Multimedia Databases Systems Summary

§1 Introduction

Definition

Multimedia can be defined in various ways, such as

- The meaning of its parts. Multi, meaning many/various, and media (as defined by the MHEG) meaning the distribution and presentation of information.
- Any combination of digitally manipulable types of media, which requires a mixing of both continuous and discrete media, and a significant degree of independence between these media.
- Any element produced by means of authoring systems, which are programs that provide means to create complete multimedia presentation by interlinking objects.

Multimedia has attributes, like

- being interactive, when the user is able to control which elements are delivered and when. When multimedia is interactive and has a navigable structure, it's called hypermedia.
- begin linear (generally non interactive), when it can only be represented in a single continuous flow over time, or non-linear, when it consists of a set of elements that may be presented according to different flows.

It can be classified by kinds of perception, representation (main interpretation for the lecture), presentation, storage, transmission and information exchange.
Multimedia can also be represented in terms of Space and Time, like text and graphs, that are time independent and videos or audios, that are time dependent.

Multimedia System

Definition: A Multimedia System is characterized by the computer-controlled generation, manipulation, presentation, storage, and communication of a set of independent media, which include at least one continuous (time-dependent) and one discrete (independent of time) medium. It should have no rigid connection between the combined media and can be able to communicate to other distributed environments.

Domains of Multimedia

- Basic: includes principles of the processing of digital data, such as Nyquist-Shannon sampling theorem and the Pulse Code Modulation. For audio technology there are techniques such as music and speech processing, and for video, technology based on television technology.
- System: it refers to the quality of service of Multimedia Systems, including transaction (according to a certain abstraction), storage and communication of media.
- Services: meaning the ready-to-use services provided as integrated functions for users, including user-to-user communication, temporal synchronization, security measures, document structuring, or content analysis.
- Usage: meaning the literal user interfaces and user experience applications represented into MM systems.

Data Streams

A data stream is a sequence of individual packets (continuous or discrete) that are transmitted under time-dependent constraints. These can by asynchronous, in which packets should reach the destination as soon as possible (perfect for discrete media), synchronous, which defined a maximal end-to-end delay, and isochronous, which defines maximal and minimal end-to-end delay. In data streams, not only the transmission times can vary, but also the packets size, having either fixed bitrate or variable (weakly even or uneven) bitrate.
Recap: Types of media can be discrete (text, graphics, pictures) or continuous (audio or video).

Multimedia Databases

As well as just relational databases, MM databases must provide some core functionalities, but the most relevant ones are the transparence of all physical aspects, search functionalities based on content, and access structures for multimedia and their descriptive data (metadata).
Retrieval of data in MM databases can be done depending on the type of databases, being these based on structure data with deterministic querying and matching, or based on unstructured data, typically applying a fuzzy matching according to the artifacts inside the content. A integral MMDB management system should combine both technologies, and for this, a good option is to use a Object Oriented relational model.
Queries in MMDB management systems are used to browse and navigate through the dataset of a MMDB, and can be predicated by time (such as temporal relations), by space, a combination of both, by semantics, by example or by question-answering.

Exercise Notes

Pulse-code modulation (PCM) is a method used to digitally represent sampled analog signals. In this context, sampling applies a uniform grid over the analog signal, and quantization assigns a unique digital level. Maximum quantization error $Q : \frac{\Delta x}{2^{N+1}}$
Nyquist-Shannon Sampling Theorem states that "If a function (signal) with the highest occurring frequency f_g is sampled at a sampling ratef_s, so that $f_s > 2f_g$, then this function can be reconstructed from the sampled values without losing the underlying information". Aliasing is a phenomenon that arises when a signal is sampled at a rate that's insufficient to capture the changes in the signal. Check differences between structured and unstructured data. Semantic gap is the difference between the application of low and high level features.

2 Color

The color humans perceive is only the "visible light" part of the electromagnetic spectrum, which has a wavelength between 400nm and 700nm. The human eye is composed of three kinds of cones: S-cones, that perceive light with short wavelength (430nm), M-cones (530nm) and L-cones (560nm).

Color Models

Considering that the theoretical number of colors is infinite, we use color models in order to describe colors by their characteristics and how perceivable they are. There are additive (RGB) and subtractive (CMY) color systems.

