CONTENT-BASED IMAGE RETRIEVAL: LET THE IMAGES DO THE TALKING!

NAME: BINDITA CHAUDHURI

ROLL NUMBER: 143070008

M. Tech., 2ND Year, Communication Engineering

Department of Electrical Engineering, IIT Bombay, Mumbai - 400076

- Introduction
- Details of the general methodology
- Specific examples
- Application to remote sensing images
- Description of the proposed approach
- Experimental Analysis and Results
- Future Work

What is Content-Based Image Retrieval?

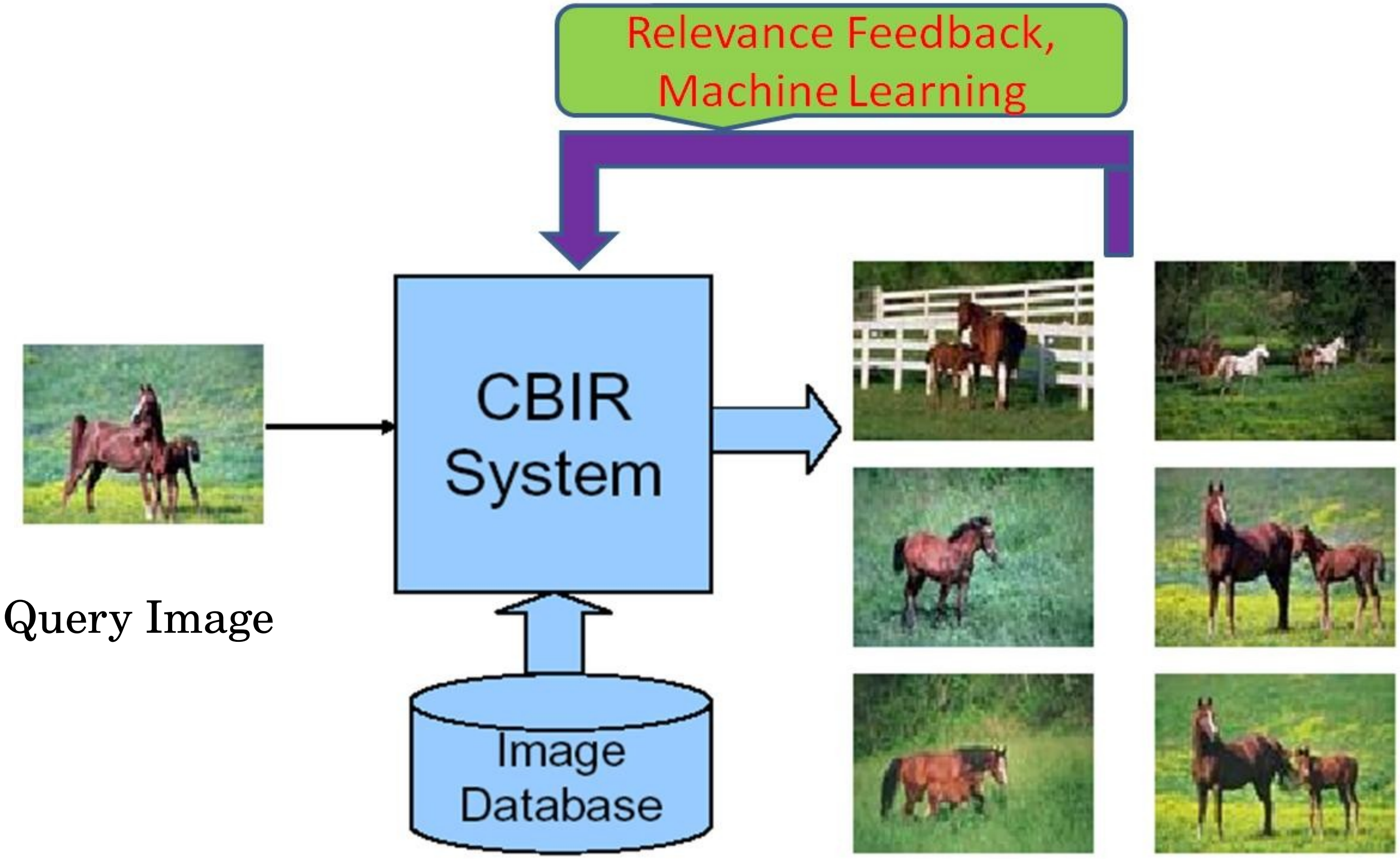
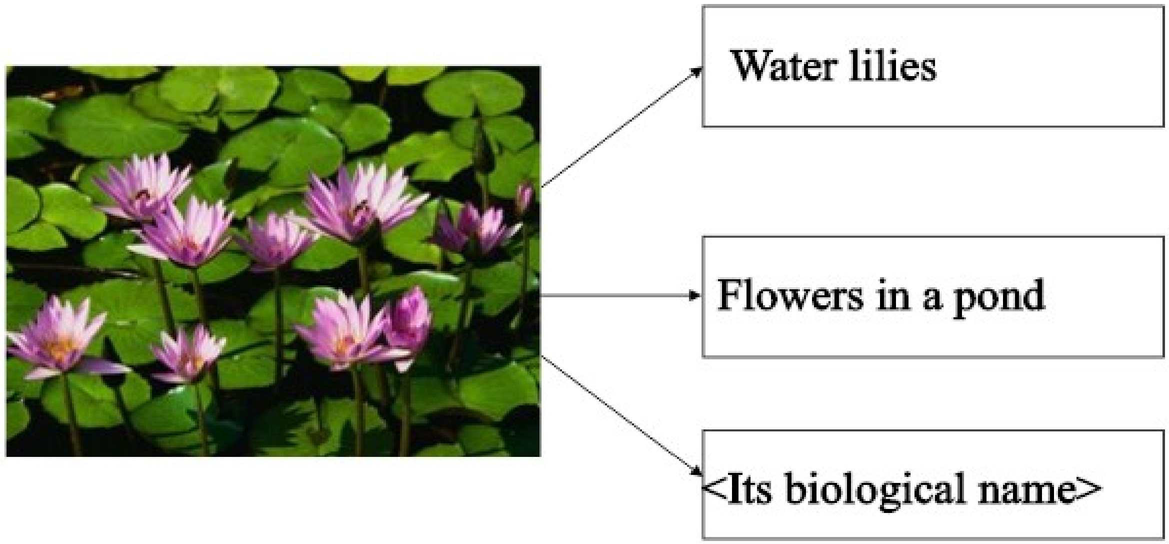


Image Courtesy: http://slideplayer.com/slide/4783566/

Why Content-Based?

Early techniques : text-based image retrieval using <u>textual annotations</u>

E.g.



Disadvantages:-

1) Annotating large databases accurately not feasible.
2) Remarks often ambiguous due to human perception
3) Valid for a particular language only

Image Courtesy: http://slideplayer.com/slide/4783566/

Example of incorrect retrieval

(a) Results for the query on a popular image search engine.

