IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 2, Ver. VIII (Mar-Apr. 2014), PP 119-125
www.iosrjournals.org

Automatic Image Retrieval through Video Authoring and Transition

Akila. R, Jayashree. B, Mr.P.Karthick, M.E
Student, CSE dept, AIHT, Chennai,
Student, CSE dept, AIHT, Chennai,
Assistant Professor, CSE dept, AIHT, Chennai,

Abstract: *An integrated system for video summarization, browsing and presentation, based on large amount of personal and web video clips. Content-consistent shots are retrieved from a video pool in order to form a descriptive long-shot video automatically by video authoring and transition to present events, persons or scenic spots captured at various remote places in an informative manner. Users accessing the application have to register their information via user authentication. Recognized users records and their personal information are secured and maintained by the administration. Input short videos are converted into frames in pre-processing where each frames are resized and merged into a single video for video categorization. Videos are categorized by using transition clues like human, object. The frames are categorized into human and nonhuman frames by using Viola-Jones algorithm where Region Of Interest is used to separate the frames. Human frames are composed with and without reference image by using Back Propagation Network with the help of Trainee Database which consists of list of human frames with different angles in a specific Group. Non Human frames are composed with and without Reference image by using Speeded up Robust Features where object and sequence matching process are performed. Object frames and related sequence frames are categorized into a separate folder and converted into separate single shot videos. The system provides an efficient video browsing mode to generate matching graph of videos.*

Key-words : *video authoring, video transition, long-shot video, user authentication, ROI.*

I. INTRODUCTION

In recent years, multimedia data has received continuously increasing interest in humans and this content gets bigger day by day with the advances in technology. Most of this content is related to visual information including video data produced by filmmakers, TV channels, amateur camera users etc. Due to advancement in technology, it becomes very easy to record huge volume of videos. A huge bulk of digital contents such as news, movies, sports, and documentaries is available. Moreover the need for surveillance has increased significantly due to increase in demand of security. Thousands of video cameras can be found at public places, public transport, banks, airports, etc. resulting in large amount of information. Extracting specific information from such a huge amount of video content creates some difficulties to search whole media like limitations on time consuming in browsing and retrieving of relevant data. Furthermore, storage of huge amount of data is not that easy. It is very important to quickly retrieve and browse huge volume of data effectively because the end user wants to get all important aspects of data. To overcome the drawbacks, the current trend is to develop algorithms capable of parsing them by segmenting and then indexing. On the other hand, temporal segmentation of a video is needed to enable an efficient indexing procedure for localizing and accessing the source of relevant information. Video summarization plays an important part in this regard, as it helps the user to navigate and retrieve through large sequence of videos. A complete video is constructed by shots, which are the collection of consecutive frames that are recorded in one camera record time. Capturing a high-quality long-shot video needs an accurate coordination between the camera movement and the captured object for a long period, which is usually difficult even for professionals. Users often need to maintain their own video clip collections captured at different locations and time. These unedited and unorganized videos bring difficulties to their management and manipulation. When users want to share their story with others over video sharing websites and social networks, such as YouTube.com and Facebook.com, they will need to put more efforts in finding, organizing and uploading the small video clips. This could be an extremely difficult task for users. Previous efforts towards efficient browsing such large amount of videos mainly focus on video summarization. These methods aim to capture the main idea of the video collection in a broad way, which, however, are not sufficiently applicable for video browsing and presentation.

We propose a structure to compose descriptive long-take video with content-consistent shots retrieved from a video pool. For each video, frame-by-frame search is performed over the entire pool to find start-end content correspondences through a coarse-to-fine partial matching process. The content correspondence here is general

and can refer to the matched regions or objects, such as human body and face. The content consistency of these correspondences enables us to design several shot transition schemes to seamlessly stitch one shot to another in a spatially and temporally consistent manner. The entire long-take video thus comprises several single shots with consistent contents and fluent transitions. Meanwhile, with the generated matching graph of videos, the proposed system can also provide an efficient video browsing mode. Experiments are conducted on multiple video albums and the results demonstrate the effectiveness and the usefulness of the proposed scheme.

Human frames

Object and reference frames

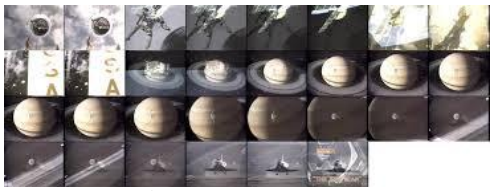

Fig.1.1. An illustration of the presentation scheme, which generates Frames from videos by selecting and composing short video clips.

