Metadata for object-oriented television

Olena Osharovska
Dept. Radio and Television Broadcasting
O.S. Popov Odessa National Academy of Telecommunications
osharovskaya@gmail.com

***Abstract*—In broadcast multimedia metadata covers both the description of services and multimedia control. Each media object has a set of descriptors associated with it. Object descriptor identifies all streams associated with the media object. TV mesh objects are combining several structures, in particular: polygonal mesh cage; a set of matching texture; a set of normal vectors to faces; the color histogram about the content. The model of the content-oriented metadata being generated must be smaller than the original data of mesh object. In the article, it is proposed to supplement Language's Web Ontology Dictionary with terms on multimedia object-oriented broadcasting.**

***Keywords*—TV, mesh-object, metadata, broadcasting service, semantic relation, ontology**

I. Introduction

For almost a century, the broadcasting television was based on the concept of simultaneous transmission of socially significant information to all viewers. With the introduction of digital broadcasting, there is a rapid transition to new types of consumption of personalized video content at a convenient time for the subscriber. In addition, television becomes interactive with the possibility of active participation of the viewer in the formation of programs. The active position of viewers demanded not only to change the system of delivery and storage of media information, but also a new approach to the process of producing video content. One of the innovative approaches to creating programs is object-oriented creation of scenes, considered in the Recommendations MPEG-4, MPEG-7 and MPEG-21 [1, 2, 3]. The main purpose of the MPEG-4 standard in working documents of the MPEG group is formulated as follows: it sets the principles of working with content (digital representation of media data) for three areas: actually, interactive multimedia (including products distributed on optical disks and via the Network), graphics applications (synthetic content) and digital television – DTV. MPEG-4 is not only a standard, in fact it sets the rules for the organization of the environment, and the environment is object-oriented. It deals not just with flows and arrays of media data, but also with media objects this is the key concept of the standard. Objects can be audio, video, audiovisual, graphic (flat and three-dimensional), text. They can be either "natural" (recorded, filmed, scanned, etc.), or synthetic (i.e., artificially generated). Examples of objects can serve as a fixed background, video characters separate from the background, text-synthesized speech, musical fragments, and a three-dimensional model that can be moved and rotated in the frame. Media objects can be streamed by any video-pass. Each media object has a set of descriptors associated with it, where all its properties are defined, the operations necessary to decode the associated streaming data, placement in the scene, as well as the behavior and permissible reactions to the user's actions. There are advanced tools for working with two-dimensional polygonal models, adapting them to existing video content for later animation. Using these tools allows you to perform many functions, for example, representation of the contours of objects using grid vertices (instead of bit ones masks), replacement in the scene of "live" video objects by synthetic ones, etc. Material of this work is a continuation of previous studies in this area [4].

II. Mesh object metadata

The use of metadata or data about data, provides answers to how to handgrip the diverse services and contents of the new digital TV platform efficiently and a consumer-friendly way. However, the use of metadata is broadcast multimedia should not be seen as just a tool to cope with the challenges inherent in a complex networked multimedia environment. Instead, it inputs up new possibilities for the development of new innovative services.

In broadcast multimedia metadata covers both the description of services and multimedia control [5]. Metadata integrates into the broadcasting value-chain with considerations for each step in the development of a digital TV broadcasting service. Any different metadata, standards related to the digital broadcasting life-cycle. The concept bears the idea of companying on insight into syntonic and semantically complex data by refining their essence into a set of simple descriptors. Metadata also helps to arrangement and accomplish information in varied setting. The unit of interchange is a structured digital object (digital item (DI) described by metadata and referring related multimedia content assets. Media objects may need a data stream that is converted into one or more elementary streams. This allows you to hierarchically process the encoded data, as well as the associated media information about the content (called "object content information").

Each stream is characterized by a set of descriptors for configuring information, for example, to determine the necessary resources of the recording device and the accuracy of the encoded time information. Moreover, the descriptors may contain hints regarding the quality of service (QoS) that is required for transmission (e.g., the maximum number of bits/s, BER, priority, etc.). Binary format for Scenes (BIFS) describes the space-time relationships of objects on the stage. Viewers can have the ability to interact with objects, for example, by moving them on the stage or by changing their position of the observation point in a 3D virtual environment. The scene description provides a wide range of nodes for compositional 2-D and 3-D operators and graphic primitives [6, 7].

