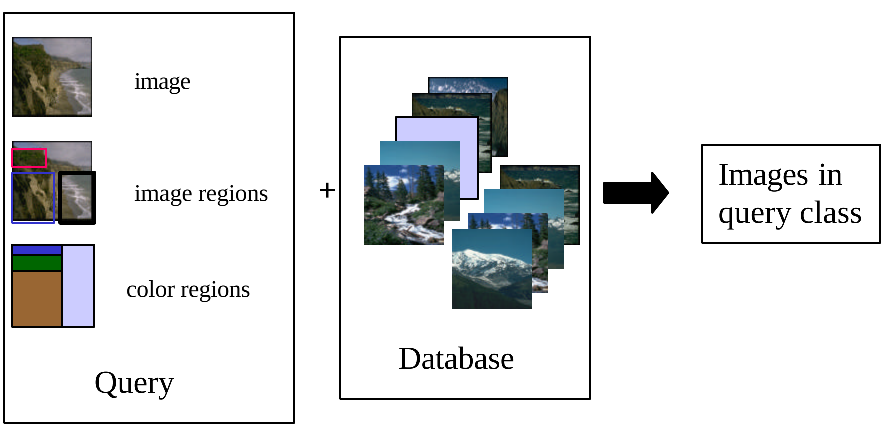
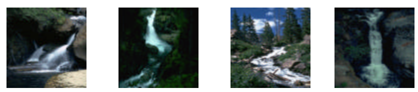
Extracting templates for Natural Scene Classification

Content Based Image Retrieval



Difficulty

- Variations in color, texture, spatial properties



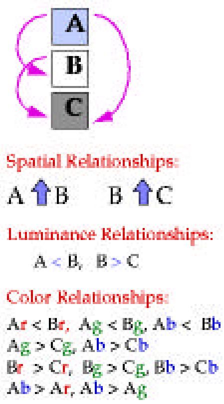
Examples from the “waterfalls class”.

Configural Recognition (Lipson 1996)

- Predefined templates :
  - relative spatial and photometric properties between image regions
  - low resolution
  - global configuration of image regions

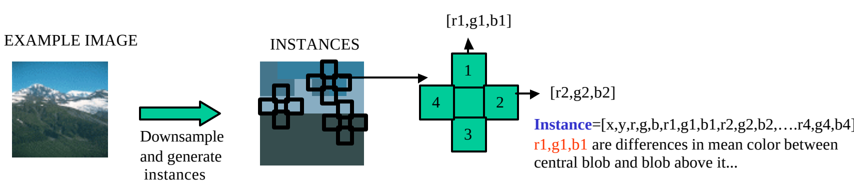
Pre-defined Template (Lipson 1996)

Snowy mountain concept



captures

Template Extraction



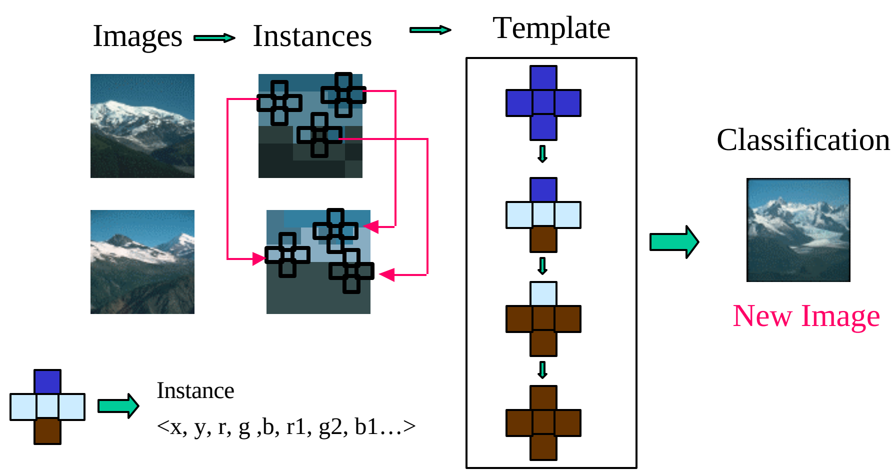
TEMPLATE EXTRACTION

1. Downsample example images
2. Extract all instances from each example image
3. Find the set of all instances that are common (reinforced) in all the positive examples (discrete intersection)
4. The common instances and their spatial relationship is the extracted template

CLASSIFICATION

For every instance in the template find closest match in new image that satisfy color and spatial relationships.

