Person-Independent Facial Expression Recognition with Histograms of Prominent Edge Directions

Farkhod Makhmudkhujaev, Md Tauhid Bin Iqbal, Md Rifat Arefin, Byungyong Ryu and Oksam Chae*
Department of Computer Science and Engineering, Kyung Hee University,
Yongin-si, Gyeonggi-do 17104, Republic of Korea
[e-mail: farhodfm@khu.ac.kr, tauhidiq@khu.ac.kr, rifat.arefin@khu.ac.kr,
read100nm@khu.ac.kr, oschae@khu.ac.kr]
*Corresponding author: Oksam Chae

Received April 24, 2018; revised June 19, 2018; accepted July 20, 2018; published December 31, 2018

Abstract

This paper presents a new descriptor, named Histograms of Prominent Edge Directions (HPED), for the recognition of facial expressions in a person-independent environment. In this paper, we raise the issue of sampling error in generating the code-histogram from spatial regions of the face image, as observed in the existing descriptors. HPED describes facial appearance changes based on the statistical distribution of the top two prominent edge directions (i.e., primary and secondary direction) captured over small spatial regions of the face. Compared to existing descriptors, HPED uses a smaller number of code-bins to describe the spatial regions, which helps avoid sampling error despite having fewer samples while preserving the valuable spatial information. In contrast to the existing Histogram of Oriented Gradients (HOG) that uses the histogram of the primary edge direction (i.e., gradient orientation) only, we additionally consider the histogram of the secondary edge direction, which provides more meaningful shape information related to the local texture. Experiments on popular facial expression datasets demonstrate the superior performance of the proposed HPED against existing descriptors in a person-independent environment.

Keywords: HPED, automatic facial expression recognition, feature extraction, person-independent, sampling error

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program (IITP-2018-2015-0-00742) supervised by the IITP (Institute for Information & Communications Technology Promotion), and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2015R1A2A2A01006412).

http://doi.org/10.3837/tiis.2018.12.021 **ISSN : 1976-7277**

1. Introduction

Understanding human behavioral characteristics is one of the integral parts of biometric research, especially in biometric authentication, affective computing, and human-computer interaction (HCI) research [1]. One of the better ways of understanding behavioral characteristics is the analysis of facial expressions, since a person's expression to a specific change reflects his behavioral traits. However, facial features change differently in different expressions, and hence, a robust description of such changes is vital in recognizing facial expressions. Therefore, recognition of facial expressions largely depends on the representation of face images, which in turn should robustly describe facial appearance changes against pose, noise and illumination variations.

Existing methods for the description of the human face can mainly be categorized into two broad categories: geometric feature-based approaches and appearance feature-based approaches. The geometric-feature-based approaches [2], [3] represent facial geometry using the shape and location of the facial components. Hence, these methods become dependent on accurate and reliable facial landmark detection, where the performance of the recognition may degrade in the case of incorrectly detected landmarks [1], [4]. On the contrary, appearance-based methods represent the appearance changes of the facial image, which can be further classified into two categories: global (holistic) and local. The global appearance-based methods represent the face globally by applying various techniques such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Linear Discriminant Analysis (LDA) [5], [6]. However, such global representation of the face may not be robust in the presence of micro-level appearance changes or pose and illumination variations, as indicated in [4]. The local appearance-based methods represent micro-level feature information by describing the appearance changes from the local region. This approach can also be categorized into two more groups: texture-based and edge-based methods. Among the texture-based methods, the Local Binary Pattern (LBP) [1] is most often used due to its computational efficiency and robustness to monotonic illumination changes.

On the contrary, the edge-based methods, such as the Local Directional Pattern (LDP) [7], Local Directional Number Pattern (LDN) [4], Local Principal Texture Pattern (LPTP) [8] and Positional Ternary Pattern (PTP) [9] utilize the local edge directional information to represent local texture. Roughly speaking, these methods apply eight-directional Kirsch compass masks [10] in the local neighborhood of a pixel and select the positions of the top few edge responses as major directions of local edge-shape. These major directions efficiently represent the shape of the local texture-primitives, such as edges and corners [11], which are crucial in representing expression changes. Apart from the above descriptors that generate code for each pixel, the well-known Histogram of Oriented Gradients (HOG) [12] represents the texture variation with the histogram of the primary edge direction (i.e., gradient orientation) over small spatial regions. Information from such regions provides significant spatial information related to the face. Moreover, HOG uses quantized orientation-bins, which decreases the possibility of generating useless empty bins in the code-histogram and avoids sampling errors.

1.1 Problem Statement

One major aspect of the above local descriptors, including LBP, LDP, LDN, LPTP, and PTP is the generation of code histogram, where the face image is divided into $p \times q$ uniform

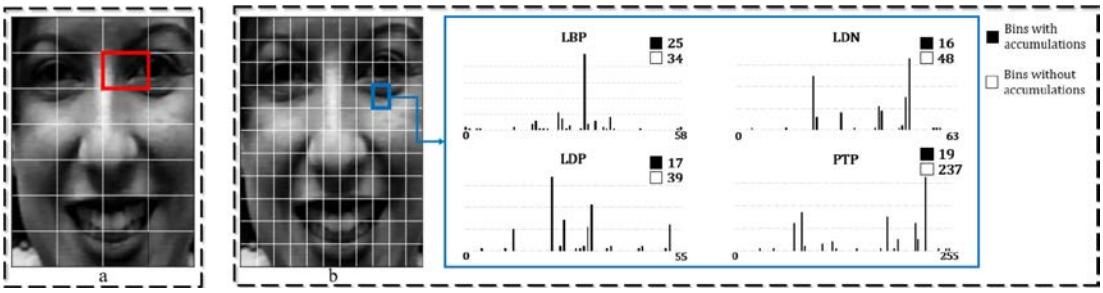

Fig. 1. (a) Face with larger grid-size. The red-marked grid contains multiple face-components (part of eye and nose-bridge). (b) Face with smaller grid-size, where local descriptors produce a number of empty bins for each grid (see the blue-marked grid).

regions (grids). Then, codes generated from all these regions are concatenated to form the final feature-vector. However, selecting an optimal grid-size against the wide number of code bins for the descriptors is always a critical issue. Maintaining a smaller grid-size preserves more spatial information from the face image. However, since small grids do not possess enough samples (i.e., pixels), code-histograms having a wide number of code-bins are prone to sampling error and may not provide the correct interpretation of that region. On the contrary, a grid-size that has fewer grids may avoid sampling error. But, this loses important spatial information due to its global nature. **Fig. 1 (a)** shows that a face with comparatively larger grid-sizes may include different face components in one grid (as in grid indicated in red). Thus, it loses the individual spatial information of each of the face components. **Fig. 1 (b)** shows the same face image with small grids, which, in contrast, do not include multiple face components in one grid, and resulting in more detailed spatial information. However, the local descriptors use a wide number of bins, but the small region does not possess enough pixels, resulting in sampling error. Moreover, the code-histograms from this small region end up with a large number of empty bins having no meaningful information, as shown in the blue-marked grid of that image. This affects the recognition performance. Therefore, preserving spatial information while avoiding sampling error is one of the key issues in representing expression-based changes in the face.

