1. Emotion Recognition using Adaptive Motion Analysis on Facial Features

2. - Preliminary study on emotional communication between performer and audience in
3. Virtual Environment System of Korean Dance -

4. 노성규, 박한훈, 진윤종, 엄태영, 박종일*

차 례

1. Introduction

Facial expression is an important element in human communication and the studies have shown that facial expression reveals the underlying emotion of the human [5]. Since facial expression provides significant clues about one's emotional state, interests have been risen [1, 2, 9, 15] on human-computer interaction (HCI) of how machines can understand facial expressions of the human. However, it always has been a great challenge to build a machine that recognizes human facial expressions effectively and reliably. The limitation of the automated machine for recognizing facial expressions is that complex human facial features are represented with "limited" verbal descriptions (i.e. any descriptions that can be described verbally) of facial expressions. We say "limited" because the human language may not describe every little details

* 한양대학교 전자통신컴퓨터공학과 가상현실연구실 / uty02@mr.hanyang.ac.kr

perceived by the human visual system. The best known facial expression analyzer i.e. the human visual system is “trained” by the tremendous amounts of data over the substantial time with the best known parallel learning system (i.e. the human neurons and their network), and it is safe to say that artificially reproducing the complex feature representation methods used by the human visual system is near impossible.

Previous researches have heavily relied on accurate estimates of facial feature movements on the face. They used facial expression representations based on Facial Action Coding System (FACS) [4, 6], 3D modeling of the human face [7], Gabor wavelet representation [1, 12, 13], or other geometric face model [8, 17] in an attempt to represent facial features as closely as possible to meet the verbal descriptions of human facial expressions. By considering many features, they may represent low-level details of the facial expression; however, they tend to be complicated and time-consuming to process.

In this paper, we present emotion recognition system using adaptive-motion analysis on facial features. Our method is simplicity-oriented but cost- effective. Simple face model is used to reduce the computational complexity required to analyze facial expressions. Adaptive motion analysis on facial features is effective way of analyzing facial expression by assigning more computational complexity on important facial features only. Lastly, rule-based facial expression classification is performed by using ID3 decision tree. ID3 decision tree classifies the given facial expression using minimal Boolean comparisons and thus fast classification is achieved.

This paper is presented in following order. Section 2 shows the overall framework of the proposed facial expression recognition system and Section 3 presents our method to analyze and recognize facial expression into emotional categories. Experiments and analysis is done in Section 4 and the conclusion of this paper is drawn in Section 5 along with future works.

2. Overall Framework

Our facial expression recognition system is composed of 2 parts: facial expression analysis and facial expression recognition. Facial expression analysis can be divided into 4 logical processing steps. First, a frontal static face is detected from an arbitrary scene. Second, facial features are detected based on simple face model. Third, detected facial features are further categorized into the facial features that are candidate for the orientation evaluation and those that are not. Lastly, adaptive motion analysis on facial features is performed. Facial expression recognition is basically classifying the analyzed facial expression into one of the 6 emotions dealt in the scope of this paper. After adaptive motion analysis is performed on the given facial expression, a feature descriptor of that facial expression is generated. Then, ID3 decision tree uses this feature descriptor to classify the given facial expression.

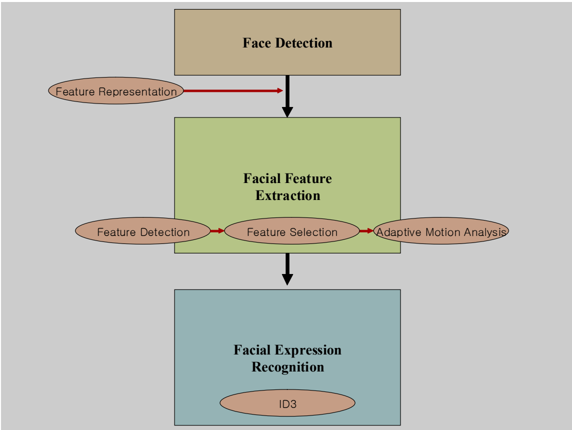

Figure1: Block Diagram of the process sequence of the proposed facial recognition system.