The system automatically generates a virtual one-shot presentation from multiple video clips. Given a messy collection of video clips, it can select a clip subset with consistent major topic (similar with finding the clues [3]). The topic can refer to a person, object, or a scene here. It can be specified by users or found with an automatic discovery method. Video Puzzle provides a novel presentation of video content that enables users to have a deeper impression of the story within the video collection.

II. Related Work

Our system concentrates on how to provide consecutive smooth video while other approaches also endeavor to classify the video segments by film theory, and compose them into a story. However, the target of the method is for professional videos which capture the whole story of a certain event while home videos and web videos often have no fixed single story [2]. Although our work can provide aesthetically pleasing videos among the user's video collection, the targets and methodologies used are totally different. It aims to automatically discover content-consistent video shots and compose into a virtual long-take video with spatial and temporal consistency. It also provides a tree structure collection with temporal smoothing for ease of video browsing. In comparison with the conventional video abstraction and presentation techniques, we not only provide a novel presentation approach but also facilitate further services such as editing.

A. ***Summarizing Videos.***

Video summarization plays an important role in this context. It helps in efficient storage, quick browsing, and retrieval of large collection of video data without losing important aspects. Video summarization has become an emerging field of research. Video summaries are becoming increasingly popular as a way to convert large chunks of video into smaller, more comprehensible units [7]. Video summarization methods are categorized on the basis of methodology used. Many video summarization techniques rely on finding features in the video or audio streams, using them to identify clips, and then assembling the most salient clips into a shorter version of the original. An alternative approach is to match and take advantage of the known structure of a video sequence to find salient elements, and then match those segments into a summary design pattern. With these techniques, the final product is more elegantly watchable, with fewer artifacts of concatenation, and the final product is usually more content bearing than summaries composed with simple signal analysis techniques.

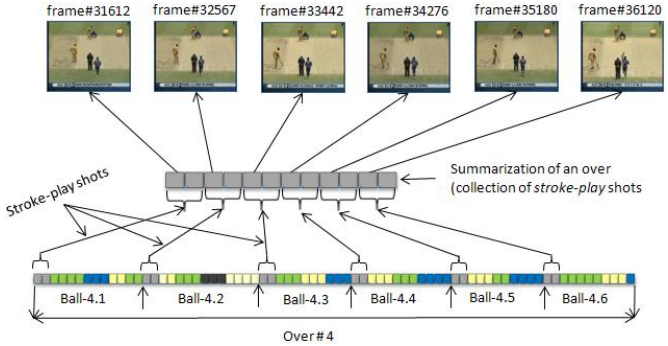

Fig.1.2. Shot transition probability graph.

B. Feature Based Video Summarization

Video summarization is usually seen as a complementary to video retrieval, as it makes browsing of retrieved videos faster. Especially when content-based video retrieval fails and the user need to browse through a large set of video sequences [1]. Digital video contains many features like color, motion, dynamic contents, gesture, audio-visual, speech transcript, object and voice etc. these techniques work well if user wants to focus on features of video. For example, if user wants to see color features then it's good to pick color based video summarization techniques. User-generated video content, which has its own distinct characteristics compared to other video types such as professionally, produced video content. As a consequence, the summarization of user-generated content is more challenging than for more constrained video content. Due to its unconstrained nature the summarization techniques must be generic in such that they can cope with all the infinite number of variations in which user-generated content exists.

C. Static and Dynamic Presentation

Video summarizations are commonly presented as set of static keyframes or dynamic video skims. After extracting the keyframes of video sequence there are different options for presenting them to the user. One of the most common video summarization presentation techniques is a storyboard, which is usually a static grid of extracted keyframes [3]. According to a recent study on evaluation of video summarization techniques the storyboard has a capability to give an informative summary of the original video content. However, according to the user studies the storyboard lacked in their representativeness and ability to replace the original video content.

Dynamic video skimming is a technique that condenses the original video into a shorter version, while preserving important content with its time-evolving properties. Hence, video skims are practically short video clips cut from the original video sequences [8], [13]. Preservation of motion information is one of the greatest advantages of video skims, in addition to the aural information, which can both enhance the expressiveness of the video summary.

