Modern Education
and Computer Science
PRESS

Dynamic Summarization of Video Using Minimum Edge Weight Matching in Bipartite Graphs

Shanmukhappa Angadi[1]
[1] Department of Computer Science and Engineering,Centre for Post Graduate Studies, VTU. Belagavi, India
Email: vinay_angadi@yahoo.com

Vilas Naik[2]
[2] Department of Computer science and Engineering Basaveshwar Engineering College, Bagalkot, India.
Email: vilasnaik_h@rediffmail.com

Abstract—To select the long-running videos from online archives and other collections, the users would like to browse, or skim through quickly to get a hint on the semantic content of the videos. Video summarization addresses this problem by providing a short video summary of a full-length video. An ideal video summary would include all the important segments of the video and remain short in length. The problem of summarization is extremely challenging and has been a widely pursued subject of recent research. There are many algorithms presented in literature for video summarization and they represent visual information of video in concise form. Dynamic summaries are constructed with collection of key frames or some smaller segments extracted from video and is presented in the form of small video clip. This paper describes an algorithm for constructing the dynamic summary of a video by modeling every 40 consecutive frames of video as a bipartite graph. The method considers every 20 consecutive frames from video as one set and next 20 consecutive frames as second set of bipartite graph nodes with frames of the video representing nodes of the graph and edges connecting nodes denoting the relation between frames and edge weight depicting the mutual information between frames. Then the minimum edge weight maximal matching in every bipartite graph (a set of pair wise non-adjacent edges) is found using Hungarian method. The frames from the matchings which are represented by the nodes connected by the edges with weight below some empirically defined threshold and two neighbor frames are taken as representative frames to construct the summary. The results of the experiments conducted on data set containing sports videos taken from YOUTUBE and videos of TRECVID MED 2011 dataset have demonstrated the satisfactory average values of performance parameters, namely *Informativeness* value of 94 % and *Satisfaction* value of 92 %. These values and duration (MSD) of summaries reveal that the summaries constructed are significantly concise and highly informative and provide highly acceptable dynamic summary of the videos.

I. INTRODUCTION

Advances in multimedia data acquisition and storage technologies have led to very rapid growth of large databases. Since it is time consuming to download and browse a long video, lot of research attention has been focused on effective video summarization in order to help users to grasp the video content quickly. Thus automatic video summarization is a technique defined as selection and representation of still images or video segments to convey complete content of the video in concise form. Depending on how this concise form is constructed and presented the summaries are classified as static and dynamic. The static summaries are constructed as group of key frames extracted from video and presented in the form of a story board**.** The dynamic summaries are the concatenation of sufficient number of representative frames or representative segments extracted from video.

Video summarization, as the name implies, is creation of a short summary of the content of a longer video document. A video abstract is a sequence of still (static video summary) or moving images (dynamic video summary) representing the content of a video in such a way that the target group is rapidly provided with concise information about the content while the essential message of the original is well preserved.

Dynamic video summary, is also called moving-image abstract, moving storyboard, summary sequence or video skim. This type of abstract consists of a collection of represetative frames or video segments extracted from the original video and converted into a video clip, but of significantly shorter duration.

There are basically two types of dynamic summaries: summary sequence and highlight. A summary sequence renders the impressions of the content of entire video by using various techniques. On the other hand, a highlight shows the most interesting parts of an original video.

Video summarization aims to provide a brief introduction of the original video to users by removing redundant content and keeping essential content in compact form. The task of identifying essential content to make summary extremely informative is a major challenge in video summarization research.

A preview sequence (summary sequence) is made with the objective of reducing a long video into a short sequence that is often used to help the user to determine if a video program is worth viewing in its entirety. It either provides an impression about the entire video content or contains only the most interesting video segments. Key frames are most suitable for content-based video browsing, where they can be used to guide a user to locate specific video content of interest. Furthermore, key frames are also effective in representing visual content of a preview sequence. In the case of preview sequences, key frames can also be extracted in two different ways, either to capture the most memorable scenes of a video referred as the *highlighting*, or to summarize the entire video in terms of its visual content known as *summarizing* key frames .There are various algorithms presented to extract key frames from videos to construct summary sequence based on various features like color, texture, motion, audio, text and speech. The various models like clustering model, HMM model and Graph models have been used to represent structure of the video and to extract key frames from video structure representation. The graph theory techniques that are employed and bipartite graph modeling for video representation are briefly described in the following.

The graph model based methods are motivated by graph theoretical techniques. The graph theory techniques like graph cut, graph optimization, graph clustering, graph matching and graph similarity are successfully employed in to video summarization and retrieval. The work proposed is a key frame based approach motivated by the capability of graph theoretical techniques to represent the video structure. The proposed approach is based on bipartite graph representation of the video and minimum edge weight maximal matching for dynamic video summarization.

In the proposed method every 20 frames of the video are collected as one side of the bipartite graph the next consecutive 20 frames form the other part of the bipartite graph. Each frame is represented by a node in the graph and the edges of the bipartite graph are edges from the first set of 20 nodes to the second set. These edges represent the mutual information between corresponding frames. The minimum edge weight maximal matching of the bipartite graph is found and the nodes of the edges in the matching with weight values below a empirically fixed threshold represent the key frames. The process is repeated for other frames of the video by taking the second set of 20 frames to form the first(left) side of the next bipartite graph and further 20 frames form the other side. The process is repeated over complete video and representative frames selected are concatenated to build a moving image clip to form dynamic summary sequence. The values of informativeness(94%) and satisfaction

(92%) is found in experiments on videos from YOUTUBE and TRECVID MED 2011 data set reveal the suitability of proposed algorithm for summarization.

