Proc. Fourth IEEE International Conference on Semantic Computing (ICSC 2010), Pittsburgh, PA, USA, September 2010, pp. 79-84.

A joint content-event model for event-centric multimedia indexing

Nikolaos Gkalelis, Vasileios Mezaris, Ioannis Kompatsiaris
Informatics and Telematics Institute
Centre for Research and Technology Hellas
6th Km Charilaou-Thermi Road, Thermi 57001, Greece
Email: {gkalelis, bmezaris, ikom}@iti.gr

***Abstract*—In this paper, a joint content-event model for indexing multimedia data is proposed. The event part of the model follows a number of formal principles to represent several aspects of real-life events, whereas the content part is used to describe the decomposition of any type of multimedia data to content segments. In contrast to other event models for multimedia indexing, the proposed model treats events as first class entities and provides a referencing mechanism to link real-life event elements with content segments at multiple granularity levels. This referencing mechanism has been defined with the objective to facilitate the automatic enrichment of event elements with information extracted by automatic analysis of content segments, enabling event-centric multimedia indexing in large-scale multimedia collections.**

***Keywords*-events; multimedia content; content-event model; event-based indexing;**

I. Introduction

In our days, a vast amount of multimedia data is daily produced and consumed within the framework of networked media, e.g., social web, web search engines, news organizations, and other. However, the quality of service of multimedia indexing tools, which are necessary for navigating within these data, is still far from reaching the required level. This is mainly due to the so-called semantic gap between the semantic descriptions of multimedia data provided by automatic analysis tools and the interpretations of the same multimedia data by humans [1]. Indeed, most machine algorithms decompose multimedia data to content segments, e.g., shots, scenes, etc., and index them using low-level feature vectors or limited higher-level metadata (e.g. visual concepts), while humans remember real life using past experience structured in events [2]. The necessity of formal event models for the description of real life events has been recently acknowledged, and a number of such models have been developed [3]–[12]. These however either lack or offer little support for describing and indexing multimedia content, or treat events as second class entities, i.e., the existence of events depends on the content they describe.

In this paper, we propose a joint content-event model, to address the limitations of the current event models and promote automatic event-centric multimedia indexing. The event part of the model allows representing real life events and their elements, i.e., where the event occurred, participants of the event and so on, while the content part is used to describe the decomposition of the multimedia data to content segments. The main contributions of the model are summarized below:

- In contrast to other event models for multimedia indexing, it treats events as first class entities.
- A referencing mechanism is provided for the automatic enrichment of event elements with information extracted by automatic analysis of multimedia content.

In section II a review of the state of the art in multimedia indexing is provided, while in section III a set of event model requirements identified by studying the state of the art is summarized. In section IV the proposed joint content-event model is described in detail. A brief example of the usage of the proposed model is shown in section V. Finally, concluding remarks are given in section VI.

II. State of the art in multimedia indexing

In order to put our model in context, we review related multimedia indexing methods, metadata standards and models with emphasis on semantics and event-based indexing. For a general review of multimedia indexing strategies the interested reader may refer to [1], [13].

In general, a piece of multimedia data may consist of multiple sources of information, such as text, images, audio, video. Moreover, access to it at different granularity levels may be required, e.g., to search or retrieve the entire video or just a specific video shot or scene. To address these issues, a large number of content-based indexing techniques have been proposed; these can be categorized as follows.

1) Indexing using perceptual information: A large fraction of content-based indexing approaches use media segmentation algorithms together with objective measurements at the perceptual level, i.e., derive features by processing the low-level visual or audio information within each content segment. These features are then used for indexing the data, e.g., MPEG-7 color and texture features or SIFT points for images [14]. Such approaches, although they are a necessary part of any multimedia indexing scheme, when used in isolation present several limitations, the most important being that they fail to capture the conceptual and contextual information conveyed by the multimedia content.

2) Concept-based indexing: Many works have appeared on combining the aforementioned low-level features with machine learning algorithms in order to achieve the association of content with concepts such as “person”, “outdoors”, etc., or different actions [15], [16]. The content segments can then be retrieved with the use of the detected concepts [13]. These methods represent a significant improvement over the methods of the previous category. However, they still do not fully capture the meaning that the content has to a human observer, who typically “sees” in it more than just a few depicted objects or elementary actions. For this to happen, the automatically detected concepts need to be seen in combination and be used for deriving higher-level interpretations of the content.

