International Journal of Computer Science
and Engineering (IJCSE)
ISSN(P): 2278-9960; ISSN(E): 2278-9979
Vol. 6, Issue 1, Dec - Jan 2017; 43-50

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

A SURVEY ON FACIAL EXPRESSION RECOGNITION

P. DINESH KUMAR[1] & B. ROSILINE JEETHA[2]

[1]Research Scholar, PG and Research Department of Computer Science,
Dr. N.G.P. Arts and Science College, Coimbatore, India

[2]Professor and Head, PG and Research Department of Computer Science,
Dr. N.G.P. Arts and Science College, Coimbatore, India

ABSTRACT

Computer vision is one among the thrust research area in the field of Image processing. Facial expression recognition is one among the thrust research dimension in computer vision. The process of recognition and identification is important due to the similarity of facial expressions. The motivation behind this research area is its capability to resolve an image processing problem and its wide range of applications. The main rationale of all images processing and computer vision algorithms is to build the visual data in a useful manner. For this reason in the domain of computer vision, the facial expression recognition begins with its applications in the HCI (Human Computer Interaction) where visual look of human, sight and touch sensations (also known as moods) and voice are utilized at the same time. This research article reviews several literatures pertaining to facial expression recognition.

KEYWORDS: Computer Vision, Image Processing, Machine Learning Algorithms, Facial Expression Recognition, Human-Computer Interaction

1. INTRODUCTION

Facial expression recognition is a process that contains recognition of cognitive action, warping of facial feature and also facial movements. This process is carried out by making use of static images and their series or even with videos. The rationale of FER is to sort out the images or videos into dissimilar abstract classes which depends on the visual facts only. Classically, human faces reproduce the inner feelings or emotions which therefore are defenseless to modifications among the environment. There exists vast research scope in facial expression recognition that aids in interpreting the states of mind and classifies several facial gestures.

Existing facial expression databases are classified into two classes: Lab-based database, where the emotions are intentionally expression under controlled environment and realistic database, where the emotions naturally occur in an uncontrolled environment (i.e. real-world conditions). Most of the existing facial expression database belong to the first class which include JAFFE, BU-3DFE, CK+, Semaine, SAL, MMI, AAI and NVIE. In contrast to lab-based emotions, realistic facial expression contains big variations in illumination face pose, size and facial occlusions etc and thus they are more challenging to categories and have more importance in real-world applications.

In spite of the variation in age, ethnicity, gender there is similarity in facial expressions, there exist six expressions named as sadness, pleasure, fury, surprise, panic and repulsion. These six emotions are recognized as basic emotions with their own and diverse nature. This natural similarity in facial expressions of human is making use of by each facial expression recognition system. The motivation behind any research area is its capability to determine a problem and its applications. The major intention of all images dispensation and computer visual algorithms is to create the visual data functional. Therefore in the domain of computer visualization, the facial expression recognition begins by means of the identical function. The significance and necessity of this research area is enhanced because of its applications in the HCI (Human Computer Interaction) wherever visual look of human, sight and touch sensations (also known as modes) and voice are employed at the same time. In addition, social psychology states that facial expressions chains in synchronized conversation. It is noteworthy that Facial expression produces 55 % of the result of spoken message if it delivered along with visual information. The involvement of verbal words is 7% and vocal tone supplies 38%. Therefore facial expression is the most considerable along with all of HCI modes. This constructs the research on facial expression recognition necessary. This paper is organized as follows. Section 2 gives a brief description of the related works carried out in facial expression recognition research. Section 3 provides findings and conclusions.

2. RELATED WORKS

Facial expression recognition engages recognition of cognitive activity, deformation of facial feature and facial movements. This is done with the help of static images and their sequences or videos. The purpose is to categorize them into different abstract classes is foundation on the visual facts only. Obviously, human faces generally reflect the inner feelings/emotions and therefore facial expressions are susceptible to modification in the environment. This makes a human face index of mind; consequently, expression recognition supports in interpreting the states of mind and distinguishes between various facial gestures. The process of recognition and identification is important due to the similarity of facial expressions. The deformation happens due to expressed emotions on the human faces.