The remaining part of the paper is organized in to 4 sections. Section II presents the related work and background of the algorithms used. Section III describes the proposed bipartite graph matching based algorithm. In Section IV experimentation and results are presented. Section V brings up the conclusion.

II. RELATED WORK

The detailed literature review on video summarization approaches is presented in this section. Basically there are two main approaches for video synopsis or video abstraction. The first approach is key frame based approach in which the salient images (key frames) are selected from the original video sequence. In the second approach a collection of sufficient number of representative frames or short sequences are selected from original video to build an abstract in the form of a compact video clip. A comprehensive survey on video abstraction is presented in [3]. To bring cartoon and comic like dynamics in summary a video summarization algorithm is presented in [4]. Histogram and transform based summarization are described in [5] and a DWT and stastical feature based summarization method is presented .

A dynamic video abstraction approach based on information theory with minimal visual content redundancies is presented in [6]. The method first clustered the key frames and then concatenated short video segments of key frames to construct a dynamic video summary. A video summarization approach to produce highlights of sports video employing audio cues is found in [7].

Generally, for video summarization based on video content there is a need to effectively eliminate redundancy to provide a preview of video in compact form by extracting representative frames as key frames of segments. In [8], a content-adaptive video summarization with key frame selection is presented. A scenario based dynamic video abstraction using a graph matching algorithm is proposed in [9]. A video summarization algorithm based on video clips comparison is provided in [10]. In this method, video clip similarity is measured using graph matching algorithms for maximal matching and optimal matching.

In [11] a novel approach for video summarization based on graph optimization is presented. The approach emphasizes both a comprehensive visual-temporal content coverage and visual coherence of the video summary. The approach works by selecting a candidate shot from shots of video and models it as directional graph. A dynamic programming based approach is used to find longest path which will represent final summary. The work in [12] proposes a new approach for event detection and summarization of news videos. The approach is mainly based on two graph algorithms optimal matching (OM) and normalized cut (NC).

Initially, OM is employed to measure the visual similarity between all pairs of events. under the one-to-one mapping constraint among video shots. Then, events are represented as a complete weighted graph and NC is carried out to globally and optimally partition the graph into event clusters. Finally, based on the cluster size and globality of events, significant events are automatically detected

The authors of [13] present a method for summarizing multi-view videos by constructing a spatio-temporal shot graph and to formulate the summarization problem as a graph labeling task. The spatio-temporal shot graph is derived from a hypergraph. The shot graph identifies clusters of event-centered shots, and picks out the candidates for summarization by random walks.

The work reported in [14] is an approach for video summarization based on the analysis of the video structure. The method uses FCM algorithm and a minimum spanning tree to cluster shots. Then the algorithm works on detection of video shots and the selection of key frames from each shot.

In [15]an algorithm for video summarization with a two-level redundancy detection procedure is presented. The algorithm removes redundant video content using hierarchical agglomerative clustering in the key frame level. A part of key frames are selected to generate the initial video summary. Finally, a repetitive frame segment detection procedure is designed to remove redundant information in the initial video summary.

A video can be represented as a graph by representing frames or shots of video as vertices and edges denoting relation between frames or shots. Exploiting this some graph theoretical approaches for shot boundary detection are found in literature. A graph theoretical technique based on *normalized cuts* reported in[16] is a computational technique to obtain optimized graph partitions and key frames are selected from each partition for story board construction. Another video shot boundary detection algorithm based on the novel graph theoretic concept, namely dominant sets is presented in [17].

Graphs can be excellent modeling tools which are used to model many type of relations amongst various data objects or patterns that are part of real world scenarios. Graphs are mathematical structures used to model pair wise relations between objects. Graph theory has been developed into a useful tool in solving computing problems as many types of data can be represented as a graph. In this representation, vertices correspond to different observations, or data points and edges represent connections, similarities, or correlations among those points. The solutions for many classical problems in computer science are defined using various concepts of graph theory. As the video is collection of large number of visual patterns (frames) and audio patterns, the frames or shots can be expressed as nodes and their interrelation as edges of the graph. The procedures like graph clustering, graph partitioning, graph cuts can help in clustering patterns with maximum connectivity. The representation of video frames in the form of clusters and

representation in groups by graph representation and graph theoretical processes will help for extracting representative frames from each cluster or group to build the summary. The proposed work is benefited by the capacity of bipartite graph to represent two set of data elements as two vertex sets and each edge from every node in one set to every node in second set as relation between data elements represented as nodes. The bipartite graph is used to model the intergroup similarity relationships and maximal matching in bipartite graphs is used to measure the pair wise similarity between the frames of video in the two sets shown as nodes of bipartite graph. The content similarity between any two frames is represented as mutual information in terms of edge weights in bipartite graph. Lesser the mutual information, more dissimilar the frames and hence are better candidates for key frames.The detailed description of the methodology is given in subsequent sections.

