ISSN 2229-5518

Efficient Semantic Video Data Extraction By Sequential VISCOM Mining

P.Kanagavalli

Abstract—The use of video-based applications has revealed the need for extracting the content in videos. Raw data and low-level features alone are not sufficient to fulfill the user 's needs; that is, a deeper understanding of the content at the semantic level is required.Here, propose a semantic content extraction system that allows the user to query and retrieve objects, events, and concepts that are extracted automatically. an ontology-based fuzzy video semantic content model that uses spatial/temporal relations in event and concept definitions.

Index Terms— Semantic content extraction, video content modeling, fuzziness, ontology,spatial extraction,Temporal Extraction Event Extraction.

1 INTRODUCTION

THE rapid increase in the available amount of video data has caused an urgent need to develop intelligentmethods to model and extract the video content. Typical applications in which modeling and extracting videocontent are crucial include surveillance, video-on-demand systems, intrusion detection, border monitoring, sportevents, criminal investigation systems, and many others.

The ultimate goal is to enable users to retrieve some desired content from massive amounts of video data in an efficient and semantically meaningful manner. There are basically three levels of video content which are raw video data, low-level features and semantic content.First, raw video data consist of elementary physical video units together with some general video attributes such as format, length, and frame rate. Second, low-level features are characterized by audio, text, and visual features such as texture, color distribution, shape, motion, etc. Third, semantic content contains high-level concepts such as objects and events.

The first two levels on which content modeling and extraction approaches are based use automatically extracted data, which represent the low-level content of a video, but they hardly provide semantics which is much more appropriate for users. Users are mostly interested in querying and retrieving the video in terms of what the video contains. Therefore, raw video data and low-level features[1] alone are not sufficient to fulfill the user's need; that is, a deeper understanding of the information at the semantic level is required in many video-based applications.However, it is very difficult to extract semantic content directly from raw video data. This is because video is a temporal sequence of frames without a direct relation to its semantic content.

Therefore, many different representations using different sets of data such as audio, visual features, objects, events, time, motion, and spatial relations are partially or fully used to model and extract the semantic content. No matter which type of data set is used, the process of extracting semantic content is complex and requires domain knowledge or user interaction.There are many research works in this area. Most of them use manual semantic content extraction methods

2 VIDEO SEMANTIC CONTENT MODEL

Ontology-based fuzzy VIdeo Semantic COntent Model (VISCOM) that uses objects and spatial/temporal relations[2] in event and concept definitions.VISCOM is a metaontology for domain ontologies provides a domain-independent rule construction standard. VISCOM has a number of classes representing semantically meaningful components of videoVCxname ={Component; Object;Event;Concept; Similarity;.}.Domain-independent V ISCOM class individuals are grouped under movement, temporal, structural, and spatial relation types. DII = MRI U TRI U OCTI U SRI, MRI = fdown; up; right; leftg is the set of movement relation types, TRI = {before; meets; starts; finish; overlaps; equal; during} is the set of temporal relation types, OCTI = {composedOf; isA; partOf; substanceOf} is the set of relation types used to define concept inclusion, membership andstructural object relations, and SRI = DSRI UPSRI UTSRI is the set of spatial relation types, TSRI = {finside; partiallyInside; disjoint; touch}is the set of topological, PSRI = {right; left; above; below}is the set of positional, and DSRI = {far; near}is the set of distance spatial relation types.

2.1 Ontology Based Modelling

The linguistic part of VISCOM contains classes and relations between these classes. Some of the classes[14] represent semantic
content types such as Object and Event while others are used in the automatic semantic content extraction process.

2.2 Rule Based Modelling

Additional rules are utilized to extend the modeling capabilities. Each rule has two parts as body and head where body part contains any number of domain class or property individuals and head part contains only one individual with a value, _, representing the certainty of the definition given in the body part to represent the definition in the head part where0 _ _ _ 1. The basic syntax of rules has parentheses and logical connectives (^; _; .; 8; 9) in both body and head parts.

