Multiple Feature Fusion Based Facial Recognition in Video and Audio

[1]Miss Aparna Gautam, [1]Miss Anwesha Mukhopadhyay, [1]Miss Steffi Jerald, [2]Mrs R.Deepa,
[1]Student, Department Of CSE,SRM IST,SRM, Vadapalani
[2]Assistant Professor,Department Of CSE,SRM IST,SRM, Vadapalani

Abstract: There has been a long existing issue of video based facial expression recognition and attracted burgeoning attention lately. The pivotal to a purposeful identification structure is due to the usage of all the dimensions of the techniques for audio as well as video along with construction of sturdy functionalities that adequately differentiate the changes caused in the configuration and the landmarks of the face due to the movements on the face. Our aim is to establish productive bodywork addressing the concern in this script. All the dimensions of the techniques for audio as well as video are explored with this paper. The active patterns are excerpted using the arrays of video to portray aspects of the variations in the face which is then projected using a feature descriptor that takes care of all the three planes of the orthogonal and is known as histograms of oriented gradients. The transformation of the landmarks of the face that seizes the configurational changes of the face using warp helps to infer a purposeful geometric feature. Additionally, the dimensions of the audio are scrutinized in this examination. The complications of the video based expression identification of face are tackled using the numerous feature combinations under lab-controlled environs as well as in the wild. Trials guided with the database manifest the fact that the system forms its basis on vigorous designs which deals with issues relating the recognition of facial aspects obtained from video covered by regulations of the lab conditions.

***Index Terms* –Recognition of face, Warp Transformation, Histogram of Gradients, Multiple Feature Fusion,Mfcc**

INTRODUCTION

The processing of a 2D image by a digital computer is referred as digital image. Basically it means digitally processing any 2D data. An arrangement of actual or composite figures presented by a definite digits is termed as a automated design. The computer system saves the picture as model which is initially digitized. The high-resolution television monitor exhibits and/or processes the digitized image. In the display part, a rapid-access buffer memory stores the image and turns on the screen at a rate of 25 fabric per second to acquire a viewable uninterrupted picture.

1. PROCESSING OF IMAGE:

An model processor carries out the function of image procurement, stockpile, pre-processing, subdivision, presentation, perception, and apprehension and eventually shows or stores the output picture .The image processing system contains the following fundamental sequence which is shown by block diagram. As showed in the image, the initial step in the process is image procurement by an sensing machine in co-operation with a digitizer which digitizes the picture. Followed by the pre-processing for improvising the image then forward the image as an input to the other processes. Pre-processing manages enhancing, removal of sound, domain, etc. Subdivision divides a picture into its respective parts. The raw pixel values, having either the regional pixels themselves or the periphery of the region forms the output of segmentation. The computerized technique of changing the raw pixel records into an arrangement helpful for successive operations is termed as representation. Description handles extraction of features which are fundamental in segregating among object class. Based on the data provided by the descriptors, recognition assigns a label to an object. A collection of recognized objects is assigned description on interpretation. The knowledge base contains the knowledge concerning a problem domain. The operation and interaction of each processing module is not only dictated but also controlled by the knowledge base. The presence of each module is not needed for a specified function. Based on the application the configuration of the picture transforming machine relies on. The rate of frames for the picture transformer is approximately around 25 fabric/second.

2. DIGITIZER:

A picture is transformed into a numerical representation using a digitizer which is an appropriate input for a automated system. Few of the digital systems are:

1. Videocon camera
2. Spot scanner
3. Photosensitive solid- state arrays.
4. Microdensitometer
5. Image dissector

3. PRINCIPLES OF IMAGE PROCESSING:

Transforming of a picture in automated form refers to digital image processing. Cameras might directly capture the picture in automated type however pictures are initiated in visual form. They are taken by visual cameras and digitalized. Digitalization incorporates examining as well as performing. Five fundamental processes are used for processing these images, and minimum of an image is necessary.

A. ENHANCEMENT OF IMAGE:

The picture improvement enhances the calibre of an picture like increasing the picture's comparison and lightness features, decreasing its sound data, or hone the data. The image is enhanced therefore revealing the similar data in more comprehensible picture, not adding any details to it.

B. RESTORATION OF IMAGE:

The quality of picture is improved by image restoration like enhancement although the actions are primarily relies on predefined measures, or degeneration of the initial image. Images are restored using image restorations with issues such as mathematical deformation, inappropriate target, recurrent sound, and camera movement. It is used for rectifying pictures for familiar disruption.

