

Multiple Video Instance Detection and Retrieval using Spatio-Temporal Analysis using Semi Supervised SVM Algorithm

R. Kousalya, PhD
Assistant Professor,
HOD Department of Computer Applications
Dr. N.G.P. Arts and Science College,
Coimbatore - 48, Tamil Nadu

S. Dharani
Research Scholar, Department of Computer
Science,
Dr. N.G.P. Arts and Science College,
Coimbatore - 48, Tamil Nadu

ABSTRACT

Object instance search aims to not solely retrieve the pictures or frames that contain the query, however additionally find all its occurrences. During this work, we tend to explore the utilization of spatio-temporal cues to enhance the standard of object instance search from videos. To the present finish, the work to formulate this drawback because the spatio-temporal trajectory search downside, wherever a trajectory may be a sequence of bounding boxes that find the thing instance in every frame. The goal is to seek out the top- trajectories that are possible to contain the target object. The work tends to solve the key bottleneck in applying the approach to object instance search by leverage a randomized approach to change quick marking of any bounding boxes within the video volume.

Keywords

Key-point localization, SIFT descriptor, Orientation Assignment, Key-points descriptors, Scale-space extrema detection.

1. INTRODUCTION

1.1 Spatial Data Mining

Spatial data mining is the process of discovering interesting, useful, non-trivial patterns from large spatial datasets. Spatial data mining is the application of data mining methods to spatial data. The end objective of spatial data mining is to find patterns in data with respect to geography. So far, data mining and Geographic Information Systems (GIS) have existed as two separate technologies, each with its own methods, traditions, and approaches to visualization and data analysis. Particularly, most contemporary GIS have only very basic spatial analysis functionality. The immense explosion in geographically referenced data occasioned by developments in IT, digital mapping, remote sensing, and the global diffusion of GIS emphasize the importance of developing data-driven inductive approaches to geographical analysis and modeling.

Data mining offers great potential benefits for GIS-based applied decision-making. Recently, the task of integrating these two technologies has become of critical importance, especially as various public and private sector organizations possessing huge databases with thematic and geographically referenced data begin to realize the huge potential of the information contained therein. Among those organizations are offices requiring analysis or dissemination of geo-referenced statistical data, Public health services searching for explanations of disease clustering Environmental agencies assessing the impact of changing land-use patterns on climate

change Geo-marketing companies doing customer segmentation based on spatial location.

1.2 Challenges in spatial mining:

Spatial data repositories tend to be very large. Moreover, existing GIS datasets are often splintered into feature and attribute components that are conventionally archived in hybrid data management systems. Algorithmic requirements differ substantially for relational (attribute) data management and for topological (feature) data management. Related to this is the range and diversity of geographic data formats, which present unique challenges. The digital geographic data revolution is creating new types of data formats beyond the traditional "vector" and "raster" formats. Geographic data repositories increasingly include ill-structured data, such as imagery and geo-referenced multi-media.

There are several critical research challenges in geographic knowledge discovery and data mining. Miller and Han offer the following list of emerging research topics in the field.

1.3 Developing and Supporting Geographic Data Warehouses (GDW's)

Spatial properties are often reduced to simple spatial attributes in mainstream data warehouses. Creating an integrated GDW requires solving issues of spatial and temporal data interoperability – including differences in semantics, referencing systems, geometry, accuracy, and position.

1.4 Better spatio-temporal representations in geographic knowledge discovery

Current geographic knowledge discovery (GKD) methods generally use very simple representations of geographic objects and spatial relationships. Geographic data mining methods should recognize more complex geographic objects (i.e., lines and polygons) and relationships (i.e., non-Euclidean distances, direction, connectivity, and interaction through attributed geographic space such as terrain). Furthermore, the time dimension needs to be more fully integrated into these geographic representations and relationships.

1.5 Geographic knowledge discovery using diverse data types

GKD methods should be developed that can handle diverse data types beyond the traditional raster and vector models, including imagery and geo-referenced multimedia, as well as dynamic data types (video streams, animation).

2. RELATED WORK

2.1 Automatic Semantic Content Extraction in Videos Using a Fuzzy Ontology and Rule-Based Model [1]

This study proposes an automatic semantic content extraction framework. This is accomplished through the development of an ontology-based semantic content model and semantic content extraction algorithms. Our work differs from other semantic content extraction and representation studies in many ways and contributes to semantic video modeling and semantic content extraction research areas. First of all, we propose a metaontology, a rule construction standard which is domain independent, to construct domain ontologies. Domain ontologies are enriched by including additional rule definitions.

