Visual Attention Key Frame Extraction for Video Annotations.

Prof.Archana.V.Potnurwar
Research Scholar Computer Science &Engg Department [1]
P.I.E.T
Nagpur,India
archanapotnurwar@gmail.com1

Dr Mohammad Atique
Associate Professor P.G.Department Of Computer Science[2]
S.G.B.A.U,Amravati

***Abstract*— The insufficiency of labeled training data for representing the distribution of the entire dataset is a major obstacle in automatic semantic annotation of large-scale video database the objective is to represent the most "important" or "meaningful" scenes of the large amount of visual information by only a few images:the key frames. First, the image sequences are temporally segmented into continuous segments called shots.Then a few frames of each shot are selected as key frames.In this paper ,retrieving videos using key words requires obtaining the semantic features of the videos. Most work reported in the literature focuses on keyframe extraction methods in the video shot.The focus is on visual attention Keyframe extraction(VAKE).**

Keywords- *Visual attention ,keyframe, time constraint clustering.*

I. INTRODUCTION

In the recent literature, more attention is on two basic forms of video summaries: key frames and video skims (Truong and Venkatesh, 2007). Key Frames, also called representative frames or R-frames, is a collection of salient images extracted from the underlying video source. On the other hand, video skims, also called a moving-image abstract, moving storyboard or summarysequence, consist of a collection of video segments, and consequently their corresponding audio, extracted from the original video. The majority of techniques focus on the extraction of key rames. Moreover, it is worth noting that the authors in Hanjalic nd Zhang (1999) combine these two approaches Automatic creation of tags for searching of video through audio text and this is motivated by the ever-increasing pace at which video content is generated, rendering any annotation scheme that requires human labor unrealistically expensive and unpractical for use on anything but a very restricted subset of the generated content, which may be of unusually high value or importance (e.g. cinema productions, medical content).

II. PREVIOUS STUDY

Different Key frame extraction techniques is to

A) Compute the frame differences based on some criteria and then discard the frames whose difference with the adjacent frames are less than a certain threshold.Various low level features have been applied for this purpose including color histograms, frame correlations, edge histogram,etc. [12]

B). For instance, Pal and Leigh [13] used fuzzy geometrical and information measures to develop an algorithm to estimate the difference between two consecutive frames. The similarity between the frames was measured in terms of weighted distance infuzzy feature space. Hanjalic et al. [14] compared the difference in color histograms of consecutive frames with a threshold to obtain key frames.

C)Introduced by DeMenthon et al. [15]. The key frames were extracted by finding discontinuities on a trajectory curve, which represent video sequence.

D)n the Flexible Rectangles algorithm [16],the frame differences were used to build a ''Content Development Curve'' from a curve composed of a predefined number of rectangles through the use of an error minimization algorithm.

E)The Adaptive Sampling algorithm [17] extracted key frames by uniformly sampling the y-axis of the curve of cumulative frame differences .The resulting sampling on the x-axis represented the key frames.

F)The Shot Reconstruction Degree Interpolation [18] selected the key frames based on the ability of frames to reconstruct the original video shot using frame interpolation.

G)Ciocca and Schettini[5] extracted key frames by first finding the cumulative frame differences based on certain frame descriptors such as color histogram,histogram of edges and wavelets. Next, a curve of cumulative frame differences was sketched, and then the midpointsof two curvature points on this curve were selected as key

frames. A curvature point is a point on the curve where the angle changes are drastic. The frame difference based methods are intuitive and simple in nature. These properties make them suitable for many real-time and/or online applications. However, for extracting a particular key frame, these techniques only consider sufficient content change between the consecutive frames (or between current frame and last key frame). Therefore, a key frame that is extracted by these methods does not fully represent the portion of the video preceding it .

Some researchers used clustering for extracting key frames by treating video frames as points in the feature space.

a) The core idea behind such techniques is to cluster the frames based on some similarity measure and then select one key frame from each cluster. Yeung and Yeo [11] proposed a method to generate a pictorial summary of a video sequence which consists of a set of video posters, each representing a scene in the sequence. The key frames were extracted using a time-constrained clustering method which takes into account both visual properties and temporal locality of the shots. The video posters were generated by combining key frames based on their dominant scores assigned during the clustering phase. Zhuang et al. [19] presented a technique for key frame extraction based on unsupervised clustering using a color histogram as the visual content. A node is added to a cluster only if the similarity measure between the frame and the cluster centroid is greater than a certain threshold.

b) Doulamis et al. 9] presented a technique for summarizing stereoscopic videos which used clustering of shots to reduce redundancy. The clustering was performed based on the multidimensional fuzzy classification of segment features extracted from stereoscopic frames.

c) A motion based clustering algorithm was introduced by Zhang et al. [8] in which the clustering was done based on the motion compensation error.

d) Furini et al. [7] proposed a summarization technique called ‘‘STIMO’’ (ST Ill and MOving Video Storyboard) based on the clustering of HSV color descriptors. Avilaet al. [3] presented a method ‘‘VSUMM’’ (Video Summarization)which extracted color features from the frames after pre-sampling the frames from video. After removal of useless frames, the rest ofthe frames are clustered based on the k-means clustering algorithm.