(b) Expected results for the query.

Department of Electrical Engineering, IIT Bombay, Mumbai - 400076

METHODOLOGY

Typical CBIR

↓

Feature Extraction (Color, Texture, Shape, Spectral)

↓

Image Matching (query matching with database)

Steps involved in CBIR:

1) Features are extracted from each image in the database
2) Same features are extracted from the query image
3) Similarity between query features and database features are compared
4) Images with maximum similarity are retrieved
5) Users' relevance feedback is used to refine the retrieval results.

- Introduction
- Details of the general methodology
- Specific examples
- Application to remote sensing images
- Description of the proposed approach
- Experimental Analysis and Results
- Future Work

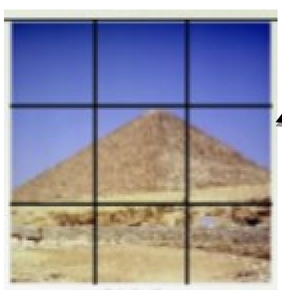
What Are Image Features?

❖ Represent the visual characteristics of an image

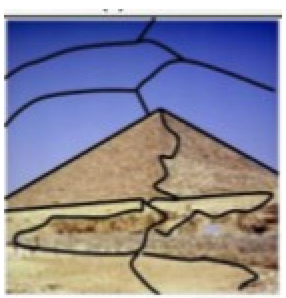
❖ **Low level** features: example – round ball
❖ **High level** features: example – white golf ball/red cricket ball

❖ Two major types of features based on method of extraction:-

• **Global** features: 1 vector from whole image
1) Computationally simple
2) Unable to capture spatial information - semantic gap
•) **Local** features: 1 vector per image segment (tile or region or object)



Tiling (regular)



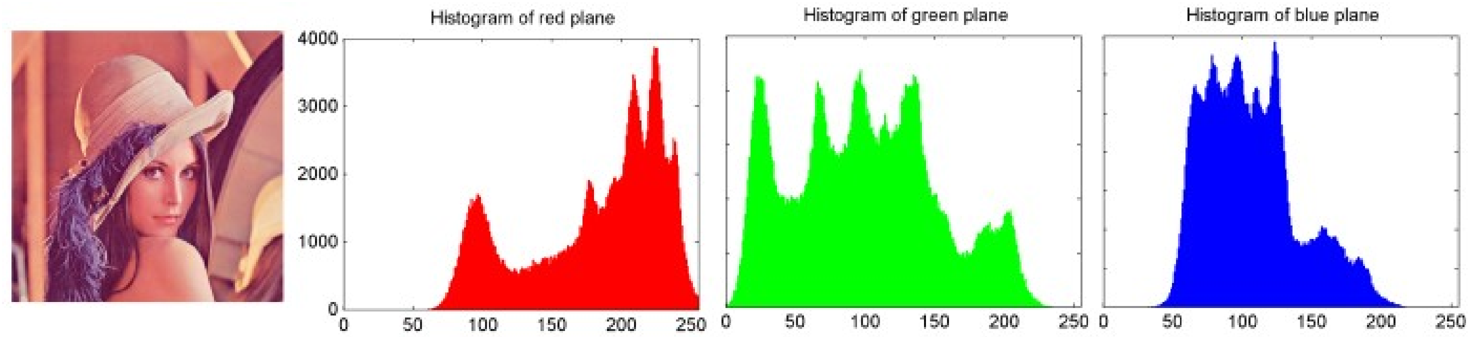
Regioning (irregular)

COLOUR FEATURE

- Each pixel in an image is represented as a point in the 3-D colour space.

- Commonly used colour spaces: **RGB** (Red, Green and Blue), **HSV** (Hue, Saturation and Value), **YCbCr** (luminance and chrominance) etc.

- Commonly used color features: color histogram, color moments

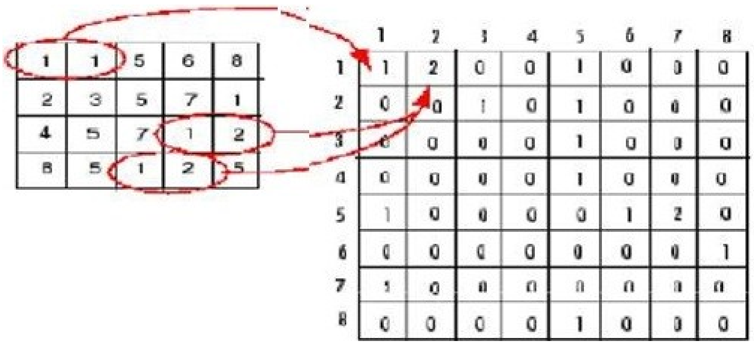


- **Colour Correlogram** incorporates *spatial correlation* of colour pairs in the *colour histogram*

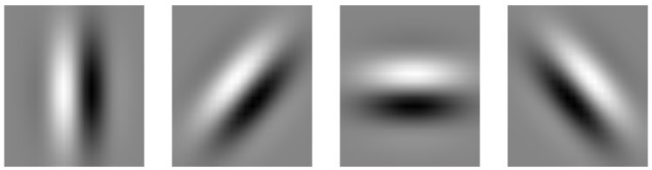
TEXTURE FEATURE

- **Texture**: Regular repetition of an element or pattern on a surface.

- Commonly used methods:
  - **GLCM:**
    Gray-level co-occurrence matrix



  - **Gabor** transformation:

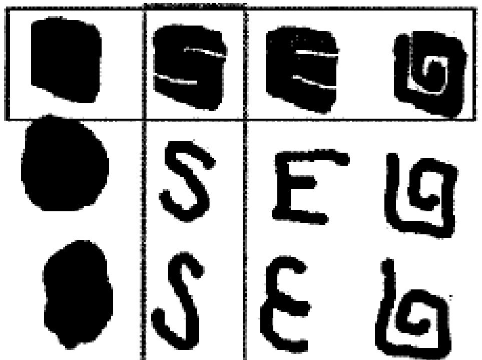


  - **Wavelets** like Discrete Wavelet Transform (DWT), Curvelet transform, Ridgelet transform, Fast wavelet histogram techniques (FWHT) etc.

SHAPE FEATURE

➢ Inherent property; unaffected by external conditions like illumination variations, 3D viewpoint changes etc.

➢ **Common contour-based descriptors**: Area, Perimeter, Bounding Area, Solidity etc.



➢ **Effective descriptors**: *Fourier descriptors*

❑ Fourier transform of the boundary pixels.

$$s(t) = ([x(t) - x_c]^2 + [y(t) - y_c]^2)^{\frac{1}{2}}$$

❑ Translation-invariant contour representations.
❑ Rotation invariance - by taking only amplitudes of coefficients.
❑ Scale invariance - by normalizing the coefficients.

OTHER FEATURES

- **SPECTRAL FEATURES:** Remote Sensing images are generally *multispectral* or *hyperspectral*. Such images are captured by measuring the reflected amount of signal of a range of frequencies.

