



UNIVERSITY
OF WOLLONGONG
AUSTRALIA

University of Wollongong
Research Online

Faculty of Informatics - Papers (Archive)

Faculty of Engineering and Information Sciences

2006

Facial expression recognition for multiplayer online games

Ce Zhan

University of Wollongong, czhan@uow.edu.au

Wanqing Li

University of Wollongong, wanqing@uow.edu.au

Philip O. Ogunbona

University of Wollongong, philipo@uow.edu.au

Farzad Safaei

University of Wollongong, farzad@uow.edu.au

Publication Details

Zhan, C., Li, W., Ogunbona, P. O. & Safaei, F. (2006). Facial expression recognition for multiplayer online games. Joint International Conference on CyberGames and Interactive Entertainment (pp. 52-58). Western Australia: Murdoch university.

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library:
research-pubs@uow.edu.au

Facial expression recognition for multiplayer online games

Abstract

The Multiplayer Online Game (MOG) becomes more popular than any other types of computer games for its collaboration, communication and interaction ability. However, compared with the ordinary human communication, the MOG still has many limitations, especially in communication using facial expressions. Although detailed facial animation has already been achieved in a number of MOGs, players have to use text commands to control avatars expressions. In this paper, we briefly review the state of the art in facial expression recognition and propose an automatic expression recognition system that can be integrated into a MOG to control the avatar's facial expressions. We evaluate and improve a number of algorithms to meet the specific requirements of such a system and propose an efficient implementation. In particular, our proposed system uses fixed and less facial landmarks to reduce the computational load with little degradation of the recognition performance.

Keywords

Facial, expression, recognition, for, multiplayer, online, games

Disciplines

Physical Sciences and Mathematics

Publication Details

Zhan, C., Li, W., Ogunbona, P. O. & Safaei, F. (2006). Facial expression recognition for multiplayer online games. Joint International Conference on CyberGames and Interactive Entertainment (pp. 52-58). Western Australia: Murdoch university.

Facial expression recognition for multiplayer online games

Ce Zhan, Wanqing Li, Philip Ogunbona, and Farzad Safaei
University of Wollongong
Wollongong, NSW 2522, Australia
{cz847, wanqing, philipo, farzad}@uow.edu.au

ABSTRACT

The Multiplayer Online Game (MOG) becomes more popular than any other types of computer games for its collaboration, communication and interaction ability. However, compared with the ordinary human communication, the MOG still has many limitations, especially in communication using facial expressions. Although detailed facial animation has already been achieved in a number of MOGs, players have to use text commands to control avatars expressions. In this paper, we briefly review the state of the art in facial expression recognition and propose an automatic expression recognition system that can be integrated into a MOG to control the avatar's facial expressions. We evaluate and improve a number of algorithms to meet the specific requirements of such a system and propose an efficient implementation. In particular, our proposed system uses fixed and less facial landmarks to reduce the computational load with little degradation of the recognition performance.

1. INTRODUCTION

A Multiplayer Online Game (MOG) is a genre of computer game that is capable of supporting multiple users simultaneously, and is played on the Internet. The collaboration, communication and interaction ability of MOGs enable players to cooperate or compete with each other on a large scale. Thus, players could experience relationships as real as those in the real world. The "real feeling" makes MOGs more popular than any other types of computer games, despite significant amounts of time and money required for playing. Taking youths in china for example, according to "Pacific Epoch's 2006 Online Game Report" [12], China has 25.3 million online gamers in 2005 and will hit 30.4 million users by the end of 2006.

However, although interactive realism is the most advantage of MOGs, compared with the ordinary human communication in real world, there are still many limitations. For example, in most existing MOGs, text-chat is used rather than

real time voice chatting; during a conversation, avatars have no activities related to natural body gestures, facial expressions, etc.

