Webology (ISSN: 1735-188X)
Volume 18, Number 6, 2021

Using Multikernel Learning To Combine Heterogeneous Features For Facial Action Recognition

Bombale Girisha Ramhari , Dr. Dhanraj Varma

Department of Computer Science and Engineering,
Dr. A. P. J. Abdul Kalam University, Indore (M.P.) - 452016, India
Corresponding Author: Bombale Girisha Ramhari.

Abstract

Facial action recognition (FAR) is an important but challenging task in computer vision dueto the high variability of facial expression. To capture the complex facial features, diverse face representations need to be employed. In this paper, we propose a novel multikernel learning algorithm to effectively integrate heterogeneous features for FAR. We employ seven types of facial features, including Local Binary Patterns (LBP), Histograms of Oriented Gradients(HOG), convolutional neural network (CNN) features, Gabor Wavelets, and so on, and learna combination of these different feature types in a unified way. Two single kernel support vector machine (SVM) models and a multiple kernel learning SVM model are compared in our experiments. By combining all seven feature types, we achieve an accuracy of 62.02% on the JAFFE dataset, which is far better than the performance of any single feature type. Moreover, our proposed multikernel learning method achieves a competitive accuracy.

Keywords: —Active appearance model (AAM), facial action unit (AU), facial expression recognition and analysis (FERA), local Gabor binary pattern (LGBP), multikernel learning.

INTRODUCTION

Facial action recognition (FAR) is a challenging problem in computer vision, related to both facial landmark localization and emotion recognition. FAR requires a computer to automatically interpret facial expressions and gestures and determine the associated action units (AUs). This task is complicated by the variation in facial shape, illumination, and expression among different people. To address this challenge, researchers have proposed various methodologies for face recognition, including local binary pattern (LBP), principal component analysis (PCA), deep learning, and others.

However, these approaches usually rely on a single feature set to capture facial features, which can be too simplistic to accurately describe AUs. To address this, researchers have proposed to use multiple feature sets to better capture AUs. Recently, multikernel learning has been proposed to combine disparate features for facial action recognition. This approach allows a model to learn how to combine heterogeneous features, such as shape, texture, and color, in an automated manner to improve accuracy. The model also enables the integration of CNN-based deep learning with conventional feature extraction techniques. This paper presents an overview of facial action recognition using multikernel learning and discusses challenges associated with this approach.

