Journal of Innovative Engineering and Research (JIER)
Vol.- 1,Issue - 1, April 2018, pp. 17-21 (5 pages)

Retrieval of Semantic Video Actions and Events Using GA and Optimize SVM

RASANA SHARMA[1], MANMOHAN SINGH[2], RAHUL SHARMA[3]

[1,2]Department of Computer Science & Engineering, Dr. A. P. J. Abdul Kalam University, Indore
Corresponding Author Email: rasana.sharm4@gmail.com

Abstract*— Video is quickly attractive due to its elevated information and amusement capability. It consists of audio, images and text together. A video database encloses collection of semantic information, the semantic information illustrates what is incidence in the video and furthermore what is evident by human user's .Video retrieval enables the user to search for particular video segment based on some description commercial increase. Other researchers have studied optimization of support vector machine using genetic algorithms through feature subset and by combining these two used this technique for Image identification. Feature selection select frame and fitness function compare each frame and select according to fitness of that frame. To identify the object we use motion vector for storing change from image to frame and camera motion feature are captured direction speed of moving object.

Index Terms*— Support Vector Machine, Genetic Algorithm, Motion Estimation, Motion Vector.

I. INTRODUCTION

Video is quickly attractive due to its elevated information and amusement capability. It consists of audio, video and text together. Verdict preferred information in a video clip or in a video record is complicated due to its semantic gap between the low-level characteristic A video database encloses collection of semantic information, the semantic information illustrates what is incidence in the video and furthermore what is evident by human users. Video browsing is very important functionality of multimedia systems which offer the user efficient way to view relevant information from large amount of video material. On the other hand, video retrieval enables the user to search for particular video segment based on some description commercial increase and SVM include for feature selection and ensemble classification. Other researchers have studied optimization of support vector machine using genetic algorithms through feature subset and by combining these two used this technique identification. Illustration SVM collect feature from entity frame, it generate and collect data set and GA identify object using genetic algorithm techniques feature selection and fitness function. Feature selection select frame and fitness function compare each frame and select according to fitness of that frame. To identify the object we use motion vector for storing change from image to frame and camera motion feature are captured direction speed of moving object. Propose framework for video event recognition is motivated by the following observations. First, video can be recognized as a set of fames; and it contain global feature, local feature and motion feature. Second, it is hard for people to decide which features are good for describing an event in a video. Third, in most existing work the event detection algorithms are embedded in systems and cannot easily be redefined. This means that users cannot adapt the event types detected or the system refined to the different editorial rules used by different broadcasting corporations. In this research we propose a framework for video event detection using SVM classifiers using GA. In our framework, unconstrained video clip can be described as global feature, local feature and motion feature; and in the training stage, each feature can train a classifier. Support Vector Machine (SVM) is employed as basic classifier due to its high precision in classification field. Considering different importance of classifier and removing redundancy, a bi-coded genetic algorithm is employed to select optimal classifier and corresponding optimal weights for every pair of classes. At event detection stage, each classifier can produce a result which is the probability format, and then a set of pre-trained classifiers are used to combine and obtain the final detection results. Because SVMs are trained off-line, it doesn't influence the real-time property of the system.

II. LITERATURE SURVEY

Literature survey gives the brief summary of current literature related to Video Retrieval, feature Selection, Pattern Recognition, and Video Event Detection. Video Event Detection using Temporal Pyramid of Visual Semantics with Kernel optimization and model Subspace Boosting [1].In this work, we present an event modeling system that uses kernel and model selection to optimize the use of temporal pyramids. Static visual semantic features are extracted from video frames and organized into a linear temporal pyramid, where the number of segments within each level incrementally increases from 1 to 10. Maximum-value, minimum value, and averaging aggregation methods are used within each pyramid segment, for a total of 30 unique feature vectors. SVM models are independently trained for each pyramid level and aggregation method, with kernels and parameters optimized by grid search with 3-fold cross validation. Bipartite matching for temporal alignment is included in the set of possible kernels, as well as other kernels that enforce fixed temporal sequence. Video Browsing by Direct Manipulation Method [1, 2] we present a method for browsing videos by directly dragging their content. This method brings the benefits of direct manipulation to an activity typically mediated by widgets. We support this new type of interactivity by: 1) automatically extracting motion data from videos; and 2) a new technique called relative flow dragging that lets users control video playback by moving objects of interest along their visual trajectory. We show that this method can out-perform the traditional seeker bar in video browsing tasks that focus on visual content rather than time.