Among the problems mentioned above, we concentrate on facial communication in particular. In everyday life, the manifestation of facial expressions is a significant part of our social communication. Our underlying emotions are conveyed by different facial expressions. To feel immersed and socially aware like in the real world, players must have an efficient method for conveying and observing changes in emotional states. Game designers use the following approaches to deal with facial expressions in existing MOGs:

- **No facial expressions:** no expressions are supported. Avatars keep the same facial expression all the time.
- **Text facial expressions:** pure text based expressions are provided. Slash commands, such as /smile, /frown, and /wink, are used to convey different emotions. The output expressions are described by text. For example, when John types the command "/smile" while having a conversation with Bruce, the text "John smiles at Bruce" appears on the screen.
- **Text facial expressions plus Facial animation:** facial expression animation is provided. The animation is also triggered by slash commands, and the corresponding text expressions accompany the animation. For example, when John types the command "/smile" while having a conversation with Bruce, John's avatar smiles, and at the same time, "John smiles at Bruce" appears on the screen.
- **Text facial expressions plus Body animation:** body gesture animation is provided to represent facial expressions. For example, (the same case as above) the slash command "/smile" produces the text expression "John smiles at Bruce" plus a swing of an arm, but the avatar doesn't smile.

According to the above rough survey, facial rendering is not the problem. A number of existing MOGs has already achieved very nice and detailed facial animation. The problem is how to control the avatar's facial expressions easily

and naturally. Text commands are simple and straightforward, but not easy to use. First, players have to memorize all the commands. Thus the more sophisticated the facial system is, the harder it is to use. Second, humans convey emotions by expressions in real time. Players can not type text commands every few seconds to update their current mood. Thirdly, facial communication should happen naturally and effortlessly, typing commands ruins the realism.

An automatic facial expression recognition system seems ideal for the purpose that “text commands” can not achieve, since the recognition results of players’ expressions could be used to control the facial system of a MOG.

In this paper, we propose an automatic facial expression recognition system for a Multiplayer Online Game. In section 2 a brief overview of the state of the art of automatic facial expression recognition is provided. Section 3 analyzes the specific requirements of a facial expression recognition system in the context of MOGs. Based on the requirements, suitable algorithms are chosen and evaluated in section 4 for an efficient implementation of the proposed system. Section 5 presents the experimental results and Section 6 concludes the paper.

2. OVERVIEW OF AUTOMATIC EXPRESSION RECOGNITION

In computer vision, a facial expression is usually considered as the deformations of facial components and their spatial relations, or changes in the pigmentation of the face. An automatic facial expression recognition system (AFERS) is a computer system that attempts to classify these changes or deformations into abstract classes automatically. A large number of approaches have been proposed since mid 1970s in computer vision. Early works have been surveyed by Samal and Iyengar [19] in 1992. Fasel et al [7] and Pantic et al [16] published two comprehensive survey papers which summarized the facial expression recognition methods proposed before 1999. Recently, Ying-Li Tian et al [20] presented the recent advances (before the year 2004) in facial expression recognition.

Generally, an AFERS consists of three processing stages: face detection, facial feature extraction and representation, and facial expression recognition. Face detection is to automatically locate the face region in an input image or image sequences. As the first step of AFERS, its reliability has a major influence on the performance and usability of the entire system. The face detector could detect faces frame by frame or just detect a face in the first frame and then track it in the subsequent images.

After the face has been detected, the next step is to extract and represent the information about the facial expression to be recognized. During this stage, pixel data of images are converted into a higher-level representation, known as “feature vectors”, which is then used for the subsequent expression classification. Geometric features which encode the shape and locations of facial components and spectral-transform-based features which are gained by applying image filters to face images are often used to represent the information of facial expressions. Irrespective of the type of

feature extraction approach used, the essential information about the displayed expressions should be preserved. The extracted features should possess high discriminative power and high stability against different expressions.

Facial expression classification is the last stage of AFERS. It is a decision procedure performed by a classifier. The facial changes can be identified as facial action units (AU) [3] or six prototypical emotional expressions [5]. Introduced by Ekman and Friesen, each of the six prototypical emotions possesses a distinctive content and can be uniquely characterized by a facial expression. These prototypical emotions are also referred to “basic emotions”. They are claimed to be universal across human ethnicities and cultures and comprise happiness, sadness, fear, disgust, surprise and anger. An AU is one of the 44 atomic elements of visible facial movement or its associated deformation. Ekman and Friesen first use AUs in their facial action coding system (FACS) [6] with the goal to describe all possible perceptible changes that may occur on the face. In applications, a facial expression is represented using a combination of AUs with respect to the locations and intensities of corresponding facial actions.

