Shallow and Depth Learning Approach Application for Facial Emotion Recognition

Tram Vu Kiet
College of Information and Communication Technology
Can Tho University
Can Tho city, Vietnam
tvvkiet@gmail.com

Truong Quoc Bao
College of Engineering Technology
Can Tho University
Can Tho city, Vietnam
tqbao@ctu.edu.vn

Le Hoang Nhi
College of Information and Communication Technology
Can Tho University
Can Tho city, Vietnam
lehoangnhiit@gmail.com

***Abstract*—In this journal, we represent the computer vision and machine learning algorithms in order to detect and recognize human emotion automatically based on facial expression, using the facial landmarks of Dlib library with Support Vector Machine - SVM. The system is able to detect human face and recognize human emotion through facial expression with six common emotions (happy, sadness, surprise, anger, fear and disgust) in real time. The practice result with 4234 images of 65 objects in Cohn-Kanade dataset shows that training with SVM machine learning and Dlib library got 94.33% accuracy on the testing set, which is higher in comparation on training with Convolutional Neutral Network – CNN, SVM machine learning using HOG features (Histogram of Oriented Gradients), k nearest neighbor using HOG features (k-Nearest Neighbor - kNN) with the accuracy are 94.30%, 45.01% và 36.43% respectivelly. The result shows the cable of integrated Support Vector Machine in order to detect human emotion in the real application.**

Keywords—Facial emotion recognition, Haar-like features, facial landmarks, Support Vector Machine (SVM).

I. INTRODUCTION

The system recognizes human emotion bases on human facial expression, extracts the facial features in image or video and recognize emotion bases on the extracted contents of human face automatically. Detecting and recognizing emotion bases on human face that support to building the artificial intelligent systems such as evaluating staff loyal, analyzing customer satisfaction, treating emotion disorders. Emotion analyzing software market is predicted reach at 10 billion USD in 2020 in comparation with 100 million in 2016 [1]. The system recognizes emotion bases on human face in two main step: locate human face in the input image and recognize emotion from that image.

There were many researches about facial emotion recognition such as facial expression recognition using Principal Component Analysis method (PCA) of Ajit P. Gosavi et al., 2013 [2]; Monika Dubey et al., 2016 [3] introduced the necessary and application of facial emotion recognition in the world; James Pao [4] used the face detector of Viola-Jones and Harris in order to get the face location and emotion, used HOG features, PCA method and Support Vector Machine - SVM.

In this article, we suggest the emotion recognition system bases on human face that using facial landmarks in Dlib library and Support Vector Machine [5]. Face locating in image from static image or camera is excuted by Haar-like feature and AdaBoost classifier. Then, the system locates facial landmarks from face in image and training the machine learning model with Support Vector Machine in order to recognition human emotion from face in image. The practice result with 4234 images of 65 objects in Cohn-Kanade dataset shows that our proposed model - training with SVM machine learning and Dlib library got 94.33% accuracy on the testing set, which is higher in comparation on training with Convolutional Neutral Network – CNN, SVM machine learning using HOG features (Histogram of Oriented Gradients), k nearest neighbor using HOG features (k-Nearest Neighbor - kNN) with the accuracy are 94.30%, 45.01% and 36.43% respectively.

The tructure of this article is organize in the following parts: Part I is the introduction. Part 2 shows the facial emotion recognition system. Part 3 presents some practice result. Part 4 shows the conclusion and development direction.

II. FACIAL EMOTION RECOGNITION SYSTEM

Human facial emotion recognition system includes two main modules: training module and recognition module as shown in image II.1.

Training module consists of some tasks: getting images from training dataset, preprocessing images, locating face area in image automatically, determing facial landmarks of the main component of human face and training the recognition model to recognize emotion bases on human face in the system.

Recognition module get the static image or camera, locating face area, locating facial landmarks and using the recognition model (which is trained from training module) in order to recognize emotion.