In [1] the authors (Yan et al, 2011) proposed a transfer subspace learning approach cross-dataset for facial expression recognition. Their chosen problem has been seldom addressed in the literature. While a lot of facial expression recognition methods have been proposed in modern years, the majority of them believe with the intention of face images in the training and testing sets are collected under the similar circumstances so that they are autonomously and indistinguishable distributed. In many real applications, this assumption will not embrace as the testing data are typically collected online and are generally more uncontrollable than the training data. Therefore, the testing samples are probably different from the training samples. The authors defined the problem as cross-dataset facial expression recognition as the training and testing data are measured to be collected from different datasets due to different acquisition conditions. In order to address this research problem they proposed a transfer subspace learning approach to study a feature subspace which transfers the knowledge expanded from the source domain (training samples) to the objective domain (testing samples) to get better in the recognition performance. To better exploit more complementary information for several feature depictions of face images, they have also developed a multi-view transfer subspace learning approach wherever multiple different yet interconnected subspaces are present to be learned to transfer information from the source domain to the objective domain. Experimental results are obtainable to demonstrate the effectiveness of these proposed methods for the cross-dataset facial expression recognition task.

In [2] the author (Yan, 2016) proposed a biased subspace learning approach for misalignment-robust facial expression recognition. Although a diversity of facial expression recognition techniques have been proposed in the literature, most of them only work well once while face images are well scheduled and supported. In many practical applications such as human robot interaction and visual surveillance, it is constantly demanding to achieve well-aligned face images for facial expression recognition because of present deficient computer vision techniques, especially under uncontrolled conditions. Motivated by the fact that interclass facial images by means of little distinction are more easily mis-classified than those with large differences, the authors proposed a technique called Biased Linear Discriminant Analysis (BLDA) by imposing huge penalties on interclass samples with little distinction and little penalties on those samples with huge distinctions simultaneously, so that discriminative features can be improved extracted for recognition. In addition, the authors also generated more virtually misaligned facial expression samples and allocate different weights to them respect to their amount of probabilities in the testing phase to learn a weighted BLDA (WBLDA) attribute space to extract misalignment-robust discriminative features for recognition. To improved exploit the geometrical information of face samples, the authors have proposed a weighted biased margin Fisher analysis (WBMFA) technique by utilizing a graph embedding criterion to extract discriminative information, in order to the theory of the Gaussian distribution of samples is not necessary. Experimental results on two extensively used face databases are open to show the efficacy of their proposed methods.

In [3] the authors (Ali et al, 2016) proposed a boosted NNE (neural network ensemble) collections based method for multicultural facial expression recognition. The boosted NNE collections based ensemble classifier engaged three steps: first and foremost step is to the training of binary neural networks, second step is to combine the predictions of binary neural networks to form NNE, and third and final step is to combine the predictions of NNE collections in sequence to sense the occurrence of an expression. The outcomes of binary neural networks are coupled to the probability value across the NNE collection. The improved technique is applied for the creation of NNEs and the final prediction is made by Naive Bayes classifier. The acted still images from three databases JAFFE, TFEID, and RadBoud derived from four different cultural and civilizing regions including Japanese, Taiwanese, Caucasians and Moroccans are united to develop the cross cultural facial expression dataset. This expression dataset of cross cultural facial is preserved for the training and testing of binary neural networks in every NNE collection. Three diversified feature extraction techniques PCA, LBP and HOG are used for sample image representation. Their experimental outcomes and statistical analysis of anticipated method for multicultural facial expression recognition constitute the involvement to the field.

