MINING MEANINGFUL PATTERNS IN LARGE SET OF IMAGES

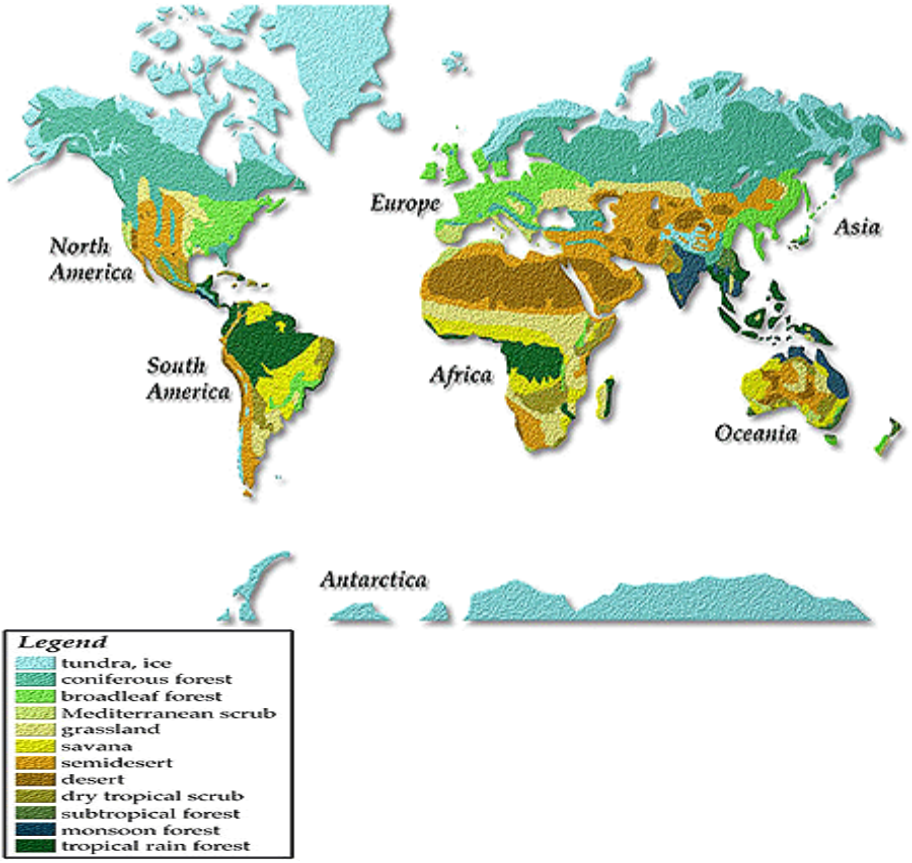
Vishwakarma Singh

2nd September, 2008

Committee: Prof. Ambuj K. Singh (Chair)
Prof. Amr El Abbadi
Prof. B. S. Manjunath

