Presented at the DLSU Research Congress 2016
De La Salle University, Manila, Philippines
March 7-9, 2016

Region-Specific Facial Recognition for Identifying Positive and Negative Filipino Laughter Emotions

Katrina Ysabel Solomon[1,*]
[1]*College of Computer Studies, De La Salle University*
**katrina.solomon@dlsu.edu.ph*

Abstract: Facial recognition is a common technique in identifying emotions exhibited by a person. It is considered as an essential factor in emotion recognition systems. To make this method more effective, particularly in differentiating positive from negative emotions, a different approach must be applied. This research presents an alternative methodology for facial recognition in classifying emotions in Filipino laughter. In analyzing a facial expression, the face will be divided into two regions namely the upper facial and the lower facial regions. The motivation for this approach is based on the notion that positive emotions are more discernable on the lower facial region while negative emotions on the upper facial region as observed from the muscle movements of the face during a laughter segment. Through several experiments, this work aims to provide support for the abovementioned idea. The facial expression analysis begins with the extraction of the facial features through the application of the Active Appearance Model (AAM) algorithm. Models that highlight the two regions were created from the extracted features. Additionally, the most relevant features were identified as well and compared to the original models in order to determine which set would produce better results. To validate the results, Support Vector Machine (SVM), a machine learning algorithm, was employed. The models yielded accuracy rates of 88% and greater. The model for the upper facial region had an accuracy rate of 93.07% and did classify negative emotions successfully. For the lower facial region, it produced an accuracy rate of 91.47% but considered to be insufficient in identifying positive emotions.

Key Words: emotion recognition; facial recognition; machine learning; laughter classification

1. INTRODUCTION

Human emotions are complex and as a result, blending of such emotions may transpire. The blending results in multiple facial expressions. There are cases in which one emotion is occurring on the upper facial region while another on the lower facial region. According to Ekman & Friesen (1975), negative emotions are observed through the upper facial region (e.g., frowning to express anger). Positive emotions are recognized, however, from the lower facial region (e.g. smiling to express joy). Each emotion is characterized by the movements of a specific group of facial muscles. Anger is characterized by pressed lips while disgust is seen from a wrinkled nose and/or raised upper lip (Petta et al., 2011).

Several related projects on blended emotion involve Embodied Conversational Agents (ECA) such as those of Buisine et al. (2006) and Mancini (2008). There are also studies focusing on vocal signals for blended emotion recognition in a call center setting such as the work of Vidrascu & Devillers (2005). The research of Hariharan (2015) makes use of blended emotion recognition for financial decision making.

Laughter is a nonverbal social signal that is a familiar occurrence in human interaction. In most cases, it is perceived as a response to a stimulus that triggers a humorous event (Truong & Trouvain, 2012). This behavior is characterized by various modalities which include vocal and facial expressions and body gestures (Griffin, 2015). Though often viewed as a means of expressing humor, laughter can also be associated with other emotional states, whether they are positive or negative. Suarez et al. (2012) presented a methodology on classifying laughter. The said work made use of the following labels to differentiate laughter: *kinikilig, mapanakit, nahihiya, nasasabik* and *natutuwa.*

Related work on blended emotion recognition is limited, more so on blended laughter emotions. The objective of this research is to build models for the upper facial region and the lower facial region through the utilization of the facial features for classifying laughter emotions. This work aims to aid in improving the classification performance of blended emotion recognition systems.

2. METHODOLOGY

2.1 Data Information

For this experiment, a collection of video clips from the work of Galvan et al. (2011) were used. The clips contain four subjects (two male and two female) exhibiting various laughter segments. There are a total of 270 usable clips in this database. The resolution is set at 720 pixels by 480 pixels and frame rate is 25 frames per second.

Two sets of labels were provided for each clip. The first set is the emotion with respect to the upper facial region while the other is the emotion observed from the lower facial region. For both sets, the following labels are utilized: *kinikilig, mapanakit, nahihiya, nasasabik* and *natutuwa.* The positive emotions are *kinikilig, nasasabik* and *natutuwa* while negative emotions are *mapanakit* and *nahihiya*.

2.2 Feature Extraction

In order to extract the facial features, the algorithm Active Appearance Model (AAM) (Cootes et al., 2001) was employed. AAM uses information such as the shape, gray-level appearance and texture of the face during its training phase. Using this algorithm allows extraction of up to 68 facial points as seen in Figure 1.

