MObile Multimedia Event Capturing and Visualization (MOME)

Amalia Panteli[#], Chrisa Tsinaraki[#], Lemonia Ragia[+], Fotis Kazasis[#], Stavros Christodoulakis[#]
[#] Department of Electronic and Computer Engineering, Technical University of Crete, Laboratory of Distributed Multimedia Information Systems and Applications (MUSIC), Chania, Greece
[+] Department of Architectural Engineering, Technical University of Crete, Chania, Greece
lia.hmmy@gmail.com, chrisa@ced.tuc.gr, ragia@isc.tuc.gr, fotis@ced.tuc.gr, stavros@ced.tuc.gr

***Abstract*— Capturing Multimedia Events such as natural disasters, accident reports, building damage reports, political events, etc., are expensive functionalities due to the number and training of the people required, as well as the time involved in the capturing and post-processing of multimedia. In addition, the captured multimedia content often fails to give the viewer a comprehensive understanding of the event captured in context. We present a model and a mobile system for multimedia event capturing by a one-man-crew. The system supports: (a) the real time capturing of complex multimedia events of different types; (b) the recording of the capturing process and the metadata associated with the events; (c) the visualization of the events and the capturing process; and (d) the learning and preparation of the one-man-crew that will do the multimedia event capturing.**

I. INTRODUCTION

Many important practical applications of multimedia are associated with events or they can be described by events. For example, reporting an accident, recording the events of a marriage or a baptism, recording an invited talk, a debate or a lecture, recording disasters like fires and floods, recording political or economic events for presentation on TV or for video on demand consumption, recording an experiment on a lab, recording the presentation of a product, producing evaluation reports for products of various types, recording damages of a building or on disaster areas, or even systematically recording the components of a complex 3D structure such as a multistore building may be modeled with multimedia events.

Typically, a multimedia event is recorded using only media such as video or pictures and audio; however, events are typically associated with some spatial information that can be described with maps or diagrams and also have some associated metadata which are used for further processing or filtering of the multimedia events.

Systematic event capturing and visualization should associate and exploit all the contextual data with the video captured in order to offer better cognitive clues to the viewer and allow exploitation of the metadata for workflow processing. For example, in an accident event the multiple media may capture the scene of the accident (cars, people involved, etc.); spatial information such as diagrams may record the location on the road of the cars involved and the metadata may be used to record day and time as well as the names of the people involved or interviewed about the accident.

Multimedia event capturing is an expensive endeavor. For example, the TV stations utilize multiple crew members to carry, install and operate the video, audio and lighting equipment that is used for multimedia capturing. This is not economically feasible for small multimedia companies or for companies that only have the multimedia information management integrated within their larger business activities. Ideally, multimedia event capturing should only involve one person. We call this requirement the "one-man-crew" requirement. Certainly, the "one-man-crew" should possess expertise from several different areas of multimedia capturing like audio, video, lights, scene composition, etc. He should also be able to work fast in real time for capturing all the multimedia events and the associated metadata.

Today all the multimedia types are recorded in a digital format. With each event, metadata values of different types will have to be captured and associated with the events. The one-man-crew will also be responsible for capturing and recording the metadata that are associated with the event. In some cases the beginning and the end of the event to be recorded is not known in advance. In this case, the person that operates the capturing equipment will have to demarcate within a longer capturing sequence the beginning and the end of the event, and associate the relevant metadata values with it in order to be able to do fast post-processing (manually or automatically). Thus, the multimedia capturing environment will have to be controlled and synchronized by a mobile computer (laptop or other) which will also provide functionality for demarcating the beginning and the end of the events and associate the appropriate metadata values with each event.

We present here a model and a system for multimedia event capturing by a one-man-crew. The first objective is to facilitate the one-man-crew to do a complete and systematic capturing of multimedia events of different types that compose a bigger event in real time. A second objective is to allow fast browsing and extraction of all the multimedia event information either in real time or at post-processing. A third objective is to allow visualization of the event process in a spatial representation together with the various shots taken related with the event so that the user viewing the multimedia captured gets a better understanding of the event context and capturing. A fourth objective is to allow the one-man-crew to be prepared through learning and reviewing for a good quality multimedia event capturing by means of event type specific learning instructions.

