The
University
of Auckland

Experimental CBIR Systems

COMPSCI.708.S1.C

A/P Georgy Gimel'farb

The University of Auckland

Classification of CBIR Systems

- **Level of automation** of feature extraction and index generation
- **Level of integration** of multimedia modalities
- **Level of adaptability** to the needs of users and applications
- **Level of abstraction** in which content is indexed
- **Level of generality** of visual information domain
- **Level of automation** of the database collection
- **Abilities to semantically categorise** visual information
- **Abilities to process compressed visual information**

The University of Auckland

Level of Automation

- Interactive content-based querying: on the quantitative features of extracted and indexed images/videos
  - Low-level features (colours, textures, shapes, and motions) are extracted by automatic methods
  - Generation of higher-level semantic indexes usually requires human input and/or system training
- In a semiautomatic system the user selects manually image objects and features which are then used by the system for generating the feature indexes

The University of Auckland

Integration / Adaptability

- Multimedia modalities: images, video, graphics, text, audio → typically, independent indexing
  - Integration of multiple modalities: under investigation but it is not yet fully exploited
- Most systems use a static set of previously extracted features included into image metadata and indexes
  - The features are selected by the designer on the basis of indexing costs and search functionalities
  - Dynamic features extraction and indexing is needed to adapt to subjective and changing needs of users

The University of Auckland

Level of Abstraction

- Level of abstraction in which content is indexed:
  - ***low-level feature level*** (colour, texture, shape)
  - ***object level*** (moving foreground item)
  - ***syntactic level*** (video shot)
  - ***high-level semantic level*** (image subject)
  - **automatic indexing**: mostly low-level features
  - **higher-level indexes**: generated manually
- Interaction between the levels: still the unsolved problem

The University of Auckland

Generality / Data Collection

- Generality of the visual information domain:
  - specific domain knowledge: special feature sets (e.g., medical and remote-sensing applications)
  - generally: indexing unconstrained visual information, such as that on the Web
- Collection of the system's database:
  - **dynamic CBIR systems**: special software ***robots*** that automatically traverse the Web
  - **other systems** (*online news archives, photo stock houses, etc.*): manual addition of visual information

Categorisation / Processing

- Semantic classes of visual information
  - Navigation through the classes in large repositories of visual information is very useful
  - Effective categorisation is not yet developed
- **Compressed Data Processing**: direct extraction of the features from compressed images and videos in order to avoid expensive data decompression
  - compression standards (JPEG, MPEG) → features can be computed in the compressed domain

The University of Auckland

IBM: CBIR System "QBIC$^{(TM)}$"

