

# COMP3013 CONFERENCE COMPUTING - RETRIEVAL IN AVALANCHE OF MULTIMEDIA

Daniel Grzywczak  
dg4g09@soton.ac.uk

---

## ABSTRACT

*Nowadays, it is increasingly difficult to find something that we are looking for in web. Every day an innumerable amount of data enters our network and litters it. To avoid noise during retrieval of desired resources, we have to look for a new way of describing files and surfing. In this paper I present some new directions in this field of retrieval.*

### Keywords

*MPEG-7, MPEG-21, retrieval, multimedia.*

## 1. INTRODUCTION

Telecommunication and Information Technology is a rapidly developing field of technology. According to data from [14] and [15] on June 2009 we had 1,668,870,408 internet users (it is about 25 % of world population) and about 240 million web pages. Fast growth of internet users, popularity of mobile technology (the number of mobile phones is getting close the to number of people in whole world, now we have more than 4 billion mobiles phones[18]), development of multimedia compression (Jpeg, Mpeg 4, Divx, Mp3), increasing popularity of peer to peer technology and file hosting, decreasing prices of hard disks and easier ways of creating and processing our movies, pictures or websites led to accumulation of information and multimedia files called information flood. On youtube.com – a popular video shearing website - 20 hours of new video material is uploaded every minute[13]. On flickr.com – a popular image sharing website - for every minute several thousands of new pictures are added[17], so we get daily several million pictures and about 30

thousand hours of video only from these two websites. Rapidshare.com – a popular file hosting website - is able to store more than 10 Petabytes( information provided at the request by[16]).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

4<sup>th</sup> Annual Multimedia Systems, Electronics and Computer Science, University of Southampton

© 2003 Electronics and Computer Science, University of Southampton

Websites to be more attractive for internet users contain more and more multimedia files. To manage with this large amount of information and multimedia files, new efficient methods of multimedia retrieval have been developed, because the old pattern searching algorithms have become inefficient.

## 2. BACKGROUND

Firstly I would like to present some basic information connected with multimedia description.

### 2.1 Features of files

Features of multimedia files can be divided into two groups. First, is signal features; these are low-level features which we can obtain from files automatically such as, colour, spectrum, motion, shape, rhythm. These features are relatively easy to extract from files but they do not contain human interpretation of picture/song/movie. The second group is semantic features; these are high-level features which are dependent on human interpretation. It is hard to obtain this information automatically and without human participation. To get satisfying results of retrieval we have to use both features in a files description. The first solution to this problem was metadata.

### 2.2 Metadata

Metadata is commonly called 'data about data'[1]. It is information that describes files in many regards. It may contain low-level features as well as high-level features as it is created by humans. It is usually connected with files or web pages. One of metadata standard is Dublin Core. It was created in the 1990s to help authors to describe and index their web resources. It is a list of basic information about resources: Title, Creator, Subject, Description, Publisher, Contribution, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, Rights[1]. Its main goal was to allow everyone to catalogue their works with simplicity. For others applications other standards were created, so in a short time many standards have spread[1]. This kind of describing files has many advantages and is still widely used. For example: after filling up information about resources we may use the well-known string searching algorithms to find our file, so it is easy in implementation. Another good point is that it can

contain low, as well as high level features. But it also has some disadvantages. Description of a resource depends on the author of the description; we could have many descriptions of one resource which may cause confusion during retrieval. The biggest disadvantage however is that putting down every key word and adding it to files takes up too much time[1], so we still keep looking for something else that's faster, more efficient and universal.

### 3. MPEG - 7

In October 1996 the Moving Pictures Experts Group (MPEG) started work at "Multimedia Content Description Interface" (MPEG - 7). Unlike its predecessors, MPEG - 7 it is a multimedia content description standard. Its main goal is to create uniform and universal description, independent of the type of file that could be widely used in retrieval of audiovisual data and search of related materials. First version of the Working Draft appeared 1999 and in 2001 it became an international standard, but the first parts of standards were not published until 2002 [2].

#### 3.1 Abilities

MPEG - 7 allows to describe audiovisual media in many ways including [2]:

- basic information about creation, production ( for instance title, authors, publish date, ...) and copyrights;
- information about type of source (format, compression, size,...);
- low-level features (colour, shape, textures, melody,...) and structure of content;
- information about objects contained in the source and the interaction between them;
- ability to connect information about set of objects;
- history of user usage.

