International Conference on Technologies for Sustainability- Engineering, Information Technology, Management and the Environment
DAV Institute of Management, Faridabad (SUSTECH-15)
28 November 2015 www.conferenceworld.in ISBN: 978-81-931039-7-5

EXTRACTION OF KEY FRAME FROM NEWS VIDEO USING FACE RECOGNITION

Sanjoy Ghatak

Department of Computer Science and Engineering, Sikkim Manipal Institute of Technology, Majitar, Rangpo, East Sikkim, (India)

ABSTRACT

With the advent of the information age and an explosion in the availability of multimedia data, news video production and its usage is growing extremely. But these require more memory and space. So in order to save space and time we need to summarize video. For this we need to perform key frame extraction into the video. This project compares with each other, the combinations of existing approaches of dimensionality reduction and different methods of clustering on face images detected from a TV news broadcast clip to perform face recognition. Key Frame extraction aims at reducing the amount of data that must be examined in order to retrieve a particular piece of information from the large database of information.

The overall process includes frames extraction, removal of non-face areas, projection of the faces into a subspace and various algorithms for the subspace projection and clustering are carried out on it

Index Terms: Face Clustering, Dimensionality Reduction, Kernel PCA, GPLVM, Tsne, Hierarchical Clustering.

I. INTRODUCTION

With the day by day increasing amount of news video, it becomes increasingly difficult to browse and retrieve them for the purpose of selection of appropriate news element. Due to the uncertain length and formats of news video, accessing them still remains a challenge. For video, a common first step is to segment the videos into temporal shots, each representing an event or continuous sequence of actions. A shot represents a sequence of frames captured from a unique and continuous record from a camera. Key frame is the frame which can represent the salient content and information of the shot. The key frames extracted must summarize the characteristics of the video, and the image characteristics of a video can be tracked by all the key frames in time sequence. Key frame extraction technique is also used for abstraction and summarizing. Shots are the building blocks of video. It is defined as a sequence of frames recorded from a single camera. Entire shots can be mapped into a small numbers of representative frames called key-frames. In order to reduce the transfer stress in network and invalid information transmission, the transmission, storage and management techniques of video information become more and more important. Video segmentation and key frame extraction are the bases of video analysis and content-based video retrieval. The use of key frames reduces the amount of data required in video indexing and provides the framework for dealing with the video content. Key frame extraction, is an essential part in video analysis and management, providing a suitable video summarization for video indexing, browsing and retrieval. In this project, performances of different key frame extraction techniques based on

different approaches are analyzed. Eventually the results obtained from these techniques are compared and tried to come up with a better and innovative approach for extraction of key frames from the video.

II. LITERATURE SURVEY

A Wide number of research efforts have been made in the area of key frame extraction which can be grouped into the following categories.

a) Shot boundary based approach: O'Connor et al. used either of the first, middle or the last frame of the shot as the shot's key frame. b) Motion analysis approach: Wolf first computes the optical flow for each frame and then computed a simple motion metric based on the optical flow. c) Visual content based approach. /hang et al. proposed color and motion features independently to extract key frames. Thresholding technique is used to find the similarity between the current frame and the last extracted key frame.

O'Connor's approach have been seen as the most easy way to extract the key frames but it lacks in capturing the visual content of the video shot. Wolf's method was quite appreciable, but computationally expensive due to motion analysis. Zhang's method is relatively fast, but their performance depends on the choice of the threshold by the user. Threshold adjustment proves a serious issue for this method.