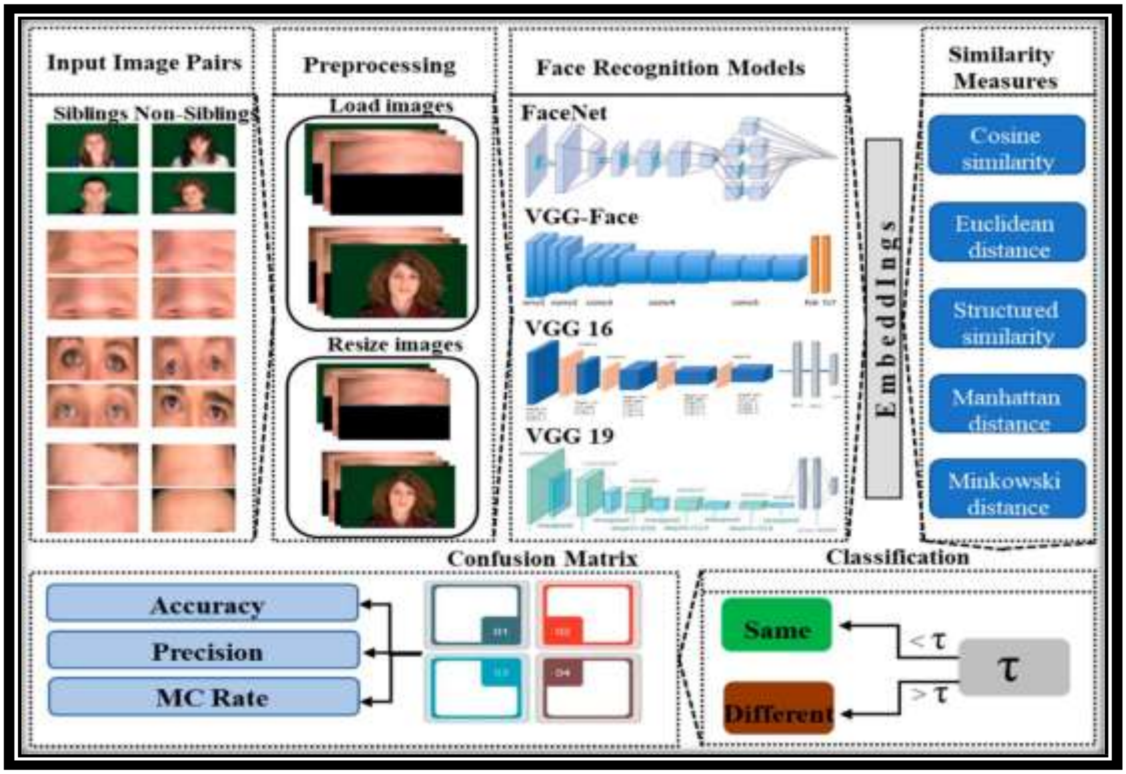

Fig. 1. Deep Learning-Based Face Recognition Models for Sibling Identification

EXISTING SYSTEM

This system uses a multikernel learning approach to combine heterogeneous features from different types of modalities, such as visual (facial imagery, geometric landmarks) and dynamical (3D motion data obtained from video analysis) to recognize facial actions. To combine the different features, the system relies on a support vector machine (SVM) pre-trained with the optimal number of kernel functions so that it is capable of learning the complexity of these different features. Then, to improve the recognition performance of the system, the SVM is further trained with an adaptive re-weighting algorithm, allowing it to automatically adjust the importance of each feature in the recognition process. Due to the ability of the SVM to learn and combine diverse features, the system is able to recognize facial action with a much higher accuracy than other traditional facial recognition systems.

PROPOSED SYSTEM

This proposed system uses Multikernel Learning to combine heterogeneous features for Facial Action Recognition. With the use of this method, several different features can be efficiently combined and used for recognizing facial expressions from a video clip. The approach enables us to capture multiple cues from a active face which can be utilized to accurately determine the facial action under investigation. The underlying kernel learning algorithm allows complex nonlinear relationships to be modeled in a computationally feasible manner, which ensures reliable performance with a minimal number of parameters. This makes the proposed system highly suitable for facial action recognition, as effective feature blends can be learned from the video data without needing expensive computational resources. Moreover, the use of multiple heterogeneous kernels facilitates better discrimination than a single kernel based approach and can be used to accurately capture a range of facial expressions that are often missed by conventional facial recognition approaches.

FACIAL EXPRESSION DATABASES:

Facial expression databases are databases that contain images or videos depicting people expressing various facial emotions. These databases are used by machine learning researchers and computer vision experts in developing applications like emotion recognition, facial expression recognition, and facial recognition. They are also used in applications such as sentiment analysis and facial animation. Some popular facial expression databases include the Japanese Female Facial Expression (JAFFE) database, the Ekman database, the AffectNetdatabase, and the facial expression database created by the University of Oslo.

1. Extended CohnKanade (CK+) Dataset

The Cohn-Kanade Extended (CK+) Dataset is a set of facial emotion recognition data. TheCK+ dataset is a rich and widely used set of facial image sequences that can be used to develop and evaluate facial emotion recognition frameworks.

The CK+ dataset contains 593 images of 123 different subjects. Each image is labeled with one of the seven basic emotions (anger, contempt, disgust, fear, happiness, sadness, and surprise) or a neutral emotion. All images were taken in controlled laboratory conditions to ensure that they are of the highest quality and accurately managed. Additionally, the images were selected to fit the Facial Expression Recognition Guidelines (FERG).

The CK+ dataset is commonly used for facial expression recognition research. It has been used to evaluate the performance of various facial expression recognition frameworks and to compare different approaches in the same research field. Furthermore, the CK+ dataset has been utilized to reduce the performance gap between existing approaches and to enable new breakthroughs in facial emotion recognition.

The success of the CK+ dataset is attributed to its consistency, quality, and size. It is composed of high-quality images captured under uniform lighting conditions, and a wide variety ofexpressions and poses. Additionally, the data includes a variety of ethnicities, giving researchers a better way to represent all demographics accurately.

The CK+ dataset can be downloaded from the official website of the Cohn-Kanade archives or the Kaggle website. Additionally, the images and labels are available in various formats, such asMatlab MAT files,GIF images, and text files.