3. FACIAL EXPRESSION RECOGNITION SYSTEM FOR MOG

As mentioned above, our goal is to automatically recognize the player’s facial expressions, so that the recognition results can be used to drive the “facial expression engine” of a multiplayer online game. Before such a system can be developed, its functionality has to be specified. An ideal case is that the system could perform recognition accurately and without delay just like human beings. In practice, the task has high computational complexity and may not be 100% accurate. In this section, we analyze the requirements for an AFERS in the context of multiplayer online game.

Before a recognition process begins, facial expression images have to be acquired. For a multiplayer online game, players should input their expressions in a natural way without any constraints, the acquisition process should perform automatically and the acquired images should be image sequences in real-time. The “obvious” solution for these requirements is to use cameras to capture the users. “Multiple camera techniques” [17] and “facial colored markers” [13] methods can not be used in this case, since it is impractical to expect a game player to buy more than one camera or paint his face in order to show expressions in a game. Therefore, facial expressions are expected to be captured by a fixed single camera (webcam in most cases).

After the video sequence is obtained, the recognition process described in section 2 can be started. The most important requirement is that all the stages (face detection, facial feature extraction and representation, and facial expression recognition) have to be performed automatically and in real-time. Meanwhile, the players’ appearances vary with different sexes, ages, and ethnicities. The lighting conditions change and background can be complex. The property of players’ camera also varies. Consequently, the resolutions of the face images are different. All these differences may influence the accuracy of face detection, feature

tracking, and the final expression recognition. The system should be able to handle all these problems and perform a robust recognition.

The recognition process is taking place when a player is playing an online game. During setup, the camera is attached to some part of the computer screen, facing the player. We can assume that the camera's field of view allows a frontal pose of the player most of the time. It is because player must stare at the screen to play game. Still, in-plane rotations and small scale out-of-plane rotations of the face would appear frequently. So face normalization has to be introduced after face detection.

We have addressed in Section 2 that facial expressions can be classified as either facial action units (AU) or six prototypical emotional expressions. Some researchers consider that AUs perform better as classification classes since they can describe almost all possible facial changes especially the subtle ones. However, in the context of multiplayer online game, subtle facial changes may not be required. Players may do not have enough time to perceive those small scale changes. The simple but meaningful basic emotions may be a better choice for most situations.

4. PROPOSED SYSTEM

In this section, we propose an automatic facial expression system for a MOG and evaluate existing algorithms based on the specific requirements discussed in Section 2. Figure 1 shows the block diagram of the system. It consists of four components: face detection, landmark point localization, feature extraction and classification of the expressions.

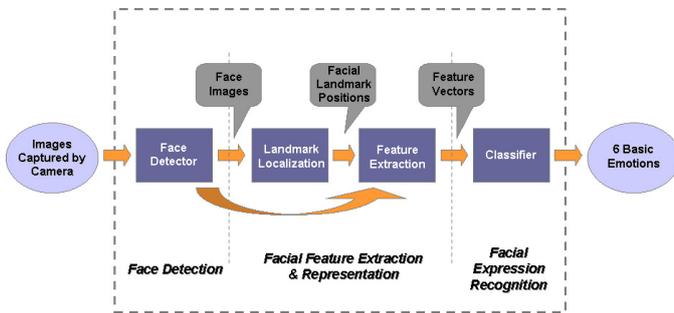


Figure 1: the proposed AFERS for MOG

4.1 Face detection

As stated in Section 3, real-time detection is the most important requirement for face detector in AFERS for MOGs. Among a large number of technologies that have been proposed so far to detect human face, boosting methods seem to be the most suitable one. Boosting is a powerful iterative procedure that builds efficient classifiers by selecting and combining very simple classifiers. The combination of simple classifiers may intuitively give a rapid detection without deteriorating the detection rates. So it achieves the best compromise between detection efficiency and speed.

