ECE271A – Fall 2007

Content based Image Retrieval

(at SVCL)

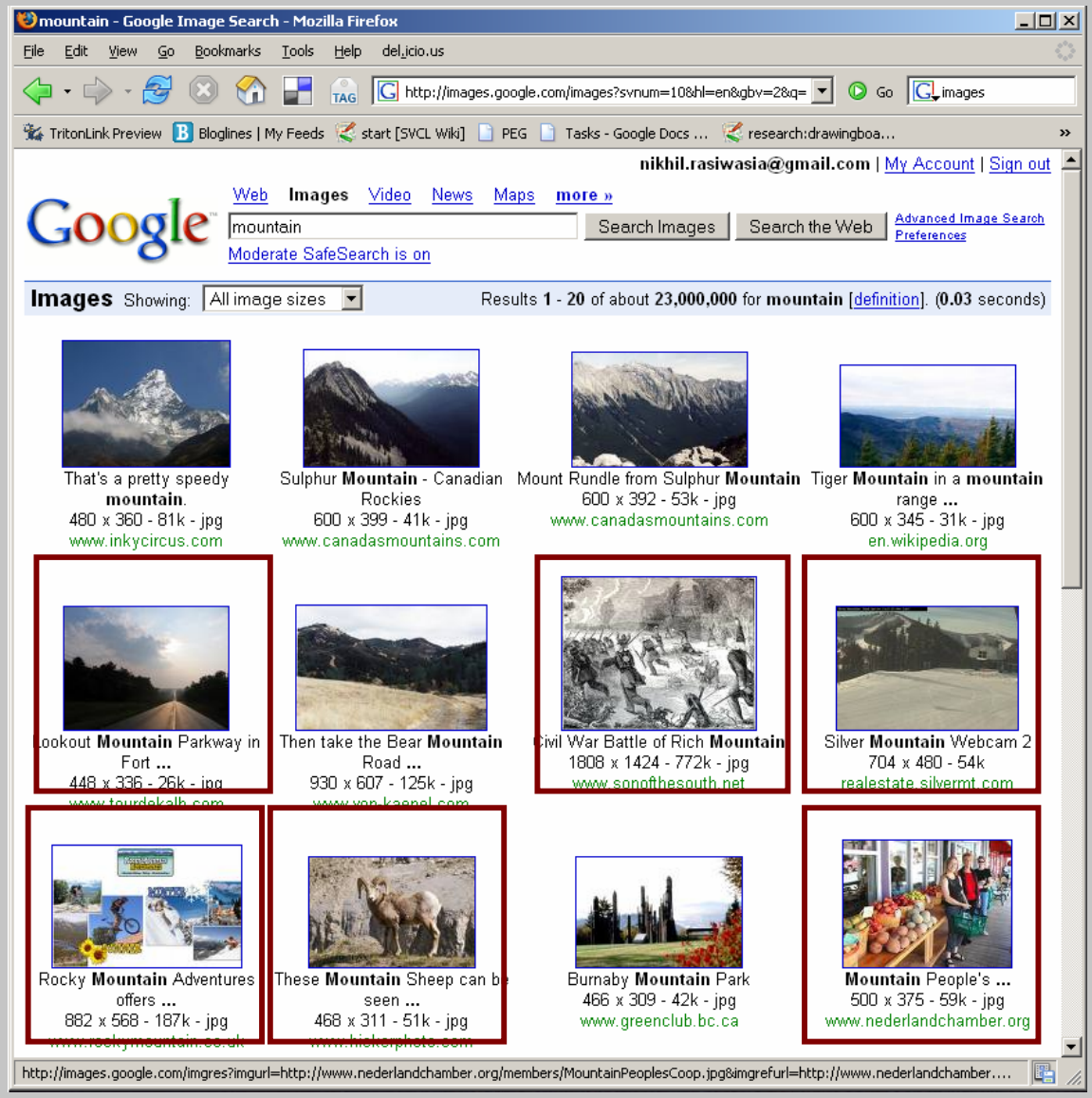
Nikhil Rasiwasia, Nuno Vasconcelos
Statistical Visual Computing Laboratory
University of California, San Diego

Image retrieval

- Metadata based retrieval systems
  - text, click-rates, etc.
  - Google Images
  - Clearly not sufficient

- what if computers understood images?
  - Content based image retrieval (early 90's)
  - search based on the image content

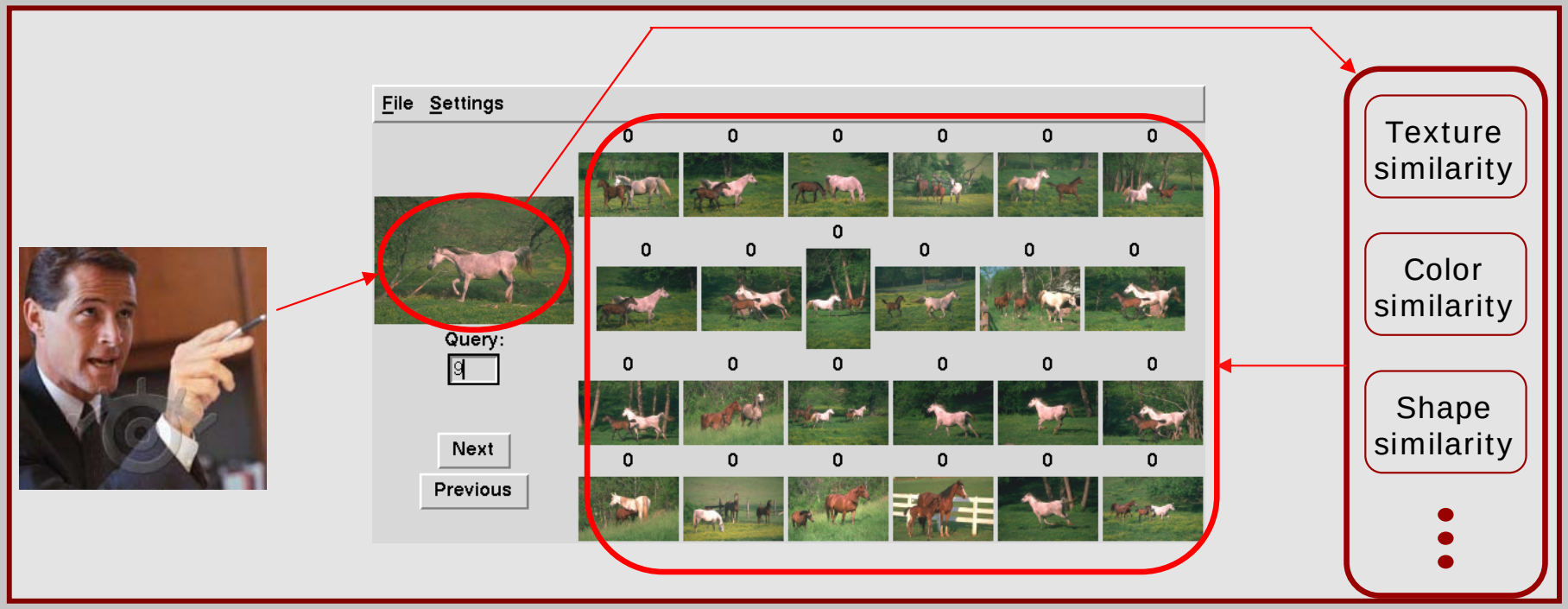
Metadata based retrieval systems



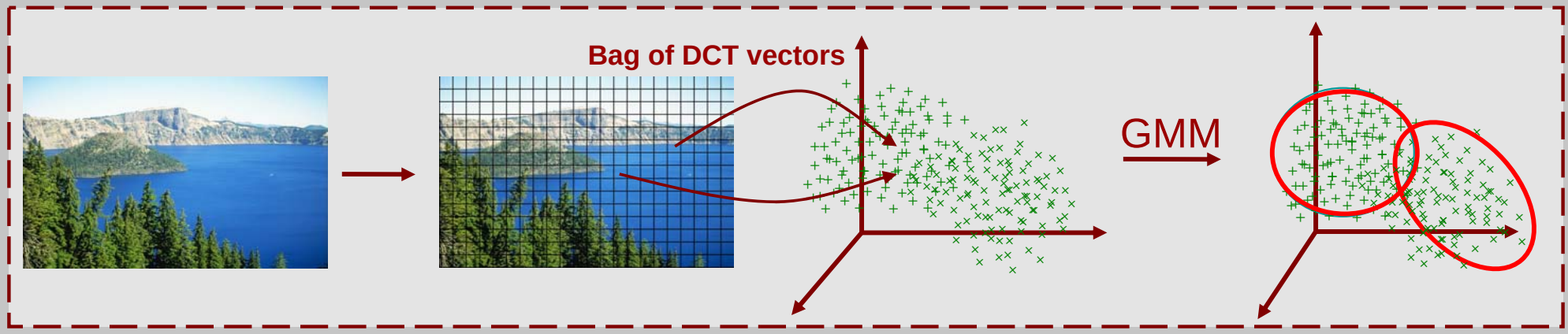
Top 12 retrieval results for the query 'Mountain'

Content based image retrieval -1

- Query by Visual Example(QBVE) 
  - user provides query image
  - system extracts image features (texture, color, shape)
  - returns nearest neighbors using suitable similarity measure