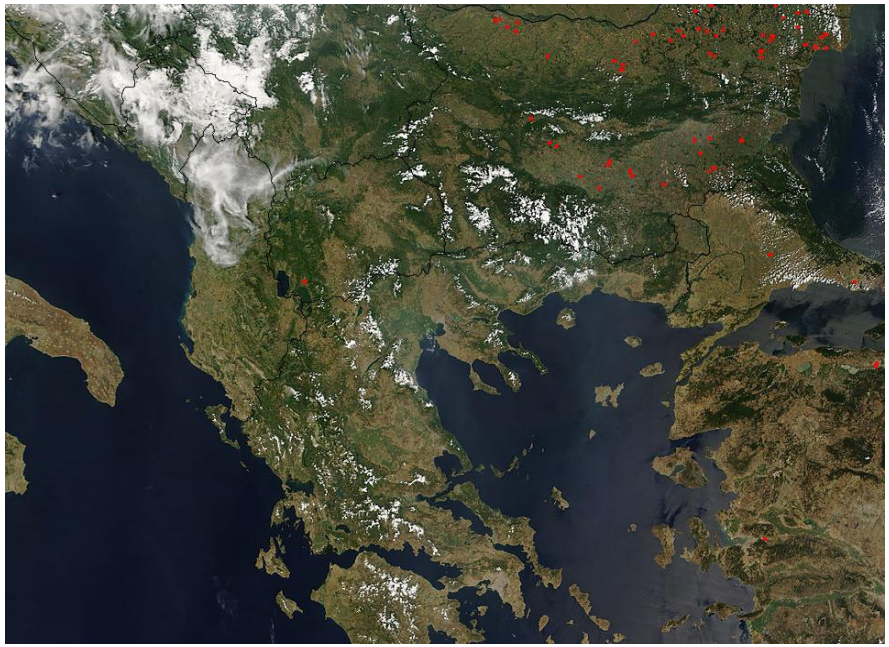
Map Vs Image

Vegetation Map of World

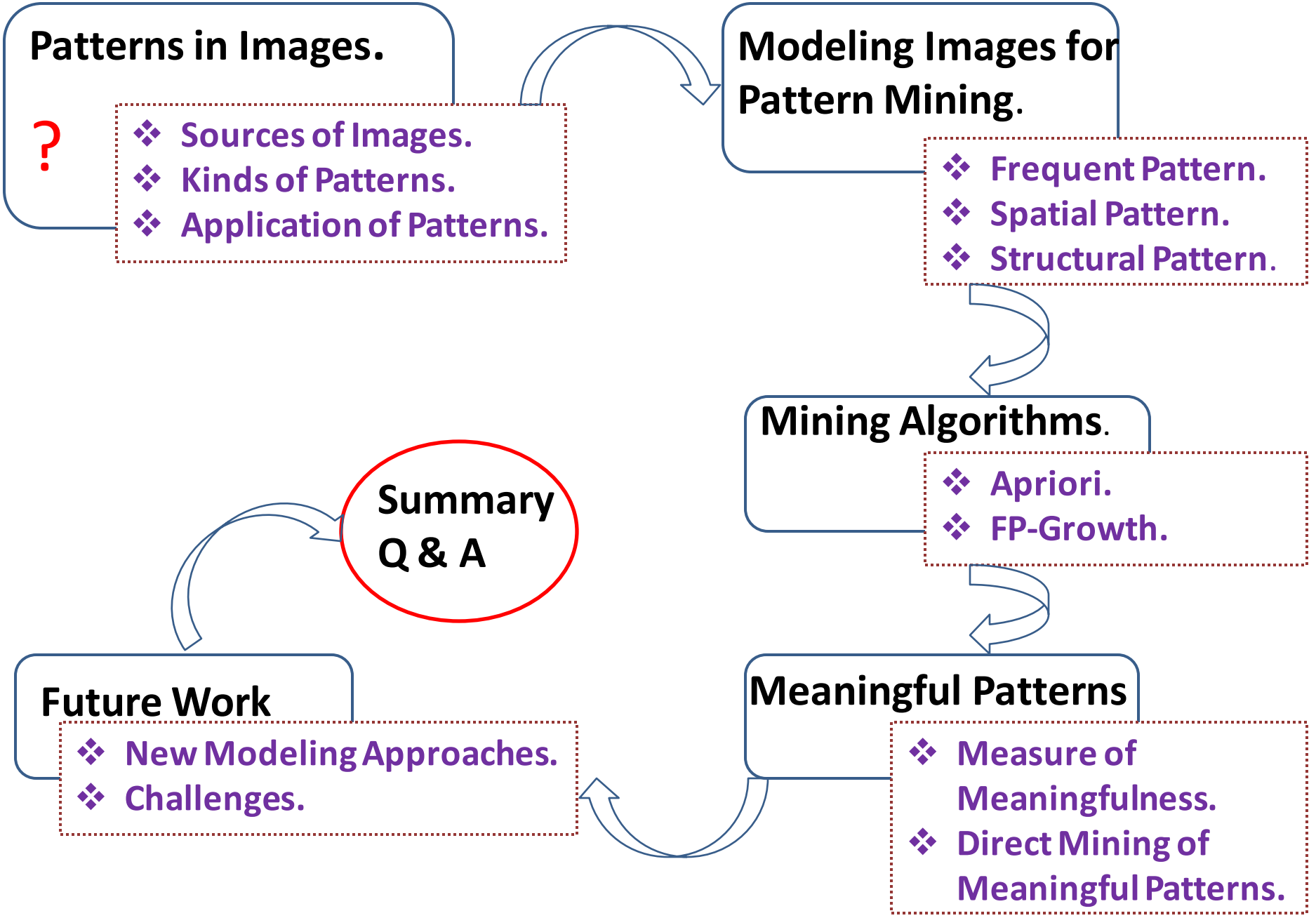


Regions/Objects are predefined with spatial location and semantic attributes.

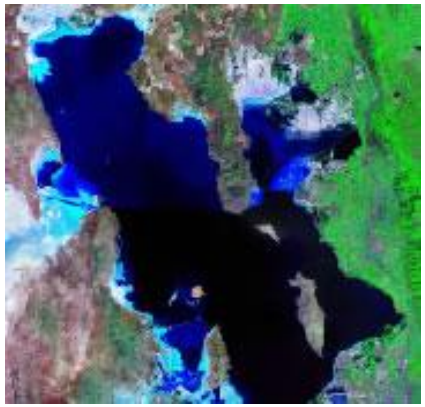
Satellite Image of Greece



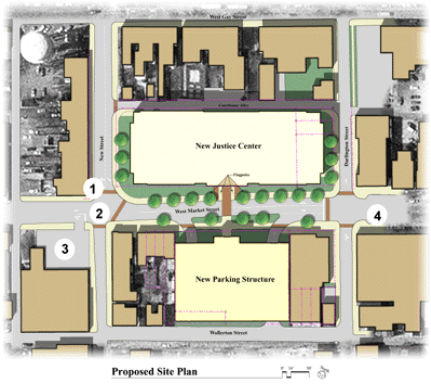
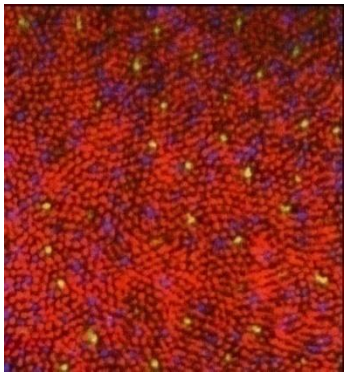
Low level processing of image required to define regions or model image for mining.


