Hybrid Facial Representations for Emotion Recognition

Woo-han Yun, DoHyung Kim, Chankyu Park, and Jaehong Kim

Automatic facial expression recognition is a widely studied problem in computer vision and human-robot interaction. There has been a range of studies for representing facial descriptors for facial expression recognition. Some prominent descriptors were presented in the first facial expression recognition and analysis challenge (FERA2011). In that competition, the Local Gabor Binary Pattern Histogram Sequence descriptor showed the most powerful description capability. In this paper, we introduce hybrid facial representations for facial expression recognition, which have more powerful description capability with lower dimensionality. Our descriptors consist of a block-based descriptor and a pixel-based descriptor. The block-based descriptor represents the micro-orientation and micro-geometric structure information. The pixel-based descriptor represents texture information. We validate our descriptors on two public databases, and the results show that our descriptors perform well with a relatively low dimensionality.

Keywords: Facial expression recognition, Histograms of Oriented Gradients, HOG, Local Binary Pattern, LBP, Rotated Local Binary Pattern, RLBP, Gabor filter, GF.

Manuscript received Mar. 31, 2013; revised Aug. 29, 2013; accepted Sept. 23, 2013.

This work was supported by the R&D program of the Korea Ministry of Knowledge and Economy (MKE) and the Korea Evaluation Institute of Industrial Technology (KEIT) [10041826, Development of emotional features sensing, diagnostics and distribution s/w platform for measurement of multiple intelligence from young children].

Woo-han Yun (phone: +82 42 860 5804, yochin@etri.re.kr), DoHyung Kim (dhkim008@etri.re.kr), Chankyu Park (parkck@etri.re.kr), and Jaehong Kim (jhkim504@etri.re.kr) are with the IT Convergence Technology Research Laboratory, ETRI, Daejeon, Rep. of Korea.

I. Introduction

Facial expression is a natural and intuitive means for humans to express and sense their emotions and intentions. For this reason, automatic facial expression recognition has been an active research field in computer vision and human-robot interaction for a long time [1], [2]. In the case of robots living with a family, it is very useful to sense the family members' emotions through facial expressions and respond appropriately.

There are three stages in the general automatic facial expression recognition systems. The first stage is to detect the faces and normalize the photographic images of the faces. This stage may be based on a holistic facial region or on facial components such as the eyes, nose, and mouth. The next stage is to extract the facial expression descriptors from the normalized faces. Finally, the system classifies the facial descriptors into the proper expression categories.

In this paper, we introduce new facial expression descriptors. These descriptors adopt two representations, a block-based representation and a pixel-based representation, to reflect the micro-orientation, micro-geometric structure, and texture information. The descriptors show more powerful description capability with low dimensionality than the state-of-the-art descriptors.

II. Previous Work

Many researchers have shown a range of approaches to construct an automatic facial expression recognition system. Geometric approaches and texture-based approaches are the most prominent types. Texture-based approaches have generally shown a better performance than geometric approaches in previous research [3], [4]. In texture-based

approaches, two descriptors have been widely used for facial expression recognition, that is, Gabor filter (GF) descriptor and Local Binary Pattern (LBP) descriptor. A GF is inspired by a human visual system and has sufficient description capability to represent the texture representation of an image [5], [6]. An LBP has the capability to represent a micro-geometric structure by comparing the pixel intensity with its neighbors [7], [8]. Later, a compact version, that is, Uniform LBP (ULBP), was designed and utilized in the research field [9]. Histograms of Oriented Gradients (HOG) is another feature descriptor used to represent the gradient orientation information of an image [10]. HOG was first designed for pedestrian detection and later used for other purposes [11].

In the first facial expression recognition and analysis challenge (FERA2011), a contest for automatic facial expression recognition, Local Gabor Binary Pattern Histogram Sequence (LGBPHS) descriptor showed the best facial description performance in terms of facial expression recognition [3], [4]. This method is a kind of hybrid approach because it serially combines the above two approaches, GF and ULBP. Even though this descriptor has a very powerful description capability, the dimensionality of the descriptor reaches 236,000 (59 bins × 100 blocks × 40 GFs). This high dimensionality has a shortage when extracting and storing facial expression information, and it makes it difficult to construct a final lightweight facial expression recognition system.

