Feature Extraction for Non-frontal Faces

Anh Tuan Do
Department of Information Technology, Hanoi University of Mining and Geology, Hanoi, Vietnam
Email: anhtuan140782@gmail.com

Thi Duyen Ngo and The Duy Bui
Human Machine Interaction Laboratory, University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam
Email: {duyennt, duybt}@vnu.edu.vn

***Abstract*—One of the most challenge tasks in building a face recognition system is how to represent and extract good quality features from face images. The difficulties come from variations in head poses, illumination conditions, and facial expression. Although many researches have been done, most were carried on under constrained environments. Most researches concentrated on dealing with frontal faces. Processing non-frontal faces encounters more challenge because some features on faces become occluded dramatically. In this paper, we propose two models to extract features from non-frontal faces in the range of 30^o to 90^o. First, we use the Viola-Jones detection method to identify the pose of face images. Then, we use Active Appearance Model (AAM) to interpret face images. Lastly, the models are trained to know how to fit new images. To improve the efficiency of fitting, we apply a nonlinear parameter update method. Experimental results show that using nonlinear fitting for non-frontal can increase the accuracy of the AAM fitting, compared with some previous methods.**

***Index Terms*—active appearance model, facial feature extraction, non-frontal faces, nonlinear aam fitting, face recognition, expression recognition**

I. Introduction

Human face plays an essential role in human communication. As indicated by Mehrabian [1], in face-to-face human communication only 7% of the communicative message is due to linguistic language, 38% is due to paralanguage, while 55% of it is transferred by facial expressions. Therefore, in order to facilitate a more-friendly human–machine interface of new multimedia products, vision-based facial gesture analysis has been studied worldwide in the last ten years. Many applications in human-computer interaction require tracking a human face and facial features. It is prerequisite for further studies such as face recognition or facial expression analysis.

Facial feature extraction encounters difficulties due to variations in head poses, lighting conditions, and feature occlusion. Therefore, this research field is becoming an active area with many methods proposed to increase the effectiveness of extraction process [2]-[8]. However, most researches were carried on in strictly controlled environments such as frontal face images. In real world scenarios, it requires capturing faces in a wide range point of views. In such cases, occlusion occurs due to large facial rotation and it makes extraction process more difficult. To deal with this challenge, there are few research proposed with different approaches.

Proposed by Cootes *et al.* [9], AAM is one of the most powerful methods for modeling deformable objects. The construction of an AAM uses a set of training images annotated with landmark points to build a statistical model of shape and texture, then they are combined to form an appearance model. In the next step, the combined model is trained to know how to fit a new image by learning the relationship between the texture residuals and the model parameters. The efficiency of the face image representation of AAM has been proved by many researches, and it has been used widely for face recognition [10]-[15], and expression recognition [16]-[20]. In such systems, AAM plays an important role in facial feature extraction processes, which provide input data for classifiers.

In this paper, we propose two AAM models for extracting facial features from non-frontal face images: a model for half-profile (30^o-60^o) face and a model for profile (60^o-90^o) face. In order to determine which model is used for a face, we build a multi-view Viola-Jones detector. We perform experiments using different fitting methods applied to the two models. Our models show that using the nonlinear discriminative method in the fitting task for non-frontal face obtains good results, compared with two other linear methods. The improvements are not only in the success rate of matching but also the computational cost. Our methods also work in a wider range of the initial positions used for the fitting process than the linear AAM fitting algorithms do.

The rest of the paper is organized as follows. In Section 2, we show a summary on related works. Then, in Section 3, we propose two models for extract facial feature points from non-frontal face images. Experiments and analytic results are presented in Section 4. Finally, Section 5 includes conclusions and future works.

Manuscript received February 1, 2015; revised May 21, 2015.

doi: 10.12720/joace.4.2.171-176

II. Related Work

Variations in head pose are still a challenging task in feature extraction, so most of works focus on frontal faces, only are few researches done on non-frontal face. In order to represent a wide range of pose, two existing approaches have been used: 2D based approach and 3D based approach.

The 2D based approach uses the view-based method which combines different viewpoints to represent faces in a wide range of pose. In 1994, Pentland *et al.* [21] introduced view-based eigenface models to present a wide variety of viewpoints (poses from -90^o to +90^o). They described faces with a set of eigenspaces. The best eigenspace describing faces was determined by calculating the residual description error ("the distance-from-face-space" metric), then it was used to encode the face image. However, this work is sensitive with illumination variations. Romdhani *et al.* [22] extended the Active Shape Model to deal with full 180^o rotation of a face using a non-linear model, but the method is slow to match a new image.

