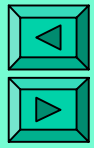
The University of Auckland

Features, Syntax, Semantics…

COMPSCI.708.S1.C

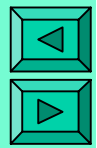
A/P Georgy Gimel'farb



The University of Auckland

Multiple Levels of Visual Data

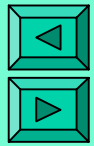
- Basic levels with no special knowledge involved:
  - Pixel-wise patterns (intensities, colours)
  - More general low-level elements: colour regions, texture, motion (inter-frame changes in a video sequence), shapes (object boundaries), etc.
- The most complex level:
  - Images represent abstract ideas depending on individual knowledge, experience, and even on a particular mood



The University of Auckland

Multiple Levels of Visual Data

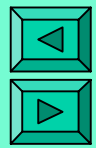
- Image ***syntax***: perceived visual elements and their spatial - temporal arrangement with no consideration to the meaning of the elements or arrangements
- Syntax exists at several perceptual levels - from simple global colour and texture to local geometric forms, such as lines and circles
- Image ***semantics*** deals just with the meaning
- Semantics can also be treated at different levels



The University of Auckland

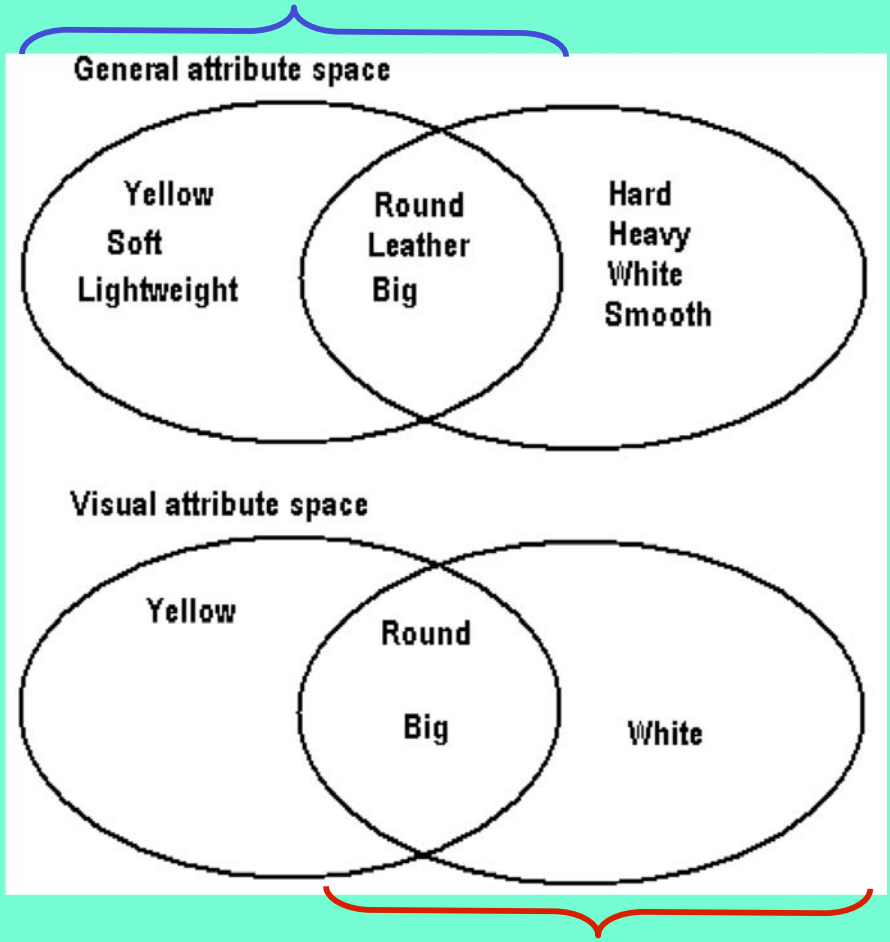
Multiple Levels of Visual Data

- Objects in images are characterised both with ***general concepts*** ("*What is it*?“) and ***visual concepts*** ("*What does it look like*?“)
- QBE relates primarily to the ***visual content***, but the ***non-visual*** content is also of big interest
- The visual content is a multilevel structure: the lower levels refer to syntax, rather than semantics