III. Proposed Descriptors

In this paper, we suggest using new hybrid descriptors, which include two types of representations: a block-based representation and a pixel-based representation. A block-based representation describes micro-geometric structure information and micro-orientation information. We first encode the orientation information using a HOG-like method in the entire image of the face. The micro-geometric structure information is encoded through one of three descriptors: LBP, ULBP, or Rotated LBP (RLBP). LBP and ULBP are popular geometric structure encoding methods. RLBP is a modified version of LBP that removes the orientation information by rotating the micro-patch to align with the canonical direction and by extracting the micro-geometric structure information at each pixel. Micro-geometric structure information and micro-orientation information are then represented in a block-based histogram representation. A pixel-based representation describes the texture information with multi-frequency and multi-orientation using GFs. To integrate three descriptors in two representation types into one, we use a principal component analysis (PCA) on each descriptor and then concatenate them. For additional tests, instead of using PCA, we use feature extraction and fusion methods: PCA + linear discriminant analysis (LDA) and PCA + canonical correlation analysis (CCA) [12].

1. Block-Based Representation

A. Orientation Information

To encode micro-orientation information, we follow a HOG-like approach. First, to remove the noise, an image of a face is blurred using a 3×3 Gaussian kernel. We then compute the gradients in the x and y directions at each pixel using a 3×3 Sobel operator. The orientation at each pixel is calculated using (1). The obtained direction information at each pixel is saved into one image, called an orientation map or O_{map}.

$$O_{map}(x,y) = \tan^{-1}\frac{g_y(x,y)}{g_x(x,y)}, \quad (1)$$

where $g_x(x, y)$ and $g_y(x, y)$ define the gradient value at position x, y in the x and y directions, respectively, after a Sobel operation. We encode the orientation map in a block-based histogram representation. The orientation map is divided into several blocks. We then obtain the histogram of the gradient orientation of each block and concatenate those histograms into one feature vector.

To improve the description power, we divide 360 degrees into 16 bins of the histograms instead of 180 degrees into 8 bins in the experiment. This means that 60 degrees and 240 degrees belong to different bins in terms of orientation. We divide an image of a face with 100×100 pixels into 20×20 blocks instead of the more commonly used division of 10×10 blocks. These two factors improve the performance by around 2%. Figure 1 shows an overview of the process used to obtain the orientation information.

B. Geometric Structure Information

To represent the micro-geometric information in the local region, we choose one of three descriptors: LBP [7], ULBP [9], or RLBP. We propose RLBP as a modified version of LBP. Because we have the orientation information from the previous stage, we do not need to use the original LBP. To obtain only the micro-geometric structure information without redundant orientation information, we rotate the local patch to follow a canonical orientation using the orientation information we obtained in the previous stage. We then obtain the binary pattern at each pixel.

Next, we adopt a ULBP-like method. In the training stage, we investigate the binary patterns to know what patterns are popularly used and what patterns are rarely used. We only

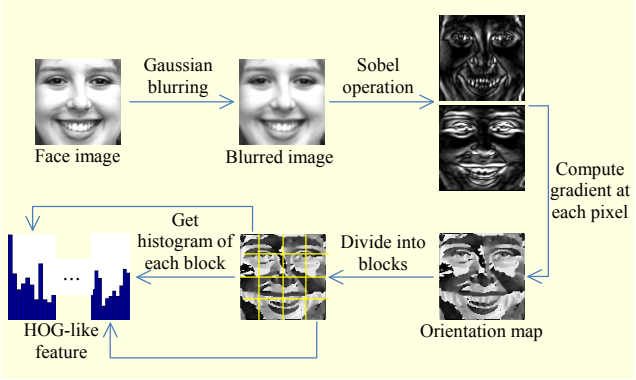


Fig. 1. Overview of process used to obtain orientation information.

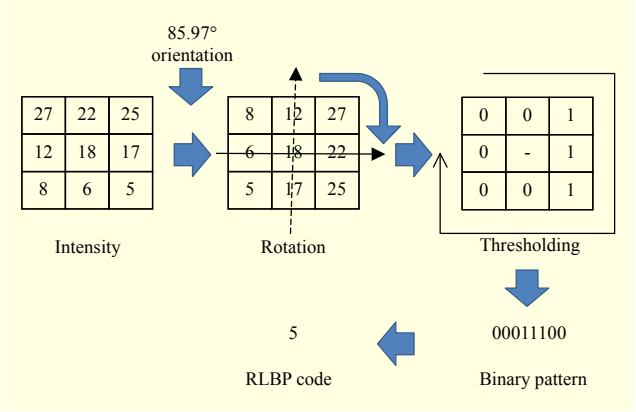


Fig. 2. Process used to obtain RLBP code from 3×3 patch of blurred image of face.

utilize popularly used patterns as separate RLBP codes. The rarely used patterns are labeled with a single code. To determine the number of labels in RLBP, we test various numbers of labels, 6, 9, 14, 18, 24, 35, and 48. In the experiment, RLBP with 14 labels shows the best performance and 13 of 14 labels retain around 95% of the patterns. The process used to obtain the RLBP code from the 3×3 patch of the blurred image of the face at each pixel is summarized in Fig. 2.