#### 3.2 Main components

- Descriptor Tools – are responsible for multimedia description. They consist of [2]
  1. Descriptors – syntax definition of description( how to describe each feature);
  2. Descriptors Schemes – define syntax and relationship between Descriptors (features relation) and/or others Descriptors Schemes (DS). Using DS we could present our description as a tree structure where leaves are descriptor, nodes are DS and root is or description of source. This tree structure makes description easier in analysis ;
- Descriptors Definition Language (DDL) - definition of syntax of language which

allows the creation of Descriptors and its Schemes. It is base on Extensible Markup Language (XML) Schema which goal is presenting data as a structure. Mpeg group adjusted; this language to MPEG - 7 demands.

- System Tools – supports storage and transmission .

What is important in MPEG - 7 is that it defines only structure of description and does not define feature extraction and retrieval process (see figure 1). It allows us to improve feature extraction and searching algorithms and apply them without changing standard[3] (in part 7 of standard we have some recommendation on extracting features, but it is only recommendation ).

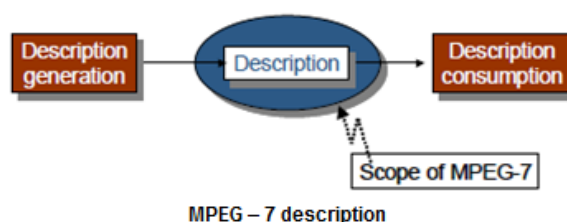


figure 1 from [1]

#### 3.3 Structure of Standard

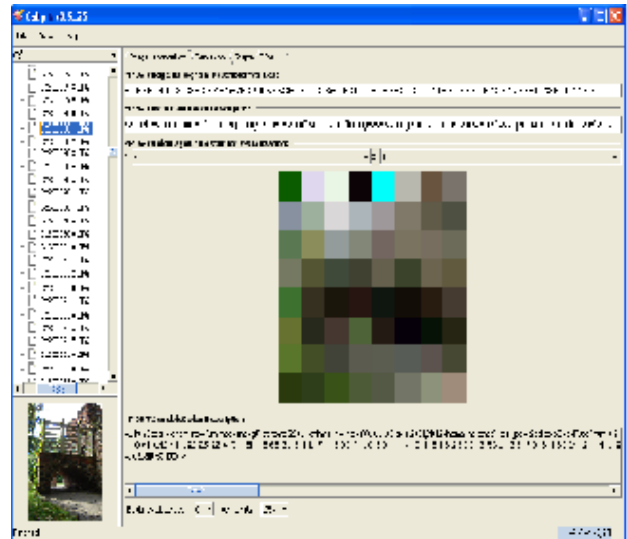
MPEG - 7 is divided into many parts. I will try to present most significant[2]:

- Part 1 – System. It specifies way of storage and transport descriptions. It is responsible for synchronization description with content;
- Part 2 - Description Definition Language. One main component, it has been briefly described above;
- Part 3 – Visual. Defines a set a visual descriptors divided in 7 categories:
  1. Basic Structures – describe structure of description and content. It allows us to divide our picture into some smaller parts and describe each separately (this idea is also used in mpeg video codecs);
  2. Colour Descriptors – give information about colours which are used. Defines colour space, quantization, colour histogram and distribution of colours.
  3. Texture Descriptors – describe texture and edges of picture;
  4. Shape Descriptors – contain information about shape. It is also prepared for 3D shapes descriptions.
  5. Motion Descriptors – it can describe motion of object or camera also in 3D. It can also give some information about distribution of motion localisation (where is object move ) and how much it moves ('quantity of action ');
  6. Localization Descriptors – define region

- or object on or video
- 7. Others – this is a place for unclassified or not invented ways of description; Especially high-level features like face recognition;
- Part 4 – Audio. Defines a set of audio descriptors which describe signal properties like power, envelope and its spectrum. It also contains some specialized descriptors which are used in sound recognition (for example AudioHarmonicity or LogAttakTime descriptors) and high-level features of description like melody or word and speaker recognition;
- Part 5 – Multimedia Description Scheme. Definition of syntax of DS;
- Part 6 – Reference Software. Is a type of software eXperimental Model (XM) created for testing MPEG- 7;
- Part 7 – Conformance testing. Includes testing procedures which check conformance with MPEG – 7;
- Part 8 – Extraction and use of descriptors. This part is a set of documents and guides concerning extraction of features and descriptors usage;
- Part 9 – Profiles and levels. This part (analogically as MPEG 4) is used to profile creation of different applications.