2.3 Domain ontology Construction with VISCOM

VISCOM is utilized as a metamodel to construct domain ontologies. Basically, domain specific semantic contents are defined as individuals of VISCOM classes and properties

Algorithm 1 presents the steps followed to construct a domain ontology by using VISCOM. For the evaluation purposes, we have constructed an Office Surveillance Ontology,a Basketball Ontology[16] and a Football Ontology by usingVISCOM. A small portion of the basketball ontology is illustrated in Fig. 2 for Rebound event, as an example

$$DomainOntology: \left\{ \begin{array}{l} \left[\begin{array}{l} metaModel \Rightarrow [VISCOM], \\ domain \Rightarrow \{D_i\}, \\ classInds \Rightarrow \{CI_0 \cdots CI_k\}, \\ dataPropInds \Rightarrow \{DP_0 \cdots DP_l\}, \\ objectPropInds \Rightarrow \{OP_0 \cdots OP_m\} \end{array} \right] \\ where \\ CIs\ are\ domain\ class\ inds, \\ DPs\ are\ domain\ data\ property\ inds, \\ OPs\ are\ domain\ object\ property\ inds, \\ ind(D_i, Domain). \end{array} \right.$$

Algorithm 1. Ontology Construction with VISCOM
Require. VISCOM
Ensure. Domain Ontology
1. define O, E and C individuals.
2. define all possible SR's occuring within an E.
3. define all possible OM's occuring within an E.
4. use SR's and M's to define SC's.
5. describe temporal relations between SC's as TSCC's.
6. make EDs with SC's, SR's and TSCC's.
7. for all E's do
8. if an event can be defined with an event def then
9. define E in terms of ED's.
10. end if
11. if an event can be defined with temporal relations between other events then
12. define E's in terms of ETR's.
13. end if
14. end for
15. for all C's do
16. construct a relation with the C that can be placed in its meaning.
17. end for
18. define S's.

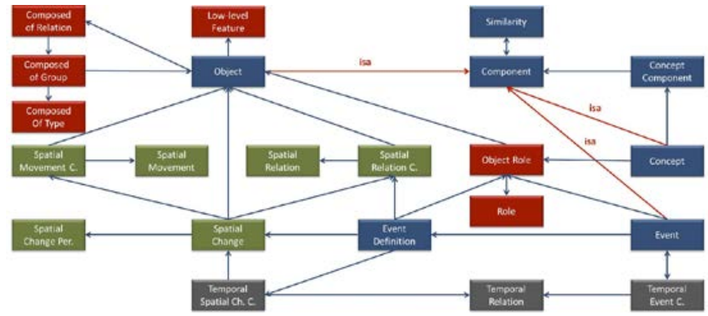

Fig 1.VISCOM classes and Reations

3 AUTOMATIC SEMANTIC CONTENT EXTRACTION FRAMEWORK

The ultimate goal of ASCEF is to extract all of the semantic content existing in video instances. Inorder to achieve this goal, the automatic semantic content extraction [17]framework takes Vi, ONTi, and Ri, where Vi is a video instance, ONTi is the domain ontology for domain Diwhich Vi belongs to, and Ri is the set of rules for domain Di.The output of the extraction process is a set of semantic contents, named VSCi, and represented as VSC ,Vi=OIi; EIi;KIii. OIi = OIi0 ; . . .;OIi is thset of object instances occurring in Vi, where an object instance is represented as OI.. MBR is the minimum bounding rectangle surrounding the object instance._ represents the certainty of the extraction, where 0 _ _ _ 1. type is an individual of a class CI in ontology ONT. EI isEI0 ; . . . ; EIing is the set of event instances

3.1 Object Extraction

Object extraction [18]is one of most crucial components in the framework, since the objects are used as the input for the extraction process.

3.2 Spatial Extraction

Every spatial relation extraction is stored as a SpatialRelation Component [8] instance which contains the framenumber, object instances, type of the spatial relation, and a fuzzy membership value of the relation.