4.1.1 Viola & Jones's methods

We built our face detector by implementing one of the boosting methods proposed by Viola and Jones [21]. The method

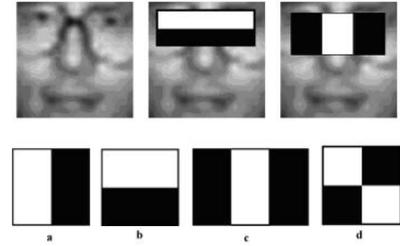


Figure 2: The four rectangle features used for face detection



Figure 3: The detection result using Viola & Jones' methods

could achieve real time detection by using very simple and easy to compute rectangle features. Four kind of these features (a, b, c, d), as shown in figure 2, are used to represent the image information in a sub-window. The feature value in each case is simply the difference between the sum of the pixel intensities in the white section and the sum of the intensities in the black section. A good detection rate was obtained by the use of the fundamental boosting algorithm, AdaBoost [8], which selects the most representative features in a large set. The detector scans an image by the sub-window at different scales. Each sub-window is tested by a classifier made of several stages (cascade). If the sub-window is clearly not a face, it will be rejected by one of the first steps in the cascade while more specific classifier (later in the cascade) will classify it if it is more difficult to discriminate. The detector can process 15 384x288 frames per second on a conventional PC with 700 MHz Intel Pentium. A detection example is presented in figure 3.

4.2 Landmark localization and feature extraction

4.2.1 Gabor wavelets based feature extraction

Many algorithms [7] have been proposed in the past, including Principal Components Analysis (PCA), Singular Value Decomposition (SVD), wavelet transformation etc. Research has demonstrated that Gabor filters are more discriminative for facial expressions and robust against various types of noise than other methods [7]. A Gabor filter is a complex sinusoid modulated by a 2D Gaussian function. We choose to apply Gabor filters to a set of landmark points to extract facial features for the following reasons with respect to the

requirements specified in Section 3:

- It can meet the demand of real time since the feature extraction is limited at the landmark points.
- It is insensitive to illumination variations and the limited localization in space and frequency yields a certain amount of robustness against translation, distortion, rotation and scaling of the images.
- If the following recognition approach is based on Euclidean distance, the observed expression image needn't correspond to the expression template rigidly. The dimensions of the observed image and the template are not required to be the same. Thus, the robustness of the whole system is expected to be improved.

A 2-D Gabor function is a plane wave with wavevector \mathbf{k} , restricted by a Gaussian envelope function with relative width σ :

$$\Psi(\mathbf{k}, \mathbf{x}) = \frac{\mathbf{k}^2}{\sigma^2} \exp\left(-\frac{\mathbf{k}^2 \mathbf{x}^2}{2\sigma^2}\right) \left[\exp(i\mathbf{k} \cdot \mathbf{x}) - \exp\left(-\frac{\sigma^2}{2}\right) \right] \quad (1)$$

In our implementation, we set $\sigma = \pi$ [15]. A set of Gabor kernels which comprise 3 spatial frequencies ($k = \frac{\pi}{4}, \frac{\pi}{8}, \frac{\pi}{16}$) and 6 different orientations ($\frac{\pi}{6}, \frac{2\pi}{6}, \frac{3\pi}{6}, \frac{4\pi}{6}, \frac{5\pi}{6}, \pi$). Each image is convolved with both even and odd Gabor kernels at facial feature landmarks (as shown in figure 4). Thus, 18 complex Gabor wavelet coefficients are gained at each landmark. Since only magnitudes of these coefficients are used, each face image is represented by a vector of 360 (3x6x20) when 20 landmarks are used.

4.2.2 Facial feature landmark localization

As we discussed above, Gabor features are produced by convolving the face image with a set of Gabor filters at facial landmarks. To extract the facial feature automatically, facial landmarks also need to be detected without manual efforts.

Automatic facial landmark localization is also a complex work. Generally, there are two categories of methods based on features they concern. Appearance-based approaches aim to find basis vectors to represent the facial features. Dealing with these vectors, machine learning techniques are used to get final results. Geometric-based methods use prior knowledge about the face position, and constrain the landmark search by heuristic rules that involve angles, distances, and areas. To find accurate position of landmarks, most of landmark detection methods involve multiple classification steps and a great number of training samples are needed. Although coarse-to-fine localization is widely used to reduce the computational load, the detection process is still too complex and time consuming for MOGs. To find a solution meeting the real-time requirement, before we choose our landmark localization method, a facial landmark location tolerance test is conducted. The aim of this test is to evaluate the variation of landmark positions in the normalized images of different faces.

