Fuzzy Ontology and Rule based Model for Automatic Semantic Content Extraction from Videos using k-Means Algorithm

Priyanka Nikam
Student
Late G.N. Sapkal COE,
Anjeneri, Nashik, Maharashtra, India

B.R. Nandwalkar
Professor
Late G.N. Sapkal COE,
Anjeneri, Nashik, Maharashtra, India

ABSTRACT

Video based applications disclosed need for efficiently extracting and modeling the video contents. The video features can be classified into normal data, relative features and logic content. Semantic level understanding is required for core content of video. So to get video content automatic semantic content framework is proposed. In proposed system a semantic content extraction system that allows the user to query and regain objects, events, and concepts that are extracted automatically. VISCOM is a video semantic content model which contains classes and relations between classes. Objects and events are represented by some VISCOM classes and other classes are used in the automatic semantic content extraction framework. VISCOM classes collect the semantic content types and relations.

Ontology based fuzzy video data semantic model which uses spatial and temporal relations in event and concept definition is proposed. Extracted objects from consecutive representative frames are processed to extract temporal relations. Additional rules to lower spatial relation computation cost and to define some difficult situations more successfully are used. To extract objects from video we apply k-means clustering algorithm. By, which we get the more relevant objects related to user query.

Keywords

Fuzziness, Ontology, Semantic Content Extraction, Spatial Relations, Video Content Modeling.

1. INTRODUCTION

As a large amount of video data is available it is necessary to model them so, that it enables users to retrieve some desired contents from video in semantically meaningful manner. Video content modeling and extraction is required in the applications like surveillance, crime investigation, sports, border monitoring etc. It helps to give desired content from video.

Video content has 3 levels which are normal data, relative features and logic content. Normal data is elementary physical units with attributes frame rate, format, length. While audio, text, and visual features like texture, color distribution, shape, motion, etc. are relative features. And high level concepts such as objects, events are logic content. First two levels are inefficient to extract semantic content as the extraction is based on only low level features. They hardly provide semantics to the user in which he is interested. So, it is necessary to develop intelligent methods for semantic content extraction in videos.

Manual techniques are used for extraction but they are subjective and tedious. Some system uses automatic or semi automatic approach but they are unable to give appropriate solutions. Although there are several studies implemented different methodologies such as object detection and tracking multi-modality and spatio-temporal derivatives, the most of these studies propose techniques for specific event type extraction or work for specific cases and assumptions. To solve such problem some system uses ontology for representation and uses both techniques ontology and spatio temporal relations for semantic content extraction. These systems are designed for specific domain some work is to be done manually.

Integration of both approaches is proposed. System uses both ontology construction and spatio-temporal relations, for extraction of events and concepts from video. VISCOM (Video Semantic Content Model) is proposed for domain independent ontology construction. It is meta-ontology. For object extraction k-means clustering algorithm is used. It gives more appropriate objects in video. As, clustering is used only query related and identical objects are get extracted. Events and concepts are extracted using spatial and temporal relations between objects. Also, VISCOM which is ontology is used in the extraction process.

In the automatic event and concept extraction process, objects, events, domain ontologies, and rule definitions are used. The extraction process starts with object extraction. Specifically, a semiautomatic k-means Algorithm-based object extraction approach is used for the object extraction and classification needs of this study. For each representative frame, objects and spatial relations between objects are extracted. Then, objects extracted from consecutive representative frames are processed to extract temporal relations, which is an important step in the semantic content extraction process. In these steps, spatial and temporal relations among objects and events are extracted automatically allowing and using the uncertainty in relation definitions. Event extraction process uses objects, spatial relations between objects and temporal relations between events. Similarly, objects and events are used in concept extraction process.

2. LITERATURE SURVEY

We categorize video content extraction evolution process in manual techniques, spatiao-temporal relations, ontology based video content extraction.

In content-based video retrieval modeling and querying capabilities are provided [2]. Video modeling makes distinction between the structure and content of video. Modeling can be done on the basis of either feature-based or semantic-based. System is unable to integrate both the feature-based and semantic modeling. A framework for mapping from features to high-level concepts is used in layered video data model [3]. Flexibility is provided for using different video processing techniques. For mapping low-level features to high-level concepts a new video data model which supports the integrated use of two different approaches. Model is unable automatically extract high level semantics from video data.