- Image retrieval is done on the basis of relationship among different constituent bands in the spectral curve.

- **SPATIAL FEATURES:** Similar objects are distinguished using spatial location (coordinate), local/global spatial relationships etc.

- **METADATA:** Metadata is data about data. Examples are date and time of image capture, photographer's name etc. It can be used as a preliminary coarse filter to reduce retrieval scope and improve efficiency.

IMAGE MATCHING

Similarity measurement: Feature vectors extracted from query image are compared with the features vectors of the database by measuring the distance between the vectors. Smaller distance implies higher similarity.

Euclidean (or Cartesian) distance	$D_{[2]}(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$
Chebyshev distance	$D_{[\infty]}(\mathbf{x},\mathbf{y}) = \max_{i=1}^{n} \lvert x_i - y_i \rvert$
Manhattan (city-block) distance	$D_{[1]}(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n} \lvert x_i - y_i \rvert$
Minkowsky distance	$D_{[p]}(\mathbf{x},\mathbf{y}) = \left[\sum_{i=1}^{n} \lvert x_i - y_i \rvert^p\right]^{\frac{1}{p}}$
Weighted Minkowsky distance	$D_{[p,w]}(\mathbf{x},\mathbf{y}) = \left[\sum_{i=1}^{n} w_i \lvert x_i - y_i \rvert^p\right]^{\frac{1}{p}}$
Mahalanobis distance	$D(\mathbf{x},\mathbf{y}) = \lvert \det \mathrm{C} \rvert^{1/n} (\mathbf{x}-\mathbf{y})^{\mathrm{T}} \mathrm{C}^{-1}(\mathbf{x}-\mathbf{y})$
Generalised Euclidean (quadratic) distance	$D(\mathbf{x},\mathbf{y}) = (\mathbf{x}-\mathbf{y})^{\mathrm{T}} \mathrm{K}(\mathbf{x}-\mathbf{y})$
Correlation coefficient	$\rho(\mathbf{x},\mathbf{y}) = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_i)(y_i - \bar{y}_i)}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x}_i)^2 \sum_{i=1}^{n}(y_i - \bar{y}_i)^2}}$
Relative entropy (Kullback-Leibler divergence)	$D(\mathbf{x} \Vert \mathbf{y}) = \sum_{i=1}^{n} x_i \log \frac{x_i}{y_i}$ when $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i = 1$
χ^2-Distance	$D_{\chi^2}(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n} \frac{(x_i - y_i)^2}{y_i}$ when $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i = 1$

POSSIBLE IMPROVEMENTS

❖ Use of improved features:

- Scale-invariant feature transform (**SIFT**) (described later)
- Histogram of Oriented Gradients (**HOG**)
- Local Binary Patterns (**LBP**) and Local Phase Quantization (**LPQ**)
- **GIST** descriptors

❖ Dimension reduction of feature vectors using -

- Principal Component analysis (PCA) (*linear*)
- Laplacian eigenmaps (*non-linear*)

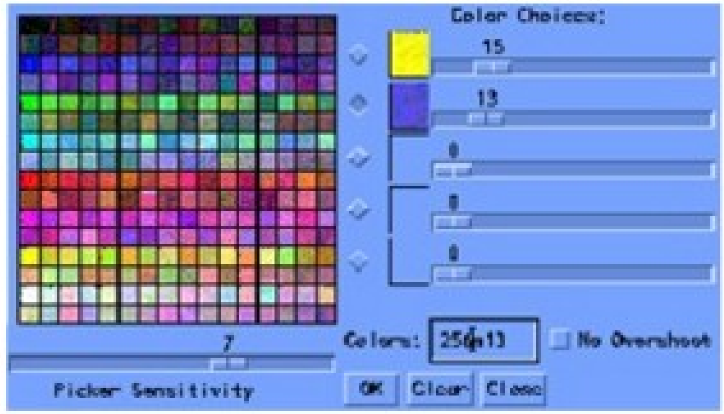
❖ Active Learning using **Relevance feedback**: User interacts with the system to **refine** the results until he/she is satisfied.

❖ Either user explicitly marks images as relevant/irrelevant

❖ Or implicitly inferred by the system from previous settings of the user.

- Introduction
- Details of the general methodology
- Specific examples
- Application to remote sensing images
- Description of the proposed approach
- Experimental Analysis and Results
- Future Work

