International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

Preprocessing Step- Review of Key Frame Extraction Techniques for Object Detection in video

Aziz Makandar†, Daneshwari Mulimani†,* and Mahantesh Jevoor‡

†Department of Computer Science, Karnataka State Women's University, Bijapur, India
‡Department of Computer science, BLDEA's KCP science college Bijapur, India

Accepted 12 June 2015, Available online 17 June 2015, **Vol.5, No.3 (June 2015)**

Abstract

Videos are often constructed in the hierarchical fashion: [Frame]->[Shot]->[Scene]->[Video]. As a preprocessing step frames have to be extracted for object recognition, detection and then for tracking process. To do this, choosing an efficient technique to extract the Key Frames from the video is essential. This paper contains the brief representation and comparison of effective Key Frame Extraction(KFE) methods like cluster-base analysis, Generalized Gaussian density method(GGD), General-Purpose Graphical Processing Unit(GPGPU), Histogram difference, which results in high performance and more accuracy in extracting the key frames from the video.

Keywords: *KFE, Clustering, Discrete contour evaluation, MCMC, GGD, GPGPU, X^2.*

Introduction

Video is a collection of frames. Each frame represents the unique state of object status. Object detection from the frame and keep Tracking the particular object throughout video sequence became a challenging research area in computer vision system. Tracking is a problem of estimating the trajectory of an object in the image plane as it moves around a scene. Detection and Tracking of moving objects in a video sequence can be complex due to:

- Noise in images, • Complex object motion, • Non-rigid or articulated nature of objects,
- Partial and full object occlusions, • Complex object shapes, • Scene illumination changes, and
- Real-time processing requirements.

To overcome these problems, it is very essential to use an efficient KFE technique on video processing system. The Object detection and Tracking Follows the preprocessing steps as shown in the flow chart.

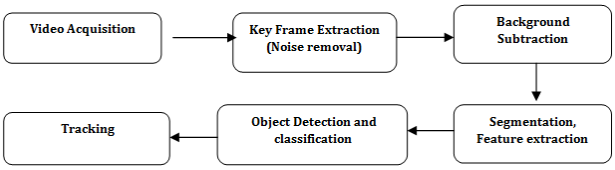

Figure 1: Steps for Object Tracking in video

*Corresponding author **Daneshwari Mulimani** is a Research Scholar; **Dr. Aziz Makandar** is working as Associate Professor and **Mahantesh Jevoor** as Assistant Professor

Key Frame Extraction

In an object detection process the preprocessing is done with extracting the Key Frame from the video sequence. Technologies for video segmentation and key-frame extraction have become crucial for the development of advanced digital video systems.

KFE plays an important role in many video processing applications such as video compression, retrieval, skimming, editing, etc. Each video sequence is a combination of shots". A shot is defined as an unbroken sequence of frames recorded from a single camera, which forms the building block of a video. KFE generally involves selecting one frame from each shot segment called cluster", which represents that video segment. A key frame should follow two main rules to adequately represent its cluster : 1) it should be similar enough to the frames in its cluster and 2) it should tolerably differ from frames in other clusters. In a typical KFE process, first, a set of features are extracted from each frame to form a feature vector. Next, a suitable distance measure is applied to the feature vectors to examine similarity/dissimilarity between frames. Finally, based on the distance measurement in the selected feature space, shot and cluster boundaries are detected and the key frames are extracted. These frames are very necessary for recognition of object and to keep track of that required object.

Many Techniques were introduced in the previous work for KFE, out of that we discuss more efficient techniques for future reference. Those are listed below with results.