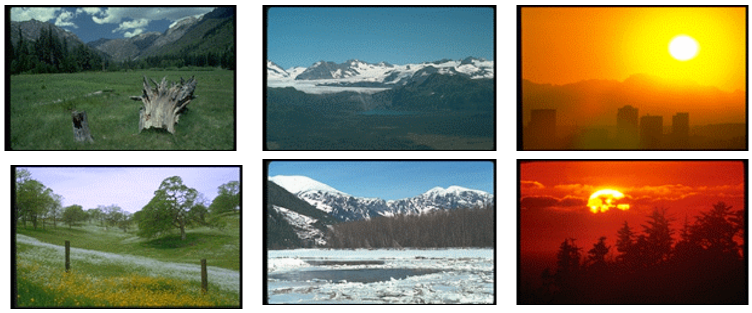
Example:
Extracting the Snowy Mountain Template



Multiple-Instance learning for Natural Scene Classification

Oded Maron and Aparna Lakshmi Ratan

LEARNING NATURAL SCENE CONCEPTS



Images are inherently ambiguous since they can represent different things.

We apply Multiple-Instance to the natural scene classification problem.

Multiple Instance Learning

- Regular learning: learn concept form leabeled examples
- MI learning: learn concepts from labeled bags
- a bag is a collection of examples
- positive bag has at least one positive example
- negative bag has only negative examples

Natural Scene Classification

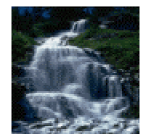
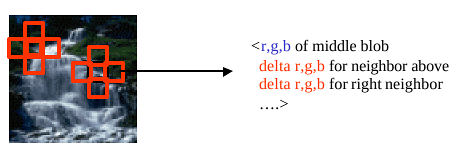
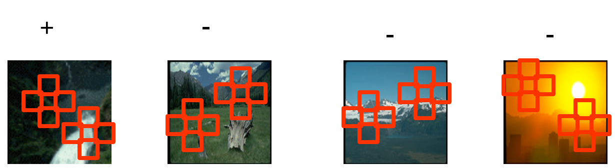
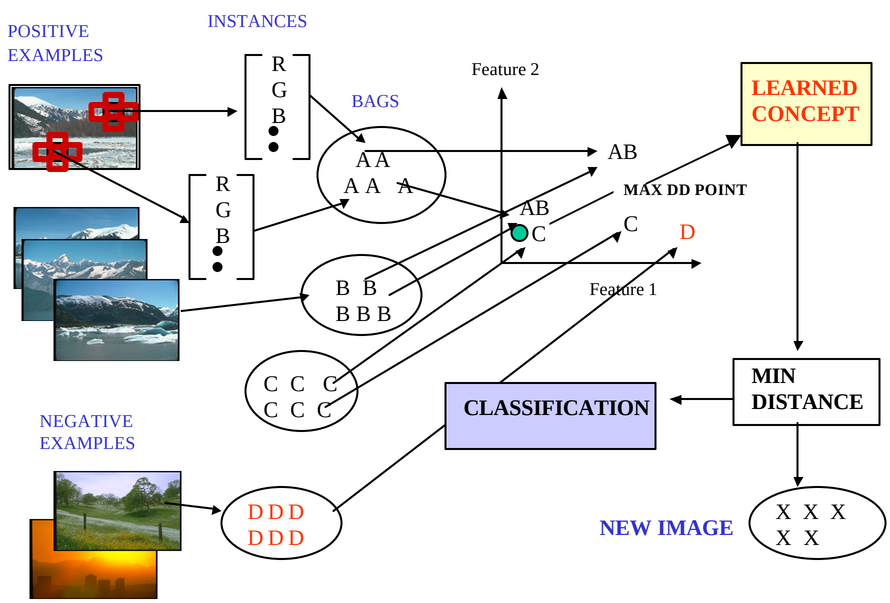
- Give me more images like this 
- Images are inherently ambiguous
- Be explicit about the ambiguity: an image is a bag and each instance is something that possible represents the image

