PVR System: Personalized Video Recommendation

Mrunal Borgaonkar, Mrunali Katta, Prajakta Kudale, Vaishnavi Deshpande, Prof. Vanita Babanne
RMDSSOE, Warje, Pune- 411058

Abstract- Emotions are an integral part of the human psychology. These emotions are of various kinds. To differentiate these is one of the objectives of our proposed system. The proposed system is based on the recommendation framework of media contents taking into consideration the human emotions. These human emotions captures only the facial gestures of an individual. In existing work, recommendation system is based on user's previous history or log. The recommendation technique used to recommend contents which are not fully personalized or which do not consider users current point of interest or emotion. So there is need of such system which recommends video contents to users. Hence solution to content recommendation is the personalized content recommendation based on emotional characteristics. In correspondence our system considers aspects such as age, gender and emotional characteristics in real time facial expressions. Our approach is to overcome the accuracy problem in existing applications.

Keywords- OpenCV, Machine Learning, Convolutional Neural Network, LBPH

I. INTRODUCTION

Emotions are nothing but a metal and physiological state which are both subjective and private to an individual. Facial expressions play important role in detecting human emotions and current state of mind. It is a form of non-verbal communication technique. And Machine Learning is a domain that has the ability to automatically learn and experiences an improvisation of the system. Machine Learning can detect emotions by capturing facial expressions and learning what each expression means. Also applies that knowledge to new information presented.

The Automatic recommendation is one of those considerable technique. Various media sharing platforms are experimental area which are automatic machines to collect human-like characteristics. Among many non-verbal and involuntary channels through which humans express themselves, facial expressions hold paramount importance. By calculating human facial expressions, we are going to develop emotion based video recommendation system. There is a challenge to recommend emotional based video contents to users. Solution to this is the personalized video recommendation system based on emotional characteristics. To overcome existing history based recommendation technique by providing real time emotion based recommendation system.

The problem to be solved by this work is to present a work with the following characteristics:

1) Take real time face as an input from camera module and using open compute vision library of machine learning and image processing input is processed to recognize current emotion.

2) The system runs using machine learning algorithms such as LBPH, CNN.

3) The system can control changes at run time.

4) For the real time accuracy in the face expression recognition able regulations model creation process must be completely spontaneous. This feature is to make the system absolute alternative to current face expression evaluation techniques, especially where an artistically untrained user is concerned.

II. LITERATURE SURVEY

Abhishek Tripathi et.al [1] states, EmoWare: A Context-Aware Framework for Personalized Video Recommendation Using Affective Video Sequences. In this paper, they harness the potential of the two techniques and propose Emo-Ware (emotion-aware), a personalized, emotionally intelligent video recommendation engine. They proposed a personalized, emotionally intelligent video recommendation engine, employing a novel context-aware collaborative filtering approach, where the intensity of users' spontaneous non-verbal emotional response toward the recommended video is captured through interactions and facial expressions analysis for decision-making and video corpus evolution with real-time feedback streams.

Paul Covington et.al [2] Deep Neural Networks for YouTube Recommendations, describe the system at a high level and focus on the dramatic performance improvements brought by deep learning. In this split according to the classic two-stage information retrieval dichotomy: first, detail a deep candidate generation model and then describe a separate deep ranking model. The candidate generation network takes events from the user's YouTube activity history as input and retrieves a small subset (hundreds) of videos from a large corpus.

Carlos a. Gomez-uribe et.al [3] proposed article discusses the various algorithms that make up the Netflix recommender system, and describes its business purpose. They explain the motivations behind and review the approach that they use to

improve the recommendation algorithms. They are inventing Internet television. Their product and source of revenue is a subscription service that allows members to stream any video in our collection of movies and TV shows at any time on a wide range of Internet-connected devices. As of this writing, we have more than 65 million members who stream more than 100 million hours of movies and TV shows per day.