RLBP is quite different from ULBP. In ULBP, the same geometric structures with different orientations are labeled with separate codes. Contrary to ULBP, in RLBP, the same geometric structures with different orientations are labeled with a same code. This modification has several strong points. First, we can save much more dimensional space by removing useless bins in an LBP. We use only limited bins because we align the patch with the main orientation. We can also represent the same geometric information in a more compact way.

After we obtain the RLBP code for every pixel, we save the RLBP code at each pixel and obtain a geometric structure map. The geometric structure map is divided into several blocks.

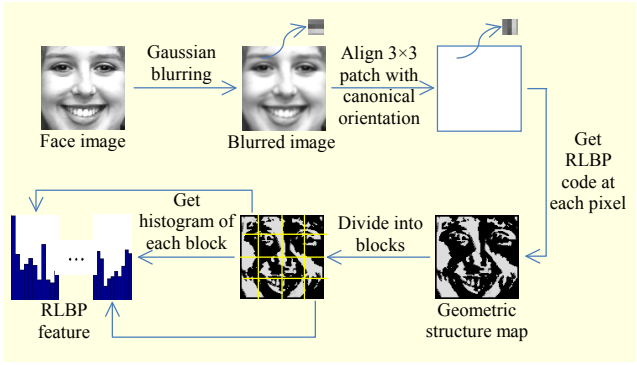


Fig. 3. RLBP process.

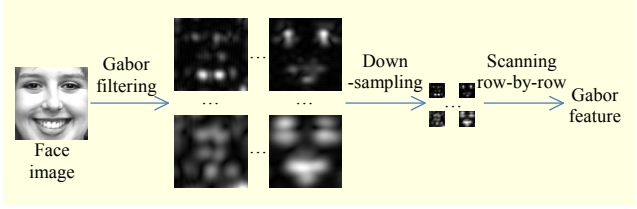


Fig. 4. Process to obtain pixel-based representation.

We then obtain the histogram of the RLBP code of each block and concatenate those histograms into one RLBP feature vector. An overview of this step is shown in Fig. 3. In the experiment, we divide an image of a face with 100×100 pixels into 20×20 blocks.

2. Pixel-Based Representation

We apply the multi-frequency and multi-orientation GFs to extract the texture information from a face image [6]. The Gabor representation is derived by convolving the face image with the GFs. Let $I(\mathbf{z})$ be an image of a face and $\mathbf{z}=(x, y)$. The convolution of $I(\mathbf{z})$ using the 2D GFs, $g_{\mu,v}(\mathbf{z})$, can then be defined as follows [6]:

$$G(\mathbf{z}, \mu, v) = \int_{\mathbf{z}} I(\mathbf{z}) g_{\mu,v}(\mathbf{z}-\mathbf{z}')d\mathbf{z}, \quad (2)$$

with a family of GFs, $g_{\mu,v}(\mathbf{z})$, and the characteristic wave vector $\mathbf{k}$ [5]:

$$g_{\mu,v}(\mathbf{z}) = \frac{\|\mathbf{k}\|^2}{(2\pi)^2}\exp\left(-\frac{\|\mathbf{k}\|^2\|\mathbf{z}\|^2}{2(2\pi)^2}\right)\left(\exp(i\mathbf{k}^T\mathbf{z})-\exp\left(-\frac{(2\pi)^2}{2}\right)\right), \quad (3)$$

$$\mathbf{k} = \left[2^{-\frac{v+2}{2}}\pi\cos\left(\mu\frac{\pi}{8}\right), 2^{-\frac{v+2}{2}}\pi\sin\left(\mu\frac{\pi}{8}\right)\right]^T. \quad (4)$$

The parameters μ and v denote the orientation and frequency of the GF, respectively. In the experiment, we use three frequencies (v=0, 1, 2) and four orientations (μ=0, 2, 4, 6). We

only use the magnitude of the GF output. Because of a huge dimensionality of the Gabor filtered image, the magnitudes of the twelve GF outputs are downsampled into an image of 20 pixels × 20 pixels. Each downsampled image is scanned row by row and concatenated into one long feature vector of 4,800 × 1. The summarized process is shown in Fig. 4.