- CIE-Normalized color table; follows the value of colors in a cartesian table, using a function $R^2 \to R$. With this, we can define a color Gamut, which is the subdomain of colors that can be accurately reproduced by a certain device. e.g. inkjet printers can't reproduce all the spectre of colours.
- RGB color cube. Additive color model.
- CMYK. Subtractive color model, adds black.
- HSB/HSV/HSI. in these, H is hue, S is saturation, and B/V/I can be brightness, value or intensity. It can be transformed from and to RGB.
- HMMD (Hue max min diff). Cone shaped model.
- CIE-Lab-System. 3 Dimentional color space. Enhancement of the CIE color model, and oriented towards physiological properties of human perception. Also enables lossless conversions of color information from system to system.
- YUV. Includes luminance (Y) and chrominance (U, V) values of the color space, just like a human eye does (its also the closest to it). It separates intensity from color information.

Exercises Notes

Metamers are two same perception of colors produced by different spectral radiant power distributions. A color model is an abstract method for representing color information, and can use characteristics of the human vision system in order to do it. There are additive and subtractive types.
Converting from RGB to CMYK

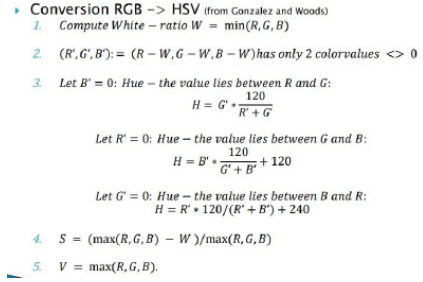
$$\begin{pmatrix} \bar{C} \\ \bar{M} \\ \bar{Y} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} 0.8 \\ 0.4 \\ 0.7 \end{pmatrix}$$

$$K = min(\bar{C}, \bar{M}, \bar{Y}) = 0.4,$$

$$C = \frac{\bar{C} - K}{1 - K}, M = \frac{\bar{M} - K}{1 - K}, Y = \frac{\bar{Y} - K}{1 - K}$$

$$C = 0.67, M = 0, Y = 0.5$$

Converting from RGB to HSV



Check differences between CIE XYZ and CIE L*a*b

3 & 4 Image Medium

Basics

Digital images are usually bitmap represented, and can be from sources such as digital cameras, digital drawings, scanners, etc. Their disadvantage is the need for memory, specially in cases of escalation of images. While editing means the modification of digital images using specialized software, image processing includes the use of mathematical algorithms for either the edition, modification or analysis of images.

Resolution

Device Resolution defines how accurately is a device display at approximating an image representation. It can be represented in dots per inch, pixel dimension, or else. An image is basically an array of pixels without defined physical dimension, so we can give one to the 2D array by computing $\frac{\text{Pixel dimension value}}{\text{Physical dimension value}}$

Bit depth $\rightarrow 2^b = \#$ colors of the display

Then, b equals to the number of bits per pixel needed to represent the colors.

When the image resolution is lower than the device resolution, interpolation is required, which can lead to loss of quality. On the other case, we apply downsampling.

Considering a scaling example, there is a s factor, to which the initial resolutions x and y are transformed into $\frac{x}{s}$ and $\frac{y}{s}$. These values are not always integers.

Kinds of interpolations

- Nearest neighbor: The new pixel gets its value from the closest pixel from the original image.
- Bilinear interpolation: Linear interpolations from both dimensions (in case of 2D images) are used.
- Bicubic interpolation: Same as before, but using cubic splines.

Image organization

After transmission, images can be stored bitwise in different manners. Taking this into care, the information from every channel can always be retrieved by a logic filtering (AND) and a byte shifting operation.

Quantization of Color

When storing images, one has to take into account the amount of memory needed to represent all the colors of it, this is color quantization. There are many approaches to it, such a direct assignment (arbitrary 32 bits) or a look up table. The last method only includes a color index, instead of the full RGB value. When a CLUT doesn't have enough colors, one can also interpolate them to the closest one (causes posterization).

Dithering is the name of the phenomenon that occurs when there's an apparent increase in the number of perceivable colors, when applying spatial displacement. In other words, it adds arbitrary noise into the neighboring pixels of an image in order to avoid quantization errors.

Formats

Usual image formats have information regarding: image parameters, metadata, functionalities, types of compression, java or browser support, etc.