3. Methodology

For any facial expression analyzer, including the best known facial expression analyzer i.e. the human visual system, it should first perceive the face and then its appearance to guess the underlying emotion of the facial expression. Our method attempts to imitate the human visual system and thus it is composed of three steps: face detection, facial feature extraction, and facial expression classification. In this section, details of each step are covered and we introduce our approaches for recognizing facial expression. Our new approaches include: feature representation, feature selection, adaptive motion analysis for the facial feature extraction, and rule-based classification using ID3 tree for the facial expression classification.

3.1. Face Detection

The first step to recognize the facial expression is detecting the human face, "Where is the face?" There have been various efforts on detecting the human face, namely face detection in facial images and in arbitrary images [15]. Although the static frontal-view facial images (i.e. images from JAFFE database) are dealt in the scope of this paper, we used the face detection method in arbitrary images for the future work of implementing the real-time system. Viola et al. [18] proposed the rapid object detection, including the human face, using a boosted cascade of simple features. Lienhart et al. [10] improved the performance by extending [18]'s rapid object detection framework. We adopted Lienhart's method into our system for the face detection.

3.2. Facial Feature Extraction

After the face is detected, the next step for the facial expression recognition is to extract facial expression information or facial features. In our approach, facial features are extracted in the following order. Predefined features are first detected and then carefully "selected" (by using the ID3 entropy, or information gain, of the feature), for differentiating features with high variability from the ones with low variability, to effectively calculate the motion energies. For each facial feature, adaptive motion analysis is performed. In the sections 3.2.1 through 3.2.4, feature representation as well as how the system is trained, feature detection, feature selection, and adaptive motion analysis on features are discussed in detail.

3.2.1. Facial Feature Representation

One of the most critical issues in the automatic facial expression recognition is the representation of features, "Which features are used to represent the geometric face model and to classify the facial expression?" and "How are we going to represent these features?" The human visual system is "trained" by the tremendous amounts of data over the substantial time with the best known parallel learning system (i.e. the human neurons and their network), and it is safe to say that artificially reproducing the complex feature representation methods used by the human visual system is near impossible.

The limitation of the automated system for recognizing the facial expressions is that the features are represented with "limited" verbal descriptions (i.e. any descriptions that can be described verbally) of the facial expressions. We say "limited" because the language itself may not describe every little details perceived by the human visual system. The facial expression representations based on Facial Action Coding System (FACS) [4, 6], 3D modeling of the human face [7], Gabor wavelet representation [1, 12, 13], or other geometric face model [8, 17] attempt to represent facial features as closely as possible to meet the verbal descriptions of human facial expressions. By considering many features, they may represent low-level details of the facial expression; however, they tend to be complicated and time-consuming to process.

Based on the verbal descriptions of the facial expressions from DataFace [5, 19], our approach used 9 facial features to represent the geometric model of the face and classify the 6 emotional facial expressions. We first considered initial 15 features for the face model. 10,368 possible facial expressions, which are automatically generated based on the verbal descriptions of the expressions, are used for the initial training of the initial ID3 tree, and the initial ID3 tree is generated.

The subsequent result, i.e. initial ID3 tree, indicated that some features tend to move together (ex. left eyebrows and right eyebrows) and thus these features are united as one feature (ex. eyebrows for left eyebrows and right eyebrows) for the sake of reducing the complexity. ID3 is well-suited for this task because it searches for the attributes (i.e. features) that best classify the data (i.e. the emotion) and finds "unnecessary" features that may be discarded. After the removal of "unnecessary" features, the final of 9 features are used to represent the geometric model of the face and classify the 6 emotional facial expressions. Using the final 9 features and their "actions," 1728 possible facial expressions (that are obtained by considering every possible actions of each feature) are used to train the final ID3 tree for classifying the facial expression.

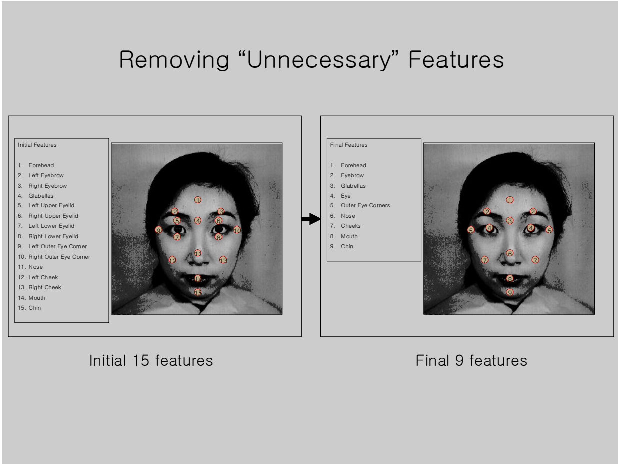

Figure2: For the facial expression analysis, the final of 9 features are obtained from the initial 15 features.