3. Descriptor Integration

To integrate three descriptors in two representation types into one, we use PCA on each descriptor and concatenate them. Although the CCA shows a better performance than PCA in previous research [12] and PCA+LDA shows a better performance in our experiment with the ground truth landmarks, PCA shows the best performance in a fully automatic experiment similar to a real situation. We call each final descriptor a HOG-LBP-Gabor (HLG), HOG-ULBP-Gabor (HUG), and HOG-RLBP-Gabor (HRG) with respect to their components based on PCA. Additionally, for a combination with the feature extraction of LDA and CCA, we added the notations "-LDA" and "-CCA."

IV. Experiment Results

We validate our descriptors on the JAFFE database and CK+ database with other state-of-the-art descriptors for facial expression recognition [13], [14]. The JAFFE database includes 219 images of ten Japanese females with seven facial expressions, that is, neutral (NE), happiness (HA), sadness (SA), surprise (SU), anger (AN), disgust (DI), and fear (FE). The JAFFE database does not include eye coordinates. Therefore, we use eye coordinates obtained from the automatic eye detection algorithm using AdaBoost. The other database we use is the Extended Cohn-Kanade (CK+) database. This database includes 593 sequences across 123 persons with seven facial expressions, that is, angry (AN), contempt (CO), disgust (DI), fear (FE), happy (HA), sadness (SA), and surprise (SU), and is an extended version of the previous Cohn-Kanade database [15]. In this database, only 327 sequences were labeled with its own emotion. In the experiment, we use only the snapshot at the peak emotion, which is labeled in the database. We use a tenfold cross validation for the JAFFE database and leave-one-subject-out cross validation for the CK+ database.

In the test, we first detect the face and eye positions using AdaBoost-based detectors or utilize the eye coordinates in the database. To obtain the image of the face, we use a two-point holistic alignment translating and rotating the face image on the eye coordinates. The aligned image is scaled into a 100 pixel × 100 pixel cropped image of a face in which the intra-ocular distance is 48 pixels. The descriptors are normalized to the unit length. Since all descriptors have a high dimensionality leading to a computational burden, we apply PCA to reduce the dimensionality, retaining 95% of the energy except for our descriptors, which already have a reduced dimension using PCA in the inner process [16]. This dimensional reduction shows a better performance than those with the original dimension in the experiment. We use the LIBSVM with a linear kernel as a classifier [17]. We utilize the MATLAB toolbox for a dimensionality reduction for PCA and LDA [18].

1. Results on JAFFE Database

We first conduct the experiment on the JAFFE database. The JAFFE database includes 219 images of ten Japanese females with seven facial expressions. In the experiment, we use the eye coordinates obtained using the automatic eye detection algorithm based on AdaBoost because the database does not include the eye coordinates. We validate our descriptor on a tenfold cross validation of the subjects. We first present the comparison results of our descriptors with other state-of-the-art descriptors, as shown in Table 1. The state-of-the-art descriptors we compare include Gabor3×4 [6], Gabor5×8 [19], LBP(LBP256) [8], ULBP(LBP59) [9], HOG [10], and LGBPHS [20].

In these results, our descriptors, HRG, HUG, and HLG, show better recognition rates than the other descriptors for most facial expressions. Furthermore, considering the dimensionality of each descriptor, our descriptor has many more pros. We present the original dimensionality and a computation time for descriptor extraction and dimension reduction of the descriptors in Table 2. Even though the computational time is dependent on the experimental settings and implementations, such as the optimization level, it could be an option for

Table 1. Recognition rate of descriptors on JAFFE database.

Descriptor	NE	HA	SA	SU	AN	DI	FE	Overall
Gabor3×4	**100.00**	90.32	80.65	**96.67**	90.00	68.97	87.50	87.79
Gabor5×8	96.67	90.32	74.19	90.00	96.67	75.86	84.38	86.85
LBP256	90.00	**96.77**	70.97	90.00	90.00	68.97	68.75	82.16
LBP59	80.00	87.10	64.52	76.67	93.33	62.07	65.63	75.59
RLBP	80.00	90.32	70.97	86.67	83.33	86.21	68.75	80.75
HOG	90.00	**96.77**	74.19	90.00	86.67	72.41	75.00	83.57
LGBPHS	**100.00**	**96.77**	80.65	93.33	90.00	**96.55**	**93.75**	92.96
HRG	**100.00**	**96.77**	**96.77**	**96.67**	96.67	**96.55**	**93.75**	**96.71**
HUG	**100.00**	**96.77**	**96.77**	**96.67**	96.67	89.66	**93.75**	95.77
HLG	**100.00**	**96.77**	**96.77**	93.33	**100.00**	93.10	**93.75**	96.24

ETRI Journal, Volume 35, Number 6, December 2013
http://dx.doi.org/10.4218/etrij.13.2013.0054

Table 2. Original dimensionality of descriptors and computation time for descriptor extraction and dimension reduction.