Image ⇨ Bag   Blob ⇨ Instance

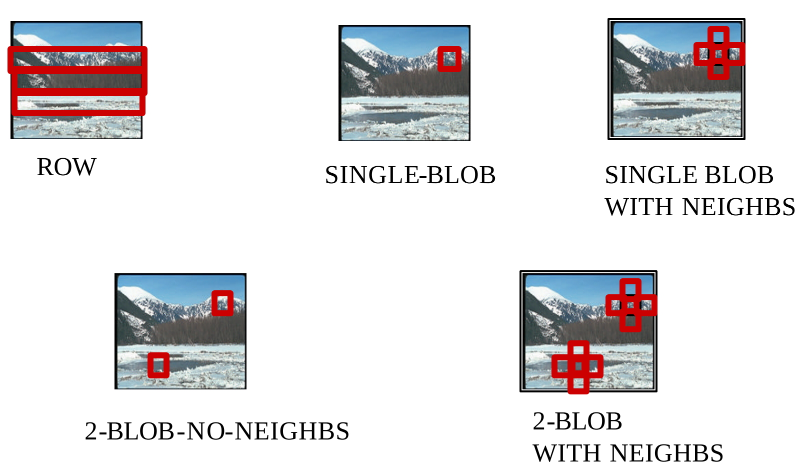




DIVERSE DENSITY FOR SCENE CLASSIFICATION

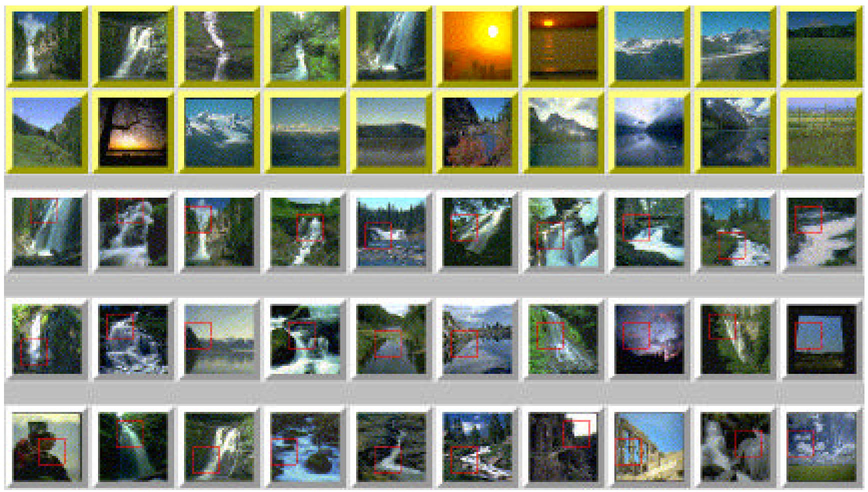


HYPOTHESIS CLASSES



SNAPSHOT OF SYSTEM

CONCEPT = WATERFALLS
DATABASE = 2600 IMAGES



DD at a point is a measure of how many instances from different bags are near that point and how far away the negative instances are.

The algorithm returns point(s) in feature space with high DD

•For a single point target concept (t) and positive and negative bags (B), we can find t by maximizing

$$\Pr(\ t \mid B_1^+, B_2^+, \ldots, B_n^+, B_1^-, \ldots, B_m^-)$$

•If we assume a uniform prior, and that bags are conditionally independent given the concept, then we maximize likelihood

$$\arg\max_t P(t \mid B_1^+, B_2^+, \ldots, B_1^-, \ldots \mid t) = \arg\max_t \prod_i P(B_i^+ \mid t) \prod_i P(B_i^- \mid t)$$

$$= \arg\max_t \prod_i P(t \mid B_i^+) \prod_i P(t \mid B_i^-)$$

•Each bag is made up of many instances

$$P(t \mid B_i^+) = P(t \mid B_{i1}^+, B_{i2}^+, \ldots B_{im}^+)$$

We use the **NOISY-OR** idea: target must be caused by one of the instances, and the causations are independent

$$P(t \mid B_i^+) = 1 - \prod_j (1 - P_c(t \mid B_{ij}^+))$$

Causal probablities are approx. by a gaussian.

$$P_c(t \mid B_{ij}^+) \approx \exp(-\| t - B_{ij}^+ \|^2)$$

SCALING FEATURES

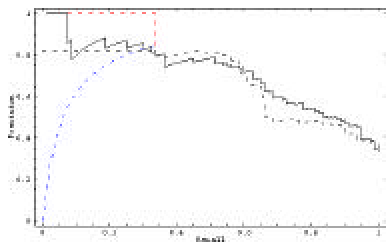
• We change the weights of features in order to increase DD. To do this we maximize over both position and weights of features.

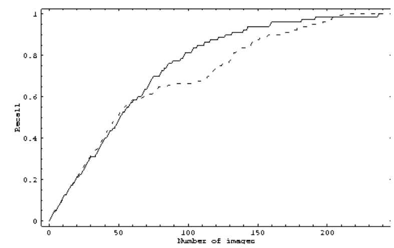
$$\max_{t,s} \prod_i P(t \mid B_i^+) \prod_i P(t \mid B_i^-) = \star \exp(-\| t - B_{ij}^+ \|^2) \star$$

$$\| t - B_{ij}^+ \|^2 = \sum_k w_k (t_k - B_{ijk}^+)^2$$

MAXIMIZING DD : max DD point will be located near a positive cluster of instances. So, if we perform gradient ascent from every positive instance, one of them will start very close to the peak.

