International Journal of Security and Its Applications
Vol.8, No.5 (2014), pp.203-212
http://dx.doi.org/10.14257/ijsia.2014.8.5.19

Robust Emotion Recognition Algorithm for Ambiguous Facial Expression using Optimized AAM and *k*-NN

Yong-Hwan Lee, Wuri Han, Youngseop Kim and Cheong-Ghil Kim

Dept. of Smart Mobile, Far East University, Chungbok, Korea
Dept. of Electronic Engineering, Dankook University, Chungnam, Korea and
Dept. of Computer Science, Namseoul University, Chungnam, Korea
hwany1458@empal.com, genasix1@gmail.com, wangcho@dankook.ac.kr and chkim@nsu.ac.kr

Abstract

Analysis of human emotion plays an important role in interaction between human and machine communication. The most expressive way to extract and understand of human emotion is by facial expression analysis. This paper proposes a novel recognition method of multiple emotions from facial expression running on mobile environments. Especially, we formulate the classification model of facial ambiguous emotions using a variance of the estimated facial feature points. First, we extract 65 landmark points from input stream using active appearance model, and we then analyze the changes of the values of the feature points to recognize a facial emotion by comparing with fuzzy k-NN classification. Finally, five types of the emotions are recognized and classified as a facial expression. To evaluate the proposed approach, we assess the ratio of success with iPhone camera views, and we achieve the best 93% accuracy in the experiments. The results show that the proposed method performed well in the recognition of facial emotion on mobile environments, and the implementation system can be represented by one of the example for augmented reality on displaying combination of real face video and virtual animation with user's avatar.

Keywords: *Emotion Recognition, Ambiguous Facial Expression, Classification Model, AAM (Active Appearance Model)*

1. Introduction

Analysis and Recognition of human facial expression and emotion have attracted a lot of interest in the past few decades, and this has been researched extensively in neuroscience, cognitive science, computer science and engineering. These researches focus not only on improving human-computer interface, but also on improving the actions which computer takes on feedback by a user. While feedback from user traditionally has been occurred by using keyboard or mouse, however in the recent, smart-phone such as iPhone with advanced camera enables the computer to see and track the user's activities. This should replace the traditional interaction, and lead that a user can easily utilize an intelligent interaction [1]. Generally, human interacts with each other not only by speech, but also with gesture, to emphasize a certain part of the speech, and to display of human emotions. User's emotions can be displayed in a number of ways visually via facial expressions, vocally, either through word choices or by using non-verbal sounds, and by other physiological means such as gestures [2]. Of these the most natural way to display emotions is the use of facial expressions, which in terms of this research are primarily conveyed by video sequences [3].

ISSN: 1738-9976 IJSIA

This paper proposes a novel extraction and recognition method of facial expression and emotion from mobile video stream. Mainly, we formulate a classification model of facial emotions using a variance of the estimated landmark points. For extracting landmark points of user's face, we extract 65 feature points from entered image stream with optimized active appearance model, and then variances of the point locations are used to recognize a facial emotion by comparing a mean shape model with fuzzy *k*-NN classification. Finally, five types of facial emotion are recognized and classified as a facial expression in happy, angry, surprise, sad and neutral.

This paper is organized as follows: Section 2 reviews related works about emotion recognition through facial expression analysis and background knowledge of active appearance model (AAM). Section 3 presents the proposed system that uses two main steps to recognize and classify the facial emotion from input sequences. Experimental results and discussion are shown in Section 4, and Section 5 summarizes this study and its future works.