Patterns in Images.
?
Sources of Images.
Kinds of Patterns.
Application of Patterns.
Modeling Images for Pattern Mining.
Frequent Pattern.
Spatial Pattern.
Structural Pattern.
Mining Algorithms.
Apriori.
FP-Growth.
Meaningful Patterns
Measure of Meaningfulness.
Direct Mining of Meaningful Patterns.
Future Work
New Modeling Approaches.
Challenges.
Summary
Q & A

Patterns in Images

Images of Sky			
 Image of Orion Constellation	1. Structural relationships invariant over time and space. 2. Recurring cluster of stars.	Field: Astronomy. *[Szalay02, Skycat96]*	Device: Satellite and Telescope. Repositories: Sloan Digital Sky Library .
Remote Sensing Images			
 Satellite Image of Salt Lake, Utah	1. Spatial relationship between bodies. 2. Temporal changes of region. 3. Pattern per pixel.	Field: Geography, Defense, Ecology, Geology, Agriculture, Metrology, Health. *[Jelena03, Ding02, Selim03, Selim07, Silva05, Conquest95]*	Device: Satellite. Repositories: geology.com, terraserver.com

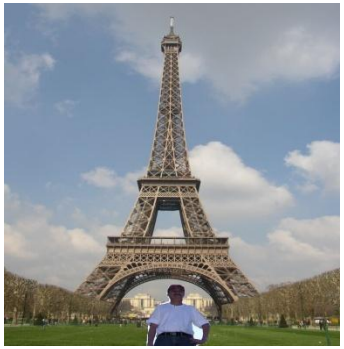
Patterns in Images

Architectural Design			
 Design of a shopping complex	Repeating structures in design and layout.	Field: Architecture, Archeology, Cross Cultural study. *[MLLee03, Shan06, Sangkyum08]*	Synthetic Data
Biological and Medical Images			
 Mosaic of photoreceptors in a Squirrel (Red – rods, Blue- cones)	Frequent relationships between cells in space or interaction.	Field: Biology, Medicine *[Oluknle02, Pan05]*	Device: Microscope, X-Ray, MRI, Ultra Sonography, Computed Tomography, Radiography

Commercial Motivation

Petronas -> Person1
Petronas -> Person2,Person3

Petronas -> Person1, Person2, Person3



Eiffel -> Person1
Eiffel -> Person2,Person3

Eiffel -> Person1, Person2, Person3

Pisa -> Person1
Person2,Person3 -> ?

Note: Frequent pattern based learning can discover potential candidate for marketing.

Repositories: Flickr, Caltech-101

[Gool08]

Application of Patterns

- Classification. *[Cheng07]*
- Clustering. *[Sangkyum08]*
- Indexing and Retrieval. *[Xifeng04, Gabor04, Petrakis97]*
- Automatic Image Captioning. *[Dorado03, Faloutsos04]*
- Market analysis. *[Sri94]*
- Learning underlying processes in medicine. *[Oluknle02, Pan05]*
- Learning Potential Customers.

Patterns in Images.

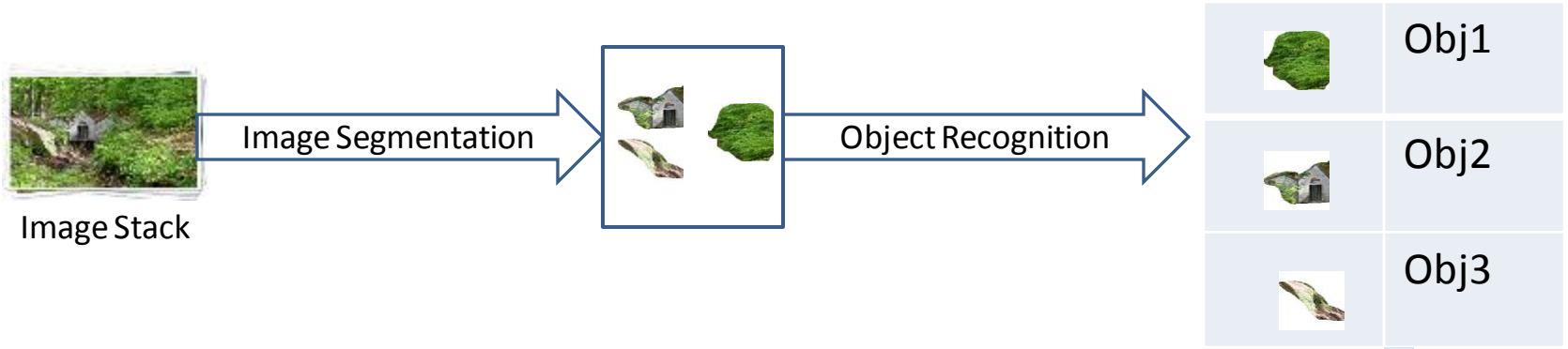
?

- Sources of Images.
- Kinds of Patterns.
- Application of patterns.

Modeling Images for pattern Mining.

- Frequent Pattern.
- Spatial Pattern.
- Structural Pattern.

