Tomograph Based Concept Detection of Video Data by means of Particle Swarm Optimization

Priyanka Lall, Prof. Swati Sorte

Abstract—In video processing, concept detection is an important research area. Many techniques already exists to improve concept detection method.. One such technique is proposed in this work in which spatio-temporal video slices or called as video tomographs are implemented. The video tomographs can be a replacement for key frames. Video tomographs are capable of capturing motion information which was not the case with the previous methods use.Feature extraction is implemented. To assemble all this particle swarm optimization technique is used . The aim of the work is to implement a system capable of, extracting features, detecting motion or non- motion related concepts and optimizing for better outcome.

***Index Terms*— video analysis, supervised learning, video tomograph, concept detection, particle swarm optimization algorithm**

INTRODUCTION

In image and video processing community, vital goal is to develop techniques that would assist in empathizing the uncertainties of video data sequence .Accurate and efficient concept detection is highly desirable.The proficient and successful detection of concepts depends on visual content is a challenging yet vital setback.

In the last few years the research community has been targeting on multiple video concept detection, that is, the advancement of systems that is capable of handling large amounts of video data and detects multiple concepts efficiently. As a result, number of powerful techniques have surfaced, which aim to combine high precision and low computational cost.

***Priyanka Lall**, Department of Electronics Engineering,, G. H. Raisoni College of Engineering, Nagpur, India, 9561506838.*

***Prof. Swati Sorte**, Department of Electronics Engineering, G. H. Raisoni College of Engineering, Nagpur, India.*

So far the existing schemes for concept detection do not take into account motion information. Normally, they are based on plain set of distinctive key frames which are obtained from the selected video frames. The method is intricate as every frame is treated to feature extraction. Disregarding motion information discards the active nature of video. Motion descriptors are another option but have comparatively more computational cost .Due to all the drawbacks mentioned tomographs offer a better alternative. Not only tomographs capture motion information but also have the computational cost as the key frames.

In the proposed work, the video tomographs are used to represent video sample. These tomographs are spatiotemporal slices having one axis in time and another in space. Tomographs extraction is similar to their visual counterparts... A video tomograph is defined as a cross-section image (i.e. an image defined by the intersection between a plane and the video volume) which is additionally smoothed using a high-pass filter.

Tomgraphs are similarly used as key frames for concept detection. More particularly, image patches are approximated, followed by descriptor extraction and vector quantization. The classifiers used are tomograph in nature, each one independently trained for particular concept using interpreted samples taken from tomographs of the respective type.

LITERATURE SURVEY

In the literature, authors have employed number of algorithms for concept detection. They also have presented with different machine learning algorithm.

In [1] authors have presented with concept detection in large scale video database. They have used combination of tomograph and key frame based technique for incrementing the detection results

Additionally, the work does not support the concept of one fits for all which eventually lead in the increment of computational cost along with compromise in accuracy. This is because of the fact that it lead to pointless selection of base detector for required concepts. The authors have come up with genetic algorithm to solve this problem .The algorithm lead to the selection of optimal detectors.

SIFT (Scale Invariant feature transform) and SURF (Speeded up Robust Features) algorithms are used for feature extraction.

However, in [2] authors have suggested a better option of linear machine algorithm which is parallel lasso. Parallel lasso is distributed machine learning algorithm. Due to the disadvantages like low scalability and robustness, parallel or distributed machine learning are used. The authors have made use of PICF (parallel incomplete cholosky

factorization) to estimate the co-variance states and thus reduce the space and time complication. The main objective of parallel lasso is to differentiate computational data among range of clusters thus making it possible to parallelize the time and space perception.

In concept detection it is important to take care of foreground and background of the video samples. The same is explained in [3].Whereas in [3] authors have stated a new method of classification based on concept codebooks which are built upon Multiple Instance Learning. The features are represented as in video tracking. The authors in their work, first extracted sample points randomly from selected video shots. Spatial temporal features are extracted which are then applied to classifiers based on MIL. Obtaining the tracks is done by image segmentation which is rather costly .Optical Flow algorithm is used to make it somewhat efficient. In this paper, large numbers of concept of interest are sampled and the obtained patches are observed. Later are quantized based on concept codebooks.

So far ranges of methods are available for concept detection. In paper [4], the relation or in simple terms the similarity between the concepts is studied in a low-level feature samples by using clustering-based method. The samples taken are evaluated by entropy-based schemes. The work defines the significance of 'gap' between different concepts which falls under the same category. Trained data provides information about independent models of distinctive concepts and are labeled under larger ontology. Clustering plays an important role in this work, as it organizes number of concepts and put them under a category which consists of similar concepts. Clustering is generalized as vector quantization, provides a codeword and hence a more compact input data is generated. The model proposed is probabilistic. The entropy based model provides visual similarities between concepts that are more relevant in nature.