3) Semantics-based indexing: A large number of multimedia indexing techniques based on technologies related to the vision of the Semantic Web [17] have recently emerged in various application domains [18]–[20] to support the higher-level interpretation of the content. These include attempts to develop an MPEG-7 multimedia ontology using RDF or OWL. However, in these latter attempts, MPEG-7 related problems often arise [21]–[23]. In [22], the core ontology for multimedia (COMM) is proposed, which builds upon the descriptive ontology for linguistic and cognitive engineering (DOLCE) [24]. However, COMM may yield very complex and large RDF graphs in order to describe large pieces of multimedia data, e.g. videos [25]. Another important drawback of COMM (and similarly of all other MPEG-7 ontologies) is that the resulting multimedia annotations are not centered around events, in contrast to the fact that human memory organizes experiences in events.

4) Event-based indexing: During the last years there is a growing interest on event-centric multimedia systems [26].

In [11], a semantic-syntactic video model (SsVM) is proposed, which provides mechanisms for video decomposition and semantic description of video segments at different granularity levels. In [12], the video event representation language (VERL) and the video event markup language (VEML) are presented for the description and annotation of events in videos, respectively. In both models, events are not treated as first class entities.

In [3], the IPTC G2 family of news exchange standards are provided, including EventML, a standard for describing events in a journalistic fashion. In [4], the conceptual reference model (CRM) ISO standard of the International Committee for Documentation (CIDOC) is described, aiming to provide formal semantics for the integration of multimedia data in cultural heritage domain applications. Both standards treat events as first class entities; however, they only provide limited support for multimedia content description.

In [27], the Event Ontology (EO) [5] is implemented in OWL and is used to describe music events in several granularity levels. In [6], [7], the event model E for e-chronicle applications is provided and the authors compile the general requirements that a common event model should satisfy. In [8], this model is further extended and specialized in order to support the description of events in multimedia data, similar to the way that COMM provides description of semantic entities in multimedia data. In [28], the event model F is proposed, which, in contrast to event model E, is based on the DOLCE foundational ontology to provide formal semantics and representation of context. In [9], the linked data event model (LODE) is designed in order to link descriptions of the ABC, CIDOC, DOLCE, and EO models. In [29], [30], two event models based on the generic models E and F, as well as novel sets of event composition operators are presented. In [10], the use of event-based ontologies is proposed to close the so-called “glocal” gap between experience captured locally in personal multimedia collections, and common knowledge formalized globally. The models reviewed in this paragraph provide little or no support for capturing the structure of multimedia content.

III. Event model requirements

From the above literature review and inspired mostly by [7], we list a number of aspects that should be covered by an event model for multimedia indexing.

1) Formality aspect: It is important that the relationships, properties and domain concepts of the event model are formally defined, e.g., by enforcing the use of foundational ontologies. Formally defined models can significantly facilitate the subsequent development of event-processing tools for querying and retrieving relevant multimedia data.

2) Informational aspect: This aspect refers to the information regarding the event itself, i.e., the name and type of the event, and may further include information regarding the participation of agentive and non-agentive entities, e.g., an amount of money or a human face.

3) Experiential aspect: Multimedia data comprise the experiential dimension of an event. Events may need to address a specific content segment of multimedia data, e.g., a scene or a shot. In addition, events should form first class entities, i.e., it should be possible to define them independently of multimedia data. For the above two reasons *media decomposition* and *media independence* should be considered when addressing the experiential aspect of an event model. Media decomposition can be accomplished by either designing a suitable content decomposition model or by providing instructions on how to use an existing one. On the other hand, media independence can be achieved by an appropriate mechanism for referencing content segments.

4) Temporal aspect: Event elements are closely related to the notion of time. Temporal information can be expressed using either *absolute* or *relative* times. There is a number of standards developed for this purpose, e.g., the OWL-Time [31], the W3C Datetime Formats [32], and other.

5) Spatial aspect: Similar to temporal information, the spatial properties of an event element should be able to capture its *absolute* or *relative* location information, e.g., using the Basic Geo (WGS84 lat/long) Vocabulary [33] or the Region Connection Calculus (RCC) [34].

6) Compositional aspect: This aspect refers to the compositional structure of events, i.e., the existence of composite events that are made of other events. This type of relationship should be explicitly defined.

7) Causal aspect: The creation of an event may alter the state of one or more events, e.g., a robbery event may result in the creation of several gunshot events. This cause-effect relationship is inherent in events and should be represented.