III. Bipartite Graph Model Based Algorithm For Dynamic Video Summarization

A bipartite graph is a connected undirected graph such that the vertices of G are partitioned into two sets U and V and every edge of G has one end point in U and the other in V. A matching M in G is a set of edges that have no end points in common [18].
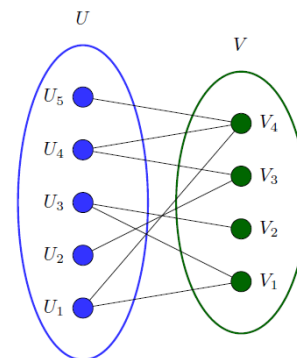

Fig.1. A typical Bipartite graph

In the proposed algorithm the video is considered in groups of 40 frames with overlap of 20 frames and each group is modeled as a bipartite graph with 20 frames each in two disjoint sets, by representing video frames as vertices of bipartite graph and edges with weights equal to mutual information between two frames that the edge connects. A minimum edge weight maximal matching is found for every bipartite graph model. From the edges (links) of *matching* M, nodes connected by links of edge weight below 70% of maximum edge weight in terms of mutual information are identified and frames represented by these nodes are extracted as key frames. The higher values of mutual information between two frames indicate that both the frames have more similar content. From the 40 frames which are considered for bipartite graph frames with nearly similar content i.e. with higher value of mutual information are to be eliminated to avoid redundant frames. At the same time the less similar

frames arranged on timeline are added to a file which when played back at high speed (25-30 frames/second) each still picture merges into the next one so they blur together to make single moving picture. From the experimentations the frames with mutual information below 70%(an empirically selected value) of maximum edge weight from a bipartite graph will retain minimum

amount of similar content temporally and construct a moving picture and at the same time attain sufficient temporal compression. The steps of complete algorithm are described in Fig. 2. The detailed description of bipartite graph construction, minimum edge weight graph matching algorithm and key frame extraction are given in following subsections.
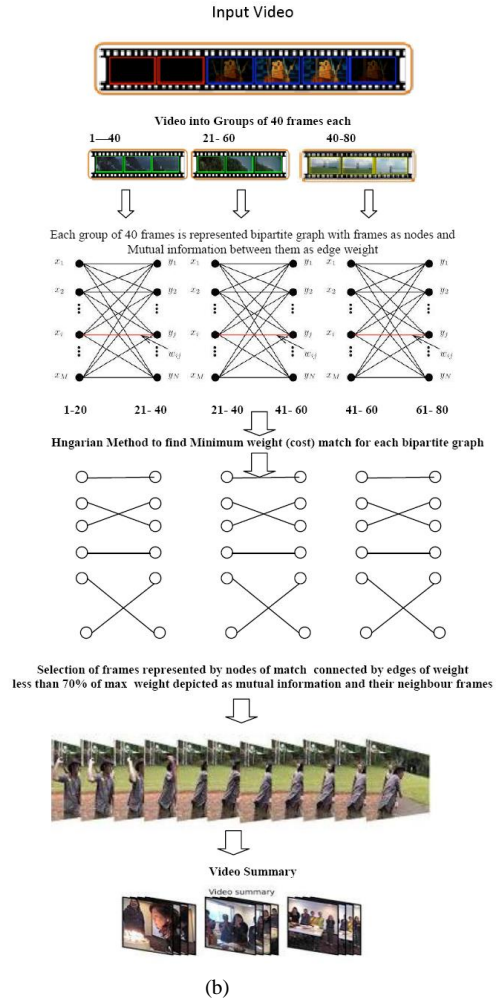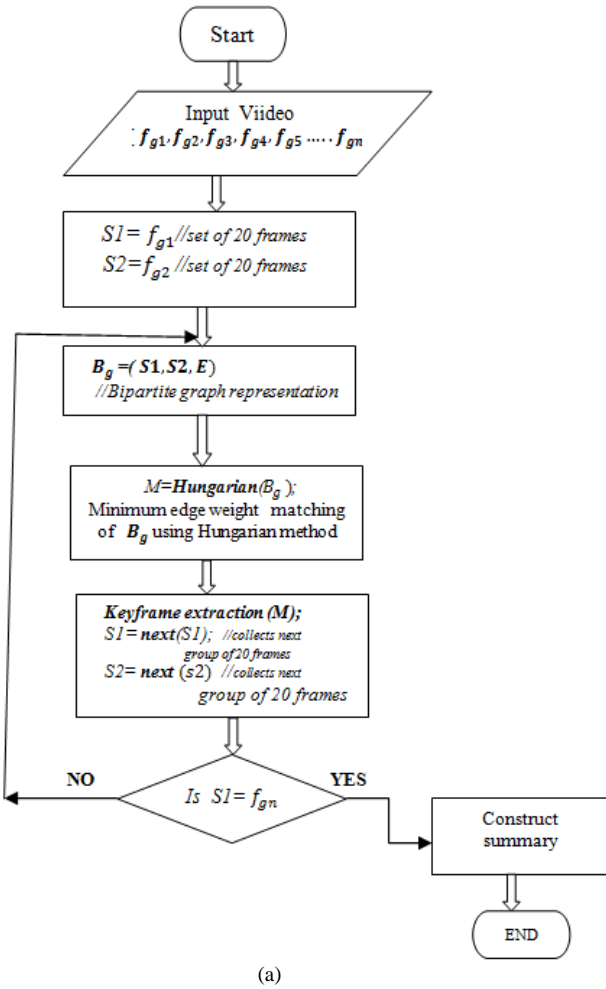

Fig.2. (a) Flow chart of proposed technique (b) The proposed Video summarization technique

A. Bipartite Graph Modeling

Let G = (U, V, E) denote a bipartite graph with n vertices and m edges. A subset M⊆E of edges is a matching if no two edges in M share an endpoint. The problem of finding maximum matching in bipartite graphs is a classical problem in combinatorial optimization. In this work a bipartite graph model is proposed to represent the correlation between two group of frames, by representing frames of each group as nodes of separate side of bipartite graph as shown in Fig.3.The links stand for the interrelation between frames of two groups and are quantified in terms of mutual information.

