A Bayesian Network Approach to Multi-feature Based Image Retrieval

Authors: *Qianni Zhang and Ebroul Izquierdo*

Affiliation: *Multimedia and Vision Research Lab*

Department of Electronic Engineering

Queen Mary, University of London

Queen Mary University of London

Multimedia and Vision Research Group

Introduction

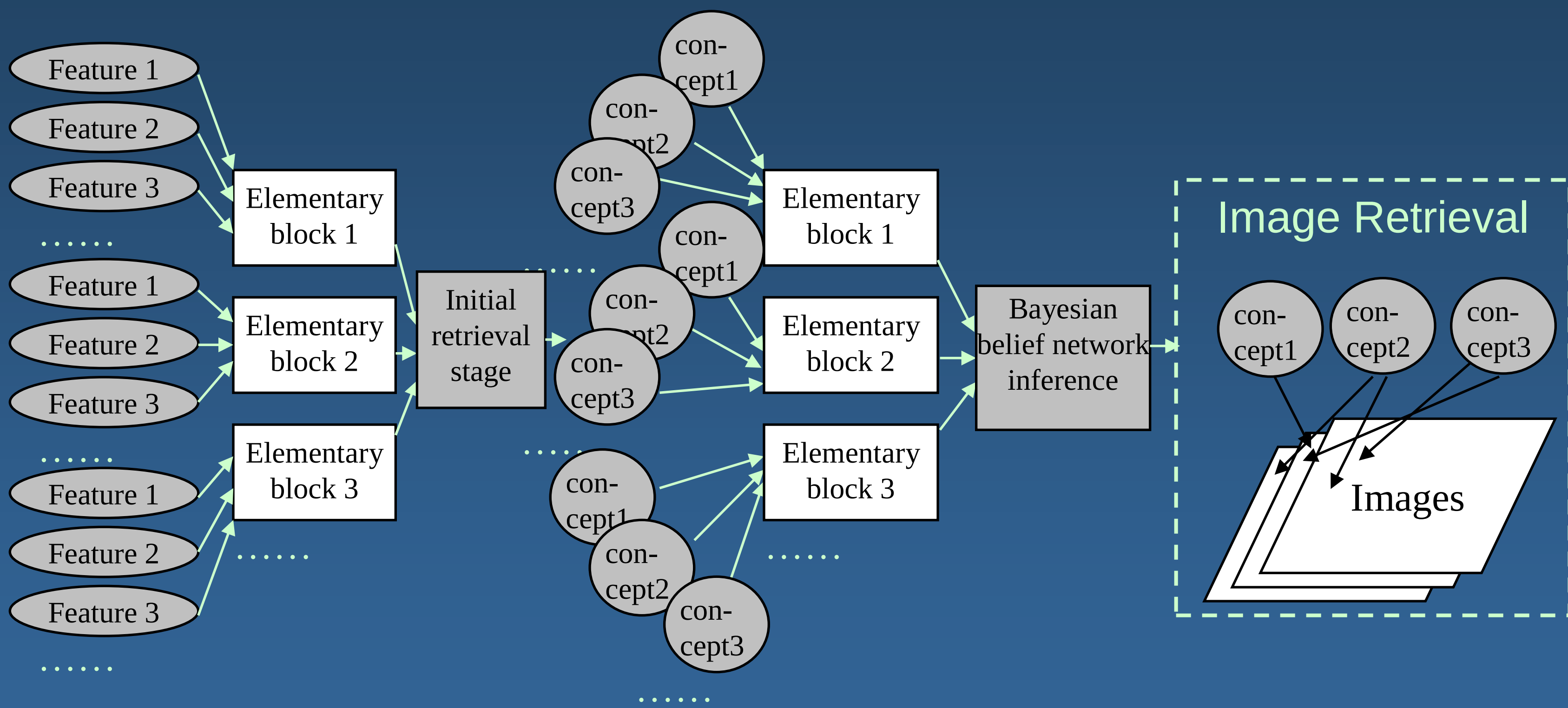
- Aims:
  - Devising a Bayesian Network approach to object centered image retrieval
  - Combining multiple low-level visual primitives as cue for retrieval.

- Two stages:
  - the initial retrieval stage is concentrated on finding an optimal multi-feature space and doing a simple initial retrieval within this space;
  - the Bayesian inference stage uses the initial retrieval information and seeks for a more precise second-retrieval.

