 했고 기존 방법 보다 제안된 방법의 성능이 우수함을 입증했다.

Local Prominent Directional Pattern for Gender Recognition of Facial Photographs and Sketches

Farkhod Makhmudkhujaev*, Oksam Chae**

ABSTRACT

In this paper, we present a novel local descriptor, Local Prominent Directional Pattern (LPDP), to represent the description of facial images for gender recognition purpose. To achieve a clearly discriminative representation of local shape, presented method encodes a target pixel with the prominent directional variations in local structure from an analysis of statistics encompassed in the histogram of such directional variations. Use of the statistical information comes from the observation that a local neighboring region, having an edge going through it, demonstrate similar gradient directions, and hence, the prominent accumulations, accumulated from such gradient directions provide a solid base to represent the shape of that local structure. Unlike the sole use of gradient direction of a target pixel in existing methods, our coding scheme selects prominent edge directions accumulated from more samples (e.g., surrounding neighboring pixels), which, in turn, minimizes the effect of noise by suppressing the noisy accumulations of single or fewer samples. In this way, the presented encoding strategy provides the more discriminative shape of local structures while ensuring robustness to subtle changes such as local noise. We conduct extensive experiments on gender recognition datasets containing a wide range of challenges such as illumination, expression, age, and pose variations as well as sketch images, and observe the better performance of LPDP descriptor against existing local descriptors.

Keywords : Local Histogram, Prominent Directions, Feature Extraction, Sketch Images, Gender Recognition

접수일(2019년 2월 17일), 게재확정일(2019년 6월 03일)

* 경희대학교/컴퓨터공학과(주저자)
** 경희대학교/컴퓨터공학과(교신저자)

1. Introduction

Gender recognition from face images is one of the dominant research areas of computer vision. Performance of many real-world applications highly relies on the proper estimation of human gender. The examples of such applications are human-computer interaction, visual surveillance, and biometric authentication. Similar to other face-related classification tasks, gender recognition is decomposed into several steps: a) face detection and registration that detects the face region and pre-processes it to reduce the effect of scale/orientation variations; b) feature extraction which focuses on generating discriminative and robust descriptions of the detected face regions; and c) classification performing a binary decision.

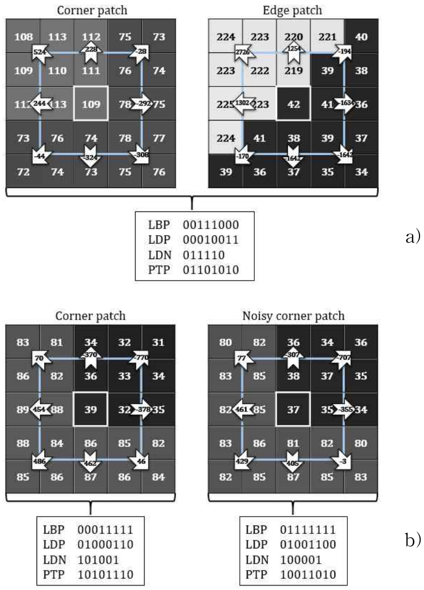
Among all these steps, feature extraction plays a vital role to achieve a favorable recognition performance. In particular, descriptors representing the face images should have distinctiveness among genders while being robust against challenges such as illumination and pose variations, expressions, occlusions, and resolutions, as pointed out in [1]. Therefore, generation of the discriminative and robust descriptor is still a challenging task of the gender recognition application.

In the past decades, we have witnessed numerous methods dealing with the description of facial features. Mainly, there are two common approaches representing face images: geometric-feature-based and appearance-based approaches. The geometric-feature-based methods [2, 3, 4] characterize facial structure by considering the shape and locations of different facial components. The methods of this category represent facial geometry well in the presence of accurate facial landmark alignment and tracking; however, due to the facial appearance variations and imaging conditions obtaining this information is difficult. Hence, geometric-feature-based methods may perform differently depending on conditions. In contrast, appearance-based methods describe facial appearance changes by performing the operation and/or transformations on the pixel level. Specifically, appearance-based methods attempt to describe facial changes in two ways: holistic (global) and local. Holistic methods capture global face information by applying projection techniques on whole face image. For instance, the method presented in [5] used principal component analysis (PCA), whereas [6] utilized more recent 2D PCA technique in their feature extraction step. Besides, [7] employed independent component analysis for gender recognition task. Nevertheless, the performance of holistic methods degrades significantly under the variations of pose and illumination [1].

In turn, local appearance-based methods aim to describe micro-level appearance changes by examining local neighborhood. To be precise, existing local appearance-based methods are divided into two approaches: texture-based and edge-based descriptors. One of the examples of texture-based descriptors is Local Binary Pattern (LBP) [8] which has notable properties such as invariance to monotonic illumination changes and computational efficiency. However, according to [9], LBP is sensitive to noise and uneven illumination variations. To improve these limitations, local edge-based descriptors have been introduced. In particular, these descriptors utilize the edge responses obtained by the convolution of Kirsch masks in eight directions centered at the target pixel and encode top k directional responses to describe local structures in the code representation.