Query by visual example

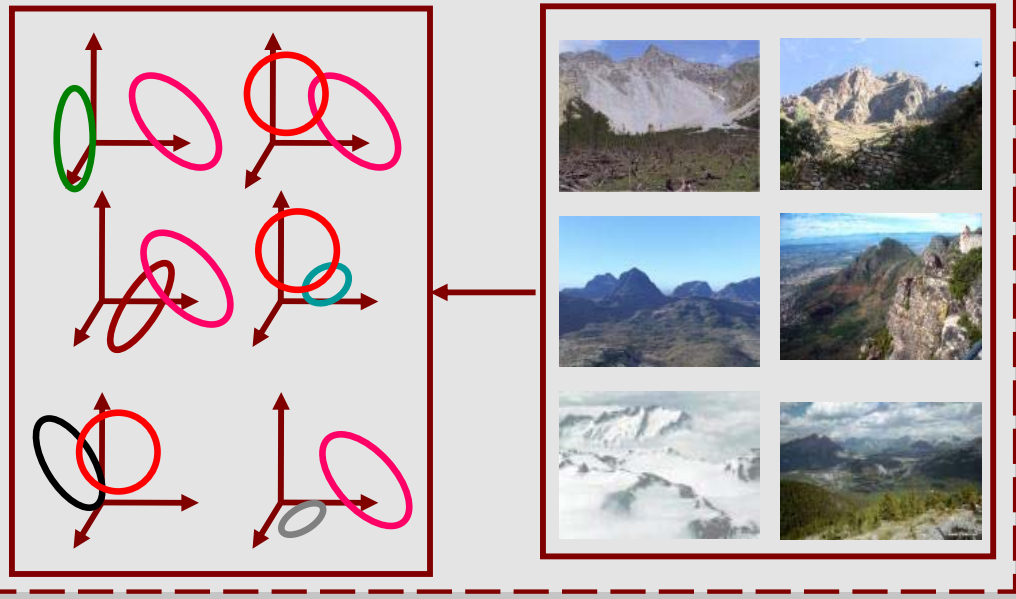


Query Image

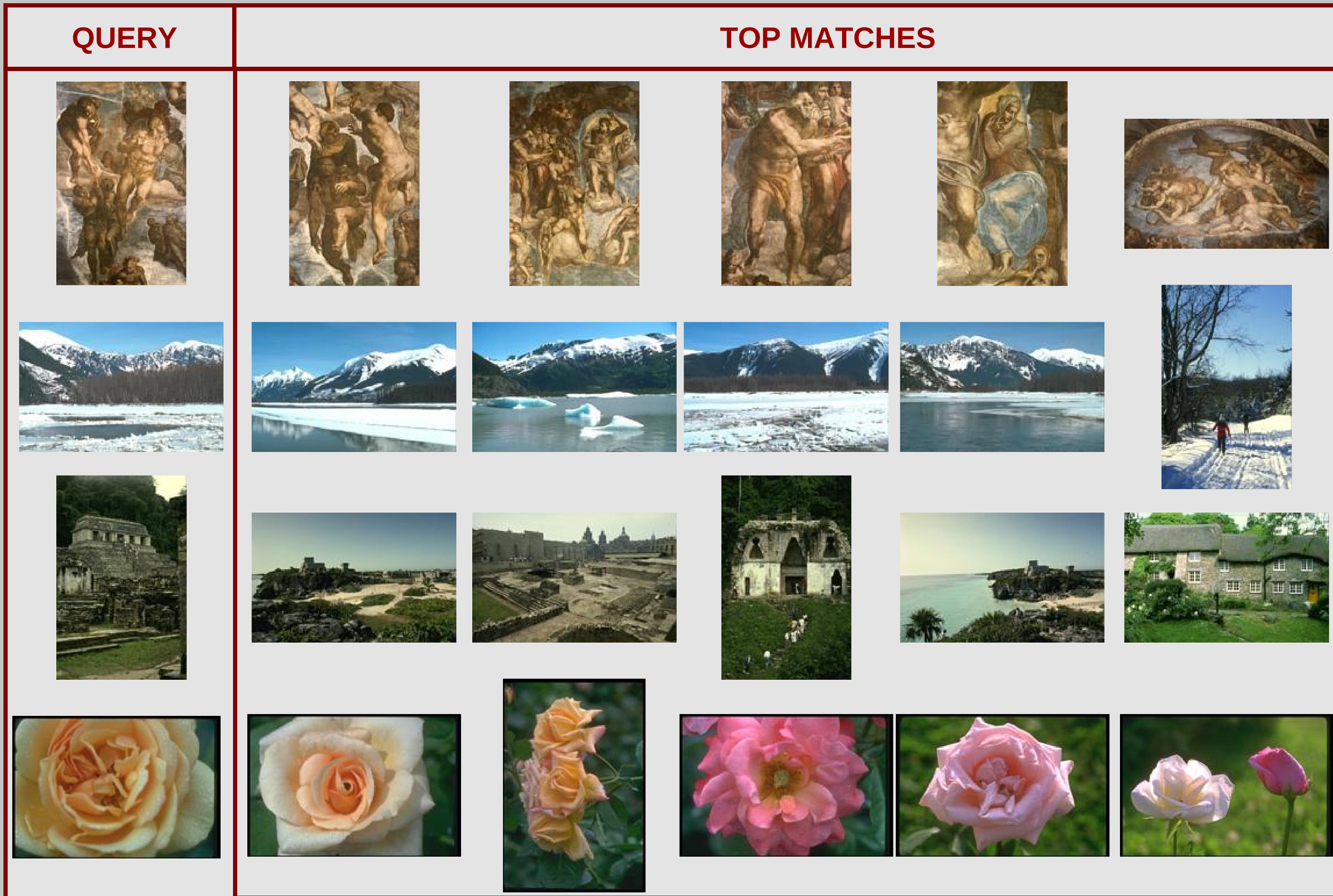
Probability under various models

Ranking

Candidate Images



Query by visual example (QBVE)



Query by visual example (QBVE)

- visual similarity does not always correlate with "semantic" similarity

Disagreement of the semantic notions of train with the visual notions of arch.

Both have visually dissimilar sky

Content based image retrieval -2

- Semantic Retrieval (SR)

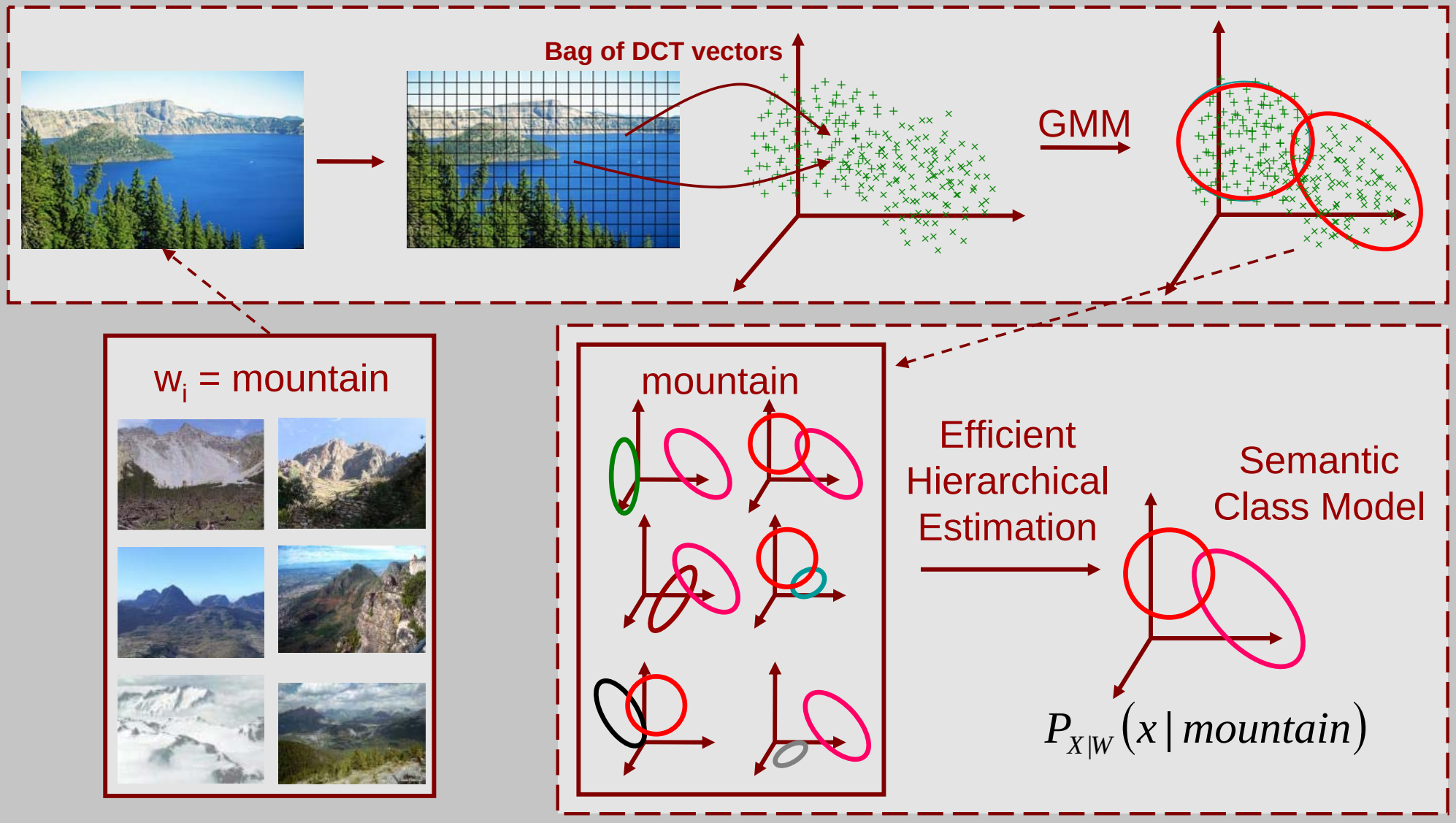
abc

  - User provided a query text (keywords)
  - find images that contains the associated semantic concept.

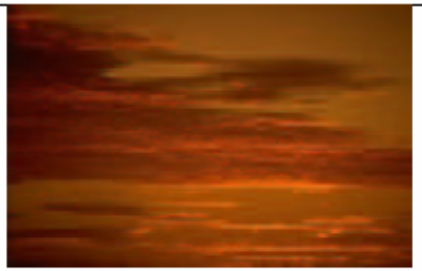
query: “people, beach”

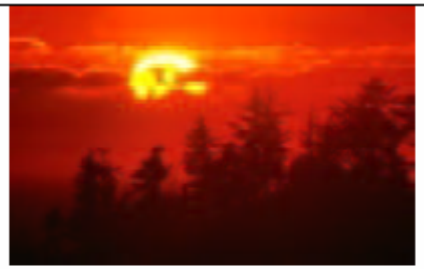
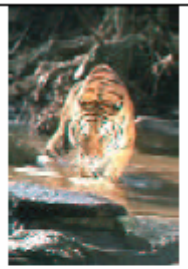
  - around the year 2000,
  - model semantic classes, learn to annotate images
  - Provides higher level of abstraction, and supports natural language queries

Semantic Class Modeling



- *"Formulating Semantics Image Annotation as a Supervised Learning Problem"* **[G. Carneiro, IEEE Trans. PAMI, 2007]**

Model				
Human Annotation	sky jet plane smoke	bear polar snow tundra	water beach people sunset	buildings clothes shops street
Mix-Hier Annotation	smoke clouds plane jet flight	polar tundra bear snow ice	sunset sun palm clouds sea	buildings street shops people skyline
Model				
Human Annotation	grass forest cat tiger	coral fish ocean reefs	mountain sky clouds tree	leaf flowers petals stems
Mix-Hier Annotation	cat tiger plants leaf grass	reefs coral ocean fan fish	mountain valley sky clouds tree	petals leaf flowers lily stems
Model				
Human Annotation	sky jet plane smoke	sky clouds formation sunset	snow fox arctic	water boats waves
Mix-Hier Annotation	plane jet smoke flight prop	sea sun sunset waves horizon	arctic snow polar fox ice	coast waves boats water oahu

Model				
Human Annotation	sky sun clouds tree	city sun water	water rocks cat tiger	coral ocean reefs
Mix-Hier Annotation	sun sea sunset clouds horizon	sun sunset city horizon clouds	rocks cat tiger water shore	ocean coral reefs fish fan
Model				
Human Annotation	tree restaurant street statue	water boats harbor skyline	people street cars festival	sky buildings street cars
Mix-Hier Annotation	statue street tree buildings castle	skyline boats coast shore water	street cars village buildings people	street buildings bridge sky arch
Model				
Human Annotation	sky people ruins temple	tree house garden lawn	flowers house garden	field horses mare foals
Mix-Hier Annotation	statue buildings temple stone ruins	garden cottage house tree path	garden plants tree house flowers	mare foals horses field meadow

First Five Ranked Results

- Query: mountain

- Query: pool

- Query: tiger

First Five Ranked Results

- Query: horses

- Query: plants

- Query: blooms

First Five Ranked Results

- Query: clouds

- Query: field

- Query: flowers

First Five Ranked Results

- Query: jet

- Query: leaf

- Query: sea

abc

Semantic Retrieval (SR)

- Problem of lexical ambiguity
  - multiple meaning of the same word
    - Anchor - TV anchor or for Ship?
    - Bank - Financial Institution or River bank?

Lake? Fishing? Boating? People?

- Multiple semantic interpretations of an image
    - Boating or Fishing or People?

Fishing! what if not in the vocabulary?

- Limited by Vocabulary size
  - What if the system was not trained for 'Fishing'
  - In other words, it is outside the space of trained semantic concepts

abc VS 

In Summary

- SR Higher level of abstraction
  - Better generalization inside the space of trained semantic concepts
  - **But** problem of
    - Lexical ambiguity
    - Multiple semantic interpretations
    - Vocabulary size

- QBVE is unrestricted by language.
  - Better Generalization outside the space of trained semantic concepts
    - a query image of 'Fishing' would retrieve visually similar images.
  - **But** weakly correlated with human notion of similarity

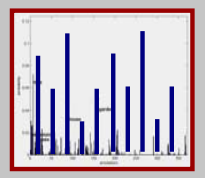
Lake? Fishing? Boating? People?

Fishing! what if not in the vocabulary?

Both have visually dissimilar sky

The two systems in many respects are complementary!

