# Emotion Recognition via Facial Expression: Utilization of Numerous Feature Descriptors in Different Machine Learning Algorithms

John Chris T. Kwong[1], Felan Carlo C. Garcia[2], Patricia Angela R. Abu[3] and Rosula S.J. Reyes[4]
Department of Electronics, Computer and Communications Engineering[1,4],
Solutions and Services Engineering Division[2] Department of Information Systems and Computer Science[3]
Ateneo de Manila University[1,3,4], Advanced Science and Technology Institute[2]
Loyola Heights, Quezon City, Philippines[1,3,4], ASTI Bldg., C.P. Garcia Ave., Quezon City, Philippines[2]
john.kwong@obf.ateneo.edu[1], felan@asti.dost.gov.ph[2], pabu@ateneo.edu[3], rsjreyes@ateneo.edu[4]

***Abstract*— Emotion Recognition has been a prominent study even before computers had the same computing power as of today. Human's emotions can be recognized through their body language, behavior and, most evidently, from the facial expression of the person. In facial image classification, each facial image can be represented through feature descriptors. Feature descriptors are simplified representations of the facial image that incorporates the essential key facial features. This study determines which feature descriptor will best fit a respective machine learning algorithm to classify facial expressions. Twelve possible combinations of Key Facial Detection, Saliency Mapping, Local Binary Pattern, and Histogram of Oriented Gradient are investigated together with six machine learning classification algorithms thus generating a total of seventy-two models. These will classify the following emotions: anger, disgust, fear, joy, neutral, sadness and surprise. A stratified ten-fold cross-validation is performed for verification on both the CK+ dataset and the locally gathered dataset for "in the wild" image testing. This study has determined that among the seventy-two models, the RBF SVM HOG+LBP model attained the highest average accuracy of 0.94 across the seven emotions with an F1 score of 0.93.**

***Keywords—emotion recognition via facial expression, feature descriptor, histogram of oriented gradient, local binary pattern, support vector machine, k-fold cross-validation***

## I. INTRODUCTION

In our day to day communication with other people, facial expression is an essential part in the means of emotion recognition towards others. Human beings display facial expressions when experiencing emotion. The human spectrum of emotion, having different intensities or levels can be shown in a lot of ways, e.g. verbal and non-verbal communications. However, But facial expressions being the main channel. Facial expression is described as the positions of the muscle beneath the skin of the face. Studies pointed out that humans are capable of showing six basic human emotions and a neutral state. The seven basic human emotions are anger, disgust, fear, joy, neutral, sadness and surprise as shown in Fig. 1.

Emotion Recognition via Facial Expressions (ERFE) studies has been increasing in the last two decades since the technology is rapidly advancing along with it. ERFE is mainly being applied in intelligent human-computer interaction [1], safety [2], medicine [3], affective computing and is deeply studied due to its non-intrusive method of recognition.

Anger Disgust Fear Joy

Neutral Sadness Surprise

*Fig. 1. Seven basic human emotions*

Though advancement has been made in both hardware and software, recognizing facial expressions with high accuracy remains challenging as there is much variability and complexity in each face. This is especially true with "in the wild" facial expression recognition. The "in the wild" facial expression recognition is classifying facial images outside the closed-set of classes or the training classes. This is a challenge to facial expression recognition systems which will most likely result in misclassification. This is the objective that many studies have been working on as it is problematic to train every system with all possible facial feature.

The goal of this study is to develop frameworks that can be used in facial expression recognition. To achieve this objective, this study determines which feature descriptor will best fit a respective machine learning algorithm to classify facial expressions by (1) using and examining different feature extraction methods and combinations of the methods as feature descriptors, (2) classifying the images using the feature descriptors by utilizing several machine learning algorithms, and to evaluate their respective performances and (3) determining how the models will perform "in the wild" by introducing locally gathered images.

978-1-5386-5457-6/18/$31.00 ©2018 IEEE

## II. Theoretical Background

### A. CK and CK+ Database

The database utilized was from the Carnegie Mellon University in Pittsburg and is one of the most comprehensive test-bed for comparative studies of facial expression analysis [1]. The images of the database are from a video sequence. Hence, each subject folder has separated emotions and consists of a sequence from a neutral face to each of the respective emotion. The CK database has 2,105 digitized image sequences from 182 adult subjects while the CK+ database increased database by 22% in images and 27% in subjects [1]. The database consists of 65% females, with 15% African-American and 3% being Latino or Asian [5].

