International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/IJSRCSEIT

Real-time Algorithms for Facial Emotion Recognition : A Comparison of Different Approaches

Gitanjali Bhujbal, Sachin Patil, Paurnima Kawale
Department of Computer Engineering, Zeal College of Engineering and Research, Pune, Maharashtra, India

Article Info
Volume 8, Issue 2
Page Number : 338-342

Publication Issue :
March-April-2022

Article History
Accepted: 10 March 2022
Published: 22 March 2022

ABSTRACT

Emotion recognition has application in various fields such as medicine (rehabilitation, therapy, counselling , etc.), e-learning, entertainment, emotion monitoring, marketing, law. Different algorithms for emotion recognition include feature extraction and classification based on physiological signals, facial expressions, body movements. In this paper, we present a comparison of five different approaches for real-time emotion recognition of four basic emotions (happiness, sadness, anger and fear) from facial images.

Keywords: Convolutional Neural Network, Emotion Recognition, Facial Expression, Multilayer Perceptron, Support Vector Machine.

I. INTRODUCTION

In recent years, there has been a growing interest for development of accurate and reliable computer algorithms for emotion recognition based on facial features acquired from camera. Facial expression is one of the most important features of human emotion recognition [1]. Nowadays, automated facial expression recognition has a large variety of applications, such as data- driven.

Facial features, such as eyes, brows, nose, mouth and chin, can be labeled in a face image and create facial feature points. These facial features can be detected in an image through the process of fitting a predefined set of facial feature points into a face image which is called Facial Feature Point Detection (FFPD).

Facial expression recognition systems can work with static images [5-7] or with dynamic image sequences [8. In static-based methods, a feature vector comprises information about the current input image only. Sequence based methods use temporal information of images to recognize the expression captured from one or more frames. Automated systems for facial expression recognition receive the expected input (static image or image sequence) and typically give as output one of the basic expressions (anger, sadness, happiness and fear), while some systems also recognize the neutral expression, surprise and disgust.

II. METHOD

A. Conventional approaches

Emotion recognition algorithms based on conventional approaches include: 1) facial landmark detection (eyes, brows, nose, mouth and chin) and face extraction, 2) feature extraction and classification.

Facial landmark detection and face extraction

Facial landmark extraction was performed on monochromatic (gray-scale) images using open source OpenFace toolkit [19] for Matlab (Mathworks, USA). Facial landmark detection was performed using the generic algorithm in OpenFace , introduced by Yu et al. [20]. This algorithm is based on Constrained Local Neural Field (CLNF) and Constrained Local Model (CLM) models. To remove non-facial information from the image, a binary mask was created by using a convex hull that surrounds facial landmarks (Fig. 1b) and then applied to extract the face (Fig. 1c,d).

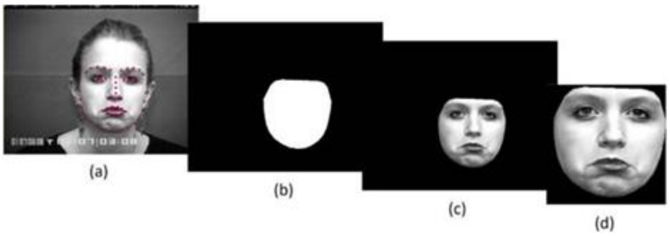

Fig. 1. Face extraction procedure: (a) facial landmark detection by OpenFace ; (b) binary mask creation; (c) removing non-facial information by masking; (d) extracted face.

III. RESULTS

A. Experiment description

Real-time testing was performed on 8 volunteers (3 male and 5 female) of age 22.75±4.62. All subjects have signed a written consent for participation. The testing was done during daylight, in a room with additional lighting. Camera Nikon D5100 was used, while the subjects' distance from the camera was 50 cm. In order to compare results of different approaches, it would be necessary to synchronously test all proposed algorithms in real-time with the same input data. As it was too complex, we acquired a video of volunteers and used recorded videos as inputs to all algorithms in the same way that real-time input data would be used. Videos of subjects' faces were recorded at a frame rate of 24 frames per second (fps) for 160 seconds, during which they had to express four emotions: happiness, sadness, anger and fear, cyclic, five times each. The specified frame rate was chosen as it was previously determined that all algorithms work without delay in real-time at this frame rate (offline frame rate was greater than or equal to 24 fps).

TABLE 1-CONFUSION MATRIX FOR SVM CLASSIFICATION :

Predicted output					
Desired output		happiness	sadness	anger	fear
	happiness	75.86%	3.45%	20.69%	0%
	sadness	3.03%	48.48%	48.48%	0%
	anger	3.23%	0%	96.77%	0%

TABLE II: CONFUSION MATRIX FOR MLP CLASSIFICATION

				Predicted output	
Desired output		happiness	sadness	anger	fear
	happiness	89.66%	10.34%	0%	0%
	sadness	6.06%	81.82%	12.12%	0%
	anger	6.45%	67.74%	22.58%	3.23%
	fear	35.71%	35.71%	28.57%	0%

In Table 3 confusion matrix for AlexNet CNN is shown, Table 4 shows confusion matrix for FER-CNN approach and Table 5 present results of commercial Affdex CNN solution.

