# Facial Expressions with AAM

Jocelynn Cu
jiji.cu@delasalle.ph

Calvin Enriquez
Calvin_enriquez@yahoo.com

Samuel Bernard Sia
Samuel_bernard_sia@yahoo.com

Gwenabel Marie Caronan
Caronan_gwenie@ymail.com

Yao Tien Huang
Yao_huang2000@yahoo.com

Merlin Suarez
merlin.suarez@delasalle.ph

## ABSTRACT

This study aims to determine the best combination of images that would enable an affect model to recognize Filipino facial expressions with minimum error. On this, several combinations of images are used such that the training set for the model will consider the various skin tones of the target user, ranging from light to brown. Results show that the AAM default training set added with images from FilMED make up the best combination for an appearance model to correctly extract facial features from a Filipino face. The study also identified relevant facial points and features to accurately recognize Ekman's six basic emotions.

## Keywords

Affective computing, facial expressions, image processing, active appearance model

## 1.0 INTRODUCTION

If you want to know how a person really feels, you just have to look at his face. A person's facial expression often complements his conversation with another person. If one knows how to recognize and interpret these expressions, it is easier to respond accordingly. For computers, the recognition and interpretation process is more complicated. Usually, the input to the computer is an image stream of a person's face. The machine would then process these images individually: first, by detecting the face from the image, then it will assign facial points on the face to mark relevant facial features, which will then be extracted to create a face model. The model will be trained, using machine learning techniques, to recognize the different facial expressions and assign the corresponding emotion label.

TALA [3] is an empathic space that provides empathic responses based on the occupant's affective state. It relies on the occupant's facial expression to know how the person feels. MRCAM [1] is a facial expression recognition system designed for TALA. MRCAM uses the Active Shape Model to assign points on the entire region of the face. Distances between the facial points are used to build the face affect model. It achieved 59% accuracy in recognizing facial expressions. SMERFS [5] is a multimodal affect recognition system designed for TALA. SMERFS determined that the eyes, the eye brows, and the mouth are important facial regions for affect recognition. SMERFS achieved 86% accuracy in recognizing facial expressions. Facial points assigned by SMERFS and MRCAM are shown in Figure1. Between MRCAM and SMERFS, both were able to perform automatic recognition of facial expressions albeit the low accuracy rate. This could be attributed to the type of image database that was used during the training phase of the model. MRCAM's affect model was trained using the Cohn-Kanade [6] and Jaffe [7] image databases. SMERFS affect model was trained using the FilMED [4] database. Since the users of TALA are Filipino, these explain why MRCAM did not achieve the expected accuracy. The skin tone and texture, shape of the face are just some of the factors that might have caused the inaccuracies in their facial affect model. For SMERFS, even though it achieved a more decent recognition rate, it is not robust enough to recognize emotions that are not included in Ekman's 6 universal facial expressions. These are the motivation for this study.

This research aims to find an Active Appearance Model (AAM) [2], which literature claims to be better than ASM, fit for occupants of TALA, who are mostly Filipinos. Various datasets will be used to train the appearance model, in an effort to find the best combination that will help it achieve minimum error in recognition. Section 2 of this paper describes the methodology and databases used in the experiments. Section 3 presents the experiments, its results and discussion. And section 4 concludes the paper.

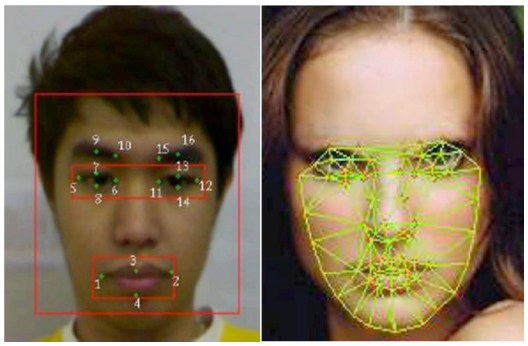


**Fig 1. Facial points assigned by (L) SMERFS and (R) MRCAM.**

## 2.0 METHODOLOGY

The objective of this research is to determine the best combination of images or database that would help an affect model recognize Filipino facial expressions with minimum error.

### *2.1 The Image Databases*

To this effort, the following databases are used: the Cohn-Kanade database and FilMED. Figure 2 shows sample images present in each type of databases.

