

## Recognition of Facial Expressions Based on Tracking and Selection of Discriminative Geometric Features

Deepak Ghimire<sup>1</sup>, Joonwhoan Lee<sup>2</sup>, Ze-Nian Li<sup>3</sup>, Sunghwan Jeong<sup>1</sup>, Sang Hyun Park<sup>1</sup> and Hyo Sub Choi<sup>1</sup>

<sup>1</sup>*Korea Electronics Technology Institute, Jeonju-si, Jeollabuk-do 561-844, Rep. of Korea*

<sup>2</sup>*Dept. of Computer Engineering, Chonbuk National University, Jeonju-si, Jeollabuk-do 561-756, Rep. of Korea*

<sup>3</sup>*School of Computing Science, Simon Fraser University, Burnaby, B.C., Canada*  
*deepak@keti.re.kr, chlee@jbnu.ac.kr, li@sfu.ca, shjeong@keti.re.kr,*  
*shpark@keti.re.kr, hschoi@keti.re.kr*

### Abstract

*In this paper, we present a method for fully automatic facial expression recognition in facial image sequences using feature extracted from tracking of facial landmarks. The facial landmarks at the first frame of the image sequence under examination are initialized using elastic bunch graph matching (EBGM) algorithm and tracked in the consecutive video frame over time. At first, the most discriminative geometric features in terms of triangle are selected using multi-class AdaBoost with extreme learning machine (ELM) classifier. The features for facial expression recognition (FER) are extracted from AdaBoost selected most discriminative set of triangles composed of facial landmarks. Finally, the facial expressions are recognized using support vector machines (SVM) classification. The results on the extended Cohn-Kanade (CK+) and Multimedia Understanding Group (MUG) facial expression database shows a recognition accuracy of 97.80% and 95.50% respectively using proposed facial expression recognition system.*

**Keywords:** *facial landmark elastic bunch graph matching, geometric features, multi-class AdaBoost, extreme learning machine, facial expression recognition*

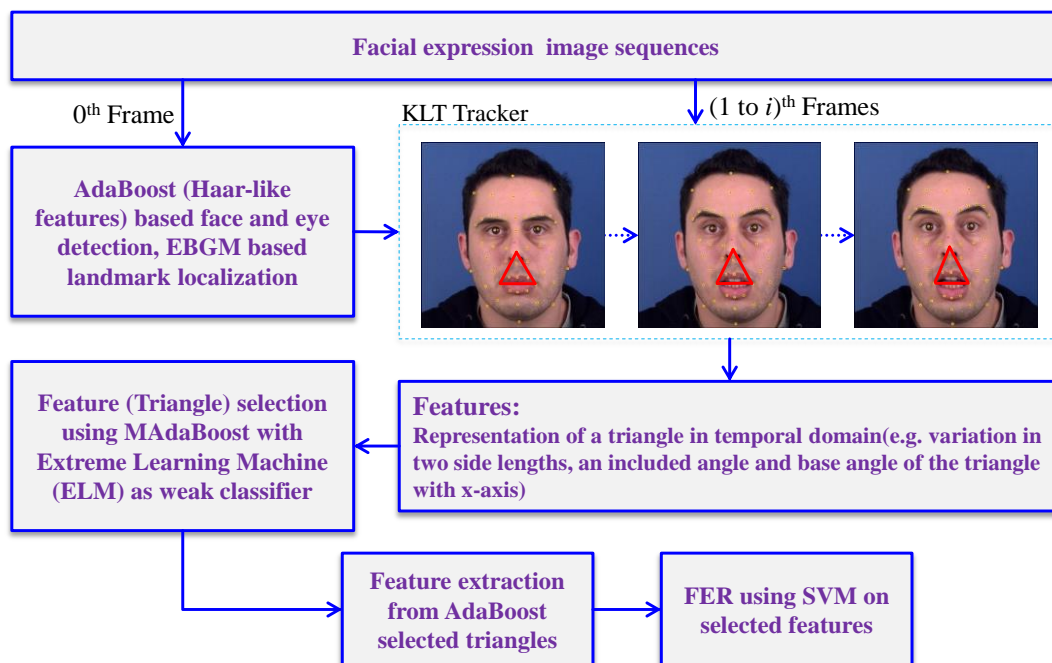
### 1. Introduction

In this paper, we address an important domain of human computer interaction (HCI), which is effective visual facial expression recognition [1]. Automatic facial expression recognition and analysis has been an active topic in the scientific community for over two decades [2]. Automated system related to facial biometrics need to be capable of adapting to particular situations. Such systems include automatic facial feature extractor and change trackers. Automated and real-time facial expression recognition would be useful in many other applications, *e.g.*, virtual reality, video-conferencing, customer satisfaction studies, *etc.* The facial expressions under examination were defined by psychologists as a set of six basic facial expressions (anger, disgust, fear, happiness, sadness, and surprise) [3].