Some examples are:

- TIF (tagged image file format): composed of a header, image data, and tags according to the kind of data used.
- GIF (graphics interchange format): has loseless efficient compression, and may contain more than a single image in a single file. composed of header, encoding of the application that created it, structure control, image data, comments and plain text information, in case the image contains it.
- PNG (portable network graphics): improvement for the two previous formats, containing 4 possible loseless data compression levels. It uses pre filters in order to improve the compressibility of the data.

Format	Storage req.	Application	Remark
JPEG	Low to average	P, D, W	Lossless compression
TIFF	hoch	M, I, P, D	Lossless compression possible
GIF	Average	M, I, W	Indexed colors, possible high loss of tones
PNG	Average (or low)	M, I, W, (P)	Lossless compression possible

Vector Graphics

Vector graphics are mathematically and programmatically defined drawing instructions within a coordinate system (vectorized images). For this kind of graphics, isometric transformations cam be easily and accurately applied. One drawback of it is that vector graphics need to be drawn, which means they have bad interpretation or reproducibility. Computer generated images are usually a collection of geometric elements, either in 2d or 3d. Some operations/transformation for graphics are: shading, mapping, lighting, displaying and rendering (generation).

For the Scalable Vector Graphics (SVG) file format, the point (0,0) in the 2d coordinate system is the top left corner, but it can be user modified.

Bezier Curves

Parametric curves are intended to provide the generality of polygons but with fewer parameters for smooth surfaces. A spline is a parametric curve defined by control points. For a Hermite spline, a user needs to provide the endpoints of the curve, and the parametric derivatives. Unlike Hermite splines, Bezier curves replace the derivative points for geometric tangents.

The user supplies d control points, $\boldsymbol{p}_i$
Write the curve as:

$$\mathbf{x}(t) = \sum_{i=0}^{d} \mathbf{p}_i B_i^d(t) \qquad B_i^d(t) = \binom{d}{i} t^i (1-t)^{d-i}$$

Properties

- First and last control points of Bezier Curves are interpolated.
- The curve lies entirely within the convex hull of its control points.
- Using them in svg files reduces the data usage.

Image Manipulation

A color histogram plots the number of pixels with a certain color value for a certain channel.
Types of image operation:

- Point-wise operations: newPixel(x,y) = f(Pixel(x,y))
- Neighborhood operations (filters): here, the new pixel value depends on the value of its surrounding neighbor pixels. A linear filter (linear combination) is also called convolution. A filter can be called "Low pass" if all frequencies of the image are reduces a certain cut-off freq, or "High pass", in the contrary case.
- Geometric operations

Exercise Notes

Vector images occupy less space as they make use of sequential 2d or 3d commands in order to represent the set of pixels. Raster images use a bitmap, ergo, more space.
All main image formats (GIF, PNG, JPEG) use the RGB color model, but only the first two use a lossless compresion. JPEG is DCT based.
Check memory usage between a 3D vs a 2D storage. Binary operators for unpacking.
Pixel density: we need to compute width and height from the aspect ratio and the diagonal screen size. Then, resolve using the pixel resolution from any of the dimensions.
Color quantization: for an image I with not enough m different colors, we set a new set of colors $C'$of size $n << m$ such that we have minimal perceptible degradation. Can be uniform (uniformly distributed along the color model) or of median cut. Noise dithering tries to relief the effects of quantization by adding random noise.
In linear point-wise operations, α is the contrast factor, and β is the brightness factor. the biggest contrast ratio for an image with minimum and maximum values a and b is given when $HK(a) = 0$ and $HK(b) = 255$ are the respective min and max value of the kind

of image. Solutions for linear filters at the boundaries are zero padding and redefinition. Weighted smoothing is usually applied using x and y into the functions as the manhattan distance from the current pixel to the target pixel. Laplacian filters are based on the derivatives of the dimensions, and usually detect edges on images.

5 Text, Video and Audio

Characters and text

Text character representation can be done by bitmaps, or by outlines of the font styles. Text as a media type can be represented in many ways, such as ASCII or ISO characters, or markup, structured or hyper text. On the other hand, operations applicable to text are, string operations, edition, formatting, pattern recognition, sorting operations, encryption, etc.
Unicode characters are coded on 16 Bits, but UTF-8 encoding allows to store these characters only using either 1, 2 or 3 bytes, saving space over not using any compression. All ASCII characters are coded with a single byte.
XML as a markup language is composed of a prologue, and a body. Throughout the body, there are elements delimited by tags. XML files must be formatted correctly, by only having a single root element, only elements that are structurally connected to this root and by not containing any invalid characters. XML files only give information about structure, not presentation, which allows to other markup languages like XSL to transform XLM files into different ones, separating both domains.