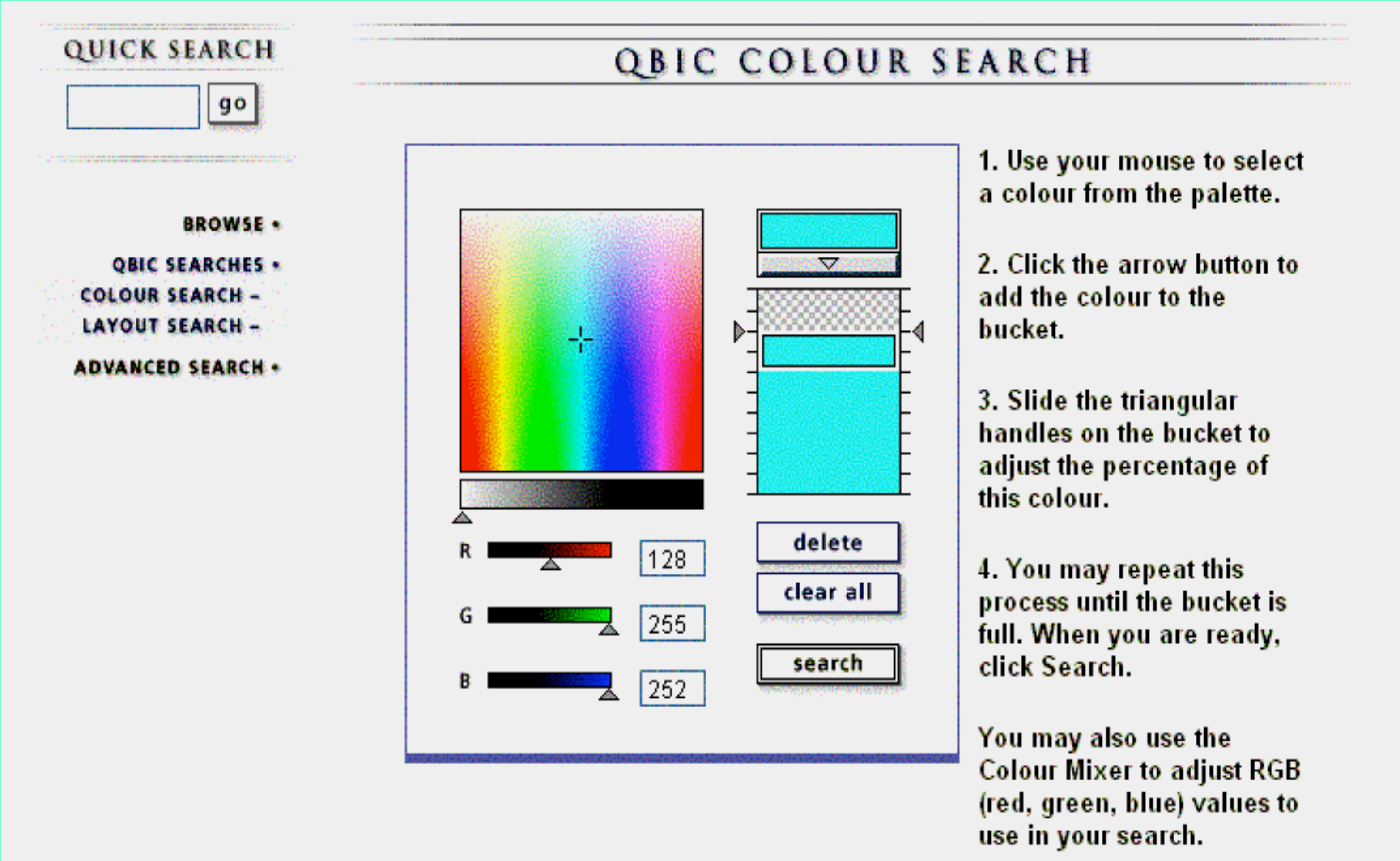
- **Q**uery **b**y **I**mage **C**ontent System
  - "…When keywords alone cannot locate that special "something" to fit a specific taste, users can turn to IBM's patented Query By Image Content or QBIC, a new search engine that sorts through database images according to colors, textures, shapes, sizes and their positions" (**from the IBM's description**)
- The first CBIR system exploring feature-based image and video retrieval