Examples of such methods include Local Directional Pattern (LDP) [9], Local Directional Number Pattern (LDN) [10], Local Principal Texture Pattern (LPTP) [11], and Positional Ternary Pattern [12]. Despite their favorable performance, representing the local structures with more discriminability and robustness is still subject under the consideration. One key concern in this regard is these descriptors evaluate edge responses of a particular pixel, whereas edge information of surrounding pixels is not considered. That is, the local structure is being encoded based on information of a single pixel. Make use of such information in the code generation causes several issues. First of all, utilizing edge responses of a single pixel which signalize to the presence of changes in certain directions might not be sufficient for shape representation. In real scenarios, this issue may yield to less discriminability among differentiable structures [1]. Fig. 1(a) provides such example where existing local descriptors generate identical codes for corner and edge structures. Second of all, missing the edge information of neighborhood pixels make the codes sensitive to local subtle changes as single pixel edge responses are more prone to distortions. Specifically, this issue can be observed in Fig. 1(b), where local region containing noise variations corrupt the coding scheme of edge descriptors and lead to produce different codes for the same structures.

Besides the aforementioned descriptors, deep learning-based methods have also demonstrated promising performances in gender recognition [13, 14, 15, 16, 17, 18]. However, practical application of deep learning-based methods in some cases might face difficulties due to the following issues. The first one is a high volume of training data that is required to train

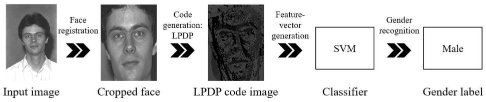


(그림 1) Shortcomings of existing local descriptors.

the network properly. In turn, expenses on collecting large face datasets might be high and raise issues on privacy protection. Based on considerations of [19], training on huge private datasets that comprise several millions of images make the obtained results of CNN's non-reproducible for the scientific community. Another issue is the high requirements on the computations and memory which create difficulties to use CNN on embedded platforms like smartphones. A more detailed discussion on this matter can be found in [19].

Recently, Local Prominent Directional Pattern (LPDP) [20] demonstrated better performance in recognizing facial expressions compared to aforementioned local descriptors. By means of encoded prominent directions that are extracted from the histogram of directional variations of the local region, LPDP descriptor has shown discriminative representations of local shapes which are also robust to noise and positional variations. However, data used in facial expression recognition has such explicit clues (i.e., distinguishable expression changes in the eye, forehead, and mouth areas) that are being described properly lead to better class separability. In contrast, gender recognition is not so straightforward as there are no such explicit and distinct variations that might cardinally improve the separation of classes, except for cases where a male has beard/mustache, and female make-up/groomed eyebrows.

In this work, we have analyzed the performance of the LPDP descriptor under the aforementioned circumstances of gender recognition as it is not studied yet in this direction. Hence, the ultimate goal of this paper is to provide performance analysis of LPDP descriptor on gender recognition. In an empiric manner, we study the efficacy of representing facial images by means of LPDP for recognizing gender classes. The performance of this descriptor is evaluated based on Support Vector Machine (SVM) that is found the most accurate classifier [1, 21] and hence became a popular choice in gender recognition [22, 23, 24]. We perform experiments with widely-utilized gender datasets such as FERET [25], LFW [26], and new dataset acquired in real scenarios UNISA-Public [27]. Besides aforementioned datasets containing facial photographs, we aim to investigate the problem of gender recognition on facial sketch datasets, CUFS and CUFSF [28], by means of LPDP descriptor as this topic is a great value for applications like a possible suspect and criminal detection. In addition, we perform another set of experiments on noisy images as surveillance systems might deal with these type of information. Through these extensive experiments, we demonstrate better capabilities of LPDP compared to existing local descriptors in terms of discriminative and robust representation of face images for more proper gender recognition on diverse/challenging facial images.



(그림 2) The workflow of the proposed method

2. Local Prominent Directional Pattern

Unlike the aforementioned descriptors that rely on single pixel information to encode the local shape at the target pixel, Local Prominent Directional Pattern (LPDP) utilizes statistical information of gradient directions calculated from the wider local neighborhood. The workflow of the proposed method is illustrated in Fig. 2. Make use of local statistics in the current context can be explained by considering the intuition behind the edges that are going through the local neighborhood. In detail, pixels on and near to edges have same and/or similar gradient direction characteristics where an accumulation of these gradient directions provide a more reliable base to generate discriminative and robust shape code at the target pixel. Herein, discriminative shape representation and robustness to local subtle changes (like noise) are achieved by allowing each pixel of the neighborhood to give a vote in the accumulation so that it could contribute to be part of the edge-like structure and demonstrate its strength.