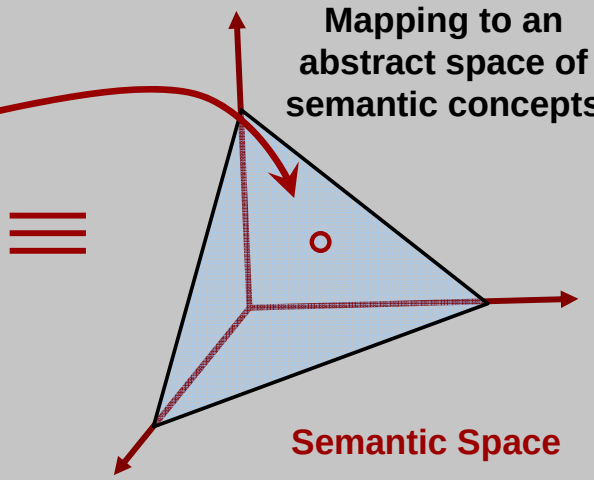
Query by Semantic Example (QBSE)



- Suggests an alternate query by example paradigm.

vector of weights or
Semantic multinomial

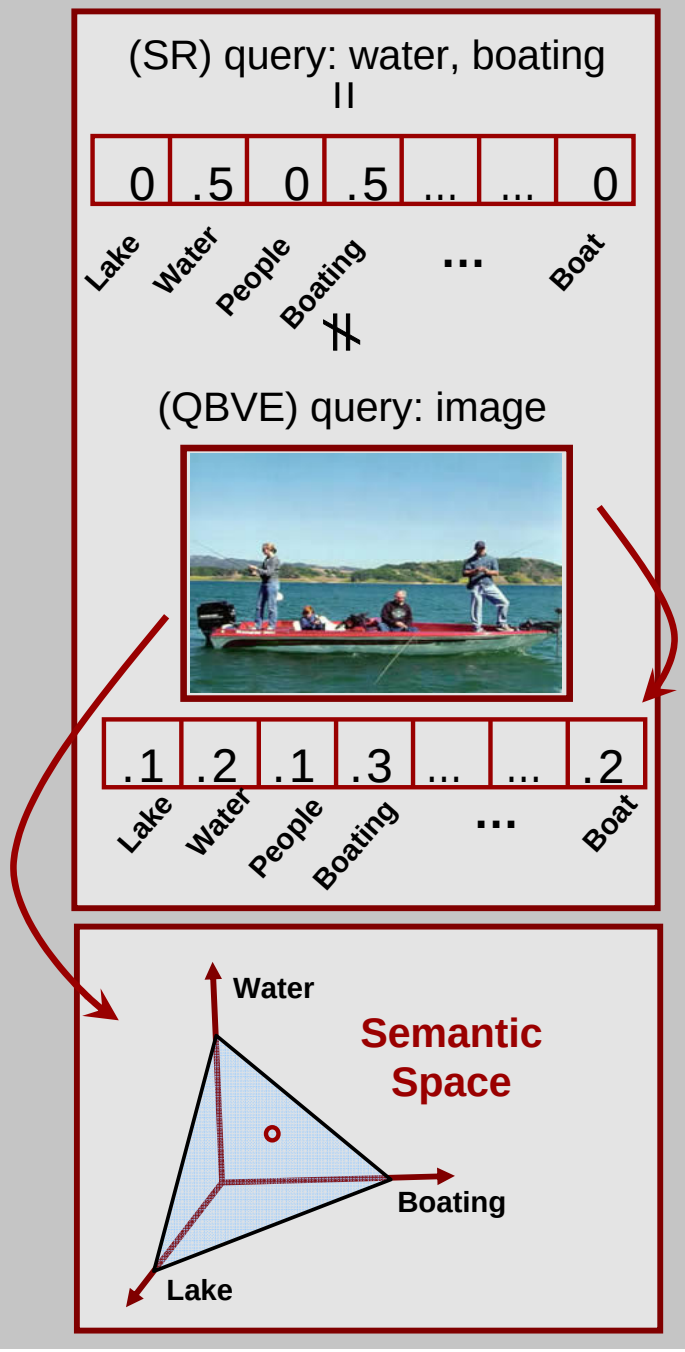
.2	.3	.2	.1	...	...	
Lake	Water	People	Sky	...		Boat



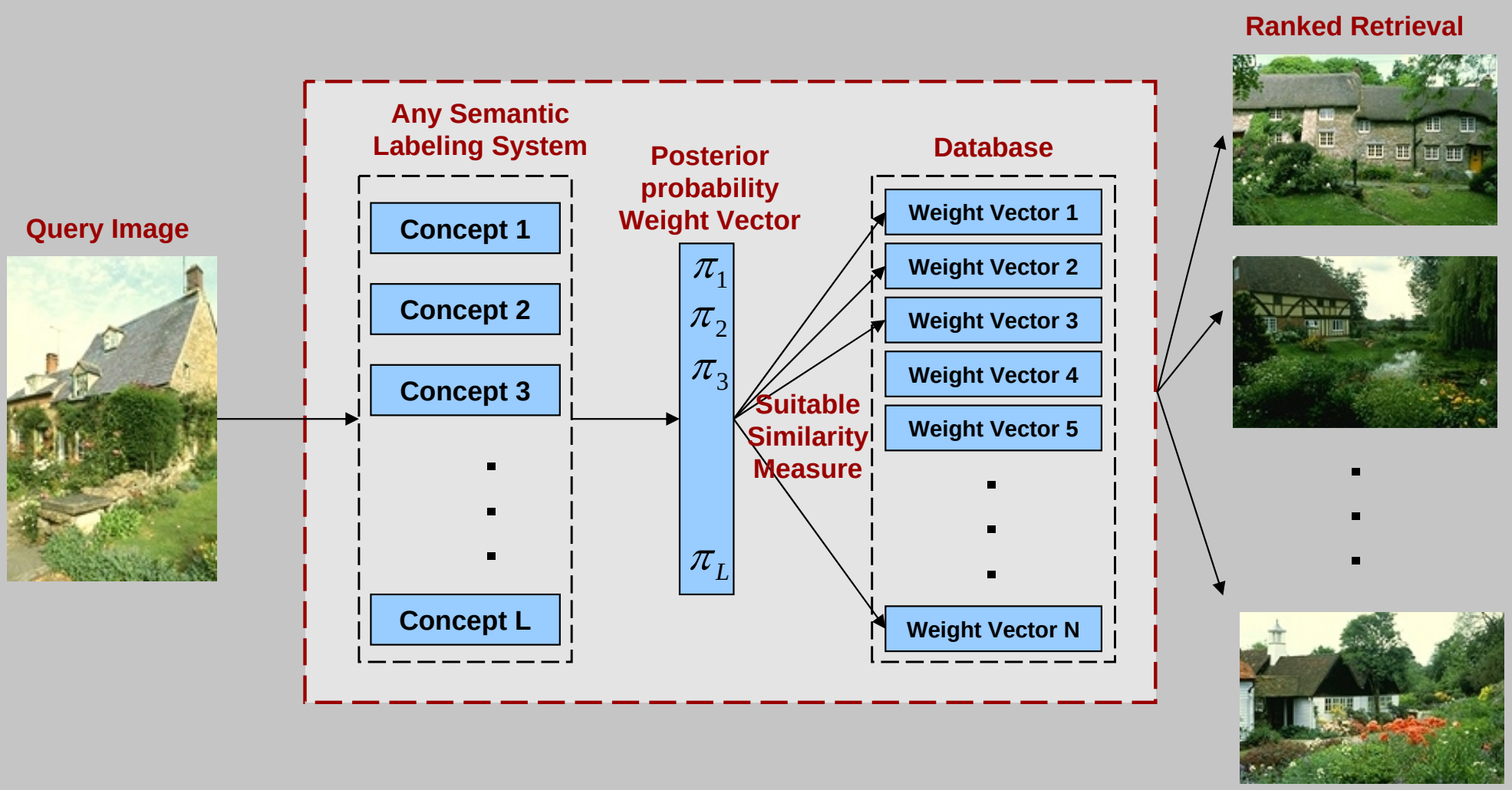
  - The user provides an image.
  - The image is mapped to vector of weights of **all** the semantic concepts in the vocabulary, using a semantic labeling system.
  - Can be thought as an projection to an abstract space, called as the semantic space
  - To retrieve an image, this weight vector is matched to database, using a suitable similarity function

Query by Semantic Example (QBSE)

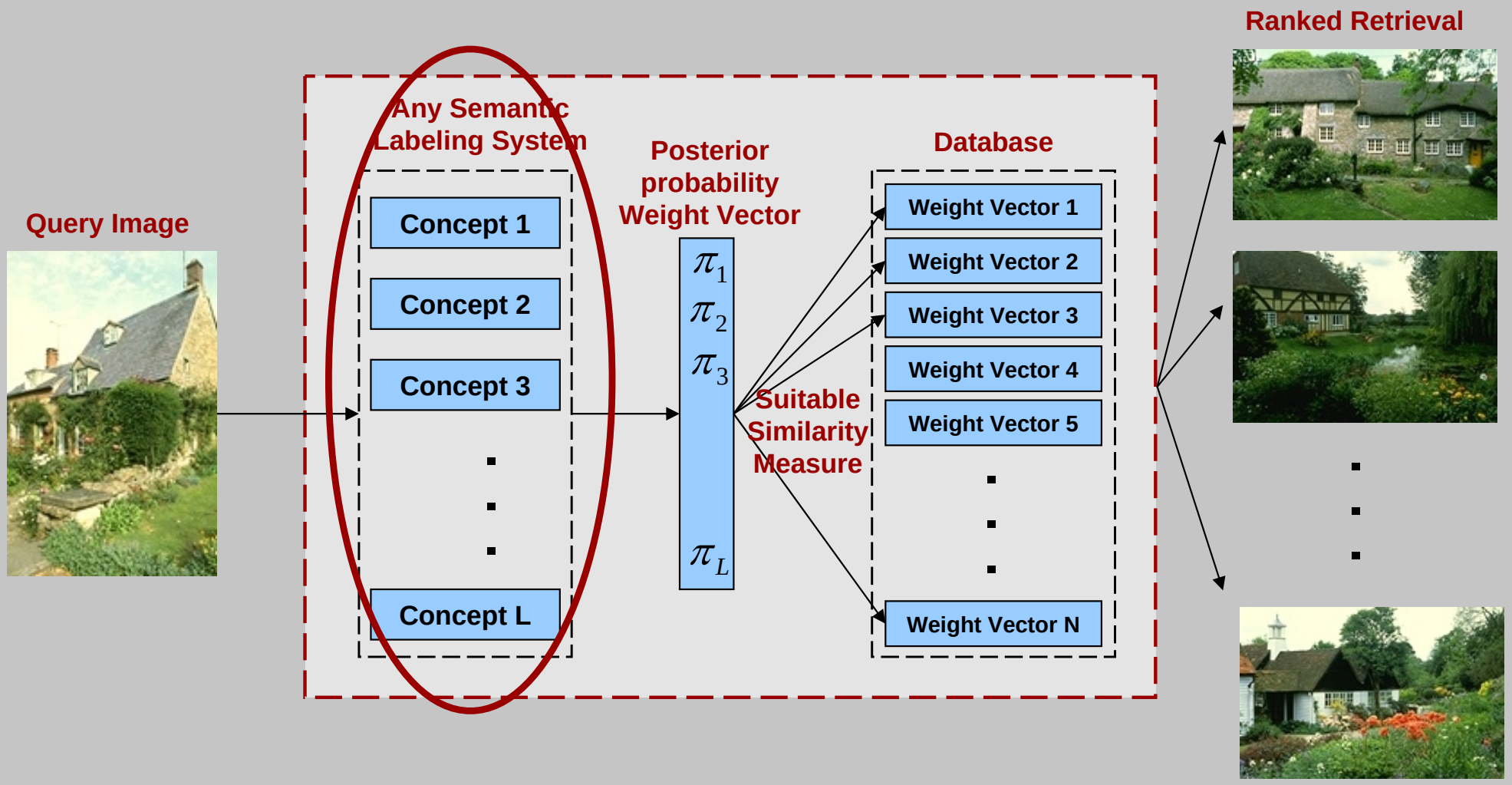
- As an extension of SR
  - Query specification not as set of few words.
  - But a vector of weights of all the semantic concept in the vocabulary.
  - Eliminates
    - Problem of lexical ambiguity- Bank+'more'
    - Multiple semantic interpretation– Boating, People
    - Outside the 'semantic space' – Fishing.
- As an enrichment of QBVE
  - The query is still by an example paradigm.
  - But feature space is Semantic.
    - A mapping of the image to an abstract space.
  - Similarity measure at a higher level of abstraction.



