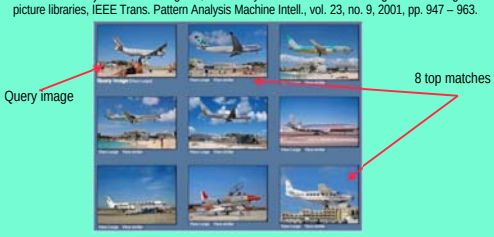
Content-based Information Retrieval

- Search in an on-line database of 800,000 aircraft related images (CBIR "Airliners.net"): see J. Z. Wang e.a., SIMPLIcity: Semantics-sensitive integrated matching for picture libraries, IEEE Trans. Pattern Analysis Machine Intell., vol. 23, no. 9, 2001, pp. 947 – 963.



Information Era

- **Digital revolution**: 1950s → 1990s → at present →…
  - Lossless / high-quality lossy data compression
  - Fast data transmission
  - Compact data storage
  - Powerful and fast data processing
- "**Information superhighway**" - Internet / WWW / multi-media systems and applications
- **Internet protocols** (standard conventions, or rules) to govern communications and data search / retrieval

Internet and the WWW

- TCP/IP Transmission Control / Internet Protocol
- FTP File Transfer Protocol
- HTTP Hypertext Transfer Protocol
- IPP Internet Printing Protocol
- SMTP Simple Mail Transfer Protocol
- IIP Internet Imaging Protocol
- MPEG / ISO multimedia access standards

Multimedia Information Retrieval

- **Multimedia databases** combine texts, graphics, animation, image / video data, speech, and non-speech audio data…
- **Information retrieval** from such databases is a multidisciplinary research / application area:
  - Computational analysis of diverse multimedia data (image / video / audio / text processing)
  - Artificial intelligence (AI) based search strategies
  - Human-computer interaction (HCI)

Multimedia Image / Video Retrieval

- Hundreds of millions still images and dozens of thousands videos are stored in the Web and new heaps of them appear everyday in various repositories
- **Content-based information retrieval** of image and video information from Web-based and other multimedia databases is very important but **extremely challenging** problem…

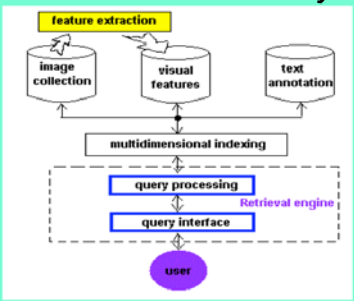
Multimedia Communications Standards

MPEG-1	ISO/IEC IS 11172	Coding of movies and audio (multimedia CD-ROMs; <1.5 Mb/s; Web video distribution
MPEG-2	ISO/IEC IS 13818	Generic video/audio coding; 2 - 50 Mb/s; high-quality digital multimedia transmissions
MPEG-4	ISO/IEC IS 14496	Video/audio object coding: interactive multimedia - distribution of and access to content on the Web
MPEG-4 VTC	Ibid, Pt.2 Visual	Visual texture coding to compress still images and video information in photorealistic 3D models
JPEG2000		Emerging standard to provide rate distortion and subject image quality superior to existing ones
MPEG-7	ISO/IEC IS 15938	**Multimedia content description interface** for the CBIR applications
MPEG-21	ISO/IEC IS 18034	**Multimedia framework** for the transparent wide-range use of multimedia resources

Content Based Image / Video Retrieval

- Hard problem: to derive searching and matching criteria for multimedia image / video data
  - These differ much from a traditional search for text
- Techniques for indexing unstructured visual data are called *content-based video information retrieval* (CBVIR) or more frequently *content-based image retrieval* (CBIR)

Architecture of a CBIR system



Descriptions & Semantics

- *Image descriptors*
  - local and global features of visual objects
  - semantic relationships for features / objects
- Semantics ('significant' in Greek): the relationships between words and meanings in *linguistics* and between signs and what they mean in *logic*
  - In relation to images, semantics is concerned with meaning of depicted objects and their features

Descriptions & Semantics

- Semantic relationships encode interpretations of images which are *relevant to the application*
  - These interpretations constitute a small subset of possible meaningful interpretations
- Automatic description of 'true' contents is an unsolvable problem (also, due to subjective human perception of images)
  - Contents is so far described with digital signatures combining recognised objects, shapes, features, and relationships
  - Images are ranked by similarity to query description in terms of objects, features, and their relationships
  - Top-rank, i.e. most similar images are retrieved and output

Features vs. Content

- *General-purpose* and *domain-specific* features
  - general-purpose features: colour, texture, geometric shape, sketch, and spatial relationships
  - domain-specific features in special applications (face recognition, remote sensing of the Earth's surface)
- Description of semantics (meaning) is a very hard problem with no universal solution!
  - Meaningful descriptions (interpretations) easily formed by human vision are typically extremely difficult for computation

Contents or Meaning: What Is It?