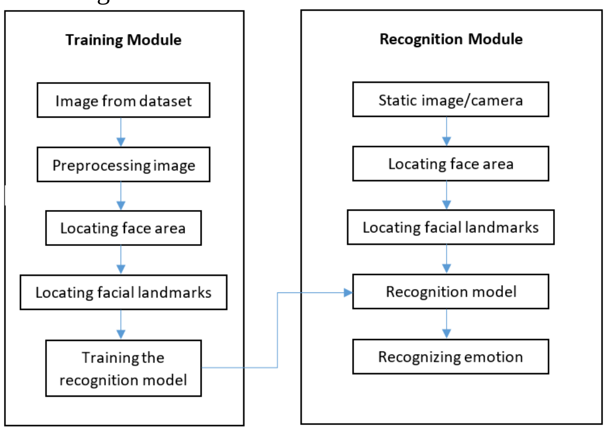


Figure II.1. Facial emotion recognition system main modules

The system needs to excute three main steps: detect face area in image from static image or camera automatically, quickly, exactly; using Haar-like features AdaBoost classifier; locating facial landmarks of the main component of the detected face; training recognition model by SVM in order to recognize emotion bases on facial expression.

A. *Detect face area with Haar-like features and AdaBoost classifier*

Diagram of face detection in images:

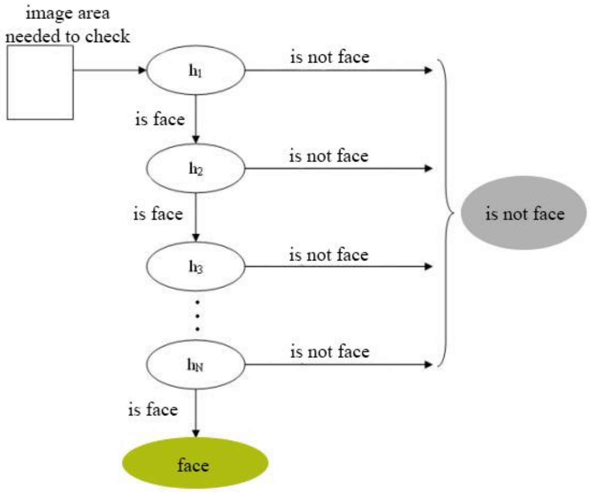


Figure II.2 Detect face area in image

In static image or camera may contain human face or not, its background and maybe other objects. In order to increase system perfomance, it is necessary to detect the candidate area which is human face and not concern the rest area which is not contain human. Using the Haar-like features (using the features as shown in figure II.3) and AdaBoost classifier in order to detect and locate human face in image quickly and exactly as figure II.2.

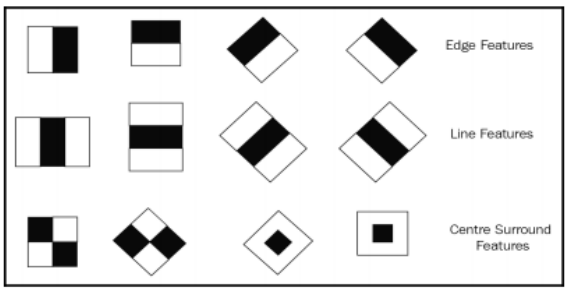


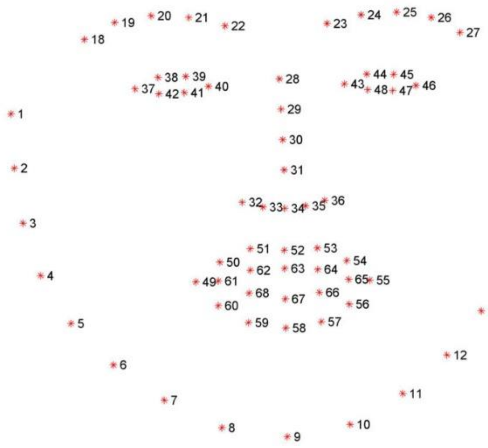
Figure II.3 The three main features of Haar-like features

The advantages of Haar-like are very quickly and its calculation is simplier than other methods, so it helps locate human face quickly and exactly in image from camera.

B. *Locating facial landmarks from image*

When the face is detected in image, the next step is extracting the facial features and calculating the relevant features in order to create facial emotion recognition model. We propose that using facial features of Dlib library because its simplification and high performance was proven in facial emotion recognition field.

The more facial landmarks were found, the higher performace in emotion recognition. In this article, we used the approach of Kazemi và Sullivan (2014) [6] in order to extract 68 facial landmarks effectively (was shown in figure II.4). The facial landmarks were divided into six main categories: eyes, eyebrows, eyelids, nose, lips and jaw. Each emotion has a relevant expression of facial landmarks.