TV mesh objects are combining several structures, in particular: polygonal mesh cage; a set of matching texture; a set of normals to landfills; the color information for each polygon.

Mesh data has spatial and temporal aspects, as well as arbitrarily high dimensionality, which aggravates the task of finding compact, accurate, and easy-to-compute data models. This approach leads to interesting challenges. The model (equivalently, the content-oriented metadata) being generated must be smaller than the original data by at least an order of magnitude. Second, the metadata representation must contain enough information to support a broad class of queries. Finally, the accuracy and speed of the queries must be within the tolerances required by users.

We approach this problem by building compact, approximate, multi-resolution models of the mesh data and then using the models to support high-fidelity ad hoc queries [8].

Statistical and mathematical data analysis and compression methods such as clustering [9, 10], spline-based fitting [11, 12] and wavelets [13, 14], are all suitable in the preprocessing phase.

Let us define mesh data as a discrete representation of continuous data, which can be defined as an ordered set of tuples as follows. (Mesh data represented in this manner is called a "point mesh" [8], which is just a collection of data points with no topological connection among them.)

$\{t, x_1, x_2, ..., x_n, v_1, v_2, ..., v_m\}$ where t denotes a temporal variable defining a time step, x_i, $i \in [1, ...n]$, denotes a spatial variable defining the geometrical coordinates in an n-dimensional space, and v_j, $j \in [1, ...m]$, denotes a field variable defined at each node (positioned at $(t, x_1, x_2, ..., x_n)$ >) or each zone in the mesh at time step t. A zone in a Regular mesh is an n-dimensional cubic bounded by the surrounding 2n mesh nodes, whereas an indeterminate number of nodes surrounds a zone in an irregular mesh. One organize the metadata in two levels. The lower level contains two elements. The first element is a multi-resolution model of mesh data, modeled per partition of mesh data, as well as a collection of summary information that is generated as the result of preprocessing. For example, the summary information includes, per variable, information on min, max, mean, median, standard deviation, and the first several moments for each variable. The actual models vary from cluster prototypes, to regression equations, to matrixes of equations, and so on. The second element of metadata is a set of indexes that define the structure that must be traversed to reach the first element data. More specifically, each node of an index contains the summary information and model information of the corresponding mesh partition. There are many challenges to overcome for this approach to work. Resolving them involves investigating and making complex trade-offs between preprocessing time, compression level, query speed, query accuracy, and the range of queries the metadata supports.

Scalable solutions: all aspects of the preprocessing phase must be scalable to terabyte mesh data, and the algorithms must be amenable to efficient parallelization on machines.

Complete models: little is known about what makes an effective model of mesh data hat is capable of supporting arbitrarily complicated ad hoc queries. There are many alternatives, which have varying degrees of accuracy and descriptive power. The models must be compact, yet contain enough information to answer a wide range of possible queries.

Compact, efficient, multi-resolution metadata: we allow the scientists to trade query response time for accuracy, allowing interactive ad hoc queries at one extreme, and slower, and more accurate responses at the other.

It can be assumed that the goal of preprocessing the mesh data is to generate a representation of model that is much smaller and yet can be used to support approximate ad hoc queries. Firstly, the entire mesh is decomposed into an appropriate partition, then, the mesh data within each partition is concisely approximated using an appropriate parametric model (e.g., splines, wavelets, or clusters). Algorithms to generate this metadata typically iterate through two phases: mesh partitioning and partition characterization. Iteration between these phases takes place to revise the actual partitioning based on a similarity (or error) metric that measures the difference between the characterization of the data in a partition and the actual mesh data itself.

III. The Query by Image Content

Content-based image retrieval (CBIR), also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR) is the application of computer vision techniques to the image retrieval problem, that is, the problem of searching for digital images in large databases. Content-based image retrieval is opposed to traditional concept-based approaches.