2.1 Clustering Technique

There is a Probabilistic Framework of Selecting Effective Key frames for videobrowsing and indexing where a new strategy to extract the most characteristic frame is proposed. The main idea is to cluster similar or redundant views within the shot together. Clusters are approximated by a mixture of Gaussians using standard Expectation Maximization algorithm. Bayes information Criterion is used to choose the appropriate number of clusters. From each obtained clusters, the closest frames to the median of its frame is taken to be a reference key frame. The cluster content is then verified against the reference key frame on the variation of time and appearance. Application of temporal filter is done on the set of all selected frames to remove the overlapping between the constructed set of frames. The most distinguishing work of this work is that the selection of key frames is a fully automated process where no user intervention is required in terms of the input parameters. The temporal variation of color histogram in RGB color space is modeled by the Gaussian mixture density. The Histogram approach has been followed in this work due to its simplicity, speed and robustness. The estimated Gaussian models of tracked objects can be used to recognize similar objects for the whole video.

approach is the combination of shot boundary detection and an intra-shot clustering of frames to find an adequate number of representative key Frames for the given shot with respect to its visual complexity. The researchers used the M1'EG-7 Color Layout Descriptor (CID) as a feature for each frame and computed the differences between consecutive frames. Adaptive Threshold Technique is used to detect shot boundaries. The nearest frame to the mean of every cluster is used as the representative for the shot after clustering. Bayesian Information Criteria (BIC) is used to 'mate the number of clusters. The shortcoming of Shot boundary detection and Intra-shot clustering was found to be the redundancy of the same content by multiple key frames due to the missing inter-shot reasoning. This redundancy occurs when the same artist is shown indifferent shots throughout the video. The researchers directly used the already extracted key frames by previously existing methods and clustered them using the k-means algorithm.

2.2 An Approach to Extract Key Frames Using K-Means Algorithm

Another one is Key-frame Extraction for video tagging and Summarization [8] where unsupervised learning for video retrieval and summarization technique was proposed. The approach uses shot boundary detection to segment the video into shots and the K-means algorithm to determine cluster representative for each shot that are used as key-frames. The new approach is the combination of shot boundary detection and an intra-shot clustering of frames to find an adequate number of representative key Frames for the given shot with respect to its visual complexity. The researchers used the MPEG-7 Color Layout Descriptor (CID) as a feature for each frame and computed the differences between consecutive frames. Adaptive Threshold Technique is used to detect shot boundaries. The nearest frame to the mean of every cluster is used as the representative for the shot after clustering. Bayesian Information Criteria (BIC) is used to 'mate the number of clusters. The shortcoming of Shot boundary detection and Intra-shot clustering was found to be the redundancy of the same content by multiple key frames due to the missing inter-shot reasoning. This redundancy occurs when the same artist is shown indifferent shots throughout the video. The researchers directly used the already extracted key frames by previously existing methods and clustered them using the k-means algorithm.

2.3 Shot Reconstruction Degree Interpolation (SRDI) Algorithm

There is a Shot Reconstruction Degree Interpolation (SRDI) algorithm of TieyanLiu [15] where motion vector are used to compute the frame's motion energy. All the motion used to build a motion curve that is passed to a polygon simplification algorithm. This algorithm retains only the most salient points that can approximate the whole curve. The frames corresponding to these points form the key frame set. If the number of frames in the final set differs from the number of key frames requested, the set is reduced or increased by interpolating frames according to the Shot Reconstruction Degree criteria. When the number of frames is lower than the number desired, the shot is reconstructed by interpolating the frames in the frame set, and the interpolated frames that have largest reconstruction errors are retained up to the number of key frames needed. When the number of frames in the frame set is greater than the number of key frames needed, the frames in the frame set are interpolated, and those with the minimal reconstruction error are removed from the set.

2.4 An Optimized Key Frame Extraction Scheme Based on Frames Extracted

In this method by Ntalianis, KlimisS an optimized and efficient technique for key frames extraction of video sequences is proposed, which leads to selection of a meaningful set of video frames for each given shot. Initially for each frame, the singular value decomposition method is applied and a diagonal matrix is produced, containing the singular values of the frame. Afterwards, a feature vector is created for each frame, by gathering the respective singular values. Next, all feature vectors of the shot are collected to form the feature vectors basin of this shot. Finally, a genetic algorithm approach is proposed and applied to the vectors basin, for locating frames of minimally correlated feature vectors, which are selected as key frames. Experimental results indicate the promising performance of the proposed scheme on real life video.