In 2000, Cootes *et al.* [23] introduced view-based active appearance models to demonstrate that a small number 2D linear statistical model are sufficient to capture the shape and appearance of a face from a wide range of viewpoints. They built five 2D AAM models to represent faces in 180^o of angle viewpoints where 0^o, ±45^o, ±90^o corresponds to fronto-parallel, half profile and full profile, respectively. The five models roughly centered on viewpoints at -90^o, -45^o, 0^o, 45^o, 90^o. Each model was trained independently, so different models had different sets of parameters. The relationships between the models of different views are learnt, then they are used to predict new views of a face seen from one view. However, the models only work well in case of initial position for starting fitting algorithms close to the optimal position. Many researches have proved that the original algorithm proposed by Coote is not efficient both in accuracy and capture range [10, 24-26].

Since 2003, 3D models have been developed to represent faces. Blanz and Vetter introduced a 3D model to deal with a wide range of viewpoint [27]. They estimated the 3D shape and texture of faces from a single image. The estimation was achieved by fitting a statistical, morphable model of 3D faces to images. The model learned from the set of texture 3D scans of heads. The disadvantage of the 3D model is high computational complexity. To share the advantages of both 2D and 3D models, Xiao describe a combined 2d+3D AAM model [28]. The model used a non-rigid structure-from-motion algorithm to build the corresponding 3D shape modes of a 2D AAM. They showed how the 3D models can be used to constrain the AAM so that it only generates model instances that can also be generated with the 3D models.

Recently, Jason Saragih and Roland Goecke [24] proposed a nonlinear discriminative approach to AAM fitting. They showed that this method was faster and more accurate than two common methods for AAM fitting, the fixed Jacobian method [9] and the project-out inverse compositional method [25]. In addition, the method also exhibits exceptionally good convergence rates with a large capture range. However, in practice, it has been applied only on frontal faces.

Techniques to extract facial features can be divided into two categories: appearance-based methods and model-based methods. The appearance-based methods need good quality images and are affected strongly by light conditions. The model-based methods try to build a model of a human face; these methods can be 2D or 3D. The 3D models are usually more complicate and require a huge computational cost. On the other hand, AAM is the most powerful 2D model but it has been used only by Cootes *et al.* [23] to represent non-frontal faces in an attempt to deal with large pose variations. Since being introduced in 1998, a number of alternatives to the original AAM exist to improve the efficiency of the model [10]. Our solution presented in this paper will eliminate the limitations of non-frontal face models proposed by Cootes *et al.*

III. Our Approach

We use a set of separate 2D AAM models to represent faces under different poses. To deal with full 180^o rotation (from left profile to right profile), we need five AAM models. Each model represents faces in a limited range of camera angles as shown in Fig. 1. Each model were built and trained independently, and it has its own set of parameters. To estimate the pose of a face, we build different Viola-Jones detectors for different views of the face. The pose estimator determines which the most adequate AAM model to a face is (see Fig. 2).

-90^0:-60^0	-60^0:-30^0	-30^0:+30^0	30^0:60^0	60^0:90^0
left profile	left half-profile	frontal	right half-profile	right profile

Figure 1. Five models for full 180^o rotation of a face.

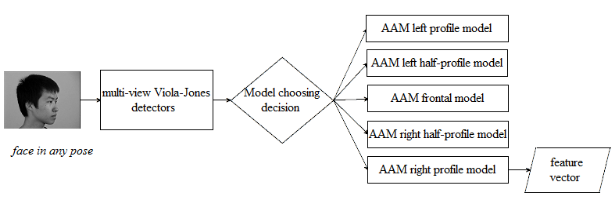

Figure 2. View-based AAM for large pose variation.

The pairs of the left profile model and the right profile model and the pairs of the left half-profile model and the right half-profile model are simply reflection of each other. Therefore, we only need to build three distinct models. There have been many researches on building frontal face model by using AAM, so in this paper, we only investigate two models: the right half-profile model and the right profile model. Our models are based on three phases: modeling appearance of face, training models and fitting a new face image (shown in Fig. 3).