Query by Colour in QBIC



** Images courtesy : Yong Rao

Query by Shape in QBIC

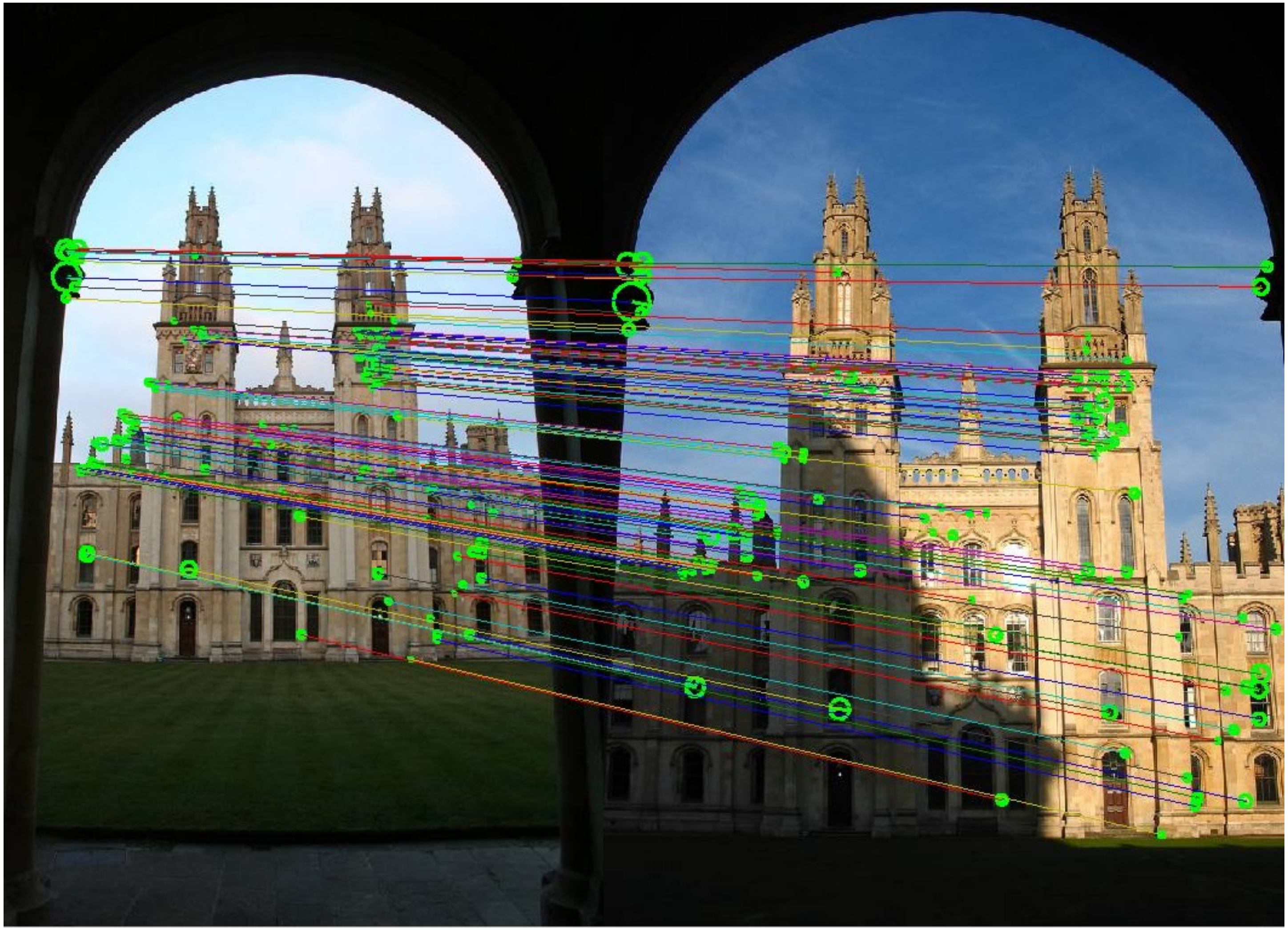
** Images courtesy : Yong Rao

Department of Electrical Engineering, IIT Bombay, Mumbai - 400076

Query by Sketch in QBIC

** Images courtesy : Yong Rao

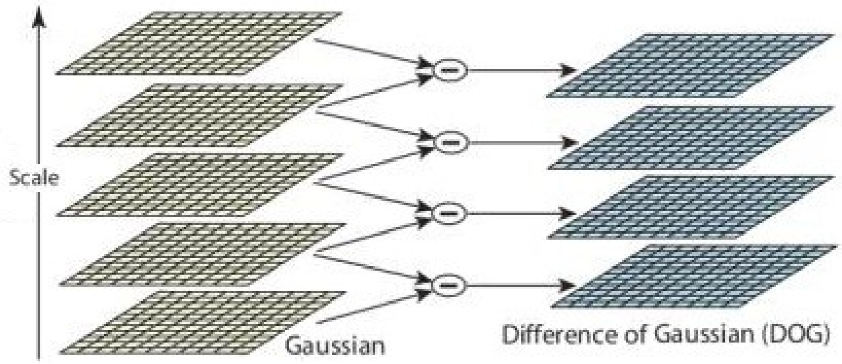
Scale-Invariant Feature Transform



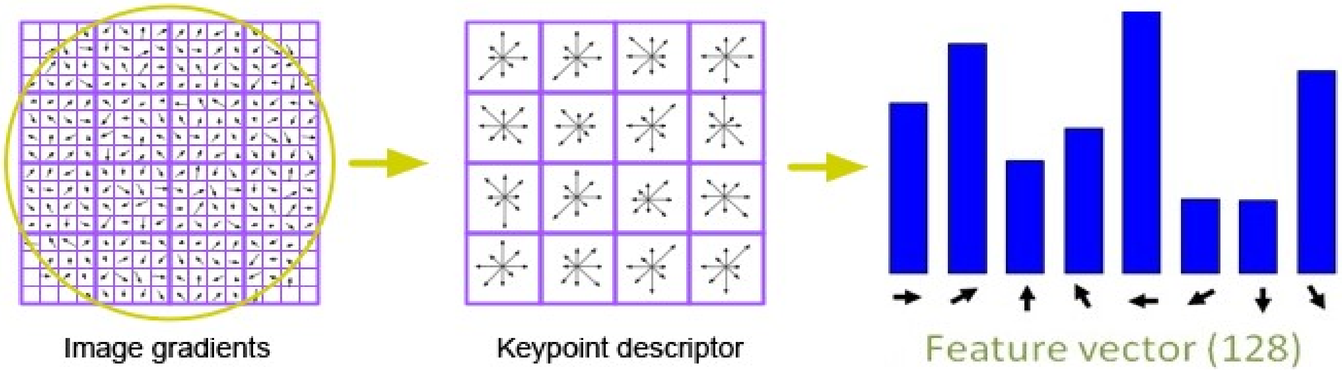
SIFT DESCRIPTORS

- Computes a **128-dimensional** feature vector at each interest point.

- Interest points detected by scale-space extrema detection.