QBIC(TM) in the Hermitage

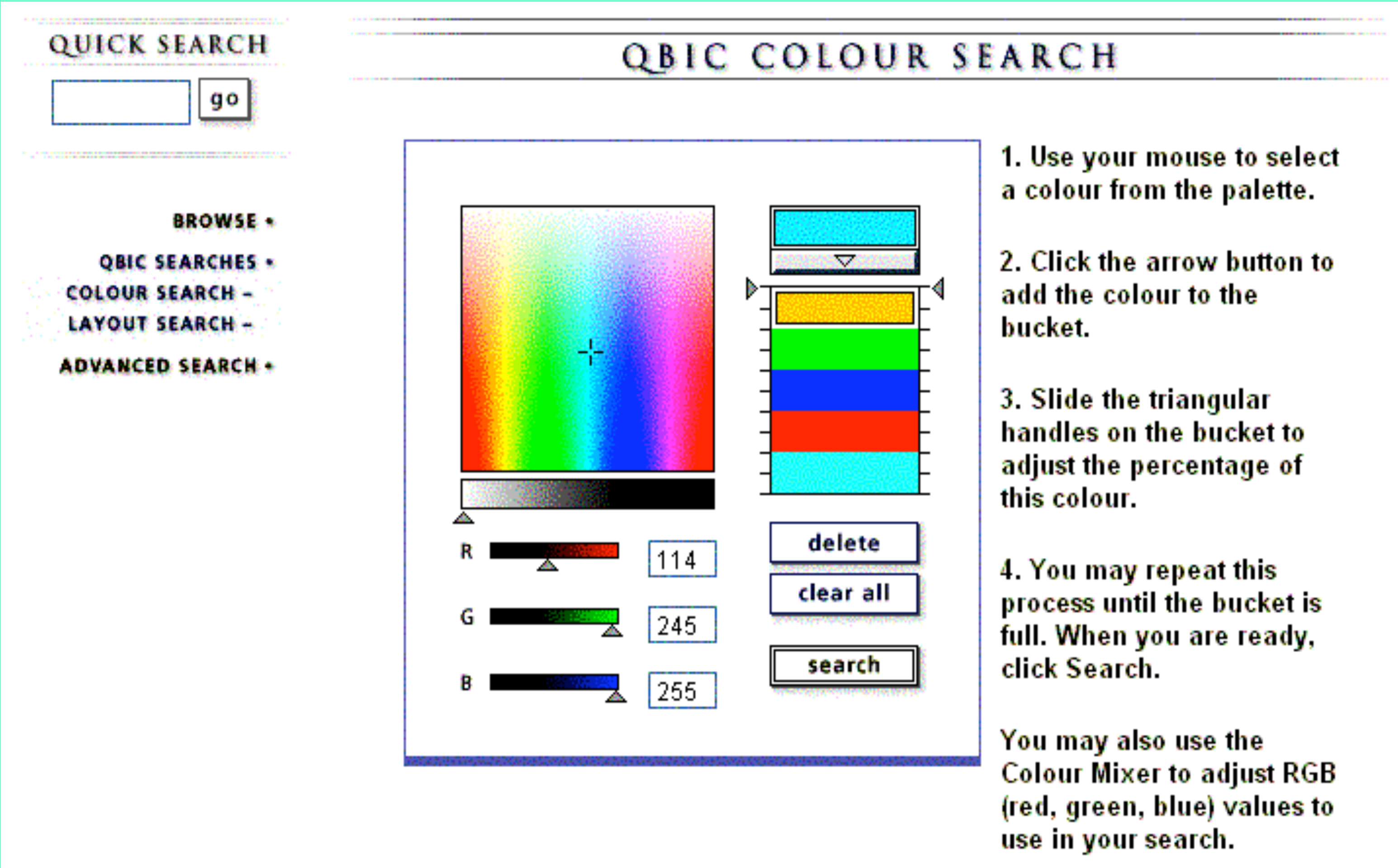
- The Hermitage (Russian Museum)
- **Browsing** of art categories: paintings, drawings, sculptures, arms, etc
- **Advanced Search**: by artist, title, subject, style, genre, theme, or date
- **QBIC**: to refine the search by requesting all artwork with comparable visual attributes