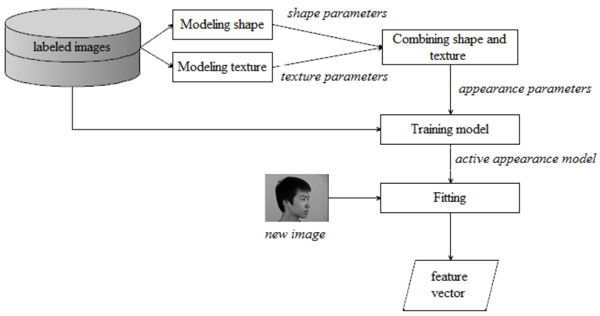

Figure 3. Our approach for non-frontal face feature extraction.

A. Building Appearance Models

Model building phase involves the following steps:

1. Designing a training set of labeled images: The half-profile model have a set of 150 face images of 30 individual; each image was annotated with 49 landmark points. The profile model was also trained on a set of 150 face images of 30 individuals but each image was annotated with 22 landmark points (see Fig. 4). All images are in gray and at 640x480 pixels resolution.

2. Building shape models: Shapes are modeled based on the labeled landmark points. The shapes are normalized using the Procrustes Analysis, then the Principal Components Analysis (PCA) technique is applied to reduce the dimensionality and obtain a model:

$$x = \bar{x} + P_s b_s \quad (1)$$

where $\bar{x}$ is the mean shape, P_s is a set of orthogonal modes of variation and b_s is a set of shape parameters.

3. Building texture models: Based on the shape-free patch, the texture can be raster scanned into a vector g. Then the texture is linearly normalized by:

$$g = (g_{im} - \beta 1)/\alpha \quad (2)$$

where values of α and β are chosen to best match the vector to the normalized mean. Applying PCA to the normalized data we obtain a linear model:

$$g = \bar{g} + P_g b_g \quad (3)$$

where $\bar{g}$ is the mean normalised grey-level vector, P_g is a set of orthogonal modes of variation and b_g is set of grey-level parameter.

4. Building combined models: the coupled relationship between the shape and the texture is analyzed by PCA. In the end, the shape and the texture can be described as follow:

$$x = \bar{x} + Q_s c \ , \ g = \bar{g} + Q_g c \quad (4)$$

where Q_s, Q_g are matrices describing the modes of variation derived from the training set, and c is a vector of appearance parameter controlling both the shape and the texture.

By retaining 95% of shape and texture variation and 98% of combined appearance variation both models, the half-profile model has 21 modes of shape and 76 modes of texture and 58 combined modes; the profile model has 14 modes of shape, 70 modes of texture and 54 combined modes. Fig. 5 shows faces created by the half-profile model.

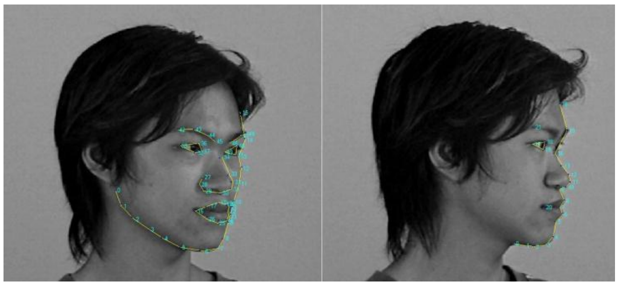

Figure 4. Landmark points of half-profile model and profile model.

Figure 5. Example of faces generated by the models.

B. Training Models

The aim of training phase is to find a strong regressor for each model parameter. Saragih [24] described the method that learns the multivariate regressor through a boosting procedure, where the update function for the k^{th} parameter takes the following form:

$$F^k(g_s) = \sum_{t=1}^{n_k} \alpha_i^k f_i^k(g_s) \quad (5)$$

where $f_i^k \in L$, F^k is the strong regressor composed of a number of n_k week learners f_i^k with corresponding weights α_i^k. L is a dictionary of weak learners.

The final update model is a concatenation of the updates for every parameter:

$$\Delta p = [F^1(g_s); \ldots; F^{N_p}(g_s)] \quad (6)$$

where N_p is the total number of parameters.

C. Nonlinear AAM Fitting

After a training phase, we apply a nonlinear fitting procedure as described below to obtain the appearance parameters considered as a feature vector of a face.

Require: image I, $\{F_1, .., F_n\}$ and p
1. for i=1 to n do
2. Get feature vector g_s
3. Calculate integral images from g_s
4. Calculate parameter updates using F^i (6)
5. Update parameters $p \leftarrow p + \eta\Delta p$
6. end for
7. return p

IV. Experiments

A. Experiment 1

We built five different Viola-Jones detectors for five views with the angle from -90° to 90°. Fig. 6 shows some examples of face images that we used to train and test. The results of the detectors are shown in Table I.