### B. Feature Descriptor

Feature descriptors in image processing is simply a thumbprint that is used to distinguish one feature from another. Ideally, a feature descriptor should be invariant. Thus, it should constitute the information if the image is transformed in some way. This study utilized twelve feature descriptors which are from four feature extraction methods, their combinations with some undergoing Principal Component Analysis (PCA) [6] for dimension reduction. The feature extraction methods used were Key Facial Landmark Detection (KFL) [7], Histogram of Oriented Gradient (HOG) [8], Saliency Mapping (SAL) [9] and Local Binary Pattern (LBP) [10].

### C. Machine Learning Classification

For classification, machine learning algorithms have an advantage over conventional pattern analysis or human-crafted rules since machine learning is data driven. Machine learning algorithms are often accurate, cheap and flexible in a sense that it can be applied to any learning task. While a study pointed out that Neural Network is one of the best performing machine learning algorithms, the "No Free Lunch" theorem in machine learning statistics is applicable [11]. Simply put, there is no universally best learning algorithm. A certain algorithm may outperform other algorithms but it does not outperform it all the time or in every problem. Thus, the researcher must be diligent in testing out other machine learning algorithms to truly determine which performs better in a given application. The machine learning algorithms used in this study are RBF Kernel Support Vector Machine (RSVM) [12], Neural Network (NN) [13], Sigmoid Kernel Support Vector Machine (SSVM) [12], Random Forest (RF)[14], Logistic Regression (LR)[15] and K-Nearest Neighbor (KNN)[16].

### D. Stratified K-fold Cross-validation

In *K*-fold cross-validation, the original sample dataset is randomly partitioned into *K* equal-sized subsamples. These *K* subsamples are then separated with a single *K* subsample acting as validation and the rest being training data. This is repeated *K* times or folds, with each *K* subsample used once as validation. The *K* results will then be averaged resulting to a single approximation. This is commonly sought out by validation practices because of how this repetition in random subsamples are used in both training and validation with each observation used for validation once. And with stratified *K*-fold cross-validation, it preserves the class proportions for each fold, resulting in better estimates. [17]. This study will be utilizing ten-fold cross-validation. The accuracy, recall, precision and F1 score of the models are measured for performance metrics evaluation.

### E. "In the wild" Facial Recognition

Most facial expression recognition studies are done in a closed-set of classes. This means that the facial features extracted are from a reference image database, thus encountering facial features that are not in the image database may result to misclassification [18,19]. Facial expression recognition studies using "in the wild" studies are more realistic and grounded as this makes it relevant in the real-world application. This study would test the models to determine how well it performs given "in the wild" images.

## III. Research Methodology

This study will undergo preparation of the dataset, training and testing of the models. The trained models will also be tested with "in the wild" images. A flow diagram of the methodology is shown in Fig. 2. The programming functions and libraries utilized for this study are implemented using Python.

### A. Image Dataset

Previous studies that utilized the CK+ dataset made use of the neutral face as a baseline for their feature descriptors. This study will train the models by using the exacted emotion in each image and not getting the difference from the neutral face thus giving the models no baseline. This study will handpick the last 3 to 4 images in a respective folder that best show the emotion. A total of 1,510 images are categorized and are as follows with its respective image count: anger (147), disgust (153), joy (228), neutral (401), sadness (162) and surprise (264).

This study will also utilize local images for "in the wild" classification. This is to introduce faces that were not part of the training set to test the performance of the models. This study will gather 10 individuals and capture the 7 emotions with their facial expressions. The photo of each individual will be taken 3 times which would provide this study a total of 210 images categorized by the 7 emotions with 30 images per emotion.

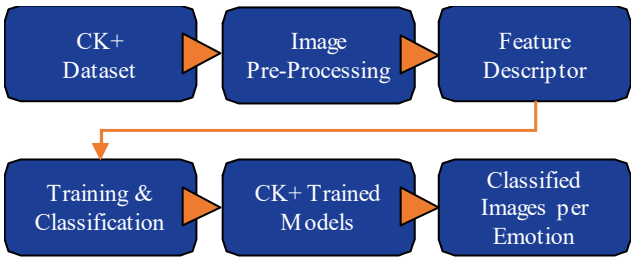
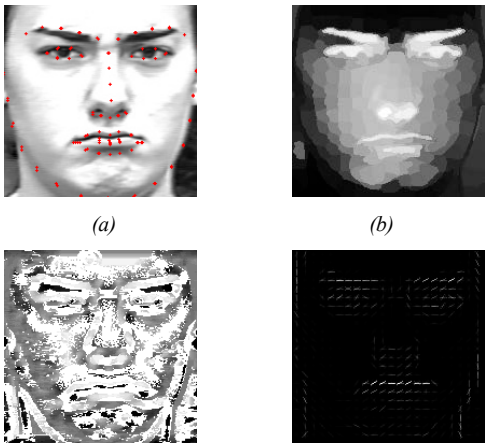



*Fig. 2. Methodology flow diagram*

(a) (b)

*Fig. 3. (a) Original Image (b) Cropped Image*



(a) (b)

(c) (d)

*Fig. 4. (a) KFL (b) SAL (c) LBP (d) HOG*

### B. Image Pre-Processing

Pre-processing the images aims to get the appropriate part of the image which is the face and remove the non-essential part e.g. background. The sorted images will undergo face detection and cropping. This is to limit the image data to the face and its facial features by removing the background as shown in Fig 3.

