ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 8, August 2019

Semantic Concept Detection in Videousing MPEG Features

Priya Kadam, Nita Patil, Sudhir Sawarkar

Dept. of Computer Engineering, Datta Meghe College of Engineering Navi Mumbai, University of Mumbai, India

Dept. of Computer Engineering, Datta Meghe College of Engineering Navi Mumbai, University of Mumbai, India

Dept. of Computer Engineering, Datta Meghe College of Engineering Navi Mumbai, University of Mumbai, India

ABSTRACT: The multimedia storage is increasing day by day. Also the cost to store these multimedia data is very less. Lots of videos available in the video warehouseare in unstructured format. As per user requirement, it is difficult to retrieve the relevant videos from such a huge video storage. Nowadays it is important to make such unstructured multimedia data easily available with flexibility.

In recent years, lots of research is going on to extract the semantic concepts from multimedia data.The main purpose of concept detection is to provide semantic concept based retrieval of multimedia content.

In previous content-based image and video retrieval systems, the retrieval was based on querying by examples and measuring the similarity of the database objects (images, video shots) with low-level features automatically extracted from the objects. The low-level features are insufficient to explain the content properly at conceptual level.The semantic gap characterizes the difference between two descriptions of an object by different linguistic representations, for instance languages or symbols. The semantic gap[14] can be defined as "the difference in meaning between constructs formed within different representation systems". This “semantic gap” isbasic problem in content-based multimedia retrieval system.

The main aim is to form semantic representations by extracting intermediate semantic levels (events, objects, locations, people, etc.) from low-level visual and audio features by using machine learning algorithms. In recent studies it is observed that the accuracy of the concept detector is far from perfection but still those detectors can be useful in high-level indexing and querying on multimedia data. This is possible because semantic concept detectors can be trained off-line with supervised learning algorithms and with positive and negative training examples which are available at query time.

Here, we are mentioned the description of our proposed system and its methodology for implementation of semantic concept detection. The main aim of this system is to improve the accuracy of concept detection.

KEYWORDS: Shot detection,Key frame extraction,Support vector machine;Concept Detection,Concept Correlation.

I. INTRODUCTION

In image processing, to detect high-level concepts within multimedia documents is the big problem.Lots of research is going on as amount of audiovisual content is increasing day by day and nowadays this is interest area for many researchers. The focus set mainly on the extraction of various low-level features i.e.audio, color, texture and shape properties of audiovisual content. Lot of new techniques such as neural networks, fuzzy logic systems and Support Vector Machines (SVM) used to find out high-level features from low level features. Lots of frameworks are already developed and every framework has their own result and accuracies. Here, purpose is to develop a system for concept detection with better accuracy.

General framework for video concept detection is shown in figure 1. Shot detection is the initial step in concept detection framework. It used to detect transition between successive shots.Key frame extraction plays a vital role to analyze the video content and its management. Key frame extraction gives the summary to retrieve video.A video shot

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 8, August 2019

is represented by a single image called as"key frame"and features are extracted locally and globally.A model vector is created and high-level concept detectors are trained using a global key frame annotation.

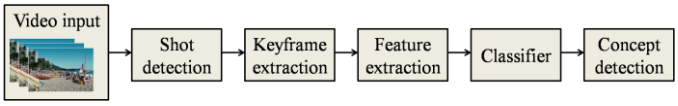

Figure 1 Semantic Concept detection framework

II. LITERATURE SURVEY

Lot of research is already done for analysis of video composition.

Shot Detection:Shot segmentation is the initial step of the concept detection framework and comes before the key frame extraction. In this shot segmentation,detection of the transition between successive shots takes place. Following are the common methods of shot detection:

1. Pixel based comparison
2. Template Matching
3. Color Histogram

To segment the shots as per the frame difference, color histogram method is accepted. The histogram based method is the most common method to calculate the frame difference. The color histogram do not relate spatial information with the pixels of a given color and only records the amount of color information, images with similar color histograms can have different appearances. To solve this issue, an improved histogram algorithm and histogram matching methods can be used [1].