The Cohn-Kanade version 2 (CK+) database includes both posed and spontaneous facial expressions. The images are labelled according to the emotions the actors are asked to show. It has a total of 593 sequences of images from 123 subjects. Emotions that can be found in the database include neutral, anger, contempt, disgust, fear, happy, sadness, and surprise. Each image has a resolution of 640x490 pixels and averages at 120KB each, stored in PNG format. The database as a whole has a total size of 1.6GB. 65% of the subjects are female, of which 15% are African-American and 3% are either Asian or Latino. This database was designed as a benchmark for research on automatic facial image analysis.

The Filipino Multimodal Emotion Database or FilMED database has at least 400 video clips taken from a Filipino reality television show. Each clip has a resolution of 720x480 pixels, stored as JPG. It includes facial expressions showing neutrality, happiness, sadness, surprise, disgust, fear and surprise. FilMED was designed specifically for SMERFS and other applications that needs to perform multimodal recognition of affect.

**Fig 3. Sample images from (L) Cohn-Kanade and (R) FilMED.**

### *2.2 Facial Feature Extraction using Active Appearance Model*

Images taken from a video stream may contain noise. Thus, several pre-processing techniques have to carried out, which includes brightness correction, gamma correction, and noise reduction. From the full frontal image capture of the face, there is a possibility that error might occur due to the uneven illumination of the light and position of the camera. Histogram equalization is just one technique that one can use to check the intensity values of each pixel in the image and adjust its contrasts to achieve uniform distribution. Sometimes, some regions of the image may be corrupted with noise in the form of speckles or grainy effects. If uncorrected, these may cause miscomputation of later image processing algorithms and cause poor system performance which is difficult to trace. For example, an image averaging filter, convolved with the image, can reduce noise and its effect on the entire image.

Once the image is enhanced and correction, the next step would be to identify the face and track its movement across several frames. The Viola and Jones face detection method is a geometric approach to face detection. It uses Haar-like features, i.e., rectangular object combinations, to determine the face in an image. It does this by subtracting the value of the dark regions and the light regions of the face to form an integral image. A set of cascaded classifier, for example AdaBoost, is then used to classify whether a set of pixels is part of the face or not.

The Active Appearance Model (AAM) has to be built prior to facial feature extraction. AAM is a statistical approach to match and track a face. It requires a training set of labelled images, where key points are marked. Given the set of statistical model of shape, variations can be generated. The strength of AAM lies in its ability to track deformed objects, which implies that it can be effectively used to track various facial expressions.

To build the model, a set of training images with main features marked are needed to generate the shape model. Then these shape models are warped to obtain a "shape-free patch". Then eigen-analysis is applied to get the texture model. The shape-free patch and the texture model are combined to create the appearance model that is used to classify images that are defined within the bounds of the training set.

**Fig 4. Example of a (L) shape-free patch and a (R) normalized patch**

The AAM is then fitted on the detected face on the image. During this fitting process, the algorithm will find the most appropriate shape and texture that fit the face as shown in Figure 5. Based on the best fit, facial points are then automatically placed on the face (shown in Figure 6) and adjusted as necessary. Then, distances between facial points are computed and used as facial features. These facial features are used to build the affective face model to differentiate facial expressions.

**Fig 5. Examples illustrating the AAM fitting process, where the most appropriate shape and texture is matched to the face. The figure on the left is an example of a bad fit, while the figure on the right is an example of a best fit.**

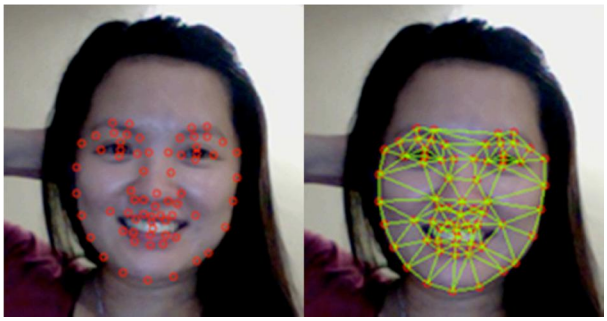

**Fig 6. When a best fit between the AAM and the face is found, facial points are automatically plotted on the face and distances between these points are computed, which serve as facial features.**

### *2.3 Relevant Facial Features for Affect Recognition*

Ekman's six basic emotions included *anger*, *disgust*, *fear*, *happiness*, *sadness*, and *surprise*. All these emotions have corresponding universal facial expressions, which can be classified using a set of facial action units. Feature points that are relevant to affect recognition are identified based on these action units. These were illustrated in Figure 7.