The University of Auckland

Colour Search Inteface in QBIC

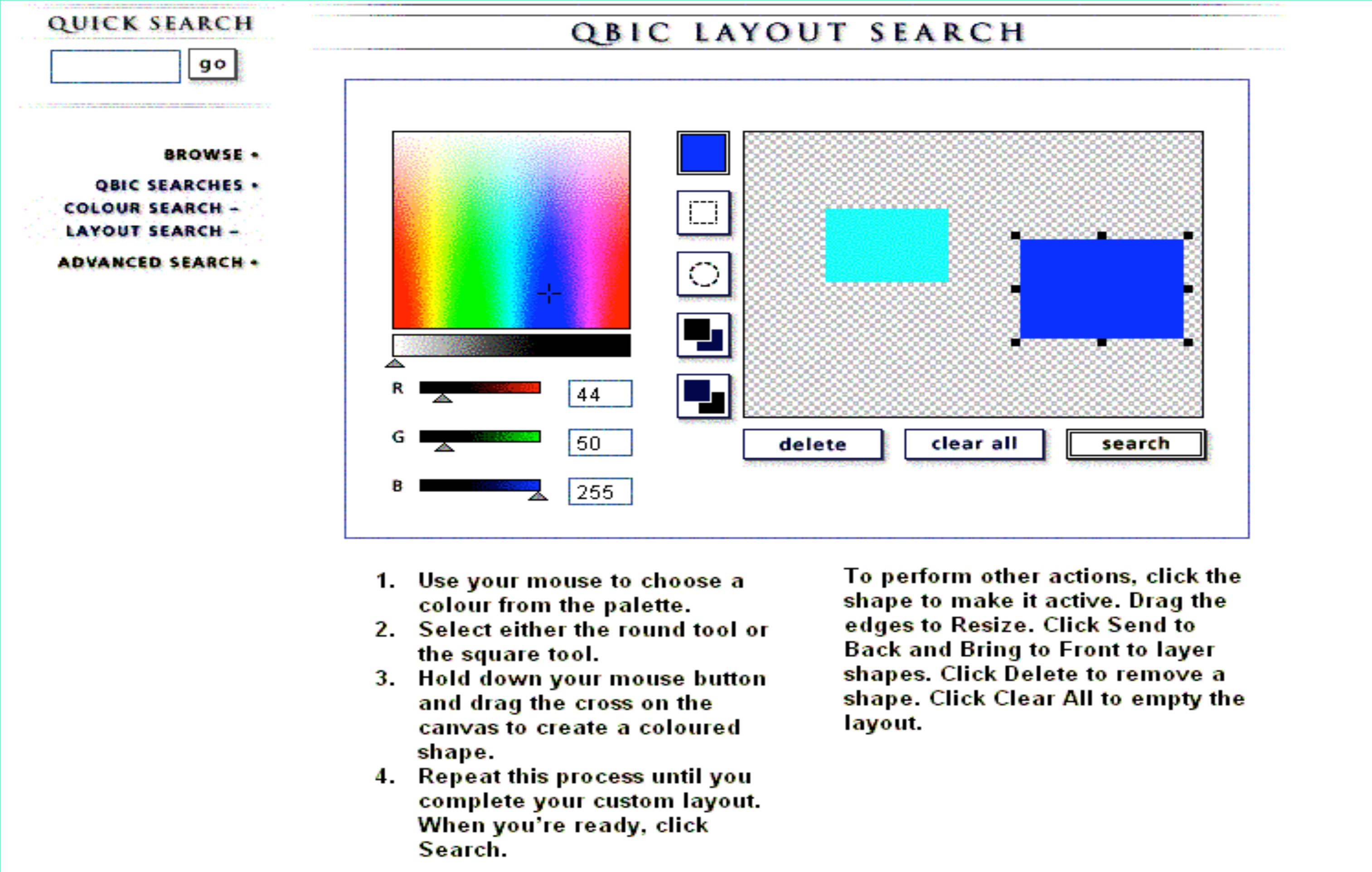


Colour Search Inteface in QBIC



The University of Auckland

Layout Search Inteface in QBIC



The QBIC Prototype System

- Two data types: a scene (image or video frame) and an object (a part of a scene)
- Objects are outlined manually and characterised by:
  - colour descriptors: average colour and histogram with 64 or 256 elements of the quantised colour space
  - modified Tamura's texture descriptors: coarseness, contrast, and directionality
  - shape descriptor (moments, heuristics, parametric splines with Hausdorff distance between the curves)

QBIC: Object / Scene Retrieval

- All objects in the images of a given database have to be outlined and features of all images and objects have to be computed at the stage of preparing the database for queries to be processed. Each query can be based on objects, entire scenes, or a combination of both.
- **Object-based query:** images containing objects with certain features such as colour percentage, colour layout, or/and texture.
- **Scene-based query:** full scenes with certain features (because the objects need not be outlined, the query interface is simpler)

The University of Auckland

QBIC: Object / Scene Retrieval

- Normalised similarity (distance) functions for each feature to be meaningfully combined
- **Weighted Euclidean distance** ⇒ for 3-component average colour vectors, 3-component texture vectors, and 20-component shape vectors
  - Weights - the inverse variances of each component in the database
- **Quadratic distance** $(\mathbf{X}-\mathbf{Y})^T\mathbf{S}(\mathbf{X}-\mathbf{Y})$ ⇒ to match colour histograms **X** and **Y**
  - $\mathbf{S}$: a symmetric colour dissimilarity matrix with the components $\mathbf{S}(i,j)$ indicating the dissimilarity of the colours i and j in the histogram