The main advantage of clustering based methods is that they generate less redundant summaries as compared to the consecutive frame difference based techniques. The problem with most of the clustering methods (less time constrained clustering)is that temporal information of the frames is not considered. In other words, the key frames are selected regardless of the temporal order of each frame. Therefore, the key frames may not preserve the temporal visual sequence of the video .

III PROPOSED METHOD

In this paper keyframe extraction is done based on visual attention which will bridge gap between low level features and semantic of video. The existing visual attention models can be grouped into two categories: bottom-up, saliency-driven and task-independent models and top-down, volition-controlled and task dependent . the main steps for key frame extraction include: modeling visual attention, fusing attention models and clustering video frames models.

A) Dynamic attention Model

Motion is an intrinsic attribute of video and human eyes are very sensitive to moving objects. Based on this, we calculate the optical flow of each frame to detect the motion conspicuous regions. In the optical flow field, each motion vector is quantized by its magnitude and orientation.

i) Motion intensity

ii)Motion orientation consistency

B) Static Attention Model

The ‘‘focus of attention’’ in the scenes is not limited to motion regions. When viewers watch a video sequence, they are also attracted by the interesting static areas, such as the red signs on the road of gray background, the flames in the forest under green background. Therefore, there is a need for a static attention model to measure which information is important or interesting.

i)Color contrast

ii)Texture contrast

In this paper saliency value is computed of each frame from a video to generate an attention curve. In this section,we mainly address the issue about how to use time-constraint cluster algorithm to select the most attractive and informative key frames. The chronological order of frames is crucial for key frame extraction based on clustering. For example, if the feature vectors are similar in a few consecutive shots, they are clustered together. Then, it is possible that only the representative frames from part of shots are selected and others from the remaining shots are missed. Therefore, a time-constraint cluster algorithm is proposed to cluster the video sequence with a natural time ordering.

C) Different clustering algorithms are available like k-means ,time, constraint, fuzzy c-means Video is divided into frames .Feature extraction is done based on low level features ,color, texture as well as motion

The following figure gives the framework of key frame extraction using visual attention model both static and dynamic are considered .the steps are given in detail below the figure given below

D) Steps for Annoting Videos

Step 1:Video is divided into frames

Steps 2.Feature Extraction

For keyframe extraction

Step2.1: Dynamic and static conspicuity maps are constructed by computing motion intensity, orientation consistency, color and texture contrast.

Step2.2: Saliency Map is formed

Step2.3: Time constraint Algorithm is applied.

Step 3.Annotation is done

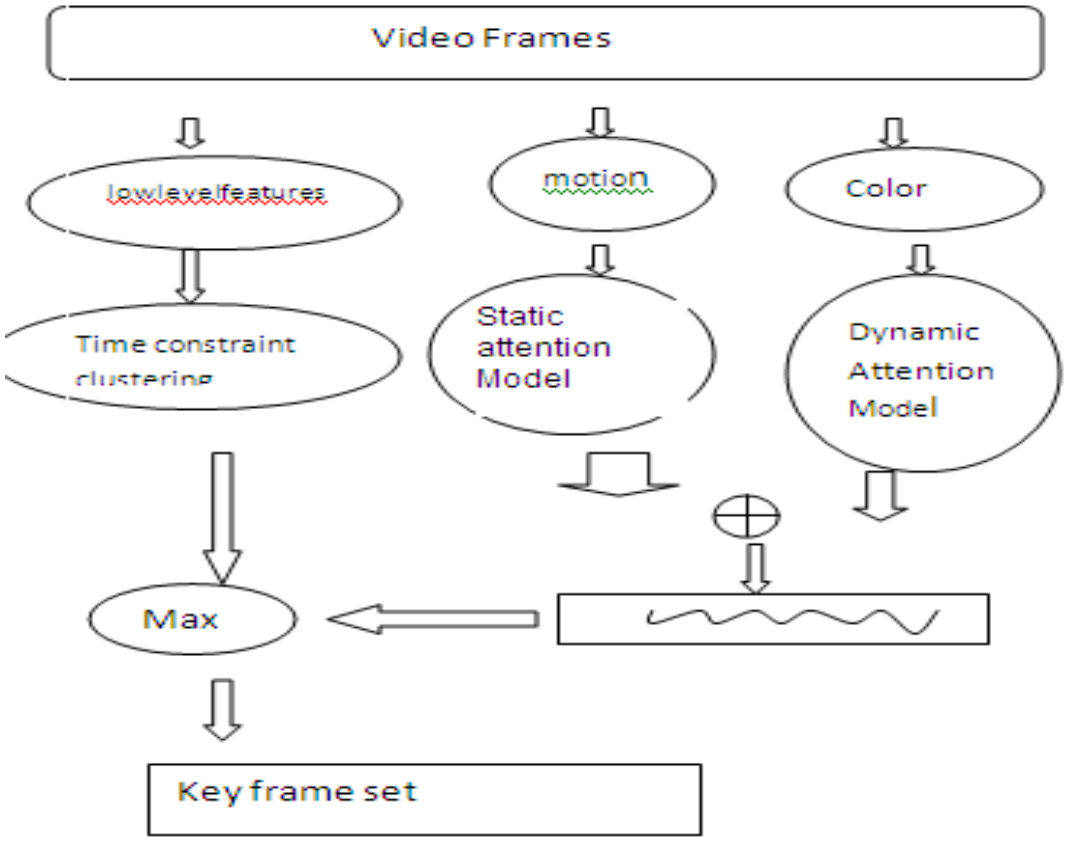

Figure: KeyFrame extraction by visual attention

CONCLUSION:

In this paper, novel framework for key frame extraction based on saliency-based visual attention modelis presented ,which is inspired by the human attention mechanism for information prioritizing and filtering. The algorithm for extracting keyframe and annoting is discussed.More experiments can be carried out for above approach.

REFERENCES:

[1] Amjad Altadmri and Amr Ahmed "Automatic Semantic Video Annotation in Wide Domain Videos Based on Similarity and Commonsense Knowledgebase's ,"in IEEE International Conference on Signal and Image Processing Applications ,2009.

[2] M. L. Shyu, Z. Xie, M. Chen, and S. C. Chen,"Video semantic event/concept detection using a subspace-based multimedia data mining framework," IEEE Transactions on Multimedia, vol. 10, no. 2, pp. 252–259, 2008.

[3] S.E.D. Avila, A.B.P. Lopes, L.J. Antonio, A.d.A. Araújo, VSUMM: a mechanism designed to produce static video summaries and a novel evaluation method, Pattern Recognition Letters 32 (1) (2011) 56–68.

[4] [4] Yuanfeng Yang1 Zhiming Cui1, Jian Wu1, , Guangming Zhang ,Xuefeng Xian1,"Traffic Video Segmentation and Key Frame Extraction UsingImproved Global K-Means Clustering",in Third International Symposium on Information Science and Engineering 2010 IEEE

[5] G. Ciocca, R. Schettini, Innovative algorithm for key frame extraction in video summarization, Journal of Real-Time Image Processing 1 (1) (2006) 69–88.

[6] B.T. Truong, S. Venkatesh, Video abstraction: a systematic review and classification, ACM Transactions Multimedia Computing, Communications and Applications. 3 (1) (2007).

[7] M. Furini, F. Geraci, M. Montangero, M. Pellegrini, STIMO: STIll and moving video storyboard for the web scenario, Multimedia Tools and Applications 46 (1) (2010) 47–69.

[8] X.D. Zhang, T.Y. Liu, K.T. Lo, J. Feng, Dynamic selection and effective compression of key frames for video abstraction, Pattern Recognition Letters 24 (9–10) (2003) 1523–1532.

[9] N.D. Doulamis, A.D. Doulamis, Y.S. Avrithis, K.S. Ntalianis, S.D. Kollias, Efficient summarization of stereoscopic video sequences, IEEE Transactions on Circuits and Systems for Video Technology 10 (4) (2011) 501–517.

[10] M.M. Yeung, B.-L. Yeo, Video visualization for compact presentation and fast browsing of pictorial content, IEEE Transactions on Circuits and Systems for Video Technology 7 (5) (1997) 771–785.

[11] R.M. Jiang, A.H. Sadka, D. Crooks (Eds.), Advances in video summarization andskimmingM. Grgic, K. Delac, M. Ghanbari (Eds.), Recent Advances in Multimedia Signal Processing and Communications, 231, Springer Berlin, Heidelberg, 2009, pp. 27–50.

[12] S.K. Pal, A.B. Leigh, Motion frame analysis and scene abstraction:discrimination ability of fuzziness measures, Journal of Intelligent and Fuzzy Systems 3 (1995) 247–256.

[13] A. Hanjalic, R.L. Langendijk, J. Biemond, A new key-frame allocation method for representing stored video-streams, in: 1st International Workshop on Image Databases and Multimedia, Search, 1996, pp. 67–74.

[14] D. DeMenthon, V. Kobla, D. Doermann, Video summarization by curve simplification, in: Proceedings of the ACM Internaternational Conference on Multimedia, NewYork, USA, 1998, pp. 211–218.

[15] A. Hanjalic, R.L. Langendijk, J. Biemond, A new method for key frame based video content representation, Image Databases and Multimedia Search, World Scientific, Singapore, 1998.

[16] S.H. Hoon, K. Yoon, I. Kweon, A new technique for shot detection and key frames selection in histogram space, in: 12th Workshop on Image Processing and Image Understanding, 2000, 217–220.

[17] L.M. Tieyan, X. Zhang, J. Feng, K.T. Lo, Shot reconstruction degree: a novel criterion for key frame selection, Pattern Recognition Letters 25 (12) (2004) 1451–1457.

[18] Y. Zhuang, Y. Rui, T.S. Huang, S. Mehrotra, Adaptive key frame extraction using unsupervised clustering, in: International Conference on Image Processing, 1998, 866-870.