An apparent solution to the above problem is to generate a histogram within small grids with a fewer code-bins. Smaller grids will preserve the important spatial information, whereas fewer code-bins will restrict the generation of empty bins, reducing the sampling error. Applying HOG can be a possible solution in this regard, as it has fewer quantized orientation-bins within smaller spatial grids. HOG considers the primary edge-direction (gradient orientation) of the pixel, which is similar to using the top Kirsch edge-direction, as in the existing edge-descriptors. However, the primary edge direction may not describe the discernible expression-affiliated textures unambiguously. For instance, in **Fig. 2**, the primary direction appears in the same position (2) in both the edge and corner patches, and it does not provide sufficient evidence to differentiate the patches. Nevertheless, the direction with the second top response (secondary direction) appears differently in these patches, providing a significant cue to differentiate the patches. Therefore, encoding the secondary direction along with the primary direction may provide more meaningful information and represent expression-related features more efficiently. Note that in the above examples, we have only considered the edge-descriptors that use Kirsch compass mask responses to generate their codes. Recently, Ryu et al. [13] proposed a Local Directional Ternary Pattern (LDTP) [13] that utilizes the top absolute responses from the symmetric four-directional Robinson compass

masks [14]. Because the absolute responses are used, LDTP misses the sign information, which is then addressed with an additional ternary pattern. On the contrary, the top positive responses from Kirsch masks, as in **Fig. 2**, directly inherit the sign information and successfully represent the texture-structure without the inclusion of any additional structural pattern. Therefore, we consider the responses of Kirsch masks in this work.

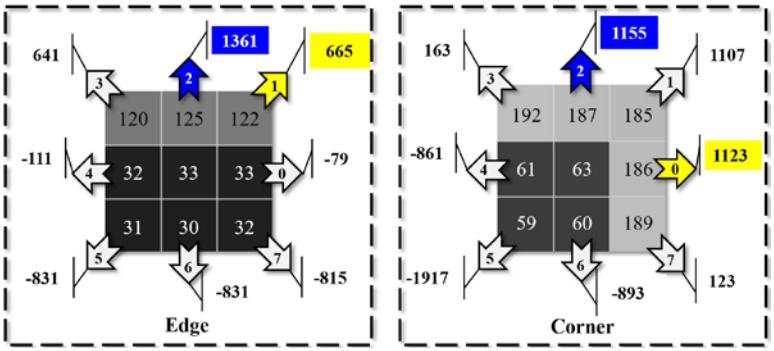

Fig. 2. Sample edge and corner patch with the corresponding Kirsch responses. Blue and yellow colors denote the position of primary and secondary direction, respectively.

1.2 Our Proposal

In this paper, we propose a novel descriptor, named *Histograms of Prominent Edge Directions (HPED)* to address the previously mentioned shortcomings of the existing descriptors in representing the face image. In contrast to the existing HOG that constructs the histogram of the primary edge direction only, we additionally consider a histogram of the secondary edge direction to achieve more meaningful information related to local shape. In the coding, we represent primary and secondary edge directions with the top two Kirsch directional responses, and construct histograms of both primary and secondary directions within smaller grids followed by concatenating all the histograms together to generate the final feature-vector. Top Kirsch directions can be any of the eight neighboring directions, and hence only 8-bin code histograms for each of the primary and secondary directions can be generated for each spatial grid of the face. This is advantageous in two ways. First, compared to the existing descriptors that suffer from sampling error for having a wide number of code-bins within limited samples, the proposed method generates a histogram in its eight code-bin only, avoiding the sampling error in the presence of limited samples. Second, this strategy reduces the possibility of encountering useless empty bins in the code-histogram, restricting the meaningless information in the feature description. We conducted experiments on well-known facial expression datasets with HPED under a person-independent environment, where HPED is found to achieve better performance than other existing descriptors. Moreover, we test the performance of HPED in the presence of noise and positional variations (as well as low-resolution, where we observe better accuracies of HPED), showing its overall efficacy in the recognition of facial expressions.

2. Methodology

The proposed HPED describes the facial appearance changes by the statistical distribution of the top two prominent edge directions from the spatial regions of the face image. In existing facial analysis research, the face is divided into several regions to represent the spatial information of the face image, and then code-histograms are generated from each of these regions. In this way, the existing descriptors may suffer from sampling error due to having fewer samples. Therefore, a meaningful description of the face image is desired where spatial information will be preserved while limiting the sampling error. In our approach, we ensure such information by generating separate histograms for the top two edge directions over smaller spatial grids. Therefore, each histogram will have a maximum of eight bins, which avoids sampling error despite having insufficient samples, and preserves the valuable spatial information through the use of smaller spatial grids.

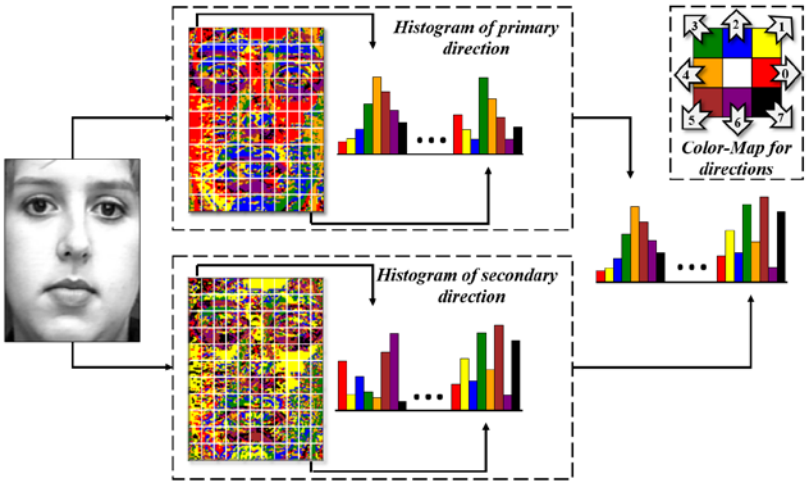

Fig. 3. HPED code generation. Primary and secondary histograms for each of the regions are generated, and then they are concatenated to form the final histogram. Eight different positions of the prominent directions are shown in different colors.