In [4] the authors (Lopes et al) proposed a simple solution for facial expression recognition that utilizes a grouping of Convolutional Neural Network and precise image pre-processing steps. Convolutional Neural Networks accomplish enhanced precision with big data. Conversely, there are no publicly accessible datasets with adequate data for facial expression recognition with deep architectures. Therefore, to tackle the problem, the authors applied some pre-processing techniques to extract merely expression precise features since a face image and investigate the appearance order of the samples throughout training. Their experiments engaged to evaluate our method were carried out using three mostly utilized public databases (CK+, JAFFE and BU-3DFE). From the results, they showed that their proposed technique accomplish aggressive results when evaluated with other facial expression recognition methods.

In [5] the authors (Zheng, 2016) proposed a multi-task facial inference model (MT-FIM) for concurrent face recognition and facial expression recognition. Specifically, face identification and facial expression recognition are learnt concurrently by extracting and utilizing appropriate shared information transversely them in the framework of multi-task learning, wherein the shared information submits to the parameter controlling the sparsely. MT-FIM concurrently reduces the within-class scatter and maximizes the distance between different classes to facilitate the vigorous performance of each individual mission. The authors conducted comprehensive experiments on three face image databases. The experimental outcomes illustrate that our algorithm do better than the state-of-the-art algorithms.

In [6] the authors (Long and Bartlett, 2016) presented a new video-based facial expression recognition technique is done by automatically learning features from video data. Exclusively, the authors employed sparse coding algorithm to discover spatiotemporal features beginning with unlabeled facial expression videos. For representing spatiotemporal layout information embedded in facial expressions to develop recognition performance, the authors extended the thought of spatial pyramid matching (SPM) addicted to video case, and carry out spatiotemporal pyramid feature pooling subsequent sparse coding feature extraction. Experimental outcomes on extensively used Cohn–Kanade database demonstrate that the classification performance can be enhanced effectively by considering spatiotemporal layout of facial expressions, and the authors also claimed that their method outperforms popular methods using hand-designed features.

In [7] the authors (Wang et al, 2017) introduced a new learning method that trains an action unit (AU) classifier using images with deficient AU annotation but with comprehensive expression labels. The goal is to use expression labels as hidden knowledge to balance the missing AU labels. Progressing towards this goal, the authors constructed a Bayesian network (BN) to capture the interaction among facial expressions and AUs. Structural expectation maximization (SEM) is used to learn the structure along with parameters of the BN while the AU labels are missing. Given the learned BNs and measurements of AUs and expression, the authors performed AU recognition within the BN through a probabilistic inference. An experimental result on the CK+, ISL and BP4D-Spontaneous databases displays the efficiency of our method for both AU classification and AU intensity estimation.

In [8] the authors (Sormaz et al, 2016) examine the surface information to be used in the recognition of facial expression. First, participants are recognized with their facial expressions (fear, anger, disgust, sadness, happiness) from images that were influenced such that they diversify mainly in shape or primarily in surface properties. The authors found that the classification of facial expression is promising in either type of image, but that different expressions are relatively dependent on surface or shape properties. Next, the authors investigated the relative contributions of shape and surface information to the categorization of facial expressions that employed a complementary method by combining the surface properties of one appearance with the shape properties from a diverged expression. Their results showed that the categorization of facial expressions in these hybrid images was similarly dependent on the surface and shape properties of the image.

In [9] the authors (Liong et al, 2016) presented a novel method for detecting and recognizing micro-expressions by utilizing facial optical strain magnitudes to build optical strain property and optical strain weighted property. The two sets of features are subsequently concatenated to form the resulting featured histogram. Experiments are performed on certain databases and the usefulness of optical strain information and added prominently, that their approaches are able to outperform the original baseline outcomes for both detection and recognition responsibilities. A comparison of the proposed method with other existing spatio-temporal feature extraction approaches is furthermore obtainable.

In [10] the authors (Shao et al, 2015) focused on the problem of 3D dynamic facial expression recognition. Their approach works directly on low-resolution RGB sequences which allows to apply their algorithm to videos recovered by extensive and standard low-resolution RGBsensors. After preprocessing both RGB and depth image sequences, sparse features are learned through spatio-temporal local cuboids. Conditional Random Fields classifier is then engaged for training and classification. Their proposed system is fully-automatic and achieves superior consequences on three low-resolution datasets constructed from the 4D facial expression recognition datasets.