TABLE III: CONFUSION MATRIX FOR ALEXNET CNN

				Predicted output	
Desired output		happiness	sadness	anger	fear
	happiness	86.21%	0.00%	13.79%	0%
	sadness	0%	69.70%	30.30%	0%
	anger	0%	3.23%	96.77%	0%
	fear	21.43%	21.43%	28.57%	28.57%

TABLE IV: CONFUSION MATRIX FOR FER-CNN

		Predicted output			
		happiness	sadness	anger	fear
Desired output	happiness	62.07%	31.03%	6.90%	0%
	sadness	0%	81.82%	18.18%	0%
	anger	0%	61.29%	32.26%	6.45%
	fear	0%	50.00%	21.43%	28.57%

TABLE V: CONFUSION MATRIX FOR AFFDEX

		Predicted output			
		happiness	sadness	anger	fear
Desired output	happiness	96.55%	3.45%	0%	0%
	sadness	3.03%	84.85%	9.09%	3.03%
	anger	0%	29.03%	70.97%	0%
	fear	0%	7.14%	0%	92.86%

Overall accuracies of all tested algorithms are shown in Table 6. Affdex CNN performs with the highest accuracy of 85.05%, followed by AlexNet, with accuracy of 76.64%.

TABLE VI: TOTAL ACCURACIES OF ALL TESTED ALGORITHMS

Total Facial emotion recognition algorithm	**accuracy [%]**
Affdex CNN	85.5
Fine-tuned AlexNet CNN	76.64
SVM classification of HOG features	63.55
MLP classification of HOG features	56.07

IV. CONCLUSION

We have presented a pilot study for real-time testing of conventional and deep learning approaches in facial emotion recognition. Preliminary results show better generalization power and better performance in real-time application of fine-tuned AlexNet CNN and Affdex CNN than SVM and MLP approaches. Commercial Affdex CNN has overall superior accuracy, but AlexNet and SVM had better “anger” recognition (96.77% vs. 70.97%). FER- CNN had the lowest overall accuracy but high accuracy for “sadness”, comparable with Affdex CNN result (81.82% vs. 84.85%). In further research we will test this fact in a larger group of volunteers and for more than four emotions.

V. REFERENCES

[1]. Y. Wu, H. Liu and H. Zha, “Modeling facial expression space for recognition,” in IEEE/RSJ International Conference on Intelligent Robots and Systems, (IROS 2005), 2005, pp. 1968-1973.

[2]. A. T. Lopes, E. de Aguiar, A. F. De Souza and T. Oliveira-Santos, “Facial expression recognition with convolutional neural networks: coping with few data and the training sample order,” Pattern Recognition, vol. 61, pp. 610-628, 2017.

[3]. J. H. Yu, K. E. Ko and K. B. Sim, “Facial point classifier using convolution neural network and cascade facial point detector,” Journal of Institute of Control, Robotics and Systems, vol. 22, no. 3, pp. 241-246, 2016.

[4]. M. Dantone, J. Gall, G. Fanelli, and L. Van Gool, “Real-time facial feature detection using conditional regression forests,” in Computer vision and pattern recognition (CVPR), 2012, pp. 2578-2585.

[5]. C. Shan, S. Gong and P. W. McOwan, “Facial expression recognition based on local binary patterns: A comprehensive study,” Image and Vision Computing, vol. 27, no. 6, pp. 803-816, 2009.

[6]. M. Liu, S. Li, S. Shan and X. Chen, “Au-inspired deep networks for facial expression feature learning,” Neurocomputing, vol. 159, pp. 126-136, 2015.

[7]. G. Ali, M. A. Iqbal and T. S. Choi, “Boosted NNE collections for multicultural facial expression

recognition," Pattern Recognition, vol. 55, pp. 14-27, 2016.

[8]. Y. H. Byeon and K. C. Kwak, "Facial expression recognition using 3d convolutional neural network," International journal of advanced computer science and applications, vol. 5, no. 12, pp. 107-112, 2014.

[9]. J. J. J. Lien, T. Kanade, J. Cohn and C. Li, "Detection, tracking, and classification of action units in facial expression," Journal of Robotics and Autonomous Systems, vol. 31, no. 3, pp. 131-146, 2000.

[10]. X. Fan and T. Tjahjadi, "A spatial-temporal framework based on histogram of gradients and optical flow for facial expression