2.5 Extraction of Key Frames Based on a Triangle Model of Perceived Motion Energy (PME)

Researchers Tianming Liu, Hon-Jiang 'Mang and Feihu Qi proposed a triangle model of perceived motion energy (PME) [14] to model motion patterns in video and a scheme to extract key frames based on this model.

The suggested key frame extraction process is threshold free and fast since the motion information in MPEG can be directly utilized in motion analysis, while the key frames are representative. The approach combines motion based temporal segmentation and color based shot detection. The turning point of motion acceleration and deceleration of each motion pattern is selected as key frame.

2.6 A Novel Inflexion Based Key Frame Selection Algorithm Using Voila Jones

Researchers Tieyan Liu,)(Wong Zhang, JianFeng, Kwok-Tung Lo recommended Shot reconstruction degree (SRD) as a novel criterion for key frame selection based on the degree of retaining motion dynamics of a video shot. According to them, the key frame set produced by SRD can capture the detailed dynamics of the shot in a better way as compared to the widely used Fidelity criterion. Using the new SRD criterion, a novel inflexion based key frame selection algorithm is developed. More emphasis is laid on the local detail and the evolution trend of a video shot as the new measure for key frame selection. The basic idea is said to be that if the shot reconstructed by interpolating the key frames can approximate the original shot well, then it can be said that the key frame can capture the detailed dynamics of the shot. Regarding this view, it can he said that the rectangular points are better than circular ones, and such criteria is suggested as SRD. According to the researchers, if all local features of the shot arc approximated well, then the global features can also be approximated well. Hence, the key frames which can reconstruct the shot well will also lead to a high fidelity. Based on the SRD criteria, a novel inflexion-based key Frame selection algorithm has been developed which led to high performance in terms of both fidelity and Shot Reconstruction degree. Future work in this field demands more human perceptions for the video summarization techniques.

III. PROBLEM DEFINITION

There are many information carriers in a video stream, as is the visual content, the narrative or speech part, possible text captions etc. Visual content remains the most important and the most difficult one to tackle as well. Simply put, video usually contains an enormous amount of visual information, since it could be stream off large number of frames every second with a high resolution and a high color depth. Now, this poses a very important consideration since it is better to keep a representative frames or key frames from a long scene with little or no change, instead of a few hundred or thousands of frames. The problem is to remove the visual content redundancy among news video frames or in other words to extract the key frames from a news video sequence and to obtain summarized output video using the key frames.

The growing interest of consumers in the acquisition of and access of visual information has created a demand for new technologies to represent model index and retrieve multimedia data. Very large database of news videos require efficient key frame extraction algorithms that enable fast browsing and access to information. Key frame extraction aim that reducing the amount of data that must be examined in order to retrieve particular piece of information from the database of information, is consequently, an essential task in video analysis and indexing. The first task is to divide the input news video shot into multiple sequential video frames. Each multiple video frames are individually analyzed for certain descriptors. Based on the descriptors, key frames arc identified from the sequential frames. As a result the non-significant frames or the redundant frames get eliminated from the

result and to obtain a summarized output video using the key frames. The final step will be the comparison of our technique with the previously existing technique.

The problem at hand can be divided into the following basic parts:

• Segmentation of the video shot into sequential frames.

• Analyzing each individual frames using different frame descriptors.

• Recognizing facial attributes from each individual frames.

• Extraction key frames from the recognized facial frames.

• Conversion of extracted key frames into a summarized output.

IV. PROPOSED SOLUTION STRATEGY

1. The strategy adopted by us would be the first to convert RGB video to gray scale video.
2. Then we would fetch out gray scale frames from the above converted video.
3. After performing the above task we use algorithm to recognize the facial attributes from the gray scale frames.
4. When the facial attributes are recognized then we extract those which act as key frames for us.
5. In our final step we use these extracted key frames and cluster them to form our reconstructed or summarized video.