In our model the multimedia events belong to complex event types that may be composed of events belonging to simpler multimedia event types. The metadata captured for each event depend on its type. Events are captured by multimedia shots which belong to shot types. Events and multimedia shots may be associated with maps and/or diagrams. Shots are described by the position (captured by a GPS device or otherwise), direction (compass) and camera movement of the shot placed on a diagram of a scene that describes the event, as well as with other parameters. The shot types are associated with visual descriptions; these can be used by the one-man-crew for learning and preparation for multimedia capturing of events of a particular type. The proposed event model can be mapped to the MPEG-7 [1] and MPEG-21 [18] standards (used, respectively, for multimedia content description and for the interaction with multimedia content).

In the next sections we first present the related research and requirements in the field of multimedia event capturing, and we then describe the proposed multimedia event capturing model and a prototype system implemented for the support of the model functionality .

II. Related Research and Requirements

Semantic modeling of events has emerged recently as a very promising research direction in business applications and business intelligence [16] [1] [2] [3]. Event Processing Architectures are becoming very important in business processing applications [16]. Although event Processing Architectures may seem to contrast to Service Oriented Architectures (SoA), researchers mostly agree for their complementarity, with the Event Processing Architectures covering better the asynchronous requirements of business processing. An Event Driven SoA is a hybrid form combining the intelligence and proactiveness of the Event Driven Architectures with the SoA organizational capabilities [17], [16].

Semantic modeling of events is also an active research area in multimedia. The Semantic Model of MPEG-7 is based on the modeling of events and includes, among others, modeling of actors that are participating in the event as well as the time and the place of the event. The Semantic Model of MPEG-7 is in a high level, but the extensibility mechanisms of MPEG-7 can be used to give powerful domain specific semantic descriptions of events in specific domains while still remaining completely within MPEG-7 [1]. The semantic multimedia event model that we present in this paper can also accommodate domain specific descriptions for the events, which include domain-specific knowledge systematically captured as described in [19]. The Semantic MPEG-7 descriptions can also be transformed in logic based languages such as OWL for further processing and inference [1]. A related recently expanding area of research is the automatic extraction of semantic events from video sequences [5].

To represent the process of multimedia capturing we need to model such aspects as camera parameters, camera location, camera movement, subject movement, light and sound locations, etc. The use of such parameters for capturing quality multimedia is studied in the cinematography area, where rich bibliography exists [7], [8], [9], [10], [13]. These principles can be taken into account when producing guidelines for taking specific shot types. Our system allows the integration of such help facilities.

Some research related to cinematography tries to capture the rules for good cinematography into language or expert system constructs, often with the objective to automate the presentation of virtual reality scenes in games and elsewhere [11], [12]. Whereas our objective is not to automate the cinematography of virtual environments, to the extent that these tools could be used for guidance to “one-man-crew” they could be also useful additions to our system.

While rich literature exists in the area of semantic event modeling and multimedia modeling based on events, as well as in cinematography principles, there is virtually no research in generic tools and methodologies for systematically capturing multimedia events and their context including the spatial context and the context of capturing. Such tools should be based on models associating composite events with the spaces where the events took place and the event evolution in time. In addition, they should facilitate the modeling of the capturing processes in space and time and its associations with the event evolution. This will facilitate the visualization of the events, the shot capturing processes and the actual multimedia captured. Without such tools the capturing process is very expensive and slow, and the visualization possibilities limited, overtaxing the cognitive capabilities of the users.

Our work is further motivated by the fact that today’s multimedia capturing devices can automatically capture a lot of contextual information (for example, using a GPS, compass, azimuth), which can be manipulated to automatically register the 3D natural environment of the picture or video [14], [15]. This eliminates the need for extensive manual editing for the systematic capturing of multimedia events and their context.