A horse & a foal & a grass field & a bush...?

↑

Objects/relations of interest in an image or video depend on an observer, time, goals, and other subjective and objective factors...

Query by Content

- The most difficult problem: how to describe what does the user need?..
  - *Example*: "Find a picture with a horse and a bush"
  - *Even a harder task*: to match a specification against a large multimedia database
- Human queries: on a cognitive level exploiting human knowledge of the context in terms of objects, persons, sceneries, scenarios, *etc*.
- In most cases, the query is not specific and is detalised only by feedback, i.e. by relevance of retrieved images

Query by Content

Humans specify image contents in different ways:

- "Find an image with a bird"
- "Find an image of an yellow-head gannet"
- "Find the scene where Titanic hits the iceberg"
- "Classify images by the place where they are taken"
- "Select aerial images of Rangitoto"
- "Find similar tornado images but for Tennessee"
- "Select a most impressive sunset image" and so on

Query by Content

Google Earth's database of high-resolution space images of the Earth: visually one can easily find a known place even without geographical co-ordinates. But how to make a formal content-based query?

Sensory and Semantic Gaps

- *Sensory gap* between the properties of an object and its description derived from images
  - Ill-posedness of the content description problem and limited capabilities of formal content representation
- *Semantic gap* between the low-level features extracted to describe the visual data and the high-level human interpretation of the same data
  - The user needs the semantic similarity, but the CBIR can provide only the similarity of signal features

3D Scenes: What to Search For?

Contents of natural 3D scenes is too manifold because interpretation depends on time, an observer, description goals, and a variety of other subjective and objective factors… Human queries are always on a cognitive knowledgeable level; CBIR queries are much more restrictive

Complexity of a Content

1. *Algorithmic (perceptual) properties* of visual information
2. *Semantic properties*, e.g. abstract primitives such as objects, roles, and scenes
3. *Subjective attributes* such as impressions, emotions and meaning associated to the perceptual properties

CBIR

Human

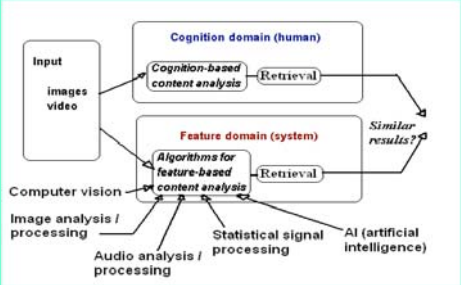
Users' Goals and Queries

- Gaps between "formal" and "human" semantics should be bridged by both extending the image descriptions and adapting the user's queries to how a CBIR system operates
- Various users' goals, e.g.
  - *search for a specific image (target-specific, or target search)*
  - *category search* , or
  - *search by association (open-ended search)*

Search by "Similarity"

- Target-specific search
  - To find a specific target image in a database; search should not return any other image, even very similar
- Category search
  - To find images from a prototypical category, i.e. "foals" or "soccer games", or images similar to a query one
- Open-ended search - browsing
  - A broad, nonspecific goal changing during the search

Algorithmic Level of a CBIR



CBIR vs. Image Analysis

- CBIR combines automatic image analysis and recognition with active user participation
- Retrieval relates inherently to image ranking by similarity to a query example, rather than to image classification by matching to a model
- The user evaluates system responses, refines the query, and determines whether the answers are relevant to that query

From Features to Semantics

- Current computer vision does not allow to easily and automatically extract semantic information
  - Ultimate goal: image encoding capturing an image's semantic content matching to human interpretation
  - Sensed encoding: the image's raw pixel values
- Low-level features - functions of the pixel values
- How to codify image semantics?
  - It is difficult to design a consistent language to express them
  - More practical: hidden languages for semantic encoding

From Features to Semantics

- Probabilistic framework for semantic indexing
  - Random field modelling of features and their spatial distributions with due account of the wide variation of visual features within the same "semantic" class
  - "Semantic" representation using effective clustering and classification techniques, e.g. Support Vector Machines (SVM) or Bayesian networks
  - Feature-based labelling of blocks (regions) of an image for interpreting its semantic content