3.2.2. Facial Feature Detection

Based on the feature representations described in 3.2.1, 9 facial features are considered in our method of facial feature extraction. The foremost step for the feature extraction is detecting the facial features. We have adopted the feature detection framework done by Park et al [16] and modified it to detect more facial features.

The feature detection proceeds as follows:

(1) Once the face is detected, detected face region is rescaled into a 100-by-100 image. The resolution and sample density of the source image is reduced to restrict the search regions of the scene where facial features are most likely to occur.

(2) The valley image [3] of rescaled image is obtained since it can clearly show main features (features around the valley regions such as eyes, nose, and mouth) and thus easy detection of main features is possible. However, the information obtained from the valley image alone may require come computational expenses to detect the features. Thus, a generic feature template [11], which segment the rescaled face image into sub-regions for the main features (ex. R1 for right eye, R2 for left eye, R3 for nose, and R4 for mouth), is applied on the valley image for even more search region restrictions and the main facial features are detected in extremely fast and effective manner.

(3) After the main features are detected, remaining features are detected based on the locations of the “main” facial features.

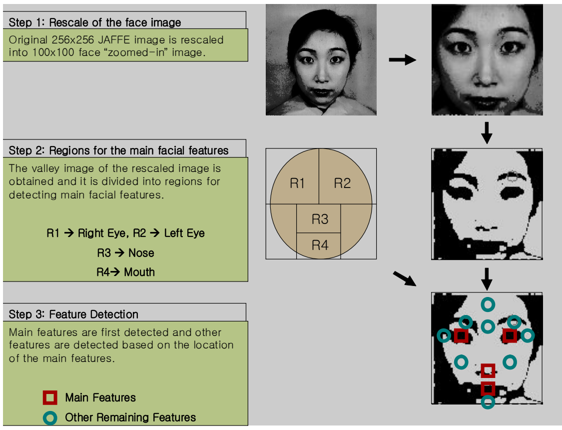

Figure 3: Feature Detection.

3.2.3. Facial Feature Selection

In our approach, each feature has two basic action states, neutral and expressed states, and expressed states are further categorized into sub-action states accordingly depending on features. Features and their "actions units" are shown in the table.

Table 1: "Action Units" of the final 9 featuresFeatures

Features	Actions
1. Forehead	- Neutral - Expressed
2. Eyebrow	- Neutral - Up - Down
3. Glabellas	- Neutral - Expressed
4. Eye	- Neutral - Narrowed - Widened
5. Outer Eye Corners	- Neutral - Expressed
6. Nose	- Neutral

	- Expressed
7. Cheeks	- Neutral - Expressed
8. Mouth	- Neutral - Open Wide - Thinned - Upper Lip Drawn up - Lip Corners Up - Lip Corners Down
9. Chin	- Neutral - Expressed

Each feature's actions, obtained from the verbal descriptions of the feature, show the variability of the feature (i.e. the motion energy orientation of the feature). They are used to differentiate the features with high variability from the ones with low variability, and by doing so we can effectively analyze the changes in motion by performing the adaptive motion analysis. Adaptive motion analysis is performing the orientation evaluation on selected features. That is, the features with one expressed states (for example, features with low variability such as forehead, cheek, etc.) are excluded from the orientation evaluation, and only the features with more than one expressed states (for example, features with high variability such as mouth, eyebrows, eye, etc.) are evaluated for their orientations.

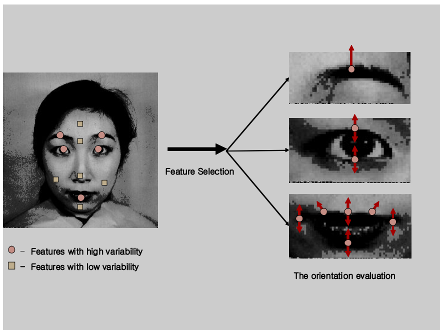

Figure4: Feature Selection.