Researchers [15, 16] mentioned three levels of queries in CBIR. Level 1: Retrieval by primitive features such as color, texture, shape, or the spatial location of image elements. Level 2: Retrieval of objects of given type identified by derived features, with some degree of logical inference. Level 3: Retrieval by abstract attributes, involving a big amount of high-level reasoning about the aim of the objects or scenes depicted. A CBIR system should provide full support in bridging the semantic gap between numerical image features and the richness of human semantics [16] in order to support query by high-level concepts.

Current systems mostly perform retrieval at Level 2. There are three fundamental components in these systems: low-level image feature extraction; similarity measure; semantic gap reduction.

Low-level image feature extraction is the basis of CBIR systems. To perform CBIR, image features can be either extracted from the entire image or from regions (region-based image retrieval, RBIR).

To perform RBIR, the first step is to implement image segmentation. Then, low-level features such as color, texture, shape, or spatial location can be extracted from the segmented regions. Similarity between two images is defined based on region features.

IV. Ontology for multimedia applications

Humans tend to use high-level features (concepts) to interpret images and measure their similarity. In general, there is no direct link between the high-level concepts and the low-level features [16]. Though many complex algorithms have been designed to describe color, shape, and texture features, these algorithms cannot adequately model image semantics and have a lot of limitations while dealing with broad content image databases [8] have not been described in the literature so far. An essential part of this methodology is to develop in OWL (Web Ontology Language) an ontology that fully captures the semantic metadata model defined in the Semantic Part of the MPEG-7 MDS [3]. The need for an ontology that fully captures the MPEG-7 metadata model has been pointed out by several research groups [7, 16, 17, 19]. Some important work in this direction has been carried out in [18] for the case of Resource Description Framework (RDF) [20], but it has some limitations: Classes corresponding to the MPEG-7 complex types have been defined, but not all of the (simple and complex) attributes of the classes are represented. In addition, typed relationships among the metadata items are not represented, although MPEG-7 provides complete support for typed relationships. Furthermore, this work has been based on RDF [21, 22].

Our approach for the integration of OWL ontologies in the framework for the support of ontology-based semantic indexing and retrieval of audiovisual content, utilizes an ontology that captures the model provided by the Semantic Part of the MPEG-7 MDS for the representation of semantic metadata for audiovisual content description. This ontology captures the semantics of the first layer of the two-layered model for semantic metadata used in the DS-MIRF framework. The second layer of the model encapsulates domain-specific knowledge, which extends the audiovisual content description standards so that they integrate transparently domain-specific ontologies. The model for the semantic description of audiovisual content provided in the Semantic part of the MPEG-7 MDS is comprised of complex data types defined, using the XML Schema Language syntax [2], in a set of Description Schemes (DSs) rooted in the Semantic Base DS.

Semantic Base DS: Semantic Base Type is the base type extended by other description schemes according to the needs for the description of semantic entities of specific types. Semantic Base Type has a set of simple attributes (ID for instance identification, time Base, time Unit, media Time Base and media Time Unit for timing support) and the following complex attributes: Abstraction Level, which represents the abstraction existing in the current semantic entity. Label, corresponding to a term that describes in brief the semantic entity. Definition, which is a textual annotation that describes the semantic entity. Property, which is a term that associates a property with the semantic entity. Relation, which relates the semantic entity with other semantic entities. Media Occurrence, which relates the semantic entity to specific media items (e.g. video segments, images etc.).

Semantic Bag DS and Semantic DS: Description schemes used for the description of collections of semantic entities. Semantic Bag Type is an abstract type, defined in the Semantic Bag DS, which extends Semantic Base Type.

Object DS: The Object Type defined here extends Semantic Base Type and is used for the description of objects and object abstractions (e.g. a table).

Agent Object DS: The actors that appear in an audiovisual segment are related with the instances of the Agent Object Type that extends the Object Type. Actors in general are represented using the Agent Type, an abstract type extending Semantic Base Type defined in the Agent DS. Person Type, Organization Type and Person Group Type extend Agent Type, are defined respectively in the Person DS, the Organization DS and the Person Group DS and are used for the representation of persons (e.g. football players), organizations and groups of persons. Event DS: The Event Type defined here extends Semantic Base Type and is used for the description of events (e.g. a goal).