8) Interpretation aspect: The way humans perceive an event depends on their past experience and the given situation. This means that a single event may have different meaning for different people. Therefore, an event model should offer a mechanism for indicating that two or more different event descriptions refer to the same event.

9) Uncertainty aspect: The automatic instantiation of events and their annotation with multimedia involves the use of a number of algorithms to analyze and associate content segments with concepts and event elements. Such algorithms, however, hardly ever produce results with a 100% confidence; instead, they typically provide multiple contradictory results, each accompanied by a different confidence score. Consequently, an event model should allow for this uncertainty to be appropriately represented.

10) Complexity: The model used for the description and indexing of multimedia content plays an essential role in the whole architecture of multimedia management applications. For this reason, it should produce descriptions of low complexity that can be efficiently processed by the respective tools.

In Table I, we summarize the properties of the event models presented in section II with respect to the event aspects identified above, extending the similar table of [28]. A cell of the table may be filled with Yes, No or Limited (Lim.), to show that an event model supports, does not support, or offers limited support regarding an aspect. We also use the values High, Low and Avg. (Average) to characterize the complexity of the model. Finally, a dash (–) denotes that this aspect is not applicable to the respective model.

IV. Proposed content-event model

The joint content-event model proposed in this work has been designed to satisfy the requirements described in the previous section, and at the same time to facilitate automatic event-centric multimedia indexing. It is made of a content part and an event part, which are appropriately linked to each other.

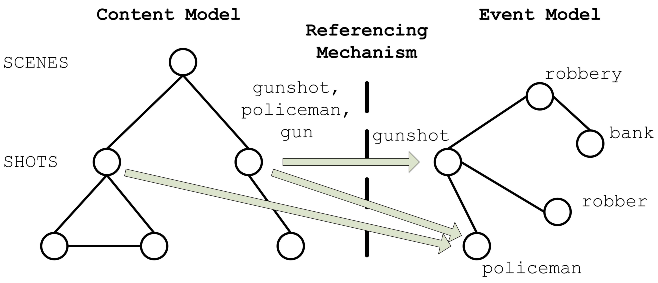

Figure 1. *Example of the proposed joint content-event model.*

The content part of the model has a hierarchical graph structure consisting of nodes and edges, e.g., as shown on the left side of Fig. 1. Content nodes are structurally alike, i.e., they all consist of the same set of properties, and one content node is used to convey information for exactly one content segment. Edges are used to connect two nodes at the same or different granularity levels, denoting a temporal or a compositional relationship, respectively.

The event part of the model has a more general graph structure, as shown on the right side of Fig. 1. Similarly to the content part of the model, an event node corresponds to one real-life event element, e.g, the event itself, a sub-event, etc., and all event nodes have the same structure. On the other hand, edges between nodes indicate a variety of relationships, e.g., spatiotemporal, causal, and other.

The properties of the event and content nodes are depicted in Fig. 2. We observe that there is a number of properties that are common in both the content and the event node. There is also a number of event node properties that their values can be inferred from the values of the respective content node properties. This set of properties constitutes the referencing mechanism of the proposed model: the values of these properties can be used to associate one or more content nodes with an event node, and, in the case of association, initialize several properties of the event node (Fig. 1).

A. Event node properties

The event node properties can be categorized with regard to the event model requirements drawn in Section III.

1) Formality aspect: An event in our model is formally defined as a graph where each node represents an element of the event, and an edge between two nodes indicates the existence of one or more relationships between these nodes. In addition, each node has a number of properties which are defined using formal classes from foundational ontologies.

2) Informational aspect: Four properties are used to cover this aspect: hasID, hasName, hasType and hasRole. The hasID property receives a URI to represent the event node in a global scope. The event element type is carried by the hasType property. Three classes from the ultra light version of DOLCE (DUL) are adopted for modelling the type of an event element, i.e., Event, Agent and Place [24]. The hasName property holds the name of the event element after its instantiation, e.g. Gunshot, Paris. The hasRole property,

Table I
Characteristics of existing event models: SsVM [11], VERL [12], EventML [3], CRM [4], EO [5], E [6], F [28], LODE [9], graph-based [29], [30].