We start computing the coding by applying eight directional Kirsch masks [10] to the pixels of an input image. The Kirsch mask rotates in a 45° direction and generates edge responses in eight different directions. Let us assume that $I(x, y)$ is a pixel value in a two-dimensional image space. We apply eight Kirsch masks on it as

$$KR_i = M_i * I(x, y), \qquad 0 \le i < 7 \tag{1}$$

where, KR_i is the i^{th} response value of the M_i Kirsch mask. However, not all the responses among these eight are significant in representing the appropriate shape structure. According to [9], [11] the top two responses contain shapes related to the most significant structural information. Therefore, we extract the position of two highest responses among the eight responses,

$$D_j^k = \arg\max_2 \{KR_i : 0 \le i < 7\} \tag{2}$$

where, D_j holds the j^{th} highest response value, where $j \in \{1, 2\}$ and k is the corresponding neighbor position. We call the positions of these highest two responses "primary" and "secondary" directions, respectively.

We now calculate primary and secondary directions for each pixel of an input image in the above way. To generate the feature vector of that image, we divide the face image into $p \times q (=z)$ regions. For each region, we generate two different histograms, one having the code-bins of the primary direction and the other the code-bins of the secondary direction. Since both primary and secondary directions appear in any of the eight neighboring pixels, the total code-bin of each histogram is eight. Histogram generation for each of the regions is formally defined as,

$$H_z^j(k) = \sum_{m \in R_z} \nabla(m,k), \quad \nabla(m,k) = \begin{cases} v, & m = k \\ 0, & m \neq k \end{cases} \tag{3}$$

where, the k is the neighbor position (code-bin) ($k \in (0, 1, 2, ..., 7)$) of the j^{th} prominent direction histogram, H_z^j, for a region, R_z in which m is a pixel that contains a possible code (direction). Note that the weight of the code bin is denoted by v. Selecting v is crucial in our method since different weighting strategies will provide different distributions. In the face image, the high-textured areas show the dominance of top m-significant edge-directions, whereas flat regions (i.e., chicks) show fewer directional variations. Therefore, the weighting should be considered in such a way that the accumulation of these less directional variations must not dominate the significant directional changes of the textured area. For this purpose, we employ a damping function on the primary and secondary edge responses, and the output of the function is used as the weight, v, for that respective directional bin. We use different types of functions like

$$\begin{aligned} &v = D_j, && v = D_j^2, && v = \sqrt{D_j}, \\ &v = \log_e D_j, && v = \frac{D_j - \arg\min\{\forall D_j\}}{\arg\max\{\forall D_j\} - \arg\min\{\forall D_j\}}, && v = 1. \end{aligned} \tag{4}$$

Here, different damping functions are shown, such as the edge response itself, its square, square root, the response of a logarithmic function, min-max normalization, and the simple occurrence of direction, respectively. The effect of each of the functions is shown in the following section. However, after generating two separate histograms for primary and secondary directions for all the regions, we combine all these histograms into one large histogram,

$$H = \underset{j=1}{\overset{j=2}{\Omega}} \; \underset{z=1}{\overset{z}{\Omega}} H_z^j \tag{5}$$

where Ω is the concatenation operator, which combines the histogram of each j^{th} direction to generate the final histogram, H. H is used as the final feature-vector for HPED. We illustrate an example of generating the HPED descriptor of an input image in **Fig. 3**.

3. Experimental Results and Analysis

In this section, we provide a comparative performance analysis of the proposed descriptor for person-independent facial expression recognition. In our approach, we conducted a *leave-one-subject-out cross-validation (LOSO)* testing scheme, where we omitted the expression images of one person from the training set and then used them for testing. The process was repeated for *N*-persons, and the average results are reported. Since no prior person information is included in the training stage, this scheme ensures person independence (PI) in our testing. We conducted our experiments in four existing expression datasets, including CK+ [15], FACES [16], BU-3DFE [17], and RaFD [18]. Sample expression-images for each dataset are presented in **Fig. 4**.

However, we started the experiments by cropping the dataset-images using either the ground-truth positions of the eyes and mouth, or manual selections. Afterward, we normalized the images to 110 × 150 resolution, as was also done in [1], [4], [7]. We now generated the HPED feature-vector for the images using the strategy described in Section 2. For classification purposes, we used the Support Vector Machine (SVM) classifier [19] with an RBF kernel, since SVM has been effective in classifying the expression classes in existing works [1], [4], [7]. However, to select the optimal parameters, we conducted a grid-search on the hyper-parameters with *leave-one-subject-out cross-validation* approach and picked the parameter values giving the best cross-validation results. We conducted our experiments using Visual Studio 2015 on a computer with an Intel core i5 @2.67GHz with 8GB RAM.

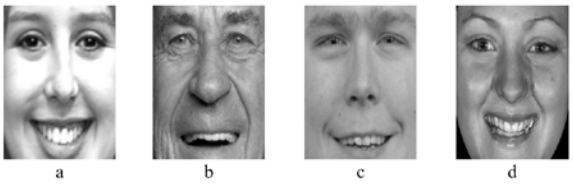

Fig. 4. Example of different facial expression images from several datasets: a) CK+, b) FACES, c) RaFD, and d) BU-3DFE.

3.1 Parameter Selection

There are several parameters for the proposed descriptor, including the histogram weight, v, as used in Eq. (3), and most importantly, the size of the grid, $p \times q$. Choosing smaller grids provides more spatial information, but it may suffer from sampling error. On the contrary, larger regions may avoid sampling error at the cost of losing valuable spatial information. Therefore, selection of the optimal grid-size is of great importance in HPED. Moreover, the purpose of the histogram weight, v, is to ensure the dominance of the bins with significant directional changes over the bins with fewer directional variations, which also plays a vital role in appropriately describing the local structural patterns.