The requirements for multimedia event capturing tools that we consider in this paper include:

1. Capturing the spatial event aspects such as integrating maps and diagrams with additional relevant details.
2. Capturing and representing the event actors, their spatial representations, locations and movements during the event.
3. Capturing the various aspects of the real scenes that relate to the event using multimedia (photos, video, and sound).
4. Capturing the event time considering absolute and relative times with respect to other events, as well as demarcating the beginning and the end of shots that relate to a specific event for easy event browsing and extraction.
5. Representing the multimedia capturing process with respect to space and time (camera location, direction, angle, movement, etc.) on top of the spatial representations for facilitating the visualization of the spatial context.
6. Representing event workflows and mechanisms to facilitate the scheduling and preparation for the event capturing and the complete coverage of all the component events.

7. Associating guidelines for what to capture and how to do the capturing (such as abstract scene descriptions, camera location and movement, shot taking, lighting conditions, etc.) with each event type.

The mobile multimedia event capturing model and system described here aim to support such functionalities.

III. The Conceptual Model

The proposed conceptual model comprises of concepts that belong to the following four major categories:

1. Concepts describing the events and the multimedia shots taken to capture those events.
2. Concepts used for the logical spatial event representation.
3. Concepts used for the logical spatial shot representation.
4. Concepts used for the physical representation of the events and shots described on top of the maps or other spatial representations like diagrams or 3D views.

A. Concepts Describing the Events and the Shots Taken

In our model, the events take place in space and time. The multimedia captured may be parts of the event itself, or multimedia related to the event. The event model integrates the event context with the multimedia capturing context and the actual multimedia captured.

We present here (see also Figure 1) the concepts of the model that describe the events and the shots taken using classes and metaclasses. The composition semantics at the metaclass level imply that the whole structure is copied at instantiation. For example, when we instantiate a composite event called "marriage" we also instantiate the events of "ceremony" and "reception" which are associated with the "marriage", as well as the roles of "bride", "groom", etc. that are the roles associated with the "marriage". At the class level, specific actors (names of individuals, etc.) are associated with the role. Roles other than single human roles can be also represented. For example the "object" role may be used to represent a car that participated in an accident, while the "crowd" role may be used to represent the approximate location of people participating in the event.

The events may be associated with structured *Spatial Objects*, the event *Time*, and *Event Shots* that represent the multimedia capturing of the event. The time associated with an event is referring to a specific time interval and it takes the relative sense of time. For example an instance of time in the marriage example could be the "Wedding day".

B. Concepts for the Logical Spatial Representation of Events

Events are associated with physical spaces that may be described by spatial objects. These objects may be represented in the form of diagrams, etc. on a canvas with a semantic meaning provided by the spatial object itself or without any semantic meaning indication at all.

The events are associated with Roles that semantically describe the event participants. The Roles are concepts that may describe humans, objects, crowd, etc. Thus, they can be seen as an Event Type ontology. A concept may have more than one visual representations to represent different states or canvas resolutions. For example, a car may have the states crashed or non-crashed. A person may be sitting or standing. The crowd may have different shapes at different times. A room may be represented simply with a square if the resolution is low or with a detailed diagram if the resolution is high.

It is important to note that roles and objects participating in an event may have specific representative graphical representations that may be reusable across events of the same type. For example, the capturing of an accident event involving two cars should record the location and movement of the cars on a map or diagram of the road. The model allows reusable car representations and motion vectors to be used for that purpose.

The spatial representation of a simple event is described by the original position, size, and orientation of the semantic concepts of the ontology on the canvas and the time that they first appear. A spatial representation of a concept is associated with the time it first appears within the event representation, and with the time that this visualization of the concept finishes as