The test was based on BioID face database [10]. The BioID database consists of 1521 gray level images with a resolution of 384x286 pixels. Each image shows the frontal view of 23 different individuals. 20 facial landmarks are manually selected on each of the 1521 images (as shown in figure 4).

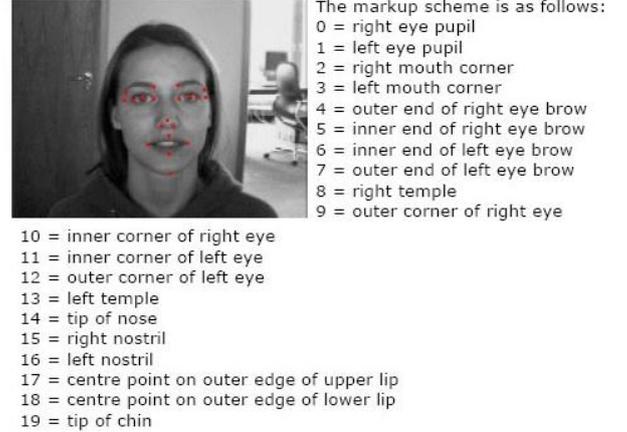


Figure 4: Landmark locations of BioID database

In the testing, all the images were geometrically normalized to align the coordinates of landmark 0 (right eye pupil) at (164, 95), coordinates of landmark 1 (left eye pupil) at (220, 95), and coordinates of landmark 17 (centre point on outer edge of upper lip) at (192, 146). (as shown in figure 5 and figure 6). The whole database was used for the test, so that each facial landmark had 1521 pairs of coordinates. During the analysis, we used following numerical features:

- Mean coordinates: the average value of all coordinates for one landmark.
- Mean distance: the average Euclidean distance from one testing coordinates to the mean coordinates.
- Ranges: difference between the maximum and minimum values of x and y respectively.
- The probability of position changes within 5 pixels around mean coordinates.

And following results were obtained:

- The mean distance is 3.1721 pixels.
- The average range of x is 17.4333 pixels.
- The average range of y is 20.2355 pixels.
- The average probability of x's changes within 5 pixels around mean coordinate is 0.9386.
- The average probability of y's changes within 5 pixels around mean coordinate is 0.8947.
- The image resolution is 384x286, 5 pixels is 1.3% of the width, and 1.75% of the height.
- The normalized distance between eyes is 56 pixels, 5 pixels is about 8.9% of the eye distance.



Figure 5: Image before alignment



Figure 6: Image after alignment

Our test has shown that facial landmark positions are relatively fixed after the geometric normalization based on the three key landmarks. It is reasonable to choose the fixed landmarks rather than performing traditional landmark detection. Thus, in our landmark localization process, only 3 points are to be searched on a face image. A SDAM (Simple Direct Appearance Models) method [9] is adopted to locate the three key points and normalize the face accordingly.

The shape with 3 key points can be described as a vector $S = (x_1, y_1, x_2, y_2, x_3, y_3)^T$ and the texture enclosed by the shape is denoted by T . It is claimed that there is a linear relationship between the shape and the texture [14]:

$$s = R * t + \epsilon \quad (2)$$

where s and t are the projection coefficient of S and T on the PCA space. R is the transformation coefficient matrix, which can be derived from a training set through linear regression. Then the face alignment progress can be described as following:

1. Initialize texture t .
2. Calculate $s = R * t + \epsilon$ to get the position of 3 key points. If s is very close to the destination position, terminate this procedure.

3. Align the image based on the detected 3 points. Gain a new texture t from the aligned image, then go to 2.

In most of the cases, one iteration is enough to align the face image. After the face image is geometrically normalized, mean coordinates obtained from our test are used as landmarks.

4.3 Classification

Several classifiers were evaluated in our proposed system. A brief description of each classifier is given below.