Transformation of an uncertain video search pipeline to Sketch based Visual Analytic Loop [3]. In this work, propose a novel visual analytics approach to sketch-based video search. We incorporate intuitive sketch-based interactions that allow a user to search queries based on spatio-temporal attributes (e.g., motion, position, distance, trajectory, and spatial occupancy), with application to sports video data. The system computes the similarity between the user sketch and the video content to present visual feedback against the video timeline, along with normalized mean over visualizations that depict segments from the video. The user can then browse the visual feedback and the normalized manevoure visualizations to explore the video data. To facilitate learning, the user can also choose to accept or reject particular results to train the system.
Mining of Semantic Context information for intelligent video Surveillance of traffic Scene [4] In this research, a novel framework is proposed to mine semantic context information for intelligent video surveillance of traffic scenes. First, we introduce how to learn scene specific context information from object-specific context information. Then, object classification is improved by combining of multiple features under a co-training framework. Based on the learnt information, we adopt it to improve object detection and tracking, and detect abnormal events.

III. METHODOLOGY

Overall system description- The purpose of our work is to provide a framework for video event detection and retrieval. Basically, this frame work consists of four layers, as illustrated in Figure 1.The raw video clip is at the left of Figure 1; this layer consists of a sequence of frames, as well as some video attributes, such as compression format, frame rate, number of bits per pixel, duration, etc. The second layer is the feature layer consisting of global feature, local feature and motion feature. Here we are extract motion features of video.
3.1 Local feature and Global feature-The first step in detecting interesting events is to abstract the raw video data into more semantically meaningful streams of information. Robust feature extraction is generally critical for image and video based recognition tasks, and video event recognition requires robust features even in greater demand due to the existing complex motions, cluttered backgrounds, object occlusions, environmental illuminations, and geometric variances of objects
3.2 Motion Feature -Motion features are especially useful for concepts with characteristic motion patterns, including the concepts "Car", "Sports", "Walking/Running". Motion vectors from MPEG macro blocks are aggregated in a two dimensional motion image, which preserves patio allocation, direction, and magnitude of the vectors. Time is compressed into intensity in the motion image. Motion vectors are present for all macro blocks (MB) in predictive (P) and bi-directional (B) frames of MPEG video. For intraframes (I), which start a group of pictures (GOP) sequence of P and B frames, motion vectors have zero magnitude. Motion of a macro block is defined as having a forward direction from a past reference frame or a backward direction from a future reference frame.

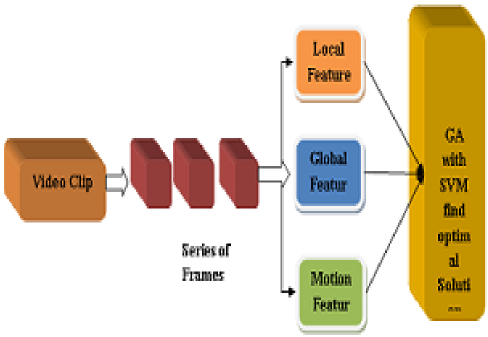


Figure 1: Proposed Work

3.3 Classifier Design using Genetic Algorithm and Support vector machine

Due to the semantic gap between low level visual features and high level concept, video event analysis based on visual features often can't achieve satisfied performance. As mentioned above, we can use 22 global features, 1 local feature and 1 motion feature construct 24kinds of SVM classifiers, each feature's SVM classifiers can label video event individually. But a single feature's SVM classifiers cannot get a better result. Mean while, different features always have different importance, thus the weights of feature's classifier need to be automatically adjust. In order to improve the video event label result, the mechanism of genetic selection is introduced to select optimal weight for feature's classifier to combine a good classifier. Genetic algorithm (GA) is a general adaptive optimization search methodology based on natural selection. GA works with a set of candidate solutions called population. It generates successive populations of alternate solutions that are represented by a chromosome, i.e. a solution to the problem, until acceptable results are obtained. A fitness function assesses the quality of a solution in evaluation step. Chromosomes are selected for reproduction by evaluating the fitness value. The fitter chromosomes have higher probability to be selected for GA operations in each evaluation step. It obtains the optimal solution after a series of interactive computations.