A meta-review of the facial expression recognition and analysis challenge has recently been published by Vlaster *et al.*, [2]. The facial expression recognition approaches can be classified into two main categories, the template-based ones and the feature-based ones. 2-D or 3-D head and facial models are used as templates in template-based methods for facial expression information extraction. On the other hand the feature-based methods use geometry-based features or appearance-based features for the representation of facial expression information. Geometry-based

features describe the shape of the face and its components, such as the mouth or the eyebrow, whereas appearance-based features describe the texture of the face, caused by expression.

In the geometric feature-based approach, first, the dense set of facial feature points are localized and tracked in the facial expression image sequences. The locations of these facial landmarks are then used in different ways to extract the shape of facial features, and movements of facial features, as some facial expression evolves over time. Kotisa *et al.*, [4] uses geometric displacement of certain selected candid nodes, defined as the differences of the node coordinates between the first and the greatest facial expression intensity frames as features and the SVM classifier is used for the recognition of facial expressions. Rudovic and Pantic [5] introduce a method for head-pose invariant facial expression recognition which is based on a set of characteristic facial points extracted using AAMs. A coupled scale Gaussian process regression (CSGPR) model is used in order to normalize the head pose to the frontal pose in terms of extracted facial landmarks. Recently, in [6], geometric features based on only eight facial feature points extracted from single frame are used to recognize the facial expression. Ghimire and Lee [7] proposed a method for facial expression recognition using geometric features extracted from salient points and lines composed of facial key points in the temporal domain, which are selected using AdaBoost algorithm. Several other methods for facial expression recognition mainly based on geometric features can also be found in the literature, such as [8].



**Figure 1. Overall Block Diagram of the Proposed Facial Expression Recognition System**

In the current paper, the geometric features are extracted based on the variation in shape of the triangles composed of facial landmarks in the facial expression sequence, which are selected using multi-class AdaBoost algorithm. The facial landmarks are automatically initialized using EBGM algorithm and tracked in the consecutive frames using Kanade-Lucas-Tomaci (KLT) tracker [9]. After face graph normalization, the multi-class AdaBoost algorithm is used to select the most

discriminative set of triangles composed of facial landmarks. Rather than extracting geometric feature based on individual landmarks, the geometric feature extracted in the form of components of triangle representation in temporal domain is found to be effective for recognizing facial expressions. The features extracted in the form of components of triangles are invariant to translation, *i.e.*, in other word, it can be treated as geometric features which are invariant to head movement. Finally the facial expressions are recognized using support vector machines classification. The overall block diagram of the proposed facial expression recognition system is shown in Figure 1.

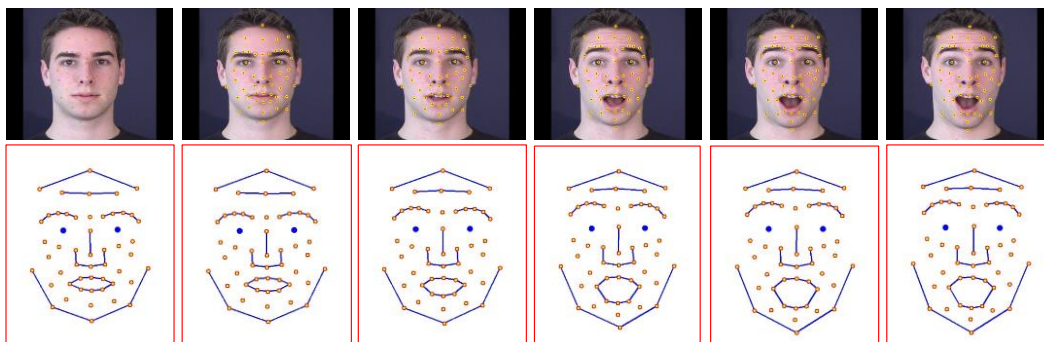
## 2. Proposed Method

The proposed facial expression recognition system is composed of three subsystems: facial feature point tracking, extracting discriminative geometric features from the landmark tracking results and classification of the extracted features into six basic facial expressions using support vector machines classification.