V. DESIGN STRATEGY FOR SOLUTION

At the Very first Phase, a video is taken as input. Thevideois sub-part of the main video on which the summarization is targeted at. Generally, a video summary is a sequence of still or moving images with or without audio. These images must preserve the overall contents of the video with a minimum of data. The focus is mainly done on the creation of a visual summary using still images or key frames. At the immediatenext phase, the input Video is disintegrated into sequential frames using code. From the sequential frames key frames are obtained using different techniques. The next Phase is the process of feature extraction from the obtained sequential frames. The extraction process is automatic and content based so that they maintain the salient content of the video while avoiding all types of visual content redundancy. Using the extracted features from the individual frames.

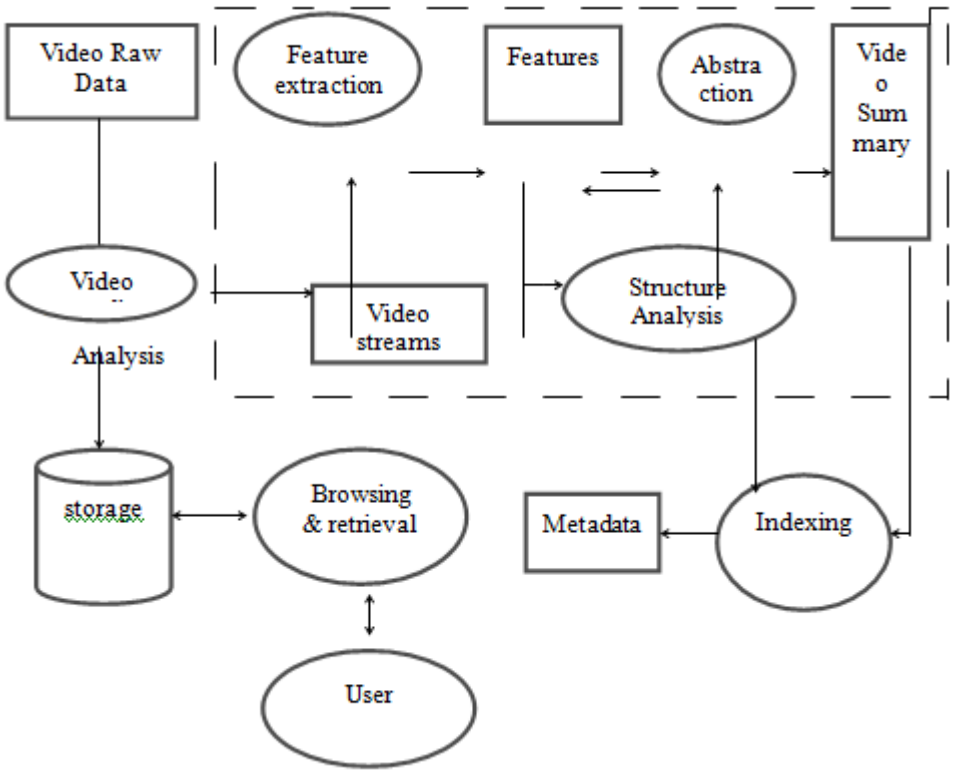

Fig 1: Block Diagram of General Video Processing.

Video summarization aims at reducing the amount of data that must be examined in order to retrieve particular piece of information in a video. The key frames when arranged in a sequential way are able to reflect the basic content of the input video shot. Key frames can Summarize the video content in more rapid and Compact way, users can grasp the overall content morequickly from key frames then by watching a set of video sequences, Key frames, which visually represent the Video content, can also be used in the indexing process, where we can apply the same indexing and retrieval Strategies developed for image retrieval to retrieve video sequences. Low level visual features can be used in indexing the key frames and thus the video sequences to which they belong. The key frames provide good bookmarking that can designate the‘key ‘contentof the input video.

