Chapter 3:

Automated Metadata Extraction from Art Images

Krassimira Ivanova, Evgenia Velikova, Peter Stanchev, Iliya Mitov

1 Introduction

Each touch to the artwork causes building the bridge between cultures and times. The unique specific of visual pieces of arts is that they are created by a cognitive process. It can therefore be instructive not to only understand the way we look at an artistic image, but also to understand how a human being creates and structures his artwork. As was mentioned in [Chen et al, 2005] "research on significant cultural and historical materials is important not only for preserving them but for preserving an interest in and respect for them".

Many art masterpieces have created over the centuries, which are scattered throughout the world. The direct touch to these treasures for most of peoples is impeded by many obstacles. On the other hand, the understanding of the art masterpiece is the process of learning not only of the artefact itself as well as the environment of its creation.

Since its first edition published in 1962 Janson's History of Art [Janson, 2004] is one of the most valuable sources of information spanning the spectrum of Western art history from the Stone Age to the 20th century. It became a major introduction to art for kids and a reference tool for adults trying to remember the identity of some embarrassingly familiar image. The colourful design and vast range of extraordinarily high-quality illustrations does not only present "dry" information, but also evokes deep emotional fulfilment by the touch to masterpieces. However, nowadays online search engines have increased

the appetite of web surfers for context and information, and there are numerous digital collections offering easy access to digital items. They present the colourfulness of art history as well as relevant metadata, provide additional information from purely technical details, ranging from the way of creating the artefacts to deeply personal details from the life of the creators, which help the observers to understand the original message in the masterpieces. For this purposes, the development of the image retrieval techniques became very important for creating appropriate and facile search engines.

The field of image retrieval has to overcome a major challenge: it needs to accommodate the obvious difference between the human vision system, which has evolved genetically over millenniums, and the digital technologies, which are limited within pixels capture and analysis. We have the hard task to develop appropriate machine algorithms to analyze the image. These algorithms are based on completely different logic and "instruments" compared to the human process of perception, but would give similar results in interpreting the input image. In the context of this thesis the challenges are even bigger because we focus our efforts on image analysis of the aesthetic and semantic content of art images. Naturally, the interpretation of what we – humans – see is hard to characterize, and even harder to teach to a machine. Yet, over the past decade, considerable progress has been made to make computers learn to understand, index and annotate pictures representing a wide range of concepts.

In spite of the fact that computers still not wield the human vocabulary and semantic, their methods and abilities for analysis information already make him irreplaceable assistant in many fields of study of art. Computers can analyze certain aspects of perspective, lighting, colour, the subtleties of the shapes of brush strokes better than even a trained art scholar, artist, or expert. As David Stork has mentioned [Stork, 2008] "the source of the power of these computer methods arises from the following:

- the computer methods can rely on visual features that are hard to determine by eye, for instance subtle relationships among the structure of a brush stroke at different scales or colours;
- the computer methods can abstract information from lots of visual evidence;
- computer methods are objective, which need not mean they are superior to subjective methods, but promise to extend the language of to include terms that are not highly ambiguous."

2 Semantic Web

During the years, the ability of processing the information as well as expanding the ways of data exchanging is increasing in parallel. The development of computing and communication capacities allows to place the user in the central point of the process of information exchange and to enable him to use all power of the intellectualized tools for satisfying his wishes. Amit Agarwal [Agarwal, 2009] provides a simple and clear comparison between Web 1.0, Web 2.0 and Web 3.0 (Table 4).

Table 4. "Comparison table" between Web 1.0, Web 2.0, Web 3.0 (excerpt from [Agarwal, 2009])

Web 1.0	Web 2.0	Web 3.0
"the mostly read only web"	"the wildly read-write web"	"the portable personal web"
Focused on companies	Focused on communities	Focused on the individual
Home pages	Blogs	Lifestream
Owning content	Sharing content	Consolidating dynamic content
Britannica Online	Wikipedia	The semantic web
Directories ("taxonomy")	Tagging ("folksonomy")	User behavior ("me-onomy")
Netscape	Google, Flickr, YouTube	iGoogle, NetVibes

Starting from read-only content and static HTML websites in Web 1.0, where people are only passive receivers of the information, Web 2.0 became as participation platform, which allows users not only to consume but also to contribute information through blogs or sites like Flickr[119], YouTube[120], etc. These sites may have an "Architecture of participation" that encourages users to add value to the application as they use it.

According to David Best [Best, 2006], the characteristics of Web 2.0 are: rich user experience, user participation, dynamic content, metadata, web standards and scalability. Further characteristics such as openness, freedom and collective intelligence by way of user participation, can also be considered as essential attributes of Web 2.0.

The pros and cons of using such paradigm as well as other one are many; for a good range of initiatives of social media outreach in the cultural heritage institutions see [WIDWISAWN, 2008]. Let´s mention alternatives, discussed from Eric Raymond in [Raymond, 1999] concerned two fundamentally different development styles, the "cathedral" model of most of the commercial world versus the "bazaar" model of the Linux open-source world, where the advantages of such social self-build

[119] http://www.flickr.com/
[120] http://www.youtube.com/

systems are shown. Here the situation is similar. For instance, while the Encyclopaedia Britannica Online[121] relies upon experts to create articles and releases them periodically in publications, Wikipedia relies on anonymous users to constantly and quickly contribute information. And, as in many examples, the happy medium is the right position. Many art repositories and portals are used for educational purposes; consequently control over the main presented text is very important. On the other hand, they are natural places for users to share their own opinion and to have a space for communication. The interest of users measured in number of hits and traces of their activity grows when they are able to add their own content or to comment on existing commentaries [Ivanova et al, 2010/Euromed].

In the area of art images social networking sites can help extent the number of users consulting an image; for example the Library of Congress explained at the American Library Association annual conference in 2010 that the number of visitors consulting images which can be both accessed on the Library of Congress website and on flickr.com attracted higher number of visitors on Flickr. The user generated comments on Flickr also helped to improve the metadata records the Library maintained.

Not much time passed before the idea of "Web 3.0" appeared. Amit Agarwal suggests that Web 3.0 is about semantics (or the meaning of data), personalization (e.g. iGoogle), intelligent search and behavioural advertising among other things [Agarwal, 2009]. While Web 2.0 uses the Internet to make connections between people, Web 3.0 will use the Internet to make connections with information. The intelligent browsers will analyze the complex requests of the users made in natural language, will search the Internet for all possible answers, and then will organize the results. The adaptation to user specifics and aptitudes (personalisation) will be based on capturing the historical information thorough searching the Web. Many of the experts believe that the Web 3.0 browser will act like a personal assistant. The computer and the environment will become artificial subjects, which will pretend to communicate in real manner as real humans. Of course, the problems of applying rights policies in such a new atmosphere are crucial. However, addressing the rights is an additional issue which needs to be solved. A core problem in this domain continues to be finding appropriate combination of retrieval methods and techniques, which can lead to high quality image discovery. In the era of Web 3.0 bridging the semantic gap stands crucial.

[121] http://www.britannica.com/

3 The Process of Image Retrieval

Information retrieval is the science of searching for digital items, based both on their content and the metadata about them. Information retrieval can be done on different levels, from personal digital collections to world repositories in the WWW. It is interdisciplinary and attracts the interest of wide range of researchers and developers from a number of domains: computer science, mathematics, library science, information science, information architecture, cognitive psychology, linguistics, statistics, physics, etc. Image retrieval is part of it; it focuses on the processes of browsing, searching and retrieving images from large collections of digital images. There are two basic methods in image retrieval: text-based retrieval and content-based image retrieval (CBIR), which are used separately or together.

Traditional text-based indexing uses controlled vocabulary or natural language to document what an image is or what it is about. Newly developed content-based techniques rely on a pixel-level interpretation of the data content of the image. The upper stage of indexing techniques – concept-based indexing is based on mixing of simple text-based and content-based tools taking into account additional information for interconnections between perceived information from the main player of this process – "the user".

3.1 Text-Based Retrieval

Search systems based on textual information contain metadata about the images such as captioning, keywords, or descriptions of the images; the retrieval is performed over the annotated words. These methods are easily implemented using already existing technologies, but require manual data input for each image in the system. Manual image annotation is time-consuming, laborious and expensive and is a potential bottleneck because the speed of manual description and data entry is lower than the speed of digitisation. This is unpractical for the huge collections or automatically-generated images. Usually text-based descriptions are not considered accurate and precise and are often incomplete. Another problem with text annotation is that it often does not conform any defined vocabulary in a particular domain and may not describe the relations of the objects in the images – besides the subjectivity of judgements of different people who entry data. Another inconvenience comes also from the lack of universal solutions for dealing with synonyms in the language, and difference of the users' languages. Additionally, the increase in social web applications and the semantic web have inspired the development of several web-based image annotation tools as crowdsourcing solutions. In the frame of text-based retrieval we

can put also context-based technique, where retrieval is based on the analysis of free textual information, which became a context of the image [Hung et al, 2007].

The current efforts in the area of structuring information in digital repositories are focused mainly in two directions:

- Assistance in the processes of ordering and classifying the meta-information (such as Getty's AAT, ULAN, TGN, CONA). The use of these ontological structures in image retrieval processing leads to a decreasing metadata amount and expands the research scope utilizing defined interconnections between concepts;
- Development of metadata schemas and structures to classify image information (for instance Dublin Core, VRA Core, CIDOC CRM). They provide conceptual models intended to facilitate the integration, mediation and interchange of heterogeneous cultural heritage information.

As example for successful project aimed to use the advantages of such directions is developed by the team of Radoslav Pavlov "Bulgarian Iconographical Digital Library (BIDL)" [Pavlov et al, 2010]. A tree-based annotation model had been developed and implemented for the semantic description of the iconographical objects. It provides options for auto-completion, reuse of values, bilingual data entry, and automated media watermarking, resizing and conversing. A designated ontological model, describing the knowledge of East Christian Iconographical Art is implemented in BIDL; it assists in the annotation and semantic indexing of iconographical artefacts. The global vision of BIDL is based on a long-term observation of the end users preferences, cognitive goals and needs, aiming to offer an optimal functionality solution for the end users [Pavlova-Draganova et al, 2010].

- *Ontologies as a Form of Ordering and Classifying the Meta-information*

A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose. Every knowledge base, knowledge-based system, or knowledge-level agent is committed to some conceptualization, explicitly or implicitly. Ontologies are used in informatics as a form of knowledge representation about the world or some part of it [Gruber, 1993]. As Gruber said "Ontology is a formal, explicit specification of a shared conceptualization". The term is borrowed from philosophy, where an ontology is a systematic account of Existence. In computer science an ontology is a formal representation of the knowledge by a set of concepts within a domain and the relationships

between those concepts. It is used to reason about the properties of that domain, and may be used to describe the domain.

According theirs specificity the ontologies are: *Generic ontologies* (synonyms are "upper level" or "top-level" ontologies), in which defined concepts defined are considered to be generic across many fields; *Core ontologies*, where defined concepts are generic across a set of domains; and *Domain ontologies*, which express conceptualizations that are specific for a specific universe of discourse. The concepts in domain ontologies are often defined as specializations of concepts in the generic and core ontologies. The borderline between different kinds of ontologies is not clearly defined because core ontologies intend to be generic within a domain.

During the image processing ontologies on different levels, includes visual, abstract and application-level concepts, can be used. The using of ontological structures in image retrieval processing allows decreasing of metadata amount and expands the research scope, using defined interconnections between concepts, specified in used ontologies.

- *Metadata Schemas and Structures*

Resource Description Framework Schema (RDFS) is a family of specifications for the description of resources by setting the metadata for resources. RDF was developed by World Wide Web Consortium (W3C)[122] and adopted by Internet Society (ISOC)[123] as a standard for semantic annotation. RDFS is used as a basic method for conceptual description or modelling of information contained in web resources with different formats and syntax.

This mechanism for describing resources is a major component of current activities of the Semantic Web, in which automated software can store, exchange and use information disseminated thorough the Internet. It gives the opportunity to the users to operate more efficiently and safely with information. The ability to model heterogeneous abstract concepts through the RDF-model leads to increasing its applying for knowledge management of the events related to the activities of the Semantic Web.

The basic idea consists in using special expressions for describing content resources. Each expression describes the relationship "subject – predicate – object", which in RDF terminology is called a triplet. The identification of the subject, predicates and objects in RDF is made by Uniform Resource Identifier (URI). URI is a string, which unique identify a resource in the digital space: document, image, file, folder, mailbox, etc.

[122] http://www.w3.org/TR/rdf-schema/
[123] http://www.isoc.org/

The most popular examples of the URI are URL (Uniform Resource Locator) and URN (Uniform Resource Name). URL is a URI, which identify the resource and in parallel provides information about its location. URN is a URI, which identifies the resource in a specific namespace (i.e. in context). In order to avoid the limitations of using only a set of Latin characters and characters W3C and ISOC gradually imposed a new standard IRI (International Resource Identifier), which are free to use any Unicode-characters.

RDF-expressions are represented by labelled directed multi-graph. Such RDF-data model is naturally suited to certain types of knowledge representation to relational model and other ontological models traditionally used in computers today. However, in practice, RDF-data often continue to be stored in relational databases or through their own descriptors, called Triplestores or Quadstores. RDFS (Resource Description Framework Schema) and OWL (Web Ontology Language)[124] indicate the possibility of using RDF as a base to build additional ontology languages.

The most critical part is to define and collect some metadata that described the analyzed object. There exists a number of text-based indexing initiatives deal with the development of metadata schemas and structures to classify image information. We could mention for example *Dublin Core*[125], which is used primarily for retrieving resources on the web, *VRA Core*[126], which has elements to describe both an original work of art and its surrogate, *CIDOC CRM*[127] that gives conceptual reference model intended to facilitate the integration, mediation and interchange of heterogeneous cultural heritage information.

3.2 Content-Based Image Retrieval (CBIR)

Content-based image retrieval, as we see it today, is any technology that in principle helps to organize digital images based on their content. By this definition, anything ranging from an image similarity function to a robust image annotation engine falls into the range of CBIR. This characterization of CBIR as a field of study places it at a unique juncture within the scientific community. While we witness continued effort in solving the fundamental open problem of robust image understanding, we also see specialists from different fields, such as, computer vision, machine learning, information retrieval, human-computer interaction, database systems, Web and data mining, information theory, statistics,

[124] http://www.w3.org/TR/owl-features/
[125] http://dublincore.org/documents/dces/
[126] http://www.vraweb.org/organization/committees/datastandards/index.html
[127] http://www.cidoc-crm.org/

and psychology contributing and becoming part of the CBIR community [Wang et al, 2006].

John Eakins and Margaret Graham affirm that Content-Based Image Retrieval is a term first used in 1992 by Toshikazu Kato [Eakins and Graham, 1999], when explaining his experiments on automatic extraction of colour and shape of paintings stored in a database [Kato, 1992]. Since then the term was used to describe the process of image retrieval from large collections using features extracted from the content of the image based on their visual similarity with a query image or image features supplied by an end user.