The University of Auckland

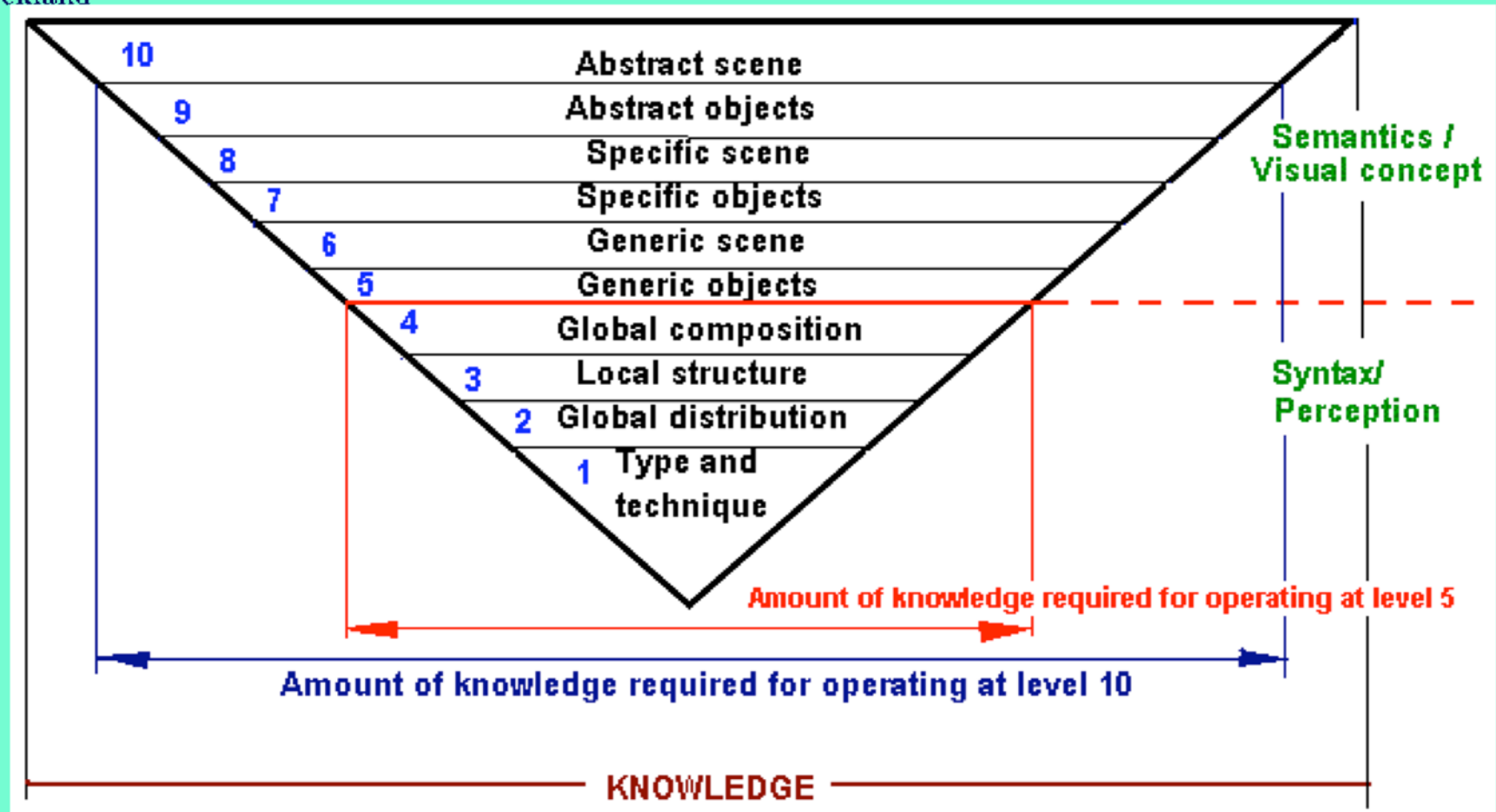
General Vs. Visual Attributes

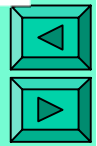


- Attributes selected by a volleyball and baseball players to describe the same object (a ball)
  - Different general concepts of a ball
  - Correlation between visual and general attributes
  - **Visual content** - relates to observed items
  - **Non-visual content** - related information absent in image