To get the best parameter values for HPED, we followed the procedure mentioned in [7]; that is, we evaluated the recognition performance for HPED using a combination of different parameter values in 500 randomly collected images from the working datasets, and we chose the values giving the best performance. Results are shown in **Fig. 5**, where the best result is observed considering the 11 × 11 grid-size.

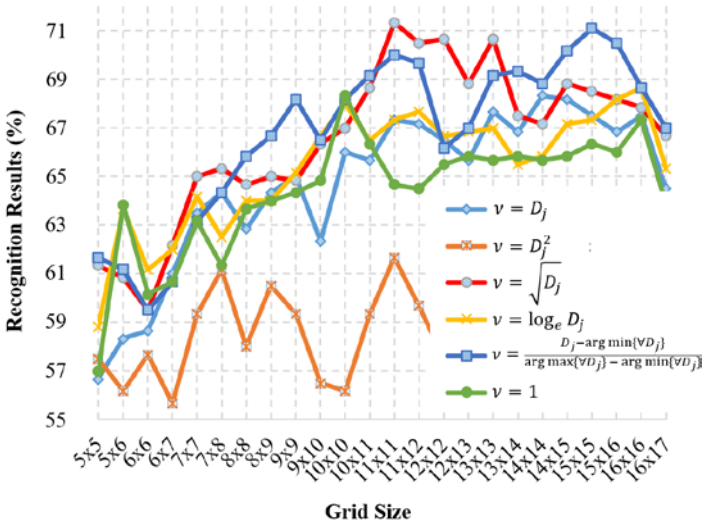

Fig. 5. Recognition results of HPED for different weighting schema, v, and grid sizes, $p \times q$ with 500 randomly collected images from working datasets.

As we see from the figure, decreasing the number of the grid decreases the performance for larger grids due to the loss of important spatial information. If we increase the number of grids beyond 11 × 11, more fine-level spatial information will be included, but the number of samples per grid will decrease at the same time, affecting the performance. Subsequently, we present the results for different weight functions in the same figure, where we observe the best performance when considering the square root of the response value. Therefore, we used 11 × 11 grid-sizes and square root of the response value as the weight, v, as the optimal parameters for HPED.

3.2 Recognition Results

CK+ Results

The Extended Cohn-Kanade (CK+) [15] dataset contains 593 image sequences, where 327 sequences with 123 subjects are labeled with one of the seven different expressions (anger, contempt, disgust, fear, happiness, sadness and surprise). For our experiments, we used the peak expression images of each labeled sequence, as done in prior works [1], [4], [7]. We conducted extensive experiments on CK+ for the given 7-expression classes to show the comparative efficacy of HPED over existing descriptors and other state-of-the-art methods. Specifically, we showed the performance of HPED under noise and positional variation. Moreover, we investigated the performance of HPED in low-resolution images. At the end, we compared the performance of HPED against other state-of-the-art results.

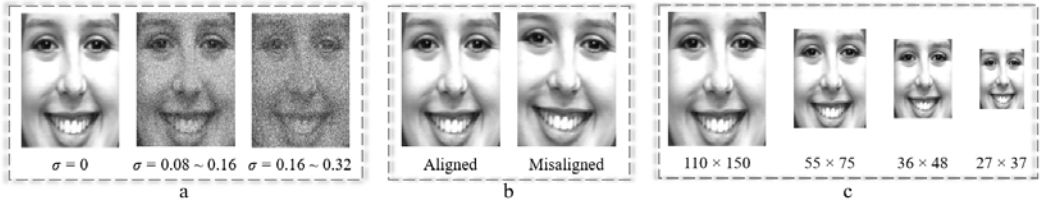

Fig. 6. Example of images with different: a) noise variation, b) positional variation (misalignment), and c) resolution.

Performance under noise: To show the robustness of the proposed descriptor against noise perturbations, we tested its performance in a CK+ dataset with artificially added noise. For this purpose, we added random Gaussian noise with zero mean and standard deviations σ with the interval of (0.08 - 0.16) and (0.16 - 0.32) for each image of the dataset. Noisy sample images are given in **Fig. 6 (a)**. We conducted a person-independent expression recognition for this noise-corrupted dataset and compared the results of HPED against other descriptors, including LBP [1], LDN [4], LDP [7], LPTP [8], PTP [9], HOG [12], and LDTP [13]. It is worth mentioning that we kept the parameters the same as stated in the respective works while generating the results for these descriptors. We present all the results in **Table 1**. Note that for the sake of comparison, we also provided results without adding noise in this table. For both of the cases, the proposed descriptor achieved higher results than other descriptors under consideration. It is important to observe that the performance gap between the proposed descriptor and other descriptors is quite large under noise. The reason for this is that HPED accumulates the histograms from a grid within a small number of bins, and the possibility of the code-value being distorted by the noise is lower compared to other descriptors, resulting in better accuracy.

Table 1. Person-independent expression recognition results on CK+ dataset by varying noise

Descriptors	Without Noise	Varying Noise	
		0.08 - 0.16	0.16 - 0.32
LBP	85.84	72.09	69.09
LDP	88.07	78.28	57.44
LDN	88.58	76.45	52.75
LPTP	91.64	87.05	77.37
PTP	91.03	88.96	82.16
HOG	92.01	88.89	84.51
LDTP	93.58	89.91	86.24
HPED	**93.74**	**91.12**	**86.93**

Performance under positional variation (misalignment): We evaluated the performance of HPED after distorting the alignment of the frontal face to test its robustness under a subtle registration error when detecting the frontal face. In practice, such positional variations of the frontal face are a very common issue that usually happens due to a face registration error (misalignment), where facial components (i.e., eye, nose, etc.) are not detected properly. Since the eye positions play an important role in cropping the face region, we add random noise with zero mean and standard deviations varying from 0.5 to 5 to the eye-positions at each expression image. This strategy artificially adds positional variations to the images. We utilized the ground-truth eye-position information of CK+ to perturb the position of the eyes, generating positional variations in the facial images. A sample position-varied (misaligned) image is provided in **Fig. 6 (b)**. We now carried out person-independent recognition in these images for the descriptors under consideration, and the results are shown in **Table 2**. As shown, HPED achieves better accuracy than all these descriptors, demonstrating its efficacy under such positional variations. The main reason for this better performance is that HPED incorporates detailed spatial information using the smaller grids. Therefore, the micro-level positional change information is included more efficiently in HPED than other descriptors that use larger grids, resulting in higher accuracy.

