



Prognosticating and Scaling of Melancholia

Akoramurthy.B¹, Thirugnanam.P², Arthi.J³
Assistant Professor^{1,3}, Senior Assistant Professor²
Department of Computer Science & Engineering^{1,2,3}
IFET College of Engineering^{1,2}, Villupuram, Tamilnadu.
Kongu Engineering College³, Perunthurai, Tamilnadu.

Abstract: Nowadays, It's common among people facing "the blues" or feeling "down in the Dump" irrespective of their ages. Such symptoms are familiar to all of us, if they persist and affect our life substantially, which is medically termed as disorder and is known as depression or melancholia. Among all the other disorders, melancholia should be treated as primer. The system proposed as "artificial intelligent system" which may be a buzz word, but the system prognosticate the depression level based on the visual and vocal expression of a person. The results shows up the need for the system where recognition accuracy of the system developed outperforms with high accuracy 88% comparably with human recognition 85%.

Keywords: Melancholia; Depression; motion history histogram; ridge regression algorithm; local binary pattern; low level descriptor; edge orientation histogram ;pitch detecting algorithm.

I. Introduction

In recent year, many problems occur due to depression. Over stress leads to depression. Due to this mental disorder people suffers a lot both in personal as well as in society side too. As we human begin can understand other human's expressions, feelings, emotion etc. But it is difficult for machine to understand. Main contribution in this paper is used to find the depression level of a human by a machine. Many varieties of videos are given as an input. As videos is a dynamic one, which contain lots of features in it. All the features are extracted by means of motion history histogram (MHH).after the extraction, it is separated into two sides, one is a visual side another is a vocal side.

Visual side is measured by means of two techniques:

- ✓ Local binary pattern (LBP), which is used to denote the images by means of 0's (when image doesn't do any actions) 1's (when image does the action).
- ✓ Edge orientation histogram (EOH) which is used to detect the outer most part of an image. The edge of an image has been denoted, in order to find the shape and expression of a human.

Vocal side is measured by means of two techniques:

- ✓ Low level descriptor (LLD) which is used to find the energy and speed level of a human in their voice.
- ✓ Pitch detecting algorithm (PDA) which is used to denote the voice tone (high pitch, medium pitch, low pitch) of a human in order to find the depression level.

Ridge regression algorithm (RRG) plays a major role in output side which is used to combine the data's got in visual side and the data's got in vocal side merge it and produce an depression level result.

II. Literature Survey

Stereotypic behaviors are present in both human and non-human primates. Stereotypic behavior is also one of the symptoms for human mental disorder. The paper studied the stereotypic behavior with the help of machine learning algorithms such as support vector machine (SVM) and Hidden Markov model (HMM). The consequence of the above mentioned algorithms are compared with up-to-the-minute data. HMM holds an upper hand of 98.89% over SVM approach [1]. [2] Fatemeh Noroozi et al discussed a multimodal emotion recognition system, grounded on the examination of audio nodes in which Mel-frequency cepstral co-efficient filter bank elans and prosodic sorts are mined, Whereas in the visual side-first, geometric relations were used and then videos are condensed into set-of key frames and those frames has been applied with Convolutional neural network in which the output is produced. [3] M. Valstar et al, emphasizes that Mood disorders are innately associated to emotion. Furthermore, the work insisted people suffered from depression (specifically unipolar one) displays a resilient relationship with the parameters such as dimensions valence, provocation and supremacy. In addition to that, the assessment of Patients level depression in the observation of expressions

(face) and vocal nods along with structured self-report questionnaires framed and used by psychologists and psychiatrists. The work majorly comes out with the challenge of Audio-Visual Emotion recognition (AVEC 2014) which intended more focused study. Consideration of labels for dominance has also been included with more annotators per clips. The paper composed of two challenges; one is to envisage the unceasing values of valence, provocation and supremacy at each instant of time and another one is to expect the value of a single self-reported sternness of melancholia sign for each hold in the dataset used.