Alan Hanjalic et al. [2] developed a statistical shot-boundary detector. This detector fuses the range information and a priori information which gives an adaptive threshold as a result and provides optimal detection performance. Mr. Hattarge A.M. et al.[3] explains an adaptive approach that combines different features of video and calculates a unique feature which is used to identify the hard cuts and gradual transitions. Yuchou Chang D.et al. [4]for video shot detection the author explains both a color feature extraction algorithm and clustering approach.Here in his experiment gives result that the proposed video shot detection has better performance than k-means single-clustering and the SIFT descriptor. Zeeshan Rasheed et al.[5] represented a high-level segmentation of videos into scenes.To cluster shots into scene, a weighted undirected graph also called as a shot similarity graph (SSG)is used. This method used to describe the content of each scene by selecting one scene key-frame as a representative image from the video Yun Zhai et al.[6]proposed a general framework for temporal scene segmentation in many video domains.The proposed method uses the Markov chain Monte Carlo (MCMC) technique to calculate the boundaries between video scenes.To achieve the correct result,the proposed method is tested for two video domains: Home Videos and Feature Films.

Key-frame Extraction:A key frame is a location on a timeline which marks the beginning or end of a transition. It holds special information that defines where a transition should start or stop. Key frames forms a compact expression of a video clip. It provides video summary to retrieve video. Key frame extraction has following methods:

1. Key frame extraction based on shot activity
2. Key frame extraction based on macro block statistical characteristics of MPEG video stream
3. Key frame extraction based on edge histogram descriptor

Here, the MPEG video compression algorithm is used. In the compression of video stream, group of frames called as GOP i.e. Group of Pictures.
Guozhu Liu et al.[1]proposed new algorithm for key frame extraction. This new algorithm reduces the constraints of other algorithms. Based on MPEG video stream,it also improves the techniques of key frame extraction.Janko Calic et al.[7]proposed novel key-frame extraction technique based on the difference metrics extracted directly from the MPEG

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 8, August 2019

compressed domain and discrete contour evolution. Pascal Kelm et al.[8]compared a key frame extraction approach against the key frame extraction of IBM Multimedia Analysis and Retrieval System. To retrieve the most stable frame,this approach extracts the key frame with the great number of visual attention and smallest amount of motion intensity. J. Calic et al.[9]explained an efficient algorithm for key-frame extraction which analyses behavior of spatio-temporal regions present in a video sequence .Experimental result gives a precise representation of a shot by maintaining the its real time processing ability.Andreas Girgensohn et al.[10] hierarchical clustering algorithm is used. To summarize videos and provide access points to key frames,the key frames used to recognize videos from each other. Temporal limitations are used to separate out some clusters and to find out the representative frame for a cluster.

Feature Extraction:The MPEG-7 descriptors are selected to find out the low level features of each region. Following are the MPEG7 color descriptors which are extracted to find out the color characteristics:

1. Dominant color descriptor
2. Color structure descriptor
3. Color layout descriptor
4. Scalable color descriptor

Following are the MPEG7 texture descriptors used to extract the texture properties:

5. Homogeneous texture descriptor
6. Edge histogram descriptor [11].

Sylvie Jeannin et al.[11]The MPEG-7 provides fully logical and useful collection of motion descriptors. It captures the various features of motion in the videos. In this the main aim is to provide brief descriptors which are easy to extract and match. In this paper how motion activity and motion trajectory both fulfill their aim is mentioned.
Jens-Rainer Ohm et al.[12]In this paper,a set of color descriptors explained which are ableto capture the important aspects of color feature and allows color similarity computations. All these descriptors are compact and hence allows efficient description of color properties. The application area of MPEG-7 color descriptorsare where there is a need of color similarity between the objects.

Swapnalini Pattanaik et al.[13]in her work explained in detail about idea of how to retrieve images using different Mpeg-7 features. Two main steps of CBIR system areFeature extraction and Similarity matching are also mentioned .The main aim of Mpeg-7 is to provide a set of technologies to describe multimedia content. Here, Color Structure Descriptor for color and Edge Histogram Descriptor for texture are explained. Both features fused to enhancethe performance of CBIR systems[13].

***Concept Detection*:**The SVM are supervised learning models with associated learning algorithm that analyze the data used for classification by constructing N-dimensional hyperplane that separates the data into separate two categories. The SVM models are closely related to neural networks. A SVM model using sigmoidal kernel function is parallel t two layer perceptron neural network.