2. Related Works

Many papers have been devoted to automatic analysis and recognition of human emotions [4]. Research on recognizing emotion by facial expression was pioneered by Ekman [5], who started their work from the psychology perspective. Cohen proposed architecture of Hidden Markov Models for automatically segment and recognize human facial expression from video sequences [2]. Yoshitomi investigated a method for facial expression recognition for human speaker by using thermal image processing and a speech recognition system [6]. He improved speech recognition system to save thermal images, just before and just speaking the phonemes of the first and last vowels, through intentional facial expression of five categories with emotion. Ioannou suggested an extraction of appropriate facial features and consequent recognition of the user's emotional state that could be robust to facial expression variations among different users [7]. He extracted facial animation parameters defined according to the ISO MPEG-4 standard by appropriate confidence measures of the estimation accuracy. However, most of these researches are used a method to recognize emotion based on the extracted frames of input videos. Several prototype systems were developed that can recognize deliberately produced action units in either frontal face view images [8] or profile view face images [9]. These systems employ different approaches including expert rules and machine learning methods such as neural networks, and use either feature-based image representation or appearance-based image representation. Valstar employed probabilistic, statistical and assemble learning techniques, which seem to be particularly suitable for automatic action unit recognition from face image sequences [10]. Bruce suggested that development of facial emotion recognition depends on task demands [11]. When children needed to point to which of two faces was happy, sad, angry, or surprised, they achieved nearly perfect accuracy by 6 years of age.

3. Active Appearance Model

Active Appearance Model (AAM) is a computer vision algorithm, used template matching a statistical model for building shape and appearance of facial features, by automatically locating landmark points that define shape and appearance of objects in an image [12]. Edwards, *et al.*, first introduced an AAM in 1988 [13], and this is widely used in facial expression analysis and medical image interpretation. Since AAM combines a powerful model of joint shape and texture with a gradient-descent fitting algorithm [14], it provides a better matching for image texture, and has a more robust method for tracking facial movements and appearance than alternative active shape model (ASM) [15].

3.1. Shape Model

Shape is a form of geometric information that is stable across an image class. Mathematically, shape, defined by *n* landmark points in *k* dimensional space, is represented by *nk* vectors. In 2-D images, *n* landmarks $\{(x_i, y_i): i = 1, \cdots, n\}$ define 2*n* vectors (*k*=2), as shown in Equation 1.

$$x = (x_1, y_1, x_2, y_2, \cdots, x_n, y_n)^T \tag{1}$$

To obtain statistical validity, it is important that all shapes are represented in terms of the same referential space. The effects of location, scale and rotation can be removed, and a Generalized Procrustes Analysis (GPA) is performed to place all shapes in a common frame. GPA consists of sequentially aligning pairs of shapes using the mean shape, and this is performed repeatedly until the mean shape no longer changes significantly within iterations. The aligned shape is then recomputed using Equation 2.

$$\overline{x_k} = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{2}$$

Then, PCA (Principal Components Analysis) is performed to reduce data dimensions by searching for the direction in the data with the largest variance and by projecting the data onto the direction. This is served as basis for the data and each point x_i can then be calculated as sum of the mean and orthogonal linear transformation as shown in Equation 3.

$$x_i = \bar{x} + \sum_{i=1}^{t} \emptyset_i b_i \tag{3}$$

where $\bar{x}$ is the mean shape vector, and $\emptyset_i$ denotes the shape parameters. Figure 1 shows 65 landmark points and their corresponding numbers.

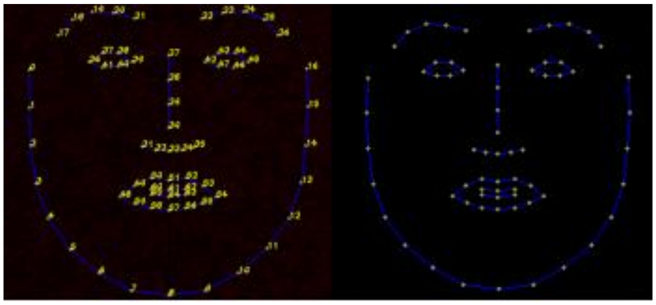

Figure 1. Shape Model by Using 65 Landmark Points

3.2. Appearance Model

Building a full model from a facial image is required to both a shape model and a texture model. The next step therefore is to build a statistical texture model, which requires an alignment of all texture samples to a reference texture frame following a similar procedure to that used for the shape model.

Appearance is composed of the pixel intensities across modeled entity of the target image, known as texture information. Since building a statistical appearance model requires warping of the color channels, control points are first matched to the mean shape. A piece-wise affine warping (for example, partitioning the convex outlines of the mean shape using a set of triangles) is performed for the texture matching. Delaunay triangulation is used to establish triangle meshes that can then be used to map the appearance vectors, after which each pixel inside a triangle is mapped into the correspondent triangle in the mean shape. Figure 2 shows a typical mesh produced by Delaunay Triangulation for facial image.

