Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7750378

A COMPREHENSIVE STUDY ON DETECTING AND RECOGNIZING EMOTIONS BY FACIAL EXPRESSIONS USING MACHINE LEARNING TECHNIQUES

[1]Dr. Balamurugan M [2]Dr. Gopinath D [3]Dr. Naveeth Babu.C [4]Dr. V. S. Prakash
[1, 2, 3, 4] Assistant Professors, Department of Computer Science (UG),
Kristu Jayanti College (Autonomous), Bengaluru, Karnataka, India.
tecbalamurugan@gmail.com, kascgopinath@gmail.com

Abstract

Facial expression is non-verbal communication which implies appearance on the face, arm movements or voice intonation conveys feelings about something without utilizing words. There are numerous uses for facial expressions. Recognition of facial expressions draws more attention to a variety of disciplines, including Computer science, Biotechnology, Psychology, Chemical and Pharmaceutical science. The facial expressions used in Human Computer Interaction (HCI) research to improve outcomes. Accurate emotional feature extraction is made possible via facial expression recognition. Facial expressions recognition approaches in static images do not fully consider the features of facial organs and muscle movements, which are static and dynamic, as well as the geometric and appearance qualities of facial expressions. This limitation is solved by using patch-based 3D Gabor feature extraction, selecting key patches, and key distance features obtained by carrying out patch matching operations. Test results produce promising results under Correct Recognition Rate (CRR), significant performance improvements in consideration of facial features and muscle movements, reduced face registration errors, and faster processing time. According to the difference in state-of-the-art performance, the proposed approach gave the highest CRR on the JAFFE and Cohn-Kanade AU-Coded Facial Expression database which is conclusive.

Keywords: Automated facial expression recognition system, face detection, emotion detection, and human-computer interaction.

1. INTRODUCTION

Facial expression analysis has been attracting considerable attention in the advancement of human machine interface since it provides a natural and efficient way to communicate between humans. Some application areas related to face and its expressions include personal identification and access control, video phone and teleconferencing, forensic applications, human-computer interaction, automated surveillance, cosmetology, and so on. But the performance of the face detection certainly affects the performance of all the applications. Many methods have been proposed to detect human face in images, they can be classified into four categories: knowledge-based methods, feature-based methods, template based methods and appearance-based methods. When used separately, these methods cannot solve all the problems of face detection like pose, expression, orientation, and occlusion. Hence it is better to operate with several successive or parallel methods. Most of the facial expression recognition methods reported to date are focused on recognition of six primary expression categories such as: happiness, sadness, fear, anger, disgust and grief.

Contrast Stretching

Contrast stretching (often called normalization) is a simple image enhancement technique that attempts to improve the contrast in an image by 'stretching' the range of intensity values it contains to span a desired range of values, e.g. the full range of pixel values that the image type concerned allows. It differs from the more sophisticated histogram equalization in that it can only apply a linear scaling function to the image pixel values. As a result the 'enhancement' is less harsh. (Most implementations accept a gray level image as input and produce another gray level image as output Fig 1.1)

Fig. 1.1: Contrast Stretching

Binary Conversion of the Image

A binary image is a digital image that has only two possible values for each pixel. Typically the two colors used for a binary image are black and white though any two colors can be used. The colour used for the object(s) in the image is the foreground colour while the rest of the image is the background colour.

Fig. 1.2: Conversion to Binary

An ideal emotion detection system should recognize expressions regardless of gender, age, and any ethnicity. Such a system should also be invariant to different distraction like glasses, different hair styles, mustache, facial hairs and different lightening conditions. It should also be able to construct a whole face if there are some missing parts of the face due to these distractions.

2. FACE DETECTION AND TRACKING

Firstly, the classifier trained for face detection searches for a face in the image. Some image processing techniques on it in order to find the face region. System can operate on static images, where this procedure is called face localization or videos where we are dealing with face tracking.

If the face is located, the classifiers for eye detections are employed only on the upper part of the face. The left and right eyes are detected separately – in left and right upper face regions. Finally, the mouth region is located with the fourth classifier which searches in the lower part of the face.

Fig. 1.3: Face detection and Face elements localization

3. RECOGNITION PERFORMANCE

JAFFE Database

CRR Correct recognition rate of all sets in 10 leave-onset-out cross-validations. Results obtained using four SVMs and four distances.

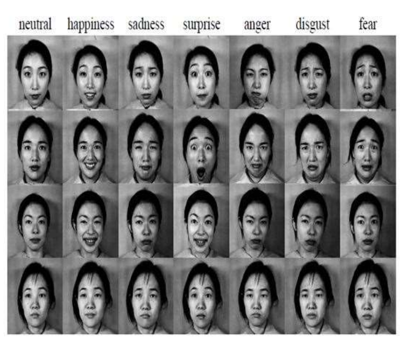

Fig. 1.4: Face Expressions Recognition Figures.