Introduction

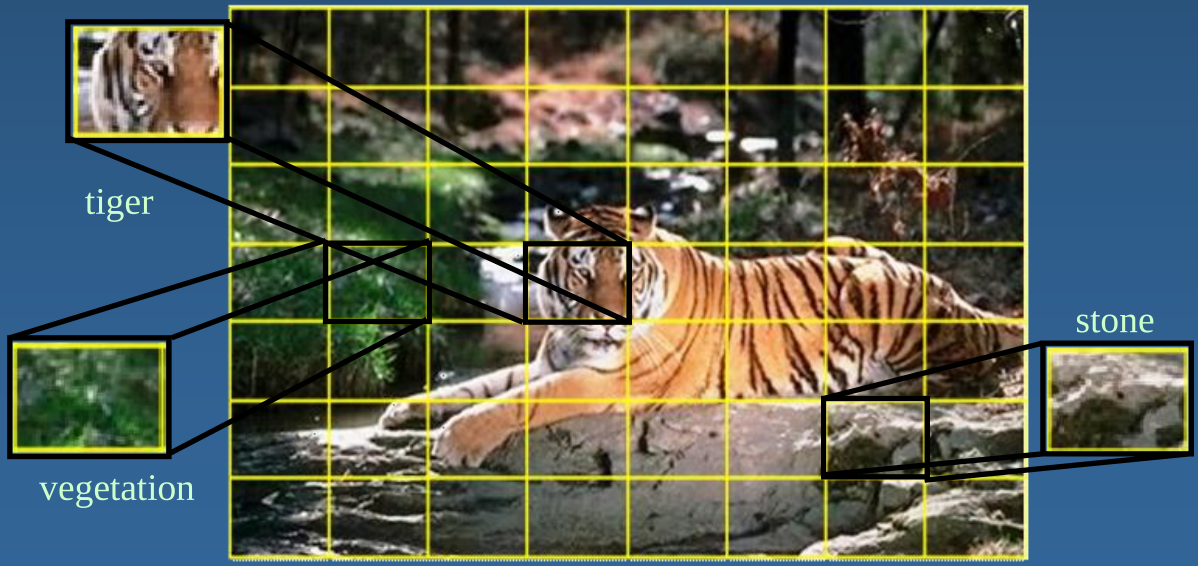
- Originalities:
  - The beliefs are formulated regarding concepts in possibly small regions of the entire image - "elementary building blocks"
  - Multi-Objective Optimization (MOO) technique is adopted for estimating the 'optimal' multi-feature metric space
  - Initial beliefs on probability distributions of concepts is modelled by the initial retrieval information
  - A global knowledge network is constructed by treating an entire image as a scenario to infer the presence of objects from the interactions between different concepts on image level.

Framework Overview



Initial Retrieval in an Optimized Multi-Feature Space

- Semantic image retrieval relies on the retrieval of semantically meaningful objects within the image
- An example of one image being divided into elementary building blocks that contain different concepts is illustrated in figure below



Feature Extraction and Distance Calculation

- MPEG-7
  - Colour Layout Descriptor (CLD)
  - Colour Structure Descriptor (CSD)
  - Dominant Colour Descriptor (DCD)
  - Edge Histogram Descriptor (EHD)
- Others
  - Gabor Filter (GF)
  - Gray-Level Co-occurrence Matrix (GLCM)
  - HSV Histogram (HSV)

- Most low-level visual descriptors show non-linear behaviors and their direct combination is meaningless.
- Thus in this paper a combination of distances with certain metric is used as a similarity measurement.

Feature Extraction and Distance Calculation

- A distance function:

- A training distance matrix can be constructed on a group of '*representative building blocks*'

- *Min-Max Normalization:*

Combining Distances and Constructing a Multiple Feature Space

- To combine the distances in different feature spaces for a element, the most straightforward candidate is the linear combination of the distances:

$$M(\mathrm{A}, D) = \begin{cases} \alpha_1 {d_1}^{(1)} + \alpha_2 {d_2}^{(1)} + \alpha_3 \\ \alpha_1 {d_1}^{(2)} + \alpha_2 {d_2}^{(2)} + \alpha_3 \\ \ldots. \end{cases}$$

