MPEG-7 DESCRIPTORS FOR EARTH OBSERVATION SATELLITE IMAGES *

X.Giró and F.Marqués
Universitat Politècnica de Catalunya
(xgiro, ferran)@gps.tsc.upc.es

F.J.Marcello and F.Eugenio
Universidad de Las Palmas de Gran Canaria
(fjmr, feugenio)@dsc.ulpgc.es

Abstract

The amount of digital multimedia information has experienced a spectacular growth during the last years thanks to the advances on digital systems of image, video and audio acquisition. As a response to the need of organizing all this information, ISO/IEC has developed a new standard for multimedia content description called MPEG-7. Among other topics, MPEG-7 defines a set of multimedia descriptors that can be automatically generated using signal processing techniques. Earth Observation Satellites generate large quantities of images stored on enormous databases that can take advantage of the new standard. An automatic indexation of these images using MPEG-7 meta-data can improve their contents management as well as simplify interaction between independent databases. This paper gives an overall description on MPEG-7 standard focusing on the low-level Visual Descriptors. These descriptors can be grouped into four categories: color, texture, shape and motion. Visual Color Descriptors represent the color distribution of an image in terms of a specified color space. Visual Texture Descriptors define the visual pattern of an image according to its homogeneities and non-homogeneities. Visual Shape Descriptors describe the shape of 2D and 3D objects being, at the same time, invariant to scaling, rotation and translation. Motion Descriptors give the essential characteristics of objects and camera motions.

These descriptors can be used individually or in combination to index and retrieve satellite images of the Earth from a database. For example, oceans and glaciers can be discerned based on their Color Descriptors, also cities and desert based on the Texture Descriptors, island images can be grouped using the Shape descriptors and cyclone trajectories studied and compared using Motion Descriptors.

1 Introduction

MPEG-7 [1] [2] is the new standard for the indexation of multimedia documents promoted by the International Organization for Standardization (ISO) [3] and the International Electrotechnical Commission (IEC) [4]. MPEG-7 is born with the will to be the standard reference in the Semantic Web era, supporting multimedia and textual queries. This indexation tool is offered to the archiving community to manage the huge amount of multimedia content generated every day, a figure growing exponentially thanks to the technological advances on the acquisition, transmission and storage of multimedia data.

The aim of MPEG-7 standard is to index any type of multimedia material (image, video, text, audio...) independently from the support media (hard disk, digital broadcast, Internet streaming...) or coding format (JPEG, BMP, MPEG-1...). MPEG-7 supports the description of multiple types of information

*This work has been partly supported by the University, Research and Information Society Department of the Generalitat de Catalunya and by the grant CICYT TIC2001-0996 of the Spanish Government.

contained or related to the document, from the lowest features (color, audio pitch, motion...) to the highest abstractions (ideas, concepts, emotions...). MPEG-7 was designed as a generic tool with an extensible architecture to provide a powerful framework for *Content Based Index Retrieval (CBIR)*.

MPEG-7 is neither a standard for the compression of multimedia content nor a toolbox for signal processing. It just establishes a set of descriptors and structures for the indexing of multimedia content, without defining how to generate them.

Satellite image databases could improve their accessibility and management by using the tools proposed by MPEG-7, among them:

- international agreed exchange format for content indexes between heterogeneous databases belonging to different agencies and companies
- binary format suitable for satellite transmission
- a common access point for accessing images stored in diverse formats
- extensible architecture to support new advances
- ability to associate extra data in any format associated to an image
- visual descriptors to support CBIR

The organization of the paper is as follows. Sections 2, 3 and 4 present the basic elements of MPEG-7. Section 5 concentrates on the textual annotation with MPEG-7, while Section 6 studies the proposed Visual Descriptors. Section 7 discusses the differences between a notation based on text or on visual descriptors and conclusions are reported in section 8.

2 Coding Schemes

Coding Schemes [5] are the MPEG-7 set of tools related to the binarization of the descriptors, satisfying requirements such as synchronization, random access, error resilience or streaming.

MPEG-7 descriptors can be embedded in the content or form an independent stream. MPEG-7 files can be stored in two formats. The *Textual Mode (TeM)* is a human-readable format mode based on XML [6]. On the other hand, the *Binary Mode (BiM)* has been created to offer a better format for transmission and storage. BiM provides compression, fast parsing and filtering without having to decompress the whole stream. BiM should be used when transmitting descriptors through satellite links, where bandwidth and bit-rate are scarce resources.