D. Hierarchical Presentation

In addition to the static and dynamic video abstracts also a technique for hierarchical abstraction has been implemented. The hierarchical abstract is a multilevel video summarization that is based on the importance of structural units in video sequence that is defined by visual and aural attention levels. A hierarchical structuring of relations may result in more classes and a more complicated structure to implement. Therefore it is advisable to transform the hierarchical relation structure to a simpler structure such as a classical flat one. It is rather straightforward to transform the developed hierarchical model into a bipartite, flat model, consisting of classes on the one hand and flat relations on the other. Flat relations are preferred at the design level for reasons of simplicity and implementation ease. There is no identity or functionality associated with a flat relation. A flat relation corresponds with the relation concept of entity-relationship modelling and many object oriented methods.

III. Video Editing

Our system aims to automatically discover content-consistent video shots and compose into a virtual long-take video with spatial and temporal consistency which provides a tree structure [11] collection with temporal smoothing for ease of video browsing. Besides, our system displays the collection with a more general graph structure and inferencing on the graph leads to the auto-discover of content-consistent video shots. Moreover, the video similarity in our system is based on mutli-cue matching and no previous similar work on video browsing has used this kind of information.

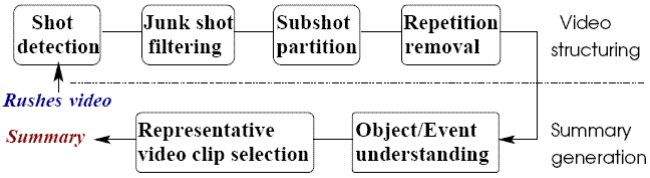

Fig.1.3 An illustration of video composition from the collection of videos

There also exist studies on video texture [5], [10] which aims to provide a continuous and infinitely varying stream of images**.** Remove unwanted footage is the simplest and most common task in editing. Many videos can be dramatically improved by simply getting rid of the flawed or unwanted bits. Choosing the best footage is common to shoot far more footage than you actually need and choose only the best material for the final edit. Editing is a crucial step in making sure the video flows in a way which achieves this goal. Adding effects, graphics, music, etc is often the best part of editing. Altering the style, pace or mood of the video includes techniques such as mood music and visual effects that can influence how the audience will react.

A. Coarse-To-Fine Partial Matching Process

We implement a coarse-to-fine partial matching scheme to generate a matching graph of the video collection. The matching scheme serves as a three-level matching, i.e., video pair selection, sequence-sequence correspondence finding, and frame-level exact matching. The video pair selection acts as an evidence for ensuring the non-redundant and complete quality of the generated one-shot video [12]. It uses a hashing-based method to quickly obtain the video similarity. We then find sequence correspondence of the selected video pairs through local key points matching. The final frame-level matching aims to find different matched objects to provide variant and rich clues for video transition generation.

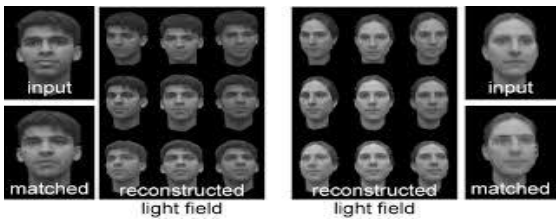

Fig.1.4 An illustration of coarse-to-fine partial matching process.

B. Graph-Based multilevel Temporal Video Segmentation

A graph-based solution approaches have become very popular for the pattern recognition research community. Graph based approaches have also been presented in structuring and summarizing of videos [10]. In each level of segmentation, a similarity matrix of frame strings is constructed by using temporal and spatial contents of frame strings. Using a priori information about a frame string, a strength factor is estimated for each frame string. The similarity matrix is reevaluated from a strength function derived by the strength factors. Then, a weighted undirected graph is constructed by the similarity matrix. The graph is partitioned by using normalized cuts algorithm with one additional constraint. Each graph cluster represents one segment of a video. Therefore, a hierarchically segmented video tree is constructed.