### C. Feature Extraction

Key Facial Landmark Detection (KFL) is performed by importing the face_utils package from the Dlib library. An image showing KFL detection is shown in Fig.4. (a). The facial image will be divided into the Eye-Aspect Ratio and Mouth-Aspect Ratio. Facial landmarks are determined from the face region (XY-coordinate) which are converted to a NumPy array. The left and right eye coordinates are extracted to compute for the Eye-Aspect Ratio for both eyes. The mouth aspect ratio is also computed using the mouth key points from in to out. The arrays are appended and labeled accordingly thus giving us values for an image with its respective emotion.

Saliency Mapping (SAL) is performed using the pyimgsaliency library. An image undergone SAL is shown in Fig.4. (b). This will compute the saliency of the image and resize to reduce data complexity. The flattened saliency image is converted to an array alongside the corresponding emotion label.

Local Binary Pattern (LBP) is performed using the LBP function from the skimage library. An image with LBP is shown in Fig.4. (c). The cropped images will be computed for LBP representation. Once the local binary representation is computed, the histogram is normalized.

Histogram of Oriented Gradients (HOG) is performed using the HOG package from the skimage library. An image with HOG is shown in Fig.4. (d). HOG feature is extracted from the cropped images. The HOG values are converted to arrays alongside the respective emotion label.

### D. Feature Combination

In selecting the combinations for the feature descriptors, other possible combinations were filtered out due to poor accuracy in the initial simulations and results. One evident initial observation is that adding KFL to other combinations resulted to a drop-in accuracy. The 12 feature descriptor combinations utilized are listed below. The feature descriptors that are concatenated, embodied by the plus sign, will have their array values linked into a long string e.g. HOG has 1,151 columns or features per image and LBP has 25 features, then HOG+LBP has 1,176 features. Some of the selected combinations are reduced using PCA, this is to reduce redundancy in data. Fig. 5 shows a visual representation of the feature combination.

(1) KFL
(2) HOG
(3) LBP
(4) SAL
(5) HOG+LBP+SAL
(6) (HOG+SAL+LBP) PCA
(7) (HOG+SAL) PCA
(8) (HOG+LBP) PCA
(9) HOG+SAL
(10) HOG+LBP
(11) (SAL)HOG
(12) (SAL)LBP

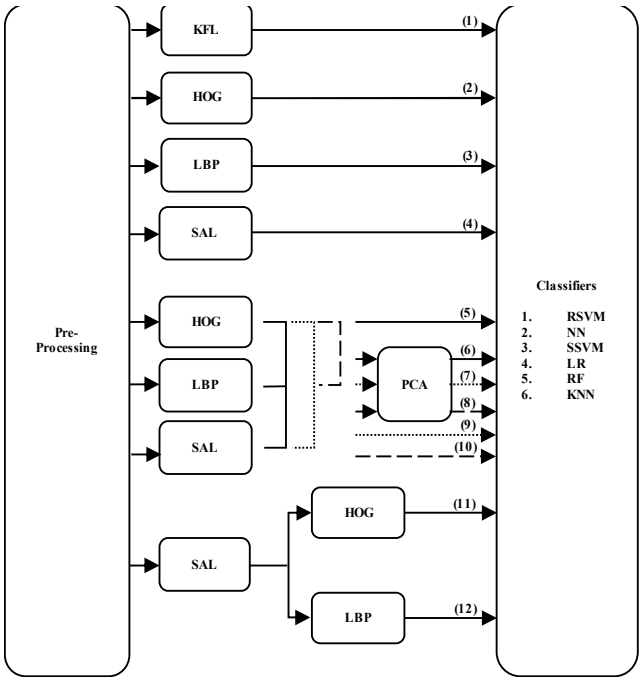



*Fig. 5. Feature descriptors utilized*

### E. *"In the Wild" Image Gathering & Testing*

In order to introduce images outside the CK+ database and determine if the models can be used "in the wild", this study would be locally gathering test images, specifically Filipino facial features.. This is also in consideration of the limitation in machine learning algorithms that the training and test set should come from the same distribution. After, the participant is asked to sit in an upright position directly facing the camera with a 1-meter distance in between. The participants are asked to mimic the emotions aforementioned.