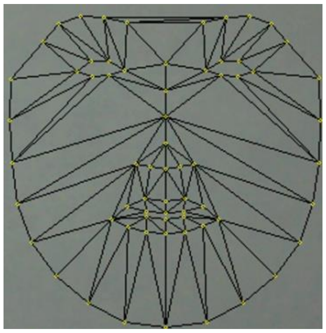

Figure 2. Delaunay Triangulation Mesh for Facial Image

Appearance model *A*(*x*) is then obtained by applying a PCA to the texture vectors, as Equation 4.

$$A(x) = A_0(x) + \sum_{i=1}^{m} \lambda_i A_i(x) \qquad (4)$$

where $A_0(x)$ represents the mean appearance vectors, λ_i is the appearance parameters, and $A_i(x)$ is the synthesized appearance vectors from the affine warping.

3.3. AAM Fitting

The parameters of generative model are necessary for estimating to fit the model to the target image. Fitting procedure is done by minimizing any error measure between appearance model and target image. The sum-of-squared error of all positions *a* is minimized to find the parameters *p*, as shown in Equation 5.

$$\underset{p}{\operatorname{argmin}} \frac{1}{2} \sum_{x \in A_0} [e(x,p)]^2 \qquad (5)$$

The error *e* at position $x \in A_0$ is computed as $e(x,p) = \alpha_\lambda(x) - (I(W(x,p)))$, where $W(x,p)$ is a non-linear warping function, and α_λ denotes the variations in the texture. An iterative style is commonly used to estimate the parameter *p*. Let assume that the current estimated parameter is *p*, and the incremental updated parameter is Δp in each iteration, the update can be calculated with the previous estimation, and simply illustrated as $\mathrm{p} \leftarrow \mathrm{p} + \Delta p$. Equation 6 shows the minimization with respect to Δp. The iteration breaks down, if the error is less than pre-determined value or there are no longer changes.

$$\underset{\Delta p}{\operatorname{argmin}} \frac{1}{2} \sum_{x \in A_0} [e(x, p + \Delta p)]^2 \qquad (6)$$

4. Proposed Emotion Recognition Approach

In this paper, we propose a new emotion recognition method based on the changes and movements of feature landmarks with human facial expressions. The proposed recognition system is composed of two major tasks: the first is detection of facial area and extraction of feature points from original input, and the second is classification and verification of facial features characteristically involved in expressing emotion. A diagram of the proposed algorithm is depicted in Figure 3.

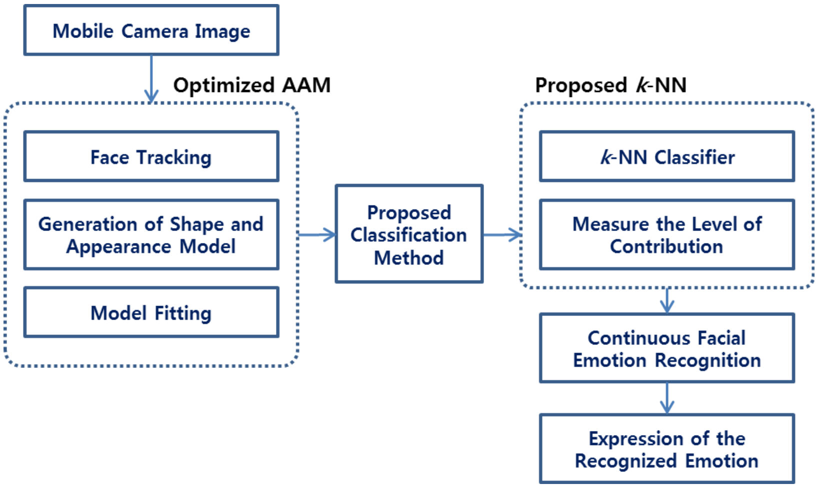
Mobile Camera Image
Optimized AAM
Face Tracking
Generation of Shape and Appearance Model
Model Fitting
Proposed Classification Method
Proposed k-NN
k-NN Classifier
Measure the Level of Contribution
Continuous Facial Emotion Recognition
Expression of the Recognized Emotion

Figure 3. Process Flow of the Proposed Method

4.1. Emotion Classification Method

To identify facial emotions, the proposed scheme requires the estimated landmark positions, provided by AAM, as well as any changes in those landmark points based on the coordinates assigned in the video frames.