2.1 The coding scheme

LPDP assigns a 6-bit binary code to each pixel of an input image. Given the image (I), LPDP starts the encoding process by calculating the gradient information of I. In fact, image gradients are effective to represent the strength and direction of edges existing in the local neighborhood as well as simple in the computation. Therefore, to extract gradient information of the image, LPDP utilizes well-known and straightforward Sobel operator. Specifically, the gradient information, i.e., magnitude (M) and direction (D) of I using Sobel operator are calculated as follows:

$$M=\sqrt{G_x^2+G_y^2},\ D=\tan^{-1}\left(\frac{G_y}{G_x}\right) \tag{1}$$

where G_x and G_y are results of convolving horizontal and vertical Sobel operators, respectively, to I:

$$G_x=I\times\begin{bmatrix}-1&0&1\\-2&0&2\\-1&0&1\end{bmatrix},\ G_y=I\times\begin{bmatrix}-1&-2&-1\\0&0&0\\1&2&1\end{bmatrix} \tag{2}$$

The existing local edge-based descriptors like LDP, LDN, and PTP while encoding the target pixel rely on edge information (strength and direction) of the single pixel without any consideration on information of other surrounding pixels. However, neighboring pixels may provide more solid information on the representation of edge-like structures. Especially, considering the fact that pixels on or near to the edges going through the local neighborhood have same and/or similar gradient direction characteristics, utilizing this information in the encoding process may improve code discriminability and robustness.

To achieve these properties, LPDP encodes the given pixel, $I_{(x,y)}$, based on a histogram of directional variations HDV that is calculated by accumulating the strength of gradient directions of all neighborhood pixels as:

$$HDV(i)=\sum_{p\in R}\Delta(D(p),i), \tag{3}$$

$$\Delta(a,b)=\begin{cases}\log_2(M(p)), & a=b\\ 0, & a\neq b\end{cases}$$

where HDV accumulates log of gradient magnitude $M(p)$ with the direction $D(p)$ of the pixel p in the 3×3 region R. Here, log operator is utilized as a damping function on gradient magnitudes due to the intensity fluctuations, such as noise, that may have dominance over accumulations of actual edge-like structures.

However, the presence of noise in some pixels of the region might still affect the accumulations of HDV. In such case, encoding all the histogram directions of HDV lead to produce noise corrupted code as noisy variations encoded along with significant edges that belong to actual local structure. In order to avoid such uncertain variations, and encode reliable gradient directions coming from actual edges, $LPDP$ selects the top k significantly accumulated directions of HDV as:

$$dir=\arg max_k\{HDV(i):0\leq i\leq q-1\} \tag{4}$$

where $\arg max_k$ is a function that extracts the indices of k significantly accumulated directions of HDV.

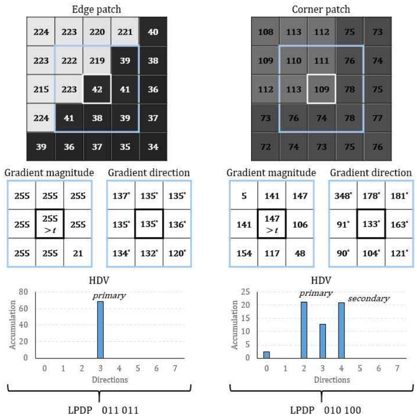
The gradient directions selected by (4) denote the prominent directions of edge(s) representing the local structure around the target pixel. Here, an important parameter of (4) is a number of k directions that are going to be encoded. LPDP utilizes set $k=2$ to pick up two significantly accumulated gradient directions to represent local structure. The main consideration of LPDP to use this number is that the face images contain mostly straight/curved edges and corners, which can be captured well by the two prominent directions. Besides, amount of junction-like structures (i.e., Y/T/X-type) are very low on the face images, and encoding more directions may introduce noise to the code representation. Thus, $LPDP$ makes use of $k=2$ and denotes two prominent directions as primary and secondary directions.

The image gradients demonstrate random variations in flat/smooth regions which yields to produce ambiguous representations. To distinguish such regions from meaningful edge regions, LPDP codes are generated only for the pixels which have gradient magnitude greater than a threshold t. As a consequence of thresholding, a default code is assigned to the pixels having gradient magnitudes lower than t. The threshold selection is an adaptive process as described in Section 2B. Conversely, LPDP calculates HDV from local neighborhood centered at the pixel that has gradient magnitude greater than t, and retrieves primary and secondary directions as follows:

$$LPDP_{(x,y)} = \begin{cases} 2^3 * dir^1 + dir^1, & HDV(dir^2) = 0 \\ 2^3 * dir^1 + dir^2, & HDV(dir^2) \neq 0 \\ Default\,code, & M_{(x,y)} < t \end{cases} \tag{5}$$

where dir^1 and dir^2 represent primary and secondary directions respectively.