1. X2(Chi Square) based Shot boundary detection method And { Histogram Difference Based}[Ganesh. I. Rathod *et al*,2013; Prajesh V. Kathiriya *et al*,2013**]:** SBD is the complete segmentation of a video into continuously imaged temporal video segments. Condensed video representation is the extraction of video frames or short clips that are semantically representative of the corresponding video. Both tasks are very significant for the organization of video data into more manageable forms. With the square histogram difference considered at block level for the video frames, a new method of extracting the keyframes based on shot type is presented. Shot boundary Detection task can be achieved using various approaches such as pixel intensity based, histogram-based, edge-based, and motion vectors based, are implemented and analyzed. Among all the approaches Histogram difference is the popular approach. Histogram based method is interested with the global percentage of colors that an image contains. This approach contains the following steps to extract the Key Frames.

a. Partitioning a frame into blocks with *m rows and n columns, and B(i, j,k) stands for the block at(i, j) in the kth frame.*

b. Computing x2 histogram matching difference between the corresponding blocks between consecutive frames in video sequence.

$$D_f(k,k+1) = \sum_{i=1}^{3} \frac{[H(k,i) - H(k+1,i)]^2}{H(k,i)}$$

c. Calculate the total difference for the total video and then calculate the mean difference(MD) using

$$MD = \sum_{k=1}^{N-1} \frac{D_f(k,k+1)}{N-1}$$

d. Compute standard variance STD of histogram difference over whole video sequence by using :

$$STD = \sqrt{\sum_{k=1}^{N-1} \frac{[D_f(k,k+1) - MD]^2}{N-1}}$$

e. Calculate the two threshold for two types of shots, i.e. Cut and Gradual Threshold by using : T = MD + STD × A where A is constant for A=1.

f. Determining ―Shot Type‖ according to the relationship between max(i) and MD:Static Shot(0) or Dynamic Shot:

$$ShotType_{C=}\begin{cases} 1 & if max(i) \geq MD \\ 0 & others \end{cases}$$

g. If $ShotType_C$=0, with respect to the odd number of a shot's frames, the frame in the middle of shot is chose as keyframe; in the case of the even number, any one frame between the two frames in the middle of shot can be chose as keyframe.

2. General Purpose Graphical Processing Unit(GPGPU)[B.Eng. Christian Kehl**]:** The GPU performance speed-up is used to compute functions for image manipulations and correction for each frame. These effects are separated into the two categories of image enhancement and artistic filters. After enhancing the picture quality, the artistic filters are applied. Graphic Processor Units (GPU)[K Wasif Mohiuddin *et al*,2011] have become popular on even personal systems and have tremendous compute power in them. Video segmentation is assisted by the GPU and the Pyramidal Histogram of Oriented Gradients (PHOG) computation is performed entirely on the GPU. The GPU also does bulk of the K-Means clustering to select representative vectors in the learning phase. In the categorization phase, the distance evaluation for the K-Nearest-Neighbor approach(K-NN) classifier for each frame is performed on the GPU. GPGPU is used for wide range of applications. GPUs are computationally powerful devices which are able to perform data intensive tasks quickly. It uses PHOG features and their clustered collection. Video frames are best distinguished based on objects present in the frames like pitch in the case of cricket, net in the case of tennis, etc. In the image, each region is represented as a Histogram Of Gradients (HOG), then K-Means clustering to group similar frames so that minor variations can be accounted. The final aggregation and decision making takes place on the CPU. This tool is able to organize a set of 100 sport videos of total duration of 1375 minutes in about 9.5 minutes. The labeling accuracy is about 96% on all videos.

3. Discrete contour evaluation: In [Janko Calic and Ebroul Izquierdo] introduces a real-time algorithm for scene change detection and key-frame extraction that generates the frame difference metrics by analyzing statistics of the macro-block features extracted from the MPEG compressed stream. The key-frame extraction method is implemented using difference metrics curve simplification by discrete contour evolution algorithm. This approach resulted in a fast and robust algorithm.