Before designing and constructing CBIR system one very important step is selecting the domain where the system will be used. Different domains would be addressed by specific functional and non-functional requirements, which have to be covered by the system. During the years a wide spectrum of areas refers to CBIR system, such as medical diagnostic, geographical information and remote sensing systems, crime prevention, the military, intellectual property, photograph archives, architectural and engineering design, art collections, etc. From the point of view of the application area the images can represent different type of sensor-related data, projected or directly received in digital formats. The digital imagery includes colour and black-and-white photographs, infra-red photographs, video snapshots, radar screens, synthetic aperture radar formats, seismographs records, ultrasound, electrocardiographic, electroencephalographic, magnetic resonance and others.

Typically, a content-based image retrieval system consists of three components:

- Feature design;
- Indexing;
- Retrieval.

The *feature design* component extracts the visual feature(s) information from the images in the image database. The *indexing* component organizes the visual feature information to speed up the query or processing. The *retrieval engine* processes the user query and provides a user interface. During this process the central issue is to define a proper feature representation and similarity metrics. CBIR systems extract visual features from the images automatically. Similarities between two images are measured in terms of the differences between the corresponding features. To take into account the subjectivity of human perception and bridge the gap between the high-level concepts and the low-level features, relevance feedback is used as a means to enhance the retrieval performance.

All these steps are highly dependent of the domain where CBIR technology is applied. For instance, in the fields such as aerial image retrieval and medicine the goal is exactly defined, the searched objects in the images has homogeneous specifics, the received results usually do not need communication with the user to refine the queries. Absolutely different is the situation in the areas that are connected with the creative side of the human beings, such as art, architecture and design. The different kinds of users also stamp different requirements into specifics of CBIR systems.

Handling with digital copies of artworks has a wide spectrum of different directions and concern different types of users:

- *Museum workers:* Analysis of the artwork itself, its lifecycle, preservation and restoration are very important but heavy tasks where automatic image processing techniques have proved theirs usability during last decades;
- *Universal citizens:* Taking into account that artwork brings a specific authors' message to the viewer the computer should provide the ability to present history, context, and relevance in order to enrich education, enhance cross-cultural understanding, and sustain one's heritage and cultural diversity;
- *Computer scientists:* Except wide standard questions for serving the processes of image analysis and managing repositories, the grand challenge is determining image semantics and automatically verbalizing it.

4 The Gaps

One of the most felicitous analogies for presenting the existing semantic gap in area of Content-Based Image Retrieval can be found in "The Hitch-Hiker's Guide to Galaxy" by Douglas Adams. In this story, a group of hyper-intelligent pan-dimensional beings demand to learn the "Answer to Life, the Universe, and Everything" from the supercomputer Deep Thought, specially built for this purpose. It takes Deep Thought 7½ million years to compute and check the answer, which turns out to be "42". The efforts of covering the semantic gap in CBIR are turned to avoid these misunderstanding between human perceiving and the ways of communications and computer manner of low-level representations [Ivanova and Stanchev, 2009].

Search in the context of content-based retrieval analyzes the actual contents of the image. The term content might refer to colours, shapes, textures, or any other information that can be derived from the image itself. Acknowledging the need for providing image analysis at semantic

level, research efforts set focus on the automatic extraction of image descriptions matching human perception. The ultimate goal characterizing such efforts is to bridge the so-called semantic gap between *low-level visual features that can be automatically extracted from the visual content* and the *high-level concepts capturing the conveyed meaning* [Dasiapoulou et al, 2007]. The semantic gap is not a unique cause of difficulties in the process of information retrieval where issues can arise on the whole range starting from the primary object' complexity and ending with end-user subjectivity. Currently different gaps are being discussed in the research literature: sensory, semantic, abstraction, and subjective.

4.1 Sensory Gap

The sensory gap is "the gap between the object in the world and the information in a (computational) description derived from a recording of that scene" [Smeulders et al, 2000].

The sensory gap exists in the multimedia world as the gap between an object and the machine's capability to capture and define that object. Digitalization has a big challenge when applied to art-works and this is to develop techniques for creating digital objects, which allow capturing the paintings in good quality. Circumstances (such as the condition of the pictures, the lighting, the capabilities of used photo-cameras or scanners, the chosen resolution, etc.) play a major role in this process.

The sensory gap in this area inevitably results in the impossibility to present real sizes of the pictures or to present all pictures in one proportional scale. One can only see the proportion of the height and length. For instance Picasso's Guernica is 3.59 m x 7.76 m, while the miniatures of Isaac and Peter Oliver not exceed 2.5 cm x 2.5 cm. This sensory gap can be omitted only with additional metadata, taken from the camera or manually added to the picture.

The granularity of digitalized sources of artworks is in accordance with their usage. Figure 16 summarizes the connections between different kinds of users and the amount and quality of corresponding digital sources.

For the purposes of professional analysis in museums, special kinds of images, received from different photographic processes such as multi-colour banding, x-rays and infra-red imaging are used. For the purposes of professional printing industry very high definition and quality images are needed [Maitre et al, 2001]. The royalties and copyright restrictions from one side [Mattison, 2004], the necessity of high speed delivery on the Internet from the other side, and the limitations of visual devices

(monitors) from the third side, impose restrictions of the sizes and resolutions of digital images.

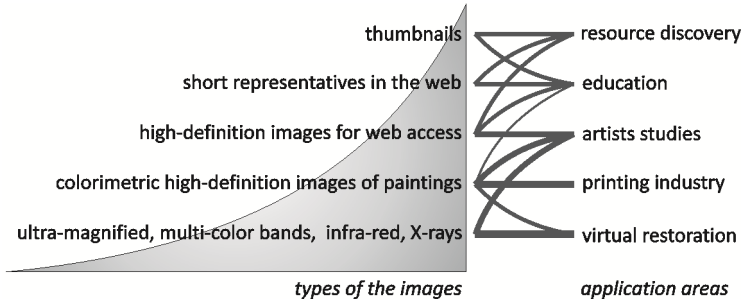

Figure 16. Digitized art images: quality and usage

Usually in the Internet space the presentations of the art paintings vary from:

- About 100x100 pixels for front presentation of the paintings as thumbnails;
- Image surrogates designed for presenting the painting on the screen, usually about 500 pixels by width or height, supplemented by additional text information, concerning the picture – author, sizes, techniques, locality, history of the creation, subject comment, etc.;
- High-definition images for web access, usually up to 1500 pixels by width or height;
- Finally, up to 4000 pixels by width or height; often watermarked items. The access to high-resolution and ultra-magnified images usually is defined by policies which set sets of use restrictions.

The fact that a digitized work of art is not the work itself but an image (instance) of this work, acquired at a certain time under specific conditions, makes semantic-based indexing and retrieval an absolute necessity in this area. For example, a query on "Mona Lisa" should retrieve all images of the painting regardless of their size, view angle, restoration procedures applied on the painting, etc. [Chen et al, 2005a].

4.2 Semantic Gap

The semantic gap is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data has for a user in a given situation [Smeulders et al, 2000]. The semantic gap is larger in visual arts images than in natural images since artworks are most often not perfectly realistic.

In simple terms, the semantic gap in content-based retrieval stems from the fact that multimedia data is captured by devices in a format, which are optimized for storage and very simple retrieval, and cannot be used to understand what the object "means". In addition to this, user queries are based on semantic similarity, but the computer usually processes similarity based on low-level feature similarity.

In early systems low-level representation was considered most reliable. Later, to bridge this gap, text annotations by humans were used in conjunction with low-level features of the objects. To extend the annotation list, different ontology systems have been used for further improving the results. It is applied generic ontologies as WordNet[128], specialized international standard structures as Dublin Core Element Set[129] and VRA Core Categories[130], or special ontologies designed for description of artefacts, such as IconClass[131] and Categories for the Description of Works of Art (CDWA)[132].

The annotations are used to group images into a certain number of concept bins, but the mapping from image space to concept space is not one-to-one, since it is possible to describe an image using a large number of words. The labelling of images is made not only for the whole image, but also for separate parts of the image [Enser et al, 2006].

The semantic gap is very critical to content-based multimedia retrieval techniques. As authors in [Smeulders et al, 2000] state: "The aim of content-based retrieval systems should be to provide maximum support in bridging the semantic gap between the simplicity of available visual features and the richness of the user semantics".

4.3 Abstraction Gap

The abstract aspects are specific to art images and differ from the semantic challenge. There are two major directions in this area, the first one addressing *cultural specifics* and the second one addressing *technical differences*.

Cultural abstraction relates to information inferred from cultural knowledge [Hurtut, 2010]. Artistic style analysis belongs to that category. According to the American Dictionary [Pickett et al, 2000] style is "the combination of distinctive features of artistic expression, execution or performance characterizing a particular person, group, school or era."

[128] http://wordnet.princeton.edu/
[129] http://www.dublincore.org/
[130] http://www.vraweb.org/vracore3.htm
[131] http://www.iconclass.nl
[132] http://www.getty.edu/research/conducting_research/standards/cdwa/

According to [Hurtut, 2010] style and semantic depiction share the same visual atomic primitives (lines, dots, surfaces, textures). Manual style recognition is a very difficult task, requiring the knowledge of numerous art historians and experts. An artwork from the blue period of Picasso for instance is recognizable not only because of its blue tonality, but because of many semantic features and iconographical cues.

Technical abstraction deals with questions about real artist of the artwork (artwork authentication), but can be focused into artistic praxis such as perspective rendering, pigment identification, searching for preliminary sketches in underlying layers, pentimenti, engraving tools, etc. These aspects can be analyzed using different imaging techniques such as X-rays, UV and infrared imaging.

4.4 Subjective Gap

The subjective gap exists due to users' aspirations and the descriptions of these aspirations. It may be difficult for a user to express what he wants from a multimedia retrieval system [Agrawal, 2009]. The subjective gap also exists due to the non-availability of any features which user wants to express. Some authors [Castelli and Bergman, 2002] identify this as an *intermediate level* of extraction from visual content, connected with emotional perceiving of the images, which usually is difficult to express in rational and textual terms.

The subjective gap is similar to the semantic gap; it refers to the lack of ability of the user to describe his needs (queries) to a retrieval system. To bridge this gap, instead of defining user's requirements at a very fine granularity level, higher level concepts can be used. The relevance feedback technique combined by neural network or fuzzy systems can bridge this subjective gap to some extent [Grosky et al, 2008].

In the recent years the term "*Emotional Semantic Image Retrieval*" enjoys growing popularity in scientific publications. The visual art is an area, where these features play significant role. Typical features in this level are colour contrasts, because one of the goals of the painting is to produce specific psychological effects in the observer, which are achieved with different arrangements of colours.

W. Bruce Croft introduced the concept "aesthetic gap" as "the lack of coincidence between the information that one can extract from low-level visual data and the interpretation of emotions that the visual data may arouse in a particular user in a given situation" [Croft, 1995]. Aesthetics is similar to quality as perceived by a viewer and is highly subjective. Modelling aesthetics of images will evolve in near future. The thesis of Rittendra Datta, presented in 2009 [Datta, 2009] is focused just to the

semantic and aesthetic inference for image search, using statistical learning approaches.

Emotional abstraction relates to emotional responses evoked by an image. These issues are addressed in the research domain called affective computing which enjoys widespread attention among computer scientists beyond those working on cultural heritage. In principal artworks by their nature are images that naturally evoke affective effects. Due to their implicit stylization, one does not look at artistic images with the same kind of attention and expectation as for natural images.

Many approaches try to bridge the gap between selected low-level features and several emotions expressed with pairs of words, e.g. warm-cool, action-relaxation, joy-uneasiness [Colombo et al, 1999]. In [Weining et al, 2006] is shown that emotional expression of the image is closely connected with such low-level characteristics as colour and luminance distributions, saturation and contrast information as well as edge sharpness in images.

5 User Interaction

Bridging the gaps is closely connected with user interaction. This is the place where the user and the system communicate.

The main focus in the creation of digital art resources has to be *user-centred* rather than *system-centred* since most of the issues around this content are related to making it accessible and usable for the real users [Dobreva and Chowdhury, 2010].

5.1 Complexity of the Queries

In the image retrieval systems, an important parameter to measure user-system interaction level is the complexity of queries supported by the system. The queries can use different modalities, such as:

- *Direct entry of values* of the desired features (query by percentage of properties). This method is not usually used in current systems because it is not particularly convenient for the users;
- *Image, also known as query by example:* Here, the user searches for images similar to a query image. Using an example image is perhaps the most popular way of querying a CBIR system in the absence of reliable metadata;
- *Graphics, or query by sketch:* This consists of a hand-drawn or computer-generated picture, where graphics are used as a query.
- *Keywords or Free-Text:* This is a search in which the user makes a query in the form of a word or group of words, selected from previously defined set or in free form. This is currently the most

popular way in web image search engines like Google, Yahoo! and Flickr. Usually this search is based on manually attached metadata or context driven information. Numerous current efforts are directed to finding the methods for automated labelling of the images – a challenge for the CBIR systems in present days;

- *Composite:* These are methods that combine one or more of the aforesaid modalities for querying a system. This also covers interactive querying such as the one implemented in relevance feedback systems.

Exploring user needs and behaviour is a basic and important phase of system development and is very informative when done as a front-end activity to system development. Currently users are mostly involved in usability studies when a set of digital resources has already been created and is being tested (for an overview on usability evaluation methods in the library domain see [George, 2008]). It would be really helpful to involve users on early stages of design and planning the functionality of the product which is being developed.

5.2 Relevance Feedback

Relevance feedback is a very important step in image retrieval, because it defines the goals and the means to achieve them. Relevance feedback provides a compromise between a fully automated, unsupervised system and one based on subjective user needs. It is a query modification technique which attempts to capture the user's precise needs through iterative feedback and query refinement. It can be thought of as an alternative search paradigm to other paradigms such as keyword-based search. In the absence of a reliable framework for modelling high-level image semantics and subjectivity of perception, the user's feedback provides a way to learn case-specific query semantics. A comprehensive review can be found in [Zhou and Huang, 2003] and [Crucianu et al, 2004]. The goal in relevance feedback is to optimise the amount of interaction with the user during a session. It is important to use all the available information to improve the retrieval results. Based on the user's relevance feedback, learning-based approaches are typically used to appropriately modify the feature set or similarity measure. In practice, learning instances are very small number. This circumstance has generated interest in novel machine-learning techniques to solve the problem, such as *one-class* learning, *active* learning, and *manifold* learning. Usually, classical relevance feedback consists of multiple rounds, which leads to loosing the patience in the user. Recent developments are directed to find techniques for minimizing the rounds. One decision is to use information of earlier user logs in the system. Another approach is

presented in [Yang et al, 2005], where a novel feedback solution for semantic retrieval is proposed: *semantic feedback*, which allows the system to interact with users directly at the semantic level. This approach is closely neighboured to the new relevance feedback paradigms aimed to help users by providing the user with cues and hints for more specific query formulation.