		[11]	[12]	[3]	[4]	[5]	[6]	[28]	[9]	[29]	[30]
Formality aspect		Lim.	No	No	No	No	No	Yes	Yes	No	No
Informational aspect		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Experiential aspect	**Media decomposition**	Yes	Yes	Lim.	Lim.	–	No	–	–	–	–
	Media independence	No	No	Lim.	Lim.	–	Yes	–	–	–	–
Temporal aspect	**Absolute**	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Relative	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes
Spatial aspect	**Absolute**	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
	Relative	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes
Compositional aspect		Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Casual aspect		Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	Yes
Interpretation aspect		No	No	Yes	No	No	Yes	Yes	No	No	No
Uncertainty aspect		No	No	No	Yes	No	Yes	No	No	No	No
Complexity		Avg.	High	Avg.	High	Low	Avg.	High	Low	Low	Low

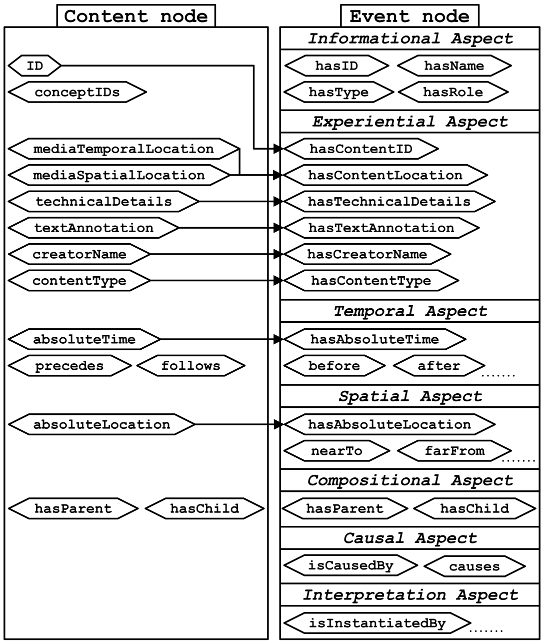

Figure 2. *Event and content node properties.*

adopted from DUL as well, is used to classify the event element in a given situation, e.g., it can classify a person as a policeman during a robbery event, or as the victim of a gunshot event.

3) Experiential aspect: The experiential aspect of an event is captured using six node properties: hasContentID, hasCreatorName, hasTextAnnotation, hasContentType, hasTechnicalDetails, hasContentLocation. These properties are automatically filled with information extracted from one or more content segments that may be associated with an event node, as shown in Fig. 2. This referencing mechanism requires that the multimedia data have been first decomposed and represented using the content part of the model, as described in section IV-B, and an association between the content and event node(s) has been established with the use of appropriate multimedia analysis tools. The hasTextAnnotation property is filled in with any textual annotation of the multimedia data, while the hasCreatorName property holds the name of the creator, extracted from the administrative information of the multimedia data. The hasTechnicalDetails property holds the technical details of the multimedia data, e.g. encoding and frame rate in case of video data. The hasContentID property is filled with the identification URI of the content node and the hasContentLocation property provides information concerning the actual position of the content segment, recorded in the mediaSpatialLocation and mediaTemporalLocation properties of the content part of the model (Section IV-B). We should note that the hasTechnicalDetails and the hasContentLocation properties capture all the necessary information to locate and use the content segment itself, avoiding the overhead for accessing the content description part of the model again.

4) Temporal aspect: We use the W3C Datetime Format profile of ISO 8601 standard [32] in order to fill the hasAbsoluteTime property and express absolute time regarding an event element. Relative time is expressed with respect to another temporal entity and captured using the temporal relations provided in Allen's Time Calculus [35].

5) Spatial aspect: We use the hasAbsoluteLocation property to capture absolute spatial location of an event element in latitude, longitude form as defined in the Basic Geo (WGS84 lat/long) Vocabulary [33]. We also use the nearTo and farFrom properties of DUL to denote relative distance between event elements, and the properties of RCC [34] to denote more complex spatial relations between two event elements.

6) Compositional aspect: We capture compositional information using the properties of hasParent and hasChild. These properties receive as values the IDs of the immediate super-events and sub-events related with the event in question, respectively.

7) Causal aspect: Causal information is captured using the properties causedBy and causes. The causes property is filled with the IDs of the events that are caused by the specific event, and the isCausedBy property is filled with the

IDs of the events that cause the current event.