3.2.4. Adaptive Motion Analysis on Features

Adaptive motion analysis is to effectively capture the patterns (namely the intensity and the orientation) of the features from the neutral face image to the expressed face image. Park et al [16] proposed the point-wise motion analysis (i.e. motion analysis is done by specifying the rectangle region on each facial features) for facial expression recognition, but their work is only concerned with the intensity of the features (i.e. "Was there an 'action' from the neutral face image to the expressed face image?").

3.2.4.1. Our Approach

Our approach extended [16]'s motion analysis framework in two ways. First, we consider not only the intensity of features but also the orientation of the features. Orientation of the feature can be thought as detailed information of the feature intensity, and in our approach the orientation of the feature is described by the expressed states as explained in Section 3.2.3. Second, we apply the adaptive motion analysis on the facial features. That is, the orientation evaluation is performed only on selected features and the intensity evaluation is performed on non-selected features.

3.2.4.2. Orientation Evaluation

The orientation evaluation is performed by pattern tracking. Pattern tracking is similar to feature tracking in a sense that patterns are first located in each image frame, then the patterns are followed from frame to frame to estimate motion [21]. We extended the pattern tracking method proposed by Park et al [16]. Two images, neutral and expressed face images, are used to estimate the motion of the features by finding the "difference image." The "difference image" is the image of difference in patterns from the neutral to the expressed image, and it represents the motion of the facial expression.

For each selected facial feature, a rectangle box is first specified for the calculation of the feature's motion. In order to evaluate the feature's orientation, the original rectangle box is divided into a number of small sub-rectangle boxes (number of small sub-rectangle boxes is closely tied to the number of expressed states of the feature), then we calculate the motion energy values for each of them, and finally the motion energy values for each sub-rectangle box is compared to determine the feature's orientation.

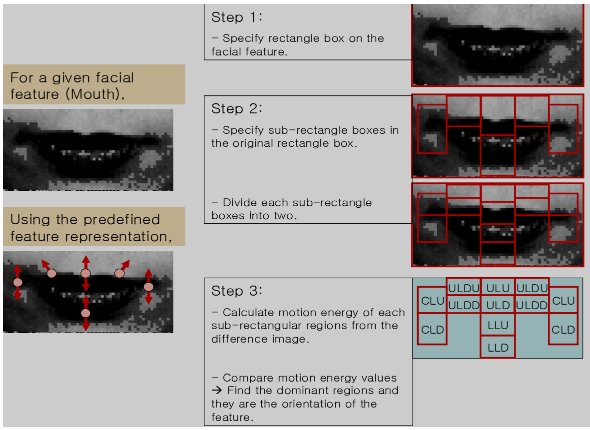

Figure5: Orientation Evaluation.

3.3. Facial Expression Classification

Once the facial features are extracted and they are represented as attributes (i.e. discrete attributes that describes the facial expression), ID3 tree can be used to classify the given facial expression.

3.3.1. ID3 requirements for the test data

ID3 algorithm has requirements (or limitations) on the sample data it can take [20], and in this section, these requirements are examined to see if our ID3 tree is suited for the facial expression recognition.

There are four requirements that the sample data used in ID3 algorithm must follow: (1) the same attributes must describe each example and have a fixed number of values. (2) An example's attributes must already be defined, i.e. they are not learned by ID3. (3) Discrete classes should be used and classes broken up into vague categories are not good. (4) There must be sufficient training examples to distinguish valid patterns from chance occurrences.

We have met all the requirements since our ID3 tree is trained with the total of 1728 possible facial expressions (satisfy 4th condition) that are generated based on 9 predefined (satisfy 2nd condition) facial features with a fixed number of actions (satisfy 1st condition). Furthermore, the "unnecessary" features from the initial 15 facial features are removed to use only the discrete facial features (satisfy 3rd condition), and the final of 9 facial features are considered to represent the face model and to generate the sample data.

3.3.2. ID3 tree for the classification

Our classification is based on rule-based method, and the ID3 decision tree is used since our ID3 tree and the training data meets the requirements imposed by ID3 algorithm. Unlike previous efforts on statistically comparing and classifying the facial expressions that are complex, computationally expensive, and time-consuming, our classification method using ID3 decision tree is based on simple Boolean comparisons. Moreover the entropy or the information gain of ID3 algorithm searches for the facial features that best classify the facial expression, and thus it forces the classification system to recognize the facial expression with minimal comparison steps.