5.3 Multimodal Fusion

Multimodal fusion is linked to the integration of information in human-machine interactive systems where several communication modalities are proposed to the user. At recent years the advance in hardware and communication techniques made possible the using of advantages of multimedia. The presentations concerning some semantic unit stands more attractive and rich, using different modalities as images, text, free text, graphics, video, and any conceivable combination of them. Thus far, we have encountered a multitude of techniques for modelling and retrieving images, and text associated with these images. The trying for solving the retrieval tasks using only independent, media-specific methods is not good decision, because the information from the context (which can be extracted from neighbouring retrieval process over another modality) can be very helpful for current retrieval process. The user can best describe his queries only by a combination of media possibilities. Here lies the need for multimodal fusion as a technique for satisfying such user queries. Research in multimodal fusion therefore attempts to learn optimal combination strategies and models.

If we observe only image retrieval process, here multimodal fusion can be considered in case of different modalities of presenting the image for the purposes of image retrieval. For example, if we select the colour as a discriminative feature, several images may have same colour, but when combined with other modality such as texture, they can be classified to their respective categories with higher confidence. Each modality extracts certain aspect of an image and they are inter-dependent to each other. In the presence of many modalities, it is important to identify the best way to fuse them.

The fusion schemes can be based on whether we fuse the image data from different modalities first and then conduct experiments or we do other way round [Snoek et al, 2005].

- *Early fusion:* Fusion scheme that integrates unimodal features before learning concepts;
- *Late fusion:* Fusion scheme that first reduces unimodal features to separately learned concept scores, then these scores are integrated to learn concepts.

In early fusion, we need one-step learning phase only, whereas late fusion requires additional learning step. According to the architecture, fusion schemes can be grouped into three main categories:

- *Parallel architecture:* all the individual classifiers are invoked independently, and their results of each of them are combined. The results may be selected based on equal weight or they may be assigned different weight based on certain user selected criteria;
- *Serial combination:* individual classifiers are applied in a sequentially ordered in increasing order of theirs computation costs;
- *Hierarchical:* the individual classifiers are placed into a decision-tree like structure.

Fusion learning is an offline process while fusion application at real time is computationally inexpensive, which makes multimodal fusion very useful method for image retrieval.

6 Feature Design

The process of feature design is achieved to make mathematical description of an image for the retrieval purposes as its signature. Most CBIR systems perform feature design as a pre-processing step. Once obtained, visual features act as inputs to subsequent image analysis tasks, such as similarity estimation, concept detection, or annotation.

The process of feature design is achieved to make mathematical description of an image for the retrieval purposes, as its *signature*. The extraction of signatures and the calculation of image similarity cannot be cleanly separated. One the one hand, the formulation of signatures determines the necessity of finding new definitions of similarity measures. On the other hand, intuitions are often the early motivating factors for designing similarity measures in a certain way, which puts requirements on the construction of signatures. In terms of methodology development, a strong trend which has emerged in the recent years is the employment of statistical and machine learning techniques in various aspects of the CBIR technology. Automatic learning, mainly clustering and classification, is used to form either fixed or adaptive signatures, to tune similarity measures, and even to serve as the technical core of certain searching schemes, for example, relevance feedback. The fixed set of visual features may not work equally well to characterize different types of images. The signatures can be tuned either based on images alone (when some property does not characterize the image – than signatures vary according to the classification of images) or by learning from user feedback (when the user is not interested in a particular feature).

In contrast with early years, where global feature representations for images, such as colour histograms and global shape descriptors were used, currently the focus shifts towards using local features and descriptors, such as salient points, region-based features, spatial model features, and robust local shape characterizations.

6.1 Taxonomy of Art Image Content

Johannes Itten [Itten, 1961] has given very good formulation of messages that one artwork sends to the viewer. He points three basic directions of evincing colour aesthetics:

- Impression (visually);
- Expression (emotionally);
- Construction (symbolically).

These characteristics are mutually connected and cannot live of full value alone: symbolism without visual accuracy and without emotional force would be merely an anaemic formalism; visually impressive effect without symbolic verity and emotional power would be a banal imitative naturalism; emotional effect without constructive symbolic content or visual strength would be limited to the plane of sentimental expression. Each artist works according to his temperament, and must emphasize one or another of these aspects [Itten, 1961].

Different styles in art paintings are connected with used techniques from one side and aesthetic expression of the artist from other side. The process of forming artist style is a very complicated one, where current fashion painting styles, social background and personal character of the artist play significant role. All these factors lead to forming some common trends in art movements and some specific features, which distinguish one movement to another, one artist style to another, one artist period to another, etc. On the other hand the theme of the paintings also stamps specifics and can be taken into account. The compositions in different types of images (portraits, landscapes, town views, mythological and religious scenes, or everyday scenes) also set some rules, aesthetically imposed for some period.

When humans interpret images, they analyze image content. Computers are able to extract low-level image features like colour distribution, shapes and texture. Humans, on the other hand, have abilities that go beyond those of computers. The humans draw own subjective conclusions. They place emphasis on different parts of images, identify objects and scenes stamping theirs subjective vision and experience. The emotion that one person gets from seeing an image, and

therefore associates with it, may differ from another person's point of view.

Trying to define some useful grounds for bridging the gaps between interpreting the information from human and from computers several taxonomies of image content as extracted by the viewer of an image had been suggested.

Alejandro Jaimes and Shih-Fu Chang focus on two aspects of image content – the received visual *percepts* from the observed images and underlying abstract idea, which corresponds to *concepts*, connected with the image content [Jaimes and Chang, 2002].

In his brilliant survey for 2D artistic images analysis Tomas Hurtut [Hurtut, 2010] expands the taxonomy suggested by Bryan Burford, Pam Briggs and John Eakins [Burford et al, 2003]. He gives profiling of extraction primitives and concepts accounting the specific of artworks, splitting image categories into three groups: *image space*, *object space* and *abstract space*.

For the purposes of this study we adopted Hurtut's proposition adjusting the distribution of features in the groups. We examine *image space*, *semantic space* and *abstract space*:

- *Image space* contains visual features, needed to record an image through visual perception. Image space includes perceptual primitives (colour, textures, local edges), geometric features (strokes, contours, shapes) and design constructions (spatial arrangement, composition);
- *Semantic space* is related to the meaning of the elements, their potential for semantic interpretation. Semantic space consists of semantic units (objects), 3D relationship between them (scene, perspective, depth cues) and context (illumination, shadow);
- *Abstract aspects* that are specific to art images and reflect cultural influences, specific techniques as well as emotional responses evoked by an image form the abstract space.

Our vision on classifying feature percepts and concepts is presented on Fig. 17 (they slightly differ from Hurtuts' proposition) while pointing examples of used techniques for extracting visual primitives as well as some of closer relationships between concepts from defined spaces [Ivanova et al, 2010/MCIS]. All concepts are mutually connected – for instance emotional abstractions depends on specific expressive power of the artists (which is closely connected with visual perception primitives), thematic of the painting (concerning objective semantics of the paintings) as well as with the viewpoint of observer with his/her cultural and psychological peculiarities. Resolving these questions come up against the problem that in real-world contexts, it is in fact dynamic in nature. The

information that one can extract from the visual data for a one-time trained image recognition model does not change, but on the other hand, the interpretation that the same data have for a user in a given situation changes across users as well as situations.

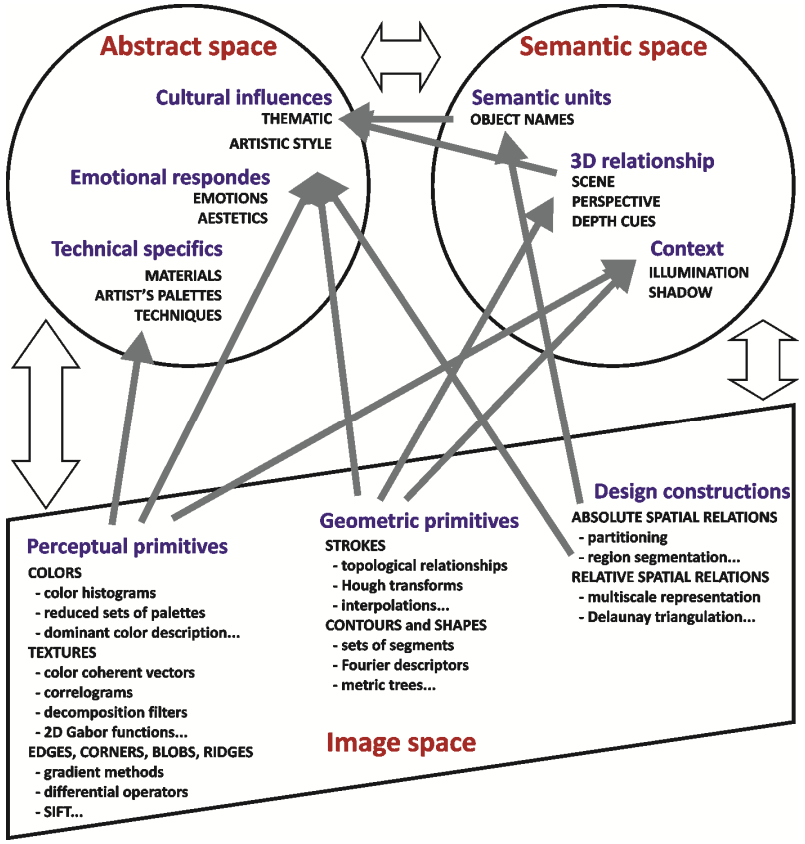

Figure 17. A taxonomy of art image content, inspired from [Burford et al, 2003] and [Hurtut, 2010]

6.2 Visual Features

According to Figure 17, main features, extracted from image space, are:

– *Perceptual features*, especially colour, texture and interesting points features;

- *Geometric features*, where the main focus for art image analysis is on contours and shapes as a source for further semantic interpretation and on strokes, which together with the previous ones are the source for extracting technical abstractions;
- *Design constructions*, connected with absolute or relative spatial relations.

Image features can be extracted at *a global level* to represent the entire image or the image can be split into parts and then features are computed *in a local level* from each part.

The most commonly used features include those reflecting *colour*, *texture*, *shape*, and *salient points* in an image. On a global level, features are computed to capture the overall characteristics of an image. The advantage is high speed for both processes: construction of signatures and computing similarity. The processing on the *local level* increases the robustness to spatial transformation of the images and gives more detailed representation of specific features of the image. Both approaches have their advantages: global features help to build an integral overview of the image as well as local ones can capture more detailed specifics.

6.2.1 Colour Features

Colour features are focused on the summarization of colours in an image. A set of colour descriptors are included in the MPEG-7 standard, which reflect different aspects of colour presence in an image. *Dominant Colour descriptor* presents the percentage of each quantized colour over the observed area. *Scalable Colour descriptor* builds a colour histogram, encoded by a Haar transformation. *Colour Layout descriptor* effectively represents the spatial distribution of colour in an image or in an arbitrary shaped region in a very compact form and is resolution-invariant. *Colour Structure descriptor* captures both colour content (similar to a colour histogram) and information about the structure of this content. Usually the exploration of colour features is attended with conversing colour representation to other colour spaces, which are more comprehensive for the human vision and in this way facilitate the choice of appropriate distance measures.

6.2.2 Texture Features

Texture features are intended to capture the granularity and repetitive patterns of surfaces within in a picture. Their role in domain-specific image retrieval is particularly vital due to their close relation to the underlying semantics. In image processing, a popular way to form texture features is by using the coefficients of a certain transformation on the original pixel values, or by statistics computed from such coefficients.

Such descriptors encode significant, general visual characteristics into standard numerical formats that can be used for various higher-level tasks. In many application areas, for example in aerial image retrieval and in medicine, thesauri for texture have been built. A thesaurus of brushwork terms concerning the annotation of paintings covering the period from Medieval to Modern art, which includes terms as "shading", "glazing", "mezzapasta", "grattage", "scumbling", "impasto", "pointillism", and "divisionism" had been proposed in the field of art image retrieval [Marchenko et al, 2007]. The brushwork is defined as a combination of colour presence and contrast features with texture features, such as directional ("impasto"), non-directional ("pointillism"), contrasting ("divisionism") and smooth ("mezzapasta"), or in case of spatial homogeneity they can be grouped into homogeneous ("mezzapasta" and "pointillism"), weakly homogeneous ("divisionism") and inhomogeneous ("scumbling", "shading" and "glazing").

6.2.3 Salient Point Features

Feature detection is a low-level image processing operation. That is, it is usually performed as the first operation on an image, and examines every pixel to see if there is a feature present at that pixel. If this is part of a larger algorithm, then the algorithm will typically only examine the image in the region of the features. There are very large number of feature detectors, which vary widely in the kinds of feature detected, the computational complexity and the repeatability. The salient point feature detectors can be divided into: *edges*, *corners*, *blobs* and *ridges* (with some overlap). A detailed review of salient point features is showed in [Ivanova et al-TR, 2010].

"*Edges*" are points where there is a boundary between two image regions. In practice, edges are usually defined as sets of points in the image which have a strong gradient magnitude. Furthermore, some common algorithms chain high gradient points together to form a more complete description of an edge. These algorithms usually place some constraints on the properties of an edge, such as shape, smoothness, and gradient value. Locally, edges have one dimensional structure.

The terms "*Corners*" or "*Interesting points*" refer to point-like features in an image, which have a local two dimensional structure. The name "corner" arose since early algorithms first performed edge detection, and then analyzed the edges to find rapid changes in direction (corners). These algorithms were then developed so that explicit edge detection was no longer required, for instance by looking for high levels of curvature in the image gradient. It was then noticed that the so-called corners were also being detected on parts of the image which were not corners in the

traditional sense (for instance a small bright spot on a dark background may be detected).

"*Blobs*" provide a complementary description of image structures in terms of regions, as opposed to corners that are more point-like. Blob descriptors often contain a preferred point (a local maximum of an operator response or a gravity centre), which means that many blob detectors may also be regarded as "point of interest" operators. Blob detectors can detect areas in an image which are too smooth to be detected by a corner detector.

The concept "*Ridges*" is a natural tool for elongated objects. A ridge descriptor computed from a grey-level image can be seen as a generalization of a medial axis. From a practical viewpoint, a ridge can be thought of as a one-dimensional curve that represents an axis of symmetry, and in addition has an attribute of local ridge width associated with each ridge point. It is algorithmically harder to extract ridge features from general classes of grey-level images than edge-, corner- or blob features. The ridge descriptors are frequently used for road location in aerial images and for extracting blood vessels in medical images.