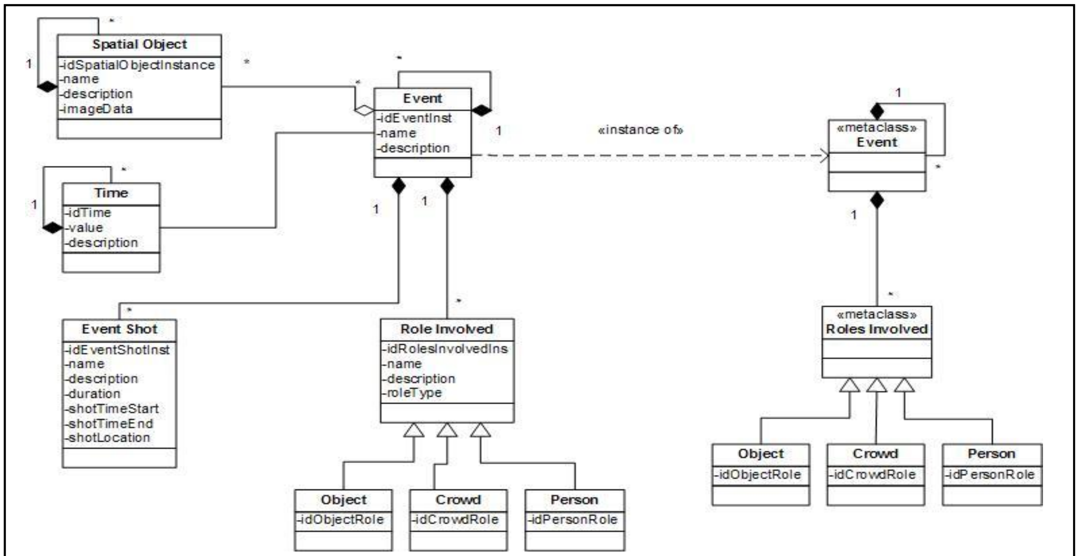

Figure 2: Concepts describing the events and the shots taken. The instantiations of an event would mean the instantiation of all the concepts associated with it.

well as with the position, size and orientation of the object at the time that it disappears. A trajectory may be associated with each object to describe the line of movement on the canvas that describes the event. Interpolation determines the position and orientation of the object in intermediate times. Optionally, an object may be declared to be deformable, in which case the visual representation of the object at the time that it disappears has to be specified and the semantics imply that the shape of the object in intermediate positions can be found via interpolation between the original and the final appearance.

For example, in a car accident scenario, we may have a spatial object representation describing the accident place, which could be a map of the street. The roles will be the cars involved in the accident and an event could describe the cars after the accident. The logical representations associated with this event will be the cars involved in the accident. The representation of a car may be a photo of the car crashed or a square placed on the car location. The original position and time of the car appearance are also part of its representation.

C. Concepts for the Logical Spatial Representation of Shots

A Shot in the media industry always refers to a simple event. It may capture the whole event or it may (typically) capture a part of the event. In some cases the event has been executed and the capturing refers to an event that has taken place. For example, assume that in an accident report the event reported has happened before, and the shooting refers to the event, but it does not capture the event itself.

In the case of shots, in addition to the actors, it represents the camera, lighting and sound parameters for the shot and their changes. The camera capturing is modeled with the location of the camera, the direction of the capturing and the viewing angle, which allows to decide which objects are viewed by the camera in a specific setting. Similar modeling is used for light and sound sources.

Event Shots are used to represent types of Shots within an event type. For example an Event shot may be a shot of the "bride arrival" event using a zoom in as she enters the church. Event shots have shot spatial representations which depict the event shot on the canvas.

A shot spatial representation describes the original and final locations and the camera parameters, the sound and light equipment on top of the canvas, in the same way that the event concepts were represented on top of it. In this case the original and final representations of the capturing parameters are described. For example, in a camera pan the original and the final position of the direction of the camera are described. This can be used to visualize the original and final footprints of the pictures taken, but also to infer the intermediate positions.

For example, in a "debate" event (which may involve shots from more than one locations), the camera location, the viewing angle, and the shot direction in relation to the two people debating may be recorded for producing a training example of capturing principles for a debate.

Note also that the representation shape may change during the capturing process. For example, the shape of the crowd involved may change. The viewing direction and the viewing angle of the camera may also change resulting from panning and zooming operations on the camera by the "one-man-crew'. Interpolation is used to infer the intermediate positions of capturing

An important application of this part of the model is that it allows the representation of the spatial evolution of an event, as well as the spatial evolution of the capturing of the event. For example, it may be used for describing the history (event) of a visit in a museum or a zoo, and the capturing of videos that the "one-man-crew" visitor took during the visit. The multimedia shots themselves are associated with the descriptions of the parameters that allow their representation on the canvas.