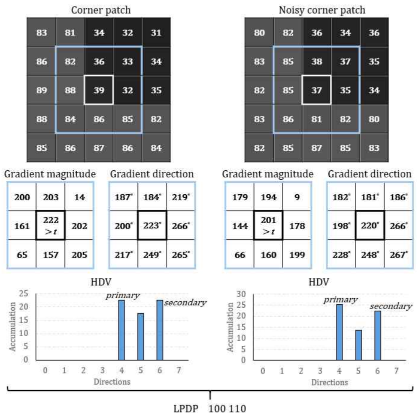
It should be noted that (4) may retrieve only one significant direction. Particularly, this case occurs in ideal edges where only one direction is accumulated in HDV. Fig. 3 demonstrates LPDP code generation on ideal edge patch as well as corner one. Although such a case provides solid evidence for the presence of an edge, they may occur rarely in the face images, and found in high contrast areas. Nevertheless, LPDP covers this type of edges, and while representing such case it repeats primary direction to avoid ambiguity among codes and maintain code uniqueness. Thus, LPDP becomes a six-bit code that ranges from 0 to 63. The default code represents pixels from featureless flat/smooth regions whereas its inclusion in the feature vector may also confuse the classifier. Therefore, LPDP discards the default code from feature vector construction.



(그림 3) Distinctive LPDP codes for edge and corner structures

Although LPDP has a similar number of codes as existing local descriptors, LPDP codes are more discriminative and robust to noise. The discriminability of LPDP codes can be seen on example patches shown in Fig. 4, where LPDP distinctively generates codes for two different local structures while existing ones generate the same code on these patches as shown in Fig. 1(a).

In addition, Fig. 4 demonstrates the robustness of LPDP that can stably generate the same code for noise-free and noisy corner patches without being corrupted by this local subtle distortion. Hence, this leads to consideration that utilizing gradient characteristics of a wider neighborhood and extraction of prominent directions based on local statistics provide more reliable and stable representations of actual local structures compared to existing descriptors that rely on either raw intensity or Kirsch mask responses of the single pixel.



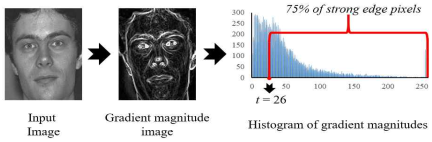
(그림 4) Robustness of LPDP codes against subtle local distortion like noise in the example of noise-free and noisy corner patches

2.2 Adaptive thresholding

The flat/smooth regions of the face image show little or even no facial appearance changes. Feature extraction on these regions lead to generate ambiguous codes which included into the feature vector may derail the classifier and further affect the performance. In order to avoid such cases, LPDP applies a thresholding mechanism that adaptively selects an appropriate threshold based on the characteristics of the image. This mechanism relies on observation in which a ratio of strong edge pixels in the face region remains similar in the images of different persons. To this purpose, LPDP utilizes histogram of gradient magnitudes H of the face image as gradient magnitude represents the strengths of facial appearance changes in each pixel whereas accumulation of them provides a base to explore a ratio of strong edge pixels that points to the proper threshold, t. For a given image, this threshold is calculated as follows:

$$t = \mathrm{arg}max_b \sum_{i=0}^{b} H(i) \le (1-\Phi)\sum_{i=0}^{255} H(i) \quad (6)$$

where Φ is the ratio of strong edge pixels with respect to the total number of pixels in a facial image. Similar concept as this was adopted in [29], where it is found that considering $0.75\% \pm 0.05$ as a ratio of strong pixels gives the best performance in facial analysis task. Example of this threshold calculation mechanism for a given image is demonstrated in Fig. 5, where t value is found as 26 for a given image assuming 75% of the pixels belong to strong edge pixels.



(그림 5) The threshold calculation steps

2.3 Face description using LPDP

After generating all the LPDP codes, the face image can be represented by LPDP histogram. However, a histogram computed over the whole face image only counts the occurrences of the edge-like structures existing in the face image without any consideration on their spatial locations. In fact, location information is very critical to represent the spatial relationship among facial features. The straightforward way to incorporate this information [9, 10, 11, 12] is to divide face image into N $(m \times n)$ number of uniform regions, $R_1, R_2, ..., R_N$ and then for each region calculate a histogram $H_{i \in [1,N]}$ as follows:

$$H_i(C) = \sum_{(x,y)\in R_i} \nabla(LDPD(x,y),c), \quad (7)$$

$$\nabla(a,b) = \begin{cases} 1, & a=b \\ 0, & a \ne b \end{cases}$$

Eventually, all the histograms are concatenated in order to generate a feature vector that represents whole face image using:

$$LPDP_H = \prod_{i=1}^{N} H_i \qquad (8)$$

3. Experimental Results

To verify the performance of LPDP descriptor, we experiment with images from two different modalities, namely facial photograph, and sketch images. The gender recognition of facial photographs is the traditional way in the literature which performed on datasets such as LFW, FERET, UNISA-Public. Fig. 6 illustrates some of the example images from all these datasets. Although the importance of recognizing genders from sketch images is high in the detection of possible suspects, this type of images is rarely considered. Thus, we consider the sketch images in our experiments which in turn should demonstrate the performance of LPDP in this direction of gender recognition. To this purpose, we perform tests on CUFS and CUFSF facial sketch datasets. Examples of sketch images from these datasets are shown in Fig. 7.