Video

The hierarchy of a video is: Frame → Scenes → Shots (General, medium or close up; Static or Dynamic) Partition of a video stream in physical and logical video segments.
A video is basically a sequence of frames, with a specific resolution and framerate. Color histograms and edge detection can help detect shots in a video (shot segmentation). Video summarization include shot segmentation (of key frames) of only relevant parts of a video. It can be independent, or user dependent by parameters.
Different video format represent architectures, or how the data in the file is structured. These also consider different compression methods (codec). Examples of formats are, AVI, MOV, or WEBM.

Audio

Audio is a one dimensional acoustic wave, that causes vibrations on a microhpone or eardrum. The range of human hearable audio goes from 20 to 20k Hz.
Analog digital transformation of audio can be done by the sample of audio waves every ΔT seconds. This can also lead to quantization errors due to the finite number of available bits per sample.

6 Compression

The principal motivation for media compression is the storage size reduction. Criteria for data compression are, compression rate, quality maintenance (lossy methods), low computing effort and low latency. This can be translated also into, independence of resolution and framerate, use of different data rates for audio and video, etc, low cost, fast computation when displaying the data, etc.
Symmetrical compression takes same amount of time for compression and decompression. Asymmetrical compression is computationally intensive in the compression, and fast at decompressing.
Data coding (compressing) can be of type Entropy Coding, where properties of the data are ignored, redundancy in the signal get reduced but has low compression factors, or Source coding, which has better compression rate and consider the specificities of the data source and recipient.
Loseless compressions can be classified as static (two pass methods), adaptative (one pass) or hybrid. Each kind of compression must provide a encode and decode function. For images, a pipeline would be Source -¿ Source Coding -¿ Channel Coding -¿ Channel -¿ Channel Decoding -¿ Source Decoding -¿ Sink/Display The previous pipeline can be generalized for image and video as: Image Preparation ¡-¿ Image Processing ¡-¿ Quantization ¡-¿ Entropy Coding

Run-length Encoding

Loseless compression method suitable for data with long sequence of identical characters. If a byte appears more often than a certain threshold, its compressed.

Statistical Encoding

Symbols are coded by sequences of varying length. For frequent symbols, the coding is shorter, and viceversa for rare symbols.
Huffman Coding Algorithm

- First a frequency table is created. Sort nodes by frequency and create the binary tree.
- Traversal of the tree.
- Coding the text.
- Storing the text.

Properties of the Huffman Code are, that it's optimal, no compression for random data, requires two traversals, and the code book must be stored in addition to coded data. This code data can be built by traversing the tree in pre order.

Transformation Encoding

Applies a transform to the data, expressing them into another mathematical space. E.g. FFT.

JPEG Coding

The first version, JPEG-1, has four modes:

- Baseline: It identifies the less important image information for human vision. This ends up discarding non relevant information (lossy) but having higher redundancy in the final data (loseless). Process: Component Decomposition -¿ Subsampling -¿ DCT -¿ Quantization -¿ Zigzag scan -¿ Run-length Encoding -¿ VLi and Huffman -¿ Goal. Here, component decomposition could mean a change in the color model. Subsampling takes just the most important information from the color model used. Discrete Cosine Transformation is applied to block of 8 by 8 for each color channel. It is reversible, but there is information loss still. Finally, zigzag scan takes first the most significative block and Huffman encodeing encodes it.
- Progressive
- Lossless
- Hierarchical

MPEG 1 & 2

The simplest form of compression is to save every frame as a JPEG. It is fast and simple, but the compression rate is low.

Kinds of compressions are: Intra coding, where spacial redundances are reduces, or inter coding, where temporal redundances are reduced. One way of comparing similarity between frames is to subtract them. An alternative to this is quantify the video's frames into block and check for motion. In an optimal case occurs, we only need to save the motion vectors (inter-coding). The calculation of the error between block is called Motion Compensation. Motion Compensation Algorithms are used to reduce the difference between the Motion vector and the actual motion.

The first MCP naive method is to try applying compensation in all possible fields, which results in a solid but slow compensation. Another one is to search for the best reference block in a given pixel window.

Block Matching can be measures using MSE or SAE, for every motion vector. One can also use a spiral search for block matching.