The VSAM proposed detection and tracking module to extract moving objects trajectory from video streams [4]. The event detection and analysis is based on the detection of moving objects and by estimating their speed and trajectory and inference of their behavior. The limitation of this model is stability as it is induced manually by observer.

To recognize the behavior of moving objects system is implemented [5]. An activity is to be considered composed of action threads and in which each thread is executed by a single actor. Multi-agent event is represented by an event graph composed of several action threads related by logical and temporal constraints [6]. The complexity depends on number of moving objects and complexity of scene. Video event graph used to learn the event structure from training videos [6]. Graph composed of temporally correlated sub-events, which is used to automatically encode the event correlation graph. Problem with graph formulation because sub-events are depend on their predecessor.

BilVideo is the video database management system provides full support for queries on spatio-temporal, low level features and semantic on the video data [7][8]. It is domain independent and to handle the spatio-temporal queries a knowledge-base is used that consists of a fact-base and comprehensive set of rules which reduce the number of facts. To reduce the work for manual selection and labelling of objects significantly by detecting and tracking the salient objects. Extended Advanced video information system (AVIS) is data model that allows efficient and effective presentation of spatial-temporal properties of objects which focused on the semantic content of video streams [9]. It is object based video data model and application independent. It supports fuzzy spatial queries including querying spatial relationships between objects.

Multilevel video database model provides a reasonable approach to bridging the gap between low-level representation features and high level semantic concepts from a human point of view [14][15]. A hierarchical semantics, sensitive video classifier is proposed to shorten the semantic gap between the low-level visual features and high level semantic concepts. Semantic content analysis framework based on domain ontology which is used to define high level semantic concepts and their relations in the context of the examined domain [10]. To enrich analysis low-level features and video content analysis algorithms are integrated into the ontology. Description logic is used to describe how video processing methods and low-level features should be applied according to different semantic content. Linguistic and dynamic visual ontology presented in [11]. The structure with proposed ontology can be used to perform higher level annotation of the video clips to generate complex queries. Logical structure of domain ontology defined in terms of linguistic concepts. Event description framework (EDF) is to capture event semantics that enables storage, inference and retrieval of events from lower level event observations [12]. EDF identify a set of classes for semantic annotation of multimedia data and describe their properties and relationships between events and entities. It is ontology for spatial-temporal relationships. But, event extraction is manually done in this framework.

Video Event Recognition Language describes event ontology [13]. It is formal language for representing events for describing ontology for application domain and for annotating data with those ontology categories. To address the problem of designing ontology for visual activity recognition [16] proposes a system. On general ontology design principles and adapt them to the specific domain of human activity ontology. Qualitative evaluation principles and provide several examples from existing ontology and how they can be improved upon are discussed. Genetic Algorithm based object extraction and classification mechanism is proposed for extracting the content of the videos [17]. The object extraction is defined as a classification problem and a Genetic Algorithm based classifier is proposed for classification.

But, manual techniques are both cost and time they are subjective also. The extraction process is manual. In some approaches automatic or semiautomatic methods with spatio-temporal relations are used. But it fails to give appropriate solution. In ontology based systems ontology is constructed for semantic content representation. Different languages are designed for representation purpose.

3. PROPOSED SYSTEM

In the proposed system input video proceed through keyframe extraction, feature extraction, object extraction and then event and concept extraction. Fig1 gives the block diagram of proposed system.

3.1 Key Frames and Feature Extraction

The process is studied as object extraction from images since videos are a set of images (key frames). Before starting with the image segmentation and object extraction from images, key frames are obtained by using a key frame extraction algorithm.

Algorithm: Key Frame Extraction

Input: Video

Output: Key Frames

1. Select video file
2. Analyze video file (Extracting information number of frames etc.)
3. For all available frames from specified video file
4. Extract first frame from frame list

After key frames extraction extracted frames are proceeds for segmentation and feature extraction. For that purpose segmentation algorithm is proposed. Firstly each segment obtained is assumed to be a candidate object. Also all combinations of neighboring segments are taken as candidate objects. All candidate objects are tried to be classified by the k-means neighboring based algorithm. Using ranking or thresholding filters, real objects are obtained. Segmentation, the features of the segments is extracted as training data or to make a comparison for decision in the querying phase. These features are MPEG descriptors, such as color, texture and shape descriptors. Segmentation and feature extraction algorithm given below.