### 2.1. Initialization, Tracking, and Normalization of Facial Landmarks

In the proposed method, facial feature point localization in the first frame is performed using method proposed in [7] which uses EBGM algorithm [10]. Now, tracking of the 52 facial landmarks in the image sequence is performed by a pyramidal variant of the well-known Kanade-Lucas-Tomasi (KLT) tracker [9]. The KLT algorithm tracks a set of facial feature points across the video frames starting from the neutral frame to the fully expressive one.

Finally, the face graph is then normalized in such a way that for each facial expression sequence the face graph in the first frame or neutral frame is at the same location with same scale and then the graph in the subsequent facial expression sequence frames evolves according to the movement of facial feature points. In the database there is variation in face resolution and head orientation. After normalization for each expression sequence the graph for the neutral frame is the average graph of all neutral frames over the database and the graph for the subsequent image frames evolves according to the facial expression. Figure 2 shows the result of the facial feature point tracking as well as result after normalization.

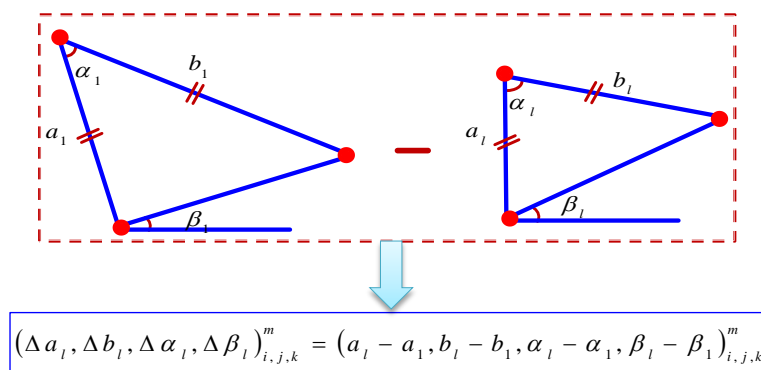


**Figure 2. Examples of Facial Feature point Tracking and Corresponding Result after Normalization for Surprised Facial Expression Sequences**

## 2.2. Triangle based Geometric Feature Extraction

In [7, 8], authors extracted the geometric features based on tracking the coordinates of an individual landmarks as well as tracking the coordinates of pair of landmarks. Basically those features are extracted based on points and lines. In the current paper, we assume three facial landmarks at a time and features are extracted in the form of triangle representation parameters in the temporal domain. The information regarding movement of facial landmarks and relationship between them while some facial expression evolves over time can be captured well by considering three landmarks at a time as compared to one or two facial landmarks. Triangle components in the  $l^{\text{th}}$  frame are subtracted with the triangle components in the first frame of the video sequence as shown in Figure 3.

Triangle components in the  $l^{\text{th}}$  frame are subtracted with the triangle components in the first frame of the video sequence. Suppose there are  $N$  frames in the sequence, then the feature vector is composed of  $(N-1) \times 4$  components, *i.e.*, if  $N = 11$ , feature dimension for a sequence extracted from the single triangle will be  $(11-1) \times 4 = 40$ .



**Figure 3. Difference in Components of Two Triangles used as Features. Vertex of Each Triangle Corresponds to the Landmarks of the Face in the Two Frames of the Video Sequence**

## 2.3. Feature Selection using Multi-class AdaBoost

The goal of this paper is first to select the most discriminative set of geometric features and use those subset of features for recognizing the six basic facial expressions. The geometric features are extracted from the triangles composed of facial landmarks. We tracked 52 facial landmarks as in [7], therefore in the feature pool, there are total of  $52! / (3!(52-3)!) = 22100$  unique triangles. Feature extracted from only a subset of those triangles will sufficiently carry the discriminative information required for efficiently recognizing the facial expressions. Therefore the feature selection scheme should select only those set of triangles which carry most discriminant information for the recognition of facial expressions.

The AdaBoost learning algorithm proposed by Freund and Schapire [11], in its original form, is used to boost the classification performance of a simple learning algorithm. AdaBoost algorithm is not only used for classification, it can also be used for feature selection. In the proposed system, multi-class AdaBoost proposed by Jhu *et al.*, [12] is used to select the triangles from which features will be extracted for facial expression recognition.