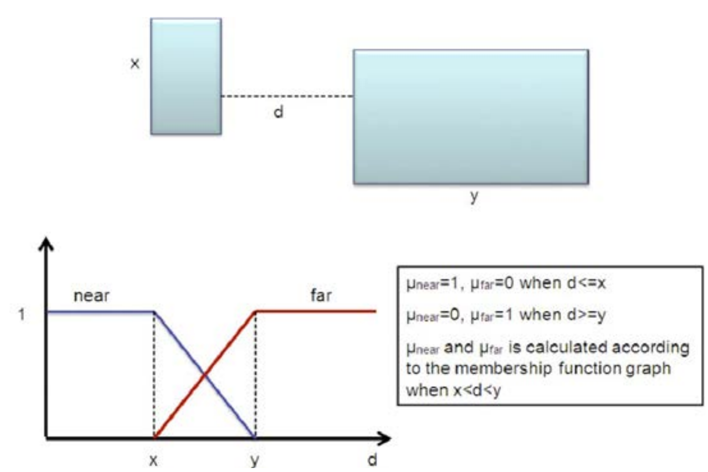

Fig. 2. Graph for distance relation membership function

3.3 Temporal Extraction

ISSN 2229-5518

In the framework, temporal relations are utilized in order toadd temporality to sequence Spatial Change or Events individuals in the definition of Event individuals

Office Surveillance Ontology Facts

Class/Rule Type	# Definitions in Ontology	Extraction Count
Object	13	15371
Spatial Relation Component	18	10342
Positional Spatial Relation	4	2704
Distance Spatial Relation	2	3960
Topological Spatial Relation	4	3678
Spatial Change	10	31
Spatial Movement Component	2	23
Event	10	44
Event Definition	11	40
Temporal Event Component	1	4
Temporal Spatial Change Component	5	4
Concept	3	6
Concept Component	5	6

Fig. 3. ontology facts

3.4 Event Extraction

Event instances are extracted after a sequence of automatic extraction processes. Each extraction process[6] output instances of a semantic content type defined as an individual in the domain ontology.

Algorithm 2. Event Extraction Algorithm
Require. Domain Ontology, Object Instances
Ensure. Event Instances
1. for all SRC individuals in the ontology do
2. extract SRC instances that satisfy the individual def.
3. execute SR rule def
4. end for
5. for all SMC individuals in the ontology do
6. extract SMC instances that satisfy the individual def.
7. end for
8. for all SC individuals in the ontology do
9. check if there are SRC or SMC instances that satisfy the individual def.
10. end for
11. for all TSC individuals in the ontology do
12. extract SC instances that satisfy the individual def.
13. end for
14. for all ED individuals in the ontology do
15. check if there are SC, SR or TSC instances that satisfy the individual def.
16. end for
17. for all E individuals in the ontology do
18. check if there are ED instances that satisfy the individual def.
19. end for
20. for all Event individuals which have Temporal Event Component individuals do
21. extract Event instances that satisfy the individual def.
22. end for
23. for all S individuals in the ontology do
24. extract E instances that satisfy the individual def.
25. end for
26. execute all rules defined for E individuals to extract additional events.