The University of Auckland

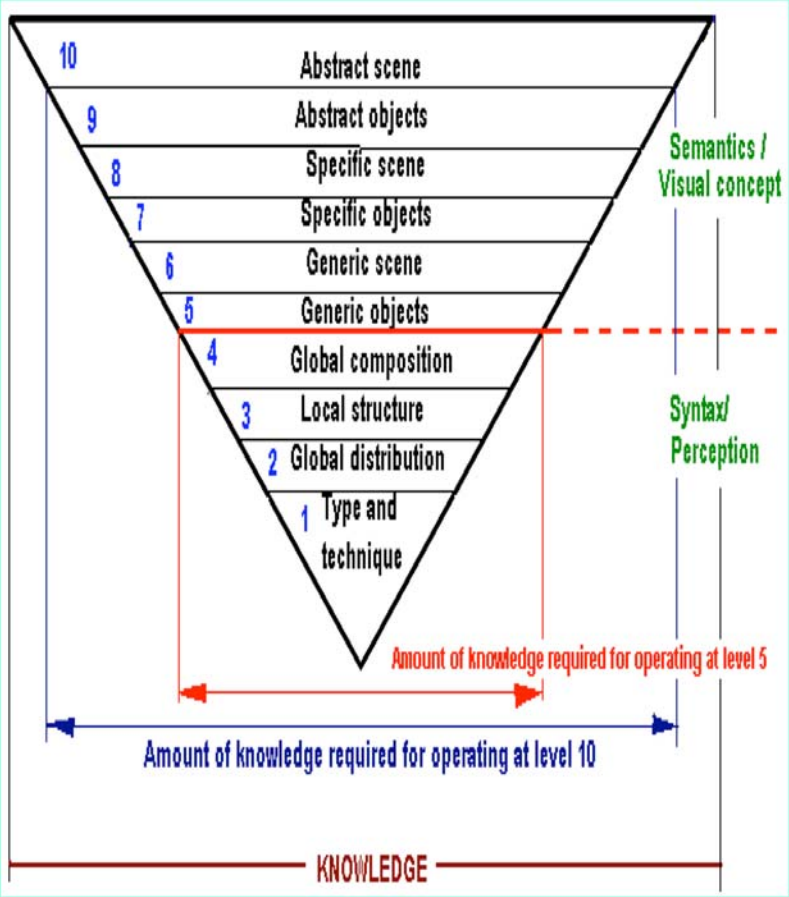
Pyramidal Indexing



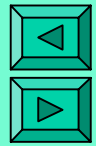


The University of Auckland

Pyramidal Indexing

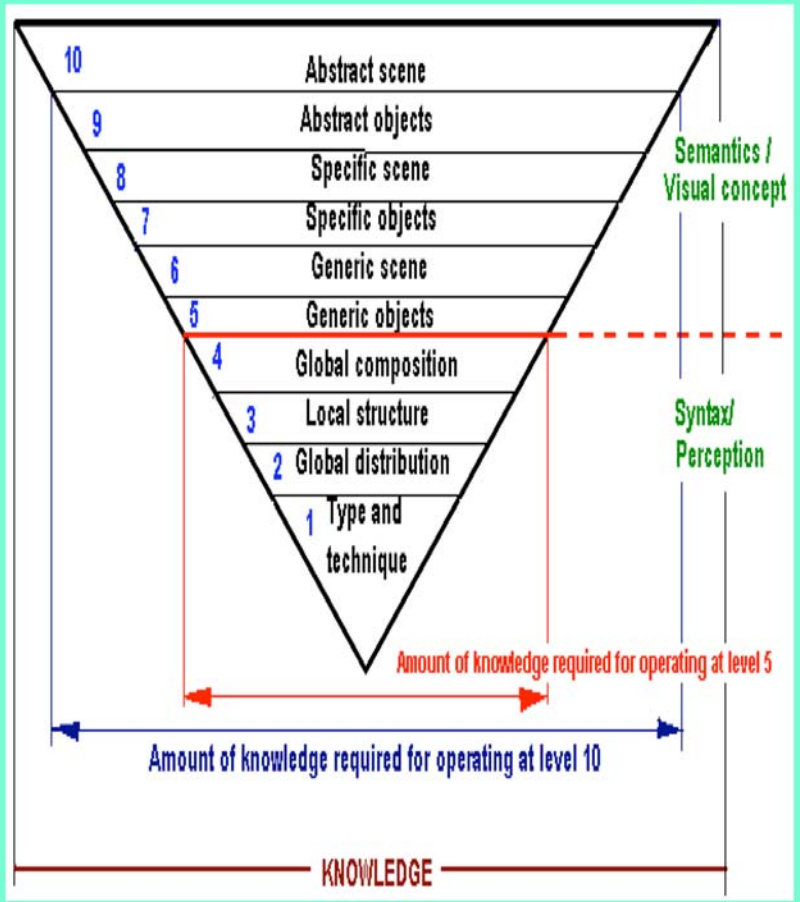


- **Four bottom perceptual levels**: no knowledge of actual objects to index an image; only low-level processing
- **Type/technique level**: general visual characteristics of images & video
- **Global distribution level**: description of a whole image (no components)
- **Local structure level**: basic syntactical symbols (by low-level processing)
- **Global composition level**: arrangement of basic elements and their groups in space



The University of Auckland

Pyramidal Indexing



Humans mainly rely on higher-level attributes to describe, classify, and search for images

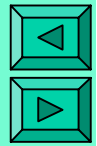
Generic objects level: attributes common to all of most members of a category (class)

Generic scene level: general knowledge of a whole image based on all its objects

Very restricted CBIR capabilities due to object detection / recognition limitations!

Specific objects / scene levels: knowledge of individual objects and their groups

Abstract objects / scene levels: subjective; the interpretative knowledge varies much



The University of Auckland

Pyramidal Indexing

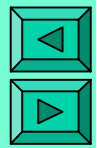
- Existing CBIR systems operate mostly at syntactic levels (i.e. **type and techniques**, **global distribution**, **local structure**, and **global composition**)