Fig. 1. Location of the 68 facial points extracted by AAM

After establishing the facial points, the facial point distances will be computed which will then be used as features to build the model. These distances are illustrated in Figure 2.

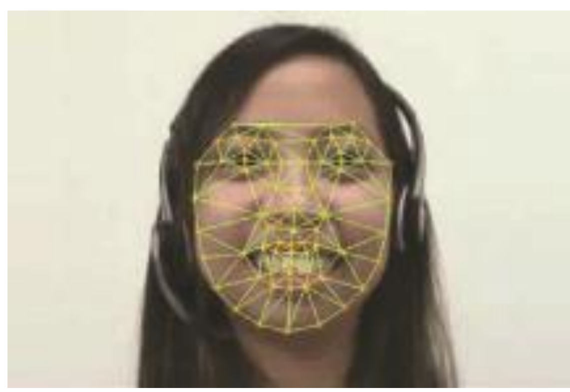

Fig. 2. 170 facial point distances computed from the 68 facial points

In order to build the region-specific models, the facial point distances must be grouped according to their location. For the upper facial region, a total of 81 features were identified. The remaining 89 distances were assigned to the lower facial region. Table 1 lists which distances fall on the upper facial region and the lower facial region. To further illustrate this, the division of the distances is presented in Figure 3.

Table 1. Categorization of the facial point distances according to region

Upper Facial Region	Lower Facial Region
1-10	11-32
33	34
35-104	105-170

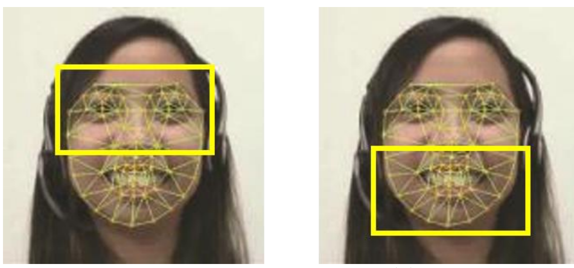

Fig. 3. Facial point distances for (a) upper facial region and (b) lower facial region

2.3 Feature Selection

Luo et al. (2011) identified the relevant features that may lead to a better classification performance. According to the said research, the movement of the eyes and the eyebrows are the most important features for the upper facial region while the mouth movement are the ones for the lower facial region. By using feature selection, another set of models were created to highlight only the relevant features previously mentioned. For the upper facial region, the features were reduced to 44 facial point distances while the lower facial region was reduced to 53 facial point distances. The list of features can be seen in Table 2 and illustrated in Figure 4.

Table 2. Categorization of the relevant facial point distances according to region

Upper Facial Region	Lower Facial Region
44-46	108-117
58-72	120-123
75-84	132-170
89-94	

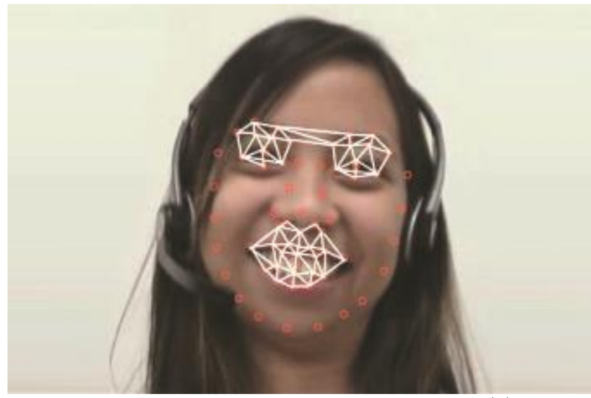

Fig. 4. Relevant facial point distances for (a) upper facial region and (b) lower facial region

2.4 Model Building and Validation

Several machine learning algorithms were tested to find the best classification performance possible. Out of the tested algorithms, the results of the Support Vector Machine (SVM) using the Pearson VII function-based universal kernel (Puk) (Üstün, 2006) will be discussed in this work. The tests were validated using the 10-fold cross validation method. Performance measures taken into consideration for analysis were accuracy rate, kappa statistic, true positive rate, false positive rate, precision, recall and F-measure.

3. RESULTS AND DISCUSSION

3.1 Data Distribution

Four data sets were created in this study. Each set has its own set of features. Two of the four sets are for the upper facial region while the other two are for the lower facial region.