In proposed solution, the mutual information between two frames is calculated separately for each of the RGB color components. In the case of the R component, the element $C_{t,t+1}^{R}(i,j)$, $0 \leq i, j \leq N-1$ (N being the number

of intensity levels in the image), corresponds to the probability that a pixel with color intensity i in frame f_t has color intensity j in frame f_{t+1}.

In other words, $C_{t,t+1}^{R}(i,j)$ equals to the number of pixels which change from color intensity i in frame f_t to color intensity j in frame f_{t+1}, divided by the total number of pixels in the video frame. The mutual information $I_{k,l}^{R}$ of frames f_k, f_l for the R component is computed as equation (1)

$$I_{k,l}^{R} = -\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} C_{k,l}^{R}(i,j) \log \frac{C_{k,l}^{R}(i,j)}{C_{k,}^{R}(i)\, C_{l}^{R}(j)} \quad (1)$$

where

$C_{k,l}^R$ *(i. j): Joint Probability that a pixel with intensity level i in frame* f_k *has intensity level j in frame* f_l *for R Component*

The mutual information $I_{k,l}^R$ is the relative entropy between the joint distribution $C_{k,l}^R$ *(i. j):* and the product distribution C_k^R (i) C_l^R (j).

Similarly the mutual information for Blue (I^B) and Green (I^G) components can be found as in (2)

$$
\left.\begin{aligned}
I_{k,l=}^G \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} C_{k,l}^G\ (i,j)\log\ \frac{C_{k,l}^G\ (i,j)}{C_{k,}^G\ (i)\ C_l^G\ (j)} \\
I_{k,l=}^B \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} C_{k,l(i,j)}^B\log\ \frac{C_{k,l(i,j)}^B}{C_{k,(i)}^B\ C_{l(j)}^B}
\end{aligned}\right\} \quad (2)
$$

The total mutual information (MI) calculated between frames f_k and f_l is defined as in equation (3)

$$I(f_k, f_l) = I_{k,l}^R + I_{k,l}^G + I_{k,l}^B. \tag{3}$$

The mutual information computation is based on the content as described in [19].

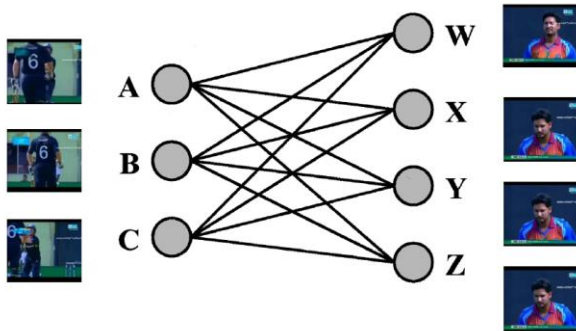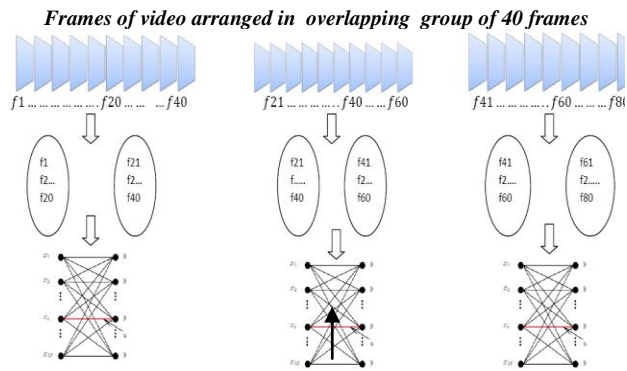The complete process of bipartite graph modeling is as shown in Fig. 4.

Fig.3. A bipartite graph model with frames as nodes on both subsets and links

Frames of video arranged in overlapping group of 40 frames

Bipartite graph representation of frames in group of 40 with 20 frames on each side

Fig.4. Process of bipartite graph modeling in proposed algorithm.

B. Minimum Edge weight Matching in Bipartite graph based Algorithm for Dynamic summarization

A graph G = (V,E) consists of a set V of vertices and a set E of pairs of vertices called edges. For an edge e = (u, v), the endpoints of e are u and v; also said that e is incident to u and v. A graph G = (V,E) is bipartite if the vertex set V can be partitioned into two sets X and Y (the bipartition) such that no edge in E has both endpoints in the same set of the bipartition. A matching M ⊆ E is a collection of edges such that every vertex of X and Y are incident to at most one edge of M. An example of match is shown in Fig. 5 and the cost of a matching M is given by equation (4)
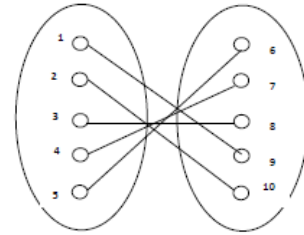

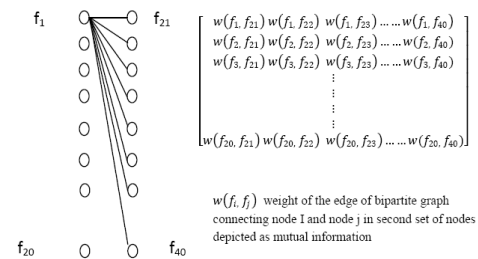Fig.5. Example Matching in bipartite the edges (1, 9), (2, 10) , (3, 8) ,(4,7) and (5,6) form a matching.

$$C(M) = \sum_{(i,j)\in M} w_{ij} \tag{4}$$

w_{ij} *: is weight/cost of the edge connecting i j(i ∈ X and j ∈ Y)*

A Maximal matching is a matching to which no more edges can be added without increasing the degree of one of the nodes to two. In the proposed work the initial task is to find the matching of minimal weight in bipartite graph modeled as two vertex subsets X and Y using group of 40 frames. The weight of a matching M is total weight of edges in matching M.