GRAPHS
Comparison with hand-crafted templates

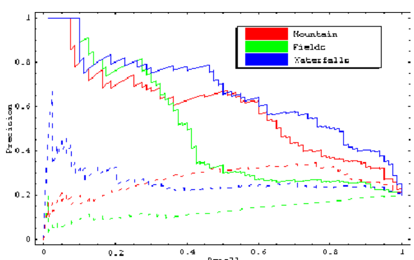


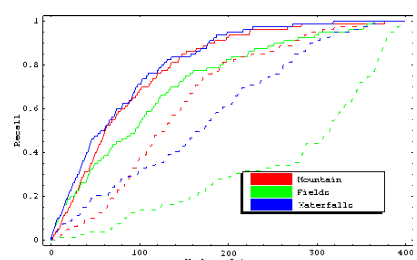


Precision-recall and recall curves comparing performance of learned concept (solid curve) with hand-crafted templates (dashed curve)

Best curves for 3 concepts (fields, mountains, waterfalls)
Dashed curves are results using global histogramming.

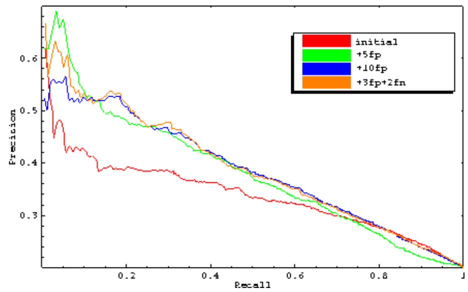
Dataset = 400 images





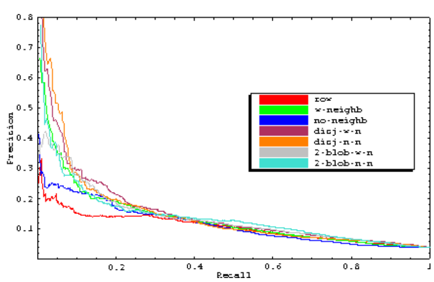
Precision curve for different training schemes averaged over concepts and hypotheses.

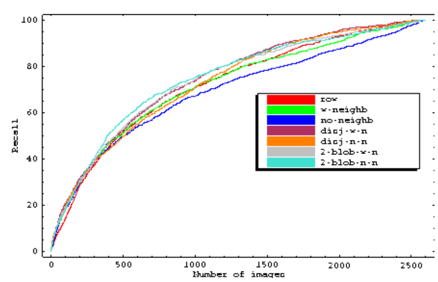
Dataset = 400 images



Different hypothesis classes averaged over concept and training schemes.

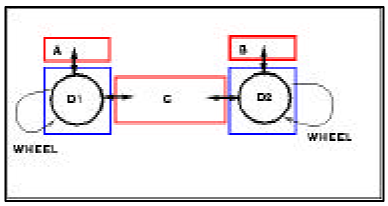
Dataset = 2600 images



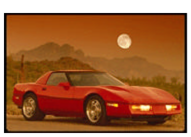


Car Detection

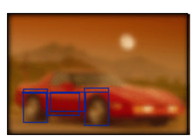
- Task:
  - to detect side-views of vehicles in images
- Basic template consists of:
  - photometric and spatial relations between image patches
  - Geometric primitives (wheels)



Vehicle detection: Example

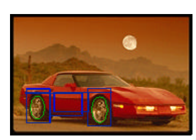


Original image



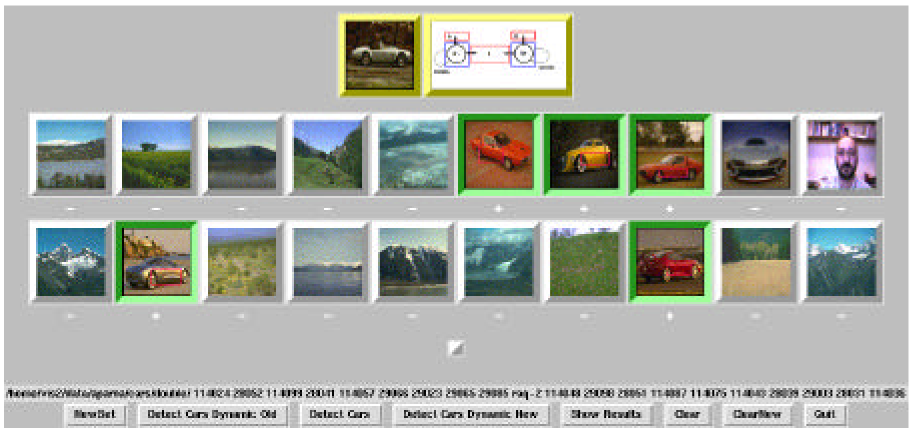
Region selection
using patch properties

Wheel detection



Model superimposed on
detected car

Some Results



Detected cars are highlighted in green.

Results on video sequence

Frames with cars are highlighted in green