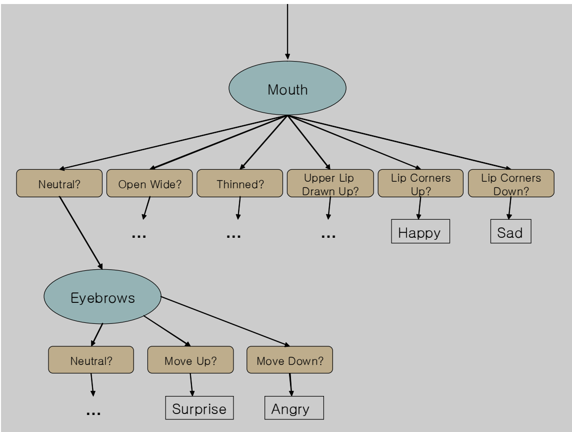

Figure6: ID3 Tree for the facial expression recognition.

As you can see from the above figure, mouth has the highest information gain in our ID3 tree and it is the first feature that the classification system will examine once the test data comes.

4. Experiments and Analysis

4.1. Experimental Results

We have tested the system on JAFFE (Japanese Female Facial Expression) database for quantitative analysis on 6 emotional states (Surprise, Disgust, Fear, Anger, Happiness, and Sadness).

95 Sample test images (Surprise: 17, Disgust: 14, Fear: 13, Anger: 14, Happiness: 21, and Sadness: 16) out of total of 193 images in the JAFFE database were used to test the system. The performance was pretty good and consistent with overall recognition accuracy of 77.6% (Surprise: 85.1%, Disgust: 77.1%, Fear: 69.6%, Anger: 82.3%, Happiness: 79.3%, and Sadness: 72.3%).

4.2. Discussion

We first compare the performance of our method with other rule-based facial expression classification system that was tested on the JAFFE database, namely point-wise motion energy analysis implemented by Park et al [16].

Table 2: Comparison of our method with other rule-based method

Methods	Scope	Database	Recognition Rate (%)
Point-Wise Motion Energy Analysis by Park et al [16].	Pattern Tracking on **Motion intensity** of the feature	JAFFE	Overall: 70% (Fear: 64%, Sad: 70%)
Our Feature-Based Adaptive Motion Analysis	Pattern Tracking on **Motion orientation** as well as **Motion intensity** of the feature	JAFFE	Overall: 77% (Fear: 69%, Sad 72%)

As you can see from the table, our method of tracking feature's motion orientation patterns performs better than [16]'s method of tracking feature's motion intensity patterns only. This indicates, as expected, that the facial feature orientation is the important element for recognizing facial expressions that should not be discarded.

It should be noted that there are “bad” samples, in the JAFFE database, that are prone to misclassify the facial expression. Certainly, we have taken into the consideration that facial expressions vary from one individual to another and the good classifier should handle well whether the sample is occurred by a chance or not. Our method handled quite well on “bad” samples. “Bad” samples can be categorized into two classes: “wrong” facial expressions that are mislabeled, and “weak” facial expressions that are misleading and hard to clearly classify so that even the human visual system may be puzzled. The examples of “wrong” facial expressions and “weak” facial expressions are shown below.

(a)

(b)

Figure7: Example of “bad,” (a) mislabeled and (b) misleading, samples in JAFFE.

When there are significant amount of “bad” samples, it may inevitably affect the recognition rate of facial expressions. In the JAFFE database, “weak” expressions tend to be distributed heavily over facial expressions labeled as fear or sad and thus the recognition rate for fear and sad expressions were relatively low compared with other recognition rate. [16]’s experimental results also support our position.

5. Conclusion and Future Works

In this paper, we have presented an emotion recognition system using adaptive motion analysis on facial features. Our method is simple because minimal reasonable facial features are selected for the generation of simple face model to reduce the computational complexity to analyze facial expressions. Furthermore, adaptive motion analysis on facial features is an effective way of estimating the facial muscle movements from the expressed facial expression by assigning more computational complexity on important facial features. Lastly, ID3 decision tree is used for the rule-based classification of facial expressions. ID3 decision tree classifies the given facial expression using minimal Boolean comparisons and thus fast facial expression classification is achieved. Our method performed quite well with 77% overall accuracy of recognizing facial expressions (Surprise, Fear, Disgust, Anger, Sad, and Happy). Our experimental result shows that the facial feature orientation plays an important role along with the facial feature intensity for recognizing facial expressions.