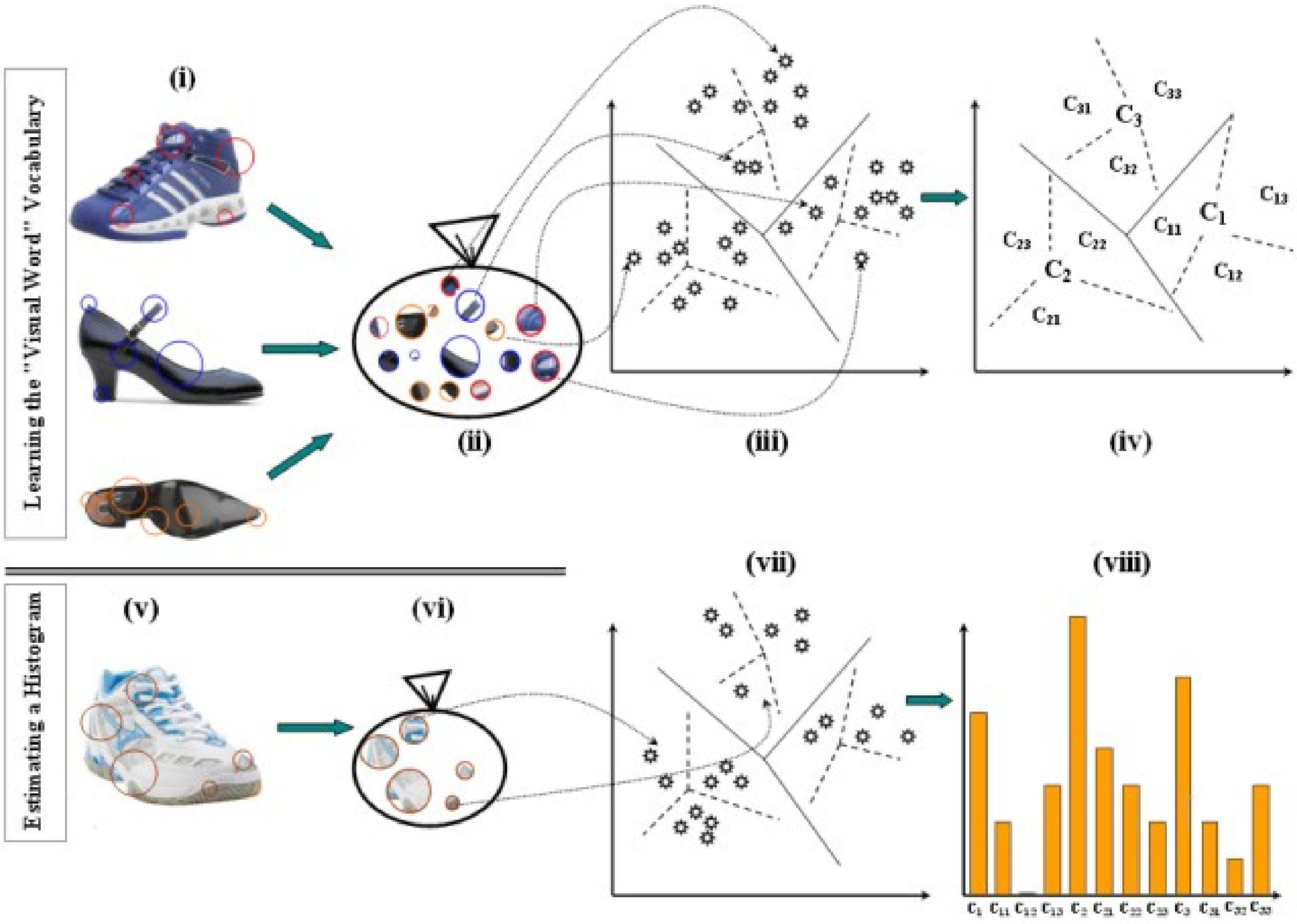


- Keypoint descriptors: (orientation histograms)



RETRIEVAL USING INTEREST POINTS

Bag-of-visual-words (BoVW) model for image representation



- Introduction
- Details of the general methodology
- Specific examples
- Application to remote sensing images
- Description of the proposed approach
- Experimental Analysis and Results
- Future Work

PROBLEM DEFINITION

❖ **Motivation:** Rapid **expansion** of Remote sensing (RS) image archives. Hence the need of a system to store, index, browse and retrieve information from large databases.

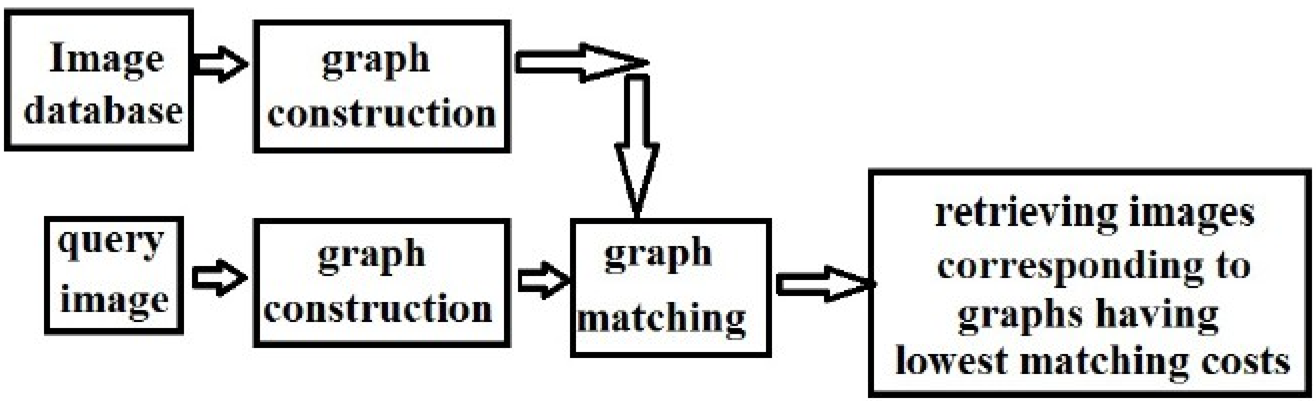
❖ **Challenge**: Unlike normal scene images, special characteristics of structures are absent in RS images since they are taken from top view.

❖ Region-based methods – group pixels with similar characteristics
A graph is created from an image with **regions** as **nodes** and their **spatial relationship** as **edges**. Image matching then becomes graph matching.

❖ **Speciality** of the proposed method – fully unsupervised. No labeled samples or training data required.

STEPS OF THE PROPOSED METHOD

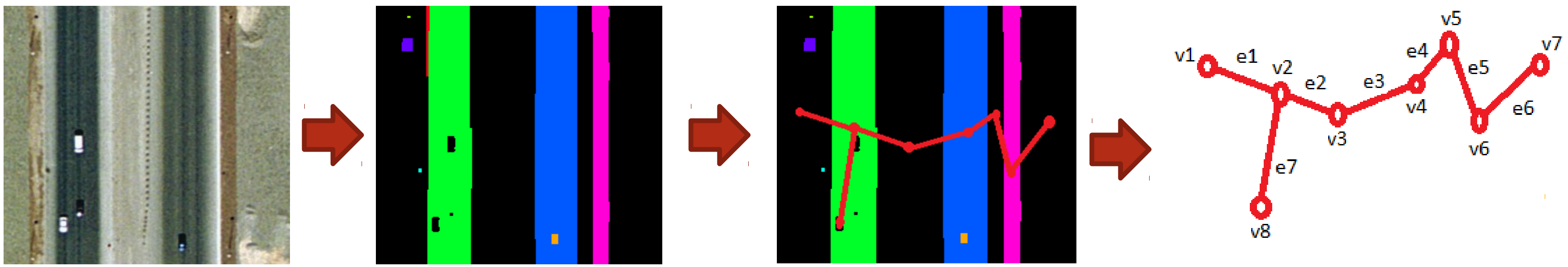
(a) Flowchart of the graph creation process



(b) Block diagram of the proposed method

- Introduction
- Details of the general methodology
- Specific examples
- Application to remote sensing images
- Description of the proposed approach
- Experimental Analysis and Results
- Future Work

Attributed Relational Graph Construction



Original image | Segmented image | Feature extraction | Graph creation

NODE FEATURES:-

- Color Moments, Contour-based Shape features & Fourier descriptors
- **Spectral Histogram**: 1) *Intensity filter* – image itself

2) *Laplacian of Gaussian* filters – variances 0.2 and 1

3) *Gabor filters* – orientation – $0^o, 45^o, 90^o, 135^o$

EDGE FORMATION:-

- Firstly, a Region Adjacency Graph **(RAG)** is created from nodes:

N × N adjacency matrix; A_{ij} = 1 if regions i and j are adjacent, otherwise 0

- **Edge Attribute definition:** $\mathbf{W}_i(v_i^s, v_i^t) = \alpha_1 \, ||c_{v_i^s} - c_{v_i^t}||_2 + \alpha_2 \, |\theta_{v_i^s} - \theta_{v_i^t}|$

Inexact Graph Matching

- **Subgraph Isomorphism**: used for **node matching** only
- Each node of query graph matched to most similar node of model graph having minimum distance measure;

$$dist = dist'_{L1_{moments}} + dist'_{L1_{fourier}} + dist'_{L2_{entropy}} + dist'_{L2_{contour}} + dist_{chi_{colorhist}} + dist_{chi_{spectral}}$$