Hình II.4 68 facial landmarks of Dlib library

C. *Traing recognition model with Support Vector Machine*

Step to extract the face landmark create n-dimensional array data (the number of images in the training data set) and each line is a set of points representing the coordinates of the corresponding face landmark of that image; The number of labels increased from 0 to the number of emotions to be identified. Specifically, the corresponding emotions and labels are assigned as follows: *anger-0*, *contempt-1*, *disgust-2*, *fear-3*, *happy-4*, *neutral-5*, *sadness-6*, *surprised-7*. SVM machine learning algorithms ranked third in the 10 most efficient machine learning algorithms [7] recommended for classifying and recognizing human emotions based on the face.

Consider for example the simple linear binary classification described in Figure II.5, with m elements $x_1, x_2, \ldots, x_m$ in n-dimensional space with the label (class) of the corresponding elements $y_1, y_2, \ldots, y_m$ has the value 1 (positive) or -1 (negative). The SVM machine tool [5] finds the optimal hyperplane (defined by the normal vector w and the deviation of the hyperplane with the quadrant b) to split the data into two classes. SVM find hyperplane separating the two classes as far as possible (optimal hyperplane) based on two parallel supports hyperplane of two class. The support hyperplane of the +1 class ($w.x\text{-}b=+1$) is a hyperplane whose x_p elements belong to the $y_p = +1$ class to the right of it, means: $w.x_p - b \geqslant +1$. Similarly, the support hyperplane of class -1 ($wx\text{-}b=\text{-}1$) is a hyperplane where the elements x_n belonging to the class $y_n = -1$ class to the left of the support hyperplane of class -1, means: $w.x_n - b \leqslant -1$. Elements in the opposite direction to the hyperplane are considered errors. The error distance represented by $z_i \geq 0$ (with x_i on the right side of its supporting hyperplane, the corresponding error distance $z_i = 0$, otherwise $z_i > 0$ is the distance from the point x_i to its corresponding hyperplane support). The distance between two supports hyperplanes is called the margin = $2 / \| w \|$, where $\| w \|$ is the magnitude (2-norm) of the vector w. The optimal hyperplane (located between two support hyperplanes) needs to maximize marginalization (greater margin, more secure classifies model) and minimize errors. The optimization problem of the SVM algorithm leads to the solution of the quadratic planning problem (1):

Quadratic planning problem recipe:

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} y_i y_j \alpha_i \alpha_j K\langle x_i, x_j \rangle - \sum_{i=1}^{m} \alpha_i \qquad (1)$$

With constraints:

$$\begin{cases} \sum_{i=1}^{m} y_i \alpha_i = 0 \\ 0 \le \alpha_i \le C \quad \forall i = 1, 2, \ldots, m \end{cases}$$

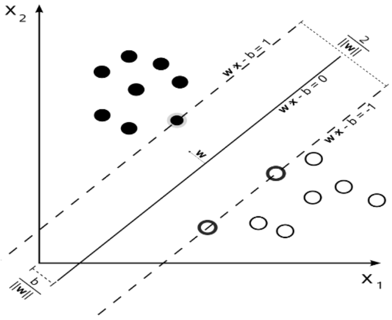


Figure II.5 Example of linear classification with SVM

Inside, C is positive constant, is used to fix the partition margin width and sum of erros distance. K<x_i, x_j> is linear kernel K<x_i, x_j> = <x_i•x_j>.

Solving the problem (1) obtained dataset *#SV* with x_i element corresponding to $\alpha_i > 0$, is called *support vectors*. Only with this *#SV* support vector we can reconstruct the classification hyperplane. SVM model classifies new element *x* by (2):

$$predict(x) = sign\left(\sum_{i=1}^{\#SV} y_i \alpha_i K\langle x, x_i \rangle - b \right) \quad (2)$$

SVM machine learning can use different kernel functions to solve nonlinear. In order to handle nonlinear classification problems, there is no need to change of the algorithms instead of just replace the kernel functions in (1) and (2) with other kernel functions. The common nonlinear kernel is the Radial Basic Function (RBF) [7]. There are also linear, sigmoid, poly kernel functions.

The SVM machine model is highly stable, with good interference tolerance and is suitable for classifying and regression problems. Many successful applications of SVM have been published in many fields such as image recognition, text classification and handwriting recognition. Usually, in a particular math problem, the number of classes will be greater than 2, and to solve the problem of multilayered class (the number of classes $c \ge 3$), SVM is usually expanded by 2 simple methods: one-vs-one (ovo) and one-vs-rest (ovr).