In Fig. 1.4 we can see that the proposed approach performs the best with a CRR of 92.93% using DL2 and linear SVM. Regarding the performance of distances, DL2 achieves higher CRRs than the other three distances for all SVMs. When L1 is used, sparse distances outperform dense distances for linear, RBF and sigmoid SVMs. On the contrary, when L2 is used, dense distances outperform spare distances for all SVMs (note that the CRR of DL2 and sigmoid SVM is not shown). For both sparse and dense distances, L2 performs better than L1 for all SVMs. Among four SVMs, linear and RBF outperform polynomial and sigmoid for all distances. More exactly, the best performance is obtained by linear, which is followed by RBF, whereas sigmoid ranks the least.

%	neutral	happiness	sadness	surprise	anger	fear	disgust
neutral	60	0	20	0	0	20	0
happiness	10	90	0	0	0	0	0
sadness	14	0	86	0	0	0	0
surprise	0	0	0	100	0	0	0
anger	0	20	20	0	60	0	0
fear	11	11	0	11	11	56	0
disgust	0	0	0	0	1	0	100

Table 1: Confusion matrix of testing with Cohn-Kanade Database

%	neutral	happiness	sadness	surprise	anger	fear	disgust
neutral	61	0	33	0	0	6	0
happiness	3	78	3	10	0	3	3
sadness	3	1	91	0	4	1	0
surprise	3	3	3	76	0	12	3
anger	0	2	36	0	51	4	7
fear	2	2	28	4	0	64	0
disgust	0	2	20	0	7	7	64

Table 2: Confusion matrix of testing with FG-NET FEED

Table 1 and 2 demonstrates the confusion matrix of six emotions using DL2 and linear SVM. Observed from this table, disgust and surprise belong to the most difficult facial expressions to be correctly recognized with the same CRR of 90.00%, whereas anger is the easiest one with a CRR of 96.67%. Regarding the misrecognition rate, anger contributes the most; as a result, it has a major negative impact on the overall performance. The emotion that follows in misrecognition rate is fear.

MFCC TRACE for happiness and anger utterances

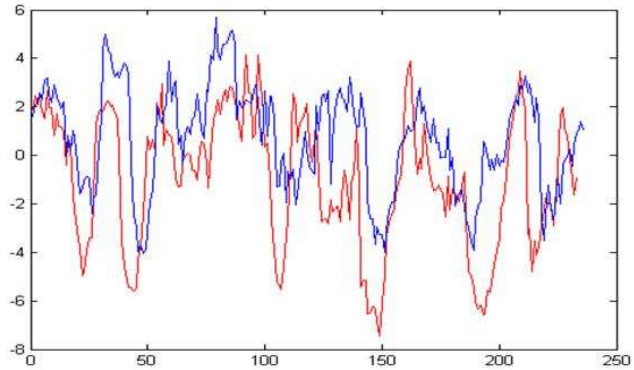

Fig. 1.5: MFCC TRACE for happiness and anger utterances

The fig 1.5 shows the following content. In addition 7 statistical features were computed for every utterance, obtaining a 27 length feature vector. Given a digital image, or a region within an image, the feature extraction task implies the taking out of a quantity of information capable to characterize the original image. Working with two-dimensional signals, the number of samples is much bigger than in the case of a one dimensional signal, thus the necessity of information quantity reduction is obvious.

CK Database

The CRRs using four SVMs and four distance metrics are shown in which the proposed approach obtains the highest CRR of 94.48% using DL2 and RBF SVM. Regarding the performance of distances, DL2 keeps the highest CRRs for all SVMs (note that the CRR of DL2 and sigmoid SVM is not shown). Moreover, dense distances have a higher overall performance than sparse distances.

This reflects that emotional information in the CK images is distributed over all orientations rather than the dominant orientation of Gabor features. As for SVMs, RBF performs the best for dense distances, while linear performs the best for sparse distances. This confirms with the results in that RBF and linear perform better than polynomial on the CK database. The confusion matrix of six emotions using DL2 and RBF SVM. As can be seen, surprise performs the best with a CRR of 100%, the following one is happy with a CRR of 98.07%. On the other hand, anger is the most difficult facial expression to be correctly recognized with a CRR of only 87.10%. The performance of surprise and anger on CK contrasts with that on JAFFE, in which surprise and anger are the most difficult and easiest emotions respectively. The reason probably is that surprise images on CK are often characterized as an exaggerated “open mouth”, while those on JAFFE are normally with a “close or slightly open mouth”. This can be seen from that the selected patches for CK focus on the mouth region, but those for JAFFE are mainly distributed around the eyes regions. Similarly, anger images on JAFFE are better expressed by the selected patches in mouth region than the selected patches are all over the face region those on CK. Among six emotions, anger and sad contribute most to the misrecognition rate.