Multiple scale-invariant features extraction algorithms such as SIFT, GLOH, SURF, LESH exist; they are widely used in current object recognition. They transform an image into a large collection of feature vectors, each of which is invariant to image translation, scaling, and rotation, partially invariant to illumination changes and robust to local geometric distortion (an overview is provided in [Ivanova et al-TR, 2010]).

6.2.4 Shape Features

Shape is a key attribute of segmented image regions, and its efficient and robust representation plays an important role in retrieval. Shape representations are closely connected with the particular forms of shape similarities used in each case. The current state of the art of this area is described in detail in [Data et al, 2008]. They have marked the shift from global shape representations which was dominant in early research to the use of more local descriptors in the last years. In MPEG-7 standard also are included Region Shape, Contour Shape and Shape 3D descriptors. The Region Shape descriptor utilizes a set of Angular Radial Transform coefficients. The Contour Shape descriptor is based on the Curvature Scale Space representation of the contour. The Shape 3D descriptor specifies an intrinsic shape description for 3D mesh models, which exploits some local attributes of the 3D surface [ISO/IEC 15938-3]. Shape features can play very significant role in semantic retrieval.

6.2.5 Spatial Relations

Representing spatial relations among local image entities plays a very important role in the process of preparing visual signatures. Several kinds of indexing methods are used for representing absolute or relative spatial relations.

- *Partitioning*

Partitioning can be defined as data-independent grouping [Data et al, 2008]. This method is not closely connected with representing absolute spatial relations, but allows a simple way for receiving more local information for the examined images. There are different methods for partitioning the image depending on the type of application. The simplest method is to divide image into non-overlapping tiles. In [Gong et al, 1996] the image is split into nine equal sub-images. [Striker and Dimai, 1997] have split image into oval central region and four corners. These methods have low computational cost and can be used for deriving more precisely (than from the whole image) low-level characteristics. They are not suitable if the goal is object segmentation.

- *Segmentation*

Segmentation is opposite of partitioning and is characterized as data-driven grouping. [Estrada, 2005] formulates segmentation as "the problem of defining a similarity measure between image elements that can be evaluated using image data, and the development of an algorithm that will group similar image elements into connected regions, according to some grouping criterion. The image elements can be pixels, small local neighbourhoods, or image regions produced by an earlier stage of processing, or by a previous step of an iterative segmentation procedure. The similarity function can use one or many of the available image cues (such as image intensity, colour, texture, and various filter responses), or be defined as a proximity measure on a suitable feature space that captures interesting image structure."

A great variety of segmentation techniques exists. Some applied approaches used either agglomerative (by merging) or divisive (by splitting) hierarchical clustering with different similarity functions (based on the entropy or statistic distance) and stopping criteria (like minimum description length, chi-square, etc.). Agglomerative algorithms are usually more frequently used than the divisive ones. An excellent example for agglomerative clustering is the algorithm where "normalized cut criterion" measures both the total dissimilarity between the different groups as well as the total similarity within the groups [Shi and Malik, 2000].

Other algorithms are not hierarchical. The simplest and widely used segmentation approach is based on *k*-means clustering. This basic approach enjoys a speed advantage, but is not as refined as some recently developed methods. Another disadvantage is that the number of clusters is an external parameter. The mean-shift algorithm is nonparametric clustering technique; it does not require prior knowledge of the number of clusters and the algorithm recursively moves to the kernel smoothed centroid for every data point looking for the point with highest density of data distribution [Comaniciu and Meer, 1999].

Amongst other approaches it is worth mentioning the multi-resolution segmentation of low-depth-of-field images [Wang et al, 2001], a Bayesian framework-based segmentation involving the Markov chain Monte Carlo technique [Tu and Zhu 2002], and the EM-algorithm-based segmentation using a Gaussian mixture model [Carson et al, 2002], forming blobs suitable for image querying and retrieval. A sequential segmentation approach that starts with texture features and refines segmentation using colour features is explored in [Chen et al, 2001]. An unsupervised approach for segmentation of images containing homogeneous colour/texture regions has been proposed in [Deng and Manjunath, 2001].

Yet another group of algorithms are the so-called model-based segmentation algorithms. The central assumption is that structures of interest have a repetitive form of geometry. These algorithms work well when the segmented image contains the search object and are widely used in medicine and radiological image retrieval.

- *Presentations of Relative Relationships*

Considering that homogeneous regions or symbolic objects have already been extracted, the relative relationships try to model or characterize the spatial relations between them, for instance "object A is under and on the left of object B" [Freeman, 1975].

Another convenient way of representing local spatial relations is Delaunay triangulation. This method was invented by Boris Delaunay in 1934 for the case of Euclidean space. In this space the Delaunay triangulation is the dual structure of the Voronoi diagram. Several algorithms can be used for computing Delaunay triangulation, such as flipping, incremental, gift wrap, divide and conquer, sweep-line, sweep-hull, etc. [de Berg et al, 2000].

6.3 MPEG-7 Standard

The Moving Picture Experts Group (MPEG) [ISO/IEC JTC1/SC29 WG11] was formed by the ISO in 1988 to set standards for audio and video compression and transmission[133]. A series of currently widely spread standards such as the standard for audio recording MP3 (MPEG-1 Layer 3, ISO/IEC 11172), the standards for transmission for over the air digital television, digital satellite TV services, digital cable television, DVD video and Blue-ray (MPEG-2, ISO/IEC 13818 and MPEG-4, ISO/IEC 14496) are outcomes of this group. In addition to the above standards, the group deals with different standards to describe content and environment. Here our interest is focused on the MPEG-7 standard ISO/IEC 15938, named "Multimedia Content Description Interface" [ISO/IEC 15938-3], which provides standardized core technologies allowing the description of audiovisual data content in multimedia environments. Audiovisual data content that has MPEG-7 descriptions associated with it may include: still pictures, graphics, 3D models, audio, speech, video, and composition information about how these elements are combined in a multimedia presentation (scenarios).

The MPEG-7 descriptions of visual content are separated into three groups:

- *Colour Descriptors:* Colour Space, Colour Quantization, Dominant Colours, Scalable Colour, Colour Layout, Colour Structure, and GoF/GoP Colour;
- *Texture Descriptors:* Homogeneous Texture, Edge Histogram, and Texture Browsing;
- *Shape Descriptors:* Region Shape, Contour Shape, and Shape 3D.

Colour descriptors used in MPEG-7 are as listed below:

- *Colour Space:* The feature colour space is used in other colour based descriptions. In the current version of the standard the following colour models are supported: RGB, YCrCb, HSV, HMMD, Linear transformation matrix with reference to RGB and Monochrome;
- *Colour Quantization:* This descriptor defines a uniform quantization of a colour space. The number of bins which the quantizer produces is configurable; this allows for great flexibility within a wide range of applications. For a meaningful application in the context of MPEG-7, this descriptor has to be combined with Dominant Colour descriptors, e.g. to express the meaning of the values of dominant colours;

[133] http://www.chiariglione.org/mpeg

- *Dominant Colour(s):* This colour descriptor is most suitable for representing local (object or image region) features where a small number of colours are enough to characterize the colour information in the region of interest. Whole images are also applicable, for example, flag images or colour trademark images. Colour quantization is used to extract a small number of representative colours in each region/image. Correspondingly, the percentage of each quantized colour in the region is calculated. A spatial coherency on the entire descriptor is also defined, and is used in similarity retrieval. The specific presentation of this descriptor allows for the variety of possibilities of using different kind of similarity measures. The Earth mover distance [Wang et al, 2003] is the most convenient for such kind of features. Other types of similarity measures are used in [Yang et al, 2008];
- *Scalable Colour:* The descriptor specifies a colour histogram in HSV colour space, which is encoded by a Haar transformation. Its binary representation is scalable in terms of bin numbers and bit representation accuracy over a broad range of data rates. The Scalable Colour descriptor is useful for image-to-image matching and retrieval based on colour feature. Retrieval accuracy increases with the number of bits used in the representation. The sum of absolute difference of coefficients can be used (L_1 metric) as a distance measure;
- *Colour Layout:* This descriptor effectively represents the spatial distribution of colour of visual signals in a very compact form. This compactness allows visual signal matching functionality with high retrieval efficiency at very small computational costs. It provides image-to-image matching without dependency on image format, resolutions, and bit-depths. It can be also applied both to a whole image and to any connected or unconnected parts of an image with arbitrary shapes. It also provides very friendly user interface using hand-written sketch queries since this descriptor captures the layout information of colour feature. The sketch queries are not supported in other colour descriptors. The colour Layout descriptor uses the YCbCr colour space with 8 bits quantization. The elements of colour Layout specify the integer arrays that hold a series of zigzag-scanned DCT coefficient values. The DCT coefficients of each colour component are derived from the corresponding component of local representative colours. For similarity measure can be used standard L_1 or L_2 metrics as well as specific functions, which takes into account the significance of the order of coefficients [Herrmann, 2002];

- *Colour Structure:* This is a colour feature descriptor that captures both colour content (similar to a colour histogram) and information about the structure of this content. Its main functionality is image-to-image matching and its intended use is for still-image retrieval, where an image may consist of either a single rectangular frame or arbitrarily shaped, possibly disconnected, regions. The extraction method embeds colour structure information into the descriptor by taking into account all colours in a structuring element of 8x8 pixels that slides over the image, instead of considering each pixel separately. Unlike the colour histogram, this descriptor can distinguish between two images in which a given colour is present in identical amounts but where the structure of the groups of pixels having that colour is different in both images. Colour values are represented in the double-coned HMMD colour space, which is quantized non-uniformly into 32, 64, 128 or 256 bins. Each bin amplitude value is represented by an 8-bit code. The Colour Structure descriptor provides additional functionality and improved similarity-based image retrieval performance for natural images compared to the ordinary colour histogram. The descriptor expresses local colour structure in an image by means of a structuring element that is composed of several image samples. The semantics of the descriptor, though related to those of a colour histogram, is distinguishable in the following way. Instead of characterizing the relative frequency of individual image samples with a particular colour, this descriptor characterizes the relative frequency of structuring elements that contain an image sample with a particular colour. Hence, unlike the colour histogram, this descriptor can distinguish between two images in which a given colour is present in identical amounts but where the structure of the groups of pixels having that colour is different in the two images. Usually the sum of absolute normalized difference of coefficients is used (L_1 metric) as a distance;
- *GoF/GoP Colour:* The Group of Frames/Group of Pictures Colour descriptor extends the Scalable Colour descriptor that is defined for a still image to colour description of a video segment or a collection of still images. The same similarity/distance measures that are used to compare Scalable Colour descriptions can be employed to compare GoF/GoP Colour descriptors.

From texture descriptors we will stop our attention on the following ones:

- *Edge Histogram:* This descriptor specifies the spatial distribution of five types of edges in local image regions (four directional edges – vertical, horizontal, 45 degree, 135 degree and one non-directional in

each local region called a sub-image. The sub-image is a part of the original image and each sub-image is defined by dividing the image space into 4x4 non-overlapping blocks, linearized by raster scan order. For each sub-image a local edge histogram with 5 bins is generated. As a result 16*5=80 histogram bins forms an Edge Histogram descriptor array. Each sub-image if divided into image-blocks. The value for each histogram bin in is related to the total number of image blocks with the corresponding edge type for each sub-image. These bin values are normalized by the total number of image blocks in the sub-image and are non-linearly quantized by quantization tables, defined in MPEG-7 standard. For this descriptor can be used each similarity measure function for histograms. [Won et al, 2002] suggests an extension to this descriptor in order to capture not only the local edge distribution information but also semi-global and global ones;

- *Homogeneous Texture:* This descriptor characterizes the region texture using the energy and energy deviation in a set of frequency channels. This is applicable for similarity based search and retrieval applications. The frequency space from which the texture features in the image are extracted is partitioned with equal angles of 30 degrees in the angular direction and with an octave division in the radial direction. As a result of applying of 2D Gabor function for feature channels and consequent quantization and coding average, standard deviation, energy and energy deviation are extracted.

The main issue with the MPEG-7 standard is that it focuses on the representation of descriptions and their encoding rather than on the practical methods on the extraction of descriptors. The creation and application of MPEG-7 descriptors are outside the scope of the MPEG-7 standard. For example, description schemes used in MPEG-7 specify complex structures and semantics groupings descriptors and other description schemes such as segments and regions which require a segmentation of visual data. MPEG-7 does not specify how to automatically segment still images and videos in regions and segments; likewise, it does not recommend how to segment objects at the semantic level [Tremeau et al, 2008].

MPEG-7 does not strictly standardize the distance functions to be used and sometimes does not propose a dissimilarity function leaving the developers the flexibility to implement their own dissimilarity/distance functions. A few techniques can be found in the MPEG-7 eXperimentation Model (XM) [MPEG-7:4062, 2001]. Apart from that, there are many general purpose distances that may be applied in order to simplify some complex distance function or even to improve the performance [Eidenberger, 2003]. A large number of successful distance measures

from different areas (statistics, psychology, medicine, social and economic sciences, etc.) can be applied on MPEG-7 data vectors [Dasiapoulou et al, 2007].

MPEG-7 is not aimed at any particular application. The elements that MPEG-7 standardizes support as broad a range of applications as possible. The MPEG-7 descriptors are often used in the processes of image-to-image matching, searching of similarities, sketch queries, etc. [Stanchev et al, 2006].

7 Data Reduction

Data reduction techniques can be applied to obtain a reduced representation of the data set that is much smaller in volume, yet closely maintains the integrity of the original data.

That is, mining on the reduced data set should be more efficient yet produce the same (or almost the same) analytical results. Strategies for data reduction include *dimensionality reduction*, where encoding mechanisms are used to reduce the data set size and *numerosity reduction*, where the data are replaced or estimated by alternative, smaller data representations. In Figure 18 a hierarchy of some data reduction techniques is presented.

7.1 Dimensionality Reduction

The "curse of dimensionality", is a term coined by Bellman [Bellman, 1961] to describe the problem caused by the exponential increase in volume associated with adding extra dimensions to a feature space. In image clustering and retrieval applications, the feature vectors tend to use high dimensional data space and in such case to fall into "curse of dimensionality" since the search space grows exponentially with the dimensions. In image databases, the volume of the data is very large and the amount of time needed to access the feature vectors on storage devices usually dominates the time needed for a search. This problem is further complicated when the search is to be performed multiple times and in an interactive environment. Thus high dimensionality of data causes increased time and space complexity, and as a result decreases performance in searching, clustering, and indexing.

When the attribute space is not high-dimensional, the standard method is representing the features as points in a feature space and using distance metrics for similarity search. The problem with this method is that with the increasing of data dimension, the maximum and minimum distances to a given query point in the high dimensional space are almost the same under a wide range of distance metrics and data distributions.

All points converge to the same distance from the query point in high dimensions, and the concept of nearest neighbours become meaningless.