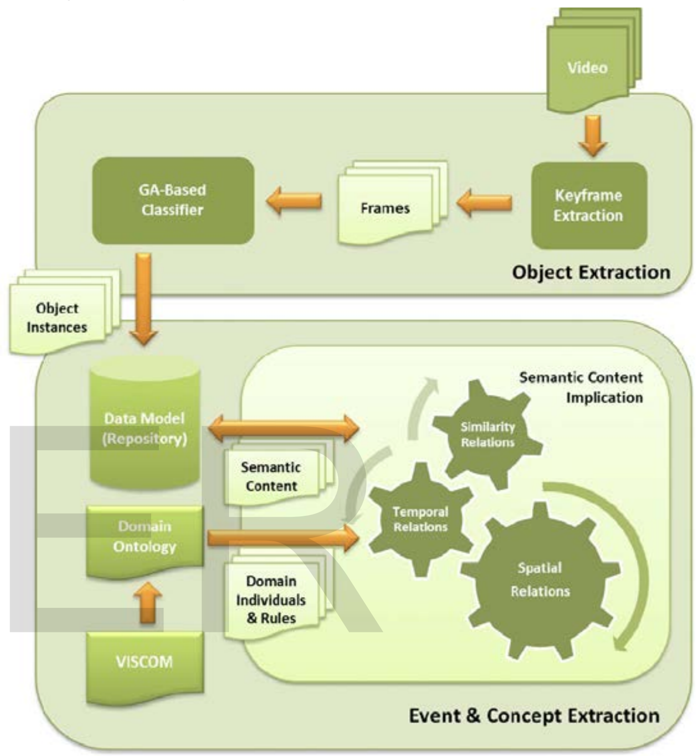

Fig. 4. Automatic semantic content extraction framework

3.5 Concept Extraction

In the concept extraction process, Concept Component individualsand extracted object, event, and concept instances are used. Concept Component individuals[5] relate objects, events,and concepts with concepts. When an object or event that isused in the definition of a concept is extracted, the relatedconcept instance is automatically extracted with the relevancedegree given in its definition. In addition, Similarity individualsare utilized in order to extract more concepts from the extracted components. The last step in the concept extraction process is executing concept rule definitions.

Algorithm 3. Concept Extraction Algorithm
Require. Domain Ontology, Object Instances, Event Instances
Ensure. Event Instances, Concept Instances
1. for all CC individuals in the ontology do
2. check is there are O or E instances that satisfy the individual def.

ISSN 2229-5518

3. end for
4. for all S individuals in the ontology do
5. extract C instances that satisfy the individual def.
6. end for
7. execute all rules defined for C individuals.

The framework is tested with three basketball videos, each being 2 minutes in length. Totally, 207 keyframes are extracted and utilized in the extraction process. The videos contain eight semantic entities, where the extraction resulted with seven correct, one wrong, and one missed entities.For this test, manually annotated object instances are utilized and membership value for object instances is defined as 90 percent. Wrong extractions in this test are the result of the unsuitable similarity class definition for the rebound event.

The framework is also tested with five football videos, each being 2 minutes in length. Totally, 312 keyframes are extracted and utilized in the extraction process. Manually annotated object instances are utilized and membership values for object instances are defined as 0.90 for football domain tests. Three event types are defined in the domain..

Scores for Basketball Videos

Name	μ	Prec(%)	Recall(%)	$Prec_{ind}$(%)	Rec_{ind}(%)	BDA(%)
Rebound(Event)	0.66	50.00	50.00	62.50	55.56	62.50
JumpBall(Event)	0.73	100.00	100.00	94.23	100.00	97.23
FreeThrow(Event)	0.72	100.00	100.00	95.00	100.00	95.00
Attack(Concept)	0.65	100.00	100.00	94.65	100.00	94.65
Total	0.69	87.50	87.50	88.42	92.36	89.34

Fig. 5.Basketball Videos

4. Conclusion

The semantic content extraction process is done automatically. In addition, a generic ontology-based semantic metaontology model for videos (VISCOM) is proposed. Moreover, the semantic content representation capability and extraction success are improved by adding fuzziness in class, relation, and rule definitions.

An automatic Genetic Algorithm-based object extraction method is integrated to the proposed system to capture semantic content. In every component of the framework, ontology-based modeling and extraction capabilities are used. 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.

Acknowledgment

. This work is supported by a research grant from TU¨ B_ITAK EEEAG with grant number 109E014.