4. CONCLUSION

In this paper, explores the issue of facial expression recognition using facial movement features. The effectiveness of the proposed approach is testified by the recognition performance, computational time, and comparison with the state-of-the-art performance. The experimental results also demonstrate significant performance improvements due to the consideration of facial movement features, and promising performance under face registration errors. The results indicate that patch-based Gabor features show a better performance over point-based Gabor features in terms of extracting regional features, keeping the position information, achieving a better recognition performance, and requiring a less number. Different emotions have different 'salient' areas; however, the majority of these areas are distributed around mouth and eyes. In addition, these 'salient' areas for each emotion seem to be not influenced by the choice of using point-based or using patch-based features. The 'salient' patches are distributed across all scales with an emphasis on the higher scales. For both the JAFFE and CK databases, DL2 performs the best among four distances. As for emotion, anger contributes most to the misrecognition. The JAFFE database requires larger sizes of patches than the CK database to keep useful information.

REFERENCES:

1. L. Sirovich and M. Kirby, "Low Dimensional Procedure for Characterization of Human Faces," J. Optical Soc. Am., vol. 4, pp. 519-524, 2017.
2. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 711–720, Jul. 2019.
3. D. Arumugam, S. Purushothaman, "Emotion Classification Using Facial Expression", International Journal of advanced Computer Science and Applications, Vol. 2, No. 7, pp. 92-98, 2020.
4. A.K. Jain, R.P.W. Duin, J Mao, "Statistical Pattern Recognition: A Review", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 22, No. 1, pp. 4-37, 2020.
5. K.schindler and L. Van Gool, "Action snippets: How many frames does human action recognition require?" in computer vision and pattern Recognition, 2018. CVPR2008.IEEEConference on, 2018, pp.18.
6. A.Hadid and M.Pietikainen, "Combining appearance and motion for face and gender recognition from videos", Pattern Recognition, vol.42, pp.2818-2827, 2019.
7. Anil Audumbar Pise, Mejdal A. Alqahtani, Priti Verma, Purushothama K, Dimitrios A. Karras, Prathibha S, Awal Halifa, "Methods for Facial Expression Recognition with Applications in Challenging Situations", Computational Intelligence and Neuroscience, vol. 2022, Article ID 9261438, 17 pages, 2022.
8. Qinying Yuan, "A Classroom Emotion Recognition Model Based on a Convolutional Neural Network Speech Emotion Algorithm", Occupational Therapy International, vol. 2022, Article ID 9563877, 12 pages, 2022.
9. Yuan Li, "Application of an Improved LSTM Model to Emotion Recognition", Journal of Electrical and Computer Engineering, vol. 2022, Article ID 3271074, 11 pages, 2022.

10. Abeer Ali Alnuaim, Mohammed Zakariah, Prashant Kumar Shukla, Aseel Alhadlaq, Wesam Atef Hatamleh, Hussam Tarazi, R. Sureshbabu, Rajnish Ratna, "Human-Computer Interaction for Recognizing Speech Emotions Using Multilayer Perceptron Classifier", Journal of Healthcare Engineering, vol. 2022, Article ID 6005446, 12 pages, 2022.
11. Zhenrong Deng, Hongquan Lin, Wenming Huang, Rushi Lan, Xiaonan Luo, "Emotional Dialogue Generation Based on Conditional Variational Autoencoder and Dual Emotion Framework", Wireless Communications and Mobile Computing, vol. 2020.
12. Hanzhong Zhang, Jibin Yin, Xiangliang Zhang, "The Study of a Five-Dimensional Emotional Model for Facial Emotion Recognition", Mobile Information Systems, vol. 2020, Article ID 8860608, 10 pages, 2020.
13. Ben, X., Ren, Y., Zhang, J., Wang, S. J., Kpalma, K., Meng, W., et al. (2021). Video-based facial micro-expression analysis: a survey of datasets, features and algorithms. IEEE Trans. Pattern Anal. Mach. Intell. doi: 10.1109/TPAMI.2021.
14. Budur, T., Demir, A., and Cura, F. (2021). University readiness to online education during Covid-19 pandemic. Int. J. Soc. Sci. Educ. Stud. 8, 180–200.
15. Cashman, D., Perer, A., Chang, R., and Strobelt, H. (2019). Ablate, variate, and contemplate: visual analytics for discovering neural architectures. IEEE Trans.Vis. Comput. Graph. 26, 863–873.