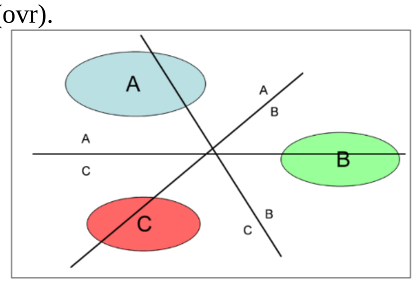


Figure II.6 One-vs-one method of multiclass SVM

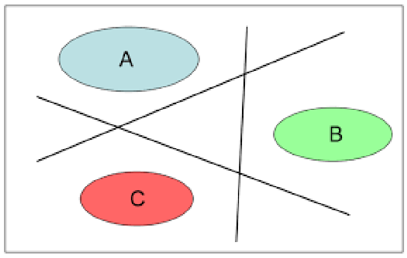


Figure II.7 One-vs-rest method of multiclass SVM

The one-vs-rest method builds c binary SVM model.
The one-vs-one method builds c*(c-1)/2 binary SVM models, each separating two classes of each other.

III. RESULTS

A. Program

We created the program to identify faces in images obtained from still images / cameras using Haar-like features and AdaBoost classifiers in the Python programming language, OpenCV library, Dlib. Extraction of the location of the face landmark is installed in the Python programming language. Use the LibSVM library to train facial emotion recognition model. Comparing the proposed method (SVM-Dlib) with CNN convolutional neural network models, SVM-based model (SVM-HOG) and k nearest neighbor model using HOG features. The training and recognizing of the CNN convolutional neural network model is installed by the Python programming language, using the TensorFlow library. The training and recognizing of SVM-HOG and kNN also uses the Python programming language, OpenCV library.

The practices were conducted on personal computers, Windows 10 operating system, Intel Core i5-4300U 1.9GHz processor, quad-core and 4GB RAM.

B. Preparing the dataset

The experiment is excuted on Cohn-Kanade data set consists 4234 images of 65 people. Start training the model by detecting the position of the face in the images. Facial images are standardized to a minimum square of 64x64 pixels to easily build a human facial emotion recognition model.

The dataset is randomly assigned at a rate of 80/20 to two sets: training set consists 4234 images and testing set consists 1058 images. The training set was then used to train the SVM-Dlib, CNN, SVM-HOG and kNN recognition models. Then, use the obtained models to identify facial emotions in the test set or any static images or images directly from the video of the computer webcam.

C. Adjusting parameters

The SVM-Dlib algorithm uses the LibSVM library to build the recognition model. We performed different experiments by adjusting kernel functions such as *rbf, poly*, parameters γ, constant *c* to find optimal results. The optimal set of parameters is $\gamma = 0.01$, *kernel = poly*, $c = 1000$.

The CNN model used for facial emotion recognition training is also adjusted to optimize the results. CNN architecture is presented as shown in Figure II.8. The optimal set of parameters is η*=0.01, batch_size=64, epochs=90.*

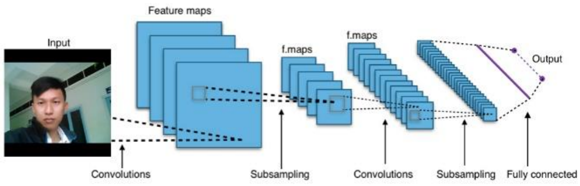


Figure III.1 Kiến trúc mạng nơ-ron tích chập

The SVM method uses HOG features and the rbf kernel function, the optimal set of parameters found *γ=0.2, c=1*.

The kNN methods uses HOG performs emotional recognition by finding the closest distance based on feature vectors from the original image to the emotion class.

D. Results

The propose model (SVM-Dlib) got 94.33% of accuracy in testing set, the highest value in four methods; the second place is Convolutional Neural Network (CNN) model with the accuracy of 94.30%; the two rest methods have lower accuracy, the third place is SVM method using HOG features with 45.01% of accuracy and the last is k nearest neighbor method using HOG features with the accuracy of 36.43%. The result is shown in table III.1, and figure III.2.