2.2 An Overview of Data Models and Query Languages for Content-Based Video Retrieval [2]

Video content-based retrieval requires many changes in a multimedia database management system relative to a traditional database management system. These changes take effect mainly in modeling and querying techniques. As stated in, a multimedia data model is different from a traditional data model. A multimedia data model should include, except the elements that represent inherent structural properties of the video, also the elements that represent the video content. It should be able to capture and represent various types of information about multimedia objects, their structures, operations and properties, as well as real-world objects and relationships among them.

2.3 Content-Based Video Retrieval by Integrating Spatio-Temporal and Stochastic Recognition of Events [3]

The current paper addresses these problems with the emphasis on the second one - recognizing semantic content in video data based on visual features. First, we propose a layered video data model that provides a framework for automatic mapping from features to concepts. The model is independent of feature/semantic extractors, providing flexibility in using different video processing and pattern recognition techniques for those purposes. At same time the model is in line with the latest development in MPEG7, differencing video content between diverse categories. Next, we extend the model with object and event grammars. These grammars are aimed at formalizing descriptions of high-level concepts, as well as facilitating their extraction based on features and spatiotemporal reasoning.

2.4 Event Detection and Analysis from Video Streams [4]

The paper presents a generic framework for event detection and behavior analysis. Motion detection is made difficult as both the observer and some elements of the scene may be moving. To handle this situation, we first compensate adjacent frames for the motion field induced by the observer which manifests it globally. The results of this ego motion estimation are used to register frames, to detect independently moving objects, and to track them. Motion by itself, however, is not a sufficient indication of a threatening or otherwise interesting activity. In most natural scenes, there are a significant number of moving objects and it is the analysis of their trajectories

and interaction with the features of the scene which allows us to classify and recognize interesting events.

2.5 Video-Based Event Recognition: Activity Representation and Probabilistic Recognition Methods [5]

On the detection of large-scale activities where some knowledge of the scene (e.g., the characteristics of the objects in the environment) is known. One characteristic of activities of our interest is that they exhibit some specific patterns of whole-body motion. For example, consider a group of people stealing luggage left unattended by the owners. One particular pattern of the “stealing” event may be: two persons approach the owners and obstruct the view of the luggage, while another person takes the luggage. The task of activity recognition is to bridge the gap between numerical pixel level data and a high-level abstract activity description. A common approach involves first detecting and tracking moving object features from image sequences.

2.6 Automatic Detection of Salient Objects and Spatial Relations in Videos for a Video Database System [6]

Multimedia databases have gained popularity due to rapidly growing quantities of multimedia data and the need to perform efficient indexing, retrieval and analysis of this data. One downside of multimedia databases is the necessity to process the data for feature extraction and labeling prior to storage and querying. Huge amount of data makes it impossible to complete this task manually. We propose a tool for the automatic detection and tracking of salient objects, and derivation of spatio-temporal relations between them in video. Our system aims to reduce the work for manual selection and labeling of objects significantly by detecting and tracking the salient objects, and hence, requiring to enter the label for each object only once within each shot instead of specifying the labels for each object in every frame they appear.

2.7 Spatio-Temporal Querying in Video Databases [7]

In this paper a video data model that allows efficient and effective representation and querying of spatio-temporal properties of objects is presented. The data model is focused on the semantic content of video streams. Objects, events, activities performed by objects are main interests of the model. The model supports fuzzy spatial queries including querying spatial relationships between objects and querying the trajectories of objects. The model is flexible enough to define new spatial relationship types between objects without changing the basic data model. A prototype of the proposed model has been implemented. The prototype allows various spatio-temporal queries along with the fuzzy ones and it is prone to implement compound queries without major changes in the data model. Present a spatio-temporal data model for video data.

2.8 Multilevel Video Content Representation and Retrieval [8]:

Several practical algorithms are proposed to support content-based video analysis, modeling, representation, summarization, indexing, and access. First, a multilevel video database model is given. One advantage of this model is that it provides a reasonable approach to bridging the gap between low-level representative features and high-level semantic concepts from a human point of view. Second, several model-

based video analysis techniques are proposed. In order to detect the video shots, we present a novel technique, which can adapt the threshold for scene cut detection to the activities of variant videos or even different video shots. A seeded region aggregation and temporal tracking technique is proposed for generating the semantic video objects.