- Total node error = sum of the matched node distances / number of nodes

- **Spectral Embedding**: Used for **edge matching** [*Umeyama, 1988*]
- W_1 and W_2 – adjacency matrices of two graphs G_1 and G_2

  P – permutation matrix

Graph error ⟵ $$||PW_1P^T - W_2||_F^2 \geq \sum_{i=1}^{n}(\alpha_i - \beta_i)^2$$

- **K = min(m, n)** smallest non-null eigenvalues taken to calculate error

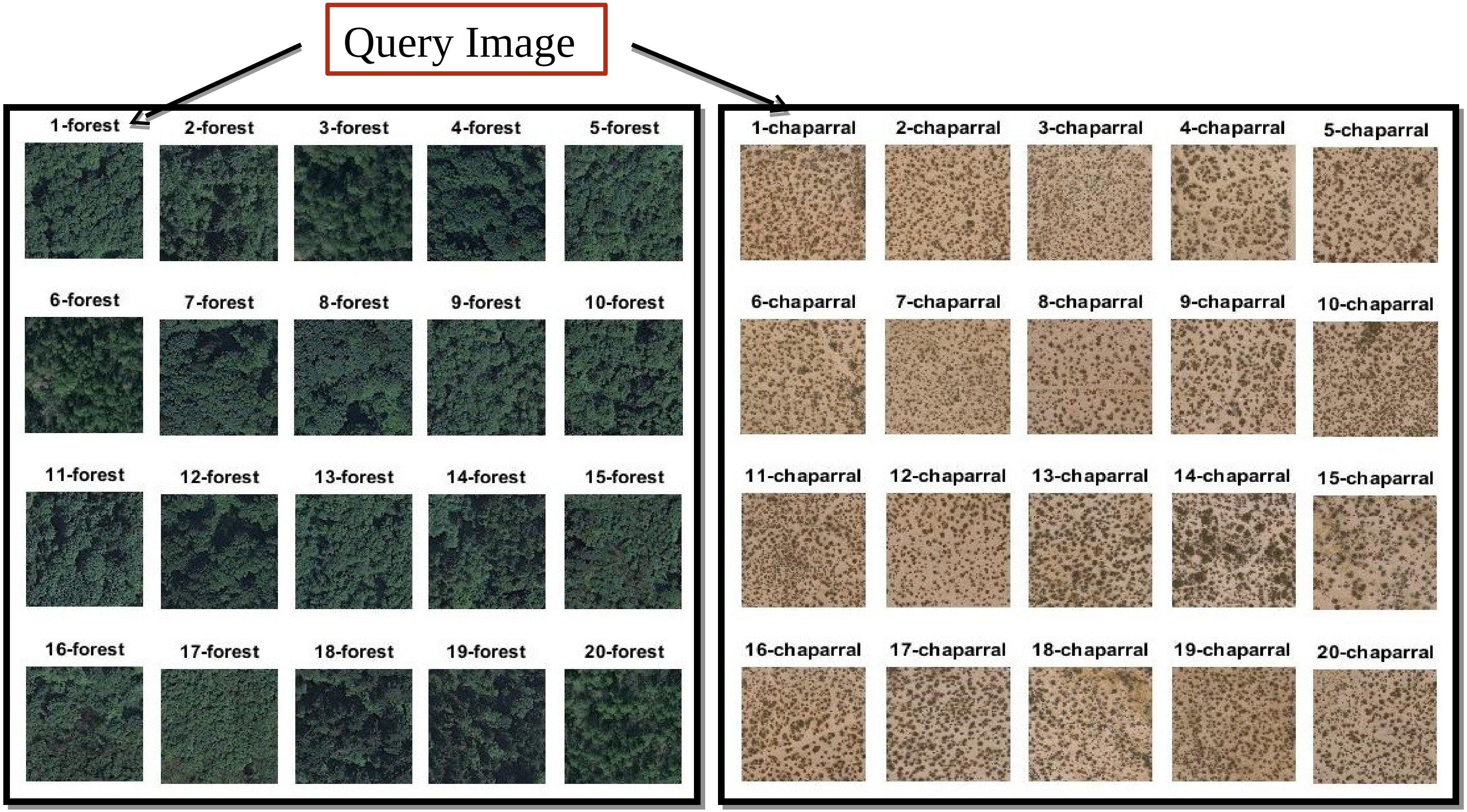
- Total graph error = node+edge error. Minimum error = maximum similarity.

- Introduction
- Details of the general methodology
- Specific examples
- Application to remote sensing images
- Description of the proposed approach
- Experimental Analysis and Results
- Future Work

DATASET DESCRIPTION

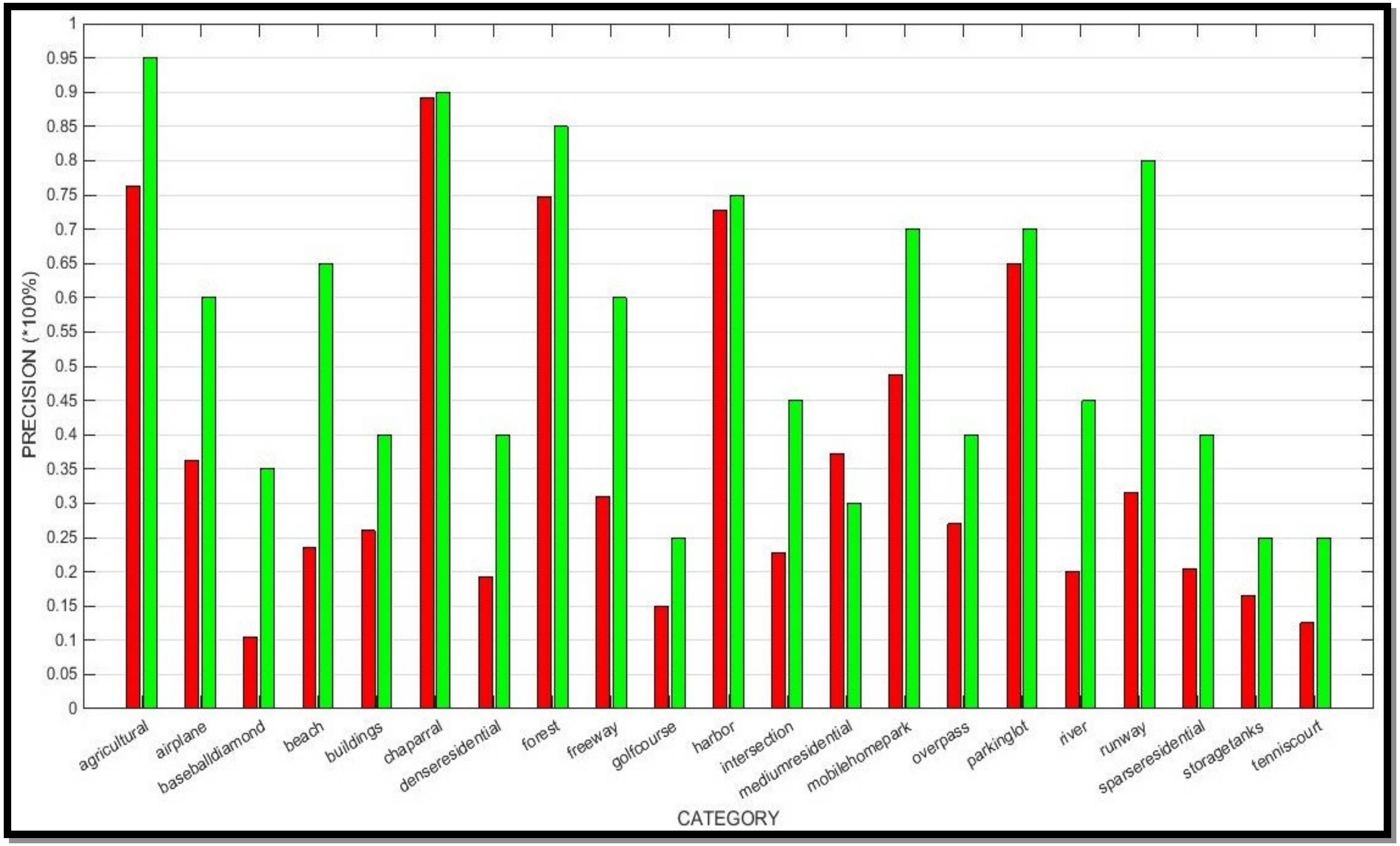
- UCMERCED archive
- 21 Land use/cover classes
- 100 images per category
- Image Size 256 × 256
- Spatial resolution: 30 cm.