In the proposed work the weighted bipartite graph is represented in the matrix form as in (5) with weight value associated with each edges of the bipartite graph, and is used as assignment matrix[20].

$$
\begin{bmatrix}
w(f_1, f_{21}) & w(f_1, f_{22}) & w(f_1, f_{23}) & \ldots \ldots w(f_1, f_{40}) \\
w(f_2, f_{21}) & w(f_2, f_{22}) & w(f_2, f_{23}) & \ldots \ldots w(f_2, f_{40}) \\
w(f_3, f_{21}) & w(f_3, f_{22}) & w(f_3, f_{23}) & \ldots \ldots w(f_3, f_{40}) \\
& & \vdots & \\
& & \vdots & \\
w(f_{20}, f_{21}) & w(f_{20}, f_{22}) & w(f_{20}, f_{23}) & \ldots \ldots w(f_{20}, f_{40})
\end{bmatrix}
$$

$w(f_i, f_j)$ weight of the edge of bipartite graph connecting node I and node j in second set of nodes depicted as mutual information

Bipartite graph with nodes representing frames of video

$$(5)$$

where

$f_1, f_2, f_3 \ldots \ldots \ldots f_{20}$ *are the frames represented as nodes on one side of bipartite graph*

$f_{22}, f_{23} \ldots \ldots \ldots f_{40}$ *are the frames represented as nodes on second side of bipartite graph*
weight $w(f_i, f_j)$ *is mutual information between ith frame and jth frame.*

The proposed technique begins with finding maximal matching to match two groups of 20 frames represented as vertex sets of a bipartite graph where the sum of the values of the edges in the matching have a minimum value. Such a matching is known as *Minimum edge weight matching*. Finding such a matching is treated as the assignment problem [21]. So Hungarian method described in[22][23] is employed to find minimum weight matching. The process of obtaining minimum edge weight matching, extraction of representative frames and construction of summary is described in *Algorithm1*.

Algorithm1: method to find minimum edge weight maximal matching of bipartite graph and key frame extraction for summarization
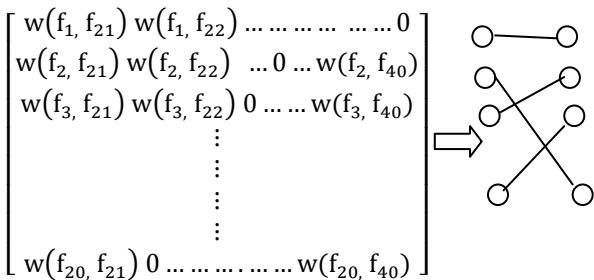

Input : The Bipartite graph represented as weight matrix with values of edge weights depicted as mutual information
Output : A Dynamic Summary clip
Step 1 . Represent Bipartite graph model for group of 40 frames as weight matrix (assignment Matrix) as in (5)
Step 2 . Apply Hungarian method to find the Minimum weight matching from matrix interpretation of bipartite graph with edge weight depicting mutual information between frames representing nodes the edge connects.
Step 3: Use the output of Hungarian method, the edge weight values of the positions indicated as zeros in weight Matrix to depict edges from a node on first side of bipartite graph to a node on second side.

$$\begin{bmatrix} w(f_1, f_{21}) \; w(f_1, f_{22}) \ldots \ldots \ldots \ldots \ldots 0 \\ w(f_2, f_{21}) \; w(f_2, f_{22}) \; \ldots 0 \ldots w(f_2, f_{40}) \\ w(f_3, f_{21}) \; w(f_3, f_{22}) \; 0 \ldots \ldots w(f_3, f_{40}) \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ w(f_{20}, f_{21}) \; 0 \ldots \ldots \ldots . \ldots \ldots w(f_{20}, f_{40}) \end{bmatrix} \Rightarrow$$

Step 4: From the edges of matching obtained find the edge with maximum edge weight and identify the edges with weight values less than or equal to 70 % of the maximum edge weight.
Step 5: Collect the frames representing nodes connected by edges found in Step4 and two of their neighboring frames, one preceding and one succeeding on time line as representative Key frames.
Step 6 : Construct a summary clip adding all the frames extracted from all the matches obtained for every group of 40 frames (as in Step 4 and Step 5)) from video along the the timeline

---

The frames with very high similarity depicted as nodes of matching (with maximum mutual information and with above 70% of maximum) are eliminated to reduce redundancy The experiments have revealed that frames connected by edge weight below 70% of maximum edge weight(highest similarity) in that bipartite graph can better represent the group of 40 frames and their immediate neighbor frames efficiently contribute to dynamic summary. The frames which are left back considerably remove redundancy and compact the content of a video segment. All such frames collected from bipartite graph representation for every group of 40 frames from video are used to create a clip. Upon adding audio frames to the clip, the resulting video is a summary with sufficient dynamics that will give complete preview with comprehensive coverage of content.

IV. EXPERIMENTATION AND RESULTS

The algorithm is implemented in Matlab and experiments are performed on a Core2 Duo 2.40 GHz Windows machine with 2GB of RAM. The experiments are conducted in order to validate the performance of the summarization algorithm and results are presented in the following. The performance of the proposed algorithm is evaluated on the YOU TUBE sports videos and the videos of TRECVID MED 2011 summaries data set. The performance of the proposed algorithm is evaluated using standard evaluation metrics.