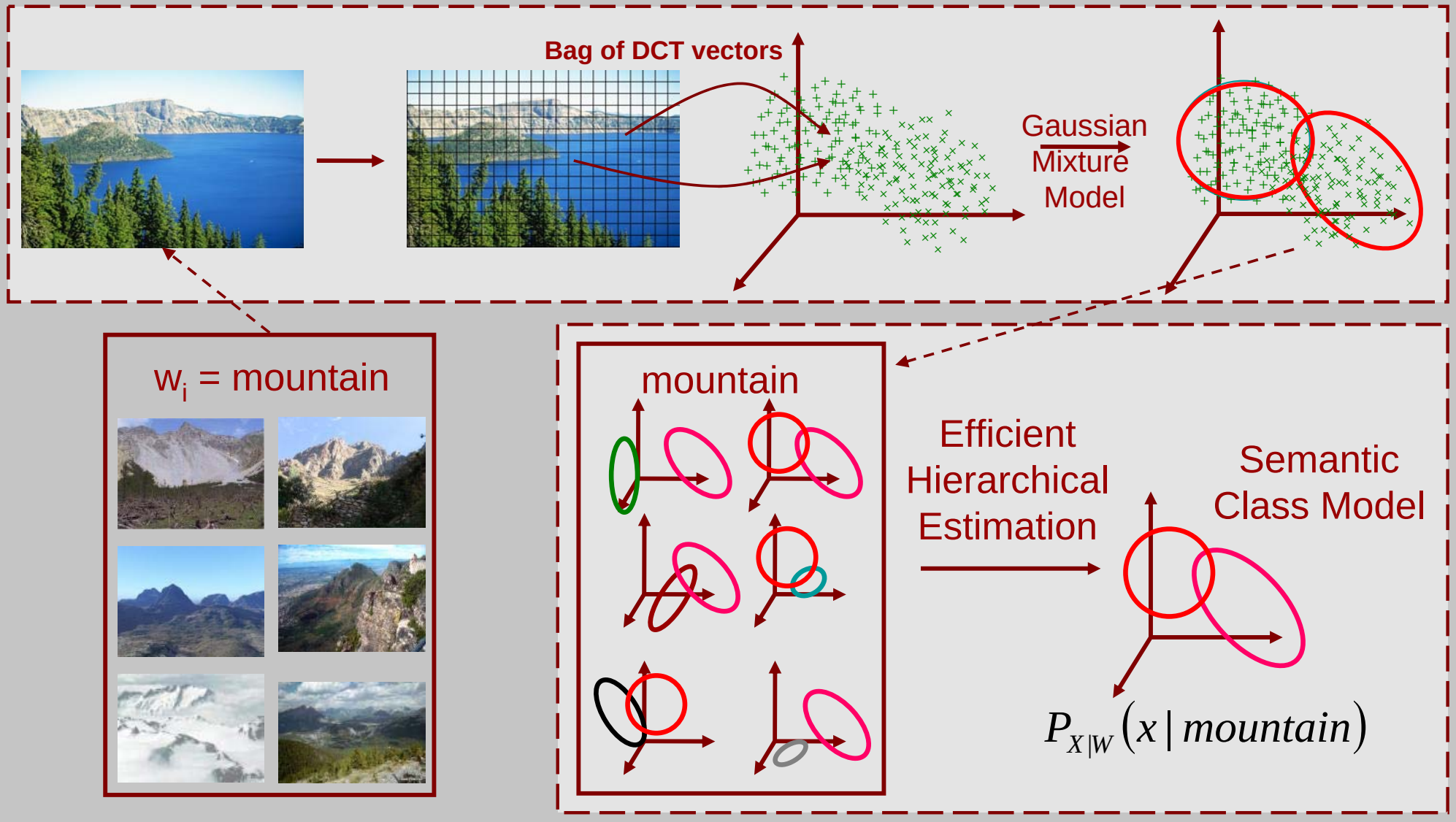
QBSE System


Query Image
Any Semantic Labeling System
Concept 1
Concept 2
Concept 3
Concept L
Posterior probability Weight Vector
π_1
π_2
π_3
π_L
Suitable Similarity Measure
Database
Weight Vector 1
Weight Vector 2
Weight Vector 3
Weight Vector 4
Weight Vector 5
Weight Vector N
Ranked Retrieval

QBSE System

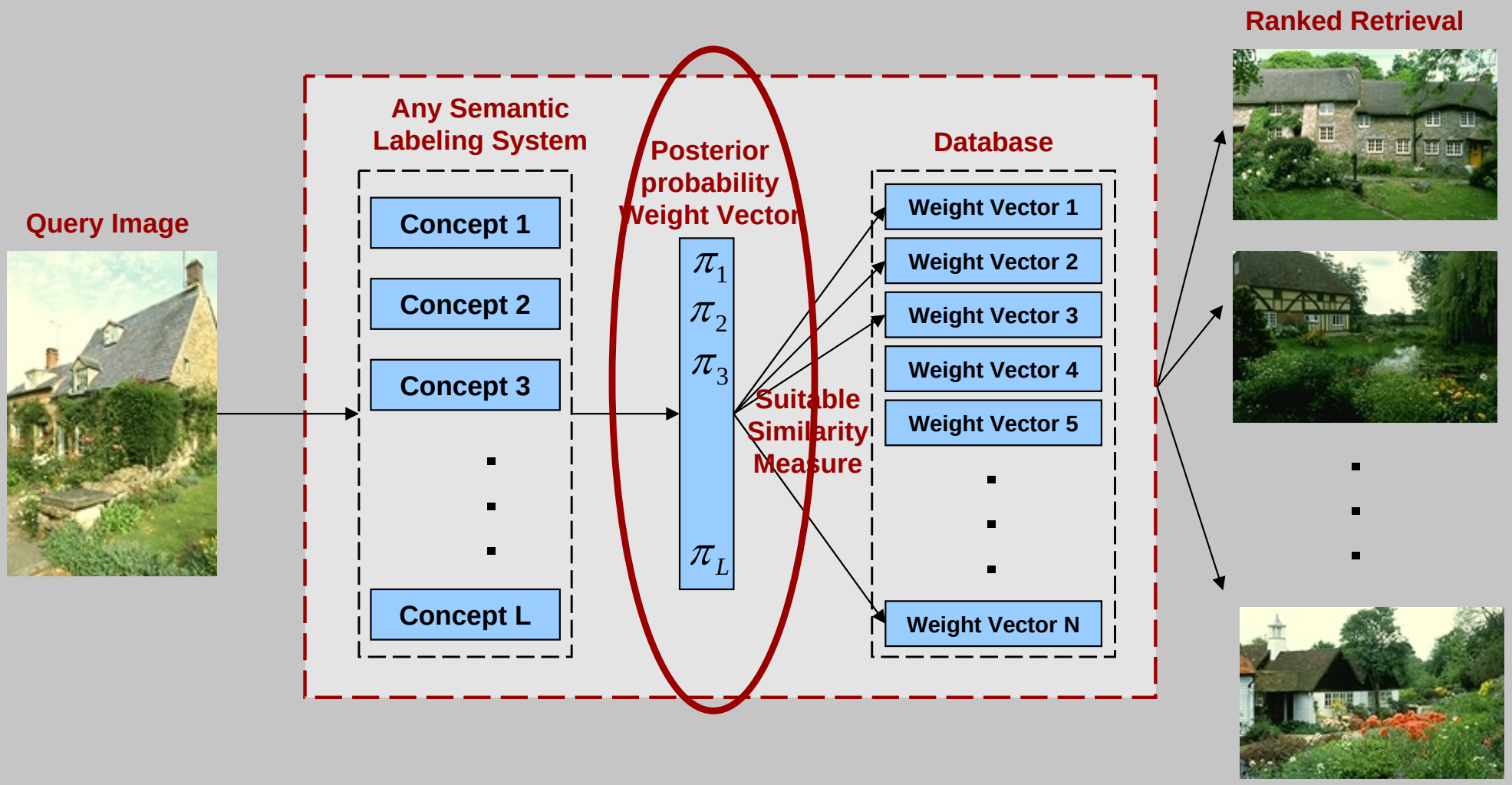

Query Image
Any Semantic Labeling System
Concept 1
Concept 2
Concept 3
Concept L
Posterior probability Weight Vector
π_1
π_2
π_3
π_L
Suitable Similarity Measure
Database
Weight Vector 1
Weight Vector 2
Weight Vector 3
Weight Vector 4
Weight Vector 5
Weight Vector N
Ranked Retrieval

Semantic Class Modeling



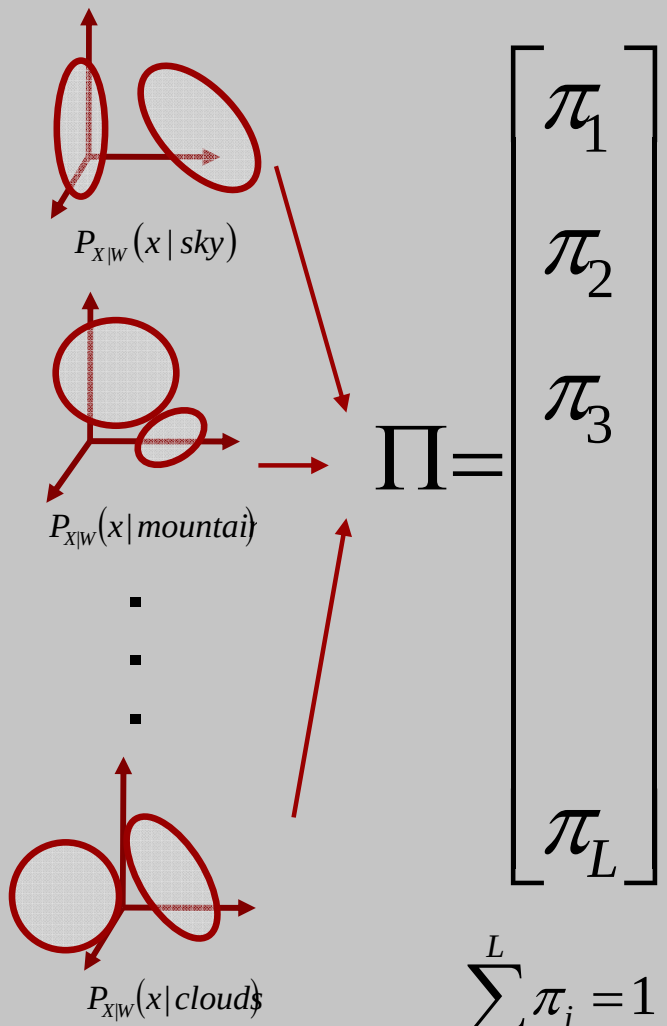
- *"Formulating Semantics Image Annotation as a Supervised Learning Problem"* **[G. Carneiro, CVPR 2005]**

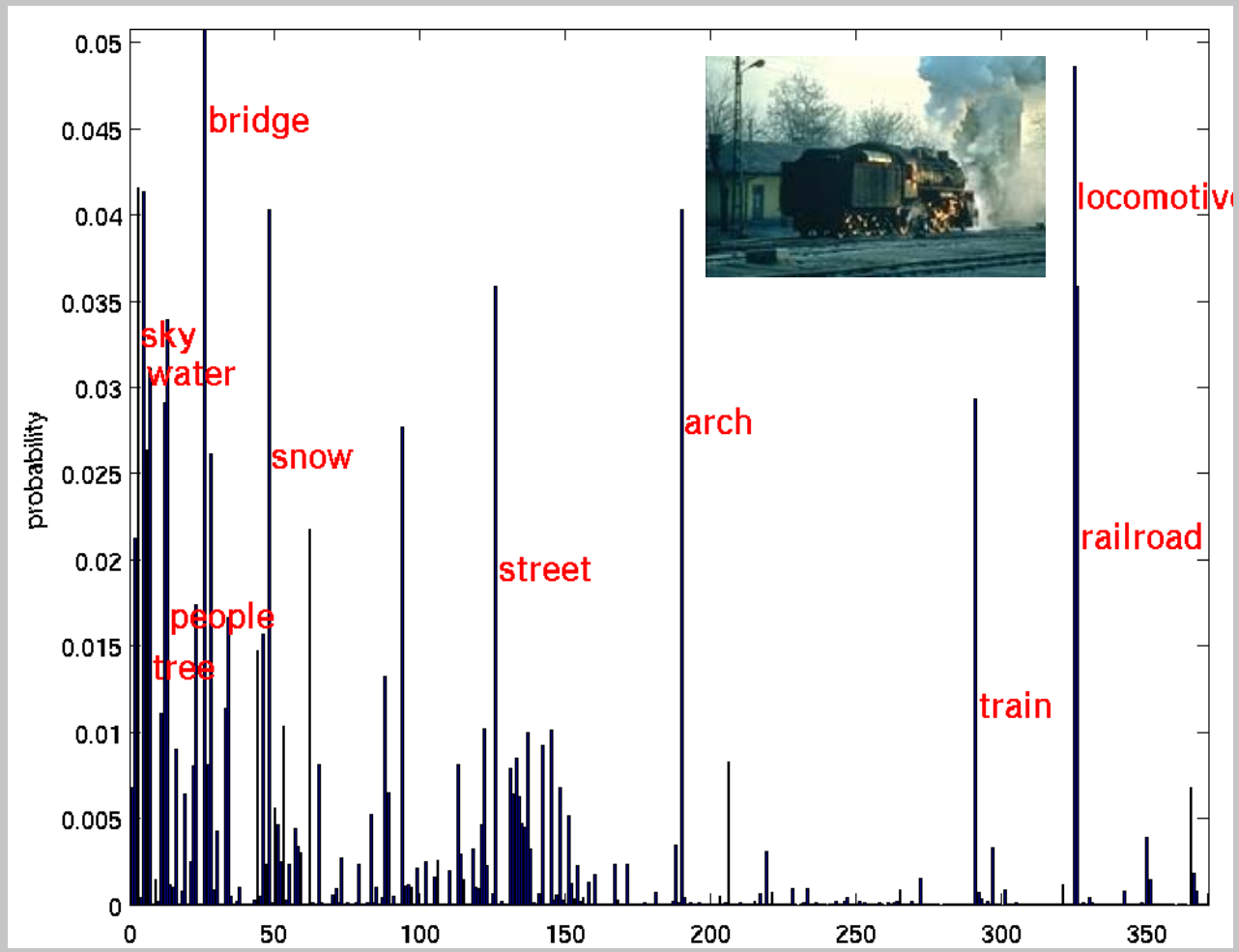
QBSE System


Query Image
Any Semantic Labeling System
Concept 1
Concept 2
Concept 3
Concept L
Posterior probability Weight Vector
π_1
π_2
π_3
π_L
Suitable Similarity Measure
Database
Weight Vector 1
Weight Vector 2
Weight Vector 3
Weight Vector 4
Weight Vector 5
Weight Vector N
Ranked Retrieval