**Fig 7. Relevant facial features for affect recognition**

## 3.0 EXPERIMENT RESULTS AND ANALYSIS

One of the objectives of this study is to determine what types of images are needed to create an active appearance model that is suited to occupants of TALA, whose skin tones vary from light to dark (or brown). Six sets of images are used in the experiments. These are summarised in Table 1.

**Table 1. Summary of image databases.**

| | Source | Images | Images selected |
|---|---|---|---|
| 1 | AAM default images | 21 | Asian, Caucasian |
| 2 | FilMED | 112 | Malay |
| 3 | AAM default + Cohn Kanade | 32 | Asian, Caucasian, Black |
| 4 | AAM default + FilMED | 133 | Asian, Caucasian, Malay |
| 5 | FilMED + Cohn-Kanade | 123 | Malay, Black |
| 6 | AAM default + FilMED + Cohn-Kanade | 144 | Asian, Caucasian, Malay, Black |

From the six sets of images, difference AAMs are built and compared. Figure 8 shows the fittings of the different models on the same image.

**Fig 8. Results of fitting different AAMs on the same image. (Top-down, left-right) the different AAMs are trained using the sets of images summarized in Table 1.**

By visual inspection, the first model (top row, leftmost) was able to assign facial points but it did not follow the contours of the subject's face. The second model (top row, center) made assumptions regarding the position of the eyebrows because the eyebrow of the subject is very light. The third model (top row, rightmost) made an almost perfect fit on the shape of the face and also assumed possible position of the subject's eyebrows. The fourth model (bottom row, leftmost) made a partial fit on the contours of the face while incurring some error on the right side of the chin. The fifth model (bottom row, center) assumed that the darker side of the jaw and cheek are also part of the face, thus included it in the fit. The sixth model (bottom row, rightmost) included the shadows on both sides of the face and portion of the neck in the fitting.

Testing these models on several images resulted in Table 2, which shows the comparison time, model size, and distance error, which is measured relative to the manually plotted points (baseline).

**Table 2. Summary of model comparison.**

| | Model Fitting Speed (in sec) | Model Size (in MB) | Average Distance Error |
|---|---|---|---|
| 1 | 40 – 60 | 44.7 | 6.76 |
| 2 | 60 – 75 | 71.0 | 4.51 |
| 3 | 95 – 100 | 106.0 | 4.59 |
| 4 | 80 – 90 | 91.2 | 4.21 |
| 5 | 80 – 85 | 89.0 | 4.86 |
| 6 | 100 – 120 | 115.0 | 4.63 |

Model #1 spent the least amount of time fitting the model and it also has the smallest model size at 44.7MB. On the other hand, it also has the worst error at 6.76. These can be attributed to the fact that the AAM was trained using its default database consisting of 21 images only. The model with the least error is Model #4, whose training set included images of people whose skin tones range from light to brown, which is also the range of skin tones for Filipinos.

Aside from identifying the relevant features, it is also important to verify if these features can be used in building an affect model. Thus, the combined relevant facial features (shown in Figure 9) identified from Ekman's six basic emotions are used in classification. Using Model #4, three machine learning algorithms: k-Nearest Neighbour (kNN), J48 decision tree, and Bayesian Network are tested for affect modeling. Summarized classification results are shown in Table 3.

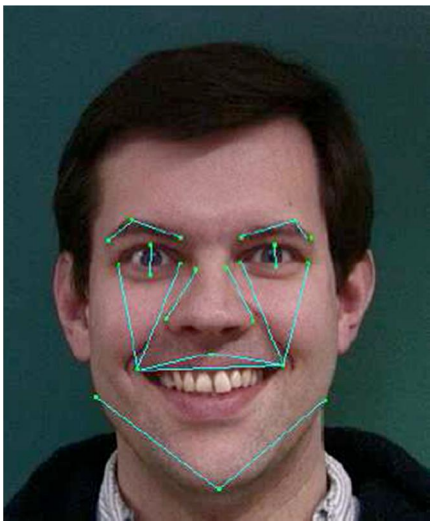

**Fig 9. Relevant facial features for classifying facial expressions.**

Model #4, trained using the combination of AAM default images and FilMED, was used to classify the six basic emotions. Results of classification shows that the k-Nearest Neighbour classifier, with k = 1, works best with a recognition accuracy of 91.28%. This may be due to the tendency of kNN to overfit. J48 also returned very good accuracy in this experiment because there are only 6 discrete classes. In this were used in classifying continuous data, such high accuracy may not be possible.