Further algorithms are, using the motion vector of neighboring blocks, sub sampling (smaller blocks) or using a mathematical transformation.

A MPEG-2 stream follows the structure (I/P [BB])* in which I is a unreferenced (independent) frame, P is a predictively coded frame using s previous P or I frame, y B are bi-directionally coded frames based on the preceding and following I or P frame.

A group of pictures (GOP) is a sequence of frames between two I-frames, which depend always on the first I-frame, and are the smallest unit of random access. Its length goes from 10 to 250 frames. Macro block from frames are also encoded in long horizontal slices, in order to avoid error propagation and to allow parallel coding/decoding. Coding of a Macro block in MPEG is similar to the one in JPEG, but they're also inter or intra coded.

The MPEG-4 compression standard uses object bases coding. Here, a video is a scene that has Scene-Graphs, that describe objects in certain frame windows.

H.264 encoding is part of MPEG4, with better quality and less transmission errors. Arrangement of partitions (sub divisions of macroblocks) depend on the pixel structure, which are optimized for motion description. It also adds a Skipped S frame of blocks, and only saves the motion vector, not any compensation. This encoding applies a neighboring pixel intra prediction in all of the 9 discrete directions.

Exercises Notes

The idea of JPEG compression is to get a reversible (but lossy) compression. In this pipeline, subsampling, DCT and quantization are the lossy steps. By changing color model, the preprocessing can subsamble the chroma coefficients, such that only the luma stay.
The idea of resizing the image before compression is so the blocks of 8 by 8 are formed and the subsampling is done correctly.
DCT transform image information of space domain into frequency domain, which is the rate at which pixel values change over spatial distance.
The most important value in a DCT transformed block is the one in the most upper-left element. This one is called DC coefficient, the other 63 are called AC coefficient.
Quantization in JPEG compression is made by dividing every element of de DCT transform element-wise by the elements of the quantization matrix. This makes a better compression, but also could worsen the quality and add quantization errors (noise).
The purpose of entropy encoding is to reduce redundancy in quantized coefficients. Run length encoding helps reducing data by compressing series of identical bits. RLE is only applied for AC coefficients, as there might be a lot of zeroes. It is coded with skips, which represent the number of zeroes, and value, the next non-zero value. For DC values, only the difference between values is used.
LZW Encoding

1. The Prefix is empty and the dictionary is initialized with all characters
2. Read the next character z from the input stream
3. Is "' Prefix + z "' in the dictionary?
   Yes: Prefix = Prefix + z
   No:
      Output Code for Prefix
      put Prefix + z in the dictionary
      Prefix = z
4. Has the end of the input stream been reached?
   No: go to step 2
   Yes: If the Prefix is not empty, output the corresponding code

LZW Decoding

Dictionary initialised with all occurring characters
Code = first code from the input stream (always one character)
Output the entry for code and
Store code in OldCode
Code = next code in the input stream
Prefix = character of OldCode
Code in dictionary?
Yes:
- Output the character of Code
- Character = first character of Code
- Enter Prefix+Character in dictionary

No:
- Character = first character of OldCode
- Enter prefix+character in dictionary AND output it

As long as characters still exist, go to 4.

Huffman Coding

- Create a leaf node for each symbol and add it to the priority queue.
- While there is more than one node in the queue:
  Remove the two nodes of highest priority (lowest probability) from the queue
  Create a new internal node with these two nodes as children and with probability equal to the sum of the two nodes' probabilities.
  Add the new node to the queue.
- The remaining node is the root node and the tree is complete.

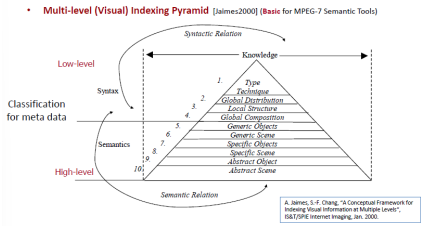
7 Multimedia Modeling

Multimedia Annotation

Multimedia annotation is the task of associating textual labels or tags to multimedia objects in order to represent their (semantic) content. Can be manual, automatic or semi-automatic, but it comes with all the problems of semantic recognition. The sensory gap is the gap between the object in the world and the information in a description derived from a recording of that scene, or rather the uncertainty about the status of an object. It measures the difference between the data and the actual knowledge.

Multimedia metadata is the information about attributes not explicitly present in the media, and can be intrinsic or infered from the actual data, or extrinstic or independent of the primary data.