^[4] Speech handling is being progressively reconnoitered in human domains; Daniel Bone et al highlighted issues in health-related speech research and talked about health applications such as addiction counseling and autism spectrum disorder in which the paper asserted a need for knowledge-based signal processing rather than existing deep learning mechanisms. ^[5]Yona Falinie et al in their work proposed an Automatic recognition system to predict the moving dimension parameters. The proposed system extracts an image frames and audio segments and then applied wavelet transform based digital filtering method which reduces inappropriate noise in the collected segments and to the best least square regression applied to predict the dimensions parameters in both modalities which was combined together in order to lift the complete performance in making decisions .

^[6]A.Pampouchidou et al in their proposed work defined depression as common mental disorder which is communal for oodles of people all over the world. The algorithm proposed in this paper dynamically analyzes a facial expression which uses robust descriptor that comprises an unusual feature selection which may lead to effective classification practice. Also, preliminary evaluation suggested a depression severity values based on local binary curve definition

III. Proposed System

The proposed system comes out with a definitive approach, which recognizes a melancholia based on both visual and vocal inklings, which combined together to provide a tremendous advantage in effective prediction. To start with Motion History Histogram (MHH) is exploited to excerpt dynamic features from videos and audios, comprehensively describes emotional fluctuation. And then the extracted features Is to be mapped with modality of melancholia state by using Ridge regression algorithm and the blending should be finally made to scale the depression level of a human.

Proposed Architecture:

For finding the depression level of a human by a machine, the following operations have been used. It is as shown by means of architecture

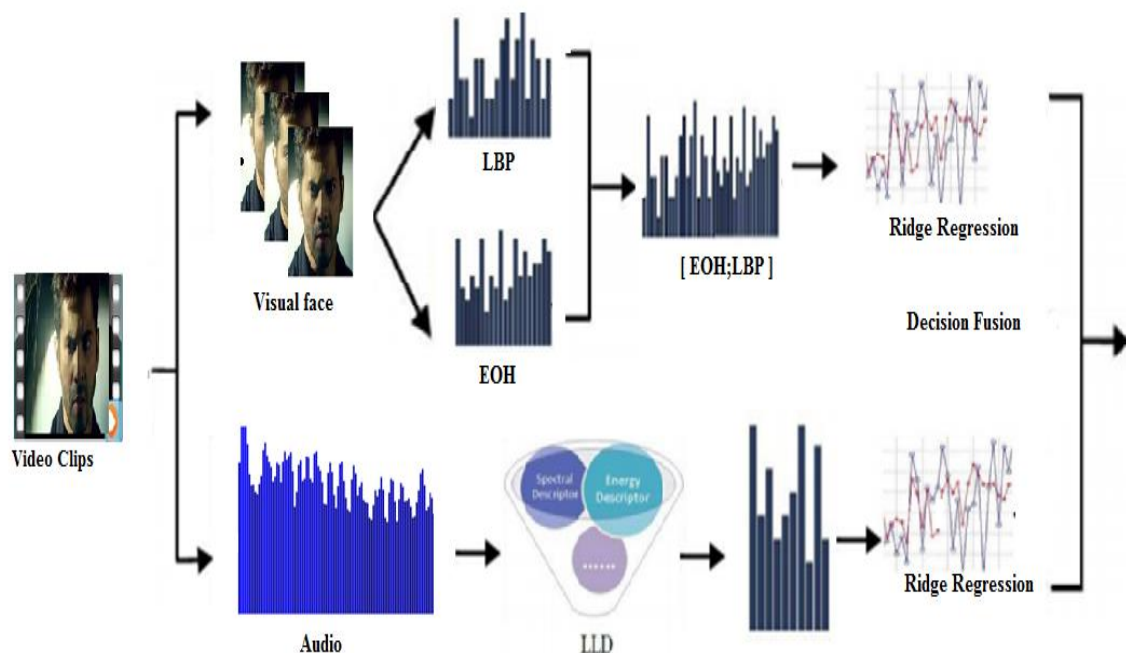


Figure 1 Architecture for Proposed System

Varies and many types of videos is given as an input. The videos can be in KB, MB and GB range. As videos is a dynamic one, which contain many features in it.