8) Interpretation aspect: For this aspect we define the properties isInstantiatedBy and hasInstantiationTime to capture the creator and the creation time of an event node, and the sameAs property to link two or more event nodes in the case that they correspond to the description of the same event by different people. The sameAs property can be also used to connect event elements representing the same entity in different context.

9) Uncertainty aspect: The properties hasID, isInstantiatedBy, hasInstantiationTime, hasTextAnnotation, hasCreatorName, hasRemoteInfo, hasContentType, hasAbsoluteTime and hasAbsoluteLocation, are filled with "crisp" values or text created during the instantiation of the multimedia content or the event element. For all other properties, each value can be accompanied by a confidence score in the range $[0, 1]$.

10) Complexity: The event part of the model uses a referencing mechanism to index only the multimedia content that is relevant to the event, ignoring the rest of the content. Consequently, the description of the entire content is not included in the event description, and, the event part description alone can be used to search and retrieve the relevant multimedia content without the use of the content part description. This is an important advantage of the proposed model against models designed primarily for annotating multimedia data, e.g., [12], which incorporate content description in the event description, thus, considerably increasing the complexity of the event representation.

B. Content node properties

The content node properties are used to encapsulate information about the corresponding content segment, starting with the ID property, which is filled with a URI to index content nodes in a uniform manner. A type taxonomy similar to the Segment Description Scheme (DS) of MPEG-7 Multimedia DS (MDS) is deployed to characterize the type of content segments, e.g., audio segment, video segment, or further specializations such as scene, shot, etc. Type information is recorded in the contentType property of content nodes. Relative position of content nodes in the content graph is captured with the properties hasParent, isChild, precedes and follows. The two former receive a URI to express compositional information between two nodes, e.g., to indicate that a shot belongs to a scene or that a face is a part of a human body. In a similar fashion, the properties precedes and follows receive a URI to express relative temporal information between content nodes, e.g., to indicate that one shot appears before another. The actual spatiotemporal position of a content segment is recorded in the mediaSpatialLocation and mediaTemporalLocation properties. The mediaSpatialLocation property is used to describe a region of an image or frame. The mediaTemporalLocation property records information regarding the temporal position of a content segment, e.g., the start and the end frame of a shot. A set of properties is filled with information extracted from the metadata accompanying the content segment. The creatorName property holds the name of the creator, extracted from the administrative metadata, while the textAnnotation property is filled with the textual annotation of the multimedia data, if any. The absoluteLocation property is used to hold geographical identification information extracted from the geospatial metadata. Similarly, the absoluteTime property is used to host time-related information extracted from the timestamp metadata of the content segment. Concepts extracted from the content segments, e.g. using a concept detection algorithm, are described in the conceptIDs property.

V. Example of event-based multimedia indexing

In this section we provide an example of using the proposed model to describe a real-life event. In textual form this event may be described as "During the robbery of bank X, the robber, named V. P., shoots a policeman, named R. J., who guarded the bank". For simplicity of illustration we describe only the gunshot of the robbery. The description of the gunshot event using the event model is shown on the right side of Fig. 3. Such a description can be generated either manually or automatically, e.g., using linguistic analysis tools. The node ① of type Event is used to represent the Gunshot event and the nodes ② and ③ of type Person (subclass of DOLCE class Agent) are used to represent the Shooter and Victim of the Gunshot respectively. Given a video that is possibly related to the event, automatic analysis techniques are initially applied to it for performing spatiotemporal decomposition to scenes, shots, etc., and trained concept detectors are used for associating content segments with concepts, e.g., "policeman", "gun", etc. The video is then represented with the use of the proposed content model. A part of the video content representation is shown on the left side of Fig. 3. Exploiting the common properties of content nodes and event nodes, and particularly the presence of common concepts ("gunshot", "shooter", "victim", "R. J.", "V. P.") as values of properties hasName, hasRole and conceptIDs in the event and content nodes, the hasContentID and hasContentLocation properties of the event nodes are filled with the corresponding data of the relevant content node, i.e., automatic event-centric indexing of the video content is achieved. The result of this process is shown in Fig. 3, where content node Ⓐ is associated with event node ① and content node Ⓑ is associated with the event nodes ② and ③.