MPEG-7



8 Content Based Image Retrieval

It is focused on low level features, by the representation of an image as a set of descriptors. The challenge here is the different perception of different users, and the lack of expert knowledge. A solution to this is a user controlled, interactive query process focused on features (tends to be inaccurate, QBF), or examples (accurate, QBE). Each result page of these queries is composed of result images, and similar features. Visual characteristics of images can be looked for until a certain threshold.

QBFs are queried on a single descriptor, and results in all images with a similarity greater than 0, or on several descriptors, combining them all. Feature indexing is a technique implemented to make QBFs faster, but it introduces different problems to the query, such as the curse of dimensionality, or that the index structure must support non-euclidean similarity measures.

A color histogram is a discrete representation of the distribution of features, and is based on pre defined prototypes per color c_n. The set of elements in a prototype is a "bin" or a "container".

Color Quantization

- For 256 values for each RGB color, we have 16777216 different colors that are distributed into n "bins". A histogram is a vector of bin values.
- Bin quantization defines the bit-coding of the value in each bin. If for each bin we need 15 Bits, we would need $15n$ Bits to store a histogram.

Distance metrics

Distance metrics are invariant to isometric operations.

$$D(I,J) = \left(\sum_i |f(i,I) - f(i,J)|^p \right)^{1/p}$$

$$D(I,J) = \sum_i \frac{(f(i;I) - f'(i))^2}{f'(i)}, \qquad f' = \frac{[f(i;I) + f(i;J)]}{2}$$

$$D(X,Y) = \sum_i f(i;X) \log \frac{f(i;X)}{f(i;Y)},$$

$$d_\cap(H,K) = 1 - \frac{\sum_\mathbf{i} \min(h_\mathbf{i}, k_\mathbf{i})}{\sum_\mathbf{i} k_\mathbf{i}}$$

Similarity's perception is subjective. Different images can have same histograms, Another approach would be to split the image into regions, and compute histograms for each region.

MPEG 7 Descriptor

CBIR systems have historically focused on the color concept, using descriptors such as color histogram, dominant color, color structure descriptor, dominant color, etc. Dominant color provides a compact description of the representative colors of an image. Spatial cohercy describes the homogeneity of the colors.
MPEG 7 can support color spaces such as RGB, HSV, HMMD, etc.

Exercise Notes

Content-Based retrieval: A content based retrieval system processes the information contained in data and creates an abstraction of its content in terms of attributes. Limitations are, problems of image annotations, with human perception for annotations themselves, or annotating images with words. The semantic gap is also taken into account.
Parts of a CBR architecture are: multimedia data, feature extraction, the database itself, a user interface composed of multimedia presentation and a multimedia query system, and finally a similarity metric system, for the query process.
A feature vector is a numerical representation of object features in a n-dimensional vector. This can bring problems at querying time, mainly while using a dimensionally large vector. They can be avoided by applying some feature selection or extraction method.
Definitions related to CBR:

- Dominant Color: The Dominant Color Descriptor allows specification of a small number of dominant color values statistical properties like distribution and variance. Not the same as color histogram. Defined as
  $F = \{(c_i, p_i, v_i,), s\}, \forall i = 1 \dots N$, where N is the number of

clusters or bins, c_i is the dominant color vector, p_i is the percentage for each dominant color, v_i is the color variance, and s is a scalar that represents the overall spatial coherency of the dominant colors.

- Spatial Coherency: The spatial coherency of a given dominant color is measured with the normalized average number of connected pixels of this color It is computed using a 3 x 3 mask.
- Distance Metrics: Degree of similarity between two points is measured by distance in data space. Several metrics
- Curse of Dimensionality: CBR performs worse when dimensionality of the FVs increases.
- Types of content-based queries: Point query, that retrieves all points with identical feature vectors; range query, which does it with a maximum distance from the query point; and k-nearest neighbor, which returns the k most similar results.

In order to realize a query by feature, one must define previously what the features actually are, like how an object must be a kind of countour, or a set of colors that are defined by the word "red", and how precise these should be.
A uniform color quantization of n bins, divides the color model (RGB for example) into n uniformly distributed sections.
The usual histogram of a 4×4 image has a quantization of 5 bits (because of 16 "blocks"), the quantization of 2 bits would have 4 values total. Usual distance metrics between histograms don't take into account the spatial distribution of color.
Statistical non-parametrical distances.