4. MCMC technique: In [Yun Zhai *et al*,2005] presented a general framework for temporal scene segmentation in various video domains. The proposed method is formulated in a statistical fashion and uses the Markov chain Monte Carlo (MCMC) technique to determine the boundaries between video scenes. In this approach, a set of arbitrary scene boundaries are initialized at random locations and are automatically updated using two types of updates: diffusion and jumps. Diffusion is the process of updating the boundaries between adjacent scenes. Jumps consist of two reversible operations: the merging of two scenes and the splitting of an existing scene. The posterior probability of the target distribution of the number of scenes and their corresponding boundary locations is computed based on the model priors and the data likelihood. The updates of the model parameters are

controlled by the hypothesis ratio test in the MCMC process, and the samples are collected to generate the final scene boundaries. The major advantage of this framework is two-fold: (1) it is able to find the weak boundaries as well as the strong boundaries, i.e., it does not rely on the fixed threshold; (2) it can be applied to different video domains. This method is tested on two video domains: Home Videos and Feature Films and accurate results have been obtained.

5. Autoregressive Parametric: In [W. Chen and Y.J. Zhang,2008] propose a parametric model for the video content analysis by using the autoregressive (AR) modeling to model the frame feature sequence over time and make the future analysis in the AR parametric space. Based on these parametric framework applications -detecting shot boundaries in video sequences, extracting key frames and combined spatial–temporal features for shot classification are proposed. Experiments show that, new parametric framework can present the video content better than the traditional color histogram.

6. Intra-frame and inter-frame motion histogram analysis: In [L. Shao and L. Ji,2009] A novel algorithm for key frame extraction based on intra-frame and inter-frame motion histogram analysis is proposed. The extracted key frames contain complex motion and are salient in respect to their neighboring frames, and can be used to represent actions and activities in video. The key frames are first initialized by finding peaks in the curve of entropy calculated on motion histograms in each video frame. The peaked entropies are then weighted by inter-frame saliency which we use histogram intersection to output final key frames. The effectiveness of the proposed method is validated by a large variety of real-life videos.

7. Clustered based analysis: In the key frame extraction method the old techniques like- 1) Shot boundary based approach , 2) Visual content based approach, 3)Motion analysis based approach and 4) Shot activity based approaches are used. The first two approaches to key frame extraction are relatively fast, however they do not effectively capture the visual content of the video shot since the first frame is not necessarily a key frame. The last two approaches are more sophisticated due to their analysis of motion and activity. However they are computationally expensive and their underlying assumption of local minima is not necessarily correct. To over-come these difficulties it is necessary to focus on base key frame selection on low level visual features such as color, texture, shape of the salient object in a shot. Then we use Clustering, a powerful technique in various disciplines such as Pattern Recognition. Speech Analysis and Information Retrieval.

In [Yueting Zhuang *et al*,2012] unsupervised clustering based approach was introduced to determine key frames within a shot boundary. Clustering based key frame extraction approach is not only efficient to compute, it also effectively captures the salient visual content of the video shots. For low activity shots it will extract less key frames or one single key frame at most of the time, while for high activity shots it will automatically extract multiple key frames depending on the visual complexity of the shot. The advantages of Clustering are Efficiency, Effectiveness, Online processing and Open framework.

8. Generalized Gaussian Density: By the motivation of these clustered-based analysis[Yueting Zhuang *et al*,2012; M. Omidyeganeh,2012], proposed new method Generalized Gaussian Density (GGD) parameters of wavelet transform sub-bands along with Kullback-Leibler distance (KLD) measurement. In an content analysis process appropriate features extraction like motion information, color histograms or features has to be employed from a 2D-transform of the video frame. However, none of the existing methods are structurally matched with the Human Visual System (HVS), which is the main sources of the error in locating the key frames. To clear this error, In GGD method, features are extracted from wavelet transformed sub-bands of each frame, leading to a better match with the HVS. GGD is used along with the KLD for texture retrieval in static images to get the excellent results. GGD parameters are helpful to construct the feature vectors and used the KLDs between two GGD feature vectors to examine similarity between two frames in a cluster and discriminate between two frames from different clusters. These results are more accurate and more subjective to the human section as compared with the previous methods.