**Table 3. Results of classification using machine learning.**

| Classifier | Recognition Accuracy |
|---|---|
| kNN (k = 1) | 91.28% |
| kNN (k = 20) | 85.85% |
| kNN (k = 40) | 81.52% |
| J48 | 87.32% |
| Bayesian Network | 78.05% |

## 4.0 CONCLUSION

This study aimed to identify the best combination of images for the extraction of facial features that would help an affect model recognize Filipino facial expressions with minimum error.

Based on experiment results, it was determined that images from the AAM default database and the images from FilMED is the best combination to train an active appearance model for application in the TALA empathic space. One must note that when designing training sets and algorithms for facial feature extraction, skin tone is a large factor. Different training sets produce different results. If automatic detection of skin tone is implemented, it will greatly improve the performance of existing algorithms; otherwise, one has to reduce all images to grayscale before performing feature extraction. Although the latter approach may improve performance, it will also introduce additional delay in computation.

Facial occlusions and head orientation are also important issues that one has to solve in order to arrive at a robust facial expression recognition system. For example, people who are showing sarcastic facial expression may have an asymmetric facial muscle movement like a sideways smile combined with one raised eyebrow. For now, these were not included in this study and all affect recognized are limited to the basics.

The relevant facial points extracted by AAM and machine learning algorithms were used to build the affect model. Although the affect model works very well, the problem with this approach is its fitting time. If the affect recognition system is to operate in real time, this approach will not work. A different approach for facial feature extraction is needed that balances real-time implementation and correct facial points extraction. Future efforts in this aspect will be focused on finding an approach that is fast and reliable.

## 5.0 ACKNOWLEDGEMENT

We would like to acknowledge the De La Salle University – University Research Coordination Office (DLSU-URCO) for the support funds; the College of Computer Studies (CCS), the Center for Empathic Human Computer Interactions (CEHCI), the Center for Language and Translation (CeLT), the Department of English and Applied Linguistics (DEAL), the Filipino Department, and the Psychology Department for providing us the theoretical background and raising research issues concerning this project. This research is supported in part by the Department of Science and Technology – Philippine Council for Industry, Energy and Emerging Technology Research and Development (DOST-PCIEERD).

## 6.0 REFERENCES

[1] Asedillo, E., Ching, M., Ribas, E., Veto, I. and Suarez, M. (2010) Mood recognition using combined algorithms and methods. Undergraduate thesis, De La Salle University.

[2] Cootes, T., Edwards, G. and Taylor, C. (1998) Active appearance models. In Proc. European Conf on Computer Vision (Burkhardt and Neumann Eds.), 2, 484-498, Springer.

[3] Cu, J., Cabredo, R., Cu, G., Legaspi, R., Inventado, P., Trogo, R., Suarez, M. (2010) The TALA empathic space: integrating affect and activity recognition into a smart space. HUMANCON 2010, 1-6.

[4] Cu, J., Suarez, M. and Sta. Maria, M. (2010) A Filipino multimodal emotion database. In Proc Int'l Workshop on Multimodal Corpora (MMC): Advances in Capturing, Coding and Analyzing Multimodality, in assoc. with the 7th Int'l Conf. on Language Resources and Evaluation (LREC 2010), 37-42.

[5] Dy, M., Espinosa, I., Go, P., Mendez, C. and Cu, J. (2010) Multimodal emotion recognition using a spontaneous Filipino emotion database. IWEC 2010, 1-5.

[6] Lucey, P., Cohn, J. Kanade, T., Saragih, J., Ambadar, Z. and Matthews, I. (2010) The extended Cohn-Kanade dataset (CK+): a complete facial expression dataset for action unit and emotion-specified expression. In 3rd IEEE CVPR for Human Communicative Behavior Analysis. http://www.pitt.edu/~jeffcohn/CVPR2010_CK+2.pdf

[7] Lyons, M., Budynek, J. and Akamatsu, S. (1999) Automatic classification of single facial images. In IEEE Trans. on Pattern Analysis and Machine Intelligence, 21(12), 1357-1362.

## 7.0 AUTHOR'S INSTITUTIONAL AFFILICATIONS

Center for Empathic Human Computer Interactions, De La Salle University, 2401 Taft Avenue, Manila, Philippines, 1004.