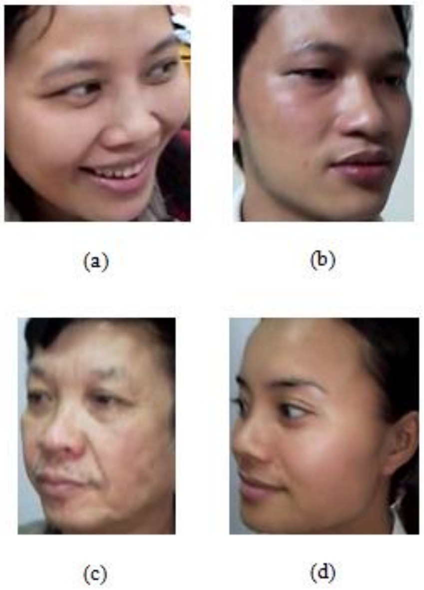

Figure 6. Example of faces in different views. (a) Right half-profile with the angle from 30-60 degree. (b) Right profile with the angle from 60-90 degree. (c) Left half-profile with the angle from 30-60 degree. (d) Left profile with the angle from 60-90 degree.

TABLE I. EXPERIMENTAL RESULTS OF DETECTING VIEWS

Views	Number of test images	Correct	Rate
Left profile (60^o-90^o)	500	472	94.4%
Half-left profile (30^o-60^o)	500	486	97.2%
Frontal	538	521	96.8%
Half-right profile (30^o-60^o)	500	488	97.6%
Right profile (60^o-90^o)	500	470	94%
All	2538	2437	96%

B. Experiment 2

We train the two proposed models using both linear and nonlinear methods. We choose two linear method called the fix Jacobian and linear regression [9] to compare with the nonlinear update method. To evaluate our models, we use different sets of manually labeled 100 unseen face images of 20 individuals for each model. All images are in gray and at 640x480 pixels resolution.

We implement two experiments for each model. First, we randomly perturb all AAM parameters of each test image from the true position 10 times. Therefore, there are 1000 searches on 100 test images for each experiment. In the first experiment, the perturbations are taken randomly within $\pm$ 10 pixels translation, $\pm 10^o$ rotation, $\pm$0.1 scale, $\pm$ 0.1 lighting gain, $\pm$ 20 lighting bias, and $\pm$ 1.5 combined appearance parameters. In the second experiment, we expand the translation to $\pm$ 20 pixels. Our goal to do this is to evaluate the ability of our models in a wider range of the initial positions. After each search we measure RMS (Root Mean Square) distance between model points and hand labeled points (pt-pt error). If the pt-pt RMS error is greater than 7.5 pixels, we consider that the fitting failed to converge. Fig. 7a shows a succesful case, and Fig. 7b shows a failure to fit.

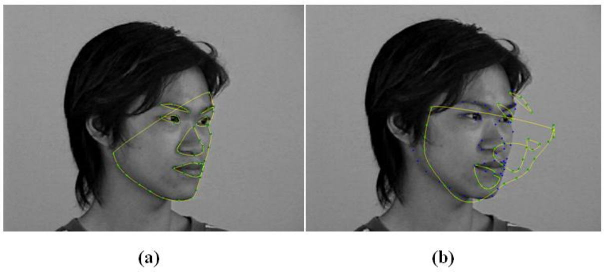

Figure 7. Example of fitting results. (a) Success. (b) Failure.

TABLE II. COMPARISION BETWEEN THE NONLINEAR AND THE TWO LINEAR UPDATE METHODS WITH INITIAL POSITION DISPLACED BY ±10 PIXELS FROM THE OPTIMAL POSITION

Fitting method	Half-profile model		
	Pt-Pt error (pixels)	Success rate (%)	Time (ms)
Fixed Jacobian	3.2	81.2	315
Linear regresion	4.9	85.3	281
Nonlinear	4.4	97.7	56
	Profile model		
Fixed Jacobian	3.4	78	147
Linear regresion	3.9	86.6	109
Nonlinear	4.3	96.4	13

TABLE III. COMPARISION BETWEEN THE NONLINEAR AND THE TWO LINEAR UPDATE METHODS WITH INITIAL POSITION DISPLACED BY ±20 PIXELS FROM THE OPTIMAL POSITION

Fitting method	Half-profile model		
	Pt-Pt error (pixels)	Success rate (%)	Time (ms)
Fixed Jacobian	3.1	40.4	295
Linear regresion	5.1	49	300
Nonlinear	4.4	96.5	54
	Profile model		
Fixed Jacobian	3.4	30	125
Linear regresion	4.3	46	102
Nonlinear	4.3	94.3	13