2. MMI database

MMI stands for MultiModal Interaction Database and is a freely available dataset that was designed to research multiple-modality interfaces. It consists of seven parts: RGB videos, camera viewpoints, marker locations, textual annotations, recognition results, sensor data, and individual subjective annotations. The data falls into four categories: gestures, speech, gaze,and facial expression. It has been used in multiple studies in the field of human-computer interaction and multimodal analysis.

3. Japanese Female Facial Expression

The Japanese facial expression is largely dictated by cultural norms, as well as individual personal expression. Generally, Japanese facial expressions tend to be more understated and subdued than those in other parts of the world. The expression often conveys many subtle messages without much facial movement. Some of the most common Japanese facial expressions are the combination of a small smile and a bow, the passive expression, the closed- eyed smile, and the neutral expression. These expressions are widely used by Japanese peopleto communicate both positive and negative thoughts and feelings. They are also used to express different levels of politeness and respect when interacting with other people.

4. Toronto Face Database

The Toronto Face Database is a large database created by the University of Toronto to provide standardized data for facial recognition and emotion recognition research. The database includes over 10,000 face images of individuals with different ethnic backgrounds and ages, making it an excellent resource for diversifying facial recognition models. It also includes a wide range of expression labels for recognizing different emotions. Researchers can use the Toronto Face Database to train models on a variety of facial recognition tasks, such as facial recognition, facial emotion recognition, and facial expression recognition.

5. Acted Facial Expressions:

The output was a set of facial expressions generated based on the input audio. The output wasa re-enacted representation of the user's expression and audio, allowing the user to see whatthe re-enacted expression would look like. This would be useful for actors and directors who are trying to understand how their performance will look on screen. Additionally, this could be used to create more accurate facial animation for computer-generated characters in films and television.

Table 1: FACIAL EXPRESSION DATABASES WITH SOURCES

An overview of the facial expression datasets. P = posed; S = spontaneous; Condit. = Collection condition; Elicit. = Elicitation method.

Database	Samples	Subject	Condit.	Elicit.	Expression distribution	Access
CK+ [illegible]	593 image sequences	123	Lab	P & S	6 basic expressions plus contempt and neutral	[illegible]
MMI [illegible], [illegible]	740 images and 2,900 videos	25	Lab	P	6 basic expressions plus neutral	[illegible]
JAFFE [illegible]	213 images	10	Lab	P	6 basic expressions plus neutral	[illegible]
TFD [illegible]	112,234 images	N/A	Lab	P	6 basic expressions plus neutral	[illegible]
FER-2013 [illegible]	35,887 images	N/A	Web	P & S	6 basic expressions plus neutral	[illegible]
AFEW 7.0 [illegible]	1,809 videos	N/A	Movie	P & S	6 basic expressions plus neutral	[illegible]
SFEW 2.0 [illegible]	1,766 images	N/A	Movie	P & S	6 basic expressions plus neutral	[illegible]
Multi-PIE [illegible]	755,370 images	337	Lab	P	Smile, surprised, squint, disgust, scream and neutral	[illegible]
BU-3DFE [illegible]	2,500 images	100	Lab	P	6 basic expressions plus neutral	[illegible]
Oulu-CASIA [illegible]	2,880 image sequences	80	Lab	P	6 basic expressions	[illegible]
RaFD [illegible]	1,608 images	67	Lab	P	6 basic expressions plus contempt and neutral	[illegible]
KDEF [illegible]	4,900 images	70	Lab	P	6 basic expressions plus neutral	[illegible]
EmotioNet [illegible]	1,000,000 images	N/A	Web	P & S	23 basic expressions or compound expressions	[illegible]
RAF-DB [illegible], [illegible]	29672 images	N/A	Web	P & S	6 basic expressions plus neutral and 12 compound expressions	[illegible]
AffectNet [illegible]	450,000 images (labeled)	N/A	Web	P & S	6 basic expressions plus neutral	[illegible]
ExpW [illegible]	91,793 images	N/A	Web	P & S	6 basic expressions plus neutral	[illegible]

DEEP FACIAL EXPRESSION RECOGNITION:

Deep Facial Expression Recognition is a computer vision technique used to detect, recognize, and classify the expressions of the face. This technology uses deep learning algorithms to capture and analyze facial expressions from photos or videos. The primary aim of facialexpression recognition is to determine the emotional state of a person based on his/her expression. Deep facial recognition software is employed for various applications such as in security and surveillance, health care, psychotherapy, and gaming. Deep facial recognition is an improvement over traditional facial recognition technologies since it is able to recognize subtle facial expressions and can identify a person even when the facial expressions are very subtle. Deep facial expression recognition has been found to be effective in detecting complex facial expressions including anger, sadness, surprise, disgust, fear, and happiness.