- Kolgomorov-Smirnov Distance:
  $KS(P,Q) = \max_i |F^r(i,P) - F^r(i,Q)|$ where $F^r(i,P)$ is the cumulative histogram.
- Chi-squared Distance: $D_\chi(P,Q) \sum \frac{(x_i - f'(i))^2}{f'(i)}$ where $f'(i)$ is the mean between the values of P and Q at index i.

Statistical parametrical distances.

- Weighted mean variance:
  $WMV(P,Q) = \frac{|\mu(P)-\mu(Q)|}{|\sigma(\mu(\text{Ref}))|} + \frac{|\sigma(P)-\sigma(Q)|}{|\sigma(\sigma(\text{Ref}))|}$

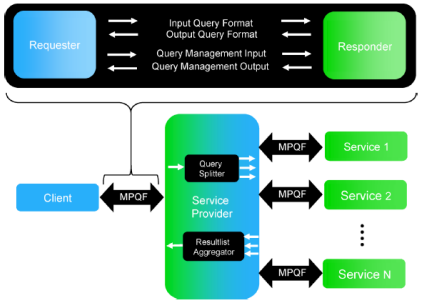
9 Databases query languages

End to end workflow of a MMDBS,
Natural language query -¿ declarative representation -¿ algebraic representation -¿ optimization strategies -¿ query execution plan
A natural language query is self explanatory, they can be classic texts, semantic based, syntactic based, similarity or content based, or correlation queries.

Categories of MMQL

Object Query Language is a general query language with multimedia extensions (MOQL), spatial or temporal relations. MOQL in theory supports all requirements of a general MM query language, but no support for audio. It had to be accepted at some point by new potential users. There is also a SQL extension for multimedia.

In the other hand, we also have versions of query languages made from scratch, like the MPEG query format. It is based on XML, and decoupled from a specific metadata standard.



In some cases the resource service is known to the user, in others, it isn't.

Exercise Notes

Relational vs object-relational DBs

- Relational
  - Only predefined data types can be used.
  - Some attributes of objects (rows) can't be expressed directly.
  - Good for managing large amounts of data.
  - Queries are easier to optimize automatically.
- Object-Relational
  - New types can be always implemented.
  - Several Object Oriented features can be part of the data mode.
  - Good at expressing complex relationships.
  - Scalable and extensible.
  - More complex queries.

Definitions

- User-defined data/object types: Objects of the database, that contemplate attributes, relationships between entities, and methods to manipulate the content.
- Inheritance: it is based on a family tree of object types that form a type hierarchy. It consists of a parent object type (supertype, defined with the keyword NOT FINAL) and one or more child object types (subtypes, defined with the keyword UNDER). Methods can also be interfaces by supertypes with NOT FINAL, and implemented or not on subtypes with OVERRIDING.
- Table of objects: In ORDB these means the storage of row objects previously defined. This also references relational tables, that define relationships between kinds of objects as column.
- Polymorphism: It means that rows of an object table of type A can contain instances of this type or any subtype. Mostly refers to methods overriding (overrides a previous definition) or overloading (containing more than one definition).
- Object identifier: Identifies objects uniquely, either by system generated ones, or by primary key based ones.
- Relationships between types: Similar to relational DBMS.
- REFERENCE (REF): logical pointers that reference a single row object of an object table. Uses the OID, as OID cannot be used directly.
- Collection: group of values where all have the same type. Used for 1:n relationships.
- DOT operator: used to navigate through type's schema.

Check the MM/SQL standard for DB formulation and queries.
Check the MPEG query format.

Indexing is applied to database structures in order to increase execution speed of queries, by enabling fast access to given data. It can increase the storage space and can be slow in cud operations, since indexes need to be recomputed.

Types of indexing

- B(alanced)-trees
- Hashes
- K-d trees
- Point-quadtrees
- R-tree is a height-balanced tree similar to a B-tree but with only index records in its leaf nodes containing pointers to data objects.

Search algorithms for similarity search

- Exact queries: exact result, without indexing the algorithm is simply a sequential search. With indexing, we start with the root node and search recursively until we find a exact match. As areas of object attributes can overlap, it is possible to get more than one result.
- Range queries: Sames as before, but with an attribute area in order to get the match. Now only the areas that intersect with the search are the ones that are retrieved as a result.
- Nearest neighbor queries: Final upgrade, made with a sequence of range queries with increasing radius.