3 Normative elements

MPEG-7 provides a wide range of tools for the managing of multimedia content. These functionalities are defined by combining three types of normative elements: *Descriptors*, *Description Schemes* and a *Description Definition Language*.

Descriptor (D) A *Descriptor* represents an attribute related to the document whose value can be evaluated for classification or analysis purposes. MPEG-7 standardizes a set of low-level descriptors such as color, motion or audio energy; as well as high-level descriptors such as author, time stamp or title. Low-level descriptors can be extracted automatically and many of them are associated to a particular media support, such as image or audio. On the other hand, high-level descriptors are textual and may require the action of a human-expert. This paper concentrates on the low-level visual descriptors and its applications on satellite images.

Description Scheme (DS) *Description Schemes* specify the relations among Ds and DSs, as well as their semantics. MPEG-7 categorizes DSs into visual, audio or generic. A DS can be related to many diverse entities like satellite images, video sequences, music CDs or authoring information.

Description Definition Language (DDL) The *Description Definition Language* [7] is the language for defining the set of description tools used in MPEG-7 (DSs, Ds, and datatypes). DDL is based on the W3C's XML, but some extensions have been added to support multimedia content description.

4 Multimedia Description Scheme tools

MPEG-7 offers a large amount of tools for the management of multimedia documents grouped under the so-called *Multimedia Description Scheme* [8]. They can be grouped in five categories: *Content management*, *Content description*, *Navigation and access*, *Content organization* and *User Interaction*.

4.1 Content management

The whole life cycle of a multimedia document can be described through the Ds and DSs standardized by MPEG-7.

Media Information Every instance of a multimedia document is associated to a *MediaInformationDS*. This DS contains features of the coded multimedia data by including one or more *MediaProfileDS* and, optionally, a *MediaIdentificationD*.
The *MediaProfileDS* describes variations of the content produced from the original or master media. Four descriptors have been defined for this purpose. The *MediaFormatD* specifies the format, size and compression rate of the content. It is common in satellite image databases to manage files in several storage formats (HDF, LAN, GeoTIFF, JPEG...). Presently, most of the software just checks for the extension of the image files to identify the format, a very simple technique that can drive to errors and problems. *MediaFormatD* provides a unique access point to the contents that simplifies the database management and software design. The *MediaInstanceDS* identifies and locates one or several instances of the document thanks to the *MediaLocator* and *LocationDesriptor* basic elements. A list of mirrored instances of the document is a possible utility of this descriptor. The *MediaTranscodingHintsD* improves the quality and reduces the complexity of the transcoding applications. For example, in video applications, information a priori about the motion vectors associated to a segment can improve the performance of the video decoder. Finally, the *MediaQualityD* describes the objective and subjective quality of the images. Some images in the satellite database could be marked as low quality to prioritize better quality images as a response to a query.
The *MediaIdentificationD* identifies the content entity independently to the different available instances. It specifies the source, acquisition and/or use of the image, audio or video content.

Creation Information MPEG-7 defines a set of DSs to include information not depicted in the content, such as title, author or genre. Three DSs fall into this category. The *CreationDS* can content textual information regarding the satellite, area displayed, date and time of the image, as well as about the company or agency who took the picture. The *ClassificationDS* offers a label for classification, for example, based on the language or genre of the document. Satellite images illustrating rare meteorological phenomenas could be classified as so with this DS. The *RelatedMaterialDS* refers to other associated documents; for example, a video about the launch of the satellite that took the picture or other images from external databases.

Usage Information The *UsageInformationDS* includes important information about how a multimedia document is used. Although MPEG-7 does not directly deal with usage rights, the *RightsD* provides a link to access the current rights owner, for example, a space agency or company. The *FinancialD* informs about the production costs or incomes resulted from the content use. The availability of the content instances, type of publication or price for using is contained in the *AvailabilityDS*. A company selling satellite images could provide pricing information to its on-line costumers with this DS. The usage history of the content might also be interesting, so the *UsageRecordDS* was created to do so. This DS could mark those images in the database that have received more accesses, helping this way to future queries.