Shang fei Wang et.al [4] has introduced a general framework for video affective content analysis, which includes video content, emotional descriptors. In this survey current research is in both direct and implicit video affective content analysis, with a focus on direct video affective content analysis. As a result of these developments, video affective content analysis is becoming increasingly important. The goal of video affective content analysis is to automatically tag each video clip by its affective content. Due to the difficulty in defining objective methods to automatically assess the emotions of a video, the research topic of video affective content analysis has not been thoroughly explored until recently.

Daniel McDuff et.al [5] states this subset consists of 242 facial videos (168,359 frames) recorded in real world conditions. They answered three self-report questions about their experience. A subset of viewers additionally gave consent for their data to be shared publicly with other researchers. Web-based framework that was used to crowd source the facial videos and the user experience. Visitors to the website opt-in to watch short videos while their facial expressions are being recorded and analyzed. Immediately following each video, visitors get to see where they smiled and with what intensity. They can compare their "smile track" to the aggregate smile track. On the client side, all that is needed is a browser with Flash support and a webcam.

M. Eckhardt et.al [6] Labeling videos for affect content such as facial expression is tedious and time consuming. Researchers often spend significant amount of time annotating experimental data, or simply lack the time required to label their data. For these reasons we have developed VidL, an open source video labeling system that is able to harness the distributed people-power of the internet.

Markus Schedl et.al [7] The purpose of this trends and survey article is twofold. They first identified and shed light on what they believe are the most pressing challenges MRS research is facing, from both academic and industry perspectives. They reviewed the state of the art towards solving these challenges and discussed its limitations. Second, they detail possible future directions and visions they contemplated for the further evolution of the field.

Oluwatobi Olabiyiet.al [8] The proposed system incorporates camera-based knowledge of the driving environment and the driver themselves, in addition to traditional vehicle dynamics. It then uses a deep bidirectional recurrent neural network (DBRNN) to learn the correlation between sensory inputs and impending driver behavior achieving accurate and high horizon action prediction. The proposed system performs better than other existing systems on driver action prediction tasks and can accurately predict key driver actions including acceleration, braking, lane change and turning at durations of 5 sec before the action is executed by the driver.

Sheena C V et.al [9] States that Key-frame extraction from video data is an active research problem in video object recognition and information retrieval. Key-frame refers to the image frame in the video sequence which is representative and able to reflect the summary of a video content. By using the key-frame it is able to express the main content of video data clearly and reduce the amount of memory needed for video data processing and complexity greatly.

Alexandre Schaefer et.al [10] proposed that using emotional film clips is one of the most popular and effective methods of emotion elicitation. The main goal of the present study was to develop and test the effectiveness of a new and comprehensive set of emotional film excerpts. Fifty film experts were asked to remember specific film scenes that elicited fear, anger, sadness, disgust, amusement, tenderness, as well as emotionally neutral scenes. For each emotion, the 10 most frequently mentioned scenes were selected and cut into film clips. Next, 364 participants viewed the film clips in individual laboratory sessions and rated each film on multiple dimensions. Results showed that the film clips were effective with regard to several criteria such as emotional discreteness, arousal, positive and negative effect.

III. EXISTING SYSTEM APPROACH

In existing approaches for web content recommendation is totally based on users search history. Such recommendation methods fail to estimate current user's point of interest. So there is need to identify user's current mindset and recommends web content accordingly.