### F. *Validation and Performance Metrics*

In determining the accuracy per model, the accuracy for the emotion are averaged. The F1 score was computed from the recall and precision scores. The accuracies per emotion, recall, precision and F1 scores were obtained after a Stratified Ten-Fold Cross Validation, and are implemented in Python.

## IV. Results

The accuracy for each emotion was obtained from the confusion matrix for each model. An RSVM HOG confusion matrix is shown in Fig. 6. It is shown that in this model, the emotion anger is misclassified 0.15 or 15% of the time as disgust.

In Fig. 7, the RSVM HOG+LBP model attained the highest average accuracy of 0.94 across the 7 emotions and is among the highest F1 scores in this study. This is mainly because of how SVM and the hyperplane in general is excellent in classifying multiple complex values even with close points. The feature descriptor that is mostly in common are combinations of HOG+LBP or partnered with PCA. This indicates that the features have minimal to non-intersecting or overlapping points resulting to better classification.

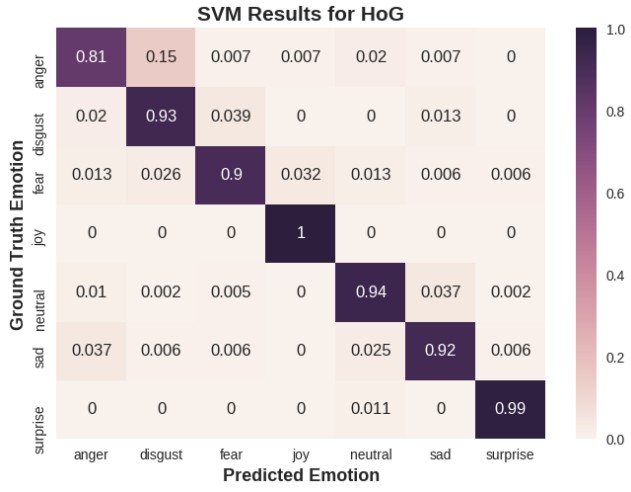



*Fig. 6. RSVM HOG Confusion matrix*


*Fig. 7. Highest accuracies & F1 scores per machine learning algorithm*

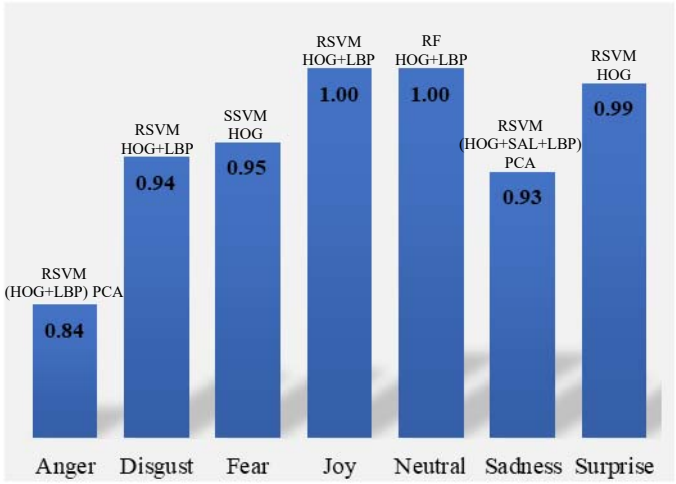



*Fig. 8. Highest accuracy models per emotion*

In the summarized model rank per emotion shown in Fig. 8, it can be seen that the RSVM machine learning algorithm achieving the highest accuracies in 4 of the 7 emotions (anger, disgust, sadness and surprise). This echoes the previous sentiment on how excellent SVM is in general in terms of handling multiple complex data. In the feature descriptors, we see HOG appear thrice as the highest feature descriptor accuracy in 3 emotions namely fear, joy and surprise.

With "in the wild" testing results shown in Fig.9, these models attained the highest average accuracies and F1 scores. The model RSVM HOG+LBP attained the highest F1 score of 0.50. While SSVM HOG+LBP and NN HOG+LBP attained 0.48 and 0.43 respectively. The decrease in accuracies are expected with "in the wild" images as this was not part of the same distribution or training images. Also, a factor to the decrease in accuracies in comparison to the CK+ dataset is that the models were trained with Western facial features and the "in the wild" images were Asian facial features or more specifically, Filipino facial features.