$$C(V) = arg \max_{P_N} \sum_{i=1}^{24} w_i \bullet C_i(V)$$

3.4 Semantic Object Recognition

Semantic video objects recognition is an important step in trajectory detection, follow-up of the semantic object tracking and semantic analysis are based on the semantic object recognition Extract a semantic object includes the following three steps:
- Video semantic segmentation, the video is divided into a series of semantic fragments, the semantic information of these split fragments stored in the database.
- Object recognition, the video frame is divided into several irregularly shaped regions based on texture and contour,

video frame image segmentation using the method of extracting contours and gets the edges of region on the frame; every region is seen as a semantic object.

• Tagging semantic for the objects and indicating that the meaning of objects. Video object tagging can be defined in two ways: one way is according to the understanding and knowledge of the objects in the feature extraction process, such as extraction of car shape, the face model, cloud, sky and other features

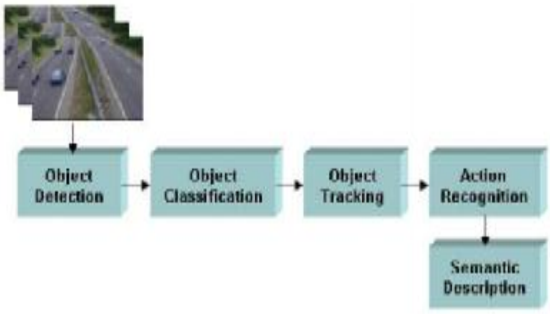


Figure 2: Object Recognition

3.5 Retrieval of Semantic Video Action

For retrieval of semantic video object motion vector is important because object are mostly moving

3.5.1 Motion Trajectory Representation

In our previous work, we have determined a new representation of the motion tracks to generate video features, which are extracted from a non-compressed video data. However, in this work we propose the direct usage of the motion vectors originally embedded in an MPEG bit- stream to construct \motion tracks" in a shot. Although the motion vectors do not always correspond to the real motion" of the objects in a video as compared with the optic low, they are relatively easy to derive. In order to extract the local motion fields from the bit- streams, we perform some preprocessing steps to compensate the camera motion; we produce a motion vector field between the last P-frame in the current GOP (group of pictures) and the I-frame in the next GOP by using the B- frames between the P- and I-frame as

3.5.2 Motion Vector Estimation

To extract reliable motion information, dense motion vectors (MVs) are first figured out from raw video data by using traditional block-matching technique, but subsequently validated for use in following camera motion estimation. Normally, MVs from compressed MPEG data are not dense enough (one MY for each 16*16 macro block) and may be erroneous since the goal of block-matching is to purchase an MV that minimizes least distortion in encoding process, rather than an MV that reflects the true motion behavior proposed .

3.5.3 Motion Vector Validation

As mentioned, MVs calculated thus far may not be correct enough due to the bias in the optimization criterion (least distortion). This situation can be more obvious for pixels in smooth or non-busy areas. To cope with this problem, a validating procedure based on the standard deviation was introduced to remove invalid MVs.

IV. EXPERIMENTALRESULT

The performance for our algorithm can be measured either by its efficiency or effectiveness. Efficiency describes by time taken in the learning the classifier and or the time taken to classify the test cases.

4.1 Experimental Performed on Various Events

In this work we choose different events for evaluating the performance of algorithm. Table 1.1 Show the performance of different events in Support vector machine and different events we are taken to analysis work is driving, bunkering, walking, running and putting. In this work we use SVM and SVM & GA.

Drive – In driving event we take motion of car and other moving object outside the car or any other vehicle.

Bunker – In Bunker event we take motion of bunking.

Walking – In this event we take motion of walking any people.