Typical key frames extraction and SVM is used in [5], however the main focus of authors in this work is to study camera motion because it relates to the interest of camera person and the viewer. Camera motion can also determine the region of interest which can be helpful in finding concepts. Colour information is ignored. Camera motion exploits temporal segmentation leading to key frame selection. Human attention model is implemented for capturing saliency information. In the study, authors have found high reliability between distribution of points recorded by eye movements and direction of camera motion.

Another paper [6] focuses on detection of topical objects i.e. the most highlighted object of a video shot. The approach used by authors has successfully detected multiple topical objects in a video. The authors have suggested that LDA (Latent Dirichlet Allocation) cannot be single handedly because of high computational cost. Instead a combination LDA-WCP can be beneficial because variant shapes and appearance can be detected and multiple objects can be detected simultaneously. In this method, key frames are extracted each key frame consists of a visual features. Further they are subjected to segmentation which is done at multiple resolutions. Clustering is performed and after that bag-of-word representation is obtained. Word co-occurrence prior parameter is obtained by studying word co-occurrence information.

Semantic gap and rare/event concept detection are some of the challenges that are faced during the concept detection. However, in paper [7] the authors have managed to solve. The initial step includes feature extraction from the highlighted video shots. The result of which is subjected to distance –based data mining consists of positive and negative samples. As in video number of irrelevant information is also present which is nullified by this technique and gives accurate result which is then subjected to filtering and reconstruction. C4.5 decision model is used for final concept detection.

In [8] the authors have studied method that could possibly bridge the gap between high-level and low-level concepts. The authors here studied wide range of concepts to detect video retrieval accuracy. They noticed that frequent concepts increase the concept detection accuracy while rare concepts does not hold good for the same case. They found that rare concepts can detect only 1% of the video collection while frequent concepts can detect 90% of them. Authors also suggested several automatic methods to find the best combination of concepts that relevant to the queried concept. Divide and conquer approach plays an intermediate role to develop several concepts that will bridge the semantic gap.

The paper [9] illustrates that invariance property is not sufficient enough, the descriptor has to be characteristics and robust. The SIFT based descriptors have proved to be more efficient and robust than most of the histogram based descriptors in the image and video category. SIFT descriptors are invariant to light intensity alterations, which is very useful in actual world applications. In the paper, performance evaluation of various kind of SIFT descriptors like opponent-SIFT, C-SIFT and RGB-SIFT were made on the grounds of illumination changes, color moments and moment invariant and histogram. In [10], the authors have compared different descriptors based on their performance. The descriptor like SIFT, PCA-SIFT and GLOH are compared on the grounds of image rotation, affine transformation, image blur, illumination changes and matching samples. Based on the result, GLOH has outperformed the other descriptors in most of the performances but SIFT proved to be more robust

PROPOSED SYSTEM

The proposed methodology will use video tomographs for selection of desired video shots in a unstructured video data set. SIFT algorithm for feature extraction. The proposed method uses the MATLAB Release 2013 version With the existing diversity of the concepts that need to be detected, and of the possible repeatability in a typical over-complete shot representation, PSO (Particle Swarm Optimization) algorithm is proposed .The algorithm selects concept autonomously the respective optimal base detector subset, instead of selecting all the possible base detectors for all concepts.

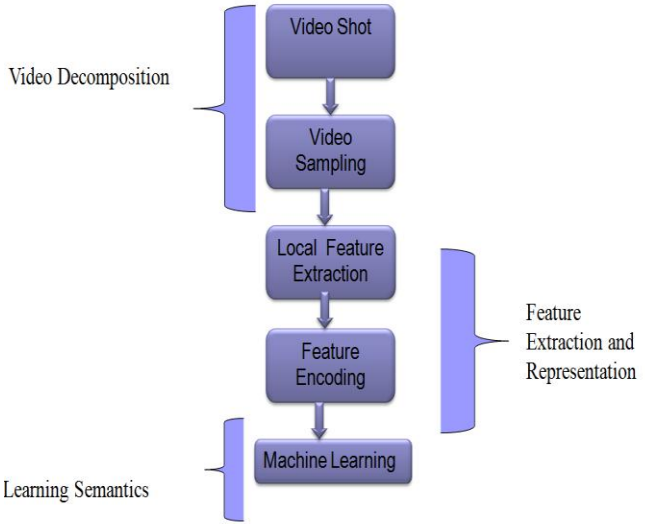
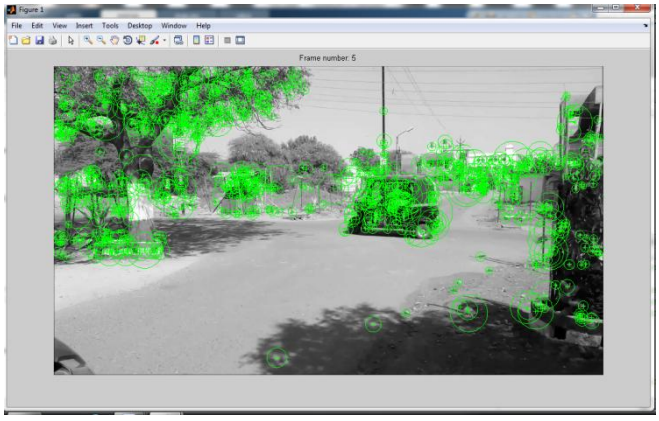