4.2 Content description

4.2.1 Structural aspects

Multimedia content can be described according to the spatial, temporal or spatiotemporal structures of the multimedia segments that form it.

Segment Entities The basic *Segment Entity (SE)* is the *StillRegionDS*, a spatial region of a 2-D image or video frame. However, more complex segments exist such as the *MovingRegionDS* for spatiotemporal intervals or the *VideSegmentDS* for temporal intervals, among others describing audio, ink, mosaics or transitions. SEs can be formed by a set of components that are not connected among them, neither in the temporal nor spatial dimensions. This is an useful issue for satellite images because some of the objects to deal with are treated as a unique entity despite not being visually connected. This is the case of archipelagos, or portions of ground or sea that might be visually disconnected because of cloud occlusions.

Segment Attributes Descriptors can be assigned to a complete document or to a single SE. Each SE can have its own creation, usage, visual descriptors or textual annotation. Moreover, the abstract *MaskD* enables the assignment of a descriptor to a SE for a limited spatial or temporal segment. The *MatchingHintD* describes the relative importance of the segments and their descriptors to improve the retrieval performance. For example, in an image with many islands, the ones in the center can be given more relevance than some others near the edges.

Segment Decomposition SEs can be decomposed by using the *SegmentDecompositionDS*. The resulting subsegments might overlap, so their union does not need to reconstruct the whole. A partition tree [9] of a still image could be described by such a tool. This would enable efficient search strategies (global or local) and allow a scalable description of the image. For example, each node in the partition tree could have its own visual descriptors associated, either by extracting them from the region or by computing them from the union of the visual descriptors of its subsegments.

Structural relation MPEG-7 supports the description of spatial and temporal relations. The *SpatialRelationCS* and *TemporalRelationCS* define a set of normative structural relations such as *south*, *left* or *below* for space and *precedes*, *during* or *overlapping* for time. The segments "oil spill" and "Lanzarote" can be related by the Spatial Relation "north of" through a *GraphDS* to describe the structural relation "oil spill on the north of Lanzarote".

4.2.2 Semantic aspects

Narrative worlds, objects, events or concepts are treated in MPEG-7 as *Semantic Entities*. An example of an image showing an intensive pollution might content the abstract concept 'disaster'. Such high-level abstraction can be represented by using graphs or with a simple textual annotation. Automatic extraction of semantic entities, usually referred as the *semantic gap* problem, is presently an important research topic for the scientific community.

4.3 Navigation and summarization

The access to multimedia content is not always done under the same circumstances. The user might choose between a thumbnail of an image or its complete version, might be accessing from a mobile terminal with bandwidth restrictions or prefer a language over the others. MPEG-7 offers different ways for accessing the different versions of the document.

Summaries Multimedia summaries provide the highlights of the content. Two types of schemes are defined; *HierarchicalSummaryDS* and *SequentialSummaryDS*. Accessing a set of Earth images through a tree organization in continents might be an example of hierarchical access, while a small set of video-shots of the evolution of a tornado is an example of sequential access to a video.

View and View Decomposition In certain cases, a unique multimedia content can be stored in different spatial, temporal or frequency points of view. MPEG-7 provides the *View* tool for accessing them. Multi-band images, so common in satellite imaging, can be described by views.

Variations Variations in the content have also been studied in MPEG-7. Compressed, low-resolution or multi-language versions are some of the example applications. For example, it would be interesting to always retrieve on screen the maximum resolution version of the image supported by the user's display, so

as not to waste bandwith. In another case, if an image contains textual notations, several language versions might be stored.

4.4 Content Organization

When having large amounts of documents in a database, it is useful to cluster them into groups according to a common feature. MPEG-7 provides tools for the creation and management of *Collections* based on *Models*.

Collections *Collections* are unordered sets of multimedia contents, segments, descriptors, concepts or a mix of them. *Collections* can be based on low-level descriptors, concepts or in a structured combination of both. For example, the retrieved images of a query can be organized as a *Collection*.