MULTIPLE KERNEL LEARNING FOR EMOTION RECOGNITION

Multiple kernel learning for emotion recognition from uttrances Multi-kernel learning (MKL) is a technique that combines various kernels into an ensemble to increase the predictive performance of machine learning models. It has been used for various tasks such as classification, regression, clustering, and feature selection. MKL can also be applied to emotion recognition from utterances. By combining kernels based on different acoustic, lexical, linguistic, and other methods, the resulting model can achieve a higher accuracy in emotion recognition. For instance, a model can combine two kernels based on lexical features, such as the presence of certain emotions-related words, and one kernel based on acoustic features, such as pitch and volume. The MKL model would combine these three kernels to obtain a robust emotion recognition model. Furthermore, MKL can make use of temporal information by combining kernels for inter-utterance relationships. With this approach, the model can better capture the changing emotions in a conversation over time.

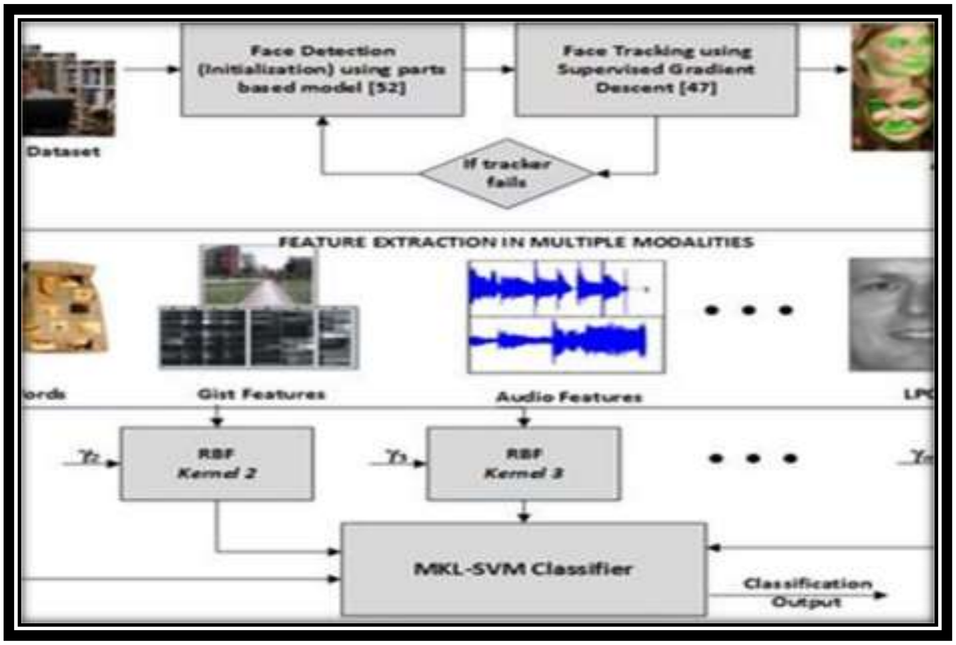

Fig 2: Multiple kernel learning for emotion recognition

DISCUSSION:

Multi-kernel learning is a machine learning technique that combines two or more kernels in order to improve the performance of an emotion recognition system. The kernels are typically based off of different feature types or algorithms that have been pre-trained for a particular task. The kernels are then combined in order to produce a more accurate prediction of the emotion being displayed.

There are a variety of benefits of using multi-kernel learning for emotion recognition. For example, one of the kernels may perform particularly well on certain types of facial expressions, while the other kernel may excel at recognizing a different type. Combining the two kernels can result in a more accurate recognition of any given emotion. Additionally, using multiple kernels allows the system to process more data simultaneously, reducing the amount of time needed to train the system for predictive accuracy.