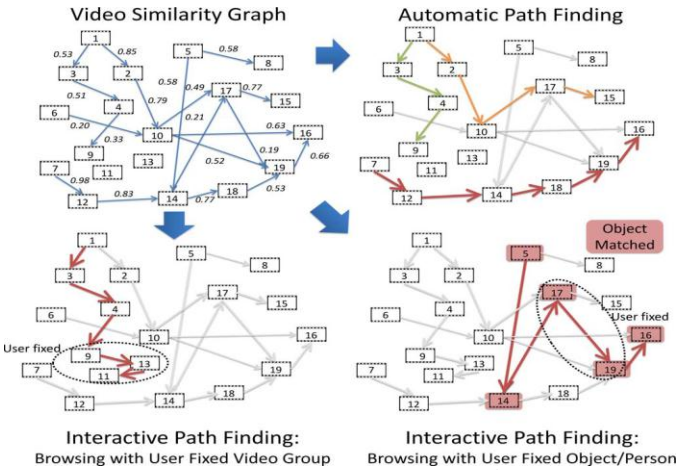

Fig.1.5. Graph Construction and Path Finding. The top-left graph is the original video similarity graph for a given video album. The videos are linked with different edge weights. Top-right graph shows the several paths of automatic path finding (indicated by different colors). Bottom-left graph shows the result of interactive path finding with user fixing a small video group (circled in the graph). Bottom-right graph shows the result of interactive path finding with user fixing the matched object/person discovered by the system (circled in the graph).

C. Edge pruning

The pruning algorithm is a technique used in digital image processing based on mathematical morphology. It is used as a complement to the skeleton and thinning algorithms to remove unwanted parasitic components. In this case 'parasitic' components refer to branches of a line which are not key to the overall shape of the line and should be removed. These components can often be created by edge detection algorithms or digitization [13]. The standard pruning algorithm will remove all branches shorter than a given number of points. The algorithm starts at the end points and recursively removes a given number (n) of points from each branch. After this step it will apply dilation on the new end points with a $(2N+1)(2N+1)$ structuring element of 1's and will intersect the result with the original image. If a parasitic branch is shorter than four points and we run the algorithm with $n = 4$ the branch will be removed. The second step ensures that the main trunks of each line are not shortened by the procedure. 5.viola-jones algorihm

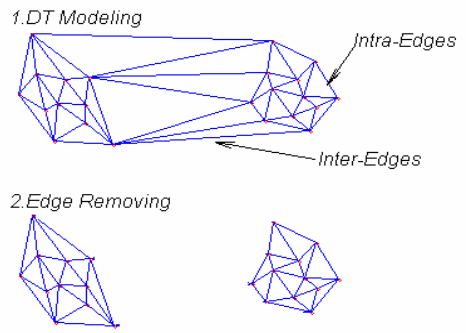

Fig.1.6.Delaunay Edges

IV. Pre- Processing

First Our Input short videos are converted into frames. Then we eliminate some frames like information less frames (Mean of Input frame<15). Each frame is resized and merged into a single video for video categorization. In this phase, humans and nonhumans frames are separated by using Viola-Jones algorithm. Using this algorithm the frames will categorized as human and nonhuman frames. If the frame has ROI(Region Of Interest) its comes under human frames or else its comes under non human frames.

V. Back Propagation Network (Bpn) Method

Videos are categorized by using transition clues like human, object. In these module human frames has to be composed with and without Reference image by using BPN with the help of Trainee Database. Trainee Database has to be created after Pre-Processing has been completed. It consists of list of human frames with different angles in a specific group. Each human frame is stored in a uniform manner. Initially each human frame has to be checked one by one with frames in Trainee Database. If the first human frame is matched with first group in

Trainee Database, then a separate folder will be created and stored. Repeating the process until all the human frames are completed and categorized. If the reference image is empty, on that condition frames in each folder will be composed for all individual humans. If the reference image is not empty, composed frames depends upon the reference image of that specific human.

VI. Overall Architecture

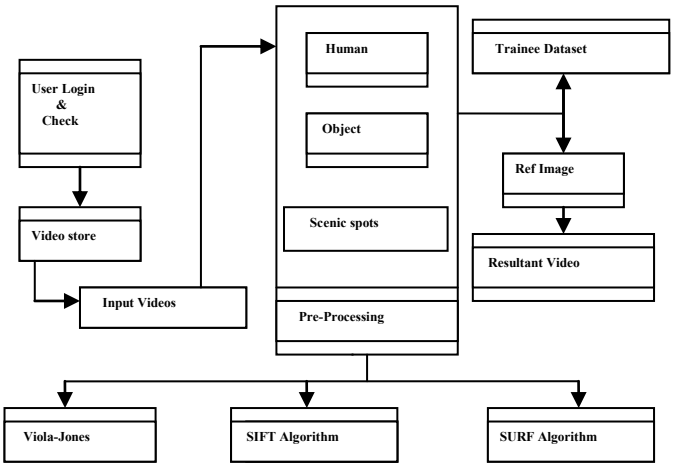

Fig. (b) Overview of architecture to retrieve single shot video

Our task is to automatically compose several related video shots into a virtual long-take video with spatial and temporal consistency, and it is different from the traditional works that try to either find a group of similar video clips. We implement a coarse-to-fine partial matching scheme to generate a matching graph of the video collection. The matching scheme serves as a three-level matching, i.e., video pair selection, sequence-sequence correspondence finding, and frame-level exact matching. The video pair selection acts as an evidence for ensuring the non-redundant and complete quality of the generated one-shot video.