We start the experiments by cropping face regions of dataset images using Viola-Jones face detector, only, if dataset images are not cropped already. Afterward, we normalize the detected face images into 110 × 150 resolution. In order to generate face description by LPDP, we divide the coded-face image into 5 × 7 regions, to generate the final feature vector, according to Section 2.3. It should be noted that while evaluating some of the existing methods, we retain their parameters same as stated in the respective works. For the classification, we use Support Vector Machine (SVM) classifier with RBF kernel, as it is found the most accurate classifier [1, 21]. In order to select the optimal parameters of SVM, we follow the considerations of [30] and conduct a grid-search on the hyper-parameters based on which pick the parameters giving the best cross-validation results.

(그림 6) Example of facial photographs from UNISA-Public, LFW, and FERET respectively.

(그림 7) Example of sketch images from CUFS (left) and CUFSF (right) datasets.

3.1 FERET

FERET [25] dataset is one of the widely used datasets in gender recognition research. The dataset contains frontal faces acquired in controlled environments with different variations in illumination, background, expression, age, and race. In this work, we use only a single facial photograph per subject taken from fa gallery of FERET. Specifically, we use 994 images in our experiments and perform five-fold cross-validation as a testing scheme.

Apart from this testing, we experiment on GENDER-FERET dataset that was created from original FERET [25] and published in [31] and further used in [32, 33]. In GENDER-FERET, the number of facial photographs in male and female classes are balanced. A total of 946 images are pre-partitioned into 474 training (237 per class) and 472 test (236 per class) images. The face images of the person are divided in a way so that they are not presented in training and test sets at the same time [33]. Same as above, we perform five-fold cross-validation on this data. The purpose of this experiment is to evaluate the performance of gender recognition on the class-balanced dataset and make a fair comparison with methods like [31, 32, 33].

Table 1 and 2 present the performance comparison of LPDP with other methods on FERET and GENDER-FERET datasets respectively. It is noteworthy to mention that LPDP performs better than other methods in both tables, and hence demonstrates its efficiency on recognizing genders of facial photographs on FERET dataset. We also found that the result of LPDP on class-balanced GENDER-FERET is slightly better than FERET. Besides, LPDP achieves higher recognition accuracies against methods presented in [31] and [33] which utilize a combination of several features like raw intensities, LBP codes and HOG descriptions, and fusion of COSFIRE- and SURF-based classifiers respectively.

To demonstrate the robustness of LPDP descriptor under the noise, we perform experiments on FERET dataset with artificially added noise variations. We generate such noise-corrupted images by adding random Gaussian noise with zero mean and two random intervals of standard deviations (σ in 0.08 - 0.16 and 0.16 - 0.32) for each facial photographs of this dataset. Fig. 8 presents examples of noisy images. In this test, we compare LPDP against existing local descriptors such as LBP, LDP, LDN, LPTP, and PTP as our intention is to demonstrate better robustness of LPDP codes. We present the results in Table 3. As can be seen, LPDP achieves higher recognition accuracies than other existing local descriptors in both noise intervals. It is also observable that the performance of existing descriptors decreases quite gradually. This indicates the stability of LPDP under the different noise variations. In turn, such performance of LPDP can be explained based on its coding scheme which uses local statistics encompassed in the histogram of directional variations providing reliable information of neighborhood to retrieve stable prominent directions. Moreover, LPDP avoids encoding the pixels of flat regions by using adaptive thresholding which makes its description less sensitive to random variations.

<표 1> Performance comparison of LPDP with other methods on FERET dataset.

Method	Recognition(%)
LBP	91.69
ELBP	93.01
LDP	89.46
EnLDP	89.97
ELDP	86.32
LDN	90.57
POEM	90.98
CoLBP	93.92
PTP	94.08
LPTP	93.98
LTP 〔34〕	93.90
MQLBP 〔34〕	95.12
Face++ 〔31〕	89.6
Luxand 〔31〕	89.2
LPDP	**95.15**

Note: We acknowledge the help from [1] to generate the results of this table.

<표 2> Performance comparison of LPDP with other methods on GENDER-FERET dataset.

Method	Recognition(%)
RAW LBP HOG 〔31〕	92.6
COSFIRE 〔32〕	93.7
COSFIRE SURF 〔33〕	94.7
LPDP	**95.74**

$\sigma = 0$ $\sigma = 0.08 - 0.16$ $\sigma = 0.16 - 0.32$

(그림 8) Sample FERET image at different noise variations σ.

<표 3> Performance comparison of LPDP with other methods on different noise intervals of FERET dataset.

Method	Noise variations	
	0.08 - 0.16	0.16 - 0.32
LBP	82.3	73.1
LDP	71.9	60.2
LDN	82.5	75.4
LPTP	87.2	78.9
PTP	88.1	79.6
LPDP	**90.4**	**84.0**

Note: These results are generated with 5-fold cross-validation testing scheme by ourselves.