The first data set contains a total of 2296 instances. This is the set that represents the upper facial region with 81 features. Data set two is for the lower facial region and contains 3025 instances and 89 features. The third data set contains the same number of instances as with the first but differ in the number of features used which has been reduced to 44. The last data set is similar to the second data set except it contains 53 features only.

The distribution of the instances for each emotion label is presented in Figure 5.

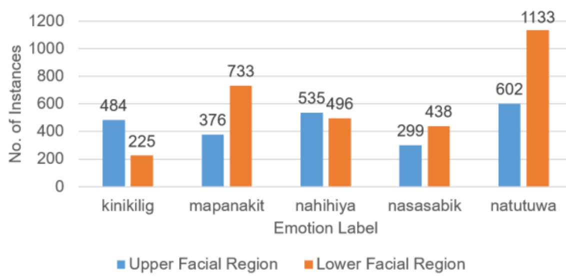

Fig. 5. Distribution of instances

3.2 Validation Results

Metrics used in order to determine the models' performance are the following: accuracy, kappa statistic, true positive rate, false positive rate, precision, recall and F-measure.

Accuracy is the rate of correctly classified test instances over all predictions.

The kappa statistic is the measure of agreement of the prediction with the actual label. A value of 1.0 is an indication of complete agreement. If there is no complete agreement, a value closest to 1.0 is ideal.

True positive rate (TP rate) is the metric demonstrating how much of the class is captured during testing. It is the proportion of instances classified as label *x* over all instances that are actually labeled as *x*. It is similar to the recall metric.

The false positive rate (FP rate) is the percentage of instances labeled as *x* but actually belong to a different label all over instances that is not labeled as *x*.

Precision is the proportion of all instances belonging to label *x* all over instances predicted to have the label *x*.

F-measure is a combination of precision and recall.

The results of the model validation using SVM with Puk are presented in Tables 3-8. Table 3 lists all the accuracy rates and kappa statistics obtained for all data sets. Tables 4-8 details the true positive rate, false positive rate, precision, recall and F-measure values for each label.

Table 3. Accuracy rates and kappa statistics for all data sets

	Data Set			
	1	2	3	4
Accuracy	93.07%	91.47%	90.20%	88.00%
Kappa Statistic	0.9122	0.8864	0.8759	0.8401

Table 4. Summary of results for the *kinikilig* label

	Data Set			
	1	2	3	4
TP Rate	0.965	0.973	0.926	0.924
FP Rate	0.025	0.005	0.032	0.005
Precision	0.910	0.944	0.885	0.937
Recall	0.965	0.973	0.926	0.924
F-Measure	0.937	0.958	0.905	0.931

Table 5. Summary of results for the *mapanakit* label

	Data Set			
	1	2	3	4
TP Rate	0.963	0.876	0.957	0.825
FP Rate	0.014	0.031	0.019	0.039
Precision	0.933	0.902	0.909	0.872
Recall	0.963	0.876	0.957	0.825
F-Measure	0.948	0.889	0.933	0.848

Table 6. Summary of results for the *nahihiya* label

	Data Set			
	1	2	3	4
TP Rate	0.845	0.988	0.821	0.974
FP Rate	0.012	0.015	0.020	0.019
Precision	0.954	0.928	0.924	0.908
Recall	0.845	0.988	0.821	0.974
F-Measure	0.896	0.957	0.869	0.940

Table 7. Summary of results for the *nasasabik* label

	Data Set			
	1	2	3	4
TP Rate	0.926	0.932	0.913	0.895
FP Rate	0.007	0.019	0.016	0.029
Precision	0.952	0.893	0.895	0.838
Recall	0.926	0.932	0.913	0.895
F-Measure	0.939	0.912	0.904	0.865

Table 8. Summary of results for the *natutuwa* label

	Data Set			
	1	2	3	4
TP Rate	0.962	0.890	0.915	0.860
FP Rate	0.030	0.047	0.037	0.071
Precision	0.919	0.920	0.897	0.878
Recall	0.962	0.890	0.915	0.860
F-Measure	0.940	0.904	0.906	0.869

The models that utilized the complete set of features produced more promising results compared to the models that used only the relevant facial point distances. This shows that the most important features alone may not be enough to build a

ISSN 2449-3309

classifier. These relevant features must be taken into consideration with respect to the shape of the face.