The semantic video scenes can then be generated from these extracted video access units (e.g., shots and objects) according to some domain knowledge. Third, in order to categorize video contents into a set of semantic clusters, an integrated video classification technique is developed to support more efficient multilevel video representation, summarization, indexing, and access techniques

2.9 Video Semantic Content Analysis Based on Ontology [9]

The rapid increase in the available amount of video data is creating a growing demand for efficient methods for understanding and managing it at the semantic level. New multimedia standard, MPEG-7, provides the rich functionalities to enable the generation of audiovisual descriptions and is expressed solely in XML Schema which provides little support for expressing semantic knowledge. In this paper, a video semantic content analysis framework based on ontology combined MPEG-7 is presented. Domain ontology is used to define high level semantic concepts and their relations in the context of the examined domain. MPEG-7 metadata terms of audiovisual descriptions and video content analysis algorithms are expressed in this ontology to enrich video semantic analysis. OWL is used for the ontology description. Rules in Description Logic are defined to describe how low-level features and algorithms for video analysis should be applied according to different perception content. Temporal Description Logic is used to describe the semantic events, and a reasoning algorithm is proposed for events detection. The proposed framework is demonstrated in sports video domain and shows promising results. The main challenge, often referred to as the semantic gap, is mapping high-level semantic concepts into low-level spatiotemporal features that can be automatically extracted from video data. Feature extraction, shot detection and object recognition are important phases in developing general purpose video content analysis. Significant results have been reported in the literature for the last two decades, with several successful prototypes. However, the lack of precise models and formats for video semantic content representation and the high complexity of video processing algorithms make the development of fully automatic video semantic content analysis and management a challenging task. And, the mapping rules often are written into program code.

2.10 Object Instance Search in Videos via Spatio-Temporal Trajectory Discovery [10]

Given a selected object as question, object instance search aims to not solely retrieve the pictures or frames that contain the question, however additionally find all its occurrences. During this work, we tend to explore the employment of spatio-temporal cues to enhance the standard of object instance search from videos. To the current finish, we tend to formulate this downside because the spatio-temporal flight searches downside, wherever a flight may be a sequence of bounding boxes that locate the thing instance in every frame.

The goal is to seek out the top- trajectories that square measure probably to contain the target object. Despite the

massive variety of flight candidates, we tend to depend upon a recent spatio-temporal search formula for event detection to with efficiency realize the best spatio-temporal trajectories in massive video volumes, with complexness linear to the video volume size. We tend to solve the key bottleneck in applying this approach to object instance search by investment a randomized approach to modify quick rating of any bounding boxes within the video volume. Additionally, we tend to gift a brand new dataset for video object instance search.

3. SYSTEM IMPLEMENTATION

Implementation is the process that actually yields the lowest-level system elements in the system. The purpose of the implementation process is to design and create a system element conforming to that element's design properties and requirements. The element is constructed employing appropriate technologies. This process bridges the system definition processes and the integration process. System Implementation is the stage in the project where the theoretical design is turned into a working system.

3.1 Phases

3.1.1 Scale-space extrema detection

3.1.2 Key-point localization

3.1.3 SIFT descriptor

3.1.4 Orientation Assignment

3.1.5 Key-points descriptors

3.1.1 Scale-space extrema Detection

Search over multiple scales and image locations. Identify locations and scales that can be repeatedly assigned under different views of the same scene or object. The scale space of an image is a function $L(x, y)$ that is produced from the convolution of a Gaussian kernel with the input image. Scale the stage where the interest points, which are called key-points in the SIFT framework, are detected. The image is convolved with Gaussian filters at different scales, and then the differences of successive Gaussian-blurred images are taken. Key-points are then taken as maxima/minima of the Difference of Gaussians (DoG) that occur at multiple scales of time. Specifically, a DoG image