3.2 LFW

Labeled Faces in the Wild (LFW) [26] dataset contains more than 13000 photographs of faces collected from 5749 subjects. It is one of the widely used datasets in the literature as it comprises real-world images captured under less controlled environments. The images of LFW were collected from the web, and thus include variations in resolution, illumination, pose, and expression. As pointed out in [1], facial texture representation is extremely challenging under such variations. Besides, LFW is an imbalanced dataset where it contains a quite significant number of images in the male class. In order to evaluate LPDP on this dataset, we perform a five-fold cross-validation testing scheme on 9641 images which are properly cropped and aligned.

We compare LPDP descriptor with a number of existing local descriptors including previously mentioned LBP, LDP, and LDN along with some other appearance-based methods utilized in combination with other approaches, e.g., Boosted LBP [22], Gabor with PCA [23], and HOG+LBP+GSS from [24]. Moreover, we compare LPDP against recent deep learning methods, such as LDNN-F [15], Local-DNN [17], and Landmark-Guided LDNN [18]. We present the comparative results in Table 4, where we visualize the better performance of LPDP descriptor against all the above methods. It is also noteworthy to mention that LPDP achieves the significant performance gain over LBP, and other local edge-based descriptors such LDP and LDN utilizing directional patterns. Remarkably, LPDP demonstrates such high recognition accuracy without going through a huge computational process and memory consumption as utilized in deep learning-based methods which indicates its efficiency in recognizing genders on real-world facial photographs without a necessity to utilize much computational effort.

<표 4> Performance comparison of LPDP with other methods on LFW dataset.

Method	Recognition(%)
LBP	83.96
ELBP	87.48
LDP	79.46
LDN	82.17
EnLDP	78.42
ELDP	82.74
POEM	87.57
CoLBP	89.70
MBP	89.82
LDNN-F [15]	95.64
DNN [17]	92.60
DCNN [17]	94.09
Local-DNN [17]	96.25
Landmark-Guided LDNN [18]	96.02
Boosted LBP [22]	94.81
Gabor+PCA+SVM [23]	94.01
HOG+LBP+GSS [24]	95.76
COSFIRE [32]	90.0
LPDP	**96.30**

Note: We acknowledge the help from [1] to generate the results of this table.

3.3 UNISA-Public

Recently, authors of [27, 33] presented a new dataset called UNISA-Public for evaluating the performance of gender recognition algorithms. This dataset contains 12 video sequences acquired in the University of Salerno. It provides already detected and cropped faces of

58 subjects in their movement toward a fixed camera. Overall, the dataset contains 406 faces captured from 42 males and 16 females in real environments. Images have different pose and expressions, and sometimes motion blur due to the sudden movements of subjects. According to [35], illumination conditions are quite controlled since videos recorded in an indoor environment close to the entrance door of a building. The dataset comes with training and test sets. In order to preserve a prior gender distribution of the whole dataset, both sets contain images of 29 subjects, i.e., 21 men and 8 women per set. Some of the example images of UNISA-Public is shown in Fig. 7. Since this dataset is released recently, we compared LPDP descriptor with the results of [33]. In particular, we compare LPDP with COSFIRE-based, SURF-based features as well as with results of stacked classification and fusion techniques such as majority and weighted voting. Table 5 provides performance comparison results on this dataset.It can be observed that LPDP descriptor performs well compared to all other methods except stacked classification scheme of [33] where this scheme fuses COSFIRE- and SURF-based classifiers. However, it is noteworthy to mention that descriptor achieves such result based on own features solely without combining other different information. This indicates to the better encoding scheme of LPDP that relies on prominent directions retrieved from the histogram of directional variations.

3.4 Sketch dataset

Gender recognition on sketch images has received less attention in the literature [1]. However, a study in this direction has a great impact on applications detecting possible suspects and criminals. Therefore, to demonstrate the efficacy of LPDP descriptor on recognizing genders from facial sketch images, we perform experiments on CUHK face sketch (CUFS) and CUHK face sketch FERET (CUFSF) datasets, both released by Chinese University of Hong Kong. The former dataset contains 606 facial sketches collected from CUHK, ARand XM2VTS databases, whereas latter one contains 1195 sketch images from FERET dataset. However, there are some duplicated images in CUFSF which we omit in our experiments. According to [1], sketches were drawn by different artists in both datasets, and there are shape exaggerations.

<표 5> Performance comparison of LPDP with other methods on UNISA-Public dataset.

Method	Recognition(%)
COSFIRE	89.9
SURF	82.9
Majority voting	90.9
Weighted voting	90.9
Stacked classification	**91.5**
LPDP	**91.5**

Note: We acknowledge the help from [32] to generate the results of this table.

<표 6> Performance comparison of LPDP with other methods on CUFS and CUFSF datasets.

Method	Recognition (%)	
	CUFS	CUFSF
LBP	87.62	92.50
ELBP	89.60	92.00
TPLBP	87.79	88.96
LDP	88.61	88.86
EnLDP	87.29	92.00
ELDP	85.47	90.27
LDN	86.80	91.19
POEM	90.10	92.91
LGBPHS	89.76	93.52
MBP	91.09	94.12
CoLBP	91.25	94.53
LPDP	**93.30**	**94.62**

Note: We acknowledge the help from [1] to generate the results of this table.