Bảng III.1 Comparing the effective of facial emotion recognition methods

ID	Methods	Accuracy (%)
1	SVM-Dlib	94.33
2	CNN	94.30
3	SVM-HOG	45.01
4	kNN	36.43

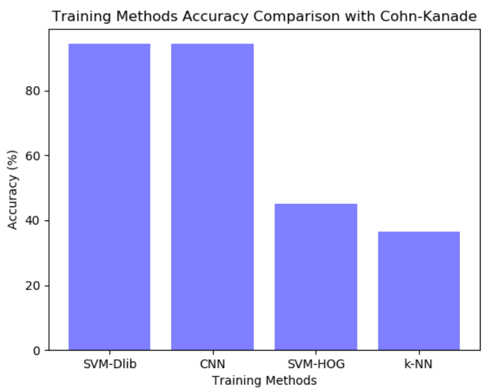


Figure III.2 Comparing the accuracy

IV. CONCLUSION AND DEVELOPMENT DIRECITON

The system recognizes human emotion bases on human face with Support Vector Machine and facial emotion recognition. Face locating in image from static image or camera is excuted by Haar-like feature and AdaBoost classifier, it meets the criteria of speed and accuracy in face detecting and locating. The system recognizes facial emotion with Support Vector Machine and facial landmarks. The practice result with 4234 images of 65 objects in Cohn-Kanade dataset shows that our proposed model - training with SVM machine learning and Dlib library got 94.33% accuracy on the testing set, which is higher in comparation on training with Convolutional Neutral Network – CNN, SVM machine learning using HOG features (Histogram of Oriented Gradients), k nearest neighbor using HOG features (k-Nearest Neighbor - kNN) with the accuracy are 94.30%, 45.01% and 36.43% respectively.

We intend to provide more experimental results on the other dataset and campare the effective of the propose model with other methods, improve recognition model in order to recognition more complex emotion.

V. REFERENCES

[1] Khanh Doan, "Soc Trang Union of Science and Engineering," 2016. [Online]. Available: http://lienhiephoi.soctrang.gov.vn/index.php/khoa-h-c-va-cong-ngh/tin-khoa-h-c-va-cong-ngh/765-cong-ngh-phan-tich-c-m-xuc-ng-d-ng-ti-m-nang.

[2] Ajit P. Gosavi and S. R. Khot, "Facial Emotion Recognition Using Principal Component Analysis," *IJSCE*, vol. 3, pp. 258-262, September 2013.

[3] Monika Dubey and Prof. Lokesh Singh, "Automatic Emotion Recognition Using Facial Expression: A Review," *IRJET*, vol. 3, no. 2, 2016.

[4] James Pao, "Emotion Detection Through Facial Feature Extraction".

[5] V. Vapnik, "The Nature of Statistical Learning Theory," *Springer-Verlag*, 1995.

[6] Vahid Kazemi and Josephine Sullivan, "One millisecond face alignment with an ensemble of regression trees," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1867-1874, 2014.

[7] T. N. Do, N. K. Pham and M. T. Tran Nguyen, "Human Face Recognition with GIST features and Support Vector Machine," *FAIR*, 2017.

[8] Xindong Wu and Vipin Kumar, The Top Ten Algorithms in Data Mining, Chapman and Hall/CRC, 2009.

[9] N Dharmesh and K. Mausmi , "Albert," *Emotion Detection: A Feature Analysis*, vol. 4, no. 7, 2015.

[10] Nishant Hariprasad Pandey, Dr. Preeti Bajaj and Shubhangi Giripunje, "Design and Implementation of Real Time Facial Emotion Recognition System," *International Journal of Science and Engineering Research*, vol. 4, no. 11, 2013.

[11] M. Albert, "Albert Mehrabian Communication Studies," *Institude of Judicial Studies.*

[12] Binh T. Nguyen, Minh H. Trinh, Tan V. Phan and Hien D. Nguyen, "An Efficient Real Time Emotion Detection Using Camera and Facial Landmarks," 2017.

[13] Dinh Viet Sang, Le Tran Bao Cuong and Vu Van Thieu, "Multi-task learning for smile detection, emotion recognition and gender classification," *SoICT2017*, p. 5, 2017.

[14] Nguyen Huu Tuan, "Building Automatic Face Recognition System Using LPQ (Local Phase Quantization)," 2016.

[15] Pham Nguyen Khang and Do Thanh Nghi, The Machine Learning Theory, Can Tho: Can Tho University Publisher, 2012.