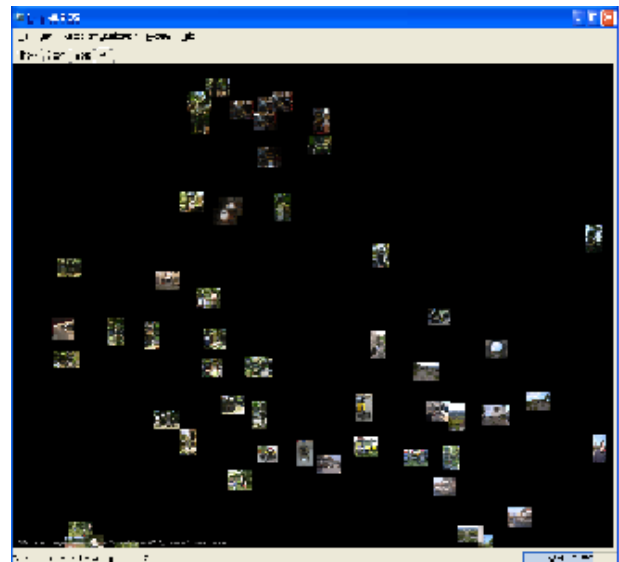
### 3.4 Application Example

Caliph and Emir [20] is a pair of programs which use MPEG – 7 technologies. Using Caliph we can create extended descriptions of images and preview some counted descriptor (see figure 2). The whole description is saved in a different file. Using Emir we can retrieve pictures using separate MPEG-7 descriptors or their combinations. Retrieval consists of comparing values of descriptors (the smaller difference, the more similar pictures are). We can also see visualisation of this difference (see figure 3). On the picture below we see visualisation of ColorLayoutDescriptor. The closer pictures are, the more similar the value of descriptor they have, so we can notice some groups of pictures with similar colour distribution.



Caliph - Mpeg - 7 description

figure 2 [screenshot from Caliph]



Emir - descriptor visualization

figure 3 [screenshot form Emir]

## 4. MPEG - 21

Another problem with large amounts of multimedia files is its delivery of every file to the final user. For instance, if we want to have some multimedia resources from a concert, we will have to look for pictures, video, music and maybe some articles. The idea of MPEG -21 is to collect all of these things and put them into one container which also includes description of content and the relation between its users(see figure 4). MPEG – 21 also contains copyright protection, which is an important feature for multimedia providers. MPEG - 21 Multimedia Framework is another standard developed by the MPEG group. It supports exchange, trade, delivery and consumption of

multimedia. Its goal is make it transparent and interoperable on all systems. It includes content description (using MPEG – 7 descriptors), right management and protection, relation between content and users adaptation and content processing[10].

### 4.1 Main components

This new standard is based on two terms[8]:

- Users – this is everyone who uses multimedia. It may be individuals, consumers, communities, organisations, corporations; in simple words everyone who uses MPEG – 21. Each user is recognized by the relationship and interaction between other users.
- Digital Item – is digital representation of multimedia. We can determine 3 parts.
  1. Resources –multimedia content ( music , pictures , video)
  2. Metadata – description of content ( MPEG -7)
  3. Structure – relation between parts of standard

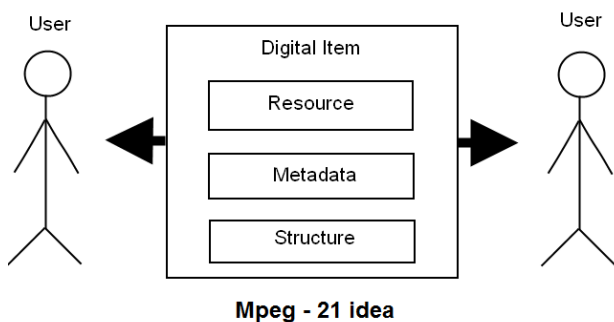


figure 4 based on [22]

### 4.2 Structure of standard

MPEG 21 standard contains 18 parts. I will list and elaborate only the most significant[10]:

- Part 1 – Vision, Technologies and Strategy – presents the idea of easy multimedia exchange using multimedia container. It also presents strategy of standard development;
- Part 2 – Digital Item Declaration (DID) definitions of the Digital Item model structure and some terms useful during creating Digital Items. It allows the creation of a multimedia set which includes many kinds of resources, description and relation between them .The most important terms are :
  - Component – is a resource connected with its description;
  - Item – group of components and/or other items. It adds description related to this set of items and/or components. In this way we achieve another level of