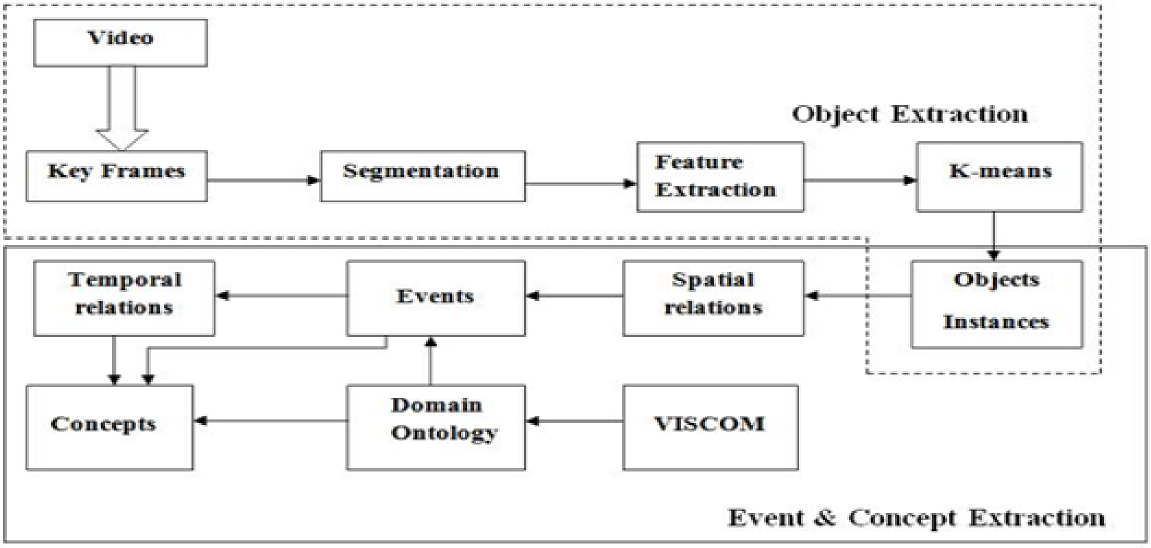

Fig 1: System Architecture

Algorithm: Segmentation and Feature Extraction

Input: Key Frames

Output: Segments and its features

1. Analyze key frame
2. Extract low level information and segmentation information
3. Loop all available segmentation from the specified key frame
4. Extract first segmentation
5. Analyze segmentation
6. Extract all feature information about specific segmentation
7. Go to step 3
8. Analyze feature information specified by key frame segmentation (apply k-means)
9. Store to database
10. Go to next key frame

3.2 Object Extraction with k-Means Algorithm

k-means Algorithms methodology is a powerful search technique for solving problems in many different research areas. k-means are mostly used for improving the performance; instead of testing all combinations, k- means ensure the most fitting-ones to be obtained in fewer tries. k-means are tried to be used in different levels/phases of object/image retrieval. k-means are used for many different purposes in object extraction; for selecting representative frame of video, for generating coordinated for objects, for feature selection, for pixel selection, etc.

Algorithm: Object Extraction

Input: Segmented and feature extracted key frames

Output: Object

1. Extracting current frames
2. Analyze feature information available on all segmentation
3. Compare segmentation into key frames (pixel wise, color wise, shape/area)
4. Check for key frame annotation
5. Create segmentation for annotation
6. Specify annotation type
7. Compare new segmentation to current segmentation
8. Store to database

K-Means – Clustering

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{ci} (||xi - vi||)^2 \qquad \textbf{……… (1)}$$

Where,
'$||x_i - v_j||$' is the Euclidean distance between x_i and v_j.

'c_i' is the number of data points in i^{th} cluster.

'c' is the number of cluster centers.

Algorithmic steps for k-means clustering

Let X = {$x_1,x_2,x_3,……..,x_n$} be the set of data points and V = {$v_1,v_2,…….,v_c$} be the set of centres.

1. Randomly select 'c' cluster centres.

2. Calculate the distance between each data point and cluster centres.

3. Assign the data point to the cluster centre whose distance from the cluster centre is minimum of all the cluster centres..

4. Recalculate the new cluster centre using:

$$\boldsymbol{vi = (\frac{1}{ci}) \sum_{j=1}^{ci} xi} \qquad \textbf{..…….. (2)}$$

where, 'c_i' represents the number of data points in i^{th} cluster.