In [11] the authors (Li et al, 2015) presented a fully automatic multimodal 2D + 3D feature-based facial expression recognition approach. Their approach combines multi-order gradient-based local texture and shape descriptors in sequence to attain efficiency and robustness. First, a huge set of fiducial facial landmarks of 2D face images along with their 3D face scans are localized by making use of a novel algorithm namely incremental Parallel Cascade of Linear Regression. Then, a novel Histogram of Second Order Gradients (HSOG) based local image descriptor in combination with the broadly used first-order gradient supported with SIFT descriptor are utilized to depict the local texture approximately with each 2D landmark. In the same way, the local geometry approximately each 3D landmark is portrayed by two original local shape descriptors constructed by means of the first-order and the second-order surface differential geometry quantities, i.e., Histogram of mesh Gradients (meshHOG) and Histogram of mesh Shape index (curvature quantization, meshHOS). To conclude, the Support Vector Machine (SVM) based recognition outcomes of all 2D and 3D descriptors are combined at both feature-level and score-level to further improve the precision. complete experimental results reveal that there exist impressive complementary characteristics between the 2D and 3D descriptors. The authors also compared their approach to the state-of-the-art ones. Our multimodal feature-based approach outperforms the others.

In [12] the authors (Wang et al, 2015) intended to get better the recognition accuracy by providing a new advance technique for facial expression recognition organized with Fuzzy Support Vector Machine (FSVM) and K-Nearest Neighbor (KNN). At first, the property of the static facial expression image is removed by the Principle Component Analysis (PCA), then, the algorithm splits the region into different types, and merges with the attribute of the FSVM and KNN, switch the classification methods to the different types. The results of their experiment showed that their proposed algorithm is capable enough to attain good identification accuracy and make things easier to the computation complexity.

In [13] the authors (Zhang et al, 2015) proposed multimodal learning for facial expression recognition (FER) method that first attempt to discover the combined representation by considering the texture in addition to landmark modality of facial images, are complementary with each other. In order to learn the demonstration of each modality in addition to the correlation and interaction connecting different modalities, the structured regularization (SR) is engaged to implement and learn the modality-specific sparsity and density of each modality. Correspondingly by launching SR, the range of the facial expression is fully taken into consideration, which can not only hold the subtle expression but in addition achieve robustly to different input of facial images. With their proposed multimodal learning network, the joint representation learning from multimodal inputs will be further appropriate for FER. Experimental outcomes on certain databases demonstrated the superiority of their proposed method.

In [14] the authors (Pu et al, 2015) proposed a novel framework for facial expressions analysis by recognizing AUs from image sequences using twofold random forest classifier. The measurement of facial motion is through tracking of Active Appearance Model (AAM) facial feature points by means of Lucas–Kanade (LK) optical flow tracker by estimating the displacements of the feature points. The displacement vectors connecting the neutral expression frame in addition to the peak expression frame are implemented as motion features of facial expression. They enforce and they are transformed to the first level random forest to verify the Action Units (AUs) of the equivalent expression sequences. Finally, the detected AUs are inputed into the second level is arbitrary forest for facial expressions classification. Their experiments on Extended Cohn–Kanade(CK+) database reveal that the proposed technique can accomplish higher performance than several other approaches on both AUs and facial expression recognition.

In [15] the authors (Wang et al, 2016) proposed a novel sparse learning method called Sparse Local Fisher Discriminate Analysis (SLFDA) for facial expression recognition. The SLFDA method is derived from the original local Fisher discriminant analysis (LFDA) and makes use of its sparse property. The sparse solution is obtained by finding the minimum l, 1 -norm solution commencing from the LFDA solutions. This difficulty is then formulated as an l 1 -minimization problem and solved by linearized Bregman iteration, which assurance convergence and is implemented. Their proposed SLFDA can deal with multi-modal troubles as well as LFDA; besides, it has more discriminate power than LFDA because the non-zero elements in the basis images are chosen from the mainly significant factors or regions. Experiments on several benchmark databases are performed to test and evaluate their proposed algorithm. Their results showed the effectiveness of SLFDA.