Evaggelos Spyrou et al.[14] in his work performed low level feature extraction, color features and texture features extraction. For color property, MPEG-7 dominant color descriptor is used.For the texture properties the actual MPEG-7 homogeneous texture descriptor is used. For region thesaurus, Subtractive clustering is used. After this a model vector is formed. The SVM are nothing but supervised learning models having associative learning algorithm which can perform data analysis for classification and regression analysis purpose.

Evaggelos Spyrou, et al.[15] in his another work for concept detection used RSST segmentation algorithm for low level feature extraction. Moreover,for the training set of images a K-means clustering algorithm is used on the feature vectors of the regions.A model vector is formed for the representation of the semantics of a key frame which are based on the set of region types, a model vector is constructed.For each high level concept,aseparate neural network based detector is trained.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 8, August 2019

Lin Lin et al.[16]here,in this work associative classification algorithm is proposed by using MCA for video concept detection. MCA is used to measure the correlation between various feature value pairs and classes to conclude the high-level concepts from the extracted low-level features. Nine concepts from the TRECVID 2007 and 2008 data are used to validate the proposed framework.

Nita S.Patil et al.[17] proposed a framework for multi-label semantic concept detection. In this framework classifiers trained on global and deep features. Also, Dataset is partitioned into three segments so as to solve the imbalance dataset issue the evaluation of this framework done on TRECVID dataset using mean average precision as predictive measure for multilabel dataset. Only CNN work better for top concepts. Both Hybrid model of CNN and SVM performed better than individual classifier for top 10 concepts.

Concept Correlation:Concept correlation refers to the relations among concepts within a video shot. By using this information it is possible to refine the predictions**.** Once features have been determined, correlation is used to find the matching of image features or to find image motion.
Correlation between concepts are determined to improve the detection scores of classifier.To exploit concept correlation,two main types of methods.Stacking based approach collects the scores generated by concept detector and in second learning step that score is refined.Inner-learning approach followsonly single step learning process which combinely uses low level visual features and concept correlation information.[18].
Foteini Markatopoulou et al. [18]used color extensions of scale invariant features transform applicable to both binary and non-binary local descriptors and pca for compaction used multilevel classification algorithm for concept correlation and the last layer of the stack. Ken-Hao Lie et al.[19]n his paperto find out the relationship between targeted concepts and remaining other semantic concepts, MCA i.e. .multiple correspondence analysis is used.These such arelationships are utilized as transaction weight and moreover to refine detection ranking scores.
Lin Lin[20] proposed a multimodal correlation based video semantic concept detection framework by using MCA. The TRECVID 2007 video data is used for validation of the detection performance of this framework. To check the detection Vegetation, Urban, Crowd, Office, Road, Sky, Waterscape, Building, Road, Outdoor, Faces concepts are used. To estimate the correlation between various items and classes, functionality of MCA is used. Hence it is possible to conclude the high level concepts from the extracted low level features.

III. PROPOSED METHOD

Figure 2 shows framework for proposed concept detection based on Mpeg features.

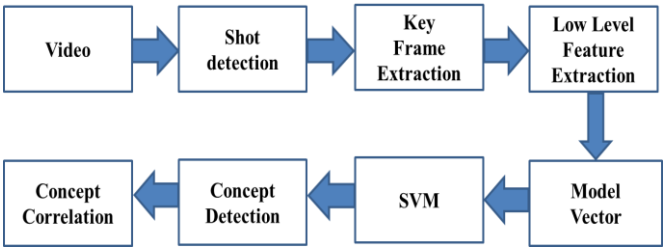

Figure 2. Proposed Semantic Concept Detection using Mpeg features

The proposed method has following steps. Initial step is shot segmentation and key frame extraction from video. The next step is low level feature extraction to produce model vector for classification. The next step is classification using SVM. Next step is to find correlation between the concepts. Correlation between concepts isused to improve the detection scores of classifier.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 8, August 2019

A. Shot segmentation:The automated detection of transition between shots in digital video with the purpose of temporal segmentation of videos.