Key frames which visually represent the video content can also be used in the indexing process, where we can apply the same indexing and retrieval strategies developed for image retrieval to retrieve video sequences. Low level visual features can be used in indexing the key frames and thus the video sequences to which they belong. Key frames often provide good bookmarking that can designate the ‘key’ contents of the input video. The key frames provide an abstract level of insight into the basic content of the original input video shot. These frames preserve the overall features of the video while removing the visually redundant frames such that the viewer is not awareof redundancy and is able to grasp the main feature and content of the original

video shot. The video obtained after summarization is essentially of lesser size as compared to the input video. Key frames, which visually represent the video content, can also be used in the indexing process, where we can apply the same indexing and retrieval strategies developed for image retrieval to retrieve video sequences. Key frames are truly content based which are dependent on the input video shot. The removal of redundant frames results in the video with the key frames.

VI. IMPLEMENTATION

For implementation, Mat labis used as a platform. At the initial stage, video is taken as an input to extract all the sequential frames that constitutes the whole video then we convert RGB video to gray scale video after performing this we would fetch the gray scale frames from the above converted Gray scale video. After this we use algorithm to recognize the facial attributes from the gray scale frames and then the facial attributes are recognized then the content are extracted in form of the key frames .In our final step we use these extracted key frames and cluster them to form our reconstructed or summarized video.

6.1 Extraction of All the Frames in the Video

A video or movie frame is a single picture or still shot, that is shown as part of a larger video or movie. Many single pictures are run in succession to produce what happens to be a seamless piece of film or videotape. Each frame can be selected on its own to print out a single photograph.

Algorithm for reading the video shots and to extract the frames:

Start

Step1: Read the input video shot

Step2:Determine the no. of frames in the video

Step3:Display the no of frames

Step4: Convert value ofnumber into string

Step5: Display the frames in sequential manner.

In this case, first input video shot should be read for calculate the no. of frames from video. Then no of frames are displays and convert the value of number into string. After this frames are display in sequential manner.

6.2 Extraction of key Frames

Key Frames are used to represent the contents of a video. Choosing an appropriate number of key frames from video is difficult since it contains both action and non-action scenes. Selecting one key frame may represent a static video;however, a dynamic video may not be represented adequately. Therefore, a method is developed to select variable number of key frames depending upon the video activity. Each shot can be represented by a set of key frames.

6.3 Conversion of Frames into Gray Scale Frames

After Extraction of key frames they are further converted into Gray scale which then used to recognition of facial attributes from the gray scale frames.

Steps for conversion of frames to gray scale:

Step1: Identify the total number of frames.

Step2: Convert frames from rgb to gray scale.

6.4 Detection of Facial Attributes from Key Frames

After performing above algorithms we need to recognize the facial attributes from the gray scale frames. When the facial attributes are recognized then we extract those which act as key frames for implementation of the project.

 ISBN: 978-81-931039-7-5

Steps for Face Recognition

Step1: Identifying the total number of frames (0 to 99)

Step2: Detection of face from video frames.

Step3: Identify the position of face.

Step4: Write this in file.

Reconstruction of video from the extracted key frames:

Step 1: Calculate number of key frames consisting of a faces.

Step2: Taking the entirekey frame from base file, reconstruct the video.

Step3: Save summarizes video in a file.

VII. RESULTS AND DISCUSSIONS

The extraction of the key-frames is done in a totally automatic way without requiring that the user specifying the number of key frames to be extracted as a parameter. Also it is flexible enough to extract the variable number of key frames. Key frames or Representative frames that are obtained are sufficient enough to represent the original shot with little or no change. Relatively lower numbers of key frames are extracted as compared to the consecutive frames obtained after disintegrating the video shot. The obtained video is able to represent the input video shot in a short and concise manner. In this method, the visually redundant frames were removed which have been termed as the non-significant frames. Variable numbers of key frames are generated depending on the size of the input video. By the inclusion of face recognition algorithm, the obtained results generated lesser number of key frames yet they are able to reflect the significant properties of the input video frames. The important part is that the results are entirely based on the visual details of the input video shot without taking into consideration the audio part.