Table II and Table III summarize the results. We compute the average pt-pt errors based on the fitting which converge successfully. The experiments were carried on a computer with Intel Core i5 1.7 GHz and 6Gb RAM. Although the average pt-pt error of the fixed Jacobian method is the smallest, its failure rate is highest, and this method also runs slowly. In the Table III, we can see that both two linear methods have very poor successful convergence rates when the initial position for starting fitting algorithms is not close to the true position (the position of the best optimal fitting). Whereas the successful convergence rate of the nonlinear method still maintain quite high, above 94% around ± 20 pixels translation. Besides, the nonlinear method is much faster than both the two linear methods.

V. Conclusion

In this paper, we proposed an approach to deal with a wide range of pose by using a set of different views. To detect a face and its pose, we built multi-view Viola-Jones detectors. Then we used AAM models to interpret faces for each view. In this research, we only focused on two non-frontal AAM models: a model for right half-profile (30^o-60^o) faces and a model for right profile (60^o-90^o) faces.

First, we show that multi-view face detection using the Viola & Jones method brought good results with great accuracy. Another advantage of the Viola & Jones method is very fast. Second, we have demonstrated that the models using the nonlinear fitting method are faster and more accurate than the models using linear fitting methods that were used in few previous works. The two linear methods are hard to fit face images to the models, especially the initial position for the search procedure is far away from the optimal position. When we test the two methods with distance of around ± 20 pixels, the linear methods have a poor success rate, whereas the nonlinear method is still good. Furthermore, the models applied the nonlinear method are much faster.

Our work is an attempt to deal with unconstrained environments. Based on the promised results of this work, we will use the models as a feature extractor for non-frontal face recognition or face expression recognition.

References

[1] A. Mehrabian, "Communication without words," *Psychology Today*, vol. 2, no. 4, pp. 53–56, 1968.

[2] J. Shelton, G. Dozier, K. Bryant, J. Adams, K. Popplewell, T. Abegaz, K. Purrington, D. L. Woodard, and K. Ricanek, "Genetic based lbp feature extraction and selection for facial recognition," in *Proc. ACM-SE the 49th Annual Southeast Regional Conference*, 2011, pp. 197-200.

[3] I. Marques, "Face recognition algorithms," Universidad del Pais Vasco - Euskal Herriko Unibersitatea, 2010.

[4] S. Chitra and D. Balakrishan, "A survey of face recognition on feature extraction process of dimensionality reduction techniques," *Journal of Theoretical and Applied Information Technology*, vol. 36, no. 1, 2012.

[5] Y. Lu, J. Zhou, and S. A. Yu, "Survey of face detection, extraction and recognition," *Computing and Informatics*, vol. 22, no. 2, 2003.

[6] R. Jaberi, "An overview on model-based approaches in face recognition," in *Proc. 7th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data (AIKED'08)*, University of Cambridge, UK, 2008, pp. 109-115.

[7] A. Hussain, M. S. Khan, M. Nazir, and M. A. Iqbal, "Survey of various feature extraction and classification techniques for facial expression recognition," in *Proc. the 11th WSEAS International Conference on Electronics, Hardware, Wireless and Optical Communications, and Proceedings of the 11th WSEAS International Conference on Signal Processing, Robotics and Automation, and Proceedings of the 4th WSEAS International Conference on Nanotechnology*, 2012, pp. 138-142.

[8] A. Nikolaidis and I. Pitas, "Facial feature extraction and pose determination," *Pattern Recognition*, vol. 33, no. 11, pp. 1783–1791, 2000.

[9] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," in *Proc. European Conf. on Computer Vision*, vol. 2, pp. 484–498, 1998.

[10] X. Gao, Y. Su, X. Li, and D. Tao, "A review of active appearance models," *IEEE Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews*, vol. 40, no. 2, pp. 145-158, 2010.

[11] O. Khayat, H. R. Shahdoosti, and A. J. Motlagh, "An overview on model-based approaches in face recognition," in *Proc. the 7th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases, Artificial Intelligence Series*, WSEAS Press, Cambridge, UK, 2008, pp. 109-115.

[12] J. Guillemaut, J. Kittler, M. T. Sadeghi, and W. J. Christmas, "General pose face recognition using frontal face model," in *Proc. the 11th Iberoamerican Congress in Pattern Recognition*, 2006, pp. 79–88.