- **Mahalanobis distance classifier (MDC)** [4] is one of the simplest classifiers which classify patterns based on minimum Mahalanobis distance. In the training process, the k th class w_k is represented by its mean vector M_k and covariance matrix Σ_k which can be estimated from the training samples:

$$M_k = \frac{1}{N_k} \sum_{i=1}^{N_k} X_i^{(k)} \quad (k = 1, \dots, c) \quad (3)$$

and

$$\Sigma_k = \frac{1}{N_k} \sum_{i=1}^{N_k} (X_i^{(k)} - M_k)(X_i^{(k)} - M_k)^T \quad (4)$$

In the classification process, a given pattern X if unknown class is classified to w_k if its Mahalanobis distance to w_k is smaller than those to all other classes.

- **KNN classifier** [1] is an instance-based learning algorithm that is based on a distance function for pairs of observations, such as the Euclidean distance. In the classification process, k nearest neighbors of a training data is computed first. Then the similarities of one sample from testing data to the k nearest neighbors are aggregated according to the class of the neighbors, and the testing sample is assigned to the most similar class.
- **Decision tree classifier (DTC)** [18] is a hierarchically based classifier which compares the data with a range of properly selected features. The selection of features is determined from an assessment of the spectral distributions or separability of the classes. There is no generally established procedure. Therefore each decision tree or set of rules should be designed by an expert.

We also evaluated Linear Discriminant Analysis classifier [2] (LDA) and Naive Bayes classifier [4] (NBC), due to space limitation, we do not provide descriptions about them, readers are referred to corresponding references.

5. EXPERIMENTAL RESULTS AND DISCUSSIONS

In our experiments we use the JAFFE database [11]. The database contains 213 images of female facial expressions.

Methods	Fixed Landmarks	Manual Landmarks
MDC	76%	83%
KNN	82%	87%
DTC	73%	75%
LDA	79%	80%
NBC	77%	83%

Table 1: recognition rate using fixed landmarks and manual selected landmarks

Landmarks Used	Results
all	82%
without 8, 13, 19	82%
without 8, 13, 19, 14, 15, 16	75%
without 8, 13, 19, 4, 5, 6, 7	71%
without 8, 13, 19, 0, 1, 9, 10, 11, 12	67%
without 8, 13, 19, 0, 1, 9, 10, 11, 12, 4, 5, 6, 7	50%
without 8, 13, 19, 2, 3, 17, 18	50%
without 8, 13, 19, 2, 10, 11, 15, 16	82%

Table 2: recognition rate using different landmarks set

The head is almost in frontal pose. Original images were rescaled and cropped such that the eyes are roughly at the same position with a distance of 60 pixels in the final images (resolution: 256 pixels x 256 pixels). The number of images for each of the 7 categories of expressions (neutral, happiness, sadness, surprise, anger, disgust and fear) is roughly the same.

Trained classifiers were used to categorize detected expression images into one of the six basic emotions or the neutral expression. The inputs of classifiers were Gabor coefficient vectors obtained in the feature extraction process, both fixed facial landmarks and manually selected landmarks are used. The recognition results are listed in Table 1. According to the results, the best recognition rate was achieved by KNN classifier. Since KNN is instance-based method, it tends to be computationally slow. So the value of K should be chosen to be small while keeping a good recognition rate. In our evaluation, we set $K=3$, the classification speed was acceptable. As another distance based classifier, Mahalanobis distance classifier also achieved a relatively good result considering its simplicity. This may be due to the property of Gabor filters as discussed in Section 4.2. As we expected, the recognition rate of using fixed landmarks is relatively close to that of using manual selected landmarks. This demonstrates that our landmark localization method is feasible.

Besides using all the facial landmarks showed in figure 4, we also conducted a test using selected facial landmarks. The test results based on KNN classifier are showed in Table 2. As it can be seen in the table, when only 12 facial landmarks (0, 1, 3, 4, 5, 6, 7, 9, 12, 14, 17, 18) are used, recognition rate is comparable to the one that all of the 20 landmarks are used. Nearly half of the computational load can be saved during the feature extraction process in this case. The testing results also indicate that the information of facial expressions are mainly conveyed by eyes and mouth.