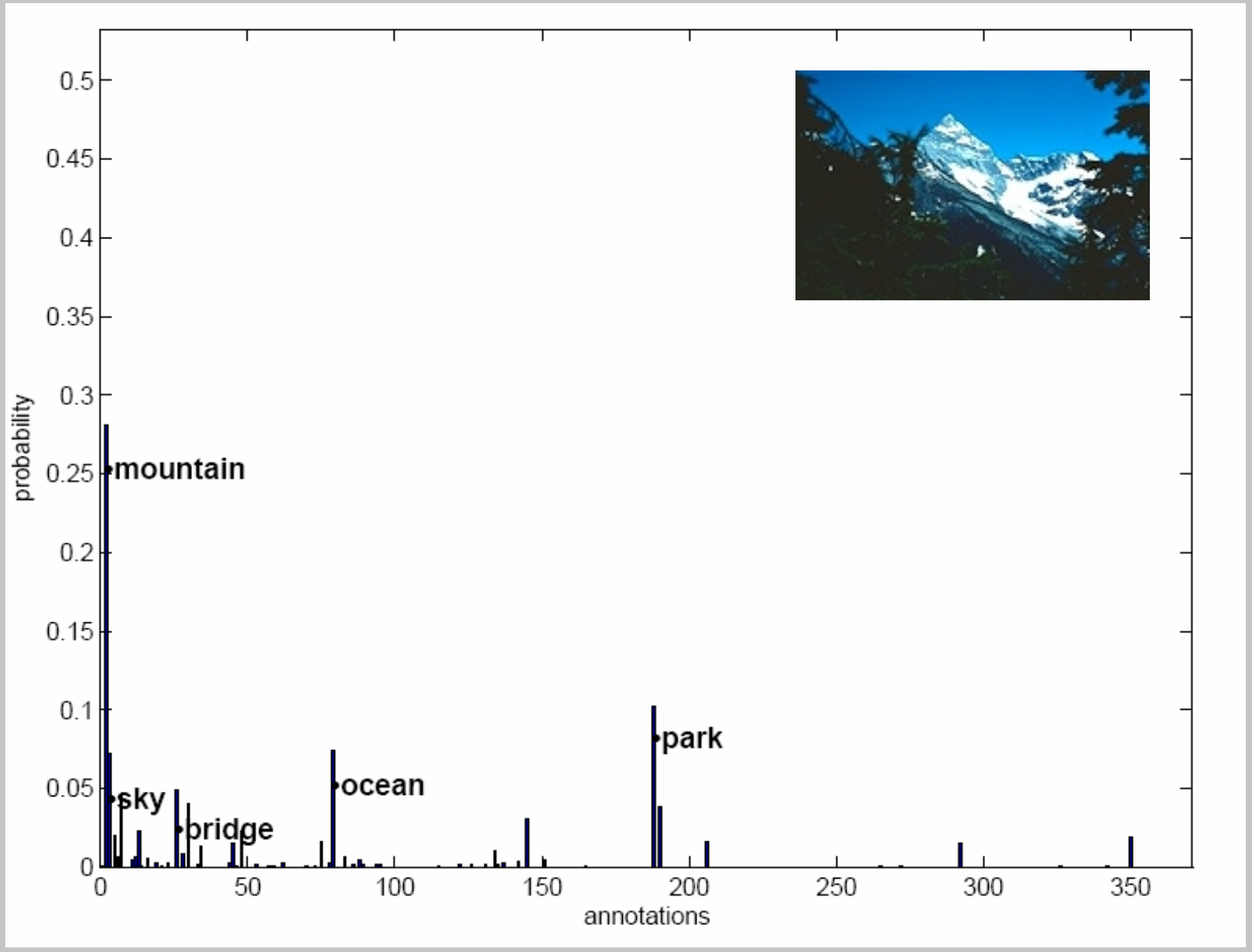
Semantic Multinomial

- Posterior Probabilities under series of L independent class models

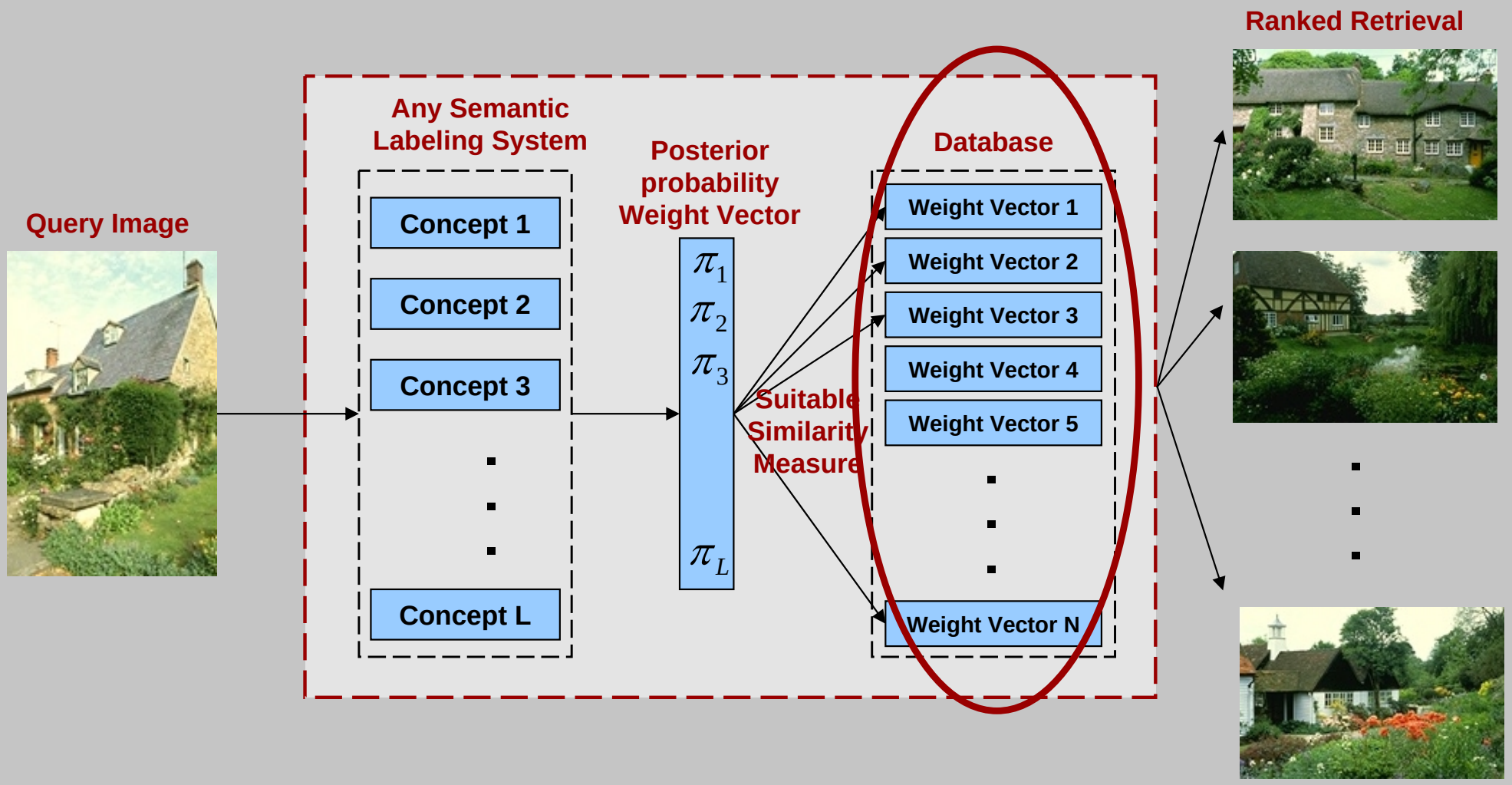




Semantic Multinomial



QBSE System

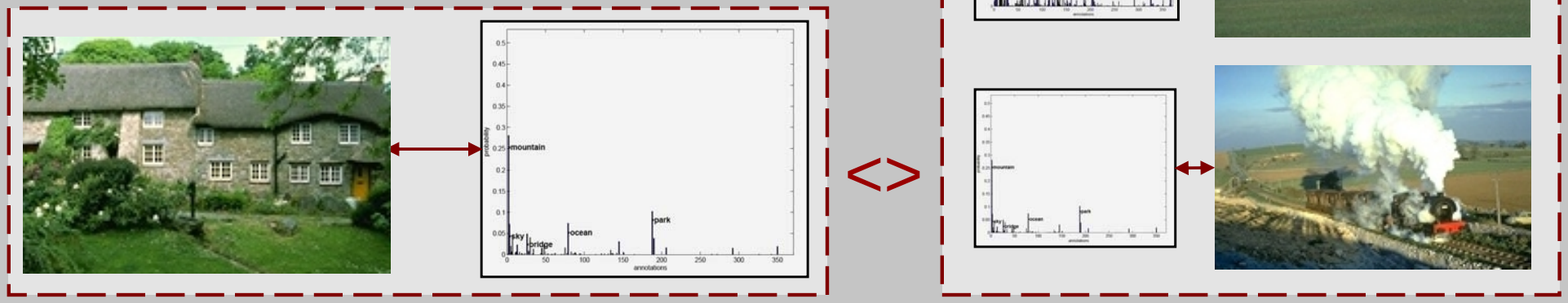

Ranked Retrieval
Query Image
Any Semantic Labeling System
Concept 1
Concept 2
Concept 3
Concept L
Posterior probability Weight Vector
π_1
π_2
π_3
π_L
Suitable Similarity Measure
Database
Weight Vector 1
Weight Vector 2
Weight Vector 3
Weight Vector 4
Weight Vector 5
Weight Vector N

Query using QBSE

- Note that SMNs are probability distributions
- A natural similarity function is the Kullback-Leibler divergence

$$f(\Pi) = \arg\max_i KL(\Pi \,||\, \Pi_i)$$

$$= \arg\max_i \sum_{j=1}^{L} \pi_j \log \frac{\pi_j^i}{\pi_j}$$



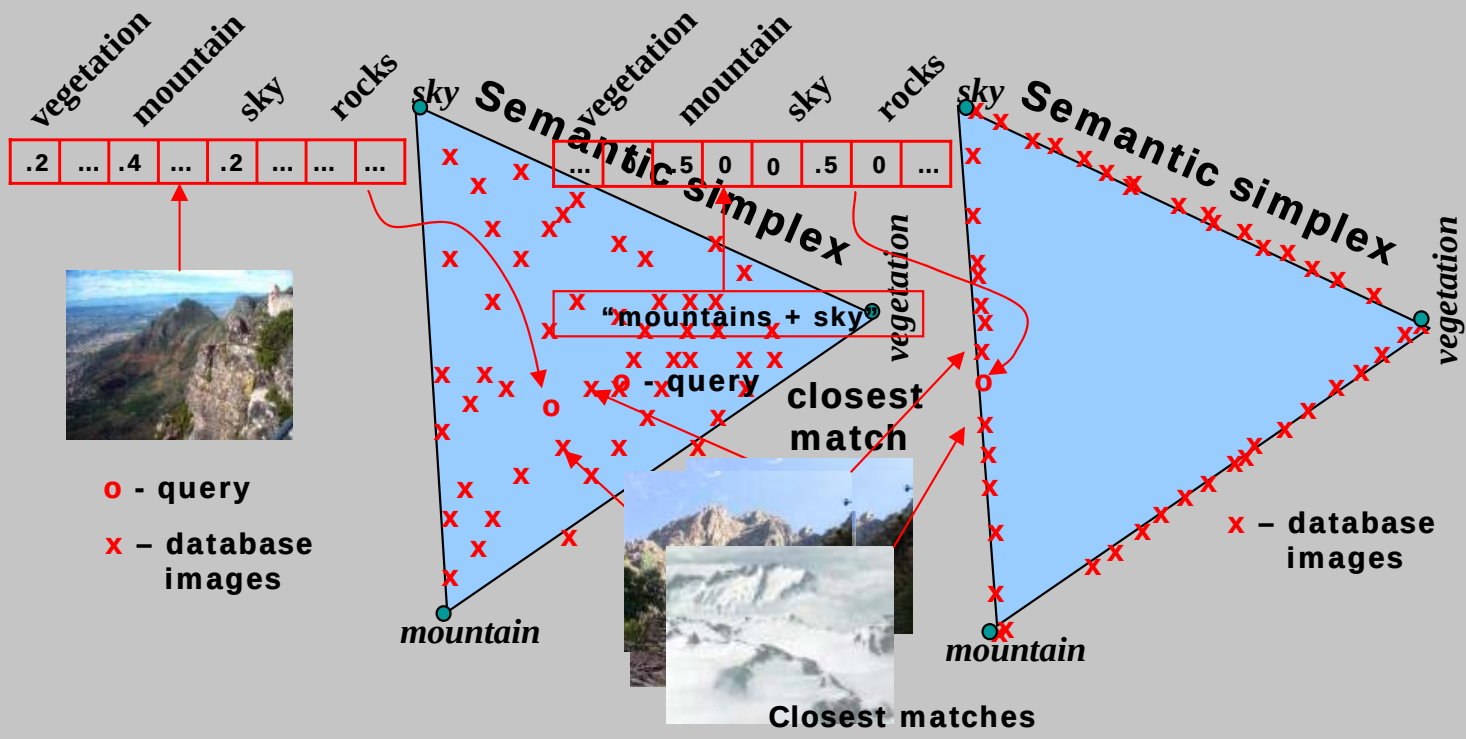
Query Query SMN

Semantic Feature Space

- The space is the simplex of posterior concept probabilities
- Each image/SVM is thus represented as a point in this simplex