[13] M. Nuruzzaman, H. T. Hatim, H. Azham, and A. S. Amir, "Feature extraction using active appearance model algorithm with bayesian classification approach," in *Proc. the 4th International Conference on Computing and Informatics*, 2013, pp. 94-102.

[14] S. Kim, S. T. Chung, S. Jung, S. Jeon, J. Kim, and S. Cho, "Robust face recognition using aam and gabor features," *PWASET 21*, 2007, pp. 493–497.

[15] F. W. Wheeler, X. Liu, and P. H. Tu, "Multi-frame super-resolution for face recognition," in *Proc. IEEE Int. Conf. BTAS*, 2007, pp. 1-6.

[16] B. Abboud, F. Davoine, and M. Dang, "Facial expression recognition and synthesis based on an appearance model," *Signal Process.: Image Commun*, vol. 19, no. 8, pp. 723 -740, 2004.

[17] J. So, S. Han, H. C. Shin, and Y. Han, "Active appearance model-based facial expression recognition using face symmetrical property in occluded face image," *International Journal of Advances in Mechanical and Automobile Engineering*, vol. 1, no. 1, 2014.

[18] T. Hong, Y. B. Lee, Y. G. Kim, and K. Hagbae, "Facial expression recognition using active appearance model," *Advance in Neural Network, Lecture Notes in Computer Science*, vol. 3972, pp. 69-76, 2006.

[19] X. Feng, B. Lv, Z. Li, and J. Zhang, "A novel feature extraction method for facial expression recognition," in *Proc. Joint Conf. Inform. Sci. Issue Adv. Intell. Syst. Res.*, 2004, pp. 371-375.

[20] Z. Riaz, C. Mayer, M. Wimmer, M. Beetz, and B. Radig, "A model based approach for expression invariant face recognition," in *Proc. 3rd International Conference on Biometrics*, Italy, June 2009.

[21] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in Proc. *IEEE Computer Vision and Pattern Recognition*, Seattle, WA, 1994, pp. 84-91.

[22] S. Romdhani, S. Gong, and A. Psarrou., "A multi-view non-linear active shape model using kernel pca," in *Proc. BMVC*, 1999, pp. 483-492.

[23] T. F. Cootes, G. V. Wheeler, K. N. Walker, and C. J. Taylor, "View-based active appearance models," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog*, 2000, pp. 227–232.

[24] J. M. Saragih and R. Goecke, "A nonlinear discriminative approach to AAM fitting," in *Proc. IEEE Int. Conf. Computer. Vis*, 2007, pp. 14–20.

[25] I. Matthews and S. Baker, "Active appearance models revisited," *Int',l J. Computer Vision*, vol. 60, no. 2, pp. 135-164, 2004.

[26] T. F. Cootes and P. Kittipanya-ngam, "Comparing variations on the active appearance model algorithm," in *Proc. Br. Machine Vision Conference, BMVC*, 2002, vol. 2, pp. 837 -846.

[27] V. Blanz and T. Vetter, "Face recognition based on fitting a 3d morphable model," in *ITEE Transactions of Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063–1074, 2003.

[28] J. Xiao, S. Baker, I. Matthews, and T. Kanade, "Real-time combined 2D+3D active appearance models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*, vol. 2, 2004, pp. 535–542.

Tuan Anh Do was born in Lang Son, Vietnam on July 14, 1982. He graduated in Information Technology in 2004 from VNU University of Engineering and Technology. Since 2007, He is with the Department of Information Technology, Hanoi University of Mining and Geology, Hanoi, Vietnam. His research interests include artificial intelligence, image processing and pattern recognition.

Thi Duyen Ngo was born in Nam Dinh, Vietnam on Feb 1, 1983. She received her Bachelor degree in Information Technology in 2005 from University of Engineering and Technology (UET), Vietnam National University – Hanoi (VNUH), where she has been working as a lecturer since 2006. She is curently completing her PhD thesis at UET, VNUH. Her research interests involve techniques for building embodied agents (creating facial emotional expressions for embodied agents, bring emotions to embodied agents' speech, etc).

The Duy Bui was born in Ha Nam Ninh, Vietnam on May 9, 1978. He got his Bachelor degree in Computer Science from University of Wollongong, Australia in 2000 and his Ph.D. in Computer Science from University of Twente, the Netherlands in 2004. He is now working at Human–Machine Interaction Laboratory, University of Engineering and Technology, Vietnam National University, Hanoi. His research includes Machine Learning, Human–Computer Interaction, Computer Graphics and Image Processing.