B. Key Frame Extraction: Following steps explains the key frame extraction method.
The types of frames I, P and B frame arranged in video stream. I frame is the first frame. I frame is intra coded frame. I and P frames are the reference frame. These frames processed with DCT(Discrete cosine transform)by using 8*8 block .DC coefficient contains main information. P and B frames are inter frame coded. P frame refer to the preceding I frame or P frame. Based on macro blocks, these are coded with forward motion compensation. The forward motion vectors for forward motion predictions are obtained. Also DCT coefficients of residual error after motion compensation are obtained.
B frames are inter frame coded for forward motion prediction, backward motion prediction and bi-directional motion prediction. Every macro block of 16*16 pixels in P and B frames search for the optimal matching macro block in reference frame then reduce predictive error of motion compensation with DCT coding. One or two motion vectors are transferred at the same time. After a shot segmentation, Key frames are extracted using features of I,P and B frames in MPEG video stream. If scene cuts occurs then first I frame will be chosen as a key frame. [1]

C. Low-Level Feature Extraction:TheMPEG-7 descriptors are selected to find out the low level features of each region. Following are the MPEG7 color descriptors which are extracted to find out the color characteristics:

1. Dominant color descriptor
2. Color structure descriptor
3. Color layout descriptor
4. Scalable color descriptor

AS the name indicates, dominant color descriptor allows specification of small number of dominant color values and its characteristics like variance and distribution over a region. The main aim is to provide a powerful, dense and instinctive representation of colors present in an image.
The color structure descriptor is based on color histogram. It uses small structuring window to find out the local color distribution. This descriptor is attached to the Hue-Min-Max-Difference color space .i.e. HMMD color space.
The Color layout descriptor holds the spatial layout of the dominant colors on a grid which is superimposed on an image or on region. Its representation is based on coefficient of DCT(Discrete cosine transform).This is a very compact and efficient descriptor and highly used in application of fast browsing and search application. It can be applied on still images as well as on video segments.

The Scalable color descriptor is obtained from a color histogram. This histogram is defined in HSV i.e. Hue–Saturation-Value color space. It uses Haar transform coefficient encoding which allows scalable representation of description and scalability of feature extraction and matching procedure.

Following are the MPEG7 texture descriptors used to extract the texture properties :

1. Homogeneous texture descriptor
2. Edge histogram descriptor[11].

The HTD gives quantitative characterization of texture for similarity based image to image matching. This descriptor is computed by first filtering the image with bank orientation
This descriptor is computed by first filtering the image with a bank orientation and scale sensitive filters .Then computes the mean and the standard deviation of output in frequency domain.
The edge histogram descriptor records the spatial distribution of edges. The distribution of edges is a good texture sign which is useful for image matching.

D. Model Vectors: A model vector is constructed to represent the semantics of a key frame which are based on the set of region types.All the features are joined to form the single feature vector.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 8, August 2019

E.Support vector machine:A set of features which describes one case is called a vector. The goal of SVM modeling is to find the optimal hyperplane that separates clusters of vector in such a way that cases with one category of the target variable are on one side of the plane and cases with the other category are on the other size of the plane.The vectors near the hyperplane are the support vectors.

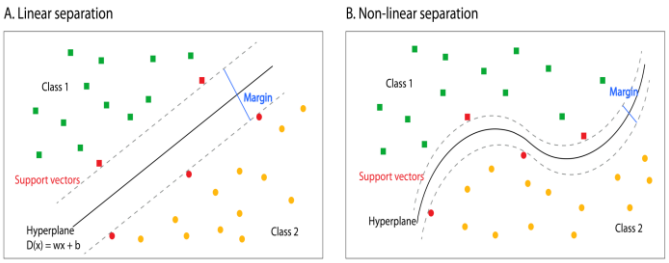

Fig 3: Linear separation and Non-linear separation

The hyperplane acts as adecision space in which margin of separation between positive and negative example can be maximized.For the detection of semantic concepts ,SVMcan be used for image classification based on its histogram.A separate support vector machine is to be used for each semantic concept which solves the binary problem. The input of the SVM is taken from the model vector[15].

F.Correlation between the concepts:The semantic concept and the temporal context are different. Both are important for visual concepts detection and based on this assumption we propose to combine the two approaches: "concepts" and "temporal context". We will use a Fusion method for merging results of both approaches by averaging the concept detection scores for each shot "concept " and the "temporal context" approaches[18].