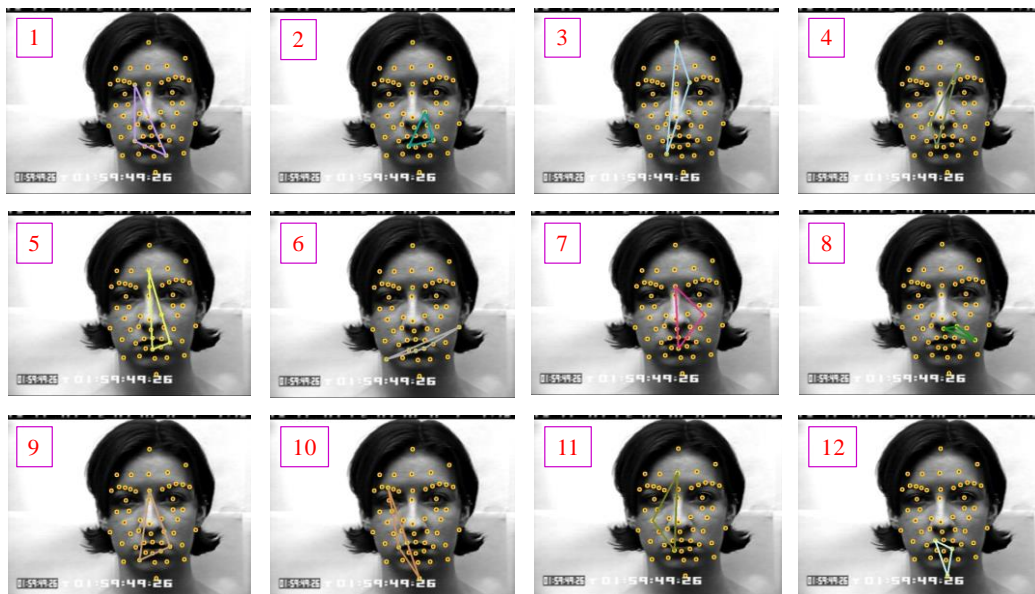
Extreme learning machine (ELM) [13] is used as a weak classifier in the AdaBoost algorithm. The ELM itself is not a weak classifier, but in our system, in terms of feature it is treated as a weak classifier, this is because each ELM will be trained using feature extracted from single triangle. ELM solves the problem of

gradient-based learning algorithm by analytically calculating the optimal weights for the SLFN. First the weights between the input layer and the hidden layer are randomly selected and then the optimal values for the weights between hidden layer and output layer are determined by calculating the linear matrix equations. The reason behind selecting ELM as a weak classifier is that, it is very fast learning algorithm, it can train almost in real time. In our feature selection scheme we need to train 22100 ELMs. In summary the ELM algorithm can be written as follows:

**ELM Algorithm:** Given a training set  $\{(x_i, t_i) \mid x_i \in R^n, t_i \in R^m, i = 1, \dots, N\}$ , hidden node output function  $g(w, b, x)$ , and number of hidden nodes  $L$ ,

- a. Randomly assign hidden node parameters  $(w_i, b_i)$ ,  $i = 1, \dots, L$ .
- b. Calculate the hidden layer output matrix  $H$ .
- c. Calculate the output weights  $\beta : \beta = H^\dagger T$ .

where  $H^\dagger$  is the Moore-Penrose generalized inverse of hidden layer output matrix  $H$ .



**Figure 4. The First Twelve Triangle Feature Selected using Multi-class AdaBoost with ELM as a Weak Classifier for CK+ Facial Expression Datasets**