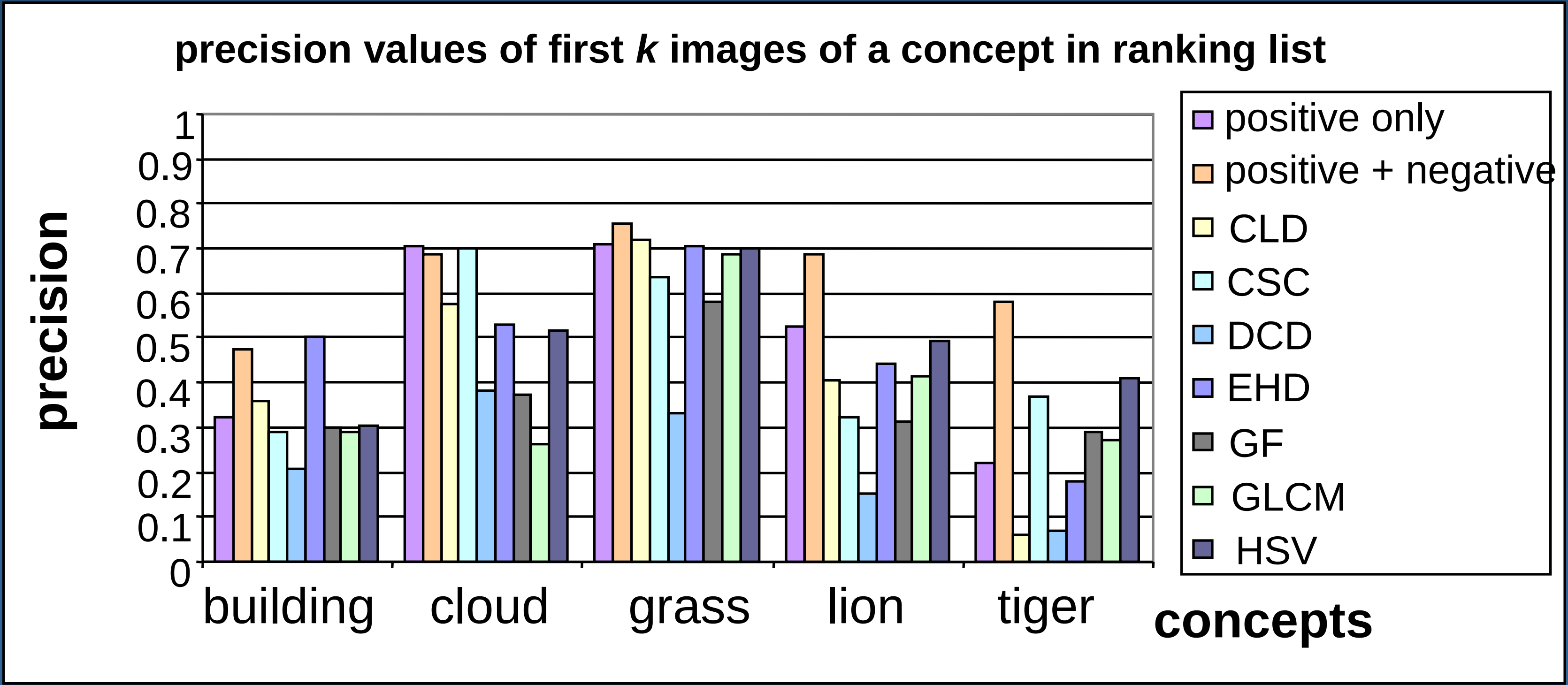
  - A is the set of weighting factors
  - *D* is the set of distance functions for single descriptors.
- The problem of finding the suitable metric consists of finding the optimal set of weighting factors $\underline{\alpha}$, where optimality is regarded in the sense of both concept representation and discrimination power.

Combining Distances and Constructing a Multiple Feature Space

- In the group of ‘representative building blocks’, each block in the representative group is used as an objective function
- Optimization can be achieved by minimizing the objective functions
- In most cases there is no way to optimize all objective functions simultaneously
- Multi-Objective Optimization (MOO) usually involves conflicting objectives
- The interaction between different objectives leads to a set of compromised solutions, largely known as the Pareto-Optimal Solutions or Pareto Front

Combining Distances and Constructing a Multiple Feature Space

- The optimal solution is to find the minimal value of <u>*M*</u> and its corresponding <u>*α*</u>, subject to constraint $\sum \alpha_j$
- The initial retrieval is done in this space of <u>*α*</u> and if any elementary block of an image is classified as relevant, the entire image is classified as relevant.