5. Recalculate the distance between each data point and new obtained cluster centers.

6. If no data point was reassigned then stop, otherwise repeat from step 3.

3.3 VISCOM (VIdeo Semantic COntent Model)

It is domain independent meta-ontology used for domain ontology construction [1]. It is solution to rule-based and domain dependent extraction methods. VISCOM provides standardized rule construction ability with the help of its meta-ontology. Rule construction process gets easier and makes its use possible on larger video data. VISCOM ontology can cover most of the event definitions for a wide variety of domains. VISCOM is composed of VISCOM classes (VC) and domain-independent VISCOM class individuals (DII). $VC_x = (VC_{xname}, VC_{xprop})$ where, VC_{xname} is the name of the class and VC_{xprop} is the set of relations and properties of class. Domain-independent VISCOM class individuals are grouped under movement, temporal, structural, and spatial relation types. DII = MRI U TRI U OCTI U SRI. For the semantic content representation, VISCOM ontology introduces fuzzy classes and properties. Spatial Relation Component, Event Definition, Similarity, Object Composed of Relation and Concept Component classes are fuzzy classes as they aim to having fuzzy definitions.

3.4 Domain Ontology

Using VISCOM domain ontology is constructed [1]. Basically, domain specific semantic contents are defined as individuals of VISCOM classes and properties. Each event definition uses different spatial and temporal relations between objects in order to define the event. The ontology developer always has a chance to add a new definition that will cover cases where existing definitions are not sufficient enough. Also has an opportunity to add new individual definitions, modify, or delete them at any time.

Domain ontology construction algorithm is given as:

Algorithm: .Domain Ontology Construction with VISCOM

Input: VISCOM

Output: Domain Ontology

1. Define O (objects), E (events) and C (concepts).

2. Define all possible Spatial Relations (SR) and Object Movements (OM) occurring within am E.

3. Use SR’s and OM’s to define Spatial Changes (SC’s). Describe temporal relations between SC’s as Temporal Spatial Change Component (TSCC’s).

4. Make Event Definition (EDs) with SC’s, SR’s and TSCC’s.

5. For all E’s if an event can be defined with an event def then define E in terms of ED’s.

6. If an event can be defined with temporal relations between other events then define E’s in terms of ETR’s.

7. For all C’s construct a relation with the C that can be placed in its meaning.

3.5 Spatial and Temporal Relations Extraction

Each spatial relation extraction is stored as a Spatial Relation Component instance contains frame number, object instances, type of spatial relation and fuzzy membership value of relation. It is calculated according to the positions of objects relative to each other. Spatial relations categorize as topological (Types- inside, partially inside, touches and disjoint), distance (Types- far, near) and positional (Types- above, below, left and right). Temporal relations are utilized to add temporality to sequence Spatial Change or Events individuals.

3.6 Event and Concept Extraction

Events are extracted using spatial and temporal relations between object instances. An extracted event instance is represented with a type, a frame set representing the events interval, a membership value and the roles of the objects taking part in the event. Frame Set is used to represent the frame interval of instances. Each extraction process outputs instances of a semantic content type defined as an individual in the domain ontology. In the concept extraction process, Concept Component individuals and extracted object, event, and concept instances are used. Concept Component individuals relate objects, events, and concepts with concepts. When an object or event that is used in the definition of a concept is extracted, the related concept instance is automatically extracted with the relevance degree given in its definition.

Algorithm: Event Extraction

Input: Domain Ontology, Object Instances

Output: Event Instances

1. Getting all object instances, which related to user input query

2. Find the spatial movement, relations and changes

3. Load all event definition, which is related to all available object instances

4. Find the similarity between 0.5 and 1, (maximum related only)

5. Display the similar event instances

Algorithm: Concept Extraction

Input: Domain Ontology, Object Instances Event Instances

Output: Event Instances, Concept Instances

1. Load all ontology and spatial, which related to object instances by user input query

2. If check is there are object or event instances that satisfy the individual definition, than extract the concept instances that satisfy the individual definition

3. Get the specified rule for the extracted concept and execute it

4. RESULT and ANALYSIS

Proposed system is compared with the genetic algorithm based system. The test results of these systems are discussed here. Fig2 gives graph of Number of key frames given by systems for football ontology. In this graph 12 videos are given for testing. Numbers of frames related to given event are then extracted. The test results shows that proposed system gives more accurate and more numbers of frames than the genetic algorithm based system.