Modeling Images as Bag of Objects



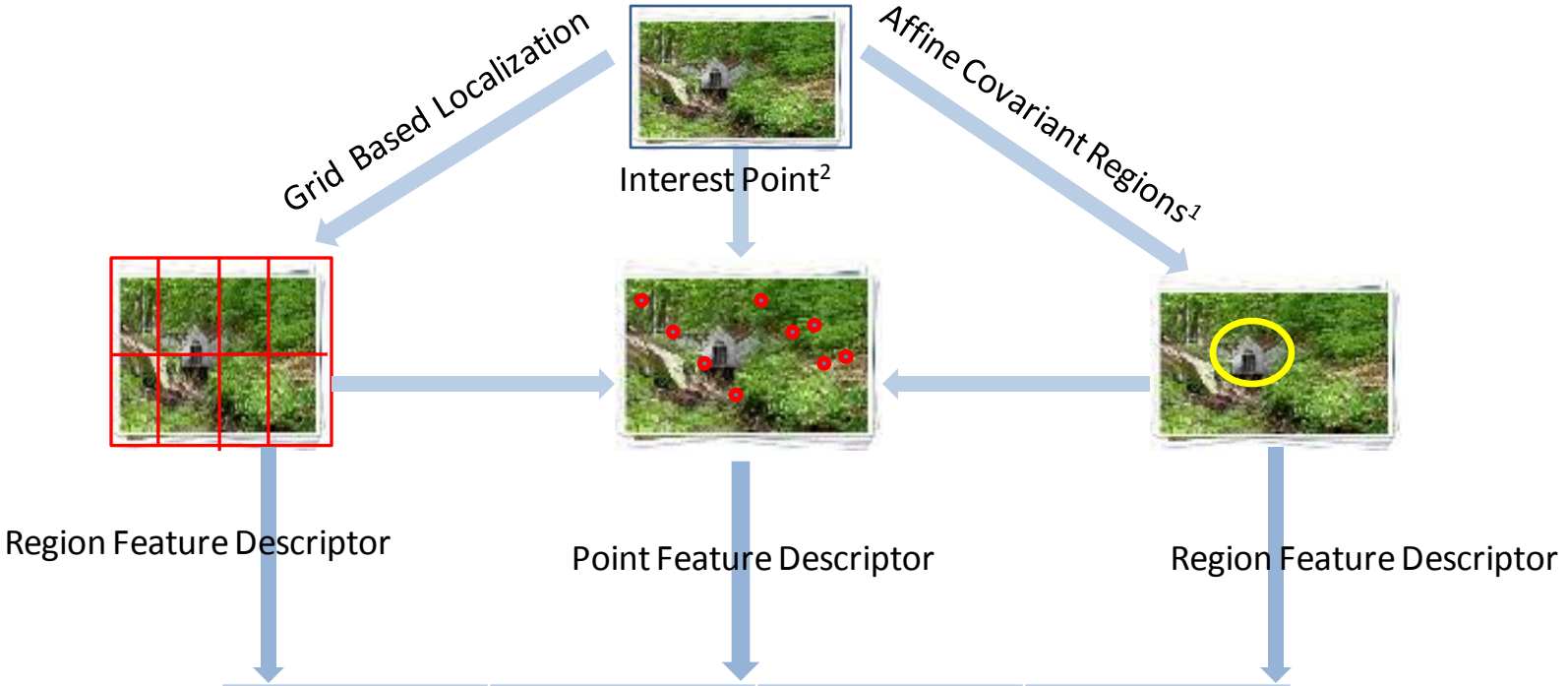
Database of images represented as histogram of objects

Image Id	Obj1	Obj2	Obj3
1	3	0	1
2	1	4	1
3	0	0	2
4	2	0	5

Apriori or **FP-Growth** based Algorithm for Frequent Pattern Mining in Histogram *[Zaine00, Pan05]*

Note: Segmentation is computationally intensive, Definition of object is subjective.

Modeling Images via Feature Descriptors



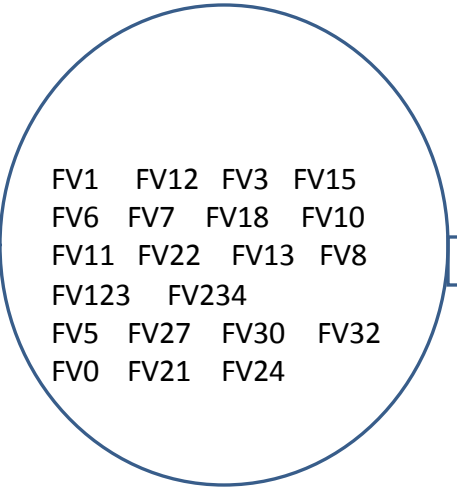
Feature Id	Image ID	Region/Point	N-Dim Feature
FV1	1	1	123,24,…
FV2	1	2	1,255,…
.		.	.
FVN	M	1	24,156,…

DB: Database of feature descriptors

Region Feature Descriptor: {CSD, CLD, HTF}[3] Point Feature Descriptor: {SIFT}[4]

1. *[Sivic04][Gool06]* 2. *[Yang07]* 3.*[Mpeg7]* 4. *[Lowe04]*

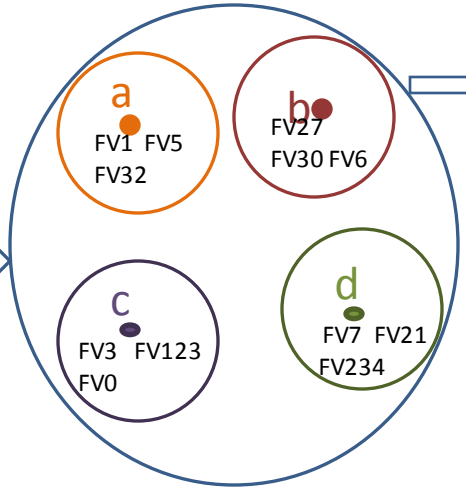
Images as Bag of Words



Set of Feature Vectors

Clustering

Visual Code Book
a
b
c
d



4 - Clusters

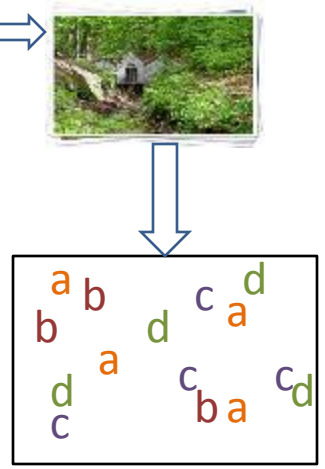


Image-> bag of words

Image ID	a	b	c	d
1	3	0	2	1
2	0	3	0	5
3	4	1	2	1
4	2	0	0	1

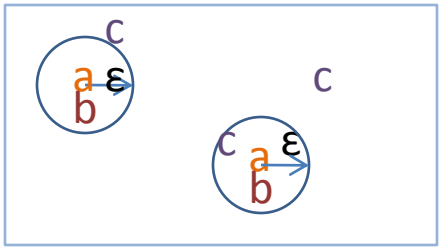
DB: Image as histogram.