We follow the testing scheme of [1] and perform five-fold cross-validation on both datasets. We compare LPDP against local descriptors like LBP and its variants as done by [1] since these methods can efficiently represent facial characteristics of sketch images. Table 6 provides a performance comparison for both datasets. Similar to [1], we found that the performance of LPDP descriptor on CUFSF dataset is higher than CUFS. Nevertheless, LPDP achieves higher recognition accuracies compared to all other descriptors in both datasets. Such results advocate efficiency of LPDP on representing facial sketches accurately which contribute to recognizing genders properly.

4. Conclusion

Local descriptors have received much attention in the past decades, and based on this, a number of facial related applications including gender recognition achieved significant performances. However, existing local descriptors due to utilizing raw intensities and/or edge responses of pixels from smaller local neighborhood limit their discrimination capability and robustness in various conditions. In particular, these issues yield to the poor performance of gender recognition methods in the images containing challenges such as variations in, noise, illumination, pose, expression, and age.

To tackle these major problems limiting the performance of existing local descriptors, we present Local Prominent Directional Pattern (LPDP) to represent facial images for gender recognition in this work. Compared to other local descriptors, LPDP relies on local statistics encompassed in the histogram of directional variations providing a reliable base to extract prominent directions for local shape representation. In order to evaluate the performance of this coding scheme, we conducted gender recognition experiments on facial photographs collected in several publicly available datasets such as FERET, LFW, and UNISA-Public. These datasets allowed us to analyze the capability of LPDP in depth as they contain images of controlled environments as well as real-world images that are full of aforementioned challenges. Besides, we performed experiments on noisy images to demonstrate the robustness of LPDP descriptor under such conditions that may occur while dealing with video inputs of surveillance systems. It is noteworthy to mention that in all of these extensive experiments, LPDP achieved better recognition accuracies compared to other existing methods including local descriptors. In addition to the experiments on facial photographs, we consider sketch images in our evaluations of LPDP descriptor. Performed experiments on two sketch datasets, namely, CUFS and CUFSF, have shown the better capability of LPDP to represent facial sketches for gender recognition owing to higher accuracies compared to other local descriptors, under our considerations. Accordingly, such results of LPDP descriptor advocates for its the face description demonstrating improvements on recognizing genders in both facial photographs and sketch images.

참고문헌

[1] Patel, B., Maheshwari, R.P., and Raman, B. “Compass local binary patterns for gender recognition of facial photographs and sketches,” Neurocomputing, 218, pp.203-215, 2016.

[1] Hong, Hai, Hartmut Neven, and Christoph Von der Malsburg. “Online facial expression recognition based on personalized galleries,” In Proceedings of Third

IEEE International Conference on Automatic Face and Gesture Recognition, pp. 354-359. 1998.

[2] Kotsia, I. and Pitas, I. "Facial expression recognition in image sequences using geometric deformation features and support vector machines," IEEE transactions on image processing, 16(1), pp.172-187, 2007.

[3] Bourbakis, N., Esposito, A. and Kavraki, D. "Extracting and associating meta-features for understanding people's emotional behaviour: Face and speech," Cognitive Computation, 3(3), pp.436-448, 2011.

[4] Turk, M.A. and Pentland, A.P. "Face recognition using eigenfaces," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.586-591. 1991, June.

[5] Yang, J., Zhang, D., Frangi, A.F. and Yang, J.Y. "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," IEEE transactions on pattern analysis and machine intelligence, 26(1), pp.131-137, 2004.

[6] Bartlett, M.S., Movellan, J.R. and Sejnowski, T.J. "Face recognition by independent component analysis," IEEE transactions on neural networks, 13(6), p.1450-1464, 2002.

[7] Ojala, T., Pietikäinen, M. and Harwood, D. "A comparative study of texture measures with classification based on featured distributions," Pattern recognition, 29(1), pp.51-59, 1996.

[8] Jabid, T., Kabir, M.H. and Chae, O. "Gender classification using local directional pattern (LDP)," In 2010 20th International Conference on Pattern Recognition, pp. 2162-2165. 2010, August.

[9] Rivera, A.R., Castillo, J.R. and Chae, O.O. "Local directional number pattern for face analysis: Face and expression recognition," IEEE transactions on image processing, 22(5), pp.1740-1752, 2013.

[10] Rivera, A.R., Castillo, J.A.R. and Chae, O. "Recognition of face expressions using local principal texture pattern," In 2012 19th IEEE International Conference on Image Processing, pp. 2609-2612. 2012, September.

[11] Iqbal, M.T.B., Ryu, B., Song, G. and Chae, O. "Positional Ternary Pattern (PTP): An edge based image descriptor for human age recognition," In IEEE International Conference on Consumer Electronics, pp.289-292. 2016, January.

[12] Levi, G. and Hassner, T. "Age and gender classification using convolutional neural networks," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 34-42, 2015.