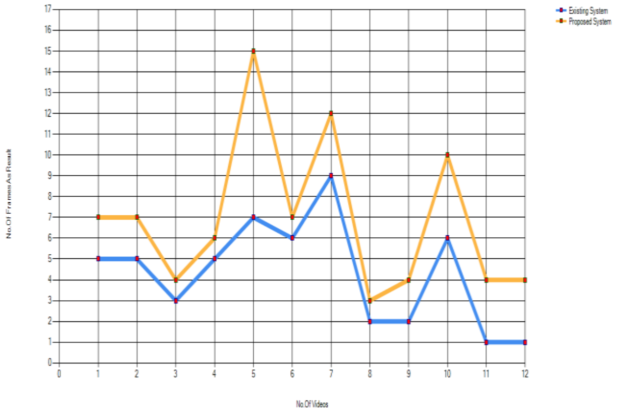

Fig 2. Number of Key Frames for Football Ontology

Fig3 shows resuts for frame segmention of a given video. i.e. number of segments done in a frame. In proposed system frame is divided into less number of segments than genetic based approach. As the number of segments increased it gives less number of objects as these segments related to same object. But, as they get segmented less relative results are get. As shwon in graph existing system creates more number of segments than the proposed system. So, number of extracted objects in proposed system are more accurate and relevant.

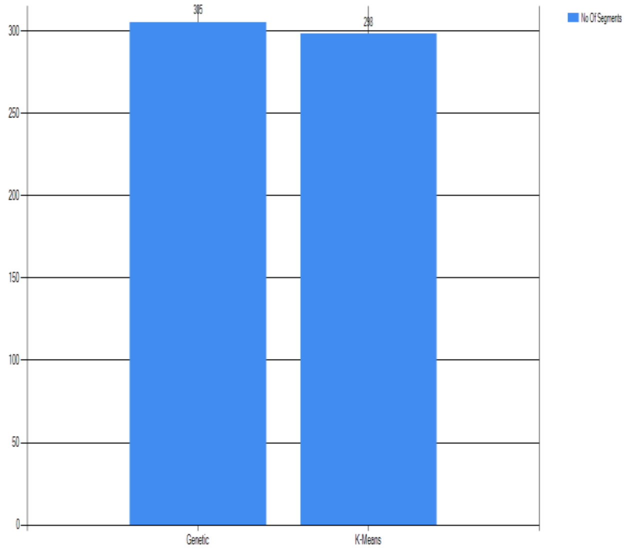

Fig 3. Frame Sgmentation

Table 1. Number of Frames

Goal (Event)			Kick(Event)			Pass(Event)			Attacking (Concept)		
P. S.	E. S.	Diff	P. S.	E. S.	Diff	P. S.	E. S.	Diff	P. S.	E. S.	Diff
7	5	2	6	5	1	12	9	3	10	6	4
7	5	2	15	7	8	3	2	1	4	1	3
4	3	1	7	6	1	4	2	2	4	1	3

As shown in the table1 results are taken from Fig2 in which first 3 videos are for goal event next 3 of kick event and next 3 of pass event last are attacking concept. From the table we analyze that there is some times there are large difference in event related frame extraction. Proposed system always gives more number of results than the exisiting ones.

From the result test cases we analyze that k-means based approach for object extraction gives more accurate and relevant results to related query. While genetic based approach gives less relevant results. For semantic content extraction and modelling based system to extract exact semantic content k-means algorithm is most useful.

5. CONCLUSION

To accomplish goal efficiently and effectively extract semantic content in video. A generic ontology-based semantic meta-ontology model for videos (VISCOM) is proposed. By adding fuzziness in class, relation, and rule definitions the semantic content representation capability and extraction success is improved. An automatic k-means based object extraction method is integrated to the proposed system to extract more relevant objects from video. Adding fuzziness to domain ontology construction more meaningful and relevant results for semantic search are obtain. It uses spatial temporal relations between objects and events for concept extraction.

K-means clustering algorithm for object extraction is proposed for semantic content. It gives more relevant and accurate results than the other approaches. For success of semantic content extraction two points must be assured. The first one is to obtain object instances correctly. Whenever a missing or misclassified object instance occurs in the object instance set, which is used by the framework as input, success of event and concept extraction decreases. The second issue is to use the proposed VISCOM meta-model effectively and construct well and correctly defined domain ontology. Wrong, extra, or missing definitions in the constructed ontology can decrease the extraction success.