The video is extracted by visual and vocal side. This extraction is done by means of motion history histogram (MHH).In visual side it denotes an eyes movement and reaction of a human. In vocal part the voice tone is detected. Local binary pattern (LBP) is used for face images. The face images are divided into many blocks and pixels. Local binary pattern is also used in each of the blocks and plotted by means of graphical manner.

Edge orientation histogram (EOH) is used to find the boundaries of objects within images. It is used to detect the outermost part of an object in order to find the expression of a human. Low level descriptor is used in audio part. It is used to find low level, normal and average range of a voice also used to find speed and energy in an audio. Pitch Detecting Algorithm (PDA) is used to find voice tone. The voice of a human is denoted by means of dB whereas if a person talks normally then his/her voice level approximates to 69-70 dB, and if the same person raises a voice then the sound pressure level increases up to 75dB-76dB. Besides if the person is talking stridently, then his /her decibel ranges from 81dB to 83dB. Finally if someone yells, then the sound pressure level of a person goes up to 88dB-89dB.

Ridge regression algorithm (RRA) is the most usually utilized technique for regularization for poorly postured issues, which are issues that don't have a one of a kind arrangement. Basically; regularization acquaints extra data with an issue to pick the best result for it.

IV. Implementation and Results

Working of Local Binary Pattern operator:

Consider a picture (that is extracted) which can be represented as $\text{Img}(x, y)$ and let g_p denotes the non RGB level of a pixel (which is arbitrary),

$$\text{I.e. } g_p = \text{Img}(x, y).$$

Besides, let g_s indicate the sampling point non RGB values which are equally spread out in neighborhood (which may be circular in nature) of K points of sample and radius R_a will be points in (x, y) range:

$$g_s = \text{Img}(x_s, y_s), \quad k=0, \dots, K-1 \quad \text{and} \quad (1.1)$$

$$x_s = x + R_a \cos(2\pi k/K), \quad (1.2)$$

$$y_s = y - R_a \sin(2\pi k/K). \quad (1.3)$$

Assume that the local texture of the picture $\text{Img}(x, y)$ is considered by the jointly distributed non RGB values of $K+1(K > 0)$ pixels:

$$T = t(g_p, g_0, g_1, \dots, g_{K-1}) \quad (1.4)$$

Then Local Binary Pattern (LBP_{K,R_a}) operator is defined as

$$\text{LBP}_{K,R_a}(x_s, y_s) = K-1 \sum_{k=0}^{K-1} s(g_k - g_s) 2^k \quad (1.5)$$

In run-through, the above equation indicates that the variances neighborhood are construed as a K -bits binary number, then the LBP operator can work in 2^K distinct values.

The non-RGB distribution level, i.e. roughness, will be roughly described with a 2^K -binary values of discrete distribution for the defined operator:

$$F \approx t(\text{LBP}_{K,R_a}(x_s, y_s)) \quad (1.6)$$

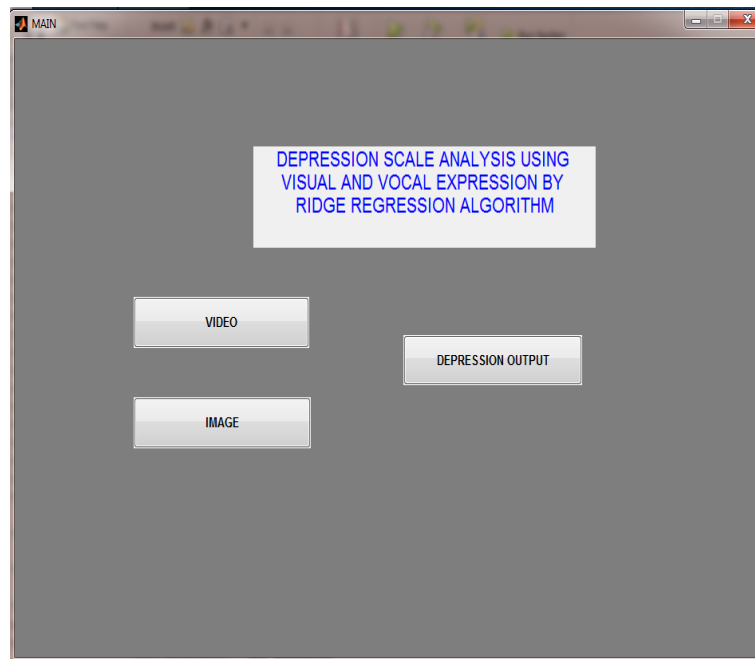


Figure2. Home page for Selecting Videos and Images

The figure 2 describes homepage for selecting videos and images. It contains video option, image option and depression level option. Once the video is clicked, it moves into the next step.

Extraction of Videos

Extraction is done in order to separate many features in videos. Extraction is done by means of motion history histogram. All the features present in the videos are extracted completely by means of MHH. Once the extraction part is completed, the videos are ready for next process.

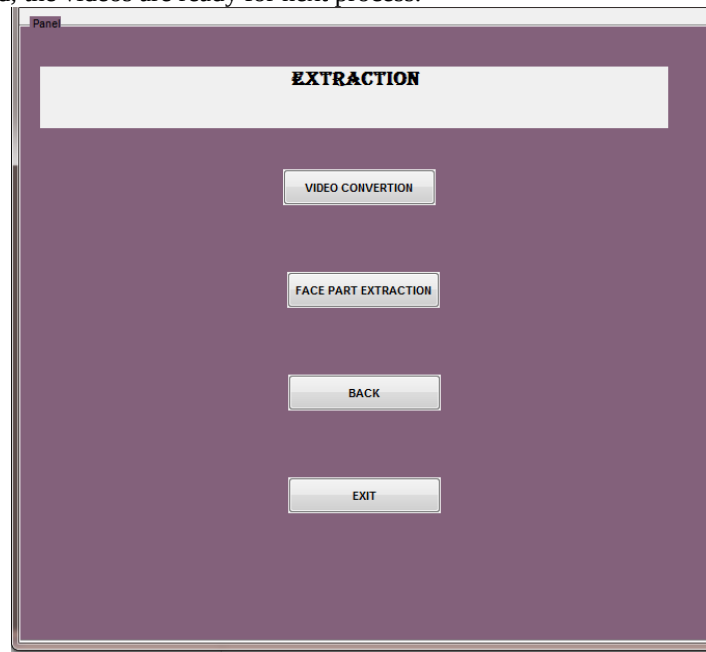


Figure3. Extraction Part

As homepage contain many numbers of videos. Particular video is selected. Once the video is selected, all the features present in that video are converted by means of frames. So frames are stored in homepage side. By doing this frame part conversion have been completed and it moves into a face part conversion.