Table 2. Person-independent expression recognition results on CK+ having position-varied (misaligned) facial images

Descriptors	LBP	LDP	LDN	LPTP	PTP	HOG	LDTP	HPED
Results	84.20	86.95	87.36	90.01	89.91	90.13	91.84	**92.05**

Performance under low-resolution: In addition, we investigated the performance of HPED descriptors in different resolution images to test its comparative efficacy in such images, especially in the low-resolution images. Achieving better performance in low-resolution images is important since most surveillance systems and real-time video analysis systems deal with low-resolution video input. To test the performance, we generated four different sets of CK+ dataset images after varying the image resolution. We divided the images into 110 × 150, 55 × 75, 36 × 48, and 27 × 37 resolutions, respectively, to generate four different CK+ image-sets. **Fig. 6 (c)** provides examples of an image at four different resolutions. As with the previous tests, we tested HPED against other descriptors including LBP, LDP, LDN, LPTP, PTP, HOG, and LDTP. Results presented in **Table 3** show that the proposed HPED achieved higher accuracy than other descriptors, showing its efficiency at different image resolutions.

Table 3. Person-independent expression recognition results on a CK+ dataset by varying image resolution

Descriptors	**Varying Resolution**			
	110 × 150	55 × 75	36 × 48	27 × 37
LBP	85.84	83.59	82.16	81.04
LDP	88.07	86.54	85.73	81.75
LDN	88.58	87.56	85.93	84.10
LPTP	91.64	89.19	88.68	85.73
PTP	91.03	90.32	89.09	85.50
HOG	92.01	90.62	88.68	85.63
LDTP	93.58	91.74	89.48	88.03
HPED	**93.74**	**93.02**	**90.06**	**88.54**

Performance against state-of-the-art results: We compared the performance of the proposed descriptor on a CK+ dataset for 7 classes against other state-of-the-art methods. In particular, we compared the performance of HPED against different appearance-based methods and deep learning based methods. Among the appearance-based methods, we considered LBP, LDP, LDN, PTP, HOG, LPQ, Gabor, and LDTP. Furthermore, we compared against a method that uses a manifold based sparse representation (MSR) [20] and approaches dealing with intra-class variations [21], [22]. It is important to note that we compared our results against some recent deep methods, including GoogLeNet [23], AlexNet [23], and the proposed network in [23]. **Table 4** provides recognition accuracies of the previously mentioned methods against the proposed HPED on a CK+ dataset for 7 expression classes. Results presented in **Table 4** show that the HPED descriptor performs better than the given local descriptors, except for the recently published $LDTP_{active}$ [13] that utilizes selected active codes to improve the performance accuracy. Nevertheless, the proposed HPED achieves better accuracy than LDTP when no active code is used, which also indicates that the performance of the proposed method can be further improved with a similar feature-selection strategy. However, we consider this as a potential improvement, and leave it as a future endeavor. Note also that HPED considerably improves accuracies compared to the deep methods [23] without using additional training samples as done in the respective works [23]. This shows the efficacy of a proposed descriptor in achieving consistent performance without the use of such

additional data and computational effort.

In this regard, we also note that some of the other state-of-the-art methods utilize temporal information in classifying the expression classes [15], [24], [25], [26], [27]. Since the proposed method uses static images to classify expression classes, we do not compare our methods against such methods to maintain fairness in the comparison.

In addition, we provide the confusion matrix of HPED for 7-class CK+ images in **Fig. 7**. The confusion matrix shows that HPED achieves convincing recognition results for disgust, happy and surprise expression classes, where expressions like anger and contempt also show promising performance. However, fear and sadness show comparatively low recognition results, which perhaps happened due to some of the less distinctive images of these class.

Table 4. Person-independent expression recognition results on CK+ dataset
[Note: results with citations are from corresponding papers]

Methods	7-class results (%)
LBP	85.84
LDP	88.07
LDN	88.58
LPTP	91.64
PTP	91.03
HOG	92.01
LBP + SRC [21]	79.97
LPQ + SRC [21]	80.78
Gabor + SRC [21]	82.82
MSR [20]	91.4
SRC + ICV [22]	90.5
Lee et al. [21]	92.34
GoogLeNet [23]	85.71
AlexNet [23]	85.87
AlexNet + SVM [23]	86.83
Wu et al. [23]	89.84
LDTP [13]	93.58
$LDTP_{active}$ [13]	94.19
HPED	93.74

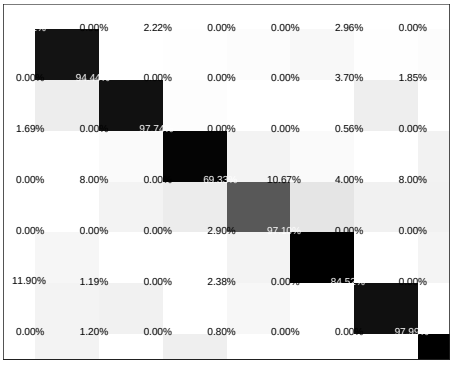

Fig. 7. Confusion matrix for CK+ dataset.

FACES Results

The FACES [16] dataset contains natural facial expressions of 171 subjects, including 58 young (19-31 years old), 56 middle-aged (39-55 years old) and 57 old subjects (69-80 years old). A total of 2052 images are assigned into one of the basic six expressions (anger, disgust, fear, happiness, sadness, and neutrality).