Apriori or **FP-Growth** based Algorithm for Frequent Pattern Mining in Histogram.

[Sivic04][Gool06] [Yang07]

Model for Spatial Pattern Mining

- **Neighborhood profiling** for each word satisfying a spatial relationship criteria to discover collocation pattern.



TID	b	c	d
1	2	3	1
2	1	0	2
3	2	3	0

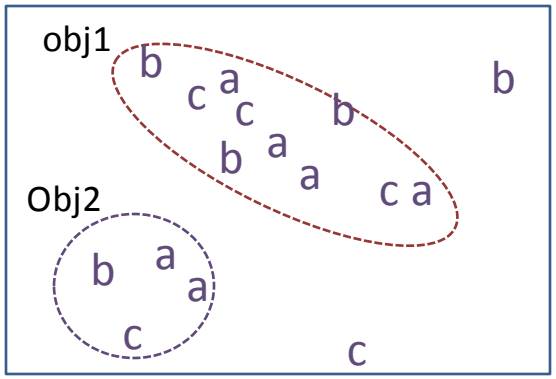
Transactions for neighborhoods of a.

[Yang07, Gool06]

- Spatial constraints *[Koperski95, Shou04]*
  - ε neighborhood of an object. (Star pattern)
  - ε distance between all members of a pattern. (Clique pattern)
  - Adjacency relationship.
- **Localized Profiling**.

a b c b	c d e		d e
d e	c d e		c f a d
f a b		d e	a b

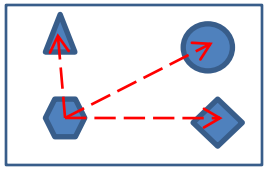
Grid Partitioning

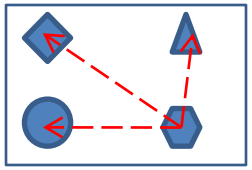


Logical Region Based Partitioning

Model for Co-Orientation Pattern Mining

- Previous models lose relative orientation of objects.
- Orientation (linear and angular distance) is important for meaningful patterns.

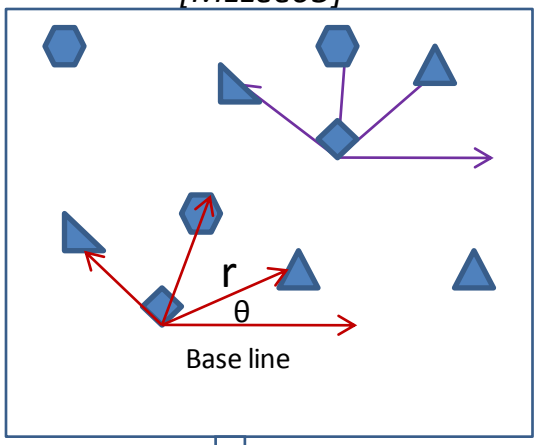




Both Images have same objects
but different configurations.

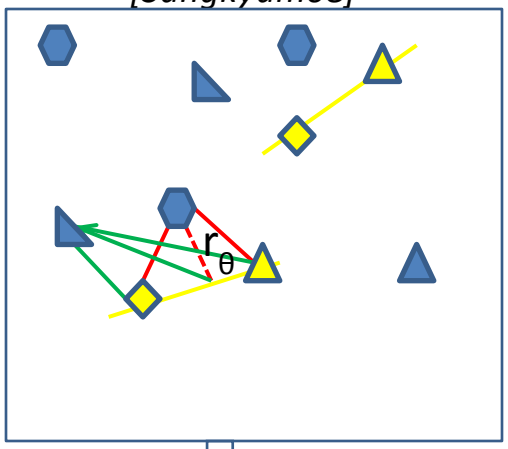
- Defining Structure

[MLLee03]



Entity = (O_1, O_2, r, θ)

[Sangkyum08]



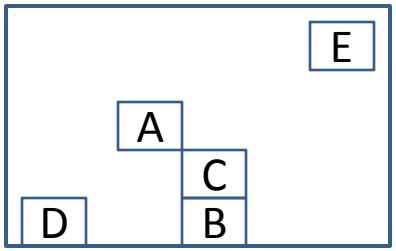
Entity = $(O_1, O_2, O_3, r, \theta)$

Reference Frame

1. Find frequent 2-size patterns.
2. Each 2-size pattern => Base.
3. Describe triplet with the base.
4. Use triplets to define complex structures.