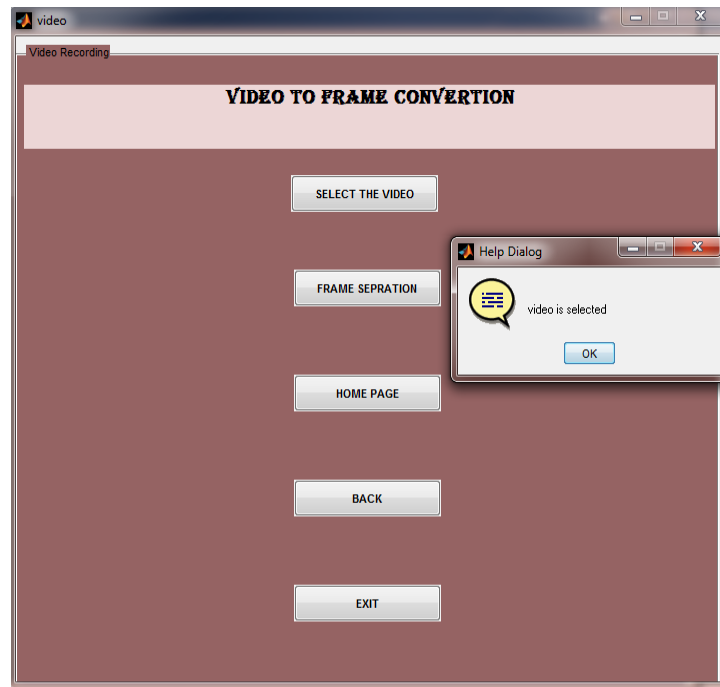


Figure 4.Video to Frame Conversion

Face Part Conversion

The image is selected and it is extracted as face part extraction by means of java program, local binary pattern and edge orientation histogram. It contain input image (image the user selects).it is then separated as face image, paired eyes, left eye brow, right eye brow. By doing this, it is easy for the machine to find the depression level of human.

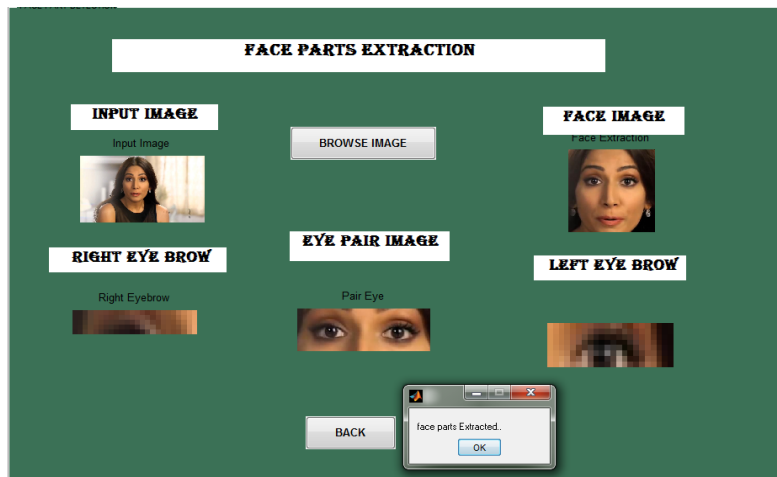


Figure5. Face Part Extraction

Results and Conclusion

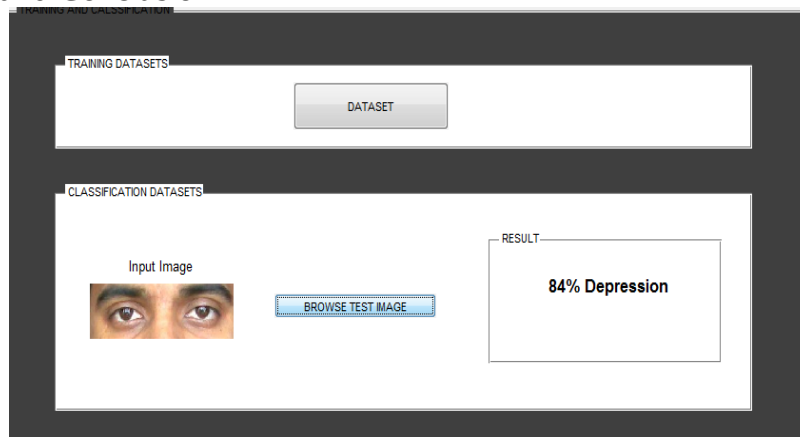


Figure 6. Melancholia Level Results

The work emerged out with the preeminent way of recognizing facial expressions, which will be comparatively higher than that of Human based facial recognition. The datasets used are created and validated by the project team itself for the various applications of emotion recognition. The impetus set deployed can test a variety of emotions and their intensity levels to examine a group differences. The depression level calculated was 85% comparably with human recognition of 81%. Hence the proposed system has upper hand comparably with human recognition and shows up the need for the automated system. Nearly 10 emotions were identified out of which only 2 emotions (Angry & Surprise) were equally identified by both proposed system and human (Figure 7). For the given datasets the depression calculated was .84, which recommend the person needs immediate attention over his/her mental health. The following graph (Figure 7) unveils the Recognition Accuracy of proposed system with human recognition.