Table 5. Person-independent expression recognition results on FACES dataset

Methods	All ages	Young	Middle-age	Old
LBP	91.52	92.67	86.16	81.29
LDP	89.57	91.38	88.09	80.70
LDN	88.69	89.37	87.79	80.12
LPTP	92.11	90.37	88.39	81.58
PTP	91.42	93.10	89.58	84.36
HOG	93.03	94.83	89.73	86.69
LDTP	93.30	94.90	91.03	86.75
HPED	**93.62**	**95.55**	**91.82**	**87.28**

Expression recognition results for all the FACES images are presented in the first column of **Table 5**, which shows that the HPED descriptor achieves better performance against the existing local descriptors. It is worth mentioning that expression-images of this dataset are divided into different age-groups. Thus, the higher result of the proposed descriptor in FACES demonstrates its efficiency under the images with age variations. Moreover, Caroppo et al. [28] shows that the expression images of older subjects are hard to detect as their facial traits exhibit less differentiation among different expressions. Hence, we conducted separate experiments for the given three age-groups, including young, middle-aged and old, to explicitly evaluate the performance of HPED in different age-groups. We provided the results in **Table 5**, where we also observed better accuracy of HPED than other descriptors in all the respective age-groups, which strongly demonstrates the efficacy of HPED in recognizing expressions under such age variations.

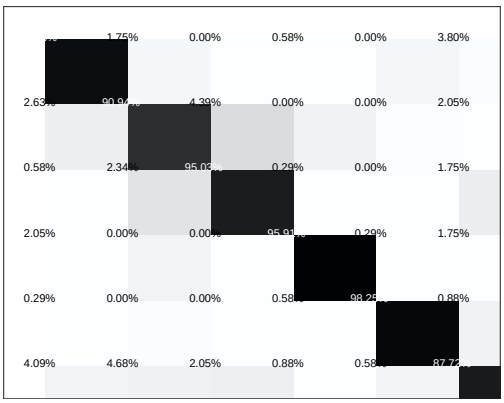

Fig. 8. Confusion matrix for FACES dataset.

Fig. 8 shows a 6-class confusion matrix for the FACES dataset when considering images of all the age-groups. Sad expressions were confused with neutral expressions. We observed that

some of the peak sad expressions showed fewer differences compared to the neutral image due to the different expression traits, which may contribute to this confusion.

BU-3DFE Results

The BU-3DFE [17] dataset consists of six prototype emotions depicted by 100 subjects where 56% of images are from females and 44% are from males. Images of BU-3DFE are labeled with one of these expression labels: anger, disgust, fear, happiness, sadness, and surprise. Since the dataset images vary in age, ethnic, racial ancestries, and intensities of expressions (i.e., 4 different intensity levels), this dataset is considered to be a challenging one.

The comparative performance is shown in **Table 6,** which demonstrates that variations of the BU-3DFE dataset play a significant role in feature extraction step, and therefore overall accuracies are much lower than the results of the above datasets. Nevertheless, the proposed HPED shows superior performance against other descriptors, which points to the efficacy of the proposed feature extraction scheme of HPED when applied to such a challenging dataset with large variations. Moreover, we provided the confusion matrix of HPED in **Fig. 9.** We observed that the fear class is most confused with all other expressions. The sad class is also confused with the other classes, especially with the anger class in BU-3DFE. One has to address these issues to achieve better performance, which we consider as a future endeavor.

Table 6. Person-independent expression recognition results on BU-3DFE dataset
[Note: results with citations are from corresponding papers]

Methods	6-class results (%)
LBP	56.2
LDP	61.3
LDN	56.5
LPTP	67.8
PTP	66.88
HOG	71.6
Hu et al. [29]	66.5
BDA/GMM [30]	68.28
Tariq et al. [31]	68.3
Moore et al. [32]	71.1
LDTP [13]	71.3
HPED	**73.39**

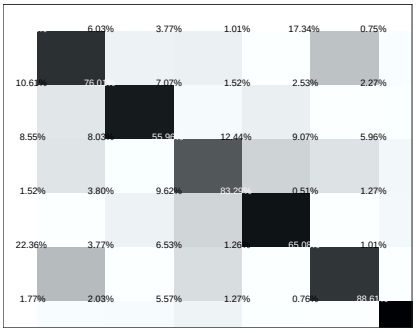

Fig. 9. Confusion matrix for BU-3DFE dataset.

RaFD Results

The Radboud Faces Databases (RaFD) [18] consists of eight expressions, including happy, sad, angry, surprise, disgust, fear, contemptuous and neutral. RaFD has images with 3 gaze directions and 5 face orientations. However, in our experiment, we used frontal face images with a frontal gaze direction.

Table 7. Person-independent expression recognition results on RaFD dataset
[Note: results with citations are from corresponding papers]

Methods	8-class results (%)
LBP	93.46
LDP	93.02
LDN	92.90
LPTP	92.53
PTP	92.90
HOG	94.15
Gabor [33]	89.78
Gabor + PCA [33]	88.80
LGBPHS [33]	94.84
LGBPHS + PCA [33]	92.56
LDTP	94.24
HPED	**95.52**

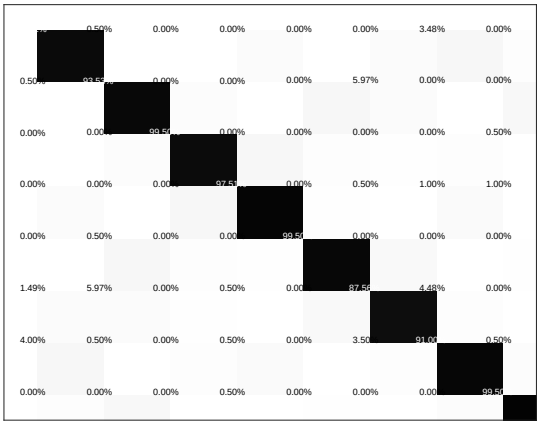

Fig. 10. Confusion matrix for RaFD dataset.

We provided the recognition performance in **Table 7**, which shows that the proposed descriptor achieves higher accuracy than other methods in this dataset. Notice that the RaFD dataset contains images of people from different races. Hence, such higher performance can also be interpreted as the efficacy of HPED for race variations. Moreover, the confusion matrix for HPED is shown in **Fig. 10** for the facial expression images in the RaFD dataset. As shown, HPED also demonstrates superior class separation in all eight expression classes.