The parameters foreseen by the model facilitate the description of the event and the shot taking in an abstract spatial space. For example, the relative positions may reflect the different GPS locations or the relative distance in meters reflected on a drawing canvas. Location in buildings, such as museums, may also be captured with respect to the location of objects that for example have been recorded in the building diagrams. This could be even automatic, for example by using RFID identifiers for the objects, exploiting their location as they have been recorded in electronic diagrams. Notice also that the direction may be determined with respect to the geographic coordinates or by recording the position of the camera and the position of another object in the camera direction (e.g. by recording its GPS position or its RFID).

D. Concepts for the Physical Event and Shot Representation

The event and shot logical representations at the spatial level that were described in the previous section may have to be viewed on top of a map. It is also possible to utilize 3D maps in order to represent on them the camera location, movement and angle of view [14], [15].

The concepts of this part of the model are responsible for doing the mapping from the logical representation to other representations as required (one or more). To do this rotation translation and scaling of the logical representations may be required. The logical representations mapped on the site map are called physical representations. The physical representations should have a location on the site map as well as shape and possible movement.

A shot site map represents a drawing canvas with x and y dimensions where the various physical representations will be located for a specific event or shot. It is used to describe the geography of the scene where a specific event or an event shot took place. The most common way to lay the physical representations on the canvas is to use a drawing (e.g. the map) of the spatial object associated with the event or shot, as a background and then to put on the canvas the physical representation of the roles involved.

A specialization of the shot site map is a geographic map. This kind of map is mostly used for exterior locations, where a specific map is chosen for usage and the detailed description of the location of the capturing is not needed. For example, in "the crowning" event shot the logical representations of the priest, the best man, the bride and the groom would be mapped to physical representations and put on top of the site map. The spatial object associated with the specific event shot would be the interior of the church. The logical representation of the church should be mapped to a physical representation and this physical representation should be used as a background to the site map. The relative position and movement of the camera capturing the shot will also be represented on the shot site map. These representations present the relative positions and orientations of the various roles, and thus they reveal the geometry of the scene.

IV. MULTIMEDIA EVENT AND SHOT VISUALIZATION

An important objective of this work is to provide the tools for fast event visualization and multimedia shots that capture the events themselves. Since the events have structure, roles, spatial and temporal aspects as well as descriptions, all those aspects should be supported in the event visualization.

Every event, complex or simple has a spatial representation on a canvas (that may include drawings or pictures). For simple events, the spatial representation is typically a representation (a map or a reasonably accurate diagram) of the inside or outside space where the event takes place, including the location and movement of the participating roles in this representation. For composite events the spatial representation of the event may be on a more abstract representation level (that may not reflect the real distances of the locations). For example in a composite marriage event that takes place in two locations (home, church) the actual distance of the two spatial locations may not be of interest. The spatial representation may only reflect the fact that there are two locations. By selecting one of them, the user can see the detailed spatial arrangements in the church ceremony for example.

Note that in case an event is associated with a spatial object which is composed of other spatial objects, the presentation system may choose to show the spatial representation of both the higher level spatial object as well as its component objects (or their representations) in a single diagram. The selection of a component object leads the visualization system to the corresponding event (if it exists). For example, when visiting a floor, the rooms may only be symbolically represented (the visits to the rooms may not be shown). Selecting the representation of the room presents the spatial representation of the room as well as the representation of the visit in it. Note also that it is possible when the user looks at the event of visiting a room to present not only the floor plan of the room and the visit in it, but also the spatial representations of the higher level spatial objects such as the floor plan and the building composition in floors.

A spatial object may be an individual which relates to a semantic concept in an ontology [14], [15]. Through this relationship, the user can for example see semantic information about the visited space (for example, when visiting a room, the user may see who resides in this room or the room functions).

The visualization of the composite events also includes the visualization of their temporal aspects. This involves the visualization of the sequence of the component events and the automatic browsing of the component events. This is achieved by connecting the spatial representations of the component events with directed lines indicating the time order of the components events. The lines may not be straight lines. This is useful in order to avoid line cluttering, or to indicate a route that is followed to visit the place where the next component event will take place (superimposed on a background map or diagram). For example, in the case of the wedding event the directed line connecting the church with the home where the reception takes place indicates that the church ceremony precedes the home reception. If this line is projected on top of a city diagram it may also indicate the route from the church to the home. The semantics of the time line only support the event ordering and any additional semantics are due to the visual interpretation in connection to the background objects.