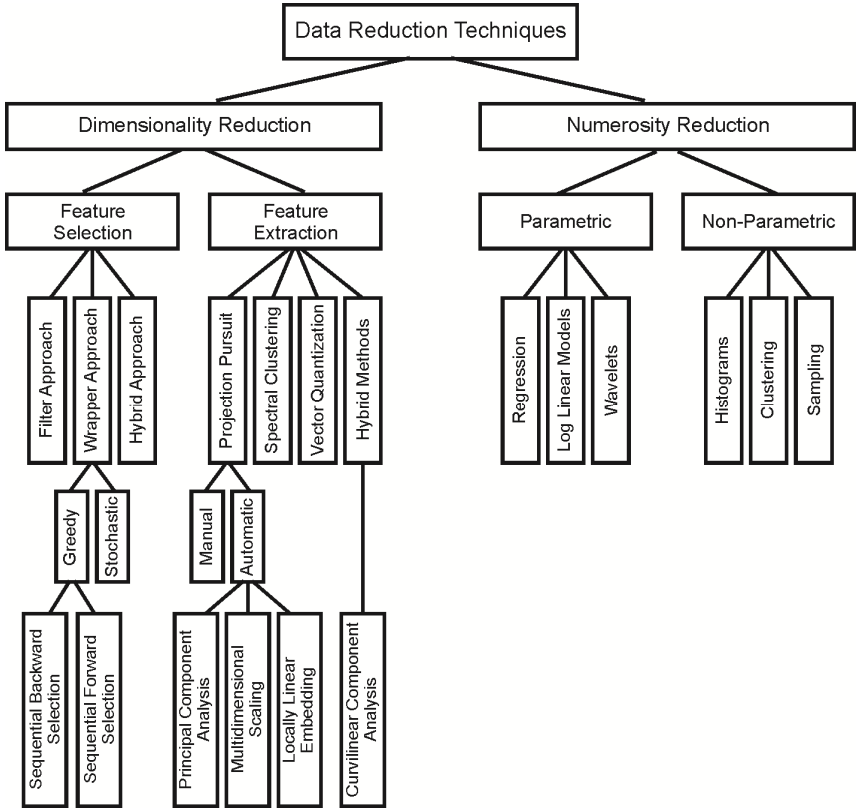

Figure 18. Data Reduction Techniques

There are two main ways to overcome the curse of dimensionality in image search and retrieval. The first is to search the approximate results of a multimedia query, and the second is to reduce the high dimensional input data to a low dimensional representation.

The dimensionality reduction techniques are based on either *feature selection* (also named *attribute subset selection*) or *feature extraction* methods.

7.1.1 Feature Selection (Attributive Subset Selection)

In feature selection, an appropriate subset of the original features is found to represent the data. This method is useful when the data available limited amount, but is represented with a large number of features [Agrawal, 2009]. It is crucial to determine a small set of relevant variables to estimate reliable parameter. The advantage of selecting a small set of features is that you need to use few values in the calculations.

Data sets for analysis may contain attributes, which may be irrelevant or redundant to the mining task. Leaving out relevant attributes or keeping irrelevant attributes may aggravate data mining process. Attribute subset selection reduces the data set size by removing irrelevant or redundant attributes (or dimensions). The goal of attribute subset selection is to find a minimum set of attributes such that the resulting probability distribution of the data classes is as close as possible to the original distribution obtained using all attributes. Finding on optimal subset is a hard computational process. Therefore, heuristic methods that explore a reduced search space are commonly used for attribute subset selection. Optimal feature subset selection techniques can be divided to: filter, wrapper and hybrid [Gheyas and Smith, 2010].

- *Filter Approaches*

In filter approaches, features are scored and ranked based on certain statistical criteria and the features with highest ranking values are selected. Usually as filter methods t-test, chi-square test, Wilcoxon Mann-Whitney test, mutual information, Pearson correlation coefficients and principal component analysis are used. Filter methods are fast but lack robustness against interactions among features and feature redundancy. In addition, it is not clear how to determine the cut-off point for rankings to select only truly important features and exclude noise.

- *Wrapper Approaches*

In the wrapper approaches, feature selection is "wrapped" in a learning algorithm. The learning algorithm is applied to subsets of features and tested on a hold-out set, and prediction accuracy is used to determine the feature set quality. Generally, wrapper methods are more effective than filter methods. Since exhaustive search is not computationally feasible, wrapper methods must employ a designated algorithm to search for an optimal subset of features. Wrapper methods can broadly be classified into two categories based on the search strategy: (1) greedy and (2) randomized/stochastic.

(1) *Greedy wrapper methods* use less computer time than other wrapper approaches. Two most commonly applied wrapper methods that use a greedy hill-climbing search strategy are:

- Sequential backward selection, in which features are sequentially removed from a full candidate set until the removal of further features increase the criterion;
- Sequential forward selection, in which features are sequentially added to an empty candidate set until the addition of further features does not decrease the criterion.

The problem with sequentially adding or removing features is that the utility of an individual feature is often not apparent on its own, but only in combinations including just the right other features.

(2) *Stochastic algorithms*, developed for solving large scale combinatorial problems such as ant colony optimization, genetic algorithm, particle swarm optimization and simulated annealing are used as feature subset selection approaches. These algorithms efficiently capture feature redundancy and interaction, but are computationally expensive.

- *Hybrid Approaches*

The idea behind the hybrid method is that filter methods are first applied to select a feature pool and then the wrapper method is applied to find the optimal subset of features from the selected feature pool. This makes feature selection faster since the filter method rapidly reduces the effective number of features under consideration [Gheyas and Smith, 2010].

7.1.2 Feature Extraction

In feature extraction, new features are found using the original features without losing any important information. Feature extraction methods can be divided into linear and non-linear techniques, depending of the choice of objective function. Some of the most popular dimensionality reduction techniques are:

- *Projection Pursuit (PP)*

Projection pursuit is a method, which finds the most "interesting" possible projections of multidimensional data. A good review of projection pursuit can be found in [Huber, 1985]. The projection index defines the "interestingness" of a direction. The task is to optimize this index. A projection is considered interesting if it has a structure in the form of trends, clusters, hyper-surfaces, or anomalies. These structures can be

analyzed using manual or automatic methods. The scatter-plot is one such manual method, which can be used to understand data characteristics over two selected dimensions at a time. There are many methods to automate this task.

- *Principal Component Analysis (PCA)*

One often used and simple projection pursuit method is the Principal Component Analysis, which calculates the eigenvalues and eigenvectors of the covariance or correlation matrix, and projects the data orthogonally into space spanned by the eigenvectors belonging to the largest eigenvalues. PCA is also called the discrete Karhunen-Loève method (K-L method), the Hotelling transform, singular value decomposition (SVD), or empirical orthogonal function (EOF) method. A good tutorial on PCA can be found in [Smith, 2002]. PCA searches for k n-dimensional orthogonal vectors that can best be used to represent the data, where $k \leq n$. The original data are thus projected onto a much smaller space, resulting in dimensionality reduction. PCA transforms the data to a new coordinate system such that the first coordinate (also called the first principal component) is the projection of the data exhibiting the greatest variance, the second coordinate (also called the second principal component) exhibits the second greatest variance, and so on. In this way, the "most important" aspects of the data are retained in the lower-order principal components.

PCA is computationally inexpensive, can be applied to ordered and unordered attributes, and can handle sparse data and skewed data. Principal components may be used as inputs to multiple regression and cluster analysis.

- *Multidimensional Scaling (MDS)*

Multidimensional scaling (MDS) is used to analyze subjective evaluations of pairwise similarities of entities. In general, the goal of the analysis is to detect meaningful underlying dimensions that allow the researcher to explain observed similarities or dissimilarities (distances) between the investigated objects. In PCA, the similarities between objects are expressed in the covariance or correlation matrix. MDS allows analyzing any kind of similarity or dissimilarity matrix, in addition to correlation matrices.

Assume, there are p items in n-dimensional space and a $p \times p$ matrix of proximity measures, MDS produces a k-dimensional representation ($k \leq n$) of the original data items. The distance in the new

k -space reflects the proximities in the data. If two items are more similar, this distance will be smaller. The distance measures can be Euclidean distance, Manhattan distance, maximum norm, or other. MDS is typically used to visualize data in two or three dimensions, to uncover underlying hidden structure.

Any dataset can be perfectly represented using $n-1$ dimensions, where n is the number of items scaled. As the number of dimensions used goes down, the stress must either come up or stay the same. When the dimensionality is insufficient the non-zero stress values occur. It means that chosen dimension k cannot perfectly represent the input data.

Of course, it is not necessary that an MDS map has zero stress in order to be useful. A certain amount of distortion is tolerable. Different people have different standards regarding the amount of stress to tolerate. The rule of thumb is that anything under 0.1 is excellent and anything over 0.15 is unacceptable.

Both PCA and MDS are eigenvector methods designed to model linear variability in high dimensional data. In PCA, one computes the linear projections of greatest variance from the top eigenvectors of the data covariance matrix. Classical MDS computes the low dimensional embedding that best preserves pair-wise distances between data points. If these distances correspond to Euclidean distances, the results of metric MDS are equivalent to PCA.

- *Locally Linear Embedding (LLE)*

Locally Linear Embedding (LLE) is also an eigenvector method that computes low dimensional, neighbourhood preserving embeddings of high dimensional data. LLE attempts to discover nonlinear structure in high dimensional data by exploiting the local symmetries of linear reconstructions. Notably, LLE maps its inputs into a single global coordinate system of lower dimensionality, and its optimizations – though capable of generating highly nonlinear embeddings – do not involve local minima. Like PCA and MDS, LLE is simple to implement, and its optimizations do not involve local minima. At the same time, it is capable of generating highly nonlinear embeddings [Saul and Roweis, 2000].

- *Spectral Clustering*

The main tools for spectral clustering are graph Laplacian matrices. The technique is based on two main steps: first embedding the data points in a space in which clusters are more "obvious" (using the

eigenvectors of a Gram matrix) and then applying an algorithm to separate the clusters, such as K-means. A good tutorial for Spectral Clustering can be found in [Luxburg, 2006]. Sometimes called Diffusion Maps or Laplacian Eigenmaps, these manifold-learning techniques are based on graph-theoretic approach. A graph is built using the data items, which incorporates neighbourhood information of the data set. A low dimensional representation of the data set is computed using the Laplacian of the graph that optimally preserves local neighbourhood information. The vertices, or nodes, represent the data points, and the edges connected the vertices, represent the similarities between adjacent nodes. After representing the graph with a matrix, the spectral properties of this matrix are used to embed the data points into a lower dimensional space, and gain insight into the geometry of the dataset. Though these methods perform exceptionally well with clean, well-sampled data, problems arise with the addition of noise, or when multiple sub-manifolds exist in the data.

- *Vector Quantization (VQ)*

Vector Quantization (VQ) is used to represent not individual values but (usually small) arrays of them. In vector quantization, the basic idea is to replace the values from a multidimensional vector space with values from a lower dimensional discrete subspace. A vector quantizer maps k *-dimensional* vectors in the vector space R^k into a finite set of vectors $Y = \{y_i : i = 1,...,n\}$. The vector y_i *is* called a *code vector* or a *codeword* and the set of all the codewords Y is called a *codebook*. Unfortunately, designing a codebook that best represents the set of input vectors is NP-hard. There are different algorithms, which try to overcome this problem. A review of vector quantization techniques used for encoding digital images is presented in [Nasrabadi and King, 1988]. VQ can be used for any large data sets, when adjacent data values are related in some way. VQ has been used in image, video, and audio compression.

- *Curvilinear Component Analysis (CCA)*

The principle of Curvilinear Component Analysis is a self-organized neural network performing two tasks: vector quantization (VQ) of the sub-manifold in the data set (input space); and nonlinear projection (P) of these quantizing vectors toward an output space, providing a revealing unfolding of the sub-manifold. After learning, the network has the ability to continuously map any new point from one space into another: forward

mapping of new points in the input space, or backward mapping of an arbitrary position in the output space [Demartines and Herault, 1997].

7.2 Numerosity Reduction

In numerosity reduction data is replaced or estimated by alternative, smaller data representations. These techniques may be parametric or nonparametric. For *parametric methods*, a model is used to estimate the data, so that typically only the data parameters need to be stored, instead of the actual data. For comprehensive data representation outliers may also be stored. Regression and Log-linear models, which estimate discrete multidimensional probability distributions, are two examples. *Nonparametric methods* for storing reduced representations of the data include histograms, clustering, and sampling. Data discretization is a form of numerosity reduction that is very useful for the automatic generation of concept hierarchies. Discretization and concept hierarchy generation are powerful tools for data mining, in that they allow the mining of data at multiple levels of abstraction.

- *Regression and Log-Linear Models*

Regression and Log-Linear models can be used to approximate the given data. They are typical examples of parametric methods [Han and Kamber, 2006].

In *simple linear regression*, the data are modelled to fit a straight line. A random variable, y (called a *response variable*), can be modelled as a linear function of another random variable, x (called a *predictor variable*), with the equation $y = wx + b$, where the variance of y is assumed to be constant. In the context of data mining, x and y are both numerical attributes. The coefficients, w and b (called *regression coefficients*), specify the slope of the line and the y -intercept, respectively. These coefficients can be solved by the *method of least squares*, which minimizes the error between the actual line separating the data and the estimate of the line.

Multiple linear regression is an extension of simple linear regression, which allows a response variable y to be modelled as a linear function of two or more predictor variables.

Log-linear models approximate discrete multidimensional probability distributions using logarithmic transformations. Given a set of tuples in n dimensions (e.g., described by n attributes), we can consider each tuple as a point in a n -dimensional space. Log-linear models can be used to

estimate the probability of each point in a multidimensional space for a set of discretized attributes, based on a smaller subset of dimensional combinations. This allows a higher-dimensional data space to be constructed from lower dimensional spaces. Log-linear models are therefore also useful for dimensionality reduction (since the lower-dimensional points together typically occupy less space than the original data points) and data smoothing (since aggregate estimates in the lower-dimensional space are less subject to sampling variations than the estimates in the higher-dimensional space).

Regression and log-linear models can both be used on sparse data, although their application may be limited. While both methods can handle skewed data, regression does it exceptionally well. Regression can be computationally intensive when applied to high dimensional data, whereas log-linear models show good scalability for up to 10 or so dimensions.

- *Discrete Wavelet Transforms*

The Discrete Wavelet Transform (DWT) is a linear signal processing technique that transforms input vector to another vector with same length, but elements are wavelet coefficients. A wavelet is a mathematical function used to divide a given function into different scale components. A wavelet transform is the representation of a function by wavelets. The wavelets are scaled and translated copies (known as "daughter wavelets") of a finite-length or fast-decaying oscillating waveform (known as the "mother wavelet"). The first DWT was invented by the Hungarian mathematician Alfred Haar in 1909. The most commonly used set of discrete wavelet transforms was formulated by the Belgian mathematician Ingrid Daubechies in 1988. Haar wavelet is the first one of the family of Daubechies wavelets. The Daubechies wavelets are a family of orthogonal wavelets defining a discrete wavelet transform and characterized by a maximal number of vanishing moments for some given support. With each wavelet type of this class, there is a scaling function (also called father wavelet) which generates an orthogonal multi-resolution analysis [Daubechies, 1988].