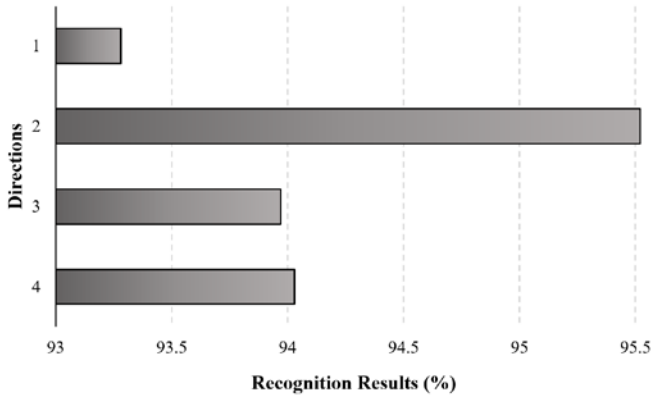

Fig. 11. Recognition results of HPED with different numbers of directions in the RaFD dataset.

To this end, one may ask why only two directions were used, that is, why the primary and secondary directions are generated in the HPED descriptor. Therefore, we would like to show the rationale of using the secondary direction along with primary direction experimentally. For this purpose, we evaluated the performance of a HPED descriptor using *n* different directions. Therefore, we varied the number of directions (*n*) from 1 to 4 and generated HPED histograms for person-independent results, respectively. Here, *n=1* denoted the generation of the HPED histogram using the top edge direction (primary direction) only. Similarly, *n=2, 3, 4* denote generation of the histogram by concatenating the top 2, 3 and 4 edge directions, respectively. Experiments were conducted on a RaFD dataset, and the results are shown in **Fig. 11**, where we observe that the method performs better when considering two directions (primary and secondary). This can be explained by the fact that using just one direction considers the edge directional information only, and hence may miss the structural details of curvature and corner-like textures, where more directions are needed. Considering two directions is more meaningful in this regard, since the primary direction captures the major edge directional axis and the secondary direction can be utilized to capture such curvature information, as was also mentioned in recent research [34]. Moreover, considering three or four types of information can be useful to capture complex junction-like textures. Nevertheless, since facial images usually do not possess such complex textures, using more directional information may generate redundant information. This may create ambiguity in the texture representation, resulting in less accuracy, as shown in **Fig. 11**.

5. Conclusion

In this paper, we present a new descriptor, Histograms of Prominent Edge Directions (HPED), for the person-independent facial expression recognition task. The proposed HPED uses the histogram of top two prominent edge directions to represent the expression related changes in face images. Such coding schemes allow HPED to avoid sampling errors while preserving valuable spatial information. Compared to HOG, which uses primary edge directions only, HPED uses secondary direction information to take advantage of important texture information. Experiments on well-known datasets demonstrate that the proposed HPED works better than other descriptors in recognizing expressions in a person-independent environment. Moreover, we show that HPED achieves better performance against noise, positional variations, and low-resolution images, showing its overall efficacy in recognizing human expressions. However, the robustness of HPED under such conditions indicates that HPED

can be applied to recognize the "in-the-wild" expressions as well as spontaneous expressions, which we leave for future endeavors.

References

[1] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on Local Binary Patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803-816, May 2009. Article (CrossRef Link).

[2] H. Hong, H. Neven, C. Von der Malsburg, "Online facial expression recognition based on personalized galleries," in *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 354-359, April 14-16, 1998. Article (CrossRef Link).

[3] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," *IEEE transactions on image processing*, vol.16, no. 1, pp. 172-87, January 2007. Article (CrossRef Link).

[4] A. R. Rivera, J. A. R. Castillo, and O. Chae, "Local directional number pattern for face analysis: Face and expression recognition," *IEEE transactions on image processing*, vol. 22, no.5, pp. 1740-1752, May 2013. Article (CrossRef Link).

[5] A. M. Martínez, A. C. Kak, "Pca versus lda," *IEEE transactions on pattern analysis and machine intelligence*, vol. 23, no. 2, pp. 228-33, February 2001. Article (CrossRef Link).

[6] K. Etemad, R. Chellappa, "Discriminant analysis for recognition of human face images", *JOSA A*, vol. 14, no. 8, pp. 1724-33, August 1997. Article (CrossRef Link).

[7] T. Jabid, Md. H. Kabir, and O. Chae, "Robust facial expression recognition based on local directional pattern," *ETRI J.*, vol. 32, no. 5, pp. 784-794, October 2010. Article (CrossRef Link).

[8] A. R. Rivera, J. A. R. Castillo, and O. Chae, "Recognition of face expressions using local principal texture pattern," in *Proc. of IEEE International Conference on Image Processing (ICIP)*, pp. 2613–2616, September 2012. Article (CrossRef Link).

[9] M. T. B. Iqbal, B. Ryu, G. Song, and O. Chae, "Positional Ternary Pattern (PTP): An edge based image descriptor for human age recognition" in *Proc. of IEEE International Conference on Consumer Electronics (ICCE)*, pp. 289-292, January 7-11, 2016. Article (CrossRef Link).

[10] R. A. Kirsch, "Computer determination of the constituent structure of biological images," *Computers and biomedical research*, vol. 4, no. 3, pp. 315–328, 1971. Article (CrossRef Link).

[11] M. T. B. Iqbal, M. Shoyaib, B. Ryu, M. Abdullah-Al-Wadud, O. Chae, "Directional Age-Primitive Pattern (DAPP) for Human Age Group Recognition and Age Estimation," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 11, pp. 2505 – 2517, November 2017. Article (CrossRef Link).

[12] P. Carcagnì, M. Coco, M. Leo, C. Distante, "Facial expression recognition and histograms of oriented gradients: a comprehensive study," *SpringerPlus*, vol. 4, no. 1, p. 645, December 2015. Article (CrossRef Link).

[13] B. Ryu, A. R. Rivera, J. Kim, O. Chae, "Local directional ternary pattern for facial expression recognition," *IEEE Transactions on Image Processing*, vol. 26, no. 12, pp. 6006-6018, December 2017. Article (CrossRef Link).

[14] G. S. Robinson, "Edge detection by compass gradient masks," *Computer graphics and image processing*, vol. 6, no. 5, pp. 492-501, October 1977. Article (CrossRef Link).

[15] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition – Workshops*, pp. 94-101, June 13-18, 2010. Article (CrossRef Link).

[16] N. C. Ebner, M. Riediger, and U. Lindenberger, "FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation," *Behavior Research Methods*, vol. 42, no. 1, pp. 351–362, February 2010. Article (CrossRef Link).