C. EXAMINATION OF IMAGE:

The mathematical or statistical data relies on features of the actual image is produced by image examination functions. They are broken into objects and then classified which is dependent on the picture data. Abstraction and contour of scene and picture characteristics, digital quantification, and grouping of objects are the common operations. System vision data are based on picture examination.

D. COMPRESSION OF IMAGE:

Compressing and decompressing the image involves the scaling down of the crucial contents of the data for defining the image. A lot of times, abundant amount of unnecessary material is present in the images, compression helps in eliminating these sort of redundancy. It just doesn't perform size reduction but also effectively stores and transports it. Decompression takes place when the compressed image is to be displayed. Though the data in the original image is preserved exactly by the compression that bears less loss, yet representation of the original image is not done by lossy compression thereby providing a compression of an admirable value.

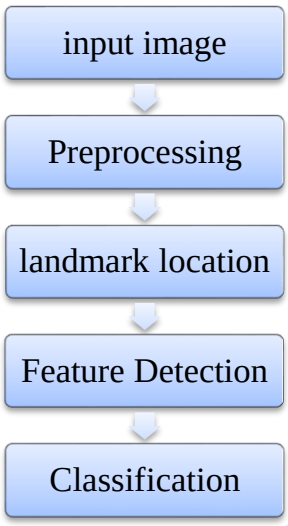
E. SYNTHESIS OF IMAGE:

Images that are formed using different images mark the function of image synthesis. Creation of images that are either physically unfeasible or inappropriate to amass are typically formed by image synthesis.

EXISTING SYSTEM

Fixed appearance design and progressive design alone depends on viewable modal quality. Nonetheless, phonics are too necessary for a person to fetch empathy and objective. Audile modal quality should give a few supportive data along with visible modal quality. Lately, audile-viewable designs for emotional perception have allured increasing awareness from the affective figure company. A lot of access has been put forward to merge audile and viewable modal qualities for emotional perception. An extensive view can be formed in. Audio features withdrawn from vocal and visible appearances withdrawn from facial pictures are merged to solve this situation. For example, vocal and speech undertaking were focused. Facial pictures and vocals were engaged. The designs expressed in, adapted different appearances algorithms like SIFT, HOG, PHOG etc. to conceal the facial pictures and merge them with audile appearances to identify the face language in extent

MODULE DIAGRAM



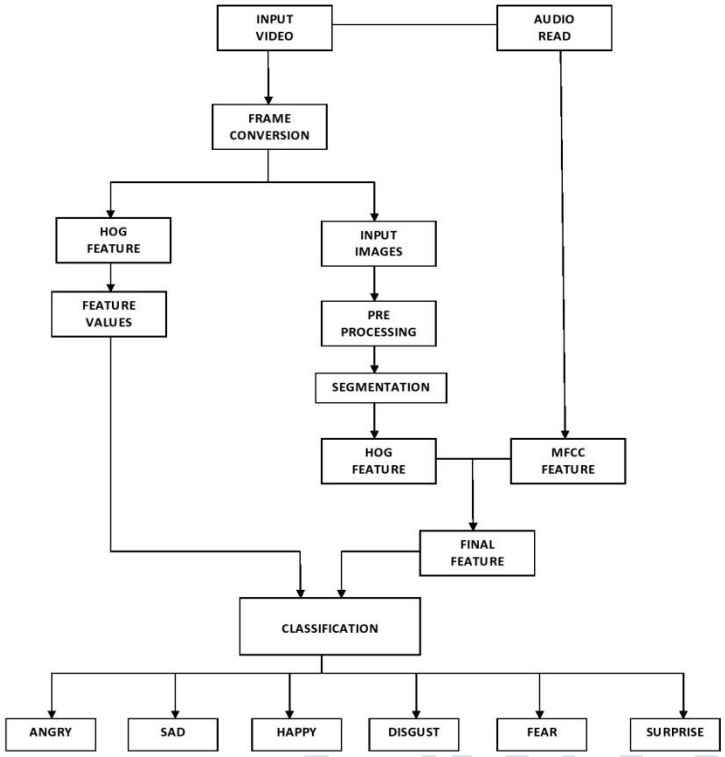
DISADVANTAGES

- Existing system focuses on fixed and mono facial pictures based face interpretation identification. Lately, face interpretation identification in viewable parts has allured huge engrossment.
- Contrasted with fixed pictures, viewable succession can not only give dimensional emergence but also adds face expressions and followed vocals.
- The solution to tackle the situation of viewable based face interpretation identification is to utilize the depiction ability of multiple modal qualities and create strong attributes to successfully distinguish the face depiction and composition adjustments caused by face interpretations.
- The drawback of the SVM algorithm is that it has several key variables that need to be set correctly to attain the best categorization results for any given situation. Variables that may result in a good categorization accuracy for the problem A, may result in a poor categorization correctness for problem B.