- description (more general);
- Container – group items in logical package;

This part allow to create tree structure that is easy in usage and analysis

- Part 3 – Digital Item Identification (DII) – responsible for identification resources, items, components. It supports many ways of identification, for example Uniform Resource Identifiers (URI), identification using descriptors, for music International Standard Recording Code (ISRC), for video International Standard Audiovisual Number (ISAN) or using special Statement in Descriptors;
- Part 4 – Intellectual Property Management and Protection (IPMP) – this part appeared in MPEG – 4 specifications. It was additional data which helped to protect copyrights. MPEG – 21 extends this idea by adding tools to retrieval and authentication. It is integrated with Rights Expression Language;
- Part 5 – Rights Expression Language (REL) – is a language which describes rights, principles – users who can use those rights, resources which are related with these rights and conditions when they can be used;
- Part 6 – Rights Data Dictionary (RDD) – dictionary with legal terms. It supports REL;
- Part 7 – Digital Item Adaptation (DIA) – The main goal of this part is to provide multimedia exchange (with contracted quality) or descriptions between Users without setting up network by them;

### 4.3 Applications

MPEG – 21 can be used mostly in entertainment. Using this standard we can store a multimedia pack with wide description, we can publish our work on the internet with some copyrights protection or usage restriction. I think that the development and popularity of MPEG – 21 will bring authors and producers the highest profits, because it contains digital right management. It can protect artists from loss caused by illegal usage and copy. But it also can be the biggest disadvantage of this standards for internet users who do not like to pay for everything.

### 4.4 Application Example

AXMEDIS Editor[21] (see figure 5) is an application based on MPEG – 21. It creates multimedia objects which can contain resources ( audio, video, pictures, text), MPEG – 7 descriptors, digital right management( we can create our license and/or make detailed usage restrictions, like the number of times we can play it), and other multimedia objects.



## 7. REFERENCES

- [1] National Information Standards Organization, *Understanding Metadata*, NISO Press, 2004
- [2] International Organisation for Standardisation Organisation, *Mpeg – 7 Overview*, Klagenfurt, July 2002
- [3] Chang S., IEEE Transaction on Circuits and Systems for Video Technology, Special Issue on Mpeg – 7, *Overview of the Mpeg – 7 standard*, June 2001
- [4] Skarbak W., *MPEG – 7*, Koscielisko Conference, October 2003
- [5] Keller T. and Skarbak W., *Standard MPEG-7 w archiwizacji multimediiów*, III Krajowa Konferencja pt. Multimedialne i Sieciowe Systemy Informacyjne , Wroclaw 2002
- [6] Chen T. and Puri A., *Multimedia systems, standards and networks*, Pereira F., Koenen R. H., 22, CRC, 2000
- [7] International Organisation for Standardisation Organisation, *Mpeg – 21 Overview*, Shanghai 2002
- [8] Wang X., *Mpeg – 21 Overview*, Workshop on Multimedia New Technologies and Applications, Shenzhen, October 2007
- [9] Kosch H., *Distributed Multimedia Database Technologies Supported by MPEG-7 and MPEG-21*, 2-3, CRC, November 2003
- [10] ISO/IEC TR 21000-1 , *Multimedia Technology – Multimedia framework (MEPG-21)*
- [11] Herman I, Introduction to Semantic Web, 2009 Semantic Technology Conference San Jose, California, USA , June 15, 2009
- [12] Berners-Lee T., Bizer C., Heath T., Idehen K., *Linked Data On The Web*
- [13] You Tube Service [www.youtube.com](http://www.youtube.com)
- [14] Netcraft [www.netcraft.com](http://www.netcraft.com)
- [15] InternetWorld Stats [www.internetworldstats.com](http://www.internetworldstats.com)
- [16] Rapidshare [www.rapidshare.com](http://www.rapidshare.com)
- [17] Flickr [www.flickr.com](http://www.flickr.com)
- [18] Central Intelligence Agency [www.cia.gov](http://www.cia.gov)
- [19] Wikipedia [www.wikipedia.org](http://www.wikipedia.org)
- [20] Caliph & Emir [www.semanticmetadata.net](http://www.semanticmetadata.net)
- [21] Axmedis Editor [www.axmedis.org](http://www.axmedis.org)
- [22] Information Technology – Mpeg-21 Multimedia Framework <http://mpeg-21.itec.uni-klu.ac.at/cocoon/mpeg21/>
- [23] Carr I., Notes form Hypertext and Web Technologies Course, University of Southampton