QBIC: Object / Scene Retrieval

- Pre-computed stored features are compared to a given query in order to determine which images match it
- **Filtering and R-trees** ⇒ to index the feature vectors in the database
  - 64-element colour histogram: access by filtering
  - 3D average colour: access by an R-tree
  - 20-dimensional moment-based shape vector ⇒ PCA to map onto a 3D space to use an R-tree index

MIT Media Lab: Photobook

The University of Auckland

- Queries based on features related to particular models fitted to each image
- Commonly: colour, texture, and shape models, although features from any model may be used (e.g., results of face recognition)
- Features comparisons with a library of matching algorithms:
  - Euclidean or Mahalanobis distance
  - Kullback-Leibler divergence
  - Vector space angle (correlation)
  - Histogram distance
  - Peaks in Fourier space
  - Wavelet tree distance, or
  - User-defined ones, as well as any linear combination of these

The University of Auckland

MIT Media Lab: Photobook

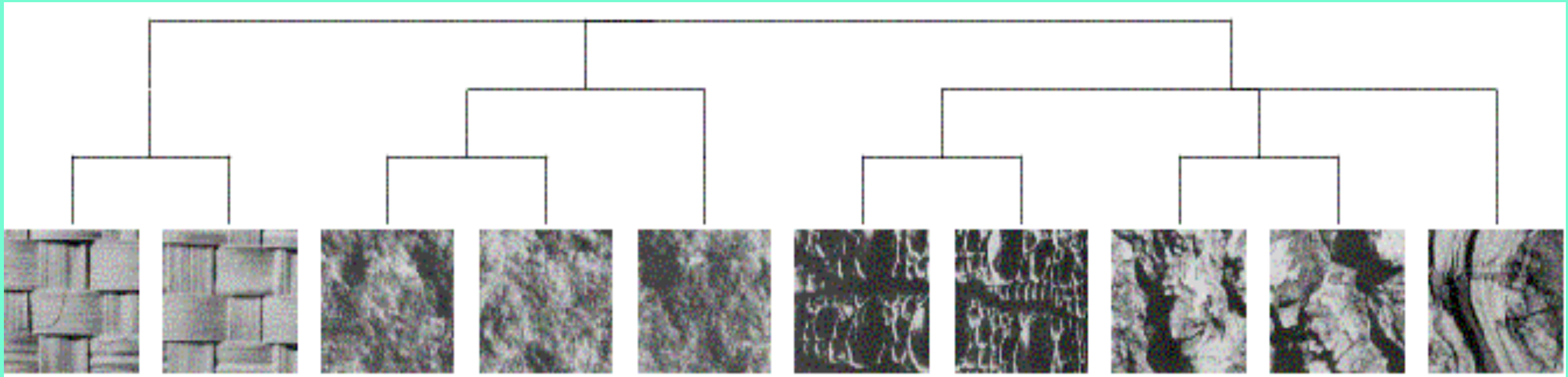
Query for images similar to the upper left image