PROPOSED SYSTEM

Feature extraction stands out to be of great significance, when it comes to facial affect recognition for a video. Drafting a productive feature is extremely essential and purposeful. In our proposed system, we aim to put forward a new feature, HOG-TOP, known to be functional and effectual; it extensively distinguishes the facial aspect changes. Furthermore, composition and figure play a vital character in terms of apprehension of facial expression by the human vision. And with no doubt, it is rest assured that the existing system has not attained and explored all the dimensions of configuration representations. Our proposed system also offers a higher potent geometric feature for apprehension of facial configuration changes.

ARCHITECTURE DIAGRAM



ADVANTAGES

1. Since the system uses HOG-TOP, it has greater feature dimensionality when compared to LBP.
2. Use of HOG-TOP makes the system more compact.
3. Use of KNN makes the cost for the learning process as null
4. KNN makes the complex concepts be dealt by local approximation and simpler procedures.

MODULE DESCRIPTION

MODULE 1

FRAME CONVERSION

It begins with the input which is in video format being transformed into a frame sequence. The system illustrates the conversion of an avi file into a succession of frames. The code is use for processing videos in MATLAB to convert them to frames.

MODULE 2

FEATURE EXTRACTION

A. HISTOGRAMS OF ORIENTED GRADIENTS

Histograms of oriented gradients were initially introduced for detection of human. The descriptor is object-deformation sensitive and hence characterizes composition and appearance of objects effectively using directions of edges as well as gradients for the localized intensity. Towards the beginning, this algorithm was restrained with static images and hence to overcome this and design dynamic patterns from a video using HOG, there has been an extension named as three orthogonal planes(TOP) XY,XT and YT to identify facial expression changes more efficiently.

B. GEOMETRIC FEATURE

Geometric warp feature, a very high potent geometric feature which finds its roots along with the transformed aspects of the face using warp. Movement of the facial muscles lead to formation of facial expressions. These movements further lead to the shift of facial landmarks. The face is divided into sub regions which are nothing but triangles having their vertexes placed at facial landmarks. The proposed system makes complete use of these deformations to present the facial composition changes.

C. MFCC FEATURE

An audio's or a sound's power spectrum is represented by the cepstral coefficients of the mel frequencies (MFCC) and it forms its basis on nonlinear scale of frequency. The frequency bands in the MFCC are spaced equally which makes the audio response more potent than the bands used in the normal cepstrum. The particular achieved frequency warp also makes a room for greater presentation of sound.

MODULE 3

CLASSIFICATION

KNN stands to be an accurate classification for data, as it chooses the number of nearest neighbors to be an odd number to eliminate the scenario of irregular data. It regulates the k-nearest neighbors by working on a minimum distance to the training set. A number of attributes are achieved from the data for the KNN method. All available cases are delivered and categorized by KNN on basis of an evaluation parameter.

CONCLUSION

Here, we explore all the dimensions of the techniques for audio as well as video and put forward a purposeful system. We make use of face image for visual modalities and speech for audio modalities. Our system also makes use of HOG-TOP descriptor, Geometric feature and mfcc feature to contribute a greater hand for detection and identification of facial expressions.

FUTURE ENHANCEMENT

The stemming algorithm can be magnified by using more number of images from the video. More images can be used for performing the sensation than the existing number of images used. The audio can be separated into bits to pin point the sensation by identifying each bit. To maximize the correctness, the pictures can be instructed for more than once by using any other algorithms. If deep learning is used instead of machine learning, the chance of getting a better output is increased.

ACKNOWLEDGMENT

The authors would like to show their gratitude to Prof.R.Deepa of SRM Institute of Science and Technology for sharing her pearls of wisdom, assisting throughout the research process and providing proper guidance to come up with a great manuscript.

REFERENCES

[1] R. A. Calvo and S. D'Mello, "Affect Detection An Interdisciplinary Review of Models, Methods, and Their Applications," IEEE Transactions on Affective Computing, vol. 1, pp. 18-37, 2010.