RETRIEVAL RESULTS (TOP 20)



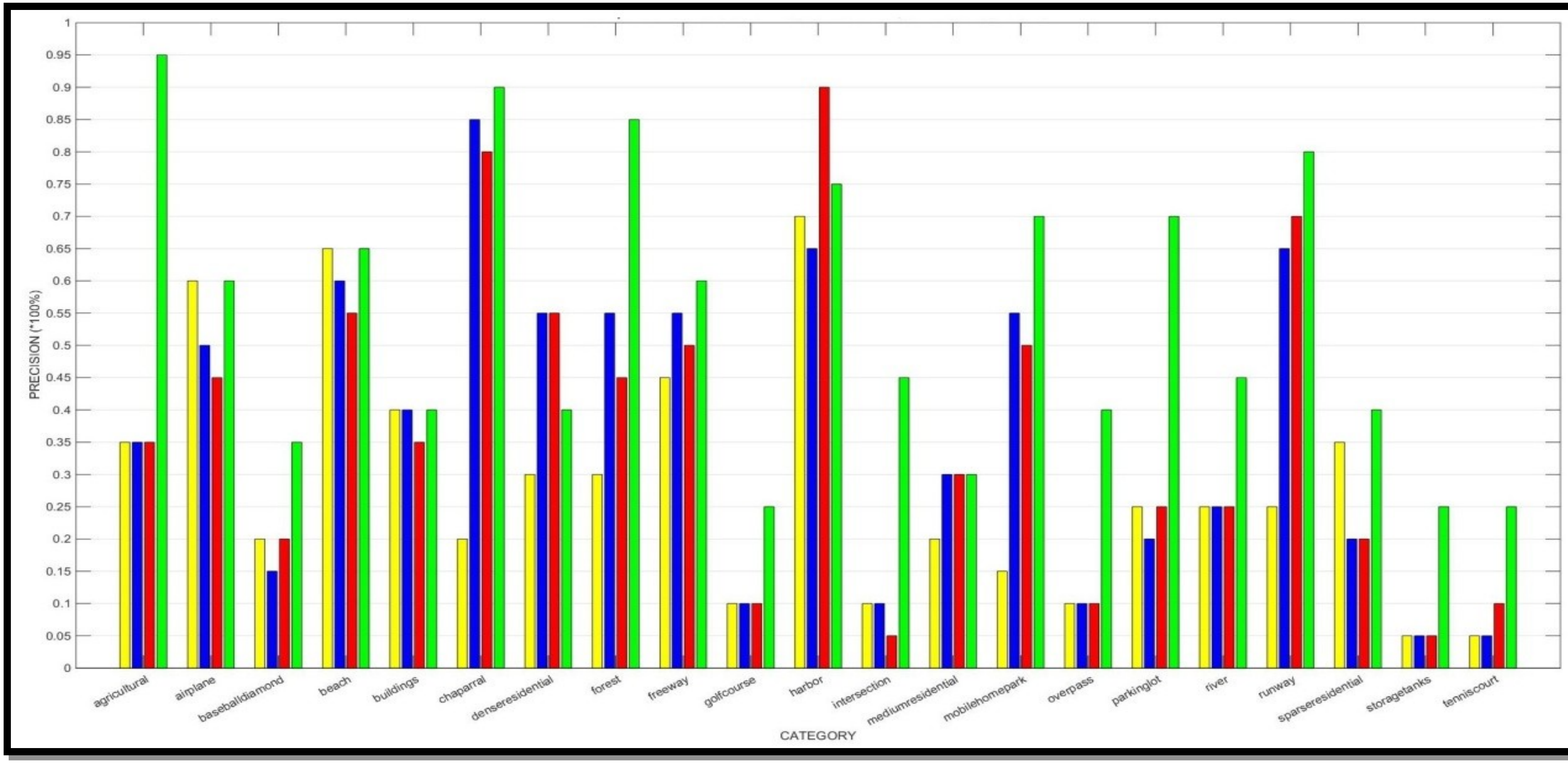
Forest category **Chaparral category**

PRECISION COMPARISONS



Comparison of SIFT-based method (red) and proposed method (green)

PRECISION COMPARISONS (cont.)



Comparison of the effect of individual features and that of all the features taken together. Yellow – color features, Blue - texture features, Red – shape features, Green – combined features

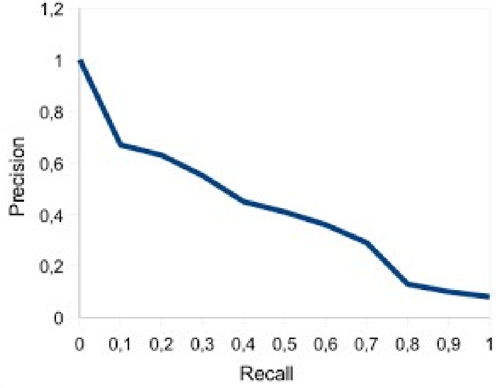
OTHER ACCURACY MEASURES

- **Confusion matrix:** given in next slide.
- **Recall**: $\frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images}}$

- **ANMRR** (Average Normalized Modified Retrieval Rank): takes into account the exact rank/position of the retrieved images.