- *k* is the number of images containing a given concept in the database according to ground truth

Bayesian Network Inference

- Decisions are inferred using Bayesian networks that are conventional directed acyclic graphs with conditional probability distributions
- All the probabilities used in the Bayesian Network are computed from information in the belief ontology which is created using the initial retrieval results.
- For a particular concept user has in mind, each image in database can be classified into two classes: “relevant” or “irrelevant”. The two possible classes are denoted as C_k, where $k \in 1,$
- In this paper C_1 corresponds to relevant and C_2 corresponds to irrelevant

Queen Mary University of London
Multimedia and Vision Research Group

Bayesian Network Inference

- The prior probability of class membership is denoted as $P(C_k$
- The features used to help the inference are denoted as a set F, and $P(\mathbf{F}$ is the evidence factor
- Inferences are based on the posterior probability function $P(C_1 \mid$
- Bayes law: $P(C_1 \mid \mathbf{F}) = \frac{P(\mathbf{F} \mid C_1}{P(\mathbf{F}}$
- The classification criterion used is *maximum a posteriori* (MAP) given by: $C_1 = \arg\max P(C$

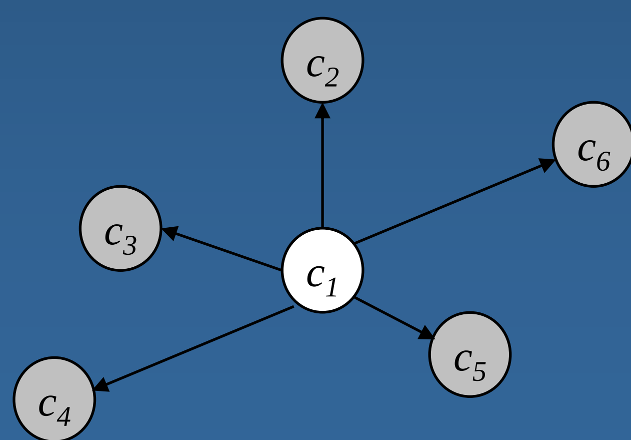
Queen Mary University of London
Multimedia and Vision Research Group

The Belief Ontology and Bayesian Network

- In this paper the belief ontology is modeled using Bayesian belief network
- They have similarity that the nodes represent propositions which are either true or false and has probabilities associated with co-occurrence relationships
- However, the co-occurrence relationships between concepts are <u>not causal</u> and the probabilities kept with these relationships simply measure statistical association.
- Why Bayesian networks:
  - A Bayesian network is naturally capable of encoding the joint probability distribution, it is considered as a representation of ontology
  - It is also an inference engine that can exploit information contained in interrelationships and dependencies between elements

Constructing the Bayesian Networks

- A small ontology containing concepts of objects that are typical in the experimental database is first pre-defined
- For each concept in the belief ontology containing p concepts $c_t, t = 1, \ldots$, a Bayesian network is constructed by considering this concept c_1 and all other concepts that are directly linked to c_1.
- A Bayesian network such as the one shown below can be constructed:



Experimental Setup

- 700 images from 'Corel' dataset
- Descriptors:
  MPEG7: CLD, CSD, DCD, EHD,
  Others: GF, GLCM, HSV
- Concepts:
  'Building', 'Cloud', 'Grass', 'Lion', 'Tiger'.
- Number of images for each concept:
  141, 264, 279, 100, 100

Evaluation

Initial retrieval result – using optimal multi-feature metric

%	Combination metric	CLD	CSC	DCD	EHD	GF	GLCM	HSV
building	70	48	24	20	**74**	40	38	42
cloud	**79**	76	70	38	68	28	34	78
grass	**92**	**92**	86	28	82	64	88	88
lion	**88**	50	36	16	50	24	40	66
tiger	**60**	2	46	7	14	26	34	57

Retrieval result using Bayesian Network

%	building	cloud	grass	lion	tiger
Initial results	70	79	92	88	**60**
Bayesian net	**72**	**84**	**94**	**92**	**60**