[2] Y. l. Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, pp. 97-115, 2001.

[3] K. Scherer and P. Ekman, "Handbook of Methods in Nonverbal Behavior Research," UK: Cambridge Univ. Press, 1982.

[4] J. F. Cohn and P. Ekman, "Measuring facial action," 2005.

[5] P. Ekman and W. V. Friesen, "Facial Action Coding System: A Technique for the Measurement of Facial Movement," Consulting Psychologists Press, 1978.

[6] P. Ekman, W. V. Friesen, and J. C. Hager, "Facial Action Coding System: The Manual on CD ROM. A Human Face," 2002.

[7] P. Ekman, "An argument for basic emotions," Cognition & Emotion, vol. 6, pp. 169-200, 1992.

[8] S. Z. Li and A. K. Jain, "Handbook of face recognition," springer, 2011.

[9] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp. 886-893.

[10] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, pp. 915-928, 2007.

[11] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+) A complete dataset for action unit and emotion-specified expression," IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2010, pp. 94-101.

[12] S. W. Chew, P. Lucey, S. Lucey, J. Saragih, J. F. Cohn, and S. Sridharan, "Person-independent facial expression detection using constrained local models," IEEE International Conference on Automatic Face & Gesture Recognition and Workshops, 2011, pp. 915- 920.

[13] S. Taheri, P. Turaga, and R. Chellappa, "Towards view-invariant expression analysis using analytic shape manifolds," IEEE International Conference on Automatic Face & Gesture Recognition and Workshops, 2011, pp. 306-313.

[14] A. Saeed, A. Al Hamadi, R. Niese, and M. Elzobi, "Effective geometric features for human emotion recognition," IEEE 11th International Conference on Signal Processing (ICSP), 2012, pp. 623- 627.

[15] K. Sikka, T. Wu, J. Susskind, and M. Bartlett, "Exploring bag of words architectures in the facial expression domain," in Computer Vision-ECCV Workshops and Demonstrations, 2012, pp. 250-259.

[16] Y. Rahulamathavan, R. C. W. Phan, J. A. Chambers, and D. J. Parish, "Facial Expression Recognition in the Encrypted Domain Based on Local Fisher Discriminant Analysis," IEEE Transactions on Affective Computing, vol. 4, pp. 83-92, 2013.

[17] L. Zhang and D. Tjondronegoro, "Facial expression recognition using facial movement features," IEEE Transactions on Affective Computing, vol. 2, pp. 219-229, 2011.

[18] S. Happy and A. Routray, "Automatic facial expression recognition using features of salient facial patches," IEEE Transactions on Affective Computing, vol. 6, pp. 1-12, 2015.

[19] M. F. Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer, "The first facial expression recognition and analysis challenge," IEEE International Conference on Automatic Face & Gesture Recognition and Workshops, 2011, pp. 921-926.
[20] A. Dhall, A. Asthana, R. Goecke, and T. Gedeon, "Emotion recognition using PHOG and LPQ features," IEEE International Conference on Automatic Face & Gesture Recognition and Workshops, 2011, pp. 878-883.

[21] X. Huang, G. Zhao, M. Pietikainen, and W. Zheng, "Expression Recognition in Videos Using a Weighted Component-Based Feature Descriptor," in Proceedings of the 17th Scandinavian conference on Image analysis, 2011, pp. 569-578.

[22] T. R. Almaev and M. F. Valstar, "Local Gabor Binary Patterns from Three Orthogonal Planes for Automatic Facial Expression Recognition," in Affective Computing and Intelligent Interaction (ACII), 2013, pp. 356-361.

[23] X. Huang, Q. He, X. Hong, G. Zhao, and M. Pietikainen, "Improved Spatiotemporal Local Monogenic Binary Pattern for Emotion Recognition in The Wild," in ACM International Conference on Multimodal Interaction, 2014, pp. 514-520.

[24] X. Huang, G. Zhao, W. Zheng, and M. Pietikainen, "Spatio temporal Local Monogenic Binary Patterns for Facial Expression Recognition," IEEE Signal Processing Letters, vol. 19, pp. 243-246, 2012.

[25] F. Long, T. Wu, J. R. Movellan, M. S. Bartlett, and G. Littlewort, "Learning spatiotemporal features by using independent component analysis with application to facial expression recognition," Neurocomputing, vol. 93, pp. 126-132, 2012.