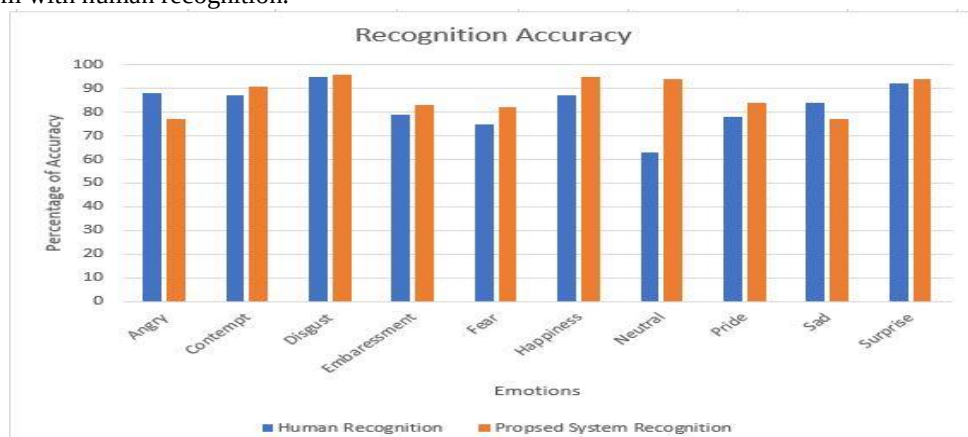


Figure 7. Human Recognition Vs Proposed System Recognition

V. References

- [1] Marcos Y. O. Camada, J'es J. F. Cerqueira, Antonio Marcus N. Lima, "Stereotyped Gesture Recognition: An Analysis between HMM and SVM", 2017.
- [2] Fatemeh Noroozi, Marina Marjanovic, Angelina Njegus, Sergio Escalera, Gholamreza Anbarjafari, "Audio-Visual Emotion Recognition in Video Clips", 2017.
- [3] M. Valstar, B. Schuller, K. Smith, T. Almaev, F. Eyben, J. Krajewski, R. Cowie, M. Pantic, "Avec 2014: 3d dimensional affect and depression recognition challenge", Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge, pp. 3-10, 2014.
- [4] Daniel Bone, James Gibson, Theodora Chaspari, Dogan Can, Shrikanth Narayanan, "Speech and Language Processing for Mental Health Research and Care", 2016.
- [5] Yona Falinie A. Gaus, Hongying Meng, Asim Jan, Fan Zhang, and Saeed Turabzadeh, "Automatic Affective Dimension Recognition from Naturalistic Facial Expressions Based on Wavelet Filtering and PLS Regression", 2015.
- [6] A.Pampouchidou, K.Marias, M.Tsiknakis, P.Simos, F.Yang, F.Meriaudeau, "Designing a Framework for Assisting Depression Severity Assessment from Facial Image Analysis", 2015.
- [7] Yanmei Chen, Bing Luo, Yen-Lun Chen, Guoyuan Liang and Xinyu Wu, "A Real-time Dynamic Hand Gesture Recognition System Using Kinect Sensor", December 6-9, 2015.
- [8] Asim Jan, Hongying Meng, "Automatic 3D Facial Expression Recognition using Geometric and Textured Feature Fusion", 2015.
- [9] Jinhyuk Kim, Toru Nakamura, Hiroe Kikuchi, Kazuhiro Yoshiuchi, Tsukasa Sasaki, Yoshiharu Yamamoto, "Covariation of Depressive Mood and Spontaneous Physical Activity in Major Depressive Disorder: Toward Continuous Monitoring of Depressive Mood", 2015.
- [10] Antonio Lanata, Gaetano Valenza, Mimma Nardelli, Claudio Gentili, Enzo Pasquale Scilingo, "Complexity Index From a Personalized Wearable Monitoring System for Assessing Remission in Mental Health", 2015.
- [11] Fiza Murtaza, Muhammad Haroon Yousaf, Sergio A. Velastin, "Multi-view Human Action Recognition using Histograms of Oriented Gradients (HOG) Description of Motion History Images (MHIs)", 2015.