Valuable future research includes a real-time system of simple rule-based facial expression classification system that has a strong tolerance to “bad” data samples. Our current research is applying the proposed facial expression classification system onto the Virtual Environment System of Korean Dance* to study the emotional communication between dance performer and audience.**

---

* Virtual Environment System of Korean Dance: To assist Korean dance performance using a camera, two projectors, AR-Toolkit, and etc.

** 논문투고일자 : 2007. 4. 30. / 심사일자 : 2007. 5. 10. / 심사완료일자 : 6. 1.

■ References

[1] M. Bartlett, G. Littlewort, I. Fasel, and J. Movellan, "Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction," In CVPR Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction, Vancouver, Canada, 2003.

[2] A.F. Bobick and J.W. Davis, "The Recognition of Human Movement Using Temporal Templates," IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 23, 257-267, 2001.

[3] G. Chow and X. Li, "Toward a System for Automatic Facial Feature Detection," Pattern Recognition, Vol. 26, 1739-1755, 1993.

[4] P. Ekman and W.V. Friesen, Facial Action Coding System, Consulting Psychologists Press Inc, 1978.

[5] P. Ekman, "Facial Expression and Emotion," American Psychologists. Col. 48, No.4, 384-392, April 1993.

[6] I. Essa and A. Pentland, "Coding, Analysis, Interpretation, Recognition of Facial Expressions," IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 19, 757-763, 1999.

[7] I. Essa and A. Pentland, "Facial Expression Recognition using a Dynamic Model and Motion Energy," In Proceedings of International Conference on Computer Vision (ICCV), , 1995.

[8] X. Feng, "Facial Expression Recognition Based on Local Binary Patterns and Coarse-to-Fine Classification," In Proceedings of International Conference of Computer and Information Technology (CIT), 178-183, 2004.

[9] S. Krinidis, I. Buciu, and I. Pitas, "Facial Expression Analysis and Synthesis: A Survey," In Proceedings of Human Computer Interaction (HCI), Vol. 4, 1432-1436, 2003.

[10] R. Lienhart and J. Maydt, "An Extended Set of Harr-Like Features for Rapid Object Detection," In Proceedings of International Conference on Image Processing (ICIP), Vol. 1, 900-903, 2002.

[11] C. Lin and L. Wu, "Automatic Facial Feature Extraction by Genetic Algorithms," IEEE Transactions on Image Processing, Vol. 8, 834-845, 1999.

[12] M. Lyons, J. Budynek, and S. Akamatsu, "Automatic Classification of Single Facial Images," IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 21, 1357-1362, 1999.

[13] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding Facial Expressions with Gabor Wavelets," In Proceedings of International Conference on Automatic Face and Gesture Recognition (FG), 200-205, 1998.

[14] T. Mitchell, "Decision Tree Learning," Machine Learning, Chapter 3, 52-81, McGraw-Hill, 1997.

[15] M. Pantic and L.J.M. Rothkrantz, "Automatic Analysis of Facial Expressions: the State of the Art," IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 22, 1424-1445, 2000.

[16] H. Park and J. Park, "Analysis and Recognition of Facial Expression Based on Point-Wise Motion Energy," In Proceedings of International Conference of Image Analysis and Recognition (ICIAR), 700-708, Lecture Notes in Computer Science (LNCS) 2004.

[17] Y. Shinohara and N. Otsu, "Facial Expression Recognition Using Fisher Weight Maps," In Proceedings of International Conference on Automatic Face and Gesture Recognition (FG), 499-504, 2004.

[18] P. Violar and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," In Proceedings of Computer Vision and Pattern Recognition (CVPR), Vol. 1, 511-518, 2001.

[19] Data Face: Facial Expressions, Emotion Expressions, Nonverbal Communication. Available at: http://face-and-emotion.com/dataface/emotion/expression.jsp

[20] The ID3 algorithm. Available at: http://www.cise.ufl.edu/~ddd/cap6635/Fall-97/Short-papers/2.htm

[21] T. Huang, P. Burt, and K. Mase, “Computer Vision and Face Processing,” III-B, NSF Report. Available at: http://face-and-emotion.com/dataface/nsfrept/computer_vision.html