This paper classifies facial emotions into five basic types; happy, sad, surprise, angry and neutral. The facial expression associated with a particular emotion is generally indicated by different characteristics and variations in facial features. We identify the type of emotions based on a combination of variances in one or more of the features commonly involved in expressing emotions, actually the eyebrows, eyes, mouth and center of eyebrow. These key areas are assigned higher weight. The new proposed model highlights not only differences in the landmarks that take place between the current frame and the previous frame, but also the correlation between the various landmark points. To establish the classification criteria, we compute the number of variations, such as distances between two landmarks, angles among three feature points, and areas of the triangles in the mesh.

Each characteristic of emotions has any rules for emotion. The happy emotion has commonly a small amount of variation in the eyebrows, and typically raises the corners of the mouth, and spreads the lips. The angry emotion is different from person in the changes of mouth shape, however has commonly a certain characteristics such as a frown. The sad emotion has similar to the angry, but has a smaller frown than the angry, and goes down on the mouth corners. The surprise emotion has the largest changes in the facial area, and raises eyebrows, as well as opens the mouth.

To build these criteria, favorable rules for emotional classification are defined, and we then estimate a variance and movement of the feature points by changing facial emotions. Table 1 shows summaries of facial emotions with the variance of the landmark points.

Table 1. Variance and Movement of the Landmark Points with Facial Emotions

Emotions	Characteristics	Movements of the landmark points
Happy	Eye opening is narrowed	38(*y*), 40(*y*), *d*(38,40)
	Mouth is opening	*d*(48,54), *d*(61,64)
	Lip corner goes up	∠(31,1,48), 48(*x*,*y*), 54(*x*,*y*), 49(*x*,*y*), 53(*x*,*y*)
Angry	Eyebrows goes down	∠(22,16,1), 21(*y*), 22(*y*)
	Eyebrows are centering	21(*x*), 22(*x*)
	Mouth is slightly opening	□(60,62,65,63), *d*(61,64)
Surprise	Upper eyelid raiser	∠(22,16,1), 21(*y*), 22(*y*)
	Eye is opening	38(*y*), 40(*y*), *d*(38,40)
	Mouth is opening	□(60,62,65,63), *d*(61,64)
Sad	Lip corner goes down	∠(31,4,48), 48(*y*), 54(*y*)
	Eyebrows go down	∠(22,16,1), 21(*y*), 22(*y*)
	Mouth stretches	61(*x*,*y*), 64(*x*,*y*), *d*(61,64), □(60,62,65,63)
	Eye is slightly closing	38(*y*), 40(*y*), *d*(38,40)

To minimize an error according to the variations of each landmark by changes in face pose, every point are re-arranged to center of the nose, which is showing non-changes with the facial expression. Since location and configuration of the estimated landmarks have a difference for a user, neutral expression is first entered to extract an initial feature points from camera view. The variances of the facial expressions from neutral emotion (*i.e.*, initial estimated points) are then measured in the face at every frame. Figure 4 illustrates major angles and the connecting points, used in the classification of the facial emotion.

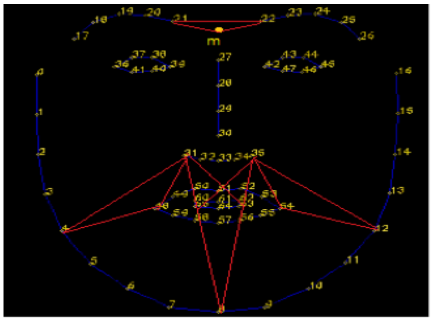

Figure 4. Key Angle Variation of Triangle Mesh Connections Among the Feature Points

The facial expression analysis by using typical variances about five basic emotions is shown in Figure 5. Independent characteristic which plays an important role in multiple emotion recognition is constructed by the proposed amount of variance of facial expression.