In the proposed work, first the video is considered in group of 40 frames each with 50% overlap and *Mutual Information (MI)* values between every frame of one subset of 20 frames and second subset a group of 20 consecutive frames are computed as explained in the section 3(a). A weight matrix is constructed for complete bipartite graph representation of smaller unit of 40 frames from video. The frames are represented as nodes of graph and edge weight is represented by Mutual information between two frames connected by that edge. A weight adjacency matrix for bipartite graph model with two subsets each of 20 nodes representing frames for a cricket video taken from YOUTUBE is presented in Fig. 6.

	f_{21}	f_{22}	f_{23}	f_{39}	f_{40}
f_1	0.1570	0.1612	0.1650	0.1221	0.1262
f_2	0.1498	0.1569	0.1539	0.1210	0.1221
f_3	0.0856	0.0859	0.0892	0.0921	0.0892
f_4	0.0756	0.0677	0.0704	0.0829	0.0834
f_5	0.0934	0.0939	0.1054	0.0821	0.0828
f_6	0.0607	0.0509	0.0636	0.0525	0.0477
f_7	0.0780	0.0718	0.0780	0.0489	0.0441
f_8	0.0798	0.0725	0.0792	0.0505	0.0445
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
f_{20}	0.1032	0.0953	0.0833 ....0.1278		0.1277

Fig.6. Cost Adjacency matrix representation of complete bipartite graph

For every weighted bipartite graph model of 40 frames with non-negative weights wij corresponding to the edge (i, j) *Minimum edge weight matching* is obtained employing Hungarian method. A complete bipartite graph for cost adjacency matrix(weight matrix) in Fig. 6 is shown in Fig. 7(a) and a typical maximal matching of it is shown in Fig.7(b), This procedure represents complete video as sequential set of bipartite graph matches as illustrated in Fig. 8.
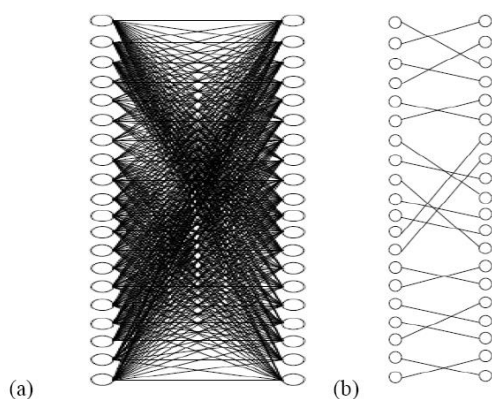

Fig.7. (a) Bipartite Graph representation of frames from Video (b) Minimum edge weight maximal match obtained for bipartite graph
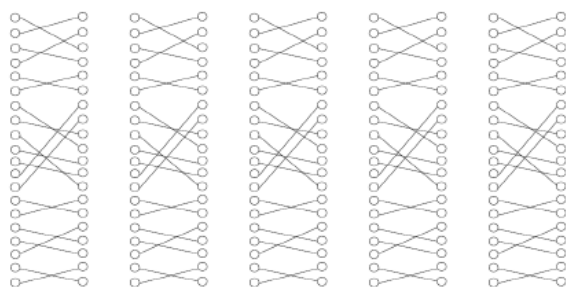

Fig.8. Sequences of minimum edge weight matches obtained from a video for every group of 40 frames

The key frames are extracted from each match as explained in *algorithm 1*. From each matching, first an edge with maximum edge weight is identified and the frames represented by nodes connected by edges of higher weight depicted as mutual information (above 70 % of the maximum edge weight) i.e with higher similarity are eliminated leading to reduction of redundancy. The frames represented by nodes connected by the edges of weight below 70 % of the maximum edge weight and their two neighboring frames are selected as representative key frames. The threshold value of 70% is empirically optimized. The experiments have shown that higher value of threshold will increase the size of summary resulting in lower conciseness and smaller value of threshold will decrease the number of key frames leading to reduction in dynamicity of summary.

In the proposed work the sports video segments taken from YOUTUBE are used in the experiments to evaluate

the performance of proposed methodology. The performance parameters namely *meaningful summary duration (MSD), Informativeness and Satisfaction* are used to demonstrate the performance of the proposed methodology the results are given in Table 1. The performance of the proposed algorithm is also evaluated by experimentation on *TRECVID MED 2011 summaries* video data set and the comparison of the results of the proposed method with results in data set is presented in Table 2.

A. Performance Evaluation of Algorithm

A challenging issue of the video summarization is related to the evaluation of the summary constructed using various algorithms, In this work a subjective evaluation method is adopted to assess the quality of video summaries There are several quality measures that can be used to evaluate the efficiency of the algorithms to produce quality summary. The measures to quantify the quality of summary are Summary duration(MSD), Informativeness and satisfaction. The MSD is duration of summary clip produced by adding all the representative frames extracted from all the segments of the video along the time line. The total duration of summary clip compared to total duration of video will quantify amount of conciseness accomplished by summarization. Informativeness assesses the capability of maintaining content coverage while reducing redundancy. Satisfaction metrics quantifies how satisfactory is the summarized video compared to the original one.

In order to quantify the performance of the video summarization in terms of Informativeness five reviewers are asked to assign scores in percentage from 0 to 100 % as informativeness to every summary. The average of the this score assigned by every reviewer for each summary is Informativeness of the each summary.

i.   The summary was easy to understand
ii.   The summary was enjoyable
iii.   The summary was informative
iv.   The summary was interesting
v.   The summary was coherent
vi.   The summary represented the video well
vii.   The summary would aid in deciding whether to watch the full video

Satisfaction measures [24] are derived from percentage rating for the above questions by reviewers and average of percentage rating demonstrates the satisfaction metrics for each video summary

The results of experimentation on 5 selected sports videos taken from YOUTUBE using proposed algorithm are presented in Table1. The comparison of the quality of summaries constructed by proposed algorithm for videos from TRECVID MED 2011 video dataset with summaries presented in dataset is also given in Table2.