References

[1] M. Petkovic and W. Jonker, "An Overview of Data Models andQuery Languages for Content-Based Vid eo Retrieval," Proc. Int'lConf. Advances in Infra structure for E-Business, Science, and Educationon the Internet, Aug. 2000.
[2] M. Petkovic and W. Jonker, "Content-Based Video Re tri eval byIntegrating Spatio-Temporal and Stochas tic Recognition of Events," Proc. IEEE Int'l Workshop Detection and Recognition oEvents in Video, pp. 75-82, 2001.
[3] L.S. Davis, S. Fejes, D. Harwood, Y. Yacoob, I. Hara toglu, and M.J.Black, "Visual Surveillance of Human Activity," Proc. Third AsianConf. Computer Vision (ACCV), vol. 2, pp. 267-274, 1998.
[4] G.G. Medioni, I. Cohen, F. Bre´mond, S. Hongeng, and R. Neva tia,"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, and F. Bre´mond, "Video- Based Eve nRecognition. Activity Representa tion and Probabilistic Recognition Methods," Computer Vision and Image Understanding, vol. 96no. 2, pp. 129-162, 2004.
[6] A. Hakeem and M. Shah, "Multiple Agent Event Detec tion andRepresentation in Videos," Proc. 20th Nat'l Conf. ArtificiaIntelligence (AAAI), pp. 89-94, 2005.
[7] M.E. Do¨nderler, E. Saykol, U. Arslan, O ¨ . Ulusoy, and U.Gu¨du¨ kbay, "Bilvideo. Design and Implementation of a VideoDatabase Management System," Multimedia Tools Applications,vol. 27, no. 1, pp. 79-104, 2005.
[8] T. Sevilmis, M. Bastan, U. Gu¨du¨ kbay, and O ¨ . Ulusoy, "Auto maticDetection of Salient Objects and Spatial Relations in Videos for a Video Database System," Image Vision Computing, vol. 26, no. 10pp. 1384-1396, 2008.
[9] M. Ko¨pru¨ lu¨, N.K. Cicekli, and A. Yazici, "Spatio-Temporal Querying in Video Databases," Information Sciences, vol. 160nos. 1-4, pp. 131-152, 2004.
[10] J. Fan, W. Aref, A. Elmagarmid, M. Hacid, M. Marzouk, and XZhu, "Multiview. Multilevel Video Content Representa tion andRetrieval," J. Electronic Imaging, vol. 10, no. 4, pp. 895-908, 2001.
[11] J. Fan, A.K. Elmagarmid, X. Zhu, W.G. Aref, and L. Wu,"Classview. Hierarchical Video Shot Classification, Iing,andAccessing," IEEE Trans. Multimedia, vol. 6, no. 1, pp. 70-86,Feb. 2004.
[12] L. Bai, S.Y. Lao, G. Jones, and A.F. Smeaton, "Video Semantic Content Analysis Based on Ontology," IMVIP '07. Proc. 11th Int'lMachine Vision and Image Processing Conf., pp. 117,124,2007.

[13] R. Nevatia and P. Natarajan, “EDF. A Framework for Semantic Annotation of Video,” Proc. 10th IEEE Int’l Conf. Computer VsionWorkshops (ICCVW ’05), p. 1876, 2005.

[14] A.D. Bagdanov, M. Bertini, A. Del Bimbo, C. Torniai, and G. Ser ra,“Semantic Annotation and Retrieval of Video Events Uing Multimedia Ontologies,” Proc. IEEE Int’l Conf. Semantic puing(ICSC), Sept. 2007.

[15] R. Nevatia, J. Hobbs, and B. Bolles, “An Ontology for Video EventRepresentation,” Proc. Conf. Computer Vision and Pattern Recognitionshop,p.119,http://ieeexplore.ieee.org/xpls/abs_all. jsp?arnumber=1384914, 2004.

[16] U. Akdemir, P.K. Tu raga, and R. Chellappa, “An On tology BasedApproach for Activity Recognition from Vid eo,” Proc. ACM Int’lConf. Multimedia, A. El-Saddik, S. Vuong, C. Griwodz, A.D. Bimbo,K.S. Candan, and A. Jaimes, eds., pp. 709-712, http://dblp.unitrier. de/db/conf/mm/mm2008.html#AkdemirTC08, 2008.

[17] Y. Yildirim, “Automatic Semantic Content Extraction in VideoUsing a Spatio-Temporal Ontology Model,” PhD dis sertation,Computer Eng. Dept., METU, Turkey, 2009.

[18] T. Yilmaz, “Object Extraction from Images/Videos Using a Genetic Algorithm Based Approach,” master’s thesis, Computer Eng. Dept., METU, Turkey, 2008.

[19] 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.

[20] Y. Yildirim, T. Yilmaz, and A. Yazici, “Ontology-Supported Objectand Event Extraction with a Genetic Algorithms Ap proach forObject Classification,” Proc. Sixth ACM Int’l Conf. Image and VideoRetrieval (CIVR ’07), pp. 202-209, 2007.