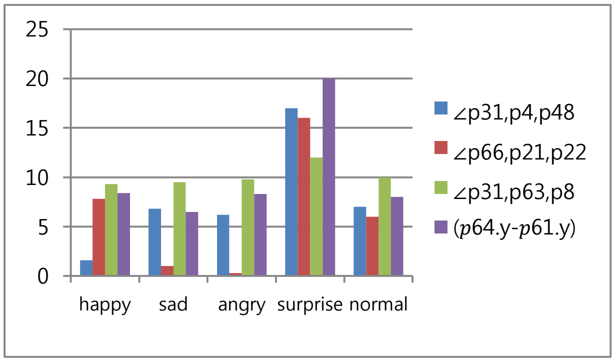

Figure 5. Independent Characteristic of Each Emotion with Variation

4.2. Classifier based on Fuzzy *k*-Nearest Neighbor

k-Nearest Neighbors (*k*-NN) algorithm is a non-parametric method for classification and regression that predicts objects values or class memberships based on the k closest training data in the feature space [11]. The fuzzy *k*-NN classifier assigns a class membership to sample vector, rather than assigning the vector to a particular class, while NN classifier requires to pre-processing of the labels sample set prior to their use.

This paper proposes a modified k-NN classifier that can recognize an ambiguous emotion with combinations of five basic emotions. The proposed *k*-NN classifier measures a level of contribution between each emotion data and input data. To determine the contribution of the individual class, a weighted distance is calculated on each neighbor's contribution to the membership value which uses Mahalanobis distance, according to the level of contribution in the range of *k*, the recognized emotion is classified with first, second and third level of the emotion. Then, the weighted contribution of each neighboring point is computed by the reciprocal of its distance from the point being classified, as Equation 7.

$$\mu_{cn} = \frac{(x - m_c)^T \sum_c^{-1}(x - m_c)}{\sum_{c=1}^{n}(x - m_c)^T \sum_c^{-1}(c - m_c)} \qquad (7)$$

where n is the number of classes, $\sum_c$ is the covariance matrix of the cth class, and m_c is the mean vector of the cth class.

Then, the weighted contribution values can be shown by combined emotions up to a maximum of three emotions, and the level of combination of each emotion is displayed with their values of percent.

5. Experimental Results

In this paper, facial expressions have been recognized from mobile video sequence, and experiments are performed on iPhone5. The proposed algorithm presented in the previous section is implemented with Objective-C and Xcode. Performance Evaluation of the proposed method is performed with five types of facial emotion, such as happy, sad, angry, surprise and neutral. The pose of the face is restricted to only in-front of the camera view, because of extracting the feature points over two eyes and eyebrows.

Simulation process in this study is carried out through the following steps: first, expression shape vectors on the current AAM fitting are compared to the previous frames. Then, movements of the landmark points are estimated with the prcrustes alignment. After measuring the variance of the facial features, mean shape vectors are computed on each of facial emotions. Then the weighted contribution of facial is calculated with fuzzy *k*-NN classifier. Finally, facial emotion is classified and recognized with *k* nearest nodes.

Table 2. Emotion Recognition Rate in the Experiments Running on Mobile Device

Emotion	Happy	Sad	Surprised	Angry	Neutral
Happy	**78%**	3%	81%	20%	72%
Sad	3%	**70%**	1%	78%	42%
Surprised	81%	1%	**93%**	76%	92%
Angry	42%	78%	76%	**74%**	26%
Neutral	62%	15%	92%	26%	**73%**

The emotions which have the clear characteristics such as happy, surprise, sad or angry show a comparatively elevated recognition rate. However, sad which has none clear characteristics reveals the lowest ratio of recognition, of 77% in the basic emotions. The recognition rate of complex emotions shows an average of approximately 45%. Irrelevant emotions such as happy-to-sad and surprise-to-sad show the lowest recognition rate.

Figure 6 shows some examples of screenshot of the implemented system. The system could easily convey the intended results with an avatar and animation of emotion effect.