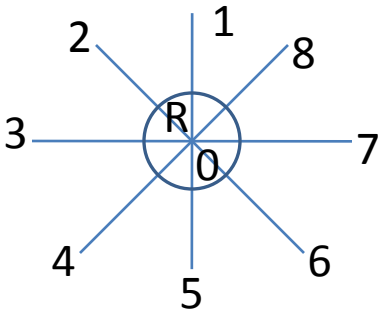
Model for Co-Orientation Pattern Mining

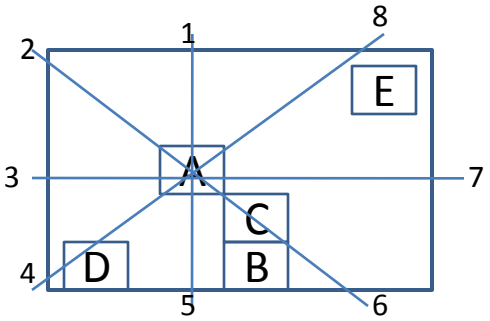
- 2D-String Representation. *[Shan06]*



(S_x,S_y)=(D<A<B=C<E, D=B<C<A<E)

- 9DLT Representation. *[JTLee07]*





(A,B,C,D,E,6,6,4,8,1,3,8,4,8,8)

- Binary Relationship.
  - H-Next-To, V-Next-To, Overlap. *[Zaine00]*
  - Adjacency. *[Jelena03]*
  - disjoined, bordering, surrounded_by, near, far, right, left. *[Selim03]*
  - top-left, top-right, bottom-left, bottom-right. *[Gool06]*

Patterns in Images.

?

- Sources of Images.
- Kinds of Patterns.
- Application of patterns.

Modeling Images for Pattern Mining.

- Frequent Pattern.
- Spatial Pattern.
- Structural Pattern.

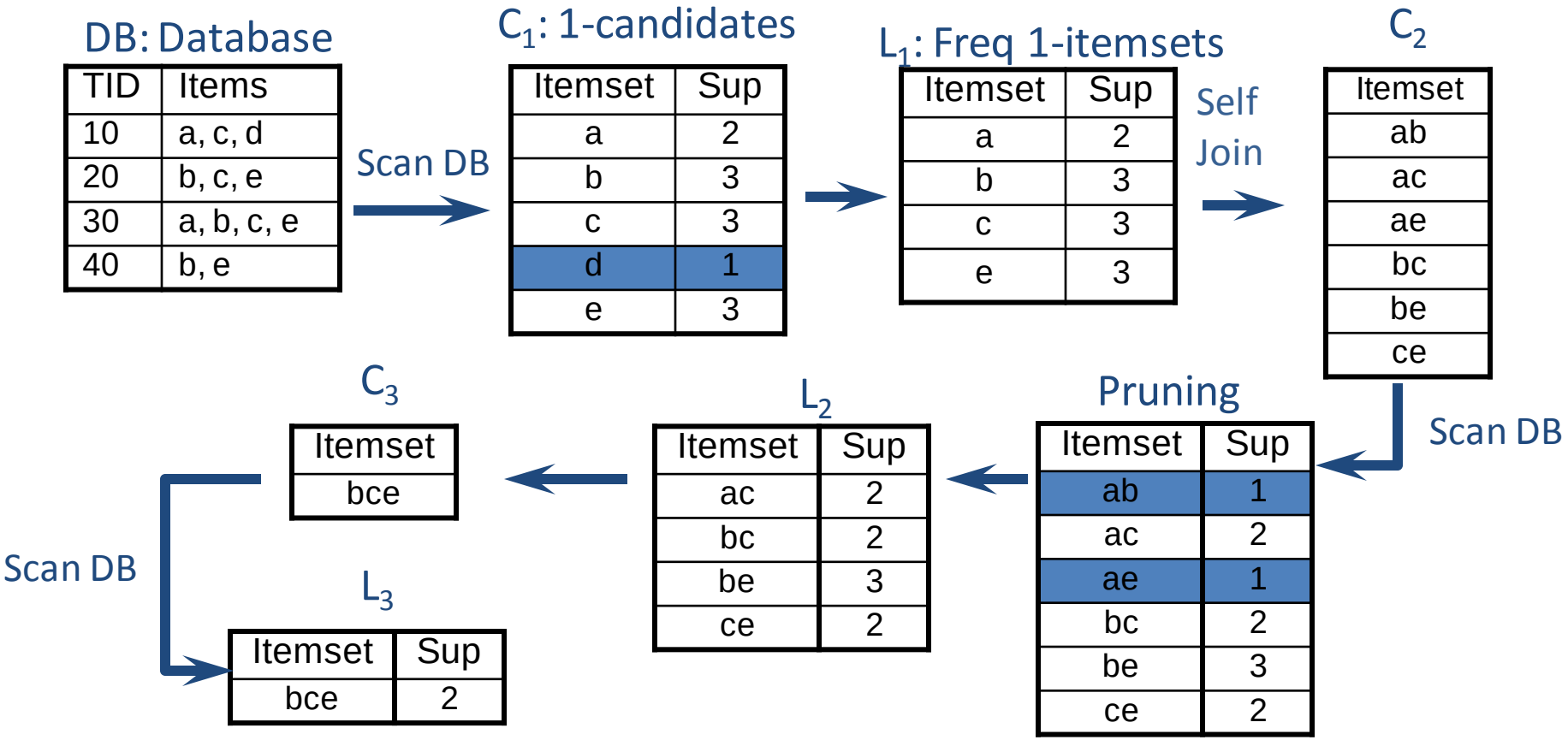
Mining Algorithms.

- Apriori.
- FP-Growth.