- *Histograms*

Histograms use binning to approximate data distributions and are a popular form of data reduction [Han and Kamber, 2006]. A histogram for an attribute partitions the data distribution of the attribute into disjoint subsets, or *buckets*. There are several partitioning rules, including *Equal-width* (where the width of each bucket range is uniform), *Equal-frequency*

(each bucket contains roughly the same number of contiguous data samples), *V-Optimal* (the histogram with the least variance) and *MaxDiff* (where a bucket boundary is established between each pair for pairs having the $b-1$ largest differences, where b is user-specified number of buckets). V-Optimal and MaxDiff histograms tend to be the most accurate and practical. Histograms are highly effective at approximating both sparse and dense data, as well as highly skewed and uniform data.

- *Clustering*

In data reduction, the cluster representations of the data are used to replace the actual data. The effectiveness of this technique depends on the nature of the data. It is much more effective for data that can be organized into distinct clusters than for smeared data. In database systems, multidimensional index trees are primarily used for providing fast data access. They can also be used for hierarchical data reduction, providing a multi-resolution clustering of the data. This can be used to provide approximate answers to queries. An index tree can store aggregate and detail data at varying levels of abstraction. It provides a hierarchy of clustering of the data set, where each cluster has a label that holds for the data contained in the cluster. If we consider each child of a parent node as a bucket, then an index tree can be considered as a *hierarchical histogram*. The use of multidimensional index trees as a form of data reduction relies on an ordering of the attribute values in each dimension. Multidimensional index trees include R-trees, quad-trees, and their variations.

Special cases of clustering are data discretization techniques, which can be used to reduce the number of values for a given continuous attribute by dividing the range of the attribute into intervals. Interval labels can then be used to replace actual data values. Replacing numerous values of a continuous attribute by a small number of interval labels thereby reduces and simplifies the original data. From point of view of using class information in the discretization process the methods are *supervised* or *unsupervised*. Supervised discretizators usually fall into the following categories: if the process starts by finding one or a few points (called *split points* or *cut points*) to split the entire attribute range, and then repeats this recursively on the resulting intervals, it is called *top-down discretization* or *splitting*. In contrast, *bottom-up discretization* or *merging* starts by considering all continuous values as potential split-points, removes some by merging neighbourhood values to form intervals, and then recursively applies this process to the resulting intervals. Discretization can be performed recursively on an attribute to provide a hierarchical or multi-resolution partitioning of the attribute

values, known as a concept hierarchy. Concept hierarchies are useful for mining at multiple levels of abstraction. We have made a brief overview of discretization techniques in [Mitov et al, 2009b].

- *Sampling*

Sampling allows a large data set to be represented by a much smaller random sample (or subset) of the data. The most common ways for receiving data reduction, using sampling, according to [Han and Kamber, 2006] are:

- *Simple random sample without replacement* (SRSWOR), where all tuples are equally likely to be sampled;
- *Simple random sample with replacement* (SRSWR), where after a tuple is drawn, it is placed back into the primary set, so that it may be drawn again;
- *Cluster sample*, where all tuples are grouped into mutually disjoint "clusters", then a simple random sample can be obtained;
- *Stratified sample*, where if source set is divided into mutually disjoint parts called strata, a stratified sample is generated by obtaining a simple random sampling at each stratum. This helps ensure a representative sample, especially when the data are skewed.

An advantage of sampling for data reduction is that the cost of obtaining a sample is proportional to the size of the sample when applied to data reduction; sampling is most commonly used to estimate the answer to an aggregate query.

8 Indexing

The second component of CBIR-systems is indexing. Efficient indexing is critical for building and functioning of very large text-based databases and search engines. Research on efficient ways to index images by content has been largely overshadowed by research on efficient visual representation and similarity measures.

In [Markov et al, 2008] we provide an expanded survey of different spatial access methods, based on the earlier analyses of [Ooi et al, 1993] and [Gaede and Günther, 1998]. The access methods are classified in several categories: one-dimensional; multidimensional spatial; metric; and high dimensional access methods (Figure 19). The article [Markov et al, 2008] includes more detailed description of interconnections between access methods as well as the references of sources, where these methods are described. Here we provide a summary of the methods.

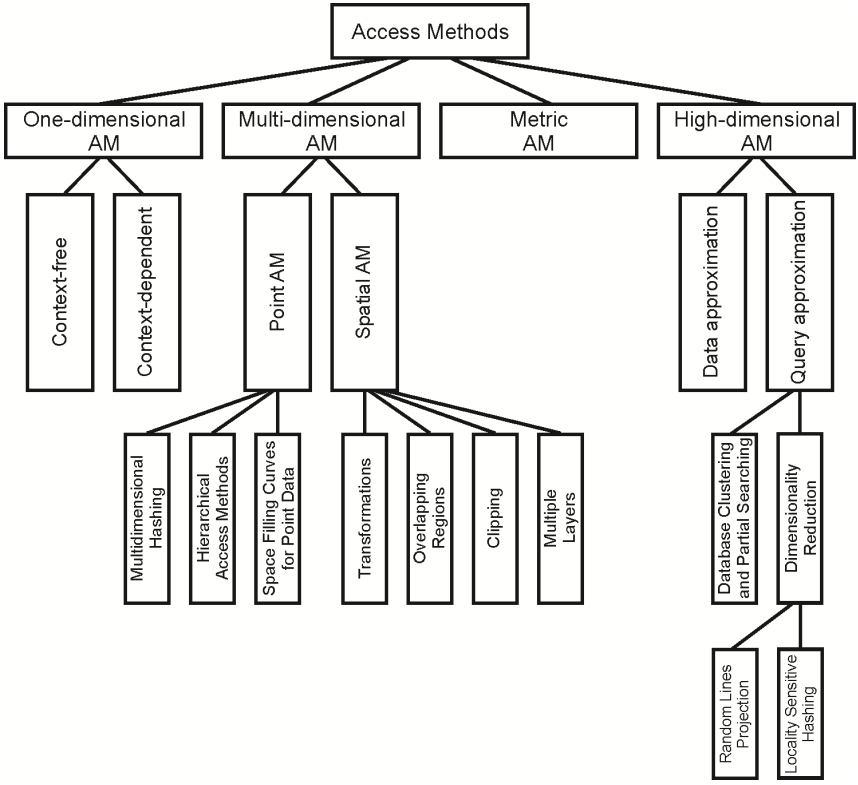

Figure 19. Taxonomy of the Access Methods

Multidimensional Spatial Access Methods are developed to serve information about spatial objects, approximated with points, segments, polygons, polyhedrons, etc. From the point of view of the spatial databases can be split in two main classes of access methods – Point Access Methods and Spatial Access Methods [Gaede and Günther, 1998].

Point Access Methods are used for organizing multidimensional point objects. Typical instance are traditional records, where one dimension corresponds to every attribute of the relation. These methods can be clustered in three basic groups: (1) Multidimensional Hashing; (2) Hierarchical Access Methods; (3) Space Filling Curves for Point Data.

Spatial Access Methods are used for work with objects which have arbitrary form. The main idea of the spatial indexing of non-point objects is using of the approximation of the geometry of the examined objects to more simple forms. The most commonly used approximation is Minimum

Bounding Rectangle (MBR), i.e. minimal rectangle, which sides are parallel of the coordinate axes and completely include the object. Approaches exist for approximation with Minimum Bounding Spheres or other polytopes, as well as their combinations. The usual problem when one operates with spatial objects is their overlapping. There are different techniques to avoid this problem. From the point of view of the techniques for organization of the spatial objects Spatial Access Methods form four main groups: (1) *Transformation* – this technique uses transformation of spatial objects to points in the space with more or less dimensions. Most of them spread out the space using space filling curves and then use some of point access method upon the transformed data set; (2) *Overlapping Regions* – here the data set are separated in groups; different groups can occupy the same part of the space, but every space object associates with only one of the groups. The access methods of this category operate with data in their primary space (without any transformations) eventually in overlapping segments; (3) *Clipping* – this technique use eventually clipping of one object to several sub-objects. The main goal is to escape overlapping regions. But this advantage can lead tearing of the objects, extending of the resource expenses and decreasing of the productivity of the method; (4) *Multiple Layers* – this technique is a variant of the technique of Overlapping Regions, because the regions from different layers can also overlap. However, there are some important differences: first, the layers are organizing hierarchically; second, every layer splits the primary space in different way; third, the regions of one layer never overlap; fourth, the data regions are separated from space extensions of the objects.

Metric Access Methods deal with relative distances of data points to chosen points, named anchor points, vantage points or pivots [Moënne-Loccoz, 2005]. These methods are designed to limit the number of distance computation, calculating first distances to anchors, and then finding the point searched for in the narrowed region. These methods are preferred when the distance is highly computational, as e.g. for the dynamic time warping distance between time series. Metric Access Methods are employed to accelerate the processing of similarity queries, such as the range and the k-nearest neighbour queries too [Chavez et al, 2001].

High Dimensional Access Methods are created to overcome the bottleneck problem, which appears with increasing of dimensionality. These methods are based on the *data approximation* and *query approximation* in sequential scan. For *query approximation* two strategies can be used: (1) examine only a part of the database, which is more probably to contain a resulting set – as a rule these methods are based

on the clustering of the database; (2) splitting the database to several spaces with fewer dimensions and searching in each of them, using Random Lines Projection or Locality Sensitive Hashing.

9 Retrieval Process

The third component of CBIR systems is served by retrieval engines. The retrieval engines build the bridge between the internal space of the system and the user making requests which need to be satisfied. Looking at the system side, the design of these engines is closely connected with the chosen feature representation and indexing schemes as well as the selected similarity metrics.

From the user point of view, in order to take into account the subjectivity of human perception and bridge the gap between the high-level concepts and the low-level features, relevance feedback has been proposed to enhance the retrieval performance. The other direction for facilitating that process is examining the image retrieval in more general frame of multimedia retrieval process, where content-based, model-based, and text-based searching can be combined.

9.1 Similarity

In the process of image retrieval, choosing the features as well as indexing the database are closely connected with the used similarity measures for establishing nearness between queries and images or between the images in a given digital resource in the processes of categorization.

The concept of similarity is very complex and is almost a whole scientific area itself. Similarity measures are aimed to give answer how much one object is close to another one. In the process of obtaining the similarity between two images several processes of finding similarity on different levels and data types need to be resolved. Image signature is a weighted set of feature vectors. In the case when a region-based signature is represented as such set of vectors, each of them would be represented in this way. A natural approach to defining a region-based similarity measure is to match every two corresponded vectors and then to combine the distances between these vectors as a distance between sets of vectors. For every level different similarities may be used. In Figure 20, similarity measures, used in CBIR for different features types, is presented.

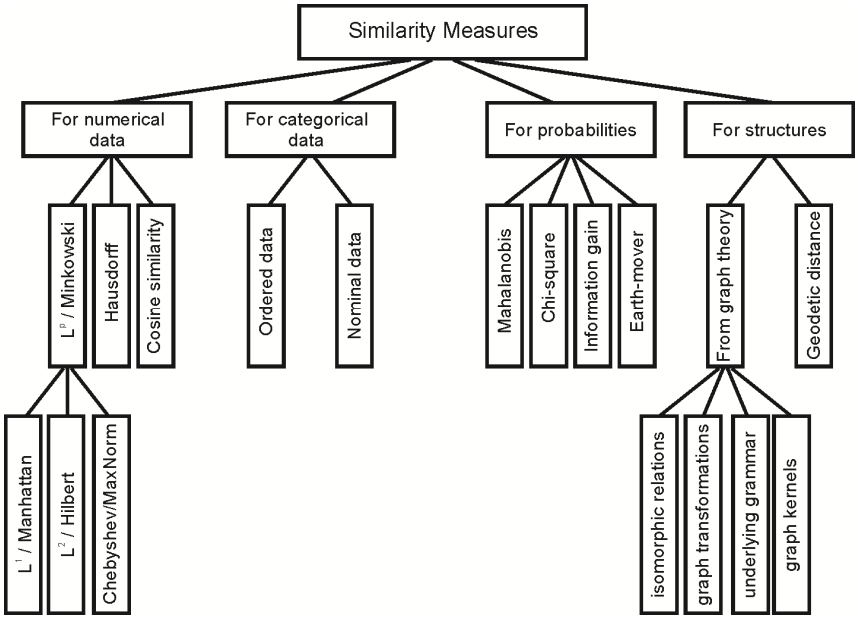

Figure 20. Different kinds of similarity measures

9.1.1 Distance-Based Similarity Measures

The most popular similarity measures are distance measures. They can be applied in each area which meets the conditions to be a metric space for equal self-similarity, minimality, symmetry and triangle inequality[134]. In mathematical notation the distance $d(X,Y)$ between two vectors X and Y is a function for which $d(X,Y) \geq 0$; if $d(X,Y) = 0$ then $X = Y$; $d(X,Y) = d(Y,X)$. Fulfilment of triangle inequality $d(X,Y) \leq d(X,Z) + d(Z,Y)$ defines distance $d(X,Y)$ as a metric. Replacing for triangle inequality with the condition $d(X,Y) \leq \max\{d(X,Z), d(Z,Y)\}$ defines an ultra-metric, which plays an important role for the hierarchical cluster analysis.

[Perner, 2003] shows a classification of some distance metrics explaining their interconnections. For two vectors $X = (x_1, x_2, ..., x_n)$ and $Y = (y_1, y_2, ..., y_n)$ we can use different metrics:

[134] http://www.britannica.com/EBchecked/topic/378781/metric-space

- The L^p -metric also called *Minkowski metric* is defined by the following formula: $d_{L^p}(X,Y)=\left[\sum_{i=1}^{n}|x_i-y_i|^p\right]^{1/p}$, where the choice of the parameter p depends on the importance of the differences in the summation;
- L^1 -metric, also known as rectilinear, taxi-cab, city-block or Manhattan metric is received for $p=1$: $d_{L^1}(X,Y)=\sum_{i=1}^{n}|x_i-y_i|$. This measure, however, is insensible to outlier since big and small difference are equally treated;
- In the case of $p=2$ the resulting spaces are the so-called *Hilbert spaces.* One of most popular of them are *Euclidean spaces*, where the distance is calculated as: $d_{Euclidean}(X,Y)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2}$. This metric gives special emphasis to big differences in the observations and is invariant to translations and orthogonal linear transformations (rotation and reflection). In image retrieval weighted Euclidean distance is also used [Wang et al, 2001];
- In the case $p=\infty$ L^∞ *-metric*, which can be also called *Chebyshev* or *Max Norm* metric is obtained: $d_{Chebyshev}(X,Y)=\max_{i=1}^{n}|x_i-y_i|$. This measure is useful if only the maximal distance between two variables among a set of variables is of importance whereas the other distances do not contribute to the overall similarity.