Table 1 Comparative study of each frame extraction methods

Author	Technique/Method	Result
B. Janvier *et al*, 2006	Information-theoretic	Results poor accuracy rate (77.2%) in shoot boundary detection
W.L. Zhao *et al*,2007	cluster-based analysis	Results high complexity and lack of broad observations(fails in comparing similarities / dissimilarities).
Yueting Zhuang *et al*	Generalized Gaussian Density	Achieved high improvements than shot boundary detection and clustering techniques.
Janko Calic and Ebroul Izquierdo	macro-block features extraction by Discrete contour evaluation	Achieves 88% of precision
K Wasif Mohiuddin, P J Narayanan, 2011	General Purpose Graphical Processing Unit(GPGPU)	Achieves 96% accuracy on all video
Ganesh. I. Rathod *et al*, 2013 Prajesh V. Kathiriya *et al* , 2013	Shot Boundary Detection and Key Frame Extraction Using Histogram Difference & x^2 method	An efficiency of almost 95% to 98% is observed using this algorithm.

Conclusion

This paper addresses about most possible and effective frame extraction techniques. Each technique is explained briefly along with their performance results and feedbacks by comparing with each other. This paper concludes that in all the older methods Histogram difference and X^2(chi square) method is more fastest and valid technique to extract the frames from the any kind of video. Moreover it is tested on more than 100 high quality sports video and results high accuracy up to 98%. This paper will definitely help researchers to initiate preprocessing of KFE for object detection in video.

References

Janko Calic and Ebroul Izquierdo Efficient Key-Frame Extraction And Video Analysis, Multimedia and Vision Research Lab, Queen Mary, University of London

Yun Zhai *et al* (2005), Video Scene Segmentation Using Markov Chain Monte Carlo, IEEE transactions on multimedia, Vol. X, No. Y

W. Chen and Y.J. Zhang (2008), Parametric model for video content analysis, Elsevier B.V., Pattern Recognition Letters, vol. 29, pp. 181–191.

L. Shao and L. Ji (2009), Motion Histogram Analysis Based Key Frame Extraction for Human Action/Activity Representation, 6th Canadian Conference on Computer and Robot Vision (CRV), pp. 88 – 92

Yueting Zhuang *et al* (2012) Adaptive Key Frame Extraction Using Unsupervised Clustering, Urbana IL 61801, USA.

M. Omidyeganeh, S. Ghaemmaghami, S. Shirmohammadi (19 April 2011) Video Keyframe Analysis Using a Segment-Based Statistical Metric in a Visually Sensitive Parametric Space, IEEE Transactions on (Volume:20,Issue:10)

K Wasif Mohiuddin, P J Narayanan (2011) A GPU-Assisted Personal Video Organizing System, *cvit.iiit.ac.in/ papers/Wasif2011.*

B. Janvier, E. Bruno, T. Pun and S.M. Maillet (Sept. 2006), Information-theoretic temporal segmentation of video and applications: multiscale keyframes selection and shot boundaries detection, Multimedia tools and application, vol.3, no. 3, pp. 273-288.

W.L. Zhao, C.W. N go, H.K. Tan, X. Wu (2007), Near-Duplicate Keyframe Identification With Interest Point Matching and Pattern Learning, Multimedia, vol. 9, no. 5, pp. 1037-1048.

Ganesh. I. Rathod, Dipali. A. Nikam (August 2013) An Algorithm for Shot Boundary Detection and Key FrameExtraction Using Histogram Difference, International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com ,ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 8.

B.Eng. Christian Kehl B.Eng. Christian Froh Development of a GPGPU Video Encoding Server Application in a Multi-GPU environment Uinversity of Applied Sciences Technology, Business and Design, Wismar, 23952 Germany

Prajesh V. Kathiriya *et al* (January 2013), X2 (Chi-Square) Based Shot Boundary Detection and Key Frame Extraction for Video, Research Inventy: International Journal Of Engineering And Science Issn: 2278-4721, Vol. 2 Issue 2