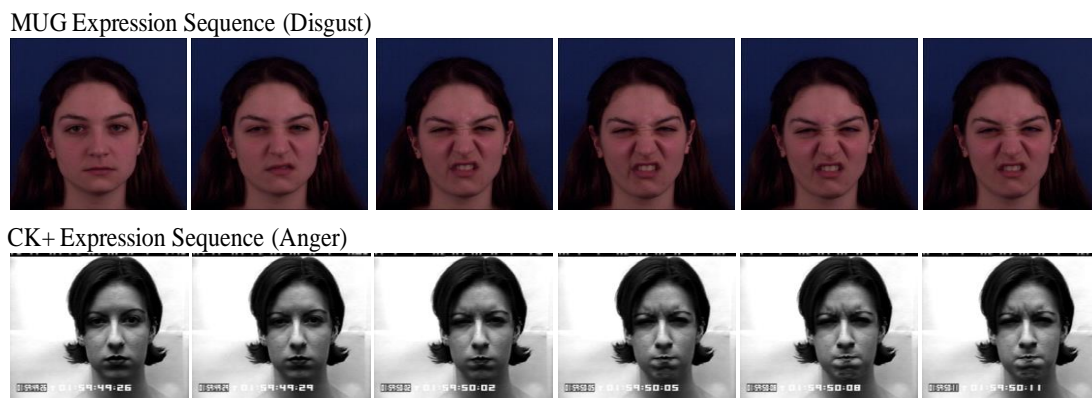
Figure 4 shows the first twelve features (triangles) selected by multi-class AdaBoost. The triangle is selected as a feature. As shown in fig. 3 in feature selection process, the geometric feature from the triangle composed of three facial landmarks consist of difference in two side lengths of triangles, difference in an included angle, and difference in base angle of the one of the side length of triangles in the neutral frame (first frame) and  $i^{th}$  frame in the video sequence. Mostly, the AdaBoost selected triangles are composed of landmarks from the eyebrow, mouth and nose position.



### 3. Experimental Results

#### 3.1. Dataset Description

To access the reliability of our approach, we evaluate the performance of the proposed facial expression recognition system on two databases: extended Cohn-Kanade dataset (CK+) [14] and Multimedia Understanding Group (MUG) dataset [15]. Out of 593 image sequences, 315 image sequences from CK+ database are used in our study, because only those sequences have a given emotional class. Again, 324 facial expression image sequences from MUG dataset are used for the evaluation of the proposed facial expression system. An example of the facial expression sequence from each dataset is given in Figure 6. Table 1 shows the number of facial expression image/video sequence for each expression and from each dataset used in our system.



**Figure 5. An Example of Facial Expression Sequence from Different Dataset**

**Table 1. Number of Facial Expression Sequence in Each Dataset**

Dataset/Expression	Anger	Disgust	Fear	Happiness	Sadness	Surprise	Total
CK+	44	62	27	69	32	81	315
MUG	56	55	51	55	51	56	324

The most usual approach for testing the generalization performance of a classifier is the  $K$ -fold cross validation approach. A five-fold cross validation is used in order to make maximum use of the available data. The results reported below are the averaged classification accuracy results from five-fold cross validation. The confusion matrices are given in order to show the better picture of the recognition accuracy of each expression type. The diagonal entries of the confusion matrix are the rates of facial expressions that are correctly classified; with the off-diagonal entries correspond to misclassification rates.

#### 3.2. Facial Expression Recognition using SVM with Boosted Features

SVM is a well-known classifier for its generalization capability. SVM classifiers maximize the hyper plane margin between classes. The publically available implementation of SVM, *libsvm* [16] is used for the classification of the facial expressions. The optimal parameter selection is performed based on the grid search strategy [17], in which radial basic function (RBF) kernel are used.

As shown in Figure 3 the features for SVM classification are extracted by subtracting triangle components composed of facial landmarks. To keep feature

dimensionality as low as possible only the maximum changes in magnitude of the four components of the triangle in the sequence with respect to the triangle components in the first frame are extracted. Therefore each triangle is composed of four features, but some triangles in the AdaBoost selected triangle set shares the common edge, therefore the total feature dimension is not equal to the number of triangles multiplied by four. The included angle value of triangle is always unique, whereas two side lengths and base angle could be in common with the other triangle in the triangle set.

To get better picture of the recognition accuracy of each expression type, the confusion matrices are given. Table 2 and 3 shows the confusion matrices for the facial expression recognition using features extracted from 160 and 98 AdaBoost selected triangles in CK+ and MUG facial expression datasets respectively. The dimensionality of the feature vector using 160 and 94 triangles in CK+ and MUG dataset is 370, 330 respectively. The average recognition accuracies are 97.80% and 95.50% respectively, which are the maximum accuracies obtained using proposed facial expression recognition system.