We note that a place may be associated with more than one events that are not necessarily sequentially ordered. The same spatial object is used for the spatial representation of the component event. The directed lines representing the sequence of execution of the component events will utilize this spatial representation more than once.

A special role (applicable to every event) is an avatar representing the narrator of the event. An animation of a composite event can be performed automatically, following the temporal ordering of the events. When the temporal ordering is a sequence, this corresponds to an automatic navigation of the event visualizations in a depth first manner. For each composite event the spatial representation of the event is shown and the narration that is associated with the immediate component events and the transitions between the component events is performed automatically.

The depth first event traversal in time order is the default automatic animation of the event presentation. Alternative automatic methods for visualization such as breadth first presentation are also possible and they may be more appropriate for certain events, like the presentation of spatial descriptions of spaces or like events that are partially ordered temporarily. For example, in a conference event it may be desirable to have a breadth first presentation of the component events at a certain level (events that happen after a certain time).

Direct access to any of the events in the event hierarchy is possible. For example, a user may directly visit the third floor of the building by selecting the appropriate rectangle from the visualization, thus bypassing the presentation of the first two floors. Then he may select and visit a specific room and listen to the narration associated to it instead of going through the narration of several rooms before. When the user selects a specific event through its representation, the system by default performs animation for all the events that belong to the subtree of the event hierarchy.

Shots are only associated with simple events. The shot visualizations are described in the same canvas with the simple events. The automatic animation of shots within a simple event is done in time order (the shots are associated with the time of capture). Not every event may have a shot associated with it. The existence of shots within an event is represented with the spatiotemporal visualization of events in order to facilitate the user navigation to events that have associated shots.

A special case of mobile multimedia event capturing is the case of systematically capturing 3D spaces, such as a building, using, for example, pictures or video. The capturing process can be very fast and accurate if diagrams of the floor footprints are available. In this case such a footprint is a multimedia object that can be associated together with its coordinates to the event canvas. Each multimedia shot is associated to a simple event, for example the capturing of a room. Alternatively many shots may be associated to a simple event (like the capturing of a whole floor).The location and direction of the camera can be either automatically recorded using its position and direction sensors or manually recorded on the map. The visualization of the result can show at the same time the spatiotemporal aspects of the capturing (such as the camera location and movement), and besides it the actual result of the capturing process (the video of the shot). This interface allows the user to understand better what the video is about. Again, higher level spatial objects (like the floor plan) can be also shown with an indication of the place where the shot was taken.

A very important class of applications of multimedia event capturing is video taking during museum visits, city tours, outside zoo visits, and archeological site visits. During these visits the users will capture objects that are not necessarily known in advance. It is however possible to capture the history of the visit through an event and associate a spatial object or map with it (possibly in an automatic manner like using GPS and compass for outside spaces). In this spatial object the multimedia capturing can be recorded as described before. An important aspect here is that if the spatial object has semantic information attached to it, then it is possible to automatically structure the visit event into more detailed events of capturing specific semantic objects which are provided by the spatial map. In this case automatic direct access to the semantic events captured is supported without much user effort.

V. EVENT CAPTURING AND VISUALIZATION

The system allows the generation of domain ontologies for capturing the various event types. With each concept, graphic avatars representing the concept can be generated and reused in many concept instances. Avatars can represent persons (e.g. a bride), objects (e.g. a car), crowd (e.g. friends), etc. More than one avatar can be designed to represent different concept states or concept representations in different resolutions. A direction is associated with the object representation. The direction is useful for example in order to orient the graphic representation towards the direction of the object motion.

The ways that multimedia capturing equipment is utilized can be also related to the graphic representations with directions associated with them. For example, a camera taking pictures towards a specific direction is related to a polygon that represents that direction. The graphical user interface also allows the generation of polygons representing locations where events take place, as well as lines representing the event execution sequence for the higher level events.

To facilitate event demarcation and collection of the event metadata in real time as well as to support fast capturing post processing, the system can trace the event execution order and automatically request the metadata values needed for the event. For events that do not occur in a particular sequence, the system provides facilities for fast browsing and search.