3. FINDINGS AND CONCLUSIONS

Among the literatures significant research contributions has been made. Many techniques and mechanism are proposed for FER which includes transfer subspace learning approach, biased subspace learning approach, boosted NNE (neural network ensemble), multi-task facial inference model (MT-FIM), sparse coding algorithm, bayesian network classifiers, fuzzy support vector machine (FSVM) and k-nearest neighbor (KNN) are reviewed. Among all the mentioned approaches the FSVM with KNN mechanism performs better in terms of accuracy. But there exists certain research scope which is capable enough to improve accuracy. This can be achieved by exploiting image processing techniques such as noise removal, feature selection and classification.

REFERENCES

1. H. Yan, M. H. Ang and A. N. Poo, "Cross-dataset facial expression recognition," IEEE International Conference on Robotics and Automation (ICRA), pp. 5985-5990, 2011.
2. H. Yan, “Biased subspace learning for misalignment-robust facial expression recognition,” Neurocomputing, vol. 208, pp. 202 – 209, 2016.
3. G. Ali, M. A. Iqbal, T. S. Choi, "Boosted NNE collections for multicultural facial expression recognition," Pattern Recognition, vol. 55, pp. 14 - 27, 2016.
4. A. T. Lopes, E. D. Aguiar, A. F. De Souza, T. O. Santos, "Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order," Pattern Recognition, (Article in Press).
5. H. Zheng, X. Geng, D. Tao, Z. Jin, "A multi-task model for simultaneous face identification and facial expression recognition," Neurocomputing, vol. 171, pp. 515 – 523, 2016.
6. F. Long, M. S. Bartlett, "Video-based facial expression recognition using learned spatiotemporal pyramid sparse coding features," Neurocomputing, vol. 173, pp. 2049 – 2054, 2016.
7. S. Wang, Q. Gan, Q. Ji, "Expression-assisted facial action unit recognition under incomplete AU annotation," Pattern Recognition, vol. 61, pp. 78 – 91, 2017.

8. M. Sormaz, A. W. Young, T. J. Andrews, "Contributions of feature shapes and surface cues to the recognition of facial expressions," Vision Research, vol. 127, pp. 1 - 10, 2016.
9. S. T. Liong, J. See, R. C. W. Phan, Y. H. Oh, A. C. L. Ngo, K. Wong, S. W. Tan, "Spontaneous subtle expression detection and recognition based on facial strain," Signal Processing: Image Communication, vol. 47, pp. 170 - 182, 2016.
10. J. Shao, I. Gori, S. Wan, J.K. Aggarwal, "3D dynamic facial expression recognition using low-resolution videos," Pattern Recognition Letters, vol. 65, pp. 157 - 162, 2015.
11. H. Li, H. Ding, D. Huang, Y. Wang, X. Zhao, J. M. Morvane, L. Chen, "An efficient multimodal 2D + 3D feature-based approach to automatic facial expression recognition," Computer Vision and Image Understanding, vol. 140, pp. 83 - 92, 2015.
12. X. H. Wang, A. Liu, S. Q. Zhang, "New facial expression recognition based on FSVM and KNN," Optik - International Journal for Light and Electron Optics, vol. 126, pp. 3132 - 3134, 2015.
13. W. Zhang, Y. Zhang, L. Ma, J. Guan, S. Gong, "Multimodal learning for facial expression recognition," Pattern Recognition, vol. 48, pp. 3191 - 3202, 2015.
14. X. Pu, K. Fan, X. Chen, L. Ji, Z. Zhou, "Facial expression recognition from image sequences using twofold random forest classifier," Neurocomputing, vol. 168, pp. 1173 - 1180, 2015.
15. Z. Wang, Q. Ruan, G. An, "Facial expression recognition using sparse local Fisher discriminant analysis," Neurocomputing, vol. 174, pp. 756 - 766, 2016.