[17] L. Yin, X. Wei, Y. Sun, J. Wang, M. J. Rosato, "A 3D facial expression database for facial behavior research," in *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 211-216, April 2, 2006. Article (CrossRef Link).
[18] O. Langner, R. Dotsch, G. Bijlstra, D. H. Wigboldus, S. T. Hawk, A. D. Van Knippenberg, "Presentation and validation of the Radboud Faces Database," *Cognition and Emotion*, vol. 24, no. 8, pp. 1377-1388, December 2010. Article (CrossRef Link).
[19] C. Cortes, V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273-97, September 1995. Article (CrossRef Link).
[20] R. Ptucha, G. Tsagkatakis, A. Savakis, "Manifold based sparse representation for robust expression recognition without neutral subtraction," in *Proc. of IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 2136-2143, November 6, 2011. Article (CrossRef Link).
[21] S. H. Lee, W. J. Baddar, Y. M. Ro, "Collaborative expression representation using peak expression and intra class variation face images for practical subject-independent emotion recognition in videos," *Pattern Recognition*, vol. 54, pp. 52-67, 2016. Article (CrossRef Link).
[22] S. H. Lee, K. N. Plataniotis, Y. M. Ro, "Intra-class variation reduction using training expression images for sparse representation based facial expression recognition," *IEEE Transactions on Affective Computing*, vol. 5, pp. 340-51, 2014. Article (CrossRef Link).
[23] B. F. Wu and C. H. Lin, "Adaptive Feature Mapping for Customizing Deep Learning Based Facial Expression Recognition Model," *IEEE Access*, vol.6, pp. 12451-12461, 2018. Article (CrossRef Link).
[24] L. A. Jeni, D. Takacs and A. Lorincz, "High quality facial expression recognition in video streams using shape related information only," in *Proc. of IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 2168-2174, November 6, 2011. Article (CrossRef Link).
[25] S. Yang, B. Bhanu, "Understanding discrete facial expressions in video using an emotion avatar image," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 4, pp. 980-92, August 2012. Article (CrossRef Link).
[26] Z. Wang, S. Wang, Q. Ji, "Capturing complex spatio-temporal relations among facial muscles for facial expression recognition," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3422-3429, June 23, 2013. Article (CrossRef Link).
[27] H. Jung, S. Lee, J. Yim, S. Park, J. Kim, "Joint fine-tuning in deep neural networks for facial expression recognition," in *Proc. of IEEE International Conference on Computer Vision (ICCV)*, pp. 2983-2991, December 7, 2015. Article (CrossRef Link).
[28] A. Caroppo, A. Leone, P. Siciliano, "Facial Expression Recognition in Older Adults using Deep Machine Learning," in *Proc. of Third Italian Workshop on Artificial Intelligence for Ambient Assisted Living 2017*, pp. 30-43, November 16-17, 2017.
[29] Y. Hu, Z. Zeng, L. Yin, X. Wei, J. Tu, T.S. Huang, "A study of non-frontal-view facial expressions recognition," in *Proc. of International Conference on Pattern Recognition*, pp. 1-4, December 8, 2008. Article (CrossRef Link).
[30] W. Zheng, H. Tang, Z. Lin, T.S. Huang, "Emotion recognition from arbitrary view facial images," in *Proc. of European Conference on Computer Vision*, pp. 490-503, September 5, 2010. Article (CrossRef Link).
[31] U. Tariq, T. S. Huang, "Features and fusion for expression recognition - A comparative analysis," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 146-152, June 16, 2012. Article (CrossRef Link).
[32] S. Moore, R. Bowden, "Local binary patterns for multi-view facial expression recognition," *Computer Vision and Image Understanding*, vol. 115, no. 4, pp. 541-58, April, 2011. Article (CrossRef Link).
[33] Z. Zhang, G. Lu, J. Yan, H. Li, N. Sun, X. Li, P. R. Zhenjiang, "Compact local Gabor directional number pattern for facial expression recognition," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 26, no. 3, pp. 1236-1248, May 2018. Article (CrossRef Link).

[34] M. T. Iqbal, M. Shoyaib, B. Ryu, M. Abdullah-Al-Wadud, O. Chae, "Directional age-primitive pattern (DAPP) for human age group recognition and age estimation," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 11, pp. 2505-2517, November 2017. Article (CrossRef Link).

Farkhod Makhmudkhujaev received the B.S. degree in information technologies and the M.S. degree in applied informatics from Tashkent University of Information Technologies, Tashkent, Uzbekistan, in 2012, and 2014, respectively. He is currently pursuing the Ph.D. degree at the Department of Computer Science and Engineering, Kyung Hee University, Yongin-si, Republic of Korea. His current research interests include facial expression recognition, background modeling, object and scene detection, and image matching.

Md Tauhid Bin Iqbal received his bachelor's degree in Information Technology from the University of Dhaka in 2012. Currently, he is pursuing a combined M.S./Ph.D. degree at the Department of Computer Science and Engineering, Kyung Hee University, Yongin-si, Republic of Korea. His current research interests include object detection, expression recognition, combined age & gender recognition, and pattern recognition.

Md Rifat Arefin received his bachelor's degree in Information Technology from the University of Dhaka in 2016. Currently, he is pursuing a M.S. degree at the Department of Computer Science and Engineering, Kyung Hee University, Yongin-si, Republic of Korea. His current research interests include object detection, expression recognition, pattern recognition, background modeling, and deep learning.

Byungyong Ryu received a B.S. degree in 2010 and a Ph.D. degree in 2017 in computer engineering from Kyung Hee University, Yongin-si, Republic of Korea, where he is currently working as a Post-doctoral fellow. His current research interests include facial expression, age, and gender recognition using face images, and image enhancement and medical image processing in dentistry, and deep learning.

Oksam Chae (M92) received a B.S. degree in electronics engineering from Inha University, Incheon, South Korea, in 1977, and M.S. and Ph.D. degrees in electrical and computer engineering from Oklahoma State University, Stillwater, in 1982 and 1986, respectively. He was a Research Engineer with the Texas Instruments Image Processing Laboratory, Dallas, TX, from 1986 to 1988. Since 1988, he has been a Professor with the Department of Computer Science and Engineering, Kyung Hee University, Yongin-si, Republic of Korea. His current research interests include multimedia data processing environments, intrusion detection systems, sensor networks, and medical image processing in dentistry. Prof. Chae is a member of the SPIE, the Korean Electronic Society (KES), and the Institute of Electronics, Information and Communication Engineers.