VI. Conclusions

We have presented a joint content-event model for indexing multimedia content. The proposed model addresses a set of requirements extracted after an extensive review of the relevant state of the art. The main advantages of the model are that it treats events as first class entities and it facilitates

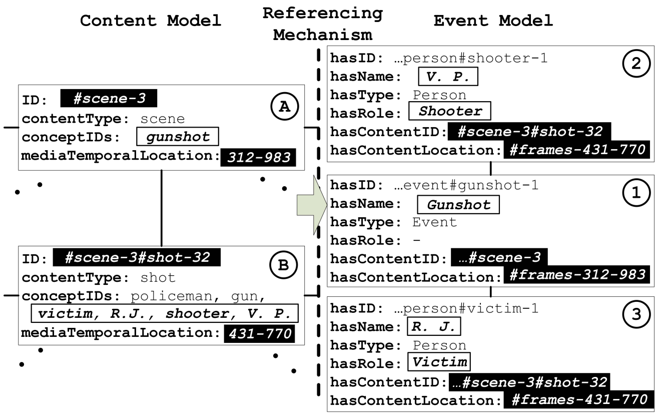

Figure 3. *Event-centric multimedia indexing example.*

automatic analysis by providing a set of common properties of event elements and content segments.

Acknowledgment

This work was supported by the European Commission under contract FP7-248984 GLOCAL.

References

[1] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, Aug. 2000.

[2] J. M. Zacks, T. Braver, M. Sheridan, D. Donaldson, A. Snyder, J. Ollinger, R. Buckner, and M. Raichle, "Human brain activity time-locked to perceptual event boundaries," *Nature Neuroscience*, vol. 4, no. 6, pp. 651–655, Jun. 2001.

[3] "International Press Telecommunications Council," http://www.iptc.org, accessed 2010-04-07.

[4] P. Sinclair, M. Addis, F. Choi, M. Doerr, P. Lewis, and K. Martinez, "The use of CRM core in multimedia annotation," in *Proc. 1st Workshop on Semantic Web Annotations for Multimedia (SWAMM)*, May 2006.

[5] "The event ontology," http://motools.sourceforge.net/event/event.html, accessed 2010-04-07.

[6] U. Westermann and R. Jain, "E - a generic event model for event-centric multimedia data management in e-chronicle applications," in *Proc. 22nd Int. Conf. on Data Engineering Workshops (ICDEW '06)*, Apr. 2006.

[7] ——, "Toward a common event model for multimedia applications," *IEEE Multimedia*, vol. 14, no. 1, pp. 19–29, Jan. 2007.

[8] A. Scherp, S. Agaram, and R. Jain, "Event-centric media management," in *Proc. IS&T/SPIE 20th Annual Symposium Electronic Imaging Science and Technology (SPIE)*, Jan. 2008.

[9] R. Shaw, R. Troncy, and L. Hardman, "LODE: Linking open descriptions of events," in *Proc. 4th Asian Semantic Web Conference (ASWC '09)*, Shanghai, China, Dec. 2009, pp. 153–167.

[10] G. Boato, C. Fontanari, F. Giunchiglia, and F. G. Natale, "Glocal multimedia retrieval," University of Trento, Tech. Rep. DISI-09-002, Jan. 2008.

[11] A. Ekin, A. M. Tekalp, and R. Mehrotra, "Integrated semantic-syntactic video modeling for search and browsing," *IEEE Trans. Multimedia*, vol. 6, no. 6, pp. 839–851, Dec. 2004.

[12] A. Francois, R. Nevatia, J. Hobbs, and R. Bolles, "VERL: An ontology framework for representing and annotating video events," *IEEE Multimedia*, vol. 12, no. 4, pp. 76–86, Oct. 2005.

[13] C. G. M. Snoek and M. Worring, "Concept-based video retrieval," *Foundations and Trends in Information Retrieval*, vol. 4, no. 2, pp. 215–322, 2009.

[14] S.-F. Chang, J. He, Y.-G. Jiang, E. E. Khoury, C.-W. Ngo, A. Yanagawa, and E. Zavesky, "Columbia University/VIREO-CityU/IRIT TRECVID2008 High-Level Feature Extraction and Interactive Video Search," in *Proc. TRECVID 2008*, USA, Nov. 2008.

[15] G. Papadopoulos, A. Briassouli, V. Mezaris, I. Kompatsiaris, and M. Strintzis, "Statistical motion information extraction and representation for semantic video analysis," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 19, no. 10, pp. 1513–1528, October 2009.

[16] V. Mezaris, A. Dimou, and I. Kompatsiaris, "Local invariant feature tracks for high-level video feature extraction," in *Proc. 11th Int. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, Desenzano del Garda, Italy, April 2010.