Descriptor	Dimensionality	Computational time	
		Descriptor extraction (C, ms)	Dimension reduction (MATLAB, ms)
Gabor3×4	4,800	42.82	0.45
Gabor5×8	400,000	203.08	22.94
LBP256	25,600	5.59	2.25
LBP59	5,900	14.16	0.77
RLBP	5,600	33.23	0.77
HOG	6,400	3.92	0.88
LGBPHS	236,000	777.20	22.02
HRG	16,800	80.28	1.85
HUG	17,100	60.91	1.88
HLG	36,800	52.33	3.44

Table 3. Confusion matrix of HRG descriptor on JAFFE database (%).

	NE	HA	SA	SU	AN	DI	FE
NE	100.00	3.23	0.00	0.00	0.00	0.00	0.00
HA	0.00	96.77	3.23	3.33	0.00	0.00	0.00
SA	0.00	0.00	96.77	0.00	0.00	3.45	0.00
SU	0.00	0.00	0.00	96.67	0.00	0.00	6.25
AN	0.00	0.00	0.00	0.00	96.67	0.00	0.00
DI	0.00	0.00	0.00	0.00	3.33	96.55	0.00
FE	0.00	0.00	0.00	0.00	0.00	0.00	93.75

(column: true class; row: predicted class)

comparing descriptors. We use C for descriptor extraction and MATLAB for dimension reduction. The computational time for classification is omitted because the time is less than 0.5 milliseconds for all cases. Our descriptors show a better performance with relatively low dimensionalities than other high dimensional descriptors, such as Gabor5×8 and LGBPHS.

Table 3 shows the final confusion matrix of the HRG descriptor showing the best accuracy on the JAFFE database. We can see that the best accuracy is for neutral, at 100%, and the worst accuracy is for fear, at 93.75%. The accuracy of the other facial expressions is about 96%.

2. Results of CK+ Database

We also conduct the experiment on the CK+ database. This database includes 593 sequences across 123 persons with seven facial expressions. In the experiment, we only use the 327 snapshots at the peak emotion with their emotion labels.

Table 4. Recognition rate of individual expressions for descriptors on CK+ database with eye coordinates from landmarks.

Descriptor	AN	CO	DI	FE	HA	SA	SU
Gabor3×4	77.78	55.56	94.92	68.00	98.55	71.43	**98.8**
Gabor5×8	86.67	72.22	**96.61**	**80.00**	**100.00**	**92.86**	**98.8**
LBP256	93.33	**88.89**	89.83	52.00	98.55	71.43	97.59
LBP59	91.11	**88.89**	88.14	48.00	**100.00**	75.00	96.39
RLBP	88.89	66.67	94.92	60.00	97.10	64.29	95.18
HOG	88.89	77.78	93.22	64.00	98.55	71.43	97.59
LGBPHS	80.00	77.78	94.92	76.00	**100.00**	**92.86**	**98.80**
HRG	91.11	**88.89**	94.92	72.00	**100.00**	78.57	97.59
HUG	**95.56**	**88.89**	94.92	**80.00**	**100.00**	78.57	97.59
HLG	**95.56**	**88.89**	94.92	**80.00**	**100.00**	78.57	97.59

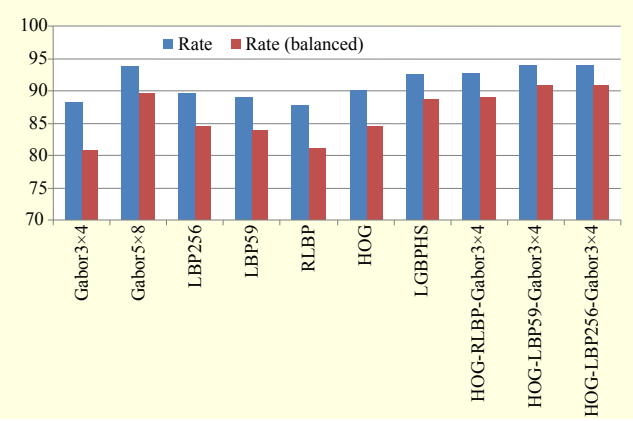


Fig. 5. Performance and balanced rate of descriptors on CK+ with eye coordinates from landmarks.

We use a leave-one-subject-out cross validation. We conduct the experiment on the CK+ database with eye coordinates from both the database and the automatic detector. For the CK+ database, we use an additional measurement, the balanced recognition rate, as well as an overall recognition rate because the CK+ has a severely different amount of facial expression data among the facial classes. This unbalanced amount of data among classes causes the overall recognition rate to become biased toward facial expression class results having many samples. A balanced recognition rate can be yielded by averaging each recognition rate over all classes.