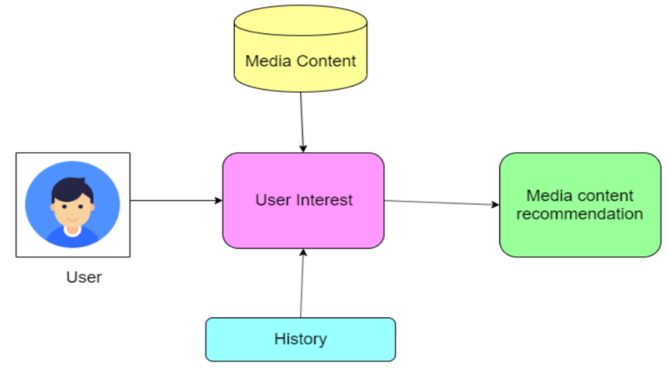

Fig.1: Block Diagram of Existing System

The overall structure of our recommendation system is illustrated in Figure 1. The system is comprised of two neural networks: one for candidate generation and one for ranking. The candidate generation network takes events from the user's YouTube activity history as input and retrieves a small subset (hundreds) of videos from a large corpus. These candidates are intended to be generally relevant to the user with high precision. The candidate generation network only provides broad personalization via collaborative filtering. The similarity between users is expressed in terms of coarse features such as IDs of video watches, search query tokens and demographics.

IV. METHODOLOGY USED

A. **Image Processing: -**
Every image is formation of RGB colours. Each and every captured image has some noise, unwanted background. Thus there is need of process those captured image before assign to our recognition module. Pre-processing unit made is up of noise removal, grey image conversion, binary image conversion of input images after that feature extraction done on those samples. In future extraction five steps applied in which finding the eccentricity. Next elongations of images are evaluated by calculating pixel segmentation as well as rotation of input images.

B. **Tensor-flow: -**
Machine learning is a complex discipline. The implementation in machine learning and creation of models is so much hard and difficult than it used to be, thanks to machine learning technologies and frameworks. Such as Google's Tensor Flow that makes our task simple. It is process of acquiring data, training models, serving predictions, and refining future results.

C. **Convolutional Neural Networks: -**
Brain-inspired systems are to replicate how humans learn. Consist of input, hidden and output layers that transform the input into something that the output layer can use. It is Excellent for finding patterns which are complex for humans to extract and teach the machine to recognize. CNN gathers their knowledge by detecting the patterns and relationships in data and learns (or is trained) through experience, not from programming. CNN takes in processed images as input.

D. **LBPH**: -
Local Binary Pattern Histogram is a popular and simple texture operator which labels the pixels of an image by thresholding the neighbourhood of each pixel and considers the result as a binary number.

E. **OpenCV**: -
Open source computer vision is a library of programming functions which are mainly aimed at real time computer vision.

F. **Machine Learning**: -
Machine Learning which is a part of AI trains the system to learn automatically and experiences to improvise without being explicitly programmed.

V. PROPOSED SYSTEM APPROACH

In a proposed system, we are proposing real time facial expression-based web content recommendation system which overcomes existing content recommendation methods.

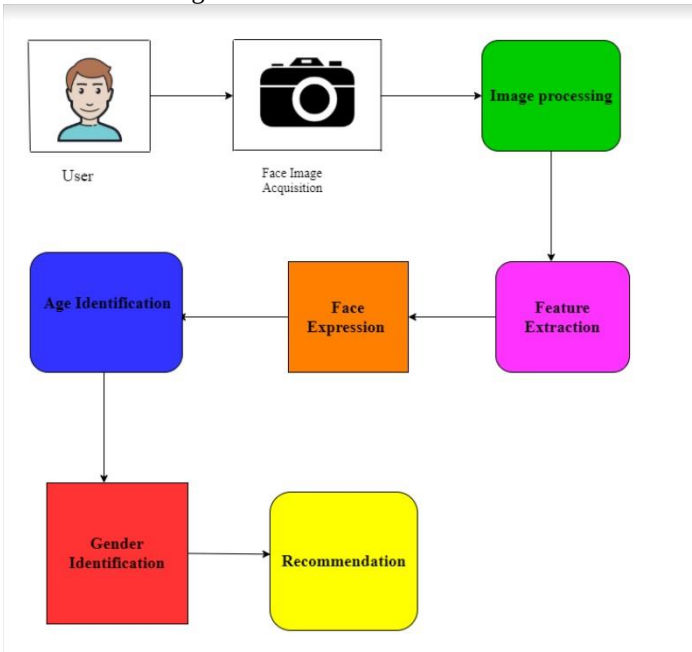

Fig.2: Block Diagram of Proposed System

In a proposed system, we are going to overcome existing drawbacks and provide features such as video acquisition by using open-cv python. Video and image processing done through algorithm and methods. Facial and age estimation done through inception modules created in tensor flow python. Personalized content recommendation using multimedia dataset.
We are going to develop following modules:

1. **Face Image Acquisition: -** Using open computer vision library. We are going to capture real time face images of user. After getting faces we are forwarding these images for feature extraction and image processing.