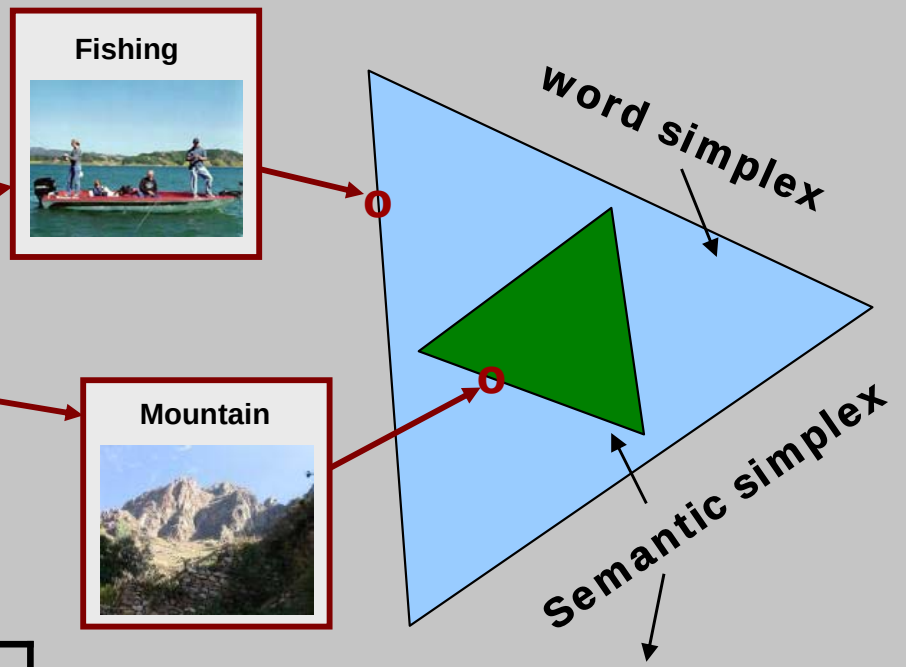
Query by Semantic Example

Traditional Text Based Query



Generalization

- two cases
  - classes outside semantic space
  - classes inside semantic space
- generalization:

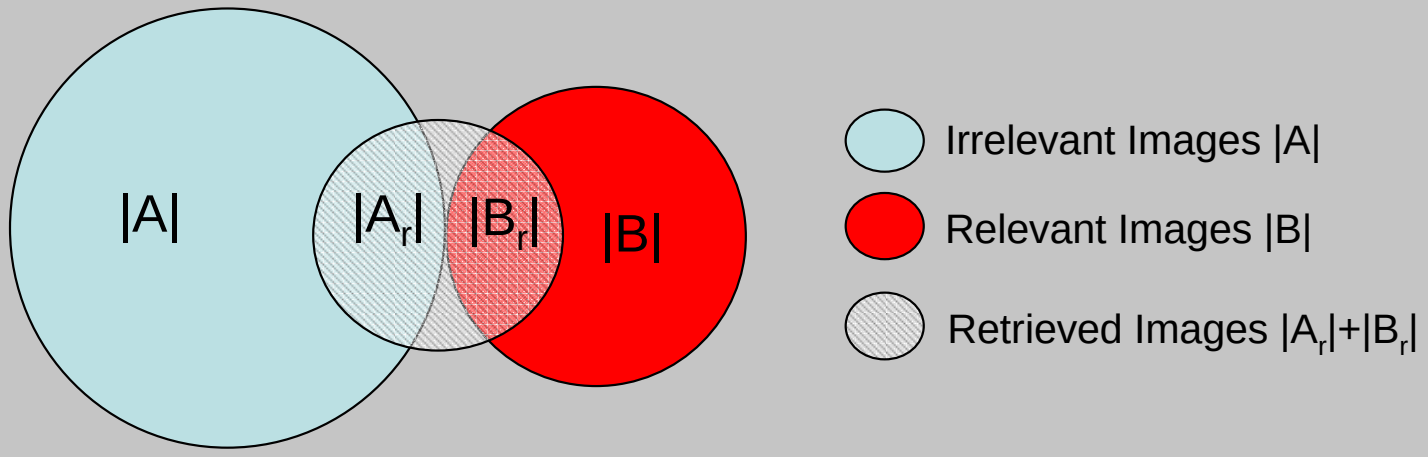


	QBVE	SR	QBSE
inside	OK	best	best
outside	OK	none	best

Fishing? ~ Lake
+ Boating
+ People

Both have visually dissimilar sky

Evaluation – Precision:Recall:Scope



$$\text{Precision} = \frac{|B_r|}{|A_r| + |B_r|}$$

The proportion of retrieved *and* relevant images to all the images retrieved.

$$\text{Recall} = \frac{|B_r|}{|B|}$$

The proportion of relevant images that are retrieved, out of all relevant images available.

$$\text{Scope} = |A_r| + |B_r|$$

The number of images that are retrieved.

Experimental Setup

- Evaluation Procedure [Feng,04]
  - Precision-Recall(scope) Curves : Calculate precision at various recalls(scopes).
  - Mean Average Precision: Average precision over all queries, where recall changes (i.e. where relevant items occur)

- Training the Semantic Space
  - Images – Corel Stock Photo CD's – Corel50
    - 5,000 images from 50 CD's, 4,500 used for training the space
  - Semantic Concepts
    - Total of 371 concepts
    - Each Image has caption of 1-5 concepts
    - Semantic concept model learned for each concept.

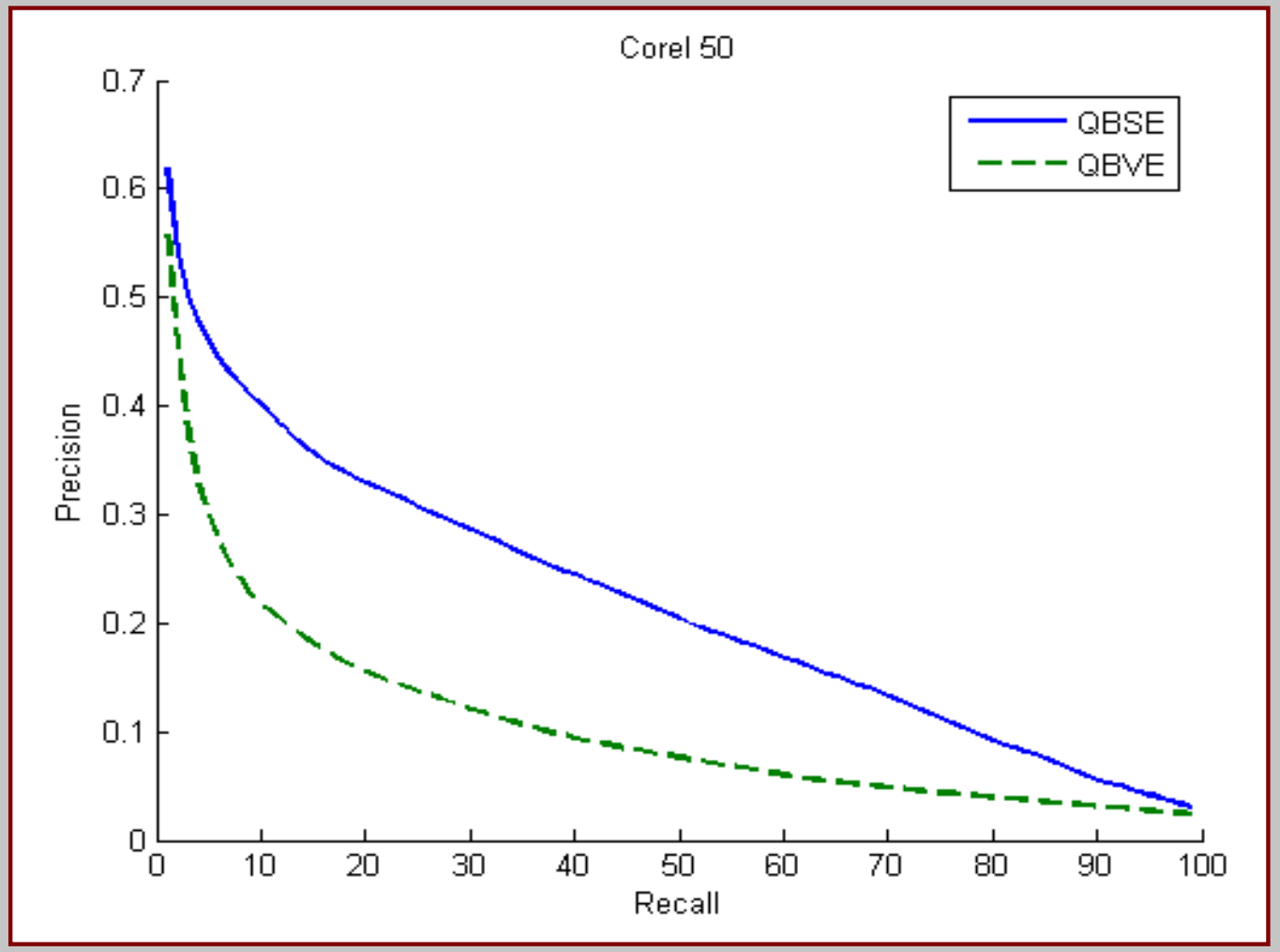
Experimental Setup

- Retrieval inside the Semantic Space.
  - Images – Corel Stock Photo CD's – same as Corel50
    - 4,500 used as retrieval database
    - 500 used to query the database
- Retrieval outside the Semantic Space
  - Corel15 - Another 15 Corel Photo CD's, (not used previously)
    - 1200: retrieval database, 300: query database
  - Flickr18 - 1800 images Downloaded from www.flickr.com
    - 1440: retrieval database, 360: query database
    - harder than Corel images as shot by non-professional flickr users