- Only few experiments have been done to take account of at least **generic** levels of semantics

	Specific	Generic	Abstract
Who	Mr. Kaplinski	Man	Serious
What action	Accessing a PC	Working	Investigation
What object	Compaq M300	Monitor, keyboard	Hi-tech
Where	Warsaw	Laboratory	Enterprise
When	March 3, 1998	Daytime	Summer
Why	Debugging	Work with computer	Professional hobby

The University of Auckland

MPEG-7 Multimedia Content Description Interface

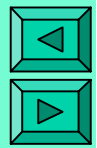
- ISO/IEC standard
  - Supports some degree of interpreting the information meaning to pass into / assess by a device / computer
  - developed by MPEG (Moving Picture ExpertsGroup)
    - MPEG also developed the Emmy Award winning standards MPEG-1 / MPEG-2, and the MPEG-4 standard
    - MPEG-1 and MPEG-2 standards made interactive video on CD-ROM and Digital Television possible
    - MPEG-4 is the multimedia standard for the fixed and mobile web enabling integration of multiple paradigms



The University of Auckland

MPEG-7 Multimedia Content Description Interface

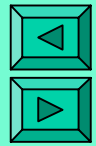
- MPEG-7 is not aimed at any particular application
- Elements that MPEG-7 standardizes support as broad a range of applications as possible
- **Goal**: fast and efficient search for multimedia data of interest to the user in spite of continuously increasing numbers of potentially interesting materials
- MPEG-7 provides semantic descriptions of multimedia data including still images, graphics, 3D models, audio, speech, video, and information about how these elements are combined in a multimedia presentation