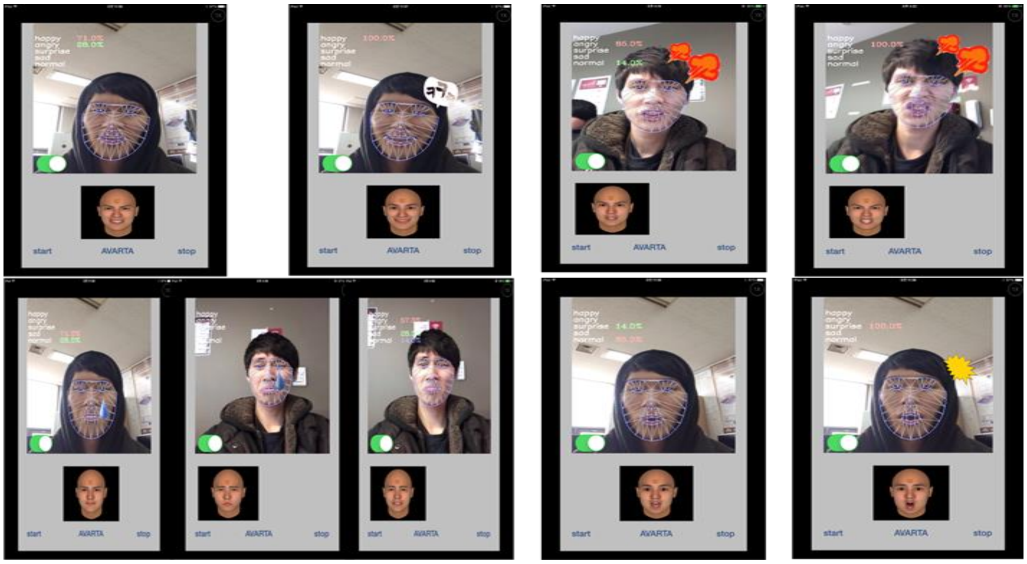

Figure 6. Screenshot of the Implemented System: Facial Emotion Recognition, Running on iPad and iPhone

6. Conclusion

This paper proposes and implements a novel emotion recognition method by continuous facial expression analysis. As well as, we formulate the classification method of facial emotion using a variance of the estimated landmark points in five types of the facial emotions like happy, sad, surprise, angry and neutral. The proposed algorithm is performed two major steps: one is a detection and extraction of facial features with active appearance model from mobile camera sequences. Another is a classification and verification of facial emotion of characteristic features by fuzzy k-NN classifiers. The simulation is performed on iPhone5 and iPad mini, and the experimental results show average successful ratio of recognition is good enough to apply to mobile devices.

The main contribution of this paper lies in the design of fuzzy k-NN classification method for the facial emotion recognition, and estimates the performance of classifier for the use of mobile emotion recognition based on AAM. As for future works, we have a plan to interact between user and machine by the recognized emotion to enable a control of the system through the face emotion.

Acknowledgements

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (NRF-2013R1A2A2A03068794).

References

[1] X. Yan and N. Aimaiti, "Gesture-based Interaction and Implication for the Future", Master Thesis, Department of Computing Science, Umea University, Sweden, **(2011).**

[2] I. Cohen, A. Garg and T. S. Huang, "Emotion Recognition from Facial Expressions using Multilevel HMM", Neural Information Processing Systems, **(2000)** .

[3] N. Sebe, I. Cohen, T. Gevers and T. S. Huang, "Multimodal Approaches for Emotion Recognition", A Servey. Proceedings of SPIE – International Society for Optical Engineering, **(2005).**

[4] S. Shinde and S. Pande, "A Survey on: Emotion Recognition with respect to Database and Various Recognition Techniques", International Journal of Computer Applications, vol. 58, no. 3, **(2012).**

[5] P. Ekman and W. V. Friesen, "Facial Action Coding System (FACS): Investigator's Guide", Consulting Psychologists Press, **(1978)**.

[6] Y. Yoshitomi, "Facial Expression Recognition for Speaker using Thermal Image Processing and Speech Recognition System", International Conference on Applied Computer Science, **(2010)**.