$D(x, y, \sigma)$ is given by:

$$D(x, y, \sigma) = L(x, y, k_i\sigma) - L(x, y, k\sigma)$$

Where $L(x, y, k\sigma)$ is the convolution of the original image $I(x, y)$ with the Gaussian blur $G(x, y, k\sigma)$ at scale, i.e.,

$$L(x, y, k\sigma) = G(x, y, k\sigma) * I(x, y)$$

Hence a DoG image between scales $k_i\sigma$ and $k_j\sigma$ is just the difference of the Gaussian-blurred images at scales $k_i\sigma$ and $k_j\sigma$. For scale space extrema detection in the SIFT algorithm, the image is first convolved with Gaussian-blurs at different scales. The convolved images are grouped by octave (an octave corresponds to doubling the value of σ), and the value of k_i is selected so that we obtain a fixed number of convolved images per octave. Then the Difference-of-Gaussian images are taken from adjacent Gaussian-blurred images per octave.

Once DoG images have been obtained, key-points are identified as local minima/maxima of the DoG images across scales. The process is done by comparing each pixel in the DoG images to its eight neighbors at the same scale and nine

corresponding neighboring pixels in each of the neighboring scales. If the pixel value is the maximum or minimum among all compared pixels, it is selected as a candidate key-point.

3.1.2 Key-point localization

After scale space extrema are detected the SIFT algorithm discards low contrast key-points and then filters out those located on edges. Resulting set of key-points are shown in a last image. Scale-space extrema detection produces a too many Key-point candidates, some of which key-points are unstable. The next process of an algorithm is to perform a

detailed fit to the nearby data for accurate location, scale, and ratio of principal curvatures. Information allows points to be rejected that have low contrast (and are therefore sensitive to noise) or are poorly localized along an edge.

3.1.3 SIFT descriptor:

Key-points of an object are first extracted from a set of reference images and stored in a database. An object is recognized in a new image by individually comparing each feature from the new image to the database and finding candidate matching feature based on their feature vector.

Query Images:



Fig.1: Some of the Query Images

Any object in an image, interesting points on the object can be extracted to provide a “feature description” of the object. Feature description is extracted from a training image, can then be used to identify the object when attempting to locate the object in a test image containing many other objects. Features extracted from the training image to detectable even under changes in image scale, noise and illumination. Such points usually lie on high-contrast regions of the image, such as object edges. Each cluster of 3 or more features that agree on an object and its pose is then subject to further detailed model verification and subsequently outliers are discarded. Object matches that pass all these tests can be identified as correct with high confidence.

3.1.4 Orientation assignment

Each key-point is assigned one or more orientations based on local image gradient directions. The key step in achieving invariance to rotation as the key-point descriptor can be represented relative to this orientation and therefore achieve invariance to image rotation.

First, the Gaussian-smoothed image $L(x, y, \sigma)$ at the key point's scale σ is taken so that all computations are performed in a scale-invariant manner. For an image sample $L(x, y)$ at scale σ , the gradient magnitude $m(x, y)$, and orientation $\theta(x, y)$, are pre-computed using pixel differences:

$$m(x, y) = \sqrt{(L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2}$$

$$\theta(x, y) = \text{atan2}(L(x, y + 1) - L(x, y - 1), L(x + 1, y) - L(x - 1, y))$$

The magnitude and direction calculations for the gradient are done for every pixel in a neighboring region around the key-point in the Gaussian-blurred image L . An orientation histogram with 36 bins is formed, with each bin covering 10 degrees. Each sample in the neighboring window added to a histogram bin is weighted by its gradient magnitude and by a Gaussian-weighted circular window within that is 1.5 times that of the scale of the key-point.

The peaks in this histogram correspond to dominant orientations. Once the histogram is filled, the orientations corresponding to the highest peak and local peaks that are within 80% of the highest peaks are assigned to the key-point. In the case of multiple orientations being assigned, an additional key-point is created having the same location and scale as the original key-point for each additional orientation.

3.1.5 3.1.5 Key-point Descriptors:

The gradient information is rotated to line up with the orientation of the key-point and the weighted by a Gaussian with

Variance of $1.5 * \text{Key-point scale}$. The key-point descriptor data is used to create a set of histograms over a window centered on the key-point. It uses a set of 16 histograms aligned in a 4×4 grid, each with orientation bins. Input arguments are calculated based on the formation of an image cells. Histograms are computed from magnitude and orientation values of samples in a 16×16 region around the

key-point value such that each histogram contains samples from a 4×4 sub region of the original neighborhood region. The descriptor then becomes a vector of all the values of these histograms. Since there are $4 \times 4 = 16$ histograms each with 8 bins the vector has 128 elements. To compute a descriptor vector for each key-point such that the descriptor is highly distinctive and partially invariant to the remaining variations such as illumination, 3D viewpoint, etc. This step is performed on the image closest in scale to the key point's scale.

Longer descriptors continue to do better but not by much and there is an additional danger of increased sensitivity to distortion and occlusion. It is also shown that feature matching accuracy is above 50% for viewpoint changes of up to 50 degrees. Therefore SIFT descriptors are invariant to minor affine changes. To test the distinctiveness of the SIFT descriptors, matching accuracy is also measured against varying number of key-points in the testing database, and it is shown that matching accuracy decreases only very slightly for very large database sizes, thus indicating that SIFT features are highly distinctive.

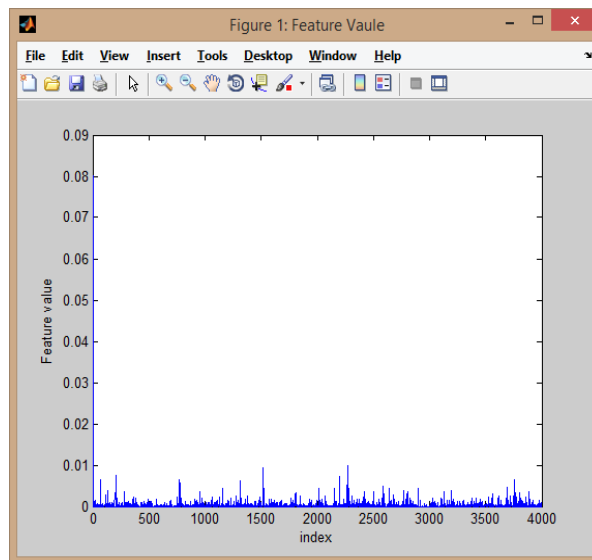


Fig.2: Histogram Computed

4. DATASET

The database in the experiments was made up of one thousand video clips downloaded from internet. The total length of video data is approximately one hundred and sixty hrs, including various contents such as news, sports, movie, cartoon, teleplay and natural scene etc. The experiments were performed on Xeon E5410 2.33GHz, 2G memory. The algorithms in were implemented for comparison Test Results. The experiments were set to the performance comparison of similarity search.

The proposed algorithm performs about twenty times faster than the other algorithms. The reason is that the proposed search is carried out in a bounded scope based on video components of relevant similar video and redundant search of dissimilar video can be avoided. Besides, the search time of other algorithms grows faster than the proposed approach. This is because the proposed search is guided into the corresponding database CIT whose volume is not proportion to that of video database. According to experimental results, the proposed method can greatly improve the efficiency of video similarity search in large database.

The system chooses the most challenging 40 queries to query the videos from a video database. Queries are selected from the videos present in the video database. Each year Google examines billions of queries that people around the world

have typed into Google search to discover the saved EVVE video Archives.

To collect the video archives data from EVVE website and choose the most complex to queries to search video. The number of downloaded videos for each query is up to 100. Extracted more than 50 videos are from the video database archives.

5. EXPERIMENTAL RESULTS

The results of recall - precision curves. The algorithm (1) [Semi-Supervised SVM] is the method and the algorithm (2) [K-Means] stands for the approach. The precision rate goes down when the recall rate rises. The proposed method holds steady precision at 1 while the others drop significantly when the recall rate rises. The algorithm (1) is slightly better than algorithm (2) in precision rate when the recall rate is smaller than 0.60. The proposed algorithm can hold precision rate at 1 when the recall rate reaches 0.75. The others have the best precision rate at 0.7 when the recall rate is 0.75. When recall rate is approaching to 1, the precision rate of the proposed method decreased a little at 0.9.

Precision: Precision is the probability that a query retrieved document is relevant.

Recall: Recall is the probability that a query is relevant document is retrieved in a searching process.

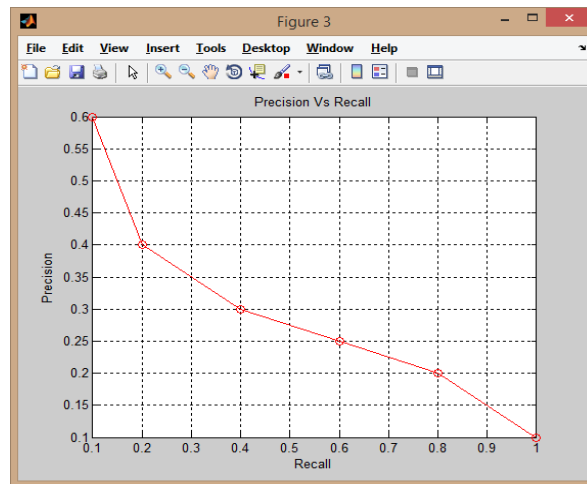


Fig.3: Precision Vs Recall

The other's precision rate are below 0.6 when their recall rate are higher than 0.8. This is because the proposed search approach is according to video components based on the statistics of STD. The similarity search is directed to the

corresponding CIT and the dissimilar video can be eliminated effectively by index clustering. The representative video shot images of comparison VSS.

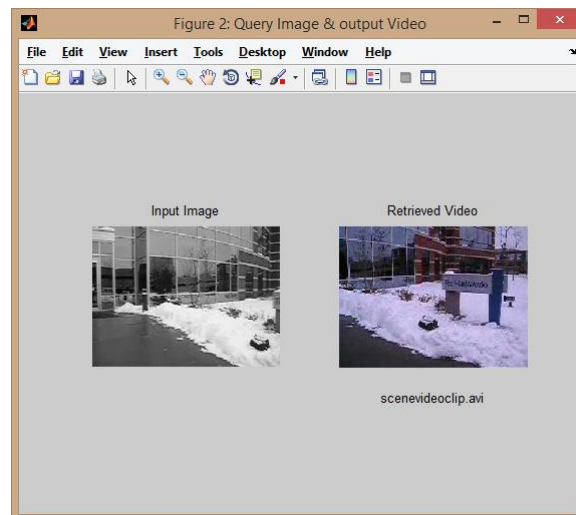


Fig.4: Video retrieved based on the Query

The proposed method can find more similar video clips with similar shots compared with the algorithm based on support vector machine. Support Vector Machine algorithm can perform twenty times faster than the other algorithms. The reason is that the proposed search is carried out in a bounded scope based on video components of relevant similar video and redundant search of dissimilar video can be avoided.

Despite the fact that the far-reaching independence assumptions are often inaccurate, the naive Bayes classifier has several properties that make it surprisingly useful in practice. In particular, the decoupling of the class conditional feature distributions means that each distribution can be independently estimated as a one dimensional distribution. This in turn helps to alleviate problems stemming from the curse of dimensionality, such as the need for data sets that scale exponentially with the number of features. Like all probabilistic classifiers under the MAP decision rule, it arrives at the correct classification as long as the correct class is more probable than any other class; hence class probabilities do not have to be estimated very well. In other words, the overall classifier is robust enough to ignore serious

deficiencies in its underlying naive probability model. Other reasons for the observed success of the naive Bayes classifier are discussed in the literature cited below.

6. COMPARISON OF TWO METHODS

Several videos have been performed using the matlab to evaluate the impact of different types of similar videos. The similar videos are performed based on the Support Vector Machine (SVM) to retrieve the correct position and frames of the similar videos. Automatic Extraction of Similar Videos using a video query clip based on Support Vector Machine contributes in several videos of several modeling and similar video extraction research area process. The Similar video extraction process is done automatically and for additional purpose (VISCOM) videos model is used for the extraction of the similar process.

By comparing the two methods of Support Vector Machine and VISCOM model the accuracy rate and time are compared to retrieve the similar video from the database. Accuracy rate and time of an each process in both the methods are reduced as compared to the previous method of an event.

Table.1: Comparison of Algorithms

Video	Support Vector Machine		Existing (K-Means)	
	Accuracy	Time	Accuracy	Time
Sports	79.32	5.44	68.50	14.26
News	91.65	1.45	85.43	4.42
Entertainment	85.76	2.82	78.89	10.28
Cartoon	93.13	3.32	86.74	4.80
Average	87.47	3.26	79.89	8.47

7. CONCLUSION

The system proposes has presented an effective query processing strategy for temporal localization of similar instances from a long un-segmented video stream using support vector machine algorithm, considering target subsequence may be approximate occurrence of potentially different ordering or length with query clip.

This work plans to further investigate the effect of representing videos by other features, such as ordinal signature and also to introduce the idea of incorporating spatial information, combining Bof model with global histogram features and expansion of queries. Moreover, the weight of each factor for measuring video similarity might be adjusted by user feedback to embody the degree of similarity more completely and systematically. Video Compression and de-noising can be added in future to the current system. The System can include the video condensation approaches to reduce the size of the video file.

8. REFERENCES

- [1] Boltz S., Debreuve E., Barlaud M. "High-Dimensional Statistical Measure for Region-of-Interest Tracking", IEEE Trans. Image Processing, vol. 55, pp. 1731-1737, 2009
- [2] G.G. Medioni, I. Cohen, F. Bre´mond, S. Hongeng, and R. Nevatia, "Event Detection and Analysis from Video Streams"
- [3] Garcia-Garcia D., Hernandez E.P., Diaz de Maria F. "A New Distance Measure for Model-Based Sequence Clustering" IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31 pp.1325-1331, 2009
- [4] Heng Tao Shen, Jie Shao, Zi Huang, Xiaofang Zhou "Effective and Efficient Query Processing for Video Subsequence Identification" IEEE Trans. on Knowledge and Data Engineering, vol.21 pp. 321-334 2009
- [5] Heung S, Zakhor "A Efficient video similarity measurement with video signature". IEEE Trans. on Circuits and Systems for Video Technology, vol.13 pp.59-74 2003
- [6] J. Fan, W. Aref, A. Elmagarmid, M. Hacid, M. Marzouk, and X. Zhu, "Multiview: Multilevel Video Content Representation and Retrieval"
- [7] JingjingMeng, Junsong Yuan, Jiong Yang, Gang Wang, and Yap-Peng Tan, " Object Instance Search in Videos via Spatio-Temporal Trajectory Discovery"
- [8] Kwang-deokSeo, Seong Park, Soon-heung Jung, "Wipe scene-change detector based on visual rhythm spectrum" IEEE Trans. Consumer Electronics. vol.55 pp.831-838 2009
- [9] L. Bai, S.Y. Lao, G. Jones, and A.F. Smeaton, "Video Semantic Content Analysis Based on Ontology"
- [10] M. Ko'pru' lu', N.K. Cicekli, and A. Yazici, "Spatio-Temporal Querying in Video Databases"
- [11] M. Petkovic and W. Jonker, "An Overview of Data Models and Query Languages for Content-Based Video Retrieval"
- [12] M. Petkovic and W. Jonker, "Content-Based Video Retrieval by Integrating Spatio-Temporal and Stochastic Recognition of Events"
- [13] Meng Wang, Xian-Sheng Hua, Jinhui Tang, Richang Hong. "Beyond Distance Measurement: Constructing Neighborhood Similarity for Video Annotation" IEEE Trans. Multimedia vol. 11, pp. 465-476, 2009
- [14] S. Hongeng, R. Nevatia, and F. Bre´mond, "Video-Based Event Recognition: Activity Representation and Probabilistic Recognition Methods"
- [15] T. Sevilmis, M. Bastan, U. Gu'du' kbay, and O' .Ulusoy, "Automatic Detection of Salient Objects and Spatial Relations in Videos for a Video Database System"

- [16] Xuefeng Pan, Jintao L, Yongdong Zhang,, Sheng Tang, Lejun Y, “Format-Independent Motion Content Description based on Spatiotemporal Visual Sensitivity” IEEE Trans. Consumer Electronics. vol.53 pp.769-774 2007
- [17] YakupYildirim, Adnan Yazici, TurgayYilmaz, “Automatic Semantic Content Extraction in Videos Using a Fuzzy Ontology and Rule-Based Model”
- [18] Yan Ke Rahul Sukthankar Larry Huston “An efficient parts-based near duplicate and sub-image retrieval system” Proc. of the 12th annual ACM international conference on Multimedia pp.869-876 2004
- [19] Yang X, Tian Qi, Chang E-C “A color fingerprint of video shot for content identification” Proc. of the 12th annual ACM int’l conf. on Multimedia pp.276-279 2004
- [20] YueGaoWeibo Wang, Junhai Yong. “A video summarization tool using two-level redundancy detection for personal video recorders”, IEEE Trans. Consumer Electronics.vol.54 pp.521-526 2008