**Table 2. Confusion Matrix for Facial Expression Recognition in Percentage using SVM with Boosted Features in CK+ dataset (160 AdaBoost Selected Triangular Features)**

	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	<b>97.5</b>	0	0	0	2.5	0
Disgust	3.33	<b>96.67</b>	0	1.67	0	0
Fear	0	0	<b>96</b>	4	0	0
Happiness	0	0	0	<b>100</b>	0	0
Sadness	3.33	0	0	0	<b>96.67</b>	0
Surprise	0	0	0	0	0	<b>100</b>

**Table 3. Confusion Matrix for Facial Expression Recognition in Percentage using SVM with Boosted Features in MUG Dataset (98 AdaBoost Selected Triangular Features)**

	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	<b>100</b>	0	0	0	0	0
Disgust	0	<b>100</b>	0	0	0	0
Fear	2.5	0	<b>85</b>	0	7.5	5
Happiness	0	0	0	<b>100</b>	0	0
Sadness	6	2	2	0	<b>90</b>	0
Surprise	0	0	2	0	0	<b>98</b>

In CK+ dataset almost all the facial expressions are recognized with high accuracy. In MUG dataset fear and sadness expressions are confused with other expressions, whereas anger, disgust, happy and surprise expressions are recognized with high accuracies.

The recognition accuracy achieved by the proposed method is comparable with the best accuracies in the literature. Recently in CK+ dataset, [8] and [7] achieved 97.34% and 83.01 % recognition accuracy, respectively. The dimensionality of feature vector in [8] is larger than in our case, whereas in [7] they used only 8 facial landmarks from the single highly expressed facial expression frame. Y. Rahulamathavan *et al.*, [18] achieved 95.24% of overall recognition accuracy in MUG facial expression database. They performed facial expression recognition on images in the encrypted domain, based on the local fisher discriminant analysis (LFDA). We achieved average recognition rate of 97.80% and 95.5% in CK+ and

MUG facial expression databases respectively with relatively lower dimension of geometric feature vectors. Therefore, one of the advantages of the proposed geometric feature based facial expression recognition system is that relatively lower dimension of features are extracted compared to the feature dimension in the state of the art methods. The recognition accuracy in CK+ and MUG database is comparable or even better with the best recognition accuracy in the literature.

#### 4. Conclusions

We have proposed a method to perform facial expression recognition in image sequences using geometric features extracted from triangles composed of facial landmarks. The EBGM algorithm is used for automatic localization of facial landmarks and then KLT tracker is used to track those landmarks in the consecutive frames in the video sequence. The triangular features are selected from the large set of triangles composed of facial landmarks in the temporal domain using feature selective multi-class AdaBoost with extreme learning machine. The extracted geometric features are somehow invariant to the head movement. Experiments on the CK+ and MUG facial expression databases show the proposed method can recognize facial expressions with high accuracies. In overall, the proposed facial system can achieve comparable or even better result of facial expression recognition as compared with the result of best methods in the literature.

#### References

- [1] M. Pantic, A. Pentland, A. Nijholt, and T. Huang, Human computing and machine understanding of human behaviour: A survey, Lecture Notes in Artificial Intelligence, vol. 44512007, (2007).
- [2] M.F. Vlaster, M. Mehu, B. Jiang, M. Pantic and K. Scherer, "Meta-analysis of the first facial expression recognition challenge", IEEE Trans. Syst. Man. Cybern. B. Cybern., vol. 42, (2012), pp. 966-979.
- [3] P. Ekman, "Strong evidence of universal in facial expressions: A reply to Russell's mistaken critique", Psychol. Bull., vol. 115, (1994), pp. 268-287.
- [4] I. Kotisa, I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines", IEEE Trans. on Image Processing, vol. 16, (2007), pp. 172-187.
- [5] O. Rudovic, M. Pantic and I.Y. Patras, "Coupled Gaussian process for pose-invariant facial expression recognition", IEEE. Trans. Pattern Anal. Mach. Intell., vol. 35, (2013), pp. 1357-1369.
- [6] A. Saeed, A. Al-Hamadi, R. Niese and M. Elzobi, "Frame-based facial expression recognition using geometrical features", Advances in Human-Computer Interaction, (2014), pp. 1-13.
- [7] D. Ghimire and J. Lee, "Geometric feature-based facial expression recognition in image sequences using multi-class AdaBoost and support vector machines", Sensors, vol. 13, (2013), pp. 7714-7734.
- [8] D. Ghimire and J. Lee, "Automatic facial expression recognition based on features extracted from tracking of facial landmarks", Fifth Int. Conf. on Graphics and Image Processing, (2013); Hong Kong, China.
- [9] J. Y. Bouguet, "Pyramidal implementation of the Lucas-Kanade feature tracker", *Tech. Rep.*, Intel Corp., Microprocessor Research Lab, (1999).
- [10] L. Wiskott, J.-M. Fellous and N. Krüger, "Face recognition by elastic bunch graph matching", IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, (1997), pp. 775-779.
- [11] Y. Freund and R.E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting", J. Comput. Syst. Sci., vol. 55, (1997), pp. 119-139.
- [12] J. Zhu, H. Zou, S. Rosset and T. Hastie, "Multi-class AdaBoost, Statistics and Its Interface", vol. 2, (2009), pp. 349-360.
- [13] G.B. Huang, Q.-Y. Zhu, C.-K. Siew, "Extreme learning machine: theory and applications", Neurocomputing, vol. 70, (2006), pp. 489-501.
- [14] P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, "The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specific expression", Proc. of the 3<sup>rd</sup> IEEE Workshop on CVPR for Human Communication Behaviour Analysis, (2010); San Francisco, CA, USA.
- [15] N. Aifanti, C. Papachristou, A. Delopoulos, "The MUG facial expression database, 11<sup>th</sup> Int. Workshop on Image Analysis for Multimedia Interactive Services", (2010); Desenzano del Garda, Italy.
- [16] C.-C. Chang, C.-J. Lin, LIBSVM: A library for support vector machines. Available online: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/> (2014).
- [17] C.-W. Hsu, C.-C. Chang and C.J. Lin, "A practical guide to support vector classification, Technical Report; Department of Computer Science", National Taiwan University, (2010).