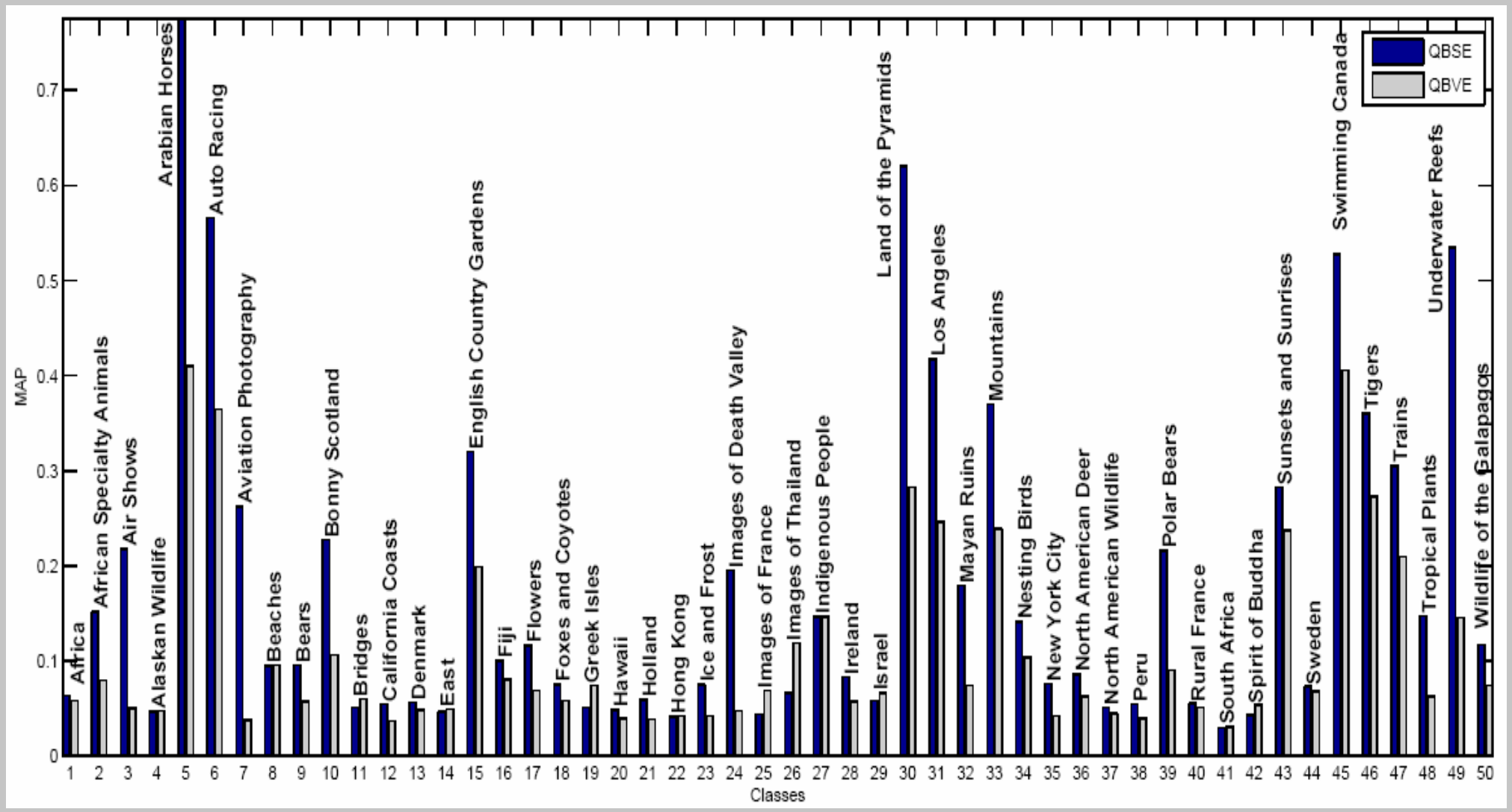
Database	Semantic Space	Source	# Retrieval Images	# Query Images	# Classes
Corel50	Inside	Corel Stock Photo CDs	4500	500	50
Corel15	Outside	Corel Stock Photo CDs	1200	300	15
Flickr18	Outside	www.flickr.com	1440	360	18

Inside the Semantic Space

- Precision of QBSE is significantly higher at most levels of recall



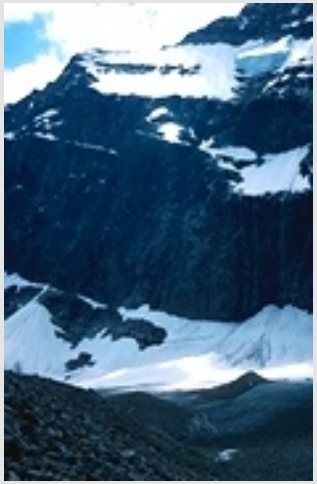
- MAP score for all the 50 classes



Inside the Semantic Space

same colors
different semantics

QBSE



QBVE

Inside the Semantic Space

"train + railroad"

"whitish + darkish"

QBSE

QBVE

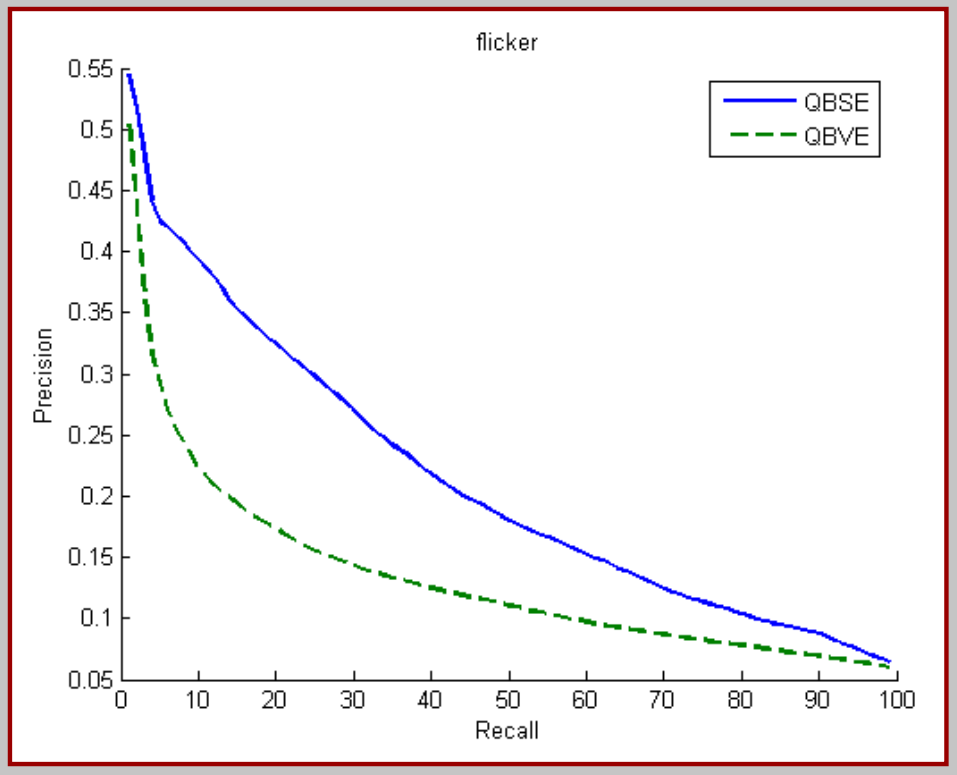


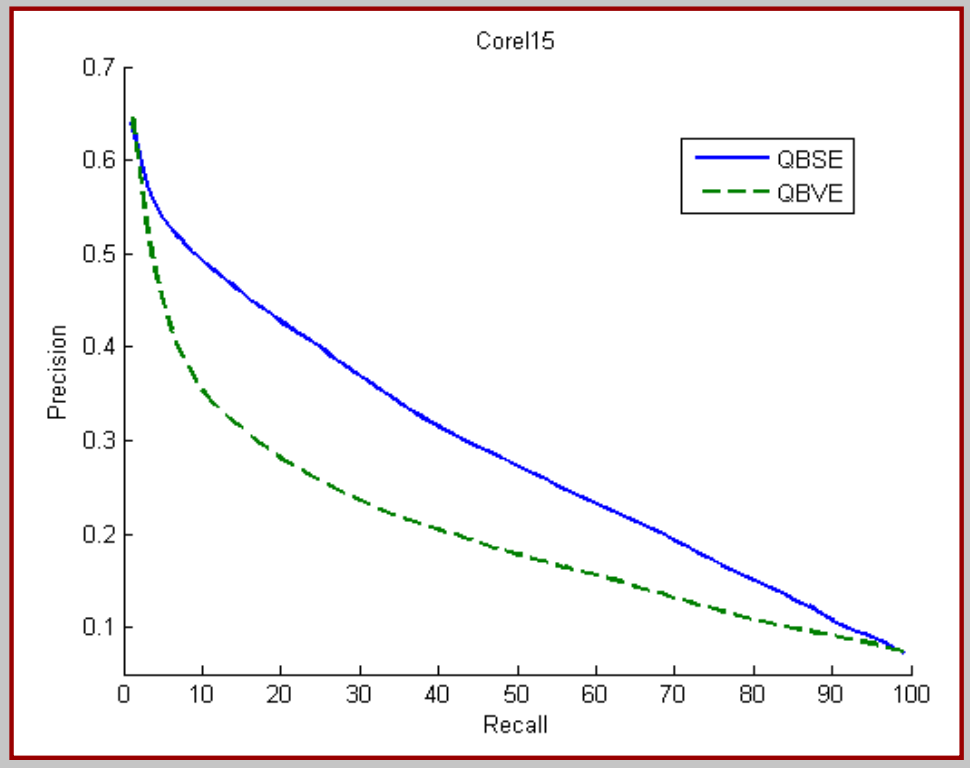





Outside the Semantic Space

Database	Semantic Space	Source	# Retrieval Images	# Query Images	# Classes
Corel50	Inside	Corel Stock Photo CDs	4500	500	50
Corel15	Outside	Corel Stock Photo CDs	1200	300	15
Flickr18	Outside	`www.flickr.com`	1440	360	18

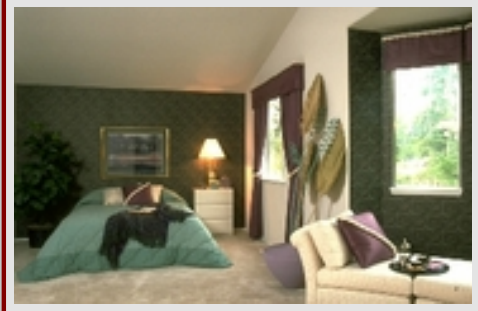


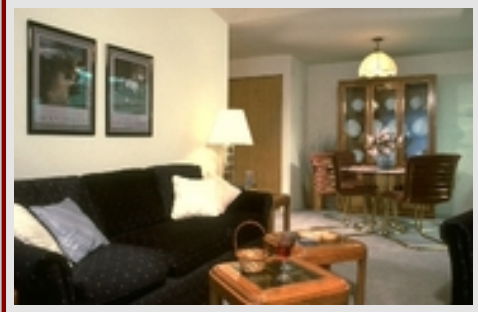


Commercial Construction

People	0.09
Buildings	0.07
Street	0.07
Statue	0.05
Tables	0.04
Water	0.04
Restaurant	0.04

QBVE





People	0.08
Statue	0.07
Buildings	0.06
Tables	0.05
Street	0.05
Restaurant	0.04
House	0.03

Buildings	0.06
People	0.06
Street	0.06
Statue	0.04
Tree	0.04
Boats	0.04
Water	0.03

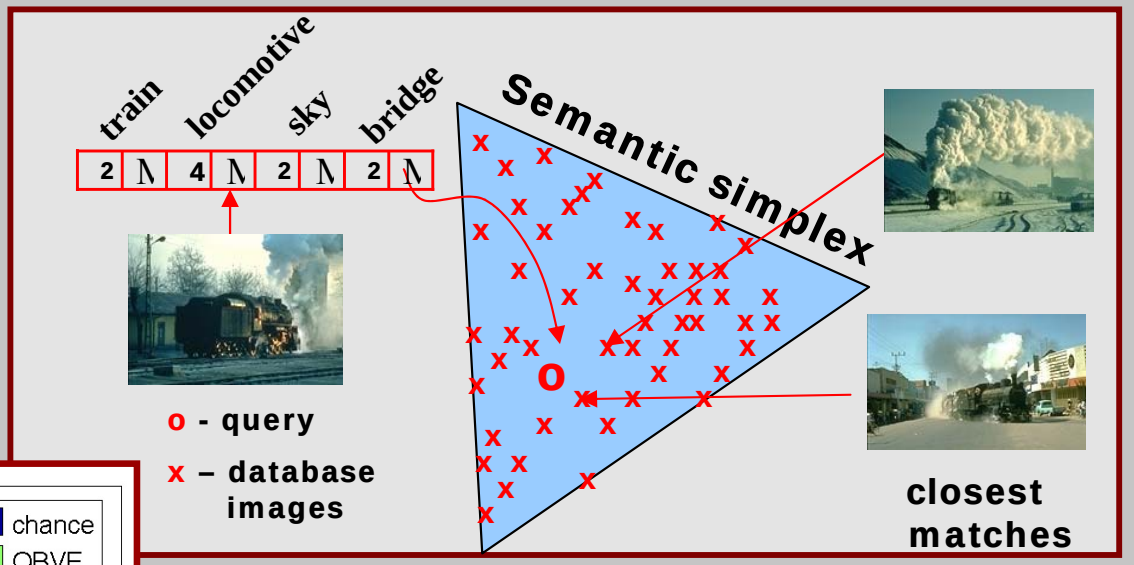
People	0.1
Statue	0.08
Buildings	0.07
Tables	0.06
Street	0.06
Door	0.05
Restaurant	0.04