Running – In this event we are taken motion of any running people.

Putting – In this event we are take motion of putting something.

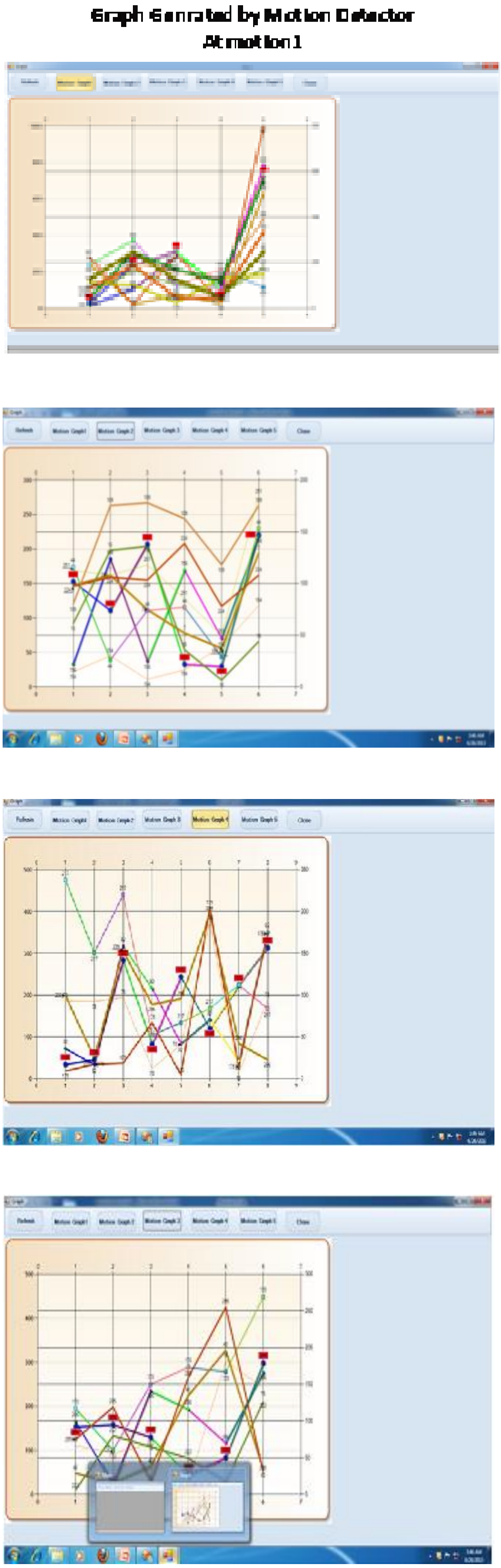
We have evaluated the effectiveness of our framework for optimizing SVM next to the SVM & GA on above video sets as shown in Table 1.1 In organize to compute the value of feature subset selection on video events, we have behavior the experiment using the similar kernel parameters for both pure SVM and GA.

Event Name	For All event	SVM		GA & SVM	
		Train%	Valid %	Train %	Valid %
Drive	20	77.12	72.33	83.82	81.71
Bunk	16	65.32	60.81	84.81	82.33
Put	20	60.78	55.55	83.07	80.43
Walk	25	66.23	52.54	89.10	86.21
Run	20	50.34	42.95	53.60	51.11
Total		63.14	58.23	81.60	77.14

Table 1 : Comparison Studies of SVM and SVM & GA

V. IMPLEMENTATION

For motion analysis we extract motion from video and generate following graphs , here we are detect total five motion for different video scenes and implementation of those result are follow:



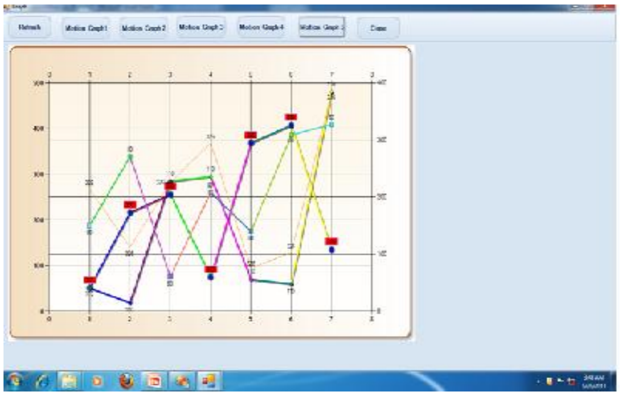


Figure 3: Graph generated by the motion 1 to 5.