Photobook + FourEyes

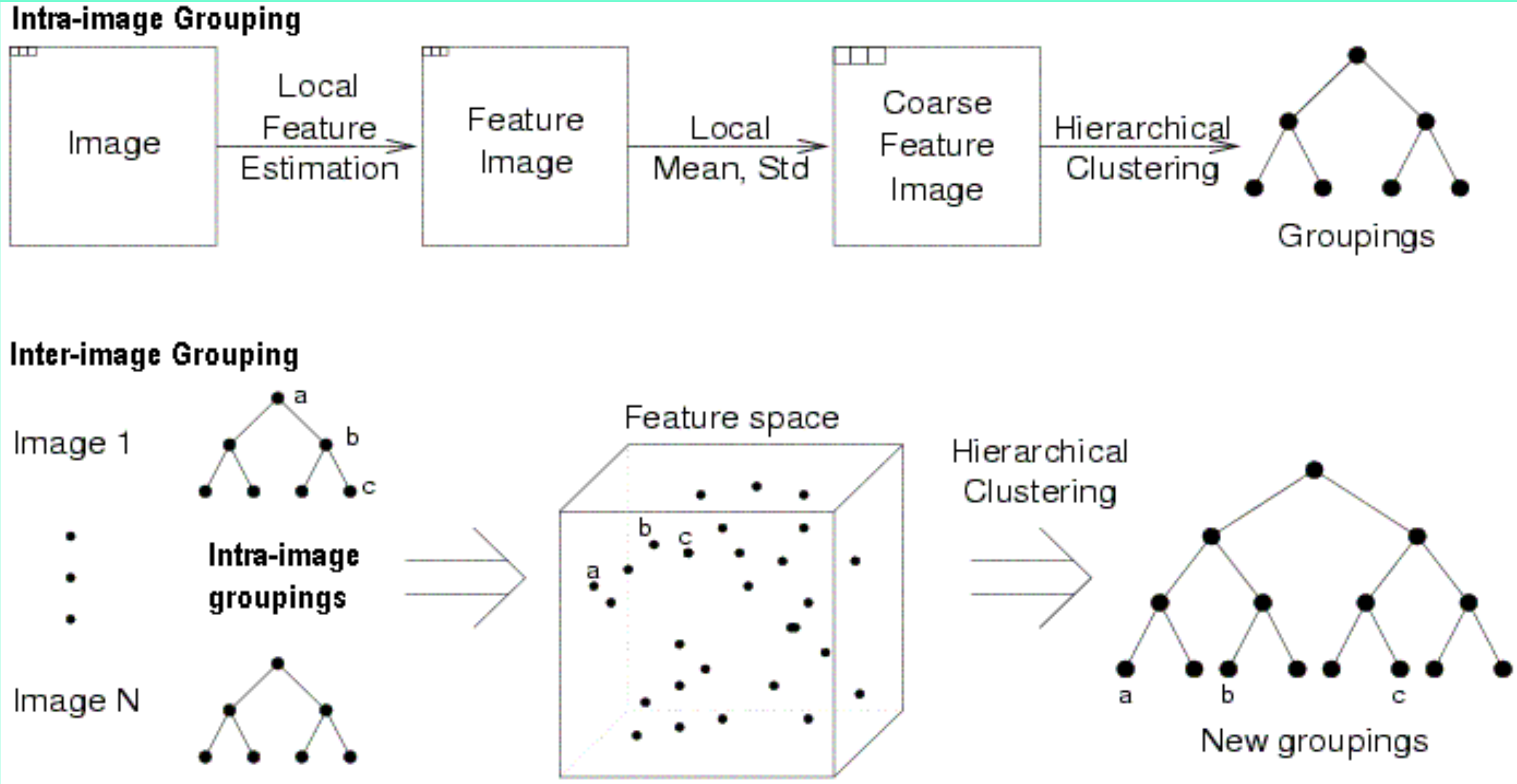
- Interactive learning subsystem
  - selecting and combining models on the basis of positive and negative examples from the user
- Due to the learning abilities of FourEyes, Photobook learns to select and combine features to satisfy a query
  - **Continuous learning**: each retrieval session adds to improvement of the system's performance
  - Learning to select and combine similarity measures, rather than design and understand them

Data Grouping in FourEyes

- The data are first organised by grouping (clustering)
- The grouping allows for different similarity measures and can be produced manually, with due account of colour / texture models, optical flow information, etc:



Data Grouping in FourEyes



The University of Auckland

Data Grouping in FourEyes

- **Intra-image** (within-image) groupings and **inter-image** (across-image) groupings composed of these:
  - Let grouping "a" in a single image contain "b", which in turn contains "c"
  - In the feature space the groupings are individual, so that the resulting clustering may specify "a" looks more similar to "b" rather than to "c"
  - The user specifies queries by examples and relates most appropriate similarity measures to different parts of the query

Computer-assisted Annotation

- The user has selected a few sky patches in the two right images and labelled them "sky“
- ***Intra-image grouping***: the labelled patches grow into larger cross-hatched "sky" regions
- ***Inter-image grouping*** places tentative labels on the two left images

Computer-assisted Annotation

- By pointing out the false labels, the user provides the negative examples to refine the decision rule
- Three decision strategies:
  - **Set covering** (SC): the smallest union of sets in the feature space which covers all of the positive examples and none of the negative ones
  - **Decision list** (DL): sets of either positive but no negative or negative but no positive examples
  - **Decision tree** (DT): the best division of examples into two parts for a possible further division