Models A *Model* is a parametrized representation of multimedia content or collections of multimedia content. Four types of models have been defined in MPEG-7. The *Probability model* describes samples of multimedia content and classes of descriptors using probabilities and statistics. For example, certain configurations of the Edge Histogram Descriptor (see section 6.2.3) may model the texture of city areas. The *Analytic model* associates labels or semantics to a collection or probability model, allowing multiple semantics associated to a single model. A particular case could be a collection of images of polluted cities, burned forests and oil spills under the tag "ecological disaster". The previous two models can be combined in a *Cluster model*, a set of analytic models and their associated probability models. *Cluster models* allow the definition of the centroid of a collection of descriptor instances to represent a semantic class. The previous texture-based definition of "City areas" could be combined with another color histogram by using a *Cluster model*. Finally, *Classification models* are used to associate labels or semantics to unknown content. The *Classification model* conatins attributes referring to the completness or redundancy by which the classification tool covers the semantic space. For example, a classification of Earth satellite images between "forests" and "cities" would be marked as *incomplete* because regions like "oceans" or "deserts" are not taken into account.

4.5 User interaction

Access to multimedia content can be improved by using user preferences. MPEG-7 defines *PreferenceDS* and *HistoryDS* for the search, filtering and browsing in databases according to the user profile. *UsageHistoryDS* collects a list of the actions performed by the user in previous database accesses. *UserPreferencesDS* expresses the preferences of the user when accessing the content, such as a supported formats, resolution or language. Preferences can be automatically inferred by analyzing the past history; for example, when a user has shown previous interest on a certain geographical region, in a new query the system might retrieve first images belonging to that same region.

5 Textual annotation

Textual annotation is the traditional method for the indexing of images. MPEG-7 offers an XML based framework to include this textual information following a structured pattern to facilitate the notation, retrieval and exchange of meta-data. Specific descriptors are defined for time, people or space, but also generic textual descriptors allow any sort of annotation related to the multimedia document.

Time The *Time datatype* is the basic element in MPEG-7 that expresses an absolute time point in the real world, a common meta-data in satellite image databases. MPEG-7 has taken the ISO8601 standard as a reference for notation with a slight modification for sub-second times.

People Three abstraction levels are defined for describing people. *PersonDS* describes a single person name with his/her contact information; for example, the last person who retrieved a certain image. *PersonGroupDS* is used with groups of persons belonging to the same collective, such as the team responsible for the image meta-data. Finally, *OrganizationDS*

refers to collectives of people whose individual members are not easily identifiable, such a space agency or a company.

Places The *PlaceDS* includes information about the shooting location and geographic position, like the longitude, latitude and altitude of the satellite when it took the picture.

Free text Textual entries with no restrictions are also possible in MPEG-7. A notation such as "Cloudy image of the Canary Islands taken from the Meteosat" is an example of free text annotation.

Keyword annotations Managing unrestricted natural language is difficult for a computer. A first approach to solve this problem is to use a simpler textual annotation by removing all kind of structure in the text. This results on a list of keywords that, despite losing some information, make the description more understandable for a computer. Following the previous example, a valid keyword annotation would include terms like "Canary Islands", "Meteosat", "clouds".

Structured annotations One of the main problems of using keywords is not knowing what kind of information they are providing. The problem is overcome by enriching the text annotation with simple structured annotations like "Where: Canary Islands" or "When: January 16, 2002".

Dependency structure An even more refined method is based on a linguistic theory called *dependency grammar*. In this case, a natural language sentence is decomposed on a set of subtrees according to its grammatical category. This type of textual annotation supports complex queries like "give me all the images of the Canary Islands".

Classification schemes Restricted vocabularies can be defined with the *Classification schemes (CS)*, a list of term definitions and, optionally, imported classification schemes. Domain-specific terms with multilingual support could be defined by a CS. For example, a Catalan toponymic annotation like "Illes Canàries" could be easily associated to the English term "Canary Islands" through a CS.

6 Visual descriptors

One of the most revolutionary tools introduced by MPEG-7 are the low-level descriptors, and among them, those aimed at visual content. MPEG-7 Visual Descriptors [10] are the first step in trying to bridge the semantic gap between the pixels contained in a digital image and the idea in the mind of the user when querying the database. Visual descriptors express certain combinations of pixels in a compact and useful way. MPEG-7 divides the Visual Descriptors in two big families: general and domain-specific. The general family contains low-level visual descriptors describing colors, textures, shapes and motions. These descriptors can be applied to any kind of image and MPEG-7 has defined a wide range of them to satisfy multiple applications. On the other hand, domain-specific descriptors have been developed for a certain type of image. Face-recognition descriptors are an example of them. Domain-specific descriptors are expected to grow in the future and new ones may be included in the standard. In fact, the Earth Observation community could consider developing their own visual descriptors if the generic ones do not cover their needs.