[17] N. Shadbolt, T. Berners-Lee, and W. Hall, "The semantic web revisited," *IEEE Intell. Syst.*, vol. 21, no. 3, pp. 96–101, May 2006.

[18] S. Dasiopoulou, V. Mezaris, I. Kompatsiaris, V. Papastathis, and M. Strintzis, "Knowledge-assisted semantic video object detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 10, pp. 1210–1224, Oct. 2005.

[19] A. Carbonaro, "Ontology-based video retrieval in a semantic-based learning environment," *Journal of e-Learning and Knowledge Society*, vol. 4, no. 3, pp. 203–212, Sep. 2008.

[20] V. Mezaris, S. Gidaros, G. Papadopoulos, W. Kasper, J. Steffen, R. Ordelman, M. Huijbregts, F. de Jong, I. Kompatsiaris, and M. Strintzis, "A system for the semantic multi-modal analysis of news audio-visual content," *EURASIP J. on Advances in Signal Processing*, 2010.

[21] R. Troncy, O. Celma, S. Little, R. Garcia, and C. Tsinaraki, "MPEG-7 based multimedia ontologies: Interoperability support or interoperability issue?" in *Proc. 1st Workshop on Multimedia Annotation and Retrieval enabled by Shared Ontologies*, Genova, Italy, Dec. 2007.

[22] R. Arndt, R. Troncy, S. Staab, L. Hardman, and M. Vacura, "COMM: designing a well-founded multimedia ontology for the web," in *The Semantic Web: ISWC 2007 + ASWC 2007*, ser. Lecture Notes in Computer Science, vol. 4825. Berlin: Springer, 2008, pp. 30–43.

[23] F. Nack, J. Ossenbruggen, and L. Hardman, "That obscure object of desire: Multimedia metadata on the Web, Part 2," *IEEE Multimedia*, vol. 12, no. 1, pp. 54–63, Jan./Mar. 2005.

[24] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider, "Sweetening ontologies with DOLCE," in *Proc. 13th Int. Conf. on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web (EKAW '02)*, London, UK, Oct. 2002, pp. 166–181.

[25] M. Vacura and R. Troncy, "Towards a simplification of COMM-based multimedia annotations," in *Proc. 5th Int. Conf. on Formal Ontology in Information Systems (FOIS'08)*, Saarbrucken, Germany, Oct. 2008, pp. 67–72.

[26] M. Das and A. Loui, "Detecting significant events in personal image collections," in *Proc. 3rd IEEE Int. Conf. on Semantic Computing (ICSC)*, Berkeley, CA, USA, Sep. 2009.

[27] Y. Raimond, S. A. Abdallah, M. Sandler, and F. Giasson, "The music ontology," in *Proc. 8th Int. Conf. on Music Information Retrieval (ISMIR '07)*, Sep. 2007.

[28] A. Scherp, T. Franz, C. Saathoff, and S. Staab, "F–a model of events based on the foundational ontology dolce+dns ultralight," in *Proc. 5th Int. Conf. on Knowledge Capture (K-CAP '09)*, Redondo Beach, California, USA, Sep. 2009, pp. 137–144.

[29] S. Rafatirad, A. Gupta, and R. Jain, "Event composition operators: ECO," in *Proc. 1st ACM Int. Workshop on Events in Multimedia (EiMM '09)*, Oct. 2009, pp. 65–72.

[30] P. K. Atrey, "A hierarchical model for representation of events in multimedia observation systems," in *Proc. 1st ACM Int. Workshop on Events in Multimedia (EiMM '09)*, Oct. 2009, pp. 57–64.

[31] "Time ontology in owl," http://www.w3.org/TR/owl-time/, accessed 2010-04-07.

[32] "Date and time formats," http://www.w3.org/TR/NOTE-datetime, accessed 2010-04-07.

[33] "Basic geo (wgs84 lat/long) vocabulary," http://www.w3.org/2003/01/geo/, accessed 2010-04-07.

[34] D. A. Randell, Z. Cui, and A. G. Cohn, "A spatial logic based on regions and connection," in *Proc. 3rd Int. Conf. on Knowledge Representation and Reasoning*, Jan. 1992, pp. 165–176.

[35] J. F. Allen, "Maintaining knowledge about temporal intervals," *Communications of the ACM*, vol. 26, no. 11, pp. 832–843, Nov. 1983.