In image matching often *Hausdorff measure* is used. The Hausdorff distance measures how far two subsets of a metric space are from each other. It turns the set of non-empty compact subsets of a metric space into a metric space in its own right. Informally, two sets are close in the Hausdorff distance if every point of either set is close to some point of the other set. The Hausdorff distance is the longest distance you can be forced to travel by an adversary who chooses a point in one of the two sets, from where you then must travel to the other set. The Hausdorff distance is symmetricized by computing in addition the distance with the role of X and Y reversed and choosing the larger of the two distances: $d_{Hausdorf}(X,Y)=\max(\max_{i=1}^{n}\min_{j=1}^{n}d(x_i,y_j),\max_{j=1}^{n}\min_{i=1}^{n}d(y_j,x_i))$.

Often techniques, used in text matching, are applied in image retrieval too. One such example is *cosine similarity*, which is a measure of similarity between two vectors of n dimensions by finding the cosine of

the angle between them, often used to compare documents in text mining. Given two vectors of attributes, X and Y, the cosine similarity is represented using a dot product and magnitude as:

$$d_{cosine}(X,Y) = \frac{X \cdot Y}{\|X\| \, \|Y\|} = \frac{\sum_{i=1}^{n} x_i * y_i}{\sqrt{\sum_{i=1}^{n} x_i^2} \sqrt{\sum_{i=1}^{n} y_i^2}}.$$

9.1.2 Distance Measures for Categorical Data

Categorical data are two types – ordered and nominal. The analysis of symbolic data has led to a new branch of Data Analysis called Symbolic Data Analysis (SDA) [Esposito et al, 2002].

The degree of dissimilarity can be defined by assigning levels of dissimilarity to all the different combinations between attribute values. The mapping can be made to discrete space [0, 1] or more complex discrete linear spaces. Special distance coefficients have been designed for nominal attributes. The basis for the calculation of these distance coefficients is a contingency table, where as columns and rows either the status "not present" (0) or "present" (1) of the property are placed. The cells of the tables contain the frequency of observations that do not share the property (N_{00}), either only one object contains the property (N_{01} or N_{10}), or both of them share the property (N_{11}).

Given that distance coefficients for nominal data can be calculated as variants of generalized formula [Nieddu and Rizzi, 2003]:

$$\frac{N_{11} + tN_{00}}{N_{11} + v(N_{10} + N_{01}) + wN_{00}}, \; t = \{0,1\}, v = \{0,1,2\}, w = \{0,1\}.$$

Several coefficients are subordinated of this formula, such as Jaccard coefficients ($t = 0, v = 1, w = 0$), Russel-Rao coefficients ($t = 0, v = 1, w = 1$), Sokal-Sneath coefficients ($t = 0, v = 2, w = 0$), Sokal-Michener coefficients ($t = 1, v = 1, w = 1$), Roger-Tanimoto coefficients ($t = 1, v = 2, w = 1$), etc.

Other similarity measures that do not fit in the previous class are considered respectively as arithmetic and geometric mean of the quantities $N_{11}/(N_{11} + N_{10})$ and $N_{11}/(N_{11} + N_{01})$ that represent the proportional of agreements on the marginal distributions: $\frac{1}{2}\left(\frac{N_{11}}{N_{11} + N_{10}} + \frac{N_{11}}{N_{11} + N_{01}}\right)$ (Kulczynski) and $\frac{N_{11}}{\sqrt{(N_{11} + N_{10})(N_{11} + N_{01})}}$ (Occhiai-Driver-Kroeber).

More sophisticated similarity measures, concerning recent data mining techniques, take into account the distribution of combinations of examined attributes as presented in [Boriah et al, 2008].

9.1.3 Probability Distance Measures

The disadvantage of the metric measures is that they require the independence of the attributes. A high correlation between attributes can be considered as a multiple measurement for an attribute. That means the measures described above give this feature more weight as an uncorrelated attribute. Some examples of the used distances of such type are shown below.

The *Mahalanobis distance* (or "generalized squared interpoint distance") is defined as: $d_i = \sqrt{(\overline{x_i} - \overline{y_i})S^{-1}(x_i - y_i)}$ and takes into account the covariance matrix S of the attributes.

The most familiar measure of dependence between two quantities is *Pearson's correlation*, which is obtained by dividing the covariance of the two variables $\mathrm{cov}(X,Y)$ by the product of their standard deviations σ_X and σ_Y : $\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}$.

Some similarity measures are defined in statistics. *Chi-square* is a quantitative measure used to determine whether a relationship exists between two categorical variables.

Other similarities come from the information theory. One example is the *Kullback-Leibler distance* also called *information divergence*, *information gain* or *relative entropy*. It is defined for discrete distributions of compared objects X and Y, which have probability functions x_i and y_i : $d(X,Y) = \sum_{i=1}^{n} x_i \log_2 \left(\frac{x_i}{y_i} \right)$. Although the information divergence is not a true metric because $d(X,Y) \neq d(Y,X)$, it satisfies many useful properties, and is used to measure the disparity between distributions.

An example of probability measure based approach is the *Earth Mover's Distance* (EMD) [Rubner et al, 1998]. EMD is a measure, which can be used for signatures in the form of sets of vectors. The concept was first introduced by Gaspard Monge in 1781. It is a mathematical measure of the distance between two distributions. Informally, if the distributions are interpreted as two different ways of piling up a certain amount of dirt over the region, the EMD is the minimum cost of turning one pile into the other. The cost is assumed to be amount of dirt moved times the distance

by which it is moved. A typical signature consists of list of pairs ((x_1,m_1) , … , (x_n,m_n)), where each x_i is a certain "feature" (e.g., colour, luminance, etc.), and m_i is "mass" (how many times that feature occurs). Alternatively, x_i may be the centroid of a data cluster, and m_i – the number of entities in that cluster. To compare two such signatures with the EMD, one must define a distance between features, which is interpreted as the cost of turning a unit mass of one feature into a unit mass of the other. The EMD between two signatures is then the minimum cost of turning one of them into the other. EMD can be computed by solving an instance transportation problem using the so-called Hungarian algorithm [Kuhn, 1955]. The EMD is widely used to compute distances between colour histograms of two digital images. The same technique is used for any other quantitative pixel attribute, such as luminance, gradient, etc. Several attempts are focused on proposing fast algorithms for calculating EMD. For instance a fast algorithm for angular type of histograms, which make good representation of hue or gradient distribution, is suggested in [Cha et al, 1999].

9.1.4 Structural Similarity Measures

Structural similarity is involved in a variety of pattern recognition problems when considered from an abstract perspective. The abstraction refers to measurements and observations whose specifics are ignored. One class of such problems is encountered in image processing, where a set of features or objects with topological interrelations is detected in several scenes. Whenever these are presumed to be similar according to position, proximity or else, the degree of similarity is of interest.

Structures are represented throughout by labelled graphs such as image graphs. In image graphs, vertices represent image edges, corners or regions of interest such as regions of constant intensity or homogenous texture. Graph edges represent relations such as neighbourhoods or concept hierarchies. Edge labels represent distances, degrees of association or else. Special branch of measures observed similarities in graph theory. Examples of such measures are given in [Dehmer et al, 2006].

Most classical methods are *based on isomorphic and sub-graph relations*. For large graphs these measures are faced with the complexity of the sub-graph isomorphism problem. Other measures are *based on graph transformations*. The graph edit distance is defined as minimum cost of transformations (deletion, substitutions, insertions) of vertices and edges, which need to transform one graph into another one. The idea of *finding underlying graph grammar*, in which both of graphs belong, is

used to define some further measures. The application of such measures is very complex, because the underlying grammar is difficult to define. *Graph kernels* take the structure of the graph into account. They work by counting the number of common random walks between two graphs. Even though the number of common random walks could potentially be exponential, polynomial time algorithms exist for computing these kernels.

From graph theory in image retrieval, often *Geodesic distances* are used to measure the similarities between images. The geodesic distance is defined as the shortest path between two vertexes of a graph.

9.2 Techniques for Improving Image Retrieval

Single similarity measure is not sufficient to produce robust, perceptually meaningful ranking of images. The results achieved with the classical content based approaches are often unsatisfactory for the user. As an alternative, learning-based techniques such as clustering and classification are used for speeding-up image retrieval, improving accuracy, or for performing automatic image annotation. Including the relevance feedback in this process allows the user to refine t query-specific semantics. Bridging the gaps between primitive feature levels, which are produced in the classic CBIR systems and higher levels, which are convenient for the user, can be made with examining the image retrieval process in more general frame of multimedia retrieval process. It needs to integrate multimedia semantics-based searching with other search techniques (speech, text, metadata, audio-visual features, etc.) and to combine content-based, model-based, and text-based searching. It can be made in two main ways – creating the semantic space through statistical pattern recognition and machine learning techniques or creating semantic concepts, which can be incorporated in an already built semantic space (for instance – created ontologies, describing interconnections between examined concepts). Several techniques in these directions are used.

- *Unsupervised Clustering*

Unsupervised clustering techniques are a natural fit when handling large, unstructured image repositories such as the Web. Clustering methods fall roughly into three main types: pair-wise-distance-based, optimization of an overall clustering quality measure, and statistical modelling. The pair-wise distance-based methods (e.g., linkage clustering and spectral graph partitioning) are of general applicability, since the mathematical representation of the instances becomes irrelevant. One disadvantage is the high computational cost. Clustering based on the

optimization of an overall measure of clustering quality is a fundamental approach used in pattern recognition. The general idea in statistical modelling is to treat every cluster as a pattern characterized by a relatively restrictive distribution, and the overall dataset is thus a mixture of these distributions. For continuous vector data, the most commonly used distribution of individual vectors is the Gaussian distribution.

- *Image Categorization (Classification)*

Image categorization (classification) is advantageous when the image database is well specified, and labelled training samples are available. Classification methods can be divided into two major branches: discriminative and generative modelling approaches. In discriminative modelling, classification boundaries or posterior probabilities of classes are estimated directly, for example, Support Vector Machines (SVM) and decision trees. In generative modelling, the density of data within each class is estimated and the Bayes formula is then used to compute the posterior. Discriminative modelling approaches are more direct when optimizing classification boundaries. On the other hand, generative modelling approaches are easier to incorporate with prior knowledge and can be used more conveniently when there are many classes.

10 Conclusion

As in other cultural heritage domains, digital art images also require methods to resolve art issues and to experiment with and implement approaches for involving the users without compromising the trustworthiness of the resources.

We believe that areas which will develop with a priority in the very near future are:

- Further refining of specialized image retrieval techniques seeking to both improve the quality of the analysis and to overcome the semantic gap;
- Defining best practices in involving the users (individual users as well as communities of users);
- Sustaining trustworthiness of the resources when social media tools are used to add user generated content;
- Improving not only the information delivery but also the user experiences and expanding the delivery of information with immersing technologies.

The ultimate goal is to facilitate the access to art objects in digital form and to convert it to fun and a great experience.

Bibliography

[Agarwal, 2009] Agarwal, A.: Web 3.0 concepts explained in plain English. 30.05.2009. http://www.labnol.org/internet/web-3-concepts-explained/8908/

[Bellman, 1961] Bellman, R.: Adaptive Control Processes: a Guided Tour. Princeton University Press, 1961.

[Best, 2006] Best, D.: Web 2.0 Next big thing or next big Internet bubble? Lecture Web Information Systems. Technische Universiteit Eindhoven, 2006.

[Boriah et al, 2008] Boriah, S., Chandola, V., Kumar, V.: Similarity Measures for Categorical Data: A Comparative Evaluation, In Proc. of 2008 SIAM Data Mining Conf., 2008, Atlanta, pp. 243-254.

[Burford et al, 2003] Burford, B., Briggs, P., Eakins, J.: A taxonomy of the image: on the classification of content for image retrieval. Visual Communication, 2/2, 2003, pp. 123-161.

[Carson et al, 2002] Carson, C., Belongie, S., Greenspan, H., Malik J.: Blobworld: image segmentation using expectation-maximization and its application to image querying. IEEE Trans. on Pattern Analysis and Machine Intelligence, 24/8, 2002, pp. 1026-1038.

[Castelli and Bergman, 2002] Castelli, V., Bergman, L. (eds.): Image Databases: Search and Retrieval of Digital Imagery, John Wiley & Sons, 2002.

[Cha et al, 1999] Cha, S.-H., Shin, Y.-C., Srihari, S.: Algorithm for the Edit Distance between Angular Type Histograms. Technical report, St.Univ. of New York at Buffalo, 1999.

[Chavez et al, 2001] Chavez, E., Navarro, G., Baeza-Yates, R., Marroquin, J.: Searching in metric spaces. ACM Computing Surveys, 33/3, 2001, pp. 273-321.

[Chen et al, 2001] Chen, Y., Zhou, X., Huang T.: One-class SVM for learning in image retrieval. Proc. IEEE Int. Conf. on Image Processing, vol. 1, 2001, pp. 34-37.

[Chen et al, 2005] Chen, C.-C., Wactlar, H., Wang, J., Kiernan, K.: Digital imagery for significant cultural and historical materials – An emerging research field bridging people, culture, and technologies. Int. J. Digital Libraries, 5/4, 2005, pp. 275–286.

[Colombo et al, 1999] Colombo, C., Del Bimbo, A., Pala, P.: Semantics in visual information retrieval. IEEE Trans. on Multimedia 6, 3, 1999, pp. 38-53.

[Comaniciu and Meer, 1999] Comaniciu, D., Meer, P.: Mean shift analysis and applications. 7th Int. Conf. on Computer Vision, Kerkyra, Greece, 1999, pp. 1197-1203.

[Croft, 1995] Croft, W.: What Do People Want from Information Retrieval? (The Top 10 Research Issues for Companies that Use and Sell IR Systems). Center for Intelligent Information Retrieval Computer Science Department, University of Massachusetts, Amherst, 1995.

[Crucianu et al, 2004] Crucianu, M., Ferecatu, M., Boujemaa, N.: Relevance feedback for image retrieval: a short survey. State of the Art in Audiovisual Content-Based Retrieval, Information Universal Access and Interaction Including Data Models and Languages (DELOS2 Report), 2004.

[Dasiapoulou et al, 2007] Dasiapoulou, S., Spyrou, E., Kompatsiaris, Y., Avrithis, Y., Stintzis, M.: Semantic processing of colour images. Colour Image Processing: Methods and Applications, Ch. 11, CRC Press, Boca Raton, USA, 2007, pp. 259-284.