Having visual descriptors associated to each image opens the door to non-textual visual queries. A generic visual query retrieves those images whose visual descriptors are more similar to the query descriptors. Database descriptors are automatically generated when an image is added to the database, while query descriptors can be manually chosen by the user or extracted from an example image.

6.1 Color

Color descriptors [11] are widely used because they are intuitive, simple and robust to viewing angle changes and rotations. MPEG-7 specifies five different descriptors and six color spaces to satisfy a wide range of queries based on color. The six color spaces supported are RGB, YCbCr, HSV, HMMD, Monochrome and a linear transformation matrix with reference to RGB.

6.1.1 Dominant Color Descriptor (DCD)

The DCD is a compact color descriptor specially designed for similarity retrieval and browsing. It can support up to eight colors but, unlike the predefined bins used in other histogram techniques, these dominant colors are computed for each image. A clustering algorithm partitions the picture following a criteria based on color homogeneity. DCD describes, for each cluster, the dominant color, the cluster percentage of pixels in the image and, optionally, the variance of the pixel values. Finally, a last parameter expresses the overall spatial homogeneity of the dominant colors. By identifying clouds with a certain color, DCD could be used to evaluate the percentage of clouds present in an image or segment.

6.1.2 Scalable Color Descriptor (SCD)

The SCD is based on the Haar transform applied on a 256-bin histogram in the HSV color space. The Haar transform is based on the sum and difference operations, which can be respectively interpreted as low and high pass filters. Scalability is achieved merging adjacent bins to create coarser representations of the histogram. To be more precise, SCD supports histograms of 128, 64 and 32 bins. Another form of scalability is also possible by truncating the coefficients in the Haar transform associated to each bin. As negative values may appear in the difference branch of the Haar transform, an extremely compact representation of SCD may consist with only the sign bit of each coefficient, though this is the extreme case and an intermediate solution can be also interesting. Scalability can speed up the retrieval process if a first search is done on the lower scales and, if necessary, later refined at higher resolutions. For example, a user could make a query based on a color to retrieve images with similar colors, with no need of that precise and patricular color. In this case, it would be useless to use the maximum resolution available at the descriptor and, in fact, a tolerance around the reference color would be welcomed by the user. The desired functionality can be achieved with a coarse representation of the SCD.

6.1.3 Group-of-Frame (GoF) or Group-of-Pictures (GoP)

GoF and GoP are special descriptors that, instead of being associated to a single image, contain the combined color characteristics of multiple frames or sets of images. They are calculated by combining the SCD descriptors of each individual frame or picture. MPEG-7 offers three operators for combining the SCD descriptors: average, median and aggregation.

GoF overcomes the problem of finding a key-frame with representative color properties in a video sequence. With the GoF descriptor, all frames count when computing the descriptor, avoiding possible errors caused by a selection of a non-representative key-frame.

GoP is a very useful descriptor when having a set of pictures with similar color properties. For example, in a meteorological sequence, where color histograms between images are very similar, storing a single GoP for the whole set of images would save storage space and avoid multiple redundant entries in the database. Another possible application for fast retrieval is to use the GoP to preselect some candidate images and later refine the search individually.

6.1.4 Color Structure Descriptor (CSD)

CSD provides information about the color content in an image, but adds some extra information about its spatial distribution. To be more precise, it expresses how clumped together pixels belonging to a certain color are. A structuring element of 8x8 scans the image and increments in one unit those histogram bins represented by one or more pixels in the structuring element. This descriptor is a special case in MPEG-7 because the standard defines how to extract it; otherwise, the matching process would be too complex. The HMMD color space is used to create histograms of 256, 128, 64 or 32 bins. CSD requires a processing effort to provide some spatial information.

In satellite images, this descriptor can discern between a picture of a continental coast from another one of an archipelago, even if they have the same proportion of land and water.