  Lower ANMRR → higher accuracy

- **P-R curve**: Plot of precision vs. recall. Gradually goes down, but higher the values, the better.



CONFUSION MATRIX

Output Class \ Target Class	agricultural	airplane	baseballdiamond	beach	buildings	chaparral	denseresidential	forest	freeway	golfcourse	harbor	intersection	mediumresidential	mobilehomepark	overpass	parkinglot	river	runway	sparseresidential	storagetanks	tenniscourt	
agricultural	19 4.5%	0 0.0%	1 0.2%	4 1.0%	0 0.0%	1 0.2%	0 0.0%	2 0.5%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	4 1.0%	0 0.0%	1 0.2%	0 0.0%	2 0.5%	54.3% 45.7%
airplane	0 0.0%	12 2.9%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	3 0.7%	0 0.0%	0 0.0%	4 1.0%	0 0.0%	2 0.5%	0 0.0%	52.2% 47.8%
baseballdiamond	0 0.0%	0 0.0%	10 2.4%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	6 1.4%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	2 0.5%	3 0.7%	45.5% 54.5%
beach	0 0.0%	0 0.0%	0 0.0%	14 3.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
buildings	0 0.0%	0 0.0%	0 0.0%	0 0.0%	8 1.9%	0 0.0%	5 1.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	4 1.0%	3 0.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	40.0% 60.0%
chaparral	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	18 4.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	4 1.0%	2 0.5%	0 0.0%	75.0% 25.0%
denseresidential	0 0.0%	0 0.0%	0 0.0%	0 0.0%	3 0.7%	0 0.0%	8 1.9%	0 0.0%	0 0.0%	0 0.0%	3 0.7%	0 0.0%	5 1.2%	2 0.5%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	38.1% 61.9%
forest	1 0.2%	0 0.0%	2 0.5%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	17 4.0%	0 0.0%	4 1.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	5 1.2%	0 0.0%	0 0.0%	2 0.5%	4 1.0%	48.6% 51.4%
freeway	0 0.0%	2 0.5%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	13 3.1%	0 0.0%	0 0.0%	2 0.5%	0 0.0%	0 0.0%	6 1.4%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	2 0.5%	50.0% 50.0%
golfcourse	0 0.0%	0 0.0%	5 1.2%	2 0.5%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	7 1.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	4 1.0%	38.9% 61.1%
harbor	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	15 3.6%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	2 0.5%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	88.2% 11.8%
intersection	0 0.0%	1 0.2%	0 0.0%	0 0.0%	2 0.5%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	9 2.1%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	1 0.2%	0 0.0%	60.0% 40.0%
mediumresidential	0 0.0%	0 0.0%	0 0.0%	0 0.0%	5 1.2%	0 0.0%	4 1.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	2 0.5%	6 1.4%	4 1.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	6 1.4%	0 0.0%	0 0.0%	22.2% 77.8%
mobilehomepark	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	3 0.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	5 1.2%	14 3.3%	0 0.0%	2 0.5%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	58.3% 41.7%
overpass	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	5 1.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	9 2.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	60.0% 40.0%
parkinglot	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	2 0.5%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	15 3.6%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	88.2% 11.8%
river	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	10 2.4%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	76.9% 23.1%
runway	0 0.0%	4 1.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	16 3.8%	0 0.0%	0 0.0%	0 0.0%	80.0% 20.0%
sparseresidential	0 0.0%	0 0.0%	0 0.0%	0 0.0%	2 0.5%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	1 0.2%	0 0.0%	2 0.5%	0 0.0%	0 0.0%	2 0.5%	0 0.0%	0 0.0%	0 0.0%	8 1.9%	4 1.0%	0 0.0%	40.0% 60.0%
storagetanks	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	6 1.4%	0 0.0%	100% 0.0%
tenniscourt	0 0.0%	0 0.0%	1 0.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	2 0.5%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	5 1.2%	62.5% 37.5%
	95.0% 5.0%	60.0% 40.0%	50.0% 50.0%	70.0% 30.0%	40.0% 60.0%	90.0% 10.0%	40.0% 60.0%	85.0% 15.0%	65.0% 35.0%	35.0% 65.0%	75.0% 25.0%	45.0% 55.0%	30.0% 70.0%	70.0% 30.0%	45.0% 55.0%	75.0% 25.0%	50.0% 50.0%	80.0% 20.0%	40.0% 60.0%	30.0% 70.0%	25.0% 75.0%	56.9% 43.1%

DRAWBACKS

Query Image
1-mobilehomepark
2-mobilehomepark
3-mobilehomepark
4-parkinglot
5-denseresidential
6-mediumresidential
7-denseresidential
8-harbor
9-mobilehomepark
10-mobilehomepark
11-denseresidential
12-denseresidential
13-mediumresidential
14-mobilehomepark
15-mediumresidential
16-mobilehomepark
17-mobilehomepark
18-mediumresidential
19-denseresidential
20-mediumresidential

- Introduction
- Details of the general methodology
- Specific examples
- Application to remote sensing images
- Description of the proposed approach
- Experimental Analysis and Results
- Future Work

FUTURE WORK

❖ Incorporate **automatic feature learning**. For e.g., understand which regions are texture-intensive and which are color-intensive and weigh the features accordingly.

❖ **Learn** an accurate representative **graph** for each category.
Given a template graph, iteratively update graph attributes using user-given positive and negative examples.

❖ Apply **weakly supervised learning** (minimal number of training samples). For example, in *visual object recognition*, instead of detailed annotation (location, shape etc.) of objects, only the presence of objects are indicated.

❖ Use **deep learning** features (like Convolutional Neural Networks) and more sophisticated computer vision techniques for improved image characterization.

REFERENCES

[1] R. Cossu, S. Chaudhuri, and L. Bruzzone. "A Context-Sensitive Bayesian Technique for the Partially supervised Classification of Multitemporal Images". *IEEE Geoscience and Remote Sensing Letters*, 2(3):352–356, July 2005.

[2] F. Long, H. Zhang, and D. D. Feng. *Multimedia Information Retrieval and Management*, chapter "Fundamentals of content-based image retrieval", pages 1–26. Springer Berlin Heidelberg, 2003.

[3] B. Demir and L. Bruzzone. "An Effective Active Learning Method for Interactive Content-based Retrieval in Remote Sensing Images". In *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 4356–4359, July 2013.

[4] Wikipedia, The Free Encyclopedia. Machine learning. http://en.wikipedia.org/wiki/Machine_learning.

[5] Wikipedia, The Free Encyclopedia. Pattern recognition. http://en.wikipedia.org/wiki/Pattern_recognition.

[6] Wikipedia, The Free Encyclopedia. Remote sensing. http://en.wikipedia.org/wiki/Remote_sensing.

ANY QUESTIONS?


Thank you...