[13] Duan, M., Li, K., Yang, C. and Li, K. "A hybrid deep learning CNN - ELM for age and gender classification," Neurocomputing, 275, pp.448-461. 2018.

[14] Liao, Z., Petridis, S. and Pantic, M. "Local Deep Neural Networks for Age and Gender Classification," arXiv preprint arXiv:1703.08497. 2017.

[15] Rodríguez, P., Cucurull, G., Gonfaus, J.M., Roca, F.X. and Gonzalez, J. "Age and gender recognition in the wild with deep attention," Pattern Recognition, 72, pp.563-571. 2017.

[16] Mansanet, J., Albiol, A. and Paredes, R. "Local deep neural networks for gender recognition," Pattern Recognition Letters, 70, pp.80-86. 2016.

[17] Zhang, Y. and Xu, T. "Landmark-Guided Local Deep Neural Networks for Age and Gender Classification," Journal of Sensors, 2018.

[18] Antipov, G., Berrani, S.A. and Dugelay, J.L. "Minimalistic CNN-based ensemble model for gender prediction from face images," Pattern recognition letters, 70, pp. 59-65. 2016.

[19] F. Makhmudkhujaev, M. Abdullah-Al-Wadud, M.T.B. Iqbal et al., "Facial expression recognition with local prominent directional pattern," Signal Processing: Image Communication, 74, pp. 1-12. 2019.

[20] Makinen, E. and Raisamo, R. "Evaluation of gender classification methods with automatically detected and aligned faces," IEEE transactions on pattern analysis and machine intelligence, 30(3), pp.541-547, 2008.

[21] Shan, C. "Learning local binary patterns for gender classification on real-world face images," Pattern recognition letters, 33(4), pp.431-437, 2012.

[22] Dago-Casas, P., González-Jiménez, D., Yu, L.L. and Alba-Castro, J.L. "Single-and cross-database benchmarks for gender classification under unconstrained settings," In 2011 IEEE International Conference on Computer Vision workshops, pp.2152-2159, 2011, November.

[23] Liu, H., Gao, Y. and Wang, C. "Gender identification in unconstrained scenarios using Self-Similarity of Gradients features," In ICIP, pp.5911-5915, 2014, October.

[24] Phillips, P.J., Moon, H., Rizvi, S.A. and Rauss, P.J. "The FERET evaluation methodology for face-recognition algorithms," IEEE Transactions on pattern analy-

sis and machine intelligence, 22(10), pp.1090-1104, 2000.
[25] G.B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, “Labeled faces in the wild: a database for studying face recognition in unconstrained environments,” Technical Report 07-49, University of Massachusetts, Amherst, 2007.
[26] Azzopardi, G., Greco, A., Saggese, A. and Vento, M. “Fast gender recognition in videos using a novel descriptor based on the gradient magnitudes of facial landmarks,” In 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 1-6. 2017, August.
[27] Wang, X. and Tang, X. “Face photo-sketch synthesis and recognition.” IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(11), pp.1955-1967, 2009.
[28] Iqbal, M.T.B., Abdullah-Al-Wadud, M., Ryu, B., Makhmudkhujaev, F. and Chae, O. “Facial Expression Recognition with Neighborhood-aware Edge Directional Pattern (NEDP),” IEEE Transactions on Affective Computing. 2018.
[29] Hsu, C.W. and Lin, C.J. “A comparison of methods for multiclass support vector machines,” IEEE transactions on Neural Networks, 13(2), pp.415-425, 2002.
[30] Azzopardi, G., Greco, A. and Vento, M. “Gender recognition from face images using a fusion of svm classifiers,” In International Conference Image Analysis and Recognition, pp.533-538, Springer, Cham, 2016, July.
[31] Azzopardi, G., Greco, A. and Vento, M. “Gender recognition from face images with trainable COSFIRE filters,” In Proceedings of 13th IEEE International Conference on Advanced Video and Signal Based Surveillance, pp.235-241, 2016, August.
[32] Azzopardi, G., Greco, A., Saggese, A. and Vento, M. “Fusion of domain-specific and trainable features for gender recognition from face images,” IEEE Access, 6, pp.24171-24183, 2018.
[33] Patel, B., Maheshwari, R.P. and Balasubramanian, R. “Multi-quantized local binary patterns for facial gender classification,” Computers & Electrical Engineering, 54, pp.271-284. 2016.

〔저자소개〕

Farkhod Makhmudkhujaev
2008년 9월 TUIT, Uzbekistan
컴퓨터과학과(학사)
2012년 9월 TUIT, Uzbekistan
컴퓨터과학과 (석사)
2014년 9월~현재 경희대학교
컴퓨터공학과
박사과정
email : farhodfm@khu.ac.kr

채 옥 삼 (Oksam Chae)
1986년 Oklahoma State University, Stillwater (공학박사)
1986년 Texas Instruments Image Processing Laboratory, Dallas, TX (연구 엔지니어)
1988년~현재 경희대학교
컴퓨터공학과 교수
email : oschae@khu.ac.kr