A. Results on CK+ with Eye Coordinates

We utilize the landmark points on the left and right eyes and obtain the eye coordinates. The facial expression recognition rates with other state-of-the-art descriptors are shown in Table 4 and Fig. 5.

Table 5. Confusion matrix of HLG descriptor (%).

	AN	CO	DI	FE	HA	SA	SU
AN	95.56	0.00	1.69	4.00	0.00	17.86	0.00
CO	0.00	88.89	0.00	0.00	0.00	0.00	1.20
DI	2.22	0.00	94.92	0.00	0.00	0.00	0.00
FE	0.00	5.56	1.69	80.00	0.00	3.57	1.20
HA	0.00	0.00	1.69	4.00	100.00	0.00	0.00
SA	2.22	5.56	0.00	0.00	0.00	78.57	0.00
SU	0.00	0.00	0.00	12.00	0.00	0.00	97.59

(column: true class; row: predicted class)

Table 6. Recognition rate of HLG descriptor with several feature extraction and fusion methods.

	HLG	HLG-LDA	HLG-CCA
Rate (%)	93.88	97.86	78.90
Rate (%) (balanced)	90.79	97.08	65.05

The proposed descriptors HUG and HLG show the best recognition rates equally: 93.88% for the overall recognition and 90.79% for the balanced recognition. For the overall recognition and balanced recognition, Gabor5×8 shows 93.58% and 89.59%, HRG yields 92.66% and 89.01%, and LGBPHS records slightly lower recognition rates of 92.35% and 88.62%, respectively. Although the recognition rates of these descriptors are similar, our descriptors have an advantage in terms of lower dimensionality.

Table 5 shows the final confusion matrix of our HLG descriptor on the CK+ database. We can observe that happiness, anger, disgust, and surprise are easily recognized with high accuracy, whereas sadness shows a relatively low recognition rate, as it is confused with other expressions, especially anger. This is because sadness and anger have similar features, such as a furrowed brow and a closed mouth. Furthermore, in some cases, a sad expression of one person looks like an angry expression of another person and vice versa. This confusion is also observed on the other descriptors. While showing a higher recognition rate for sadness, the Gabor5×8 and LGBPHS descriptors show a lower rate for anger. To increase the feature discrimination capability, we also apply LDA or CCA into the inner process. Table 6 shows the recognition rate of the HLG descriptor with feature extraction and the fusion methods, LDA and CCA. With LDA, HLG shows surprisingly increased recognition rates of 97.86% for the overall rate and 97.08% for the balanced rate. This is because LDA finds the helpful subspace reducing the variance within the facial expression classes and increasing the variance between classes.

Table 7. Recognition rate of individual expressions for descriptors on CK+ database with automatic eye detection algorithm.

Descriptor	AN	CO	DI	FE	HA	SA	SU
Gabor3×4	73.33	44.44	91.53	64.00	**100.00**	71.43	**98.80**
Gabor5×8	82.22	66.67	**96.61**	**84.00**	**100.00**	**89.29**	**98.80**
LBP256	91.11	77.78	93.22	52.00	**100.00**	67.86	**98.80**
LBP59	80.00	72.22	88.14	52.00	95.65	60.71	93.98
RLBP	82.22	77.78	94.92	68.00	98.55	53.57	93.98
HOG	84.44	77.78	94.92	68.00	98.55	78.57	97.59
LGBPHS	86.67	61.11	94.92	80.00	**100.00**	78.57	**98.80**
HRG	**93.33**	83.33	94.92	72.00	**100.00**	82.14	97.59
HUG	88.89	**88.89**	94.92	**84.00**	**100.00**	**89.29**	**98.80**
HLG	**93.33**	**88.89**	**96.61**	76.00	**100.00**	82.14	**98.80**

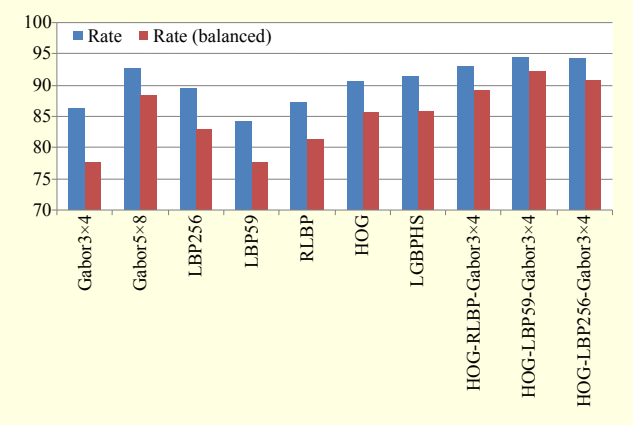


Fig. 6. Performance rate and its balanced rate of descriptors on CK+ with automatic eye detection algorithm.