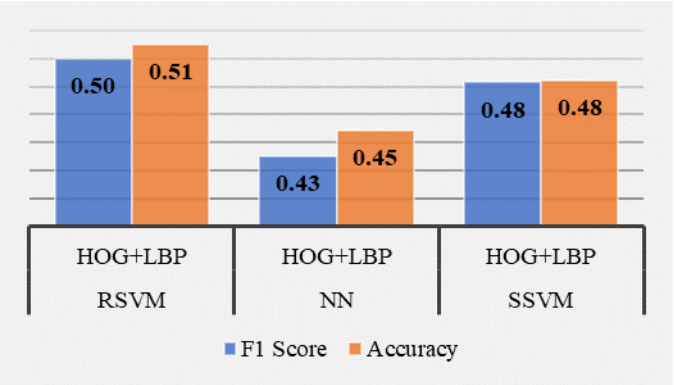



*Fig. 9. "In the wild" testing*

## V. Conclusion

With 12 feature descriptors and 6 machine learning algorithms, this study generated 72 models that classify 7 human emotions. The model that attained the highest accuracy among the 72 models was the RSVM HOG+LBP with an accuracy of 0.94. The models also displayed excellent results in classifying joy and the weakest suit being anger. This study reaches a positive conclusion with 25 out of the 72 models having 0.90 or higher average accuracies and the majority of the models can be used as frameworks for emotion recognition via facial expression.

## References

[1] "Emotional intelligence in robots: Recognizing human emotions from daily-life gestures - IEEE Conference Publication", Ieeexplore.ieee.org, 2018. [Online]. Available: http://ieeexplore.ieee.org/document/7989198/. [Accessed: 01- Apr- 2018].

[2] H. Gao, A. Yüce, J. Thiran, "Detecting emotional stress from facial expressions for driving safety", IEEE International Conference on Image Processing (ICIP), pp. 5961-5965, 2014.

[3] S. Tivatansakul, M. Ohkura, S. Puangpontip and T. Achalakul, "Emotional healthcare system: Emotion detection by facial expressions using Japanese database", 2014 6th Computer Science and Electronic Engineering Conference (CEEC), 2014.

[4] Kanade, T., Cohn, J. F., & Tian, Y. (2000). "Comprehensive database for facial expression analysis". Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), Grenoble, France, 46-53.

[5] Lin Zhong, Qingshan Liu, Peng Yang, Junzhou Huang and D. Metaxas, "Learning multiscale active facial patches for expression analysis", IEEE Transactions on Cybernetics, vol. 45, no. 8, pp. 1499-1510, 2015.

[6] K. Pearson, On lines and planes of closest fit to systems of points in space. London: University College, pp 559-572, 1901.

[7] A. Rosebrock, "Facial landmarks with dlib, OpenCV, and Python - PyImageSearch", PyImageSearch, 2018. [Online]. Available: https://www.pyimagesearch.com/2017/04/03/facial-landmarks-dlib-opencv-python/. [Accessed: 01- Apr- 2018].

[8] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection", 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, United States. IEEE Computer Society, 1, pp.886–893, 2005.

[9] W. Zhu, S. Liang, Y. Wei and J. Sun, "Saliency Optimization from Robust Background Detection", 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014.

[10] C. Shan, S. Gong and P. McOwan, "Facial expression recognition based on Local Binary Patterns: A comprehensive study", Image and Vision Computing, vol. 27, no. 6, pp. 803-816, 2009.

[11] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms", Proceedings of the 23rd international conference on Machine learning - ICML '06, 2006.

[12] C. Cortes and V. Vapnik, "Support-vector networks", Machine Learning, vol. 20, no. 3, pp. 273-297, 1995.

[13] J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities.", Proceedings of the National Academy of Sciences, vol. 79, no. 8, pp. 2554-2558, 1982.

[14] L. Breiman, Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[15] D. Freedman, Statistical Models: Theory and Practice, Cambridge University Press, pp. 26, 2009.

[16] [10]N. Altman, "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression", The American Statistician, vol. 46, no. 3, pp. 175-185, 1992.

[17] S. Raschka and R. Olson, Python machine learning. Birmingham: Packt Publishing, 2016.

[18] W. Dhifli and A. Diallo, "Face Recognition in the Wild", Procedia Computer Science, vol. 96, pp. 1571-1580, 2016.

[19] B. Sun, L. Li, G. Zhou and J. He, "Facial expression recognition in the wild based on multimodal texture features", Journal of Electronic Imaging, vol. 25, no. 6, p. 061407, 2016.

[20] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). "The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression". Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010), San Francisco, USA, 94-101.I.

[21] X. Peng, Z. Xia, L. Li and X. Feng, "Towards Facial Expression Recognition in the Wild: A New Database and Deep Recognition System", 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2016.