The University of Auckland

MPEG-7 Data Description

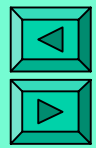
- Standard descriptors for colour, texture, shape, motion and other features of audiovisual data
  - Variety of tools to describe and structure multimedia information and facilitate its search
  - Standard means to define other descriptors, structures for descriptors, and their relationships to be associated with the content
  - The goal is to allow fast and efficient search for material of the user's interest



The University of Auckland

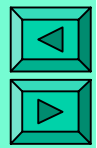
MPEG-7 Data & Features

- **Data**: audio and visual information to be described
  - Representation *regardless* of storage, coding, transmission medium, display, or technology (e.g. a video tape, an audio CD with music, speech, or sound, a picture printed on paper, or an interactive multimedia presentation on the Web)
- **Feature**: a distinctive characteristic of the data that signifies something for somebody
  - Examples: colour of an image, pitch of a speech segment, rhythm of an audio segment, camera motion in a video, style in a video, the title of a movie, etc



MPEG-7 Description Framework

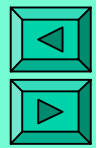
- Descriptions vary according to the data types
- Description framework:
  - a set of descriptors (D)
    - D: definition of the syntax and semantics of the feature representation
  - a set of description schemes (DS)
    - DS: specification of the structure and semantics of relationships between its components (Ds and DSs)
  - a language to specify DSs
    - DDL Description Definition Language: an XML (eXtensive Markup Language) to create new or extend existing DSs or Ds
  - one or more schemes for encoding the description



The University of Auckland

MPEG-7 Visual Description Tools

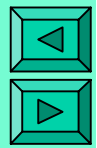
- Basic structures and descriptors (elementary and advances) for visual features
- Five basic structures:
  - **Grid layout** D: splits an image into rectangular regions of equal size
  - **Time series** D: a temporal sequence of Ds in a video segment
  - **2D / 3D multiview** D: combines 2D Ds of views of a 3D object's feature
  - **Spatial 2D coordinates** D: a 2D local or integrated coordinate frame to be used in other relevant Ds and DSs
  - **Temporal interpolation** D: polynomial interpolation of time-varying multidimensional values to reduce the D size



The University of Auckland

Visual Features for Semantic Contents

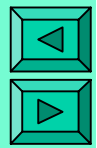
- **Colour Ds:** colour space, colour quantisation, dominant colours, scalable colour, colour structure, colour layout, and group of frames / group of pictures colour Ds
- **Texture Ds:** homogeneous texture, texture browsing, and edge histogram D
- **Shape Ds:** object region-based shape, contour-based shape, 3D shape, and 2D-3D multiple view D
- **Motion Ds:** camera motion, object motion trajectory, parametric object motion, and motion activity D
- **Localisation Ds:** region locator and spatiotemporal locator D
- **Others:** e.g. face recognition D



The University of Auckland

MPEG-7 Multimedia DSs (MMDS)

- Standard set of description tools (Ds and DSs) to deal with generic and multimedia entities
  - Generic features to all media: vector, histogram, time, etc
  - **Content description** DSs: representing perceivable data
  - **Content management** DSs: media features; the creation of the use of the audiovisual (AV) contents
  - **Content organisation** DSs: analysing several AV contents
  - **Navigation and access** DSs: summaries of the AV contents
  - **User interaction**: user preferences to the multimedia data



The University of Auckland

MPEG-21 Multimedia Framework

- A standard to define the description of content and also processes for accessing, searching, storing content and protecting the copyrights of content
- **Goal**: to enable the transparent and augmented use of multimedia resources across a wide range of networks and devices
- Basic elements of MPEG-21:
  - **Digital Items** (DI) - structured digital objects with a standard representation, identification, and description, or metadata
  - **Users** - all entities interacting with or making use of DIs