7.1 Preliminary Sequential Frame Extraction

Input: News video

Output: Sequential frames that constitute the whole video

The sequential frames which are obtained from the input news video shot have been shown below. This is the very first step, where the video shot is disintegrated into its constituent element which is known as frames. In the later phases, these frames are further processed for the identification of key frames.Fig. (4) Shows the all the key frames of video. Using these key frames we regenerated the summarized video.

Fig2: Video to RGB conversion.

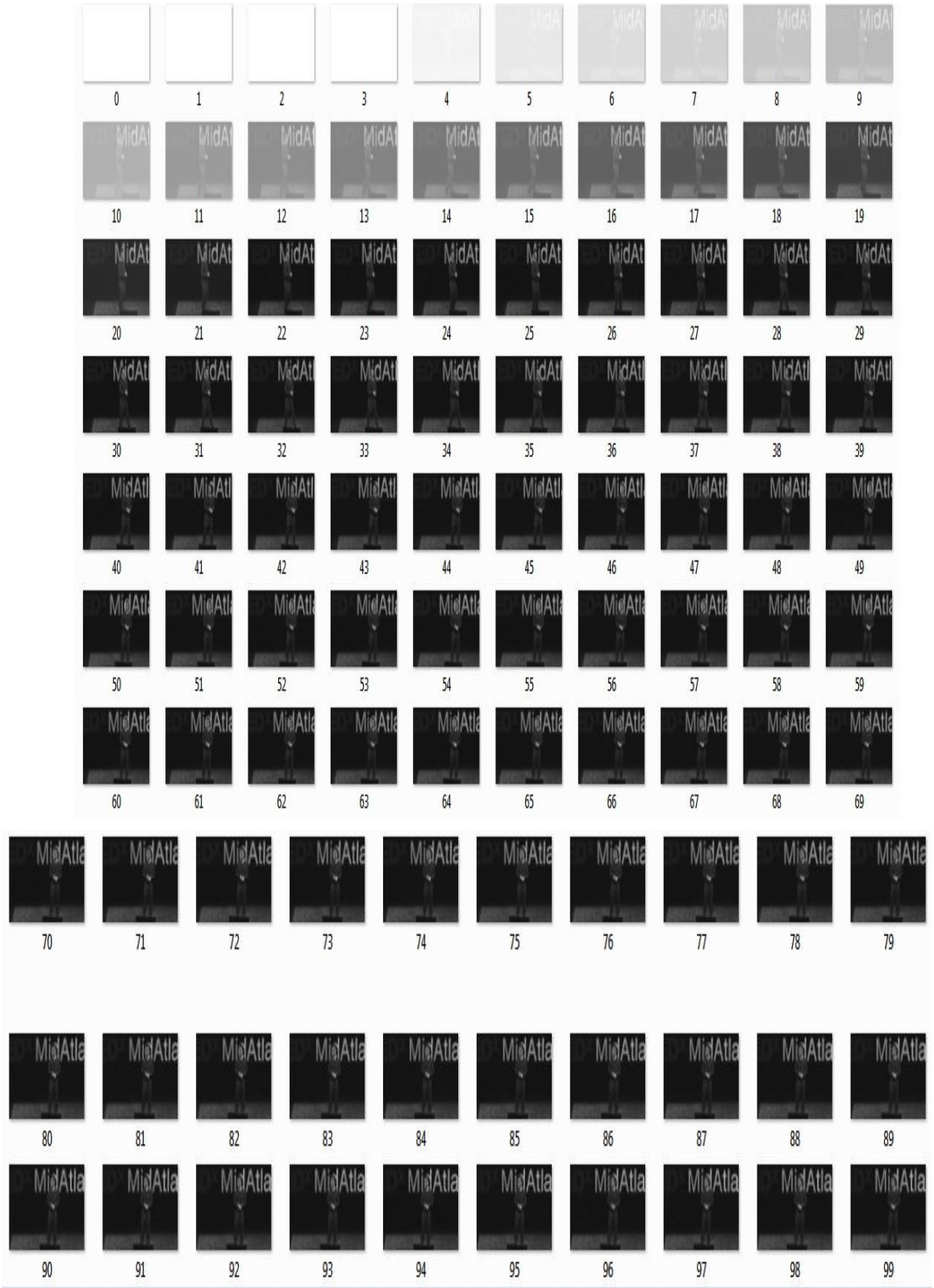

Fig3: RGB frame to GRAY scale conversion.

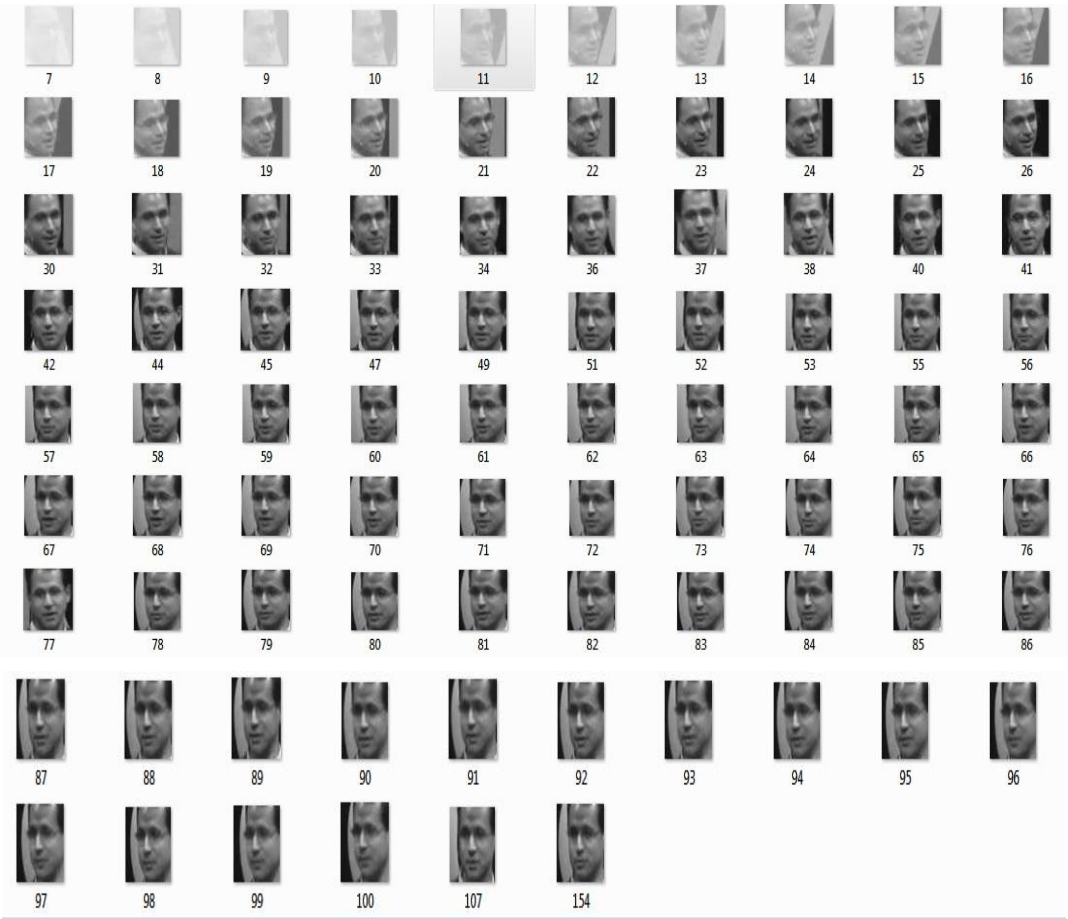

Fig4: Key frames of Videos.