[7] S. V. Ioannou, A. T. Raouzaiou, V. A. Tzouvaras, T. P. Mailis, K. C. Karpouzis and S. D. Kollias, "Emotion Recognition through Facial Expression Analysis based on a Neurofuzzy Network", Neural Networks, vol. 18, **(2005).**

[8] M. Pantic and I. Patras, "Dynamics of Facial Expression: Recognition of Facial Actions and Their Temporal Segments from Face Profile Image Sequences", IEEE Transactions on Systems, Man and Cybernetics, vol. 36, no. 2, **(2006)**.

[9] M. Pantic and L. J. M. Rothkrantz, "Facial Action Recognition for Facial Expression Analysis from Static Face Images", IEEE Transactions on Systems, Man and Cybernetics – Part B, vol. 34, no. 3, **(2004)**.

[10] M. F. Valster, M. Pantic Z. Ambadar and J. F. Cohn, "Spontaneous vs. Posed Facial Behavior: Automatic Analysis of Brow Actions", ACM International Conference on Multimodal Interfaces, **(2006).**

[11] B. V. Campbell, R. N. Doherty-Sneddon, G. S. Langton and S. McAuley, "Testing Face Processing Skills in Children. British Journal of Developmental Psychology", vol. 18, **(2000)**.

[12] P. Alexandre and D. Martins, "Active Appearance Models for Facial Expression Recognition and Monocular Head Pose Estimation", Master Thesis, University of Coimbra, **(2008).**

[13] G. J. Edwards, C. J. Taylor and T. F. Cootes, "Interpreting Face Images using Active Appearance Models", IEEE International Conference on Automatic Face and Gesture Recognition, **(1988).**

[14] L. Teijeiro-Mosquera and J. L. Alba-Castro, "Performance of Active Appearance Model-based Pose-Robust Face Recognition", IET Computer Vision, vol. 5, no. 6, **(2011).**

[15] S. Kobayashi and S. Hashimoto "Automated Feature Extraction of Face Image and its Applications", IEEE International Workshops on Robot and Human Communication, **(1995).**

[16] P. Cunningham and S. J. Delany, "*k*-Nearest Neighbor Classifiers", Technical Report UCD-CSI-2007-4, **(2007).**

[17] J. M. Keller, M. R. Gray and J. A. Givens Jr., "A Fuzzy *k*-Nearest Neighbor Algorithm", IEEE Transactions on Systems, Man and Cybernetics, vol. 15, no. 4, **(1985).**

Authors

Yong-Hwan Lee, he received the M.S. degree in Computer Science and the Ph.D. in Electronics and Computer Engineering from Dankook University, Korea, in 1995 and 2007, respectively. Currently, he is an assistant professor at the Department of Smart Mobile, Far East University, Korea. His research areas include Image/Video Representation and Retrieval, Image Coding, Face Recognition, Augmented Reality, Mobile Programming and Multimedia Communication.

Wu-Ri Han, he is in attendance at Dankook University, Korea. His current research areas include Face Detection, Emotion Recognition, and Pattern Matching.

Youngseop Kim, he received the M.S. in Computer Engineering from the University of Southern California in 1991, and the Ph.D. in Electronic Systems from Rensselaer Polytechnic Institute in 2001. He was a manager at Samsung SDI until 2003. He developed the image-processing algorithm for PDP TV while at Samsung. Currently he is an Associate Professor at Dankook University in Korea. He is the resolution member and the Editor of JPsearch part 2 in JPEG, the Co-Chair of JPXML in JPEG, and Head of Director (HOD) of Korea. He is also Editor-in-chief of the Korea Semiconductor and Technology Society. His research interests are in the areas of image/video compression, pattern recognition, communications, stereoscopic codecs, and augment reality. They include topics such as Object-Oriented Methods for Image/Video Coding, Joint Source-channel Coding, Rate Control, Video Transmission over packet wired or wireless networks, Pattern Recognition, Image Processing, and Augmented Reality.

Cheong-Ghil Kim, he received the M.S. and Ph.D. degree in Computer Science from Yonsei University, Korea, in 2003 and 2006, respectively. Currently, he is a professor at the Department of Computer Science, Namseoul University, Korea. His research areas include Multimedia Embedded Systems and Augmented Reality.