IV. CONCLUSIONS

We presented a review on various steps of concept detection framework. We discussed framework for semantic concept detection in video that has two layers for concept detection in which we focused on visual features which are extracted from compressed domain of video for concept detection and correlation between concepts. Correlation between concepts are determined to improve the detection scores of classifier. Using fusion we can improve relevancy score. Hence detecting concept and finding their correlation can improves video concept detection.

REFERENCES

[1] Guozhu Liu and Junming Zhao." Key frame extraction from MPEG video stream"Proceedings of the Second Symposium International Computer Science and Computational Technology(ISCSCT '09)Huangshan, P. R. China, 26-28,Dec. 2009.
[2] Alan Hanjalic," Shot-Boundary Detection: Unraveled and Resolved?"IEEE transactions on circuits and systems for video technology, vol. 12, no. 2, February 2002
[3] Mr. Hattarge A.M., Prof. K.S. Thakre,"Analysis and Review of Formal Approaches toAutomatic Video Shot Boundary Detection",international Journal of Advanced Research in Computer and Communication EngineeringVol. 3, Issue 1, January 2014.
[4] X. Ling, L. Chao, L. Huan, and X. Zhang, "A General Method for Shot Boundary Detection," 2008 International Conference on Multimedia and Ubiquitous Engineering (mue 2008), pp. 394–397, 2008.
[5] Zeeshan Rasheed and Mubarak Shah, "Detection and Representation of Scenes in Videos",IEEE transactions on multimedia, vol. 7, no. 6, December 2005
[6] Yun Zhai and Mubarak Shah,"Video Scene Segmentation Using Markov ChainMonte Carlo",IEEE transactions on multimedia,VOL. X, NO. Y, 2005
[7] Janko Calic and Ebroul Izquierdo,"Efficient key-frame extraction and video analysis",2002
[8] Pascal Kelm, Sebastian Schmiedeke, and Thomas Sikora,"Feature-based video key frame extraction for low quality video sequences",2009
[9] J. Calic, B. T. Thomas,"Spatial analysis in key-frame extractionusing video segmentation".
[10] Andreas Girgensohn John Boreczky Fx Palo Alto,"Time-Constrained Keyframe Selection Technique",Multimedia Tools and Applications, 11, 347-358, 2000

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 8, August 2019

[11] Sylvie Jeannin and Ajay Divakaran ,"MPEG-7 Visual Motion Descriptors", IEEE transactions on circuits and systems for video technology, vol. 11, no. 6, June 2001.
[12] Jens-Rainer Ohm,Leszek Cieplinski ,Heon Jun Kim ,Santhana Krishnamachari ,B. S. Manjunath ,Dean S. Messing ,Akio Yamada ," The MPEG-7 Color Descriptors".
[13] Swapnalini Pattanaik, D.G. Bhalke,"Efficient Content based Image Retrieval System using Mpeg-7 Features",International Journal of Computer Applications (0975 – 8887)Volume 53– No.5, September 2012
[14] Evaggelos Spyrou and Yannis Avrithis,"High-Level Concept Detection in Video Using a Region Thesaurus", 2009.
[15] Evaggelos Spyrou, Giorgos Tolias, and Yannis Avrithis,"Large Scale Concept Detection in Video Using a Region Thesaurus" Image, Video and Multimedia Systems Laboratory, School of Electrical and Computer Engineering National Technical University of Athens 9 Iroon Polytechniou Str., 157 80 Athens, Greece,2009.
[16] Lin Lin, Mei-Ling Shyu, Guy Ravitz,Shu-Ching Chen"video semantic concept detection via associative classification"
[17] Nita S.Patil, Sudhir Sawarkar ,"Semantic Concept Detection in Video Using Hybrid Model of CNN and SVM Classifiers"
[18] Foteini Markatopoulou,Vasileios Mezaris,Nikiforas Pittaras,Loannis Patras, "Local features and two layer stacking architecture for semantic concept detection in video"IEEE Trans. on Emerging Topics in Computing, vol. 3, no. 2, pp. 193-204, June 2015.
[19] Ken-Hao Lie,"Correlation based re-ranking for semantic concept detection"2014.
[20] Lin Lin, Guy Ravitz, Mei-Ling Shyu,Shu-Ching Chen,"Correlation-based Video Semantic Concept Detection using Multiple Correspondence Analysis",Tenth IEEE International Symposium on Multimedia