Apriori Algorithm

- Frequency based progressive itemsize-wise candidate-generation-and-test approach. *[Sri94]*

  Usefulness Criteria is Minimum Support = 2

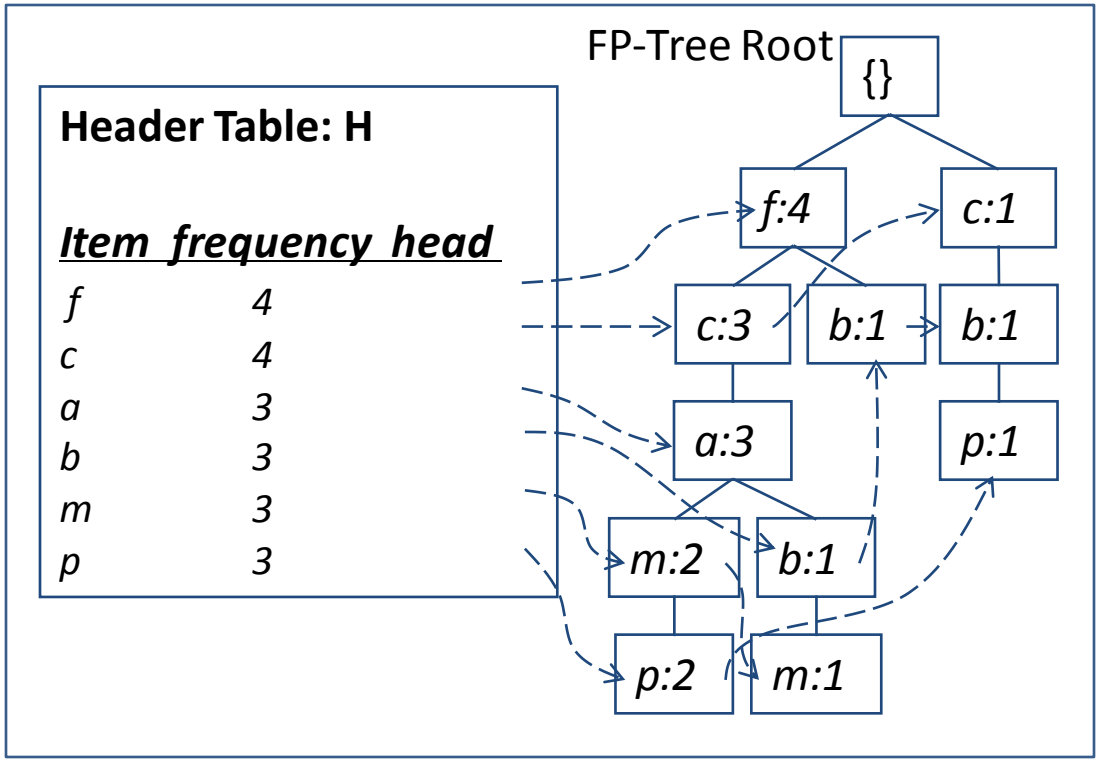


FP-Growth

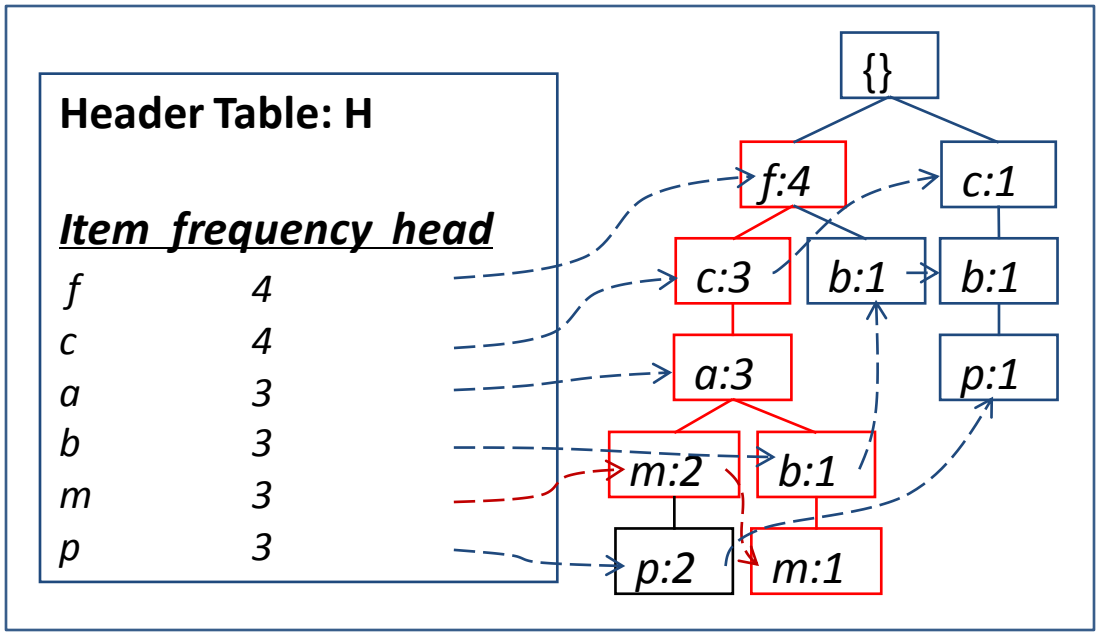
- Recursive approach to generate frequent pattern using FP-Tree. *[Han00]*