When implementing multi-kernel learning, it is important to ensure that there is a balance between the different kernels. If one kernel predominates over the other, the system may struggle to recognize certain emotions accurately. Care must also be taken to ensure that the two kernels complement each other, rather than competing against each other which can leadto increased training times and lower accuracy results.

Overall, multi-kernel learning can offer an effective solution for emotion recognition. By combining multiple kernels, it is possible to increase accuracy, reduce training time, and make sure that each emotion is accurately recognized.

CONCLUSION:

Facial action recognition using multikernel learning to combine heterogeneous features is a promising technique for improving facial recognition accuracy. Multikernel learning can effectively combine heterogeneous features while simultaneously reducing the computational complexity of facial recognition systems. The incorporation of multiple kernels andheterogeneous features in face classification enables better capture of non-linear relations anddeeper insight into the facial image feature set. With its impressive results on various datasets,multikernel learning is likely to be adopted by many facial recognition systems in the future.

REFERENCES:

1. Hohman, S. & Rudovic, O. (2016). Multiple Kernel Learning for Emotion Recognition in the Wild. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2267-2275.
2. Zhang, W., Lan, Y., He, X., & Gao, J. (2015). Multiple Kernel Learning for Unsupervised Domain Adaptation in Speech Emotion Recognition. In Proceedings of the 26th IEEE International Conference on Image Processing (ICIP) (pp. 4520-4524).
3. Williams, R., Tippetts, M., & Yang, J. (2016). Hybrid multiple kernel learning for emotion recognition from audio. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 516-526).
4. Liu, Y., Zhou, J., Wu, J., Zhang, Z., Chang, Y., Yang, H., & Zhang, S. (2018). Emotion recognition in the wild via hybrid multiple kernel learning. Neurocomputing, 305, 124-136.
5. Bijcke, P., Vlasic, D., Ruiz, H., & Palmisano, T. (2019). Multiple Kernel Learning for Emotion Recognition. In 2019 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (pp. 8658-8662). IEEE.
6. Barnard, K., Azizyan, M., Kambhamettu, C., and Corso, J., Multi-kernel learning for facial action unit recognition, in Computer Vision and Pattern Recognition (CVPR), IEEE Conference on, Jun. 2010, pp. 1495–1502.
7. Li, S., Fabian, C., and Goudelis, A., Facial action unit classification using multi-kernel learning, Image and Vision Computing, vol. 29, no. 4, pp. 177–188, Apr. 2011.
8. Ryoo, M. S., Song, H. S., and Pavlovic, V., Robust facial action unit recognition using discriminative multi-kernel learning, in Computer Vision and Pattern Recognition (CVPR), IEEE Conference on, Jun. 2012, pp. 1788–1795.
9. Kollias, D., Nikitidis, S., Tzimiropoulos, G., and Pantic, M., Real-time facial action recognition in unconstrained situations with hierarchical multi-kernel learning, in Computer Vision and Pattern Recognition (CVPR), IEEE Conference on, Jun. 2013, pp. 3510–3517.

10. Singh, M., Girish, G., Srivastava, K., and Singh, P., Hybrid facial action recognition using multi-kernel learning, Intelligent Control and Automation, vol. 5, no. 3, pp. 173–184, Jun.2014.

11. Chang, C.-T., Chiu, C.-C., Wu, M.-L., Hsu, P.-C., Ray, S., and Chen, C.-T., Facial action unit recognition with heterogeneous feature fusion and gemkl multi-kernel learning, in Proceedings of the 21st ACM International Conference on Multimedia, Oct. 2013, pp. 983– 986.

12. Wang, L., Wang, X., Zhao, J. J., Yang, M., and Han, J., Hierarchical multi-kernel learning based facial expression recognition, IEEE Transactions on Image Processing, vol. 22, no. 11,pp. 4450–4459, Nov. 2013.

13. Fookes, C., Schofield, G., Cox, S., Hatzisreatas, A., and Christmas, W., Unsupervised facialaction unit recognition using multimodal probabilisctic fusion, in Pattern Recognition, International Conference on, Sept. 2010, pp. 3297–3301.

14. Schuller, B., Wagner, M., Kovacs, I., and Puigdevall, M., Multi-kernel learning for face properties recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 8, pp. 1638–1651, Aug. 2011.