If the basis of comparison is the region, the model for the upper facial region produced better results compared to that of the lower facial region. This may be due to the nature of the data being examined. Since all data are laughter segments, regardless of the emotion, they demonstrate similar mouth movements only with varying intensities. There is a strong likelihood that the classifier had difficulty when it comes to characterizing the mouth movements of each emotion.

The models for the upper facial region were more accurate when classifying *nahihiya* and *mapanakit*, both negative emotions, based on their precision values compared to the positive emotions. This supports the idea of Ekman & Friesen (1975) that negative emotions are observable from the upper facial region.

For the lower facial region, the models classified *kinikilig* the best. This is a good indication that it can classify positive emotions well. However, the values for *nahihiya*, a negative emotion, are not far-off. The classifier is still struggling with profiling each emotion given that all instances are subsets of laughter.

4. CONCLUSIONS

The experiments were able to produce promising results as established in the metrics. The goal of building region-specific models was met. This work also provides additional support that negative emotions are more evident on the upper facial region. Improvements can still be implemented to ensure that positive emotions can easily be identified from the lower facial region.

This method of facial recognition can be applied not only by laughter recognition systems, but also by any emotion recognition system in general. This approach is more effective especially when the target labels are blended emotions.

For future work, other modalities such as voice and gestures may be taken into consideration in order to improve the classification performance. Introducing additional data may also help in creating a better profile for each type of laughter emotion.

6. REFERENCES

Buisine, S., Abrilian, S., Niewiadomski, R., Martin, J.-C., Devillers, L., & Pelachaud, C. (2006). Perception of Blended Emotions: From Video Corpus to Expressive Agent. In J. Gratch, M. Young, R. Aylett, D. Ballin, & P. Olivier (Eds.), Intelligent Virtual Agents (Vol. 4133, p. 93-106). Springer Berlin Heidelberg.

Cootes, T. F., Edwards, G. J., & Taylor, C. J. (2001, June). Active Appearance Models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(6), 681-685.

Galvan, C., Manangan, D., Sanchez, M., Wong, J., & Cu, J. (2011). Audiovisual Affect Recognition in Spontaneous Filipino Laughter. In 2011 Third International Conference on Knowledge and Systems Engineering (KSE) (p. 266-271).

Griffin, H.J.; Aung, M.S.H.; Romera-Paredes, B.; McLoughlin, C.; McKeown, G.; Curran, W.; Bianchi-Berthouze, N., "Perception and Automatic Recognition of Laughter from Whole-Body Motion: Continuous and Categorical Perspectives," in Affective Computing, IEEE Transactions on , vol.6, no.2, pp.165-178, April-June 1 2015

Hariharan, A.; Philipp Adam, M.T., "Blended Emotion Detection for Decision Support," in Human-Machine Systems, IEEE Transactions on, vol.45, no.4, pp.510-517, Aug. 2015

Luo, R. C., Huang, C. Y., & Lin, P. H. (2011, July). Alignment and Tracking of Facial Features with Component-Based Active Appearance Models and Optical Flow. In 2011 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM2011) (p. 1058-1063).

Mancini, M. (2008). Multimodal Distinctive Behavior for Expressive Embodied Conversational Agents. Universal-Publishers.

Petta, P., Pelachaud, C., & Cowie, R. (2011). Emotion-Oriented Systems. Springer-Verlag Berlin Heidelberg.

ISSN 2449-3309

Suarez, M. T., Cu, J., & Maria, M. S. (2012, May). Building a Multimodal Laughter Database for Emotion Recognition. In Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12). Istanbul, Turkey: European Language Resources Association (ELRA).

Truong, K. P., & Trouvain, J. (2012, May). Laughter Annotations in Conversational Speech Corpora - Possibilities and Limitations for Phonetic Analysis. In Proceedings of the 4th International Workshop on Corpora for Research on Emotion Sentiment and Social Signals (ES3 2012) (pp. 20–24). European Language Resources Association (ELRA).

Üstün, B., Melssen, W. J., & Buydens, L. (2006). Facilitating the application of Support Vector Regression by using a universal Pearson VII function based kernel. Chemometrics and Intelligent Laboratory Systems, 29-40.

Vidrascu, L., & Devillers, L. (2005). Annotation and Detection of Blended Emotions in Real Human-Human Dialogs Recorded in a Call Center. In IEEE International Conference on Multimedia and Expo, 2005 (ICME 2005).