[Datta et al, 2008] Datta, R., Joshi, D., Li, J., Wang, J.: Image retrieval: ideas, influences, and trends of the new age. ACM Computing Surveys, 40/2/5, 2008, 60 p.

[Datta, 2009] Datta, R.: Semantics and Aesthetic Inference for Image Search: Statistical Learning Approaches. PhD thesis, the Pennsylvania State University, 2009.

[Daubechies, 1988] Daubechies, I.: Orthonormal bases of compactly supported wavelets. Communications on Pure and Applied Mathematics. 41/7, 1988, pp.909-996.

[de Berg et al, 2000] de Berg, M., van Kreveld, M., Overmars, M., Schwarzkopf O.: Computational Geometry: Algorithms and Applications (2nd revised ed.), Springer-Verlag, 2000.

[Dehmer et al, 2006] Dehmer, M., Emmert-Streib, F., Wolkenhauer, O.: Perspectives of graph mining techniques. Rostocker Informatik Berichte, 30/2, 2006, pp. 47-57.

[Demartines and Herault, 1997] Demartines, P., Herault, J.: Curvilinear component analysis: a self-organizing neural network for nonlinear mapping of data sets. IEEE Trans. Neural Networks, 8/1, 1997, pp. 148-154.

[Deng and Manjunath, 2001] Deng, Y., Manjunath, B.: Unsupervised segmentation of colour-texture regions in images and video. IEEE Trans. Pattern Analysis and Machine Intelligence, 23/8, 2001, pp. 800-810.

[Dobreva and Chowdhury, 2010] Dobreva, M., Chowdhury, S.: A User-Centric Evaluation of the Europeana Digital Library. In: ICADL 2010, LNCS 6102, 2010, pp. 148-157.

[Eakins and Graham, 1999] Eakins, J., Graham, M.: Content-based Image Retrieval. University of Northumbria at Newcastle. Report: 39, JSC Technology Application Programme, 1999.

[Eidenberger, 2003] Eidenberger, H.: Distance measures for MPEG-7-based retrieval. Fifth ACM SIGMM Int. Workshop on Multimedia Information Retrieval, 2003, pp. 130-137.

[Enser et al, 2006] Enser, P., Sandom, Ch., Lewis, P., Hare, J.: The reality of the semantic gap in image retrieval. Tutorial, 1st Int. Conf. on Semantic and Digital Media Technologies, Athens, Greece, 2006. http://eprints.ecs.soton.ac.uk/13272/

[Esposito et al, 2002] Esposito, F., Malebra, D., Tamma, V., Bock H.: Classical resemblance measures, Analysis of Symbolic Data, Springer, 2002, pp. 139-152.

[Estrada, 2005] Estrada, F.: Advances in Computational Image Segmentation and Perceptual Grouping. PhD Thesis, Graduate Department of Computer Science, University of Toronto, 2005.

[Freeman, 1975] Freeman, J.: The modelling of spatial relations. Computer Graphics and Image Processing, 4/2, 1975, pp. 156-171.

[Gaede and Günther, 1998] Gaede, V., Günther, O.: Multidimensional access methods. ACM Computing Surveys, 30/2, 1998, pp. 170-231.

[George, 2008] George, C.: User-Centred Library Websites. Usability Evaluation Methods. Chandos publishing, 2008.

[Gheyas and Smith, 2010] Gheyas, I., Smith, L.: Feature subset selection in large dimensionality domains. Elsevier, Pattern Recognition, 43/1, 2010, pp. 5-13.

[Gong et al, 1996] Gong, Y., Chuan, C., Xiaoyi, G.: Image indexing and retrieval using colour histograms. Multimedia Tools and Applications, vol. 2, 1996, pp. 133-156.

[Grosky et al, 2008] Grosky, W., Agrawal, R., Fotouchi, F.: Mind the gaps – finding the appropriate dimensional representation for semantic retrieval of multimedia assets. In Semantic Multimedia and Ontologies, Springer London, 2008, pp. 229-252.

[Gruber, 1993] Gruber, T.: A translation approach to portable ontologies. Knowledge Acquisition, 5/2, 1993, pp. 199-220.

[Han and Kamber, 2006] Han, J., Kamber, M.: Data Mining: Concepts and Techniques, Second ed., Morgan Kaufmann Publishers, 2006.

[Herrmann, 2002] Herrmann, S.: MPEG-7 Reference Software. Munich University of Technology. http://www.lis.e-technik.tu-muenchen.de/research/bv/topics/mmdb/e_mpeg7.html

[Huber, 1985] Huber, P.: Projection pursuit. The Annals of Statistics, 13/2, 1985, pp. 435-475.

[Hung et al, 2007] Hung, Sh.-H., Chen, P.-H., Hong, J.-Sh., Cruz-Lara, S.: Context-based image retrieval: A case study in background image access for multimedia presentations. IADIS Int. Conf. WWW/Internet 2007, Vila Real, Portugal, INRIA-00192463, ver.1, 2007, 5 p., http://hal.inria.fr/inria-00192463/en/

[Hurtut, 2010] Hurtut, T.: 2D Artistic Images Analysis, a Content-based Survey. http://hal.archives-ouvertes.fr/hal-00459401_v1/

[ISO/IEC 15938-3] International Standard ISO/IEC 15938-3 Multimedia Content Description Interface – Part 3: Visual, http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=34230

[ISO/IEC JTC 1/SC29 WG11] WG11: The Moving Picture Experts Group "Coding of Moving Pictures and Audio" http://www.itscj.ipsj.or.jp/sc29/29w12911.htm

[Itten, 1961] Itten, J.: The Art of Colour: the Subjective Experience and Objective Rationale of Colour, Reinhold Publishing Corporation of New York, 1961.

[Ivanova and Stanchev, 2009] Ivanova, K., Stanchev, P.: Colour harmonies and contrasts search in art image collections. First Int. Conf. on Advances in Multimedia (MMEDIA), Colmar, France, 2009, pp. 180-187.

[Ivanova et al, 2010/Euromed] Ivanova, K., Dobreva, M., Stanchev, P., Vanhoof K.: Discovery and use of art images on the web: an overview. Proc. of the Third Int. Euro-Mediterranean Conf. EuroMed, Lemesos, Cyprus, Archaeolingua, 2010, pp. 205-211.

[Ivanova et al, 2010/MCIS] Ivanova, K., Stanchev, P., Vanhoof, K., Ein-Dor, Ph.: Semantic and abstraction content of art images. Proc. of Fifth Mediterranean Conf. on Information Systems, Tel Aviv, Israel, 2010, AIS Electronic Library, paper 42, http://aisel.aisnet.org/mcis2010/42.

[Jaimes and Chang, 2002] Jaimes, A., Chang, S.-F.: Concepts and techniques for indexing visual semantics. Image Databases: Search and Retrieval of Digital Imagery. John Wiley & Sons, 2002, pp. 497-565.

[Kato, 1992] Kato, T.: Database architecture for content-based image retrieval. Proc. of the SPIE – The International Society for Optical Engineering, San Jose, CA, USA, vol. 1662, 1992, pp. 112-113.

[Kuhn, 1955] Kuhn, H.: The Hungarian method for the assignment problem. Naval Research Logistics Quarterly, vol. 2, 1955, pp. 83-97.

[Luxburg, 2006] Luxburg, U.: A Tutorial on Spectral Clustering. Tech. Report No. TR-149, Max Planck Institute for Biological Cybernetics, 2006.

[Maitre et al, 2001] Maitre, H. Schmitt, F. Lahanier, C.: 15 years of image processing and the fine arts. Proc. of Int. Conf. on Image Processing, vol. 1, 2001, pp. 557-561.

[Marchenko et al, 2007] Marchenko, Y., Chua, T., Jain R.: Ontology-based annotation of paintings using transductive inference framework. Advances in Multimedia Modeling (MMM 2007), Springer, LNCS 4351, Part I, pp. 13-23.

[Markov et al, 2008] Markov, K., Ivanova, K., Mitov, I., Karastanev, S.: Advance of the access methods. Int. Journal Information Technologies and Knowledge, 2/2, 2008, pp. 123-135.

[Mattison, 2004] Mattison, D.: Looking for good art. Searcher – The Magazine for Database Professionals, vol. 12, 2004, part I – number 8, pp. 12-35; part II – number 9, pp. 8-19; part III – number 10, pp. 21-32.

[Mitov et al, 2009b] Mitov, I., Ivanova, K., Markov, K., Velychko, V., Stanchev, P., Vanhoof, K.: Comparison of discretization methods for preprocessing data for pyramidal growing network classification method. In IBS ICS – Book No: 14. New Trends in Intelligent Technologies, Sofia, 2009, pp. 31-39.

[Moënne-Loccoz, 2005] Moënne-Loccoz, N.: High-Dimensional Access Methods for Efficient Similarity Queries. Tech. Report No: 0505, University of Geneva, Computer Vision and Multimedia Laboratory, 2005.

[MPEG-7:4062, 2001] MPEG-7, Visual experimentation model (xm) version 10.0. ISO/IEC/ JTC1/SC29/WG11, Doc. N4062, 2001.

[Nasrabadi and King, 1988] Nasrabadi, N., King, R.: Image coding using vector quantization: a review. IEEE Trans. on Communications, 36/8, 1988, pp. 957-971.

[Nieddu and Rizzi, 2003] Nieddu, L., Rizzi, A.: Proximity Measures in Symbolic Data Analysis. Statistica, 63/2, 2003, pp. 195-212.

[Ooi et al, 1993] Ooi, B., Sacks-Davis, R., Han, J.: Indexing in Spatial Databases. Tech. Report. 1993.

[Pavlov et al, 2010] Pavlov, R., Paneva-Marinova, D., Goynov, M., Pavlova-Draganova, L.: Services for Content Creation and Presentation in an Iconographical Digital Library. Serdica J. of Computing, 4/2, 2010, pp.279-292.

[Pavlova-Draganova et al, 2010] Pavlova-Draganova, L., Paneva-Marinova, D., Pavlov, R., Goynov, M.: On the Wider Accessibility of the Valuable Phenomena of the Orthodox Iconography through a Digital Library. Third Int. Euro-Mediterranean Conf. EuroMed, Lemesos, Cyprus, 2010, Archaeolingua, pp. 173-178.

[Perner, 2003] Perner, P.: Data Mining on Multimedia Data. Springer-Verlag NY, 2003.

[Pickett et al, 2000] Pickett, J. (ed.): The American Heritage Dictionary of the English Language. Houghton Mifflin Co., 2000.

[Raymond, 1999] Raymond, E.: The Cathedral & the Bazaar. O'Reilly, 1999. http://www.catb.org/~esr/writings/cathedral-bazaar/cathedral-bazaar/

[Rubner et al, 1998] Rubner, Y., Tomasi, C., Guibas, L.: A metric for distributions with applications to image databases. In IEEE Int. Conf. on Computer Vision, 1998, pp. 59-66.

[Saul and Roweis, 2000] Saul, L., Roweis, S.: An Introduction to Locally Linear Embedding. Tech. Report, AT&T Labs and Gatsby Computational Neuroscience Unit, 2000.

[Shi and Malik, 2000] Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. on Pattern Analysis and Machine Intelligence, 22/8, 2000, pp. 888-905.

[Smeulders et al, 2000] Smeulders, A., Worring, M., Santini, S., Gupta, A., Jain, R.: Content based image retrieval at the end of the early years. IEEE Trans. on Pattern Analysis and Machine Intelligence, 22/12, 2000, pp. 1349-1380.

[Smith, 2002] Smith, L.: A Tutorial on Principal Components Analysis, 2002. http://csnet.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf

[Snoek et al, 2005] Snoek, C., Worring, M., Smeulders, A.: Early versus late fusion in semantic video analysis. In Proc. of the 13th Annual ACM Int. Conf. on Multimedia, 2005, pp. 399-402.

[Stanchev et al, 2006] Stanchev, P., Green Jr., D., Dimitrov, B.: Some issues in the art image database systems. Journal of Digital Information Management, 4/4, 2006, pp. 227-232.

[Stork, 2008] Stork, D.: Computer image analysis of paintings and drawings: An introduction to the literature. Proc. of the Image processing for Artist Identification Workshop, van Gogh Museum, Amsterdam, The Netherlands, 2008.

[Striker and Dimai, 1997] Striker, M., Dimai, A.: Spectral covariance and fuzzy regions for image indexing. Machine Vision and Applications, vol. 10, 1997, pp. 66-73.

[Tremeau et al, 2008] Tremeau, A., Tominaga, S., Plataniotis, K.: Colour in image and video processing: most recent trends and future research directions. Journal Image Video Processing, vol. 1, 2008, pp. 1-26.

[Tu and Zhu, 2002] Tu, Z., Zhu, S.: Image segmentation by data-driven Markov chain Monte Carlo. IEEE Trans. Pattern Analysis and Machine Intelligence, 24/5, 2002, pp. 657-673.

[Wang et al, 2001] Wang, J., Li, J., Wiederhold, G.: SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. IEEE Trans. on Pattern Analysis and Machine Intelligence, 23/9, 2001, pp. 947-963.

[Wang et al, 2003] Wang, S., Chia, L., Deepu, R.: Efficient image retrieval using MPEG-7 descriptors. Int. Conf. on Image Processing, 2003, pp. 509-512.

[Wang et al, 2006] Wang, J., Boujemaa, N., del Bimbo, A., Geman, D., Hauptmann, A., Tesic, J.: Diversity in multimedia information retrieval research. Proc. of the ACM SIGMM Int. Workshop on Multimedia Information Retrieval (MIR) at the Int. Conf. on Multimedia, 2006, pp. 5-12.

[Wei-ning et al, 2006] Wei-ning, W., Ying-lin, Y., Sheng-ming, J.: Image retrieval by emotional semantics: a study of emotional space and feature extraction. IEEE Int. Conf. on Systems, Man and Cybernetics, vol. 4, 2006, pp. 3534-3539.

[WIDWISAWN, 2008] Special issue on Web 2.0. vol. 6 n. 1. http://widwisawn.cdlr.strath.ac.uk/issues/vol6/issue6_1_1.html

[Won et al, 2002] Won, Ch., Park, D., Park, S.: Efficient use of MPEG-7 edge histogram descriptor. ETRI Journal, 24/1, 2002, pp. 23-30.

[Yang et al, 2005] Yang, N., Dong, M., Fotouhi, F.: Semantic feedback for interactive image retrieval. Proc. of the 12th ACM Int. Conf. on Multimedia, 2005, pp. 415-418.

[Yang et al, 2008] Yang, N., Chang, W., Kuo, C., Li, T.: A fast MPEG-7 dominant colour extraction with new similarity measure for image retrieval, Journal of Visual Communication and Image Representation, 19/2, 2008, pp. 92-105.

[Zhou and Huang, 2003] Zhou, X., Huang, T.: Relevance feedback in image retrieval: a comprehensive review. Multimedia Systems, 8/6, 2003, pp. 536-544.