- [18] Y. Rahulamathavan, R. C.-W. Phan, J.A. Chambers and D.J. Parish, "Facial expression recognition in the encrypted domain based on local fisher discriminant analysis", IEEE Trans. on Affective Computing, vol. 4, (2013), pp. 83-92.

## Authors



**Deepak Ghimire** received undergraduate degrees in Computer Engineering from Pokhara University, Nepal in 2007. He received M.S. degree and Ph.D. degree in Computer Science and Engineering from Chonbuk National University, South Korea in 2011 and 2014 respectively. He is currently a Researcher in IT Application Research Center of KETI (Korea Electronics Technology Institute), South Korea. His main research interests include image processing, computer vision, machine learning, and biometric information processing.



**Joonwhoan Lee** received undergraduate degrees in Electronic Engineering from the Hanyan University in 1980. He received M.S. degree in Electrical and Electronic Engineering from KAIST (Korea Advanced Institute of Science and Technology) University in 1982 and Ph.D. degree in Electrical and Computer Engineering from University of Missouri, USA in 1990. He is currently a Professor in the Department of Computer Engineering at Chonbuk National University, South Korea. His research interests include image processing, computer vision, emotion engineering.



**Ze-Nian Li** is a Professor in the school of Computing Science at Simon Fraser University, British Columbia, Canada. He received his undergraduate education in Electrical Engineering from University of Science and Technology of China, and M.Sc. and Ph.D. degree in Computer Sciences from the University of Wisconsin-Madison under the supervision of the late Professor Leonard Uhr. His current research interests include computer vision, multimedia, pattern recognition, image processing, and artificial intelligence.



**SungHwan Jeong** received undergraduate degrees in Computer Engineering from Jeonju University, South Korea in 2004. He received M.S. degree in Biomedical Engineering and Ph.D. degree in Computer Science and Engineering from Chonbuk National University, South Korea in 2006 and 2012 respectively. He is currently a Researcher in IT Application Research Center of KETI (Korea Electronics Technology Institute), South Korea. His current research interests include image processing, computer vision, embedded vision engineering.



**Sang Hyun Park** received his undergraduate education in Computer Science Engineering from Hankuk University of Foreign Studies, South Korea in 2000. He received M.S. degree in Computer Science Engineering from Hankuk University of Foreign Studies, Korea in 2002. Now he is Managerial Researcher in Jeonbuk Embedded System Research Center of KETI (Korea Electronics Technology Institute), Korea. His current research interests include automotive network, embedded system engineering.



**Hyo-Sub Choi** received undergraduate degree in Electronics Engineering from Kwangwoon University, Seoul, Korea, in 2004, and M.S. degree in Information Communications Engineering from Gwang-ju Institute Science and Technology (GIST), Gwang-ju, Korea, in 2006. From 2006 to 2009, he was a Researcher with the Samsung Electronics R&D Institute. He is currently working in Korea Electronics Technology Institute (KETI). His research interests include vehicle network communication and system.