VII. Scale-Invariant Feature Transform (Sift) Algorithm

In this module non human frames are composed with and without reference image by using SIFT. Initially first frame is extracted from non human frames and Compared with the remaining frames. Comparing processes are done by object matching and their characteristics. Once the process has been completed, frames are collected and stored in a separate folder and composed in a one shot video. Then the second frame is extracted and the process is repeated for the previous frame. The process is repeated for all frames present in the non human frames. If the reference image is empty, on that condition frames in the each folder will be composed for all individual nonhumans. If the reference image is not empty, composed frames depends upon the reference image of that specific nonhuman. Object frames and related sequence frames are categorized into separate folder. Finally categorized frames are converted into Separate videos. Scale-Invariant Feature Transform is replaced by using SURF algorithm (Speeded up Robust Features) for improving matching accuracy and reducing Time-consuming.

VIII. Speeded Up Robust Feature (Surf)

Object & sequence matching process are replaced by using SURF algorithm (Speeded up Robust Features) for improving matching accuracy and reducing Time-consuming. The standard version of SURF is several times faster than SIFT and claimed by its authors to be more robust against different image transformations than SIFT.

IX. Conclusion

Our work can provide aesthetically pleasing videos among the user's video collection; the targets and methodologies used are totally different. Our system aims to automatically discover content-consistent video shots and compose into a virtually long-take video with spatial and temporal consistency which provides a tree structure collection with temporal smoothing for ease of video browsing. It can facilitate family album management and web video categorization. Experimental results show that it has great potential to be used in future video management systems.

REFERENCES

[1] C. Barnes, D. Goldman, E. Shechtman, and A. Finkelstein, "Video tapestries with continuous temporal zoom," in *Proc. SIGGRAPH*, 2010.

[2] M. Everingham, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes (VOC) challenge," *Int. J. Comput. Vision*, vol. 88, pp. 303–338, 2010

[3] Video Summarization and Scene Detection by Graph Modeling Chong-Wah Ngo, *Member, IEEE*, Yu-Fei Ma, *Member, IEEE*, and Hong-Jiang Zhang, *Fellow, IEEE.*

[4] O. C. Philbin, "Near duplicate image detection: min-hash and tf-idf weighting," in *Proc. BMVC*, 2008.

[5] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal.Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645,Sep. 2010.

[6] E. Bennett, "Computational time-lapse video," *ACM Trans. Graph.*, vol. 26, no. 102, Jul. 2007

[7] J. Calic, D. Gibson, and N. Campbell, "Efficient layout of comic-like video summaries," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 7, pp. 931–936, Jul. 2007.

[8] Y. Caspi, A. Axelrod, Y. Matsushita, and A. Gamliel, "Dynamic stills and clip trailers," *Visual Comput.*, vol. 22, no. 9, pp. 642–652, Sep. 2006.

[9] C. C. Nikolaidis, "Video shot detection and condensed representation. A review," *IEEE Signal Process. Mag.*, vol. 23, no. 2, pp. 28–37, Mar. 2006.

[10] V. Kwatra, A. Schodl, I. Essa, and G. T. Bobick, "Graphcut textures: Image and video synthesis using graph cuts," *ACM Trans.Graph.*, vol. 22, no. 3, pp. 277–286, Jul. 2003.

[11] T. Wang, J. Collomosse, R. Hu, D. Slatter, D. Greig, and P. Cheatle,"Stylized ambient displays of digital media collections," *Comput.Graph.*, vol. 35, no. 1, pp. 54–66, 2011.

[12] Jane You, *Member, IEEE*, and Prabir Bhattacharya, *Senior Member, IEEE* , A Wavelet-Based Coarse-to-Fine Image Matching Scheme in A Parallel Virtual Machine Environment.

[13] Keyframe-based Video Summarization using Delaunay Clustering PADMAVATHI MUNDUR, YONG RAO, YELENA YESHA *Department of Computer Science and Electrical Engineering University of Maryland Baltimore County 1000 Hilltop Circle*.