6.1.5 Color Layout Descriptor (CLD)

CLD provides the spatial distribution of the colors in a very compact way. Images are partitioned in the YCbCr space into 64 rectangular blocks and a representative color for each block is chosen. The DCT transform is applied on each YCbCr component of the resulting 8x8-pixels image. Finally, the resulting coefficients are zigzag scanned and weighted to generate the final stream. CLD allows scalability by controlling the number of DCT coefficients. It also supports image matching as a whole or as parts of it with any arbitrary shape. This is an interesting descriptor for applications with storage and band-with limitations as well as for fast retrieval. CLD could also be used for the filtering of images from heliosynchron polar-orbiting satellites. Due to Earth rotation, these satellites sometimes provide images where the region of interest appears shifted. In addition, the Earth curvature provokes a degraded spatial resolution at the image boundaries. When both effects combine on the region of interest, the image must be discarded. CLD could help in the filtering process when the region of interest presents an important color contrast with the surrounding regions. For example, a valuable application of CLD would be the selection of images in which a specific island appears centered.

6.2 Texture

We talk about texture when referring to some basic pixel primitives that are repeated in a connected region. MPEG-7 has standardized three texture descriptors [11]; one describing the homogeneity of the texture, another one for non-homogeneities and a third one that tries to give a response similar to the human visual system.

6.2.1 Homogeneous Texture Descriptor (HTD)

The HTD provides a quantitative representation of the texture of an image through its mean energy and energy deviation in the frequency plane. The frequency plane is expressed in polar coordinates to provide the descriptor with rotation invariance. The texture descriptor of a rotated image is an angular shifted version of the non-rotated one so, during the matching process, the query descriptor is angularly shifted in all possible cases. The frequency plane is partitioned into 30 frequency sub-bands using the Gabor functions. The angular direction is uniformly split but not the radial direction, which is divided in octave scale to adapt to the human visual response. The mean energy and energy deviation in each frequency channel are calculated by combining the Gabor functions and the Fourier transform of the image. The later can be efficiently computed by using the Radon transform. The mean and the standard deviation of the image are also coded, providing a descriptor of 62 coefficients. Several applications in the field of satellite imagery can be developed based on this descriptor. This descriptor could be used to find of certain types of vegetation or oil spills on the ocean. As an example, the Alexandria Digital Library project [12] [13] in the University of California is indexed by extracting HTD from non-overlapping 128x128 tiles.

6.2.2 Texture Browsing Descriptor (TBD)

MPEG-7 presents TBD as a simple and compact descriptor that provides a characterization of the texture similar to the human perception. Fast response on retrieval applications is achieved by using only 12 bits to represent the regularity, coarseness and directionality of a texture pattern. The extraction is similar to the HTD because it performs a frequency decomposition of the image using the Gabor functions in the polar plane. By doing this, a multi-resolution decomposition in frequency is obtained, with each decomposed image providing information at a certain orientation and scale. In similarity retrieval, the TBD can be used to find a set of candidate images and the search can be later refined with the HTD.

6.2.3 Edge Histogram Descriptor (EHD)

The EHD provides information about the spatial distribution of the edges in an image. The image is split into 4x4 sub-images and an edge histogram for each

sub-image is built. The histograms are created by further subdividing the sub-images in around 1100 image-blocks and assigning a label to each of them according to its edge orientation. Five categories are defined: vertical, horizontal, 45 degrees, 135 degrees or non-directional. EHD also includes a global and a semi-global edge histograms by accumulating the local histograms for the whole image or for sub-blocks. As a result, a total of 5 global bins + 65 semi-global bins + 80 local bins are available for matching. Experiments have shown that EHD is useful in the retrieval of natural images with nonuniform textures and clip art images, as well as for sketch retrieval. The spatial distribution contained in the EHD could be useful to find certain spatial distributions of textures, for examples, cities at occidental coasts.

6.3 Shape

Shape Descriptors (SD) possess important semantic information. A proof of it is that humans are capable of recognizing objects by just seeing its shape. However, all this semantic richness can only be extracted with a previous good segmentation. Unfortunately, it does not exist yet any generic segmentation technique comparable to the human visual system, but several application specific algorithms have been developed with satisfying results. MPEG-7 provides region- and contour-shape descriptions for 2D images and a specific one for 3D volumes.