Semantic State DS: The Semantic State Type defined here extends Semantic Base Type and is used for the description of states described in an audiovisual segment and the parametric description of its features (e.g. the score in a soccer game before and after a goal).

Semantic Place DS: The Semantic Place Type defined here extends Semantic Base Type and is used for the description of places (e.g. Athens). Semantic Time DS: The Semantic Time Type defined here extends Semantic Base Type and is used for the description of semantic time (e.g. Christmas). Concept DS: The Concept Type defined here extends Semantic Base Type and is used for the description of concepts present in an audiovisual segment (e.g. cooperation).

Semantic Relation DS: The Semantic Relation Type defined here extends Semantic Base Type and is used for the description of relationships among semantic entities. The relationships may be typed, as described in the Semantic Relation DS. In addition to the attributes inherited from Semantic Base Type, Semantic Relation Type has the following attributes: Source, which is the ID of the semantic entity of the relationship. Target, which is the ID of the semantic entity that is the target of the relationship. Argument, which may be used as an alternate to source and target definition. Strength, which denotes the strength of the relationship. Name, which denotes the name of the relationship. Arity, which denotes the arity of the relationship. Properties, where the properties of the relationship are denoted. MPEG-7 Simple Datatype Representation: The simple datatypes needed are integrated in the core ontology, as OWL permits the integration in OWL ontologies of simple datatypes defined in the XML Schema Language [2]. MPEG-7 Complex Type Representation: MPEG-7 complex types correspond to OWL classes, which define groups of individuals that belong together because they share some properties. Thus, for every complex type defined in the MPEG-

7 MDS a respective OWL class is defined. Simple Attribute Representation: The simple attributes of the complex type of the MPEG-7 MDS are represented as OWL datatype properties, which relate class instances to datatype instances. Complex Attribute Representation: For the representation of the complex attributes of the complex type, OWL object properties are used, which relate class instances. An OWL class for the representation of the complex attribute instances is defined, if it does not already exist.

V. Conclusions

Metadata in the form of text and visual information constitute the most important part of digital information flows between different TV systems. The advent of information systems tools into television has led to the need for media data modeling. Within the framework of data modeling, the concepts of entities, their attributes and relationships between modeling entities are used. In TV production systems, incoming metadata allows you to find and select fragments of source materials to create new programs for use. The programs completed in production should be appropriately described in the form of an organized set of metadata. Metadata on the created programs includes a part of the input metadata and some new metadata reflecting the essence of the created programs. Thus, it can be stated that in the production systems of TV programs there are operations for the formation and conversion of metadata. Metadata in the form of text and visual information constitute the most important part of digital information flows between different TV systems. The advent of information systems tools into television has led to the need for media data modeling. Within the framework of data modeling, the concepts of entities, their attributes and relationships between modeling entities are used.

References

[1] ITU-T Recommendation H.264 *ITU. 2003-05-30. Retrieved 2013-04-18.*

[2] ISO/IEC 15938-5 *FDIS Information Technology - Multimedia Content Description Interface - Part 5: Multimedia Description Schemes* . [*July 2001, Sydney*]

[3] Jane Hunter *Reconciling MPEG-7 and MPEG-21 Semantics through a Common Event-Aware Metadata Model* . [*https://arxiv.org/html/cs.cs/0210021*]

[4] Gofaizen Oleg, Osharovska Olena, Patlayenko, Volodymyr Pyliavskyi *Associative Relations in the Assessment of TV Images* . [*Open Semantic Technologies for Intelligent Systems*]. 16-18 february 2017, report 52, 4 p.

[5] Cox M., Mulder E. and Tadic L *Descriptive metadata television: an end to end introduction* . [*Focal Pr*]. 2006 - 157 P.

[6] ISO/IEC TR 21000-1:2001(E) *Part 1: Vision, Technologies and Strategy* . [*MPEG, Document: ISO/IEC JTC1/SC29/WG11 N3939*]. <http://www.cselt.it/mpeg/public/mpeg-21_pdtr.zip>

[7] ITU-R Recommendation BT-1865 *Metadata for monitoring errors in standard definition television signals and high definition television in the broadcasting chain* . []. 2010, 46 P.