2. **Image processing: -** After getting face images by using frontal face XML only faces get cropped and further used to process or deform.

3. **Feature Extraction: -** Facial feature extraction is the step of getting face component features like eyes, nose, mouth, etc from real time face images. Facial feature extraction is very much important for the initialization of processing

techniques like face detection, facial expression recognition or face recognition. Among all facial features, eye localization and detection is essential, from which locations of all other facial features are identified.

4. **Face Expression & Age Gender Identification: -** The facial expression evaluation done by using trained inception model of face expression dataset like happy, sad, angry, confused etc. The Gender Prediction as a classification problem. The output layer in the gender prediction network is of type soft max with 2 nodes indicating the two classes "Male" and "Female". Age Prediction should be introducing as a Regression problem since we are expecting a real number as the output. However, estimating age correctly using regression is challenging. Even humans cannot accurately predict the age based on looking at a person.

5. **Automatic Content Recommendation: -** Our main aim is to develop this application that gives most personalized web contents to user based on their current mindset. Which overcomes existing history based content recommendation by our face expression and age-gender based recommendation system.

VI. CONCLUTION AND FUTURE WORK

We are going to develop facial expressions-based system which will be mostly attached to users' current emotions and interest. The proposed personalized content recommendation based on emotional characteristics evaluating with the help of face recognition and machine learning.

In future work taking audio and textual features of videos into account, a general mood-based multimedia content recommendation framework can be created which can monitor users' behavior through body language.

VII. ACKNOWLEDGMENT

This work is supported in a smart surviliance system of any state in india. Authors are thankful to Faculty of Engineering and Technology (FET), Savitribai Phule Pune University, Pune for providing the facility to carry out the research work.

VIII. REFERENCES

[1]. Abhishek Tripathi, "EmoWare: A Context-Aware Framework for Personalized Video Recommendation Using Affective Video Sequences". IEEE VOLUME 7, 2019.

[2]. Paul Covington, "Deep Neural Networks for YouTube Recommendations", RecSys '16 September 15-19, 2016.

[3]. Carlos A. Gomez-Uribe, "The Net-flix Recommender System: Algorithms, Business Value, and Innovation", ACM, December 2015.

[4]. Shangfei Wang, "Video affective content analysis: a survey of state-of-the-art methods", 1949-3045 (c) 2015 IEEE.

[5]. Daniel McDuff, "Affectiva-MIT Facial Expression Dataset (AM-FED): Naturalistic and Spontaneous Facial Expressions Collected In-the-Wild", 29 October 2015.

[6]. M. Eckhardt and R. Picard. "A more effective way to label affective expressions". In Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on, pages 1–2. IEEE, 2009.

[7]. M. Schedl, H. Zamani, C.-W. Chen, Y. Deldjoo, and M. Elahi, ''Current challenges and visions in music recommender systems research,'' Int. J. Multimedia Inf. Retr., vol. 7, no. 2, pp. 95–116, 2018.

[8]. O. Olabiyi, E. Martinson, V. Chintalapudi, and R. Guo. (2017). ''Driver action prediction using deep (bidirectional) recurrent neural network.'' [Online]. Available: https://arxiv.org/abs/1706.02257

[9]. C. V. Sheena and N. K. Narayanan, ''Key-frame extraction by analysis of histograms of video frames using statistical methods,'' Procedia Comput. Sci., vol. 70 pp. 36–40, Jan. 2015.

[10]. A. Schaefer, F. Nils, X. Sanchez, and P. Philippot, ''Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers,'' Cognition Emotion, vol. 24, no. 7, pp. 1153–1172, 2