B. Results on CK+ with Automatic Eye Detection Algorithm

We also test the fully automatic recognition rate. This recognition rate can be used in a real situation. Table 7 and Fig. 6 show the performance of individual expressions and a comparison recognition rate among several descriptors.

Proposed descriptors HUG, HLG, and HRG show the best recognition rates at 94.50%, 94.19%, and 92.97 for the overall recognition and 92.11%, 90.82%, and 89.04% for the balanced recognition, respectively. Gabor5×8 and LGBPHS are followed with slightly lower recognition rates of 92.66% and 91.44% for the overall recognition and 88.23% and 85.72% for the balanced recognition, respectively. Interestingly, our descriptors, HUG, HLG, and HRG, show slightly increased recognition rates on the automatically detected database than on the database with ground-truth landmarks. This is because

ETRI Journal, Volume 35, Number 6, December 2013
http://dx.doi.org/10.4218/etrij.13.2013.0054

Table 8. Confusion matrix of HLG descriptor (%).

	AN	CO	DI	FE	HA	SA	SU
AN	88.89	0.00	3.39	4.00	0.00	7.14	0.00
CO	2.22	88.89	0.00	0.00	0.00	0.00	1.20
DI	2.22	0.00	94.92	0.00	0.00	3.57	0.00
FE	0.00	5.56	1.69	84.00	0.00	0.00	0.00
HA	0.00	0.00	0.00	8.00	100.00	0.00	0.00
SA	6.67	5.56	0.00	0.00	0.00	89.29	0.00
SU	0.00	0.00	0.00	4.00	0.00	0.00	98.80

(column: true class; row: predicted class)

Table 9. Recognition rate of HUG descriptor with several feature extraction and fusion methods.

	HUG	HUG-LDA	HUG-CCA
Rate (%)	94.50	94.19	78.59
Rate (%) (balanced)	92.11	93.88	66.17

the block-based approaches are not affected by a small misalignment by the automatic eye detection. However, pixel-based approaches are susceptible to misalignment; for example, Gabor3×4, Gabor5×8, and LGBPHS lead to a degradation of the recognition rate of 1.83%, 0.92%, and 0.91% for the overall recognition and 3.07%, 1.37%, and 2.90% for the balanced recognition, respectively. The confusion matrix of HUG is shown in Table 8. The results show that the classifier confuses anger and sadness in some cases, which accounts for most of the confusion. We also test the best descriptor, HUG, by combining it with LDA and CCA. The results are shown in Table 9. HUG-LDA records a degradation of 0.31% in the overall recognition rate and 1.61% in the balanced recognition rate. This is because the LDA approach is susceptible to outliers and noise [21]. The automatic eye detector is not perfect and can therefore yield a misaligned image of a face leading to a degradation of the LDA.

V. Conclusion

In this paper, we suggested the use of hybrid representations for facial expression recognition. These representations consist of a block-based representation and a pixel-based representation. The block-based representation represents micro-orientation and micro-geometric structure information. The pixel-based representation represents texture information. In the experiment, we tested the descriptors with state-of-the-art descriptors on two public databases and showed that its facial expression recognition performance was comparable with a relatively low dimensionality. Our future work is to construct an automatic facial expression system that is robust to facial poses and a partial occlusion of the facial components.

References

[1] B. Fasel and J. Luettin, "Automatic Facial Expression Analysis: A Survey," *Pattern Recog.*, vol. 36, no. 1, Jan. 2003, pp. 259-275.

[2] V. Bettadapura, "Face Expression Recognition and Analysis: The State of the Art," tech report, Apr. 2012.

[3] T. Wu et al., "Multilayer Architectures for Facial Action Unit Recognition," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 4, Aug. 2012, pp. 1027-1038.

[4] M.F. Valstar et al., "The First Facial Expression Recognition and Analysis Challenge," *Proc. FG*, Mar. 2011, pp. 921-926.

[5] L. Wiskott et al., "Face Recognition by Elastic Bunch Graph Matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, July 1997, pp. 775-779.

[6] I. Kotsia, I. Buciu, and I. Pitas, "An Analysis of Facial Expression Recognition under Partial Facial Image Occlusion," *Image Vision Comput.*, vol. 26, no. 7, July 2008, pp. 1052-1067.