Table 1. Performance measures of Proposed Algorithm

Results of proposed methodology for the Videos from YOUTUBE				
Video	Length of the video in seconds	Length of the summary in Seconds(MSD)	Satisfaction	Informativeness
1	195	21	93%	95%
2	390	42	94%	95%
3	2060	460	90%	92%
4	3630	660	90%	93%
5	3680	690	92%	94%
Average			91.8%	93.8%

Table 2. The comparison of Quality of summaries produced by proposed algorithm for videos from TRECVID MED 2011 and summaries presented in data set

	Quality of summaries presented in MED summaries for Data set TrecVid MED 2011 Videos				Performance of Proposed methodology for the Data set TrecVid MED 2011 Videos		
Videos	Length of the video in seconds	Length of the summary in Seconds(MSD)	Satisfaction	Informative ness	Length of the summary in Seconds(MSD)	Satisfaction	Informative ness
1	66	15	91%	93%	13	93%	94%
2	200	42	93%	95%	38	93%	95%
3	300	60	90%	93%	55	91%	95%
4	420	80	90%	92%	68	92%	93%
5	480	90	93%	95%	83	93%	95%
	Average		91.4%	93.6%	Average	92.4%	94.4%

To quantitatively evaluate the performance of proposed video summarization method the criteria namely, *Summary duration (MSD) Informativeness* and *satisfaction* are used. The MSD is indicator of amount of conciseness attained by the proposed methodology in summarization. *Informativeness* accesses the capability of maintaining content coverage while reducing the redundancy. The *satisfaction* is measure of information coverage and visual pleasantness accomplished by summarization method. The results of this work reveal the suitability of the proposed algorithm for summarization of various length videos. The results also prove that proposed algorithm is comparable with recently presented methods.

The proposed method does not just selects one key frame to represent one group of frames from video as most of static summarization methods do. The method extracts more frames in pairs and their neighboring frames from a group of frames modeled as bipartite graphmatch. The frames thus selected from complete video of large size upon arranging on timeline presents certain amount of similarity along timeline. The video file created using frames extracted from a large video will create small concise clip with comprehensive coverage of video, making a summary. As the proposed method extracts representation from smaller group of frames

from complete video the method attains more comprehensive coverage as compared to event based methods and selected segment based methods which extracts only interested segments from video or from separated shots of video for summarization. The summaries presented in the data set are produced by one such method. The values of performance parameters namely **Informativeness** value **of 94 %** and **Satisfaction** value of 92 % and duration (MSD) of summaries of proposed method reveal that it has attained better conciseness, good satisfaction and better informativeness as compared method by which the summaries presented in data set are produced.

V. CONCLUSION

In this work a graph theoretic approach for dynamic video summarization is proposed by bipartite graph modeling of the video. The video frames are partitioned into groups of 40 consecutive frames with 50% of overlapping. Every group of frames is modeled as bipartite graph by representing 20 frames from group as vertices on one side of the bipartite graph the next consecutive 20 frames form the other part of the bipartite graph. The edges of the bipartite graph are edges

connecting a node from the first set of 20 nodes to a node in second set. The interrelation between frames from first set and second set is determined using similarity measured by obtaining maximal matching with minimum weight which is summation of weights of edges present in the match expressed in terms of mutual information between frames represented by the nodes connected by the edge. Higher the mutual information more similar are the nodes ie frames, so the frames connected by the edges with lower edge weights are taken as representative key frames. To build dynamicity into summary frames with certain similarity are also taken by considering frames connected by edges with edge weight below 70 % of maximum edge weight and their two immediate neighboring frames(one frame before and one after it). The higher edge weights are avoided to eliminate more similar frames. The results of experiments conducted on data set containing sports videos downloaded from YOUTUBE and TRECVID MED 2011 data set presented in terms of average values of performance parameters namely Informativeness value and satisfaction indicates that the proposed mechanism is suitable for video summarization to produce dynamic summaries and may be experimented on other types of videos.

REFERENCES

[1] Boreczky JS, Rowe LA 1996 Comparison of video shot boundary detection techniques. Proc Int Conf Storage Retr Still Image Video Databases 5(2):170–179.

[2] Behzard S, Gibbon DS 1995 Automatic generation of pictorial transcripts of video programs. Proc SPIE Multimedia Computer Networking 2417:512–518.

[3] Y. Pritch, A. Rav-Acha, and S. Peleg,"Nonchronological Video Synopsis and Indexing," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 11, pp. 1971-1984, 2008.

[4] Won-il hwang, pyung-jun lee, bong-kyung chun, dong-sung ryu, hwan-gue cho 2006 "cinema comics : cartoon generation from video stream"in GRAPP 2006 - COMPUTER GRAPHICS THEORY AND APPLICATIONS.

[5] J. Kavitha, P. Arockia Jansi Rani 2015, "Design of a Video Summarization Scheme in the Wavelet Domain Using Statistical Feature Extraction" , International Journal of Image, Graphics and Signal Processing, 2015, Volume 4, PP 60-67.

[6] Yue Gao Wei-Bo Wang · Jun-Hai Yong · He-Jin Gu 2009 "Dynamic video summarization using two-level edundancy detection" in Spinger Multimed Tools Appl (2009) 42:233–250.