[8] Lee S., Snapp R., Musick R., Critchlow T. *Ad hoc Query Support for Very Large Simulation Mesh Data: the Metadata Approach* . [ ]. 14 P.

[9] Hartigan John A. *Clustering Algorithms* . [*Johns Vviley & Sons*]. New York 1975

[10] Zupan J. *Clustering of Large Data Sets* . [*Johns Vviley & Sons*]. 1982

[11] C. de Boor *A Practical Guide to Splines* . [*Springer Verlag*]. 1978

[12] Dierckx P. *Curve and Surface Fitting With Splines Monographs on Numerical Analysis* . [*Oxford University Press*]. 1993.

[13] Antoniadis A. Oppenheim G. *Wavelets and Statistics* . [*Springer*]. Verlag 1998

[14] Härdle W., et all *Wavelets, Approximation and Statistical Applications* . [*Springer Verlag*]. 1998.

[15] Rui Yong, Huang Thomas S., Chang Shih-Fu *Image Retrieval: Current Techniques, Promising Directions, and Open Issues* . [*Journal of Visual Communication and Image Representation*]. 10: P. 39–62. doi:10.1006/jvci.1999.0413

[16] Nuno Vasconcelos *From Pixels to Semantic Spaces: Advances in Content-Based Image Retrieval* . [*Computer*]. 2007, pp.20-26. http://www.svcl.ucsd.edu/publications/journal/2007/computer /computer07.pdf

[17] EBU *Operating Eurovision and Euroradio* . [*TECH 3351 EBU Class Conceptual Data Model*]. Geneva 2016

[18] EBU *Operating Eurovision and Euroradio* . [*TECH 3293 EBU Core Metadata Set*]. Geneva 2016

[19] Seung-Jun Yang, Jung Won Kang, Dong- San Jun, Min Je Kim, Han-Kyu Lee *TV-Anytime Metadata Authoring Tool for Personalized Broadcasting Services* . [*Consumer Electronics, ISCE 2007. IEEE International Symposium*]. 2007. 6 p. DOI: 10.1109/ISCE.2007.4382191

[20] B. Makni et al *Semantic integration of TV data and services. A survey on challenges and approaches* . [*Web Intelligence and Agent Systems An International Journal 1*]. 2010 pp. 1-24.

[21] Tsinaraki C., Polydoros P., Christodoulakis S. *Integration of OWL Ontologies in MPEG-7 and TV-Anytime compliant Semantic Indexing* . [*International Conference on Advanced Information Systems Engineering CAiSE: Advanced Information Systems Engineering*]. 2004 pp 398-413

[22] Migual-Angel Sicilia *Handbook of Metadata, Semantics and Ontologies* . [*World Scientific Publish*]. 2014. 523 p

МЕТАДАННЫЕ ДЛЯ ОБЪЕКТНО-ОРИЕНТИРОВАННОГО ТЕЛЕВИДЕНИЯ

Ошаровская Е.В.

Одесская национальная академия связи им. А.С.Попова

В вещательных мультимедийных системах метаданные рассматриваются как описание дополнительных сервисов, так и управление самим контентом мультимедиа. Каждый медиа-объект имеет набор связанных с ним дескрипторов. Телевизионная сцена компонуется из нескольких объектов. Видео объекты целесообразно представлять триангуляционными сетками.

Модель что эквивалентно описанию метаданными, ориенти-рованными на контент, должна создавать меньший битовый поток, чем исходные данные самого сеточного объекта, по крайней мере, на порядок. Представление метаданных должно содержать достаточно информации для поддержки широкого класса запросов. Наконец, точность и скорость запросов должны быть в пределах допусков, требуемых пользователями. Метаданные органи-зованы в два уровня. Нижний уровень - это модель сетки при разном разрешении. Верхний уровень несет информацию о семантике сеточного объекта. Единица обмена представляет собой структурированный цифровой объект (цифровой элемент (DI)).

Метаданные в виде текстовой и визуальной информации составляют важнейшую часть цифровых информационных потоков между различными телевизионными системами. В этой статье предлагается использовать словарь веб-онтологии с терминами мультимедийного объектно-ориентированного вещания.