Though we didn't obtained very high recognition rates, the results are reasonable considering the limitation of database and only common and simple classifiers were used. We believe that automatic facial expression recognition for the conditions and requirements of MOG is achievable with current technologies. However, in the view of implementation, some issues have to be addressed:

- **Training data.** As just mentioned, the primary issue in expression recognition is the lack of training data. Most available databases contain only a few facial expressions from a small number of subjects. Background of the face images is fixed and clear and lighting condition is set to be good. A practical expression recognition can not be achieved purely based on those databases.
- **Robustness.** As we analyzed in Section 3, the recognition system has to be robust against various face appearance, clustered background, different lightning conditions and camera resolutions. Though, Gabor wavelet based representation can solve the problem to some extent, much more work remains to be done. We have analyzed that in most cases, a player shows a frontal face with small scale out of plane rotations. However, when some areas of the face is missing, an approach has to be found for fulfilling the missing part.

6. CONCLUSIONS

In this paper, we presented an automatic facial expression recognition system for MOGs. Existing algorithms were evaluated and tailored for an efficient implementation of the system. Although research in computer vision has advanced recently the technologies for expression recognition, building a real-time automatic facial expression recognition system still remains challenging and optimization of existing algorithms and their efficient integration are usually required.

7. ACKNOWLEDGMENTS

This work is partly supported by Smart Internet Technology (SIT) CRC Australia. Ce Zhan would like to thank SIT for providing research scholarship. Authors would also thank M.J. Lyons for providing facial expression database used in this paper.

8. REFERENCES

- [1] D. Aha. Tolerating noisy, irrelevant and novel attributes in instance-based learning algorithms. *International Journal of Man-Machine Studies*, 36(1):267–287, 1992.
- [2] J. Cohn, A. Zlochower, J. Lien, and T. Kanade. Automated face analysis by feature point tracking has high concurrent validity with manual face coding. *Psychophysiology*, 36:35–43, 1999.
- [3] G. Donato, M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski. Classifying facial actions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(10):974–989, 1999.

- [4] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley, New York, 2000.
- [5] P. Ekman. *Emotion in the Human Face*. Cambridge University Press, 1982.
- [6] P. Ekman and W. Friesen. *Facial Action Coding System (FACS): Manual*. Consulting Psychologists Press, 1978.
- [7] B. Fasel and J. Luttin. Automatic facial expression analysis: Survey. *Pattern Recognition*, 36(1):259–275, 2003.
- [8] Y. Freund and R. E. Schapire. A decision-theoretic generalization of online learning and an application to boosting. *Computational Learning Theory: Eurocolt '95*, pages 23–37, 1995.
- [9] A. Haizhou, X. Xipan, and X. Guangyou. Face detection and retrieval. *Chinese Journal of Computer*, 26(7):200–205, 2003 (in Chinese).
- [10] <http://www.bioid.com/>.
- [11] <http://www.mis.atr.co.jp/mlyons/jaffe.html>.
- [12] <http://www.pacificepoch.com>.
- [13] S. Kaiser and T. Wehrle. Automated coding of facial behavior in human-computer interactions with faces. *J. Nonverbal Behavior*, 16(2):67–83, 1992.
- [14] S. Li, Y. ShiCheng, H. Zhang, and Q. Cheng. Multi-view face alignment using direct appearance models. *Automatic Face and Gesture Recognition*, pages 324–329, 2002.
- [15] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. Coding facial expression with gabor wavelets. *Proc. Third IEEE International Conference on Automatic Face and Gesture Recognition*, pages 200–205, 1998.
- [16] M. Pantic and L. Rothkrantz. Automatic analysis of facial expressions: the state of the art. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 22(12):1424–1445, 2000.
- [17] M. Pantic and L. Rothkrantz. Expert system for automatic analysis of facial expression. *Image and Vision Computing*, 18(11):881–905, 2000.
- [18] S. Safavin and D. Langrebe. A survey of decision tree classifier methodology. *IEEE Transactions on Systems, Man and Cybernetics*, 21(3):660–674, 1991.
- [19] A. Samal and P. Iyengar. Automatic recognition and analysis of human faces and facial expressions: A survey. *Pattern Recognition*, 25(1):65–77, 1992.
- [20] Y.-L. Tian, T. Kanade, and J. F. Cohn. *Hand book of face recognition*. springer, 2004.
- [21] Viola and Jones. Robust real time object detection. *Proceedings, 2nd International Workshop on Statistical and Computational Theories of Vision*, 2001.