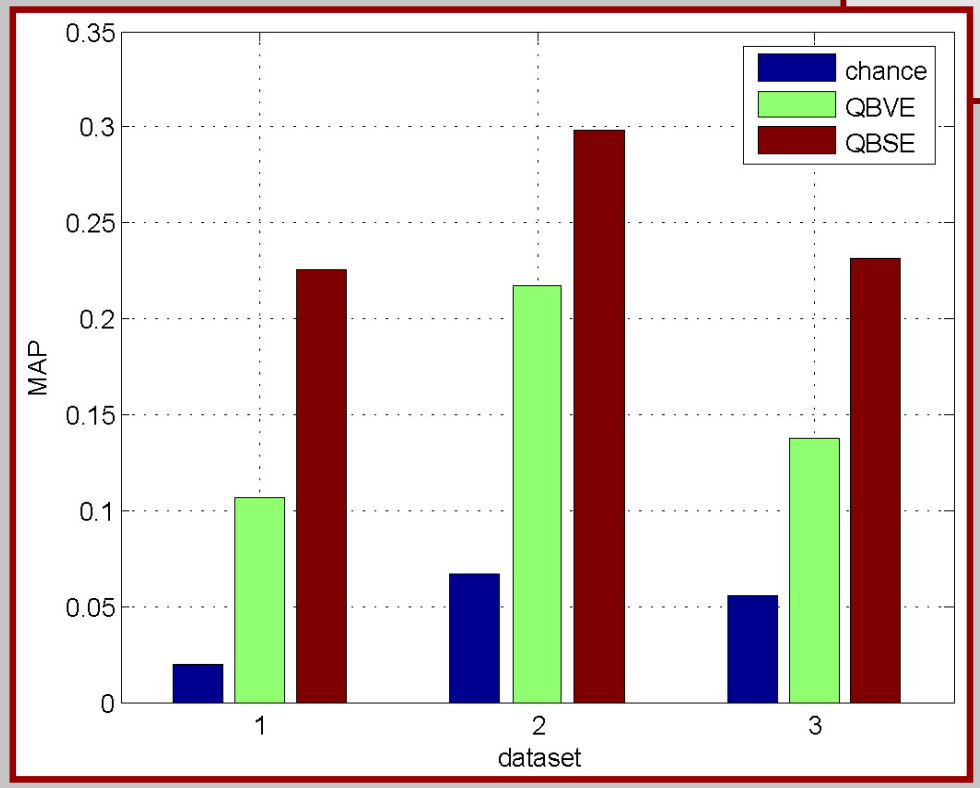
QBSE

People	0.12
Restaurant	0.07
Sky	0.06
Tables	0.06
Street	0.05
Buildings	0.05
Statue	0.05

QBSE vs QBVE

- nearest neighbors in this space is significantly more robust



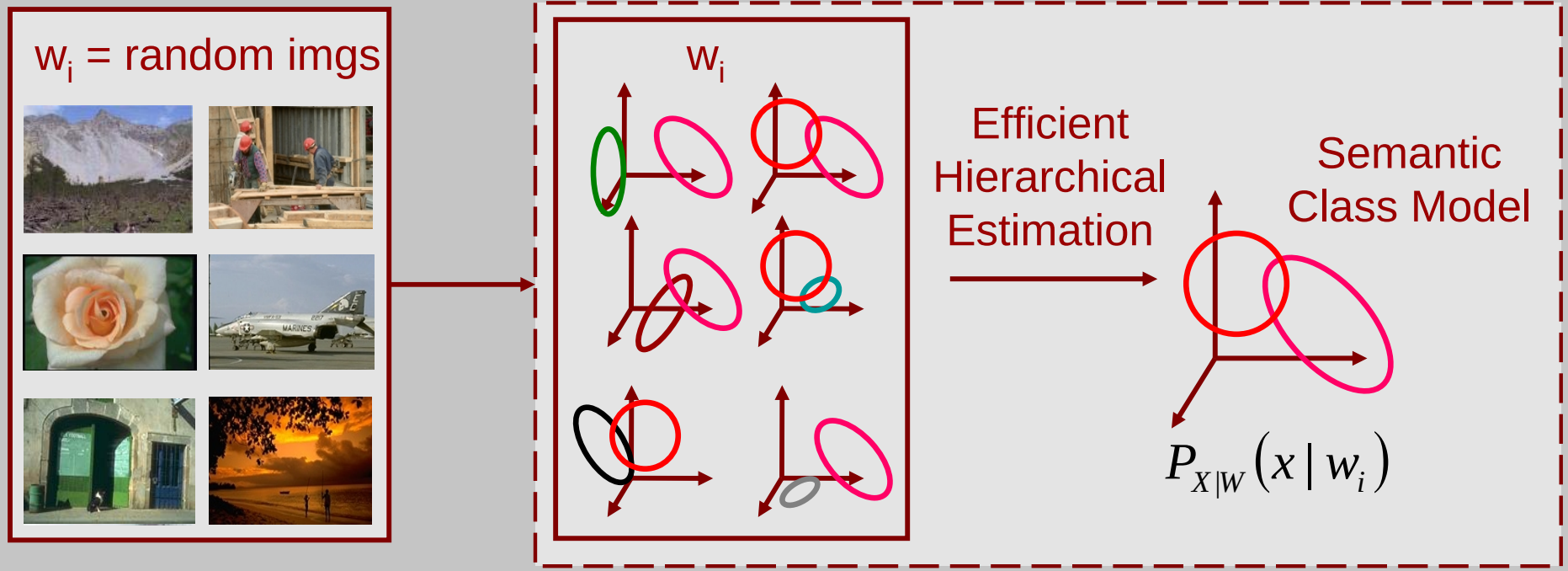


- both in terms of
  - metrics
  - subjective matching quality

- "Query by semantic example"
  [N. Rasiawasia, IEEE Trans. Multimedia 2007]

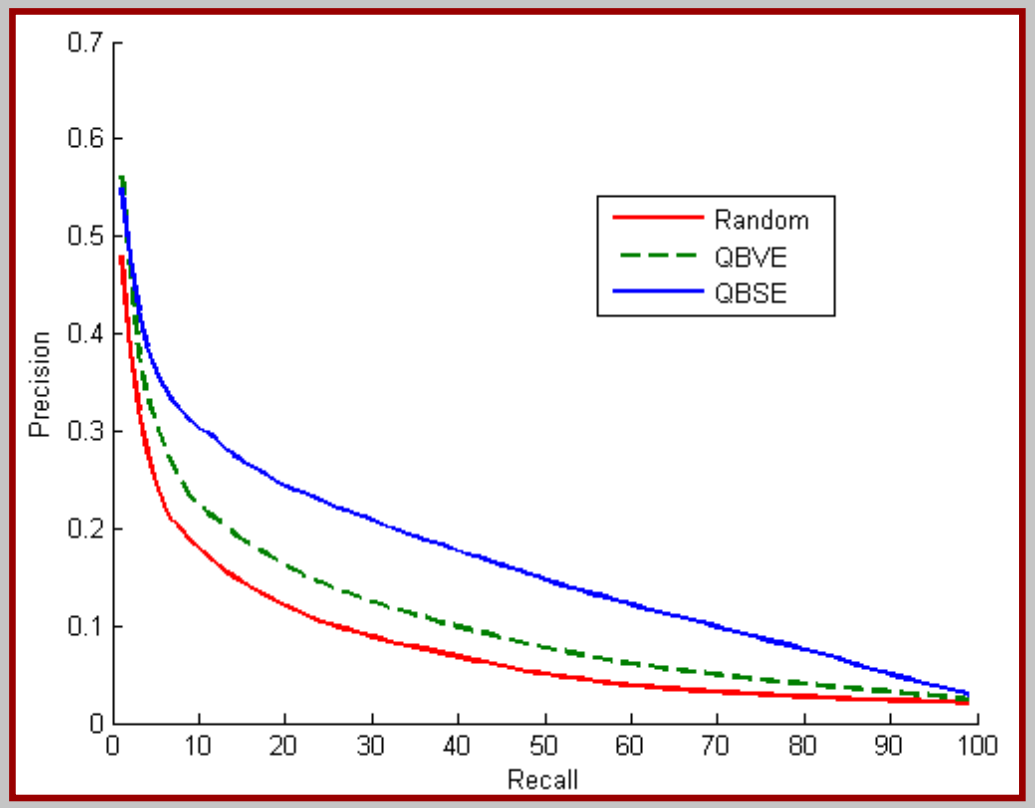
Structure of the Semantic Space

- is the gain really due to the semantic structure of the SMN space?
- this can be tested by comparing to a space where the probabilities are relative to random image groupings



The semantic gain

- with random groupings performance is
  - quite poor, indeed worse than QBVE



  - there seems to be an intrinsic gain of relying on a space where the features are semantic

Relationship among semantic features

- Does semantic space encodes contextual relationships?
- Measure the mutual information between pairs of semantic features.

$$I(w_1; w_2) = \sum_{w_2 \in \mathcal{L}} \sum_{w_1 \in \mathcal{L}} p(w_1, w_2) \log \frac{p(w_1, w_2)}{p(w_1)\, p(w_2)},$$

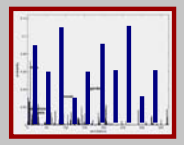
- Strong for pairs of concepts that are synonyms or frequently appear together in natural imagery.

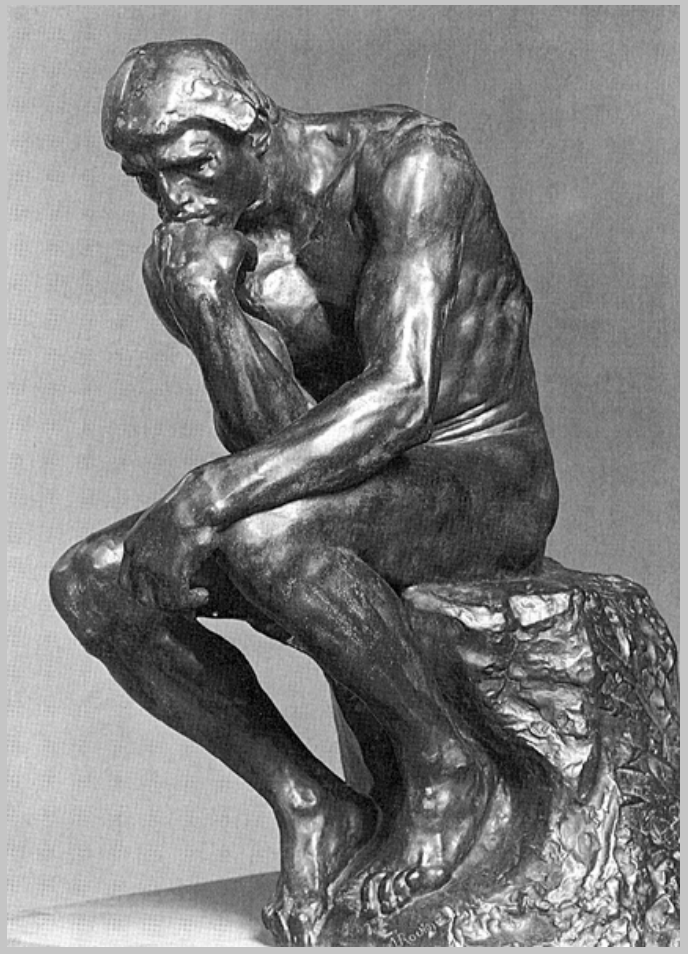
Feature Pair	MI	Feature Pair	MI
sunset-sun	0.0844	jet-plane	0.0843
coral-reefs	0.0817	ocean-reefs	0.0730
coral-ocean	0.0711	jet-flight	0.0467
sun-sea	0.0434	bear-polar	0.0392
mare-foals	0.0320	horses-foals	0.0320
cars-formula	0.0279	pool-swimmers	0.0275

Conclusion

- We present a new framework for content-based retrieval, denoted by query-by-semantic-example (QBSE), by extending the query-by-example paradigm to the semantic domain.
- Substantial evidence that QBSE outperforms QBVE both inside and outside the space of known semantic concepts (denoted by *semantic space*) is presented.
- This gain is attributed to the structure of the learned semantic space, and denoted by *semantic gain*. By controlled experiments it is also shown that, in absence of semantic structure, QBSE performs worse than the QBVE system.
- Finally, we hypothesize that the important property of this structure is a characterization of contextual relationships between concepts.

Questions?

 vs  vs **abc**



Flickr18

- Automobiles
- Building Landscapes
- Facial Close Up
- Flora
- Flowers Close Up
- Food and Fruits
- Frozen
- Hills and Valley
- Horses and Foal
- Jet Planes
- Sand
- Sculpture and Statues
- Sea and Waves
- Solar
- Township
- Train
- Underwater
- Water Fun

Corel15

- Autumn
- Adventure Sailing
- Barnyard Animals
- Caves
- Cities of Italy
- Commercial Construction
- Food
- Greece
- Helicopters
- Military Vehicles
- New Zealand
- People of World
- Residential Interiors
- Sacred Places
- Soldier

Content based image retrieval

- Query by Visual Example (QBVE)
  - Color, Shape, Texture, Spatial Layout.
  - Image is represented as multidimensional feature vector
  - Suitable similarity measure

- Semantic Retrieval (SR)
  - Given keyword **w**, find images that contains the associated semantic concept.