6.3.1 Region-based Shape Descriptor

The Region-based SD provides similarity between those images presenting close spatial distribution of pixels despite not having the same outline contours. The descriptor is both based on boundary and interior pixels, supporting disconnected regions or regions with holes. It is based on the *Angular Radial Transform (ART)*, an orthogonal unitary transform based on a set of orthonormal sinusoidal basis functions in polar coordinates. ART coefficients are computed for each image and included in the descriptor. Experiments have shown good behavior against rotation and scaling, but not so good in perspective transformations. Satellite images may present perspective problems due to the Earth curvature, so a geometric correction is necessary before applying this descriptor. This could be a good tool to retrieve archipelagos from a database or overcome segmentation disconnections caused by clouds.

6.3.2 Contour-based Shape Descriptor

Contour-based SD is suitable for those objects whose shape information is contained at the contours. It is robust to nonrigid deformations, orientations, scaling, mirroring and perspective transformations. It is based on a Curvature Scale-Space (CSS) representation of the contours. CSS decomposes the contour into convex and concave sections and expresses how prominent they are, how long compared to the full contour and their position. This is done in a multiresolution approach, by applying a smoothing process to the contour at different scales. The CSS curve is obtained by plotting the convex to concave change points on the 2-D plane, with the normalized distance along the contour on the X axis and the amount of smoothing on the Y axis. Contour-based DS is good for non-rigid deformations, so it could be used for the fast detection of islands based on its coast-line, even starting from a sketch.

6.3.3 3-D Shape Descriptor (3-D SD)

3D shape spectrum descriptor expresses volumetric features represented as discrete polygonal 3-D meshes. Because of its 3-D approach, this descriptor could be used for the retrieval of features in Digital Elevation Models (DEM) images generated from satellite data.

6.4 Motion

MPEG-7 offers four descriptors for describing the motion in video sequences. Motion can be associated either to an object present in a sequence or to the camera.

6.4.1 Motion Activity Descriptor (MAD)

The MAD describes the type of motion activity shown in a video sequence. It can be easily extracted

from the motion vectors contained in most of the modern video coding schemes, so it is not necessary to decompress the video sequence to obtain it. The descriptor defines the intensity, direction activity and its spatial and temporal distributions. Traffic control sensors could represent the vehicle fluxes through the highways with MAD.

6.4.2 Camera Motion Descriptor (CMD)

The CMD provides information of the camera motion. The trajectory of the LEO satellites recording video can be described with this descriptor.

6.4.3 Motion Trajectory Descriptor (MTD)

The MTD is applied to moving segments in a sequence, providing information about the position, speed and acceleration of a region. The center of mass of the segment is usually the reference point for the calculation of the three parameters. The evolution of an oil spill can be monitored and studied with this descriptor.

6.4.4 Warping and Parametric Motion Descriptors (WMD and PMD)

Both WMD and PMD provide a parametric model of the motion. The only difference is that WMD is applied to the global image and PMD describes a single region motion. They support five classical parametric motion models; translational, rotation/scaling, affine, perspective and quadratic. Each descriptor specifies a model, a time interval, a coordinate system and the values of the parameters.

7 Image database architectures

MPEG-7 offers two basic ways to access multimedia databases; through classic textual annotation and low level descriptors. Figure 1 shows both options in the case of an image database. Understanding the strong and weak points of the two philosophies is important to combine them effectively.

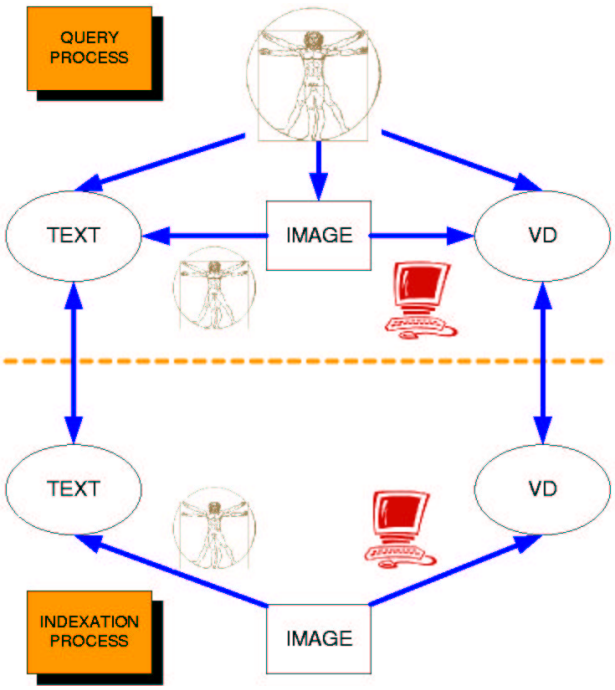

Figure 1: Acces to an image database through textual annotation or visual descriptors.