VI. CONCLUSIONS – FUTURE WORK

Event processing is very important to business. We presented here a model and a system that automate to a large extent the capturing of multimedia related to events together with the context of the events and the context of capturing. This allows powerful visualizations of the spatial and temporal evolution of the events and the capturing processes. The prototype system built supports this functionality. The system allows accommodating maps and diagrams of various types. We are currently in the process of integrating Google maps and 3D maps with our previous automatic registration algorithms [14], [15] and more extensive contextual information captured from sensors. We are also investigating the capturing and visualization of multiple concurrent events. Finally, we are now working in the definition of a detailed formal mapping of our model to the constructs of the MPEG-7 and MPEG-21 standards.

REFERENCES

[1] C. Tsinaraki, S. Christodoulakis: An MPEG-7 query language and a user preference model that allow semantic retrieval and filtering of multimedia content ACM /Verlag Multimedia Systems 13, 2: 131-153 (2007).
[2] V. Kantere,I. Kiringa,J. Mylopoulos,. Supporting Distributed Event-Condition-Action Rules in a Multidatabase Environment. s.l: World scientific, 2007. Vol. vol. 16, no. 3, pages 467-506.
[3] J. Mylopoulos, What Do Events Mean? Trento (Italy) : s.n., 2009.
[4] C Colombo, A Del Bimbo, P. Pala. Semantics in Visual Information Retrieval. IEEE MultiMedia. s.l. : IEEE Computer Society Press, July-Sept. 1999. Vol. vol. 6, no. 3, pp. 38-53. doi:10.1109/93.790610.
[5] M. Bertini, A. Del Bimbo, C. Torniai, C. Grana, R. Cuchiara: Dynamic Pictorial Ontologies for Video Digital Libraries Annotation, Workshop on Multimedia Information Retrieval on the many faces of multimedia semantics, ACM, 2007, pp47-56.
[6] D. Arijon: Grammar of the Film Language, Silman james Press edition, 1976
[7] M. Schreibman: Creative Producing from A to Z, Lone eagle PC 2001.
[8] B. Braverman: Video Shooter, storytelling with DV, HD and HDV cameras, CMP books, 2005.
[9] J. Mascelli: The Five C's of Cinematography, Silman James Press, 1965.
[10] A. Rosenthal: Writing, Directing, and Producing Documentary Films and Videos, Southern Illinois University Press, 2007.
[11] S. Drucker, T. Galyean, D. Zeltzer: CINEMA: A System for Procedural Camera Movements, Proceedings of the 1992 Symposium on Interactive 3D Graphics, ACM, 1992.
[12] M. Gleitcher, R. Heck. M. Wallick: A Framework for Virtual Videography, Proceedings of International Symposium of Smart Graphics, 2002, ACM.
[13] S. katz: Film Directing Shot by Shot, Sheridan Books, 1991.
[14] S. Christodoulakis, M. Foukarakis, L. Ragia, H. Uchiyama, T. Imai: Picture Context Capturing for Mobile Databases, IEEE Multimedia, June 2010.
[15] S. Christodoulakis, M. Foukarakis, L. Ragia, H. i Uchiyama, T. Imai: Semantic maps and mobile context capturing for picture content visualization and management of picture databases, Proceedings ACM MUM 2008.
[16] D. Luckham: The Power of Events: An Introduction to Complex Event Processing in Distributed Enterprise Systems, Addison Wesley, 2002.
[17] A. Michlmayr, F. Rosenberg, P. Leitner, S. Dustdar: Advanced Event processing and Notifications in Service Runtime Environments, Proceedings ACM DEBS 2008.
[18] F. Pereira, The MPEG-21 standard: *Why an open multimedia framework?*, in Proceedings of the 8th International Workshop on Interactive Distributed Multimedia Systems (IDMS 2001), LNCS 2158, Lancaster, September 2001, Springer-Verlag, Heidelberg, pp. 219–220.
[19] Tsinaraki C., Christodoulakis S.: "Domain Knowledge Representation in Semantic MPEG-7 Descriptions". In "The Handbook of MPEG applications: Standards in Practice", Wiley Publishers 2010.