[7] T. Ojala, M. Pietikainen, and D. Harwood, "A Comparative Study of Texture Measures with Classification Based on Featured Distribution," *Pattern Recog.*, vol. 29, no. 1, Jan. 1996, pp. 51-59.

[8] C. Shan, S. Gong, and P.W. McOwan, "Facial Expression Recognition Based on Local Binary Patterns: A Comprehensive Study," *Image and Vision Comput.*, vol. 27, no. 6, May 2009, pp. 803-816.

[9] T. Ojala, M. Pietikäinen, and T. Maenpaa, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, July 2002, pp. 971-987.

[10] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," *Proc. CVPR*, vol. 1, June 2005, pp. 886-893.

[11] W. Yun et al., "Face Recognition Using HOG Features," *Proc. URAI*, vol. 1, 2008, pp. 442-445.

[12] C. Shan, S. Gong, and P.W. McOwan, "Beyond Facial Expressions: Learning Human Emotion from Body Gestures," *Proc. BMVC*, 2007, pp. 1-10.

[13] M.J. Lyons et al., "Coding Facial Expressions with Gabor Wavelets," *Proc. FG*, 1998, pp. 200-205.

[14] P. Lucey et al., "The Extended Cohn-Kanade Dataset (CK+): A Complete Dataset for Action Unit and Emotion-Specified Expression," *Proc. IEEE Int. Works CVPR4HB*, 2010, pp. 94-101.

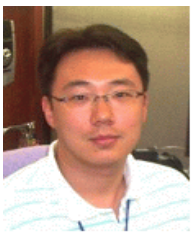
[15] T. Kanade, J.F. Cohn, and Y. Tian, "Comprehensive Database for Facial Expression Analysis," *Proc. FG*, 2000, pp. 46-53.

[16] M. Turk and A.P. Pentland, "Face Recognition Using Eigenfaces," *Proc. CVPR*, 1991, pp. 586-591.

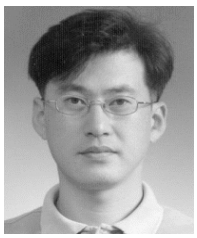
[17] C. Chang and C. Lin, "LIBSVM: A Library for Support Vector Machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, Apr.

2011, pp. 1-27.
[18] L. Maaten, E. Postma, and J. Herik, "Dimensionality Reduction: A Comparative Review," technical report, Tilburg University, 2009. http://homepage.tudelft.nl/19j49/Matlab_Toolbox_for_Dimensionality_Reduction.html
[19] M. Eckhardt, I. Fasel, and J. Movellan, "Towards Practical Facial Feature Detection," *Int. J. Pattern Recog. AI*, vol. 23, no. 3, May 2009, pp. 379-400.
[20] W. Zhang et al., "Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A Novel Non-Statistical Model for Face Representation and Recognition," *Proc. ICCV*, 2005, pp. 786-791.
[21] D.M. Hawkins and G.J. McLachlan, "High-Breakdown Linear Discriminant Analysis," *J. Am. Stat. Assoc.*, vol. 92, no. 437, Mar. 1997, pp. 136-143.

Woo-han Yun received his B.S. degree in electronics and electrical engineering from SungKyunKwan University, Suwon, Rep. of Korea, in 2004, and his M.S. degree in computer science and engineering from Pohang University of Science and Technology (POSTECH), Pohang, Rep. of Korea, in 2006. He has been a research scientist at ETRI since 2006. His current research interests include object recognition, face recognition, facial expression recognition, and human-robot interaction.



DoHyung Kim received his Ph.D. degree from Pusan National University in 2009. He has been a research scientist at ETRI since 2002. His research interests are in the area of human-robot interaction and computer vision and pattern recognition, and he mainly focuses on vision-based user recognition.



Chankyu Park received his MS degree both in electronics engineering from Kyungpook National University in 1997 and in software engineering from Carnegie-Mellon University in 2005. He joined ETRI in 1997 and has been a principal member of the engineering staff in the Intelligent Robot Research Division since 2004. He is also working toward his PhD degree in electronics at KAIST. His research interests are in computer vision, machine learning, and bio-signal processing.

Jaehong Kim received his PhD from Kyungpook National University, Daegu, Rep. of Korea, in 2006. He has been a research scientist at ETRI, Daejeon, Rep. of Korea, since 2001. His research interests include socially assistive robotics for elder care, human-robot interaction, and gesture/activity recognition.

ETRI Journal, Volume 35, Number 6, December 2013
http://dx.doi.org/10.4218/etrij.13.2013.0054