7.1 Databases relying on Textual Annotation

The traditional approach in image content indexing consists of assigning one or several textual descriptors to every document. Textual labels are extracted as a result of an analysis of the image and are stored as indexes. The generation of textual descriptors depends on a human expert, either as a manual annotator or as a designer of an automatic algorithm. This approach injects some subjectivity inherent in any linguistic annotation during the indexation process, a circumstance that might drive to problems. For example, ecologist groups or petrol companies would probably differ in labeling images as "ecological disaster". Textual queries are also restricted to those text entries which have been previously generated. On the other hand, the matching of queries and database descriptors is fast because it is performed in a limited textual space.

7.2 Databases relying on Visual Descriptors

The low-level MPEG-7 descriptors must be automatically extracted and do not need any further revision, so no human interaction is necessary. They are objective and perdurable. Flexible multimedia queries [14] are supported by them and, when a textual query is submitted, a mapping of the text to a visual descriptors can be done. In this case, the linguistic subjectivity is on the user side, a desirable property. The analysis algorithms are applied on visual descriptors instead of on the image itself, requiring a bigger computational effort during the similarity evaluation than in the textual case. Nevertheless, retrieval speed can be dramatically improved by using an intelligent indexing structure to facilitate the search of descriptors.

8 Conclusions

MPEG-7 standard is a powerful solution for the indexing of multimedia databases, and among them, those containing satellite images. MPEG-7 supports the classical textual annotation to index a wide range of information. However, one of the most revolutionary contributions are the visual descriptors. This paper has shown that relevant queries for the satellite imaging community can be performed based on automatically extractable MPEG-7 visual descriptors. As a conclusion, MPEG-7 offers a wide range of useful tools for satellite image databases, specially for those willing to support multimedia queries and interoperability with external databases.

References

[1] P. Salembier, B.S.Manjunath, T.S. Sikora, *Introduction to MPEG-7*, Wiley 2002

[2] http://mpeg.telecomitalialab.com/

[3] http://www.iso.ch

[4] http://www.iech.ch

[5] ISO/IEC 15938-1:2001, *Multimedia Content Description Interface - Part 1: Systems"*, Version 1, 2001

[6] http://www.w3.org/XML

[7] ISO/IEC 15938-1:2001, *Multimedia Content Description Interface - Part 2: DDL"*, Version 1, 2001

[8] ISO/IEC 15938-1:2001, *Multimedia Content Description Interface - Part 5:Multimedia Description Scheme"*, Version 1, 2001

[9] P. Salembier, J. Llach, L. Garrido, *Visual Segment Tree Creation for MPEG-7 Description Schemes*, Pattern Recognition, Volume 35, Issue 3, pp. 563-579, March 2002.

[10] ISO/IEC 15938-3:2001, *Multimedia Content Description Interface - Part 3: Visual"*, Version 1, 2001

[11] B.S. Manjunath, J.R. Ohm, V.V.Vasudevan, A.Yamada, *Color and Texture Descriptor*, IEEE Transactions on Circuits and Systems for Video Technology, Vol. II, No.6, June 2001

[12] http://vision.ece.ucsb.edu

[13] Manjunath, B. S., Ma, W. Y. (1996). *Browsing large satellite and aerial photographs* (invited paper), Proceedings of 3rd IEEE International Conference on Image Processing, Lausanne, Switzerland, 16-19 Sept. 1996 (pp. 765-8 vol.2). New York, NY: IEEE.

[14] R. Fagin, *Fuzzy Queries in Multimedia Database Systems*, Proc. 1998 ACM SIGACT/SIGMOD/SIGART Symposium on Principles of Database Systems. www.almaden.ibm.com/cs/people/fagin