VIII. CONCLUSION

Video Shot is segmented into a number of consecutive (or sequential) frames automatically without any problems. Properties of each individual frame are determined using the frame descriptors which are namely Color histogram, edge direction histogram and wavelet statistics, each was applied individually for ease and accuracy. Performance of the proposed method is matched with the existing algorithms. Since in the earlier works, only one or two frame descriptor(s) was/were used, the performance was not up to the mark. Hence the results obtained from the proposed work are better due to the inclusion of three frame descriptors and combining their results in the final phase to get the output. Variable numbers of key frames are extracted for different input video shot, which indicates the dynamic nature of' the code. The code may be used in future for different purposes. Key frames or the representative frames obtained are enough to represent the whole video. The main features of' the video shot are grabbed efficiently in the extracted key frames which when summarized can represent the whole video in a concise manner. Relatively lower number of key frames are extracted which reduces the size of the summarized video. The main purpose of the proposed work is fulfilled since the numbers of the extracted key frames are far lesser than the number of sequential frames obtained after the video is broken down initially.

• The proposed method basically approaches towards the extraction of key frames from the sequential frames which are obtained after disintegrating the input video shot. Based on the technique of the extraction of frames,

number of extracted key frames can vary for the same input video frame. Since the key frames are used for the summarization of the input video. So, by using different techniques and distance measures, the output video can be compressed to a variable level and it may miss out a particularly important content feature of the input video. So it is a loss full video compression technique. The work is mainly focused only on the visual content of the input video. It only focuses on the compression of the video shot without taking into account the audio part. To integrate the audio into the summarized video, different software is required which is not included in the work. The project is confined to the functionalities available in Mat lab, where the audio part is out of scope.

The future scope of this approach is very high and the improvements can be done in future. Since the method is technical based, the output performance depends purely on the developers based on their selection of distance (or difference) measure. In the future, new distance measures may be combined with appropriate frame descriptors to get a better output than the present one. The output success rate is determined by evaluating on how efficiently the extracted key frames are able to produce the detailed content of the input video shot when all of them are combined to form the video. While at the same time, emphasis is to be laid upon the minimizing the number of key frames such that the summarized output video is of small size. The audio is not embedded and only the visual details are considered in the current project. In the future, the developers may take the project to a new level by incorporating the audio feature as well.The size of any video shot can he reduced to a significant level by the key frame extraction feature, but it is a loss full compression technique. The summarized video is able to highlight the key contents of the original video shot. The number of key frames extracted is considerably less, as compared to the sequential frames that were obtained initially which led to the compression of the final video. Depending on the input video frame, variable numbers of key frames are extracted. The most significant observation is that the work does not take into consideration the audio pan that is available in the input video shot.

REFERENCES

[1] Extraction Of Key Frames from News Video Using EDF,MDF method for news video summarization, SANJOY GHATAK, DEBOTOSH BHATTACHARJEE.

[2] Hanjalic A., Lagendijk R. L., Biemond J. A new Method for Key Frame Based Video Content Representation. In: Image Databases and Multimedia Search, World Scientific Singapore, 1998.

[3] Y.Zhuang, Y.Rui. T.S. Huang, and S.Mehrotra. Adaptive key frame extraction using unservised clustering .In Proc. Of IEEE conf. on Image Processing, pages 866 870. Chicago II, October 1998.

[4] Y.S. Avrithis,A.D. Doulamis,N.D. Doulamis, and S.D. Kollias. A stochastic framework for optimal key frame extraction from mpeg video databases. Computer Vision and Image Understanding,75(1/2):3 24.July-August 1999.

[5] JankoCalic and Ebroullzquierdo.Efficient Key-Frame Extraction and Video Ananlysis.Multimedia and Vision Reasearch Lab, Queen Mary, University of London.

[6] Hoon S. H, Yoon K, and KNNeon 1. A new Technique for Shot Detection and Key Frames

[7] R. Gonzalez and R. Woods, Digital Image Processing, Addison Wesley