VI. CONCLUSION

This research focused on motion analysis, which is a most consistent feature among all low-level features for videos. The motion retrieval system proposed in the paper is a practical attempt in motion-based video retrieval. For practical video retrieval system, it should support multi-modal and multi-feature retrieval.

ACKNOWLEDGEMENT

This research was supported/partially supported by the Department of computer science. I thank to our all staff member who provided insight and expertise that greatly assisted the research, although they may notagree with all of the interpretations/conclusions of this paper.

REFERENCES

[1] N. C. F. Codella, A. Natsev, G. H. Matthew Hill, L. Cao, L. Gong and J. R. Smith, “Video Event Detection Using Temporal Pyramids Of Visual Semantics With Kernel Optimization And Model Subspace Boosting”, IEEE International Conference on Multimedia and Expo-2012.

[2] P. Dragicevic, G. Ramos , J. Bibliowicz, D. Nowrouzezahrail, R. Balakrishna, and K. Singh “ Video Browsing by direct Manipulation” In Proceedings Improved Video Navigation and Capture, CHI 2008, April 5-10, 2008 · Florence, Italy.

[3] P. A. Legg, D. H. S. Chung, M. L. Parry, R. Bown, M. W. Jones, I. W. Griffiths and M. Chen,” Transformation of an Uncertain Video Search Pipeline to a Sketch-Based Visual Analytics Loop”, IEEE Transactions on Visualization and Computer Graphics, vol. 19, no. 12, 2013.

[4].T. Zhang, S. Liu, C. Xu and H. Lu, “Mining semantic context information for intelligent video surveillance of traffic scenes” IEEE Transactions on Industrial Informatics, 2011.

[5] D. Brezeale and D. Cook, "Automatic video classification: A survey of the literature. IEEE Transactions on Systems Man and Cybernetics, vol. 38, no. 3, pp. 416–430, 2008.

[6] C. H. Ayuningtyas, G. A. Putri Saptawati, and L. E. R. Mengko, "A Comparative Study of Feature Ranking Methods as Dimension Reduction Technique in Genome-Wide Association Study". International Conference on Electrical Engineering and Informatics 17-19 July 2011, Bandung, Indonesia.

[7] D. Lin. "An information-theoretic definition of similarity", In Proceedings of the ICML'98, 1998.

[8] Haubold, "Semantic multimedia retrieval using lexical query expansion and model-based reranking", In Proceedings of ICME'06, 2006.

[9] Hauptmann, "Can high-level concepts fill the semantic gap in video retrieval? a case study with broadcast news", IEEE Transactions on Multimedia, 2007.

[10] J. Amores and N. Sebe, "Context-based object-class recognition and retrieval by generalized correlograms", IEEE TPAMI, 2007.

[11] J. Shen, D. Tao and X. Li, "Modality Mixture Projections for Semantic Video Event Detection" IEEE transactions on circuits and systems for video technology", vol. 18, no. 11, 2008.

[12] J. Lu, Y. Tian, Y. Li, Y. Zhang and Z. Lu, "A Framework for Video Event Detection Using Weighted SVM Classifiers", International Conference on Artificial Intelligence and Computational Intelligence, 2009.

[13] P. Li and Y. Guo Hanqiu " Multi-keyframe abstraction from videos" IEEE International Conference on Image Processing 2011.

[14] O. P. Popoola, and K. Wang, "Video-Based Abnormal Human Behavior Recognition - A Review" IEEE Transactions on Systems, Man and Cybernetics: Applications and Reviews, vol. 42, no. 6, 2012.

[15] T. Giannakopoulos, S. Petridis and S. Perantonis, "User-driven recognition of audio events in news videos", Fifth International Workshop on Semantics Media Adaptation and Personalization, 2010.