[7] Yongliang Xiao, Limin Xia,2010, "Key Frame Extraction Based on Connectivity Clustering" in proceedings of Second International Workshop on Education Technology and Computer Science (ETCS), volume 1 pp 174 – 177.

[8] A.M. Ferman and A.M. Tekalp, 1998"Efficient Filtering and Clustering Methods for Temporal Video Representation and Visual Summarization", J. Visual Com. & Image Rep., vol. 9, pp.336-351.

[9] JeongKyu Lee JungHwan Oh Sae Hwang,2005, "Scenario based Dynamic Video Abstractions using Graph Matching" in ACMMM 05 , November 6–11, 2005, Singapore.

[10] Girgensohn and j. Boreczky, 2000 "Time-Constrained Keyframe Selection Technique," multimedia. Tools & appl, v.11, pp.347-358.

[11] Shi Lu ; Dept. Of Comput. Sci. & Eng., Chinese Univ. Of Hong Kong, China ; Lyu, M.R.; King, I. 2004," Video Summarization By Spatial-Temporal Graph Optimization" Published in: , 2004. ISCAS '04. Proceedings Of The 2004 International Symposium On Circuits And Systems.

[12] Yuxin Peng Chong-Wah Ngo 2005 "Hot Event Detection and Summarization by Graph Modeling and Matching" CIVR 2005, LNCS 3568, pp. 257 – 266, 2005. © Springer-Verlag Berlin Heidelberg 2005.

[13] Rong Pan , Yumin Tian , Zhong Wang 2010 Key-frame extraction based on clustering in proceedings of IEEE International Conference on Progress in Informatics and Computing (ICPIC), 10-12,Dec.2010,**Volume:**2pp**:**867-871 .

[14] Besiris, D. Fotopoulou, F. ; Economou, G.; Fotopoulos, S. 2008Video Summarization By A Graph-Theoretic FCM Based Algorithm in Proceedings Of 15th International Conference on Systems, Signals And Image Processing, 2008. IWSSIP 2008.

[15] Yue Gao, Wei-Bo Wang, Jun-Hai Yong,He-Jin Gu 2008 Dynamic Video Summarization Using Two-Level Redundancy Detection Multimedia Tools and Applications April 2009, Volume 42, Issue 2, pp 233-250.

[16] Chong-Wah Ngo, Yu-Fei Ma, Hong-Jiang Zhang, 2005 "Video Summarization And Scene Detection By Graph Modeling" Ieee Transactions On Circuits And Systems For Video Technology, Vol. 15, No. 2, February 2005.

[17] D. Besiris a. Makedonas 2009 "combining graph connectivity & dominant set clustering for video summarization"Journal of multimedia tools and applications volume 44 issue 2, september 2009 pages 161 – 186.

[18] Michel X.Goemans 2009 , "Lecture notes on bipartite matching Combinatorial Optimization" Massachusetts Institute of Technology .

[19] T. M. Cover and J. A. Thomas. Elements of Information Theory. John Wiley and Sons, New York, 2006.

[20] R.E. Burkard, M. Dell'Amico, S. Martello: *Assignment Problems* SIAM, Philadelphia (PA.) 2012. ISBN 978-1-61197-222-1 .

[21] Tasorn Sornnarong 2014 "Dynamic matching in crowdsourcing platform" Thesis submitted to School of Computer and Communication Sciences Swiss Federal Institute of Technology Lausanne, EPFL .

[22] H.W. Kuhn, On the origin of the Hungarian Method, History of mathematical programming; a collection of personal reminiscences (J.K. Lenstra, A.H.G. Rinnooy Kan, and A. Schrijver, eds.), North Holland, Amsterdam, 1991, pp. 77–81.

[23] Harold W. Kuhn The Hungarian Method for the Assignment Problem 2010 50 Years of Integer Programming 1958-2008 2010, pp 29-47.

[24] Stina Westman, 2011 Evaluation of visual video summaries: user-supplied constructsand descriptions, in the Proceedings of the 14th European Conference on Digital Libraries (ECDL 2010).

Authors' Profiles

Shanmukhappa. A. Angadi is a Professor in the Department of Computer Science and Engineering, Centre for PG Studies, Visvesvaraya Technological University, Belagavi, Karnataka, India. He earned a Bachelor Degree in Electronics and Communication Engineering from Karnataka University, Dharwad and a Master Degree in Computer Engineering from University of Mysore, Mysore and a PhD in Computer science from the Department of Studies in Computer science, University of Mysore, Karnataka, India. He also completed PGDiploma in Opeartions Management (PGDOM), from IGNOU, New Delhi. He has completed many research and consulting projects of AICTE India under RPS, MODROBS, TAPTEC schemes and Research grant scheme of VTU Belagavi. His research areas are Image Processing and Pattern Recognition, Character Recognition, soft computing, Internet of things and Graph Theoretic techniques. He is a co-author of a book on C-programming language. He is life member of professional bodies like ISTE and IETE

Vilas Naik received BE(Electronics and Communication) from Karnataka University Dharwad and Master of Engineering in Computer Technology from Shri Guru Govind Singh College of Engineering Nanded under Sri Ramanand Teerth Marathawada University Nanded .India He is currently a research scholar registered to Visvesvaraya Technological University, Belagavi in the area of Image and Video processing working on the issues of Multimodal Video summarization and selection. Currently he is working as Assistant Professor in the Department of Computer science and Engineering, Basaveshwar Engineering College Bagalkot. His subjects of interest are Image and Video processing, Data Communications and Computer Networks, Computer Architectures and Multimedia computation and Communication