TID	Items bought	(ordered) frequent items
100	{f, a, c, d, g, i, m, p}	{f, c, a, m, p}
200	{a, b, c, f, l, m, o}	{f, c, a, b, m}
300	{b, f, h, j, o}	{f, b}
400	{b, c, k, s, p}	{c, b, p}
500	{a, f, c, e, l, p, m, n}	{f, c, a, m, p}

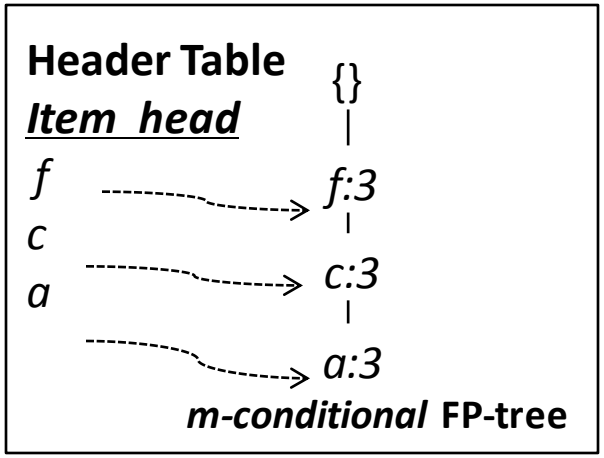
DB: A database of transactions



Frequent pattern for m



conditional **pattern base of "m":**
{fca:2, fcab:1}



Cond. pattern base of "am": (fc:3)

{}
f:3
c:3
am-conditional **FP-tree**

Cond. pattern base of "cm": (f:3)

{}
f:3
cm-conditional **FP-tree**

Cond. pattern base of "cam": (f:3)

{}
f:3
cam-conditional **FP-tree**

Patterns in Images.

?

- Sources of Images.
- Kinds of Patterns.
- Application of patterns.

Modeling Images for Pattern Mining.

- Frequent Pattern.
- Spatial Pattern.
- Structural Pattern.

Mining Algorithms.

- Apriori.
- FP-Growth.

Meaningful Patterns

- Measure of Meaningfulness.
- Direct Mining of Meaningful Patterns.

Measure of Meaningfulness *[Liqiang06]*

N : size of database, P: frequent pattern, $(p_1, p_2) \subset P$, $p_1 \cap p_2 = \Phi$

n : number of transactions having P, $p_1 \rightarrow p_2$: A discovered rule

m : number of transactions having p_1 , o : number of transactions having p_2

- Association Rules
  - Frequency (Coverage) m
  - Support (Generalizability) n/N
  - Confidence (Reliability) n/m

- Collocation Pattern *[Shou03]*
  - Participation ratio $pr(p_1, P) =$ (#instances of p_1 in any instance of P)/m
  - Prevalence $prev(P) = \min(pr(p_i, P), p_i \in P)$
  - Confidence $conf(P) = \max(pr(p_i, P), p_i \in P)$

- Significance *[Yang07]*
  - Likelihood ratio n/(m * o)
  - P-value

Measure of Meaningfulness

- Classification Rules *[Cheng07]*
  - Information Gain
  - Gini Index

- Indexing
  - Database Compression.
  - Discriminative ratio. *[Xifeng04]*

    $dr(P \mid p_1, p_2) = [(p_1\text{'s transaction set}) \cap (p_2\text{'s transaction set})]/(P\text{'s transaction set})$

Example for Mining Meaningful Patterns

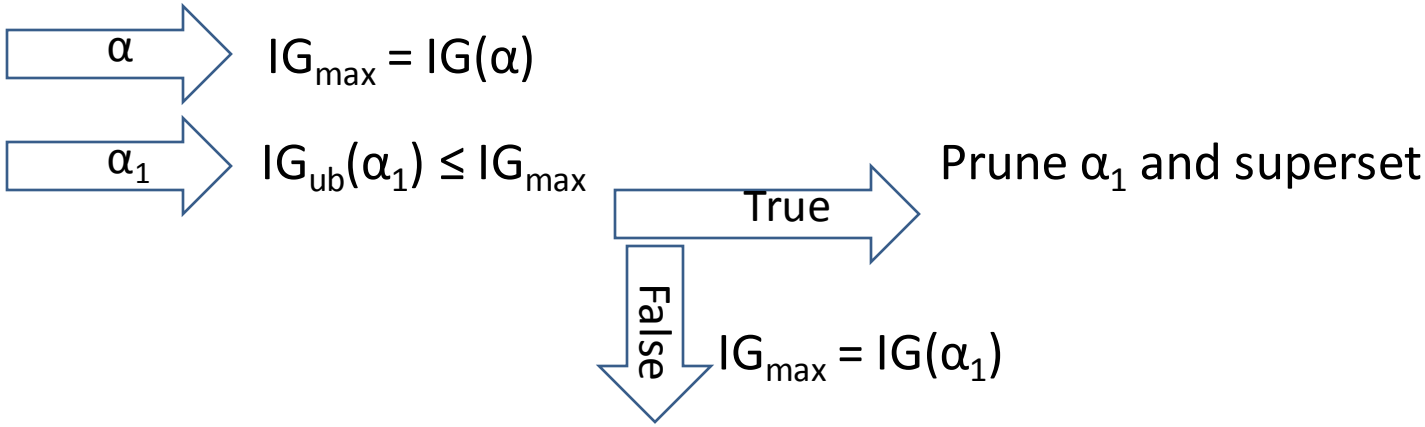
- Goal: Discriminative Pattern for Classification
- Measure: Information Gain (IG)
- Set: C, Pattern: α, and Frequency(α): f

