Mining Image Features for Efficient Query Processing

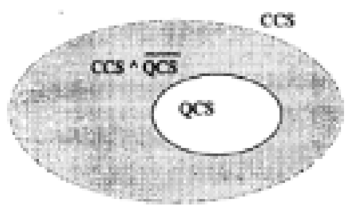
Beitao Li, Wei-Cheng Lai, Edward Chang,
and Kwang-Ting Cheng

ICDM 2001 (pp: 353 –360)

Presented by: Yi-Hung Wu
Date: 2002/6/6

Preliminaries

- **Perception-Based Image Retrieval [SIGMOD01]**
  - Query concept learner
  - Multi-resolution image characterization
- **Learning Image Query Concepts via Intelligent Sampling [ICME01]**
  - Select samples by MEGA
  - Solicit user feedback
  - Refine QCS by Vote
  - Refine CCS by Vote



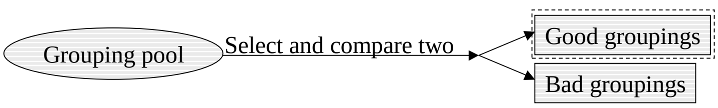
Problems

- **Query Concepts: k-CNF**
  - $C_1 \wedge \ldots \wedge C_\theta$, where $C_i=(x_1 \vee \ldots \vee x_j)$ for $j \leq K$
  - Scalability of MEGA
    - Dimensionality-curse: number of disjunctions $\sum_{i=1}^{k} \binom{M}{i}$
    - Divide-and-conquer: trade precision for speed
      - Speed up of $O(G^{k-1})$ folds: M=144, k=3, G=12 → 140 times
  - Genetic algorithm for mining feature groupings
    - Mapping between groupings and individuals
    - Genetic operators: *selection, crossover, mutation*
    - Fitness function

P.3

Solutions (1/2)

- **A Feature Grouping**
  - Exactly M features and no feature is replicated
- **Genetic Operators**
  - Tournament selection without replacement
  - Randomly exchange two features from two different feature groups for a given grouping
  - Randomly move a feature from one to another feature group according to a given probability



P.4

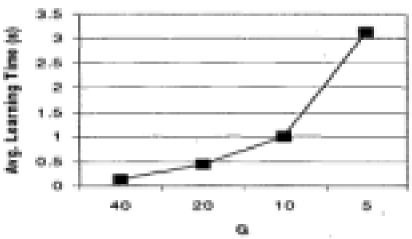
Solutions (2/2)

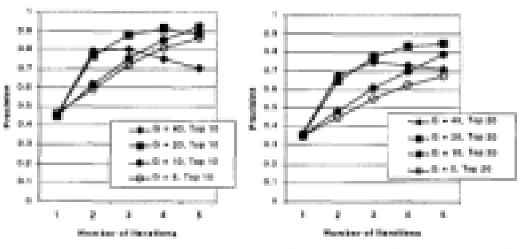
- **Fitness Function ≡ Search Accuracy**
  - 51000 images of 18 categories (concepts)
  - Use the learned concepts for top-n image retrieval
- **Mining Algorithm**
  - Initialize N groupings
  - Compute the fitness value for each grouping
  - Apply the genetic operators in series
  - Continue the next generation if necessary

P.5

Experimental Results (1/3)

- **Tradeoff between Learning Time and Accuracy**
  - Find moderate G=20



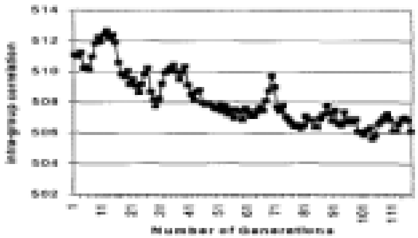


(a) Top-10 Retrieval (b) Top-20 Retrieval

P.6

Experimental Results (2/3)

- **Discovering Optimal Groupings**
  - Low intra-group correlation leads to high search precision



P.7

Experimental Results (3/3)

- **Feature - Concept Associations**

	Channels	*Ar'*	*Be'*	*Cl'*	*El'*	*Fa'*
Colors	Red color	–	–	–	–	+
	Yellow color	–	+	–	+	–
	Blue color	+	–	+	–	–
	Brown color	–	–	–	+	–
Textures	Vertical coarse		–	–	–	+
	Horizontal coarse	+	+	–	+	+
	Diagonal coarse	+	–	–	+	+
	Vertical medium	+	–	–	+	+
	Horizontal medium	+	–	–	+	–
	Vertical fine	+	–	–	+	+
	Horizontal fine	+	–	–	–	–
	Diagonal fine	+	–	–	–	+

P.8

Concluding Remarks

- **Solving Two Mining Problems**
  - Learning query concepts from user feedback
  - Discovering optimal feature groupings
- **Goodness**
  - Identify feature - concept associations
- **Weakness**
  - Genetic algorithm is not efficient but effective
  - No measure for missing concepts

P.9

Paper Scoring

- **Scores {bad, marginal, good, excellent}**
  - Originality: good
  - Technical Depth: marginal
  - Impact/Practicability: good
  - Readability: marginal
  - Overall: good

P.10