Figure 1. Flowchart of purposed work

The project is broadly divided into three modules.In module 1, as video tomography is used as replacement of key frames not all frames of video shots are decomposed. Only the frames consisting of concept are selected. As tomography is used , motion information can be contained in the further analysis. From the selected frame, features are extracted. Module 2 consists of feature extraction. As the main focus of the work is determination of concept SIFT algorithm is used. SIFT algorithm gives better result when it comes to detect key points. In module 3, the results are classified and are compared with trained data. Thus the compared result will provide with concept score. By using Particle Swarm Optimization (PSO) better convergence time and higher accuracy can be achieved.



By using the frame difference, the frames containing the motion can be determined. the same is determined in the figure 2. Only the frames containing motion will be surpassed for further analysis.

SIFT algorithm is used for determining the number of key points in a particular image .The key points being extracted is trained and is classified. The result of which is then subjected for the comparison with video data set.

CONCLUSION

The methodology will do concept detection in unstructured video by using video tomographs. As in case of concept detection, less work is done when it comes to unstructured video. For the purpose of feature extraction SIFT algorithm is proposed to be used. SIFT has the maximum key point detection capability.

Particle swarm optimization technique is proposed to be used for accurate results .The time required for optimization will be reduced and better convergence can be achieved. It is expected that with PSO better results can be obtained as compared to other evolutionary algorithm.

V. REFERENCES

[1] Panagiotis Sidiropoulos Ioannis (Yiannis) Kompatsiaris "Video Tomographs and a Base Detector Selection Strategy for Improving Large-Scale Video Concept Detection" IEEE Transactions on Circuits and Systems for Video Technology · July 2014.

[2] Bo Geng, , Yangxi Li, Dacheng Tao, , Meng Wang, ,Zheng-Jun Zha, , and Chao Xu "Parallel Lasso for Large-Scale Video Concept Detection" IEEE transactions on multimedia, vol. 14, no. 1, FebruarY 2012.

[3] Anjun Wei, Yuru Pei, Hongbin Zha "Random-Sampling-Based Spatial-Temporal Feature For Consumer Video Concept Classification" 978-1-4673-2533-2/12/$26.00 ©2012 IEEE.

[4] Golnaz Abdollahian, , Cuneyt M. Taskiran Zygmunt Pizlo, and Edward J. Delp "Camera Motion-Based Analysis of User Generated Video" IEEE transactions on multimedia, vol. 12, no. 1, January 2010.

[5] Markus Koskela, , Alan F. Smeaton, , and Jorma Laaksonen "Measuring Concept Similarities in in Multimedia Ontologies: Analysis and Evaluations" IEEE transactions on multimedia, vol. 9, no. 5, August 2007 January 2010.

[6] Gangqiang Zhao, Junsong Yuan, Gang Hua, and Jiong Yang "Topical Video Object Discovery From Key Frames by Modeling Word Co-Occurrence Prior" IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 24, NO. 12, December 2015.

[7] Mei-Ling Shyu, Zongxing Xie, Min Chen, *Member* and Shu-Ching Chen "Video Semantic Event/Concept Detection Using a Subspace-Based Multimedia Data Mining Framework" IEEE transactions on multimedia, vol. 10, no. 2, February 2008.

[8] Alexander Hauptmann, Rong Yan, Wei-Hao Lin, Michael Christel, and Howard Wactlar "Can High-Level Concepts Fill the Semantic Gap in video Retrieval? A Case Study With Broadcast News" IEEE transactions on multimedia, vol. 9, no. 5, August 2007.

[9] Koen E.A. van de Sande, Theo Gevers, Member, and Cees G.M. Snoek "Evaluating Color Descriptors for Object and Scene Recognition" 1582 IEEE transactions on pattern analysis and machine intelligence, vol. 32, no. 9, September 2010.

[10] Krystian Mikolajczyk and Cordelia Schmid "A Performance Evaluation of Local Descriptors" IEEE transactions on pattern analysis and machine intelligence, vol. 27, no. 10, October 2005.