6. REFERENCES

[1] Y. Yildirim, Adnan Yazici, TurgayYilmaz “Automatic Semantic Content Extraction in Video Using a Fuzzy Ontology and Rule-Based Model”, *in IEEE Trans. Knowledge and Data Engineering*, vol. 25, no. 1, pp. 47-61, Jan. 2013.

[2] M. Petkovic, W. Jonker, “An Overview of Data Models and Query Languages for Content-Based Video Retrieval,” *Proc. Int’l Conf. Advances in Infrastructure for E-Business on the Internet”,* Aug. 2000.

[3] M. Petkovic “Content-Based Video Retrieval by Integrating Spatio-Temporal and Stochastic Recognition of Events,” *Proc. IEEE Int’l Workshop Detection and Recognition of Events in Video*, pp. 75-82, 2001.

[4] G.G. Medioni, I. Cohen, F. Bre´mond, S. Hongeng, and R. Nevatia, “Event Detection and Analysis from Video Streams,” *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 23, no. 8, pp. 873-889, Aug. 2001.

[5] S. Hongeng, R. Nevatia, “Video-Based Event Recognition: Activity Representation and Probabilistic Recognition Methods,”*Computer Vision and Image Understanding*, vol. 96, no. 2, pp. 129-162, 2004.

[6] A. Hakeem and M. Shah, “Multiple Agent Event Detection and Representation in Videos,” *Proc. 20th*

Nat'l Conf. Artificial Intelligence (AAAI), pp. 89-94, 2005.

[7] M.E. Do¨nderler, E. Saykol, U. Arslan, OUlusoy, and U. Gu¨du¨ kbay, "Bilvideo: Design and Implementation of a Video Database Management System," *Multimedia Tools Applications*, vol. 27, no. 1, pp. 79-104, 2005.

[8] T. Sevilmis, M. Bastan, U. Gudukbay, O. Ulusoy, "Automatic Detection of Salient Objects and Spatial Relations in Videos for a Video Database System," *Image Vision Computing*, vol. 26, no. 10, pp. 1384-1396, 2008.

[9] M. Ko¨pru¨ lu¨, N.K. Cicekli, and A. Yazici, "Spatio-Temporal Querying in Video Databases," *Information Sciences*, vol. 160, nos. 1-4, pp. 131-152, 2004.

[10] L. Bai, S.Y. Lao, G. Jones, and A.F. Smeaton, "Video Semantic Content Analysis Based on Ontology," IMVIP '07: Proc. 11th Int'l Machine Vision and Image Processing Conf., pp. 117-124, 2007.

[11] A.D. Bagdanov, M. Bertini, A. Del Bimbo, C. Torniai, and G. Serra, "Semantic Annotation and Retrieval of Video Events Using Multimedia Ontologies," *Proc. IEEE Int'l Conf. Semantic Computing (ICSC)*, Sept. 2007.

[12] R. Nevatia and P. Natarajan, "EDF: A Framework for Semantic Annotation of Video," Proc. *10th IEEE Int'l Conf. Computer Vision Workshops (ICCVW '05)*, p. 1876, 2005.

[13] R. Nevatia, J. Hobbs, and B. Bolles, "An Ontology for Video Event Representation," *Proc. Conf. Computer Vision and Pattern Recognition Workshop*, p.119.

[14] J. Fan, W. Aref, A. Elmagarmid, M. Hacid, M. Marzouk, and X. Zhu, "Multiview: Multilevel Video Content Representation and Retrieval," *J. Electronic* Feb. 2004. *Imaging*, vol. 10, no. 4, pp. 895-908, 2001.

[15] J. Fan, A.K. Elmagarmid, X. Zhu, W.G. Aref, and L. Wu, "Classview: Hierarchical Video Shot Classification, Indexing, and Accessing," *IEEE Trans. Multimedia*, vol. 6, no. 1, pp. 70-86.

[16] Y. Yildirim and A. Yazici, "Ontology-Supported Video Modeling and Retrieval," *Proc. Fourth Int'l Conf. Adaptive Multimedia Retrieval: User, Context, and Feedback (AMR)*, pp. 28-41, 2006.

[17] T. Yilmaz, "Object Extraction from Images/Videos Using a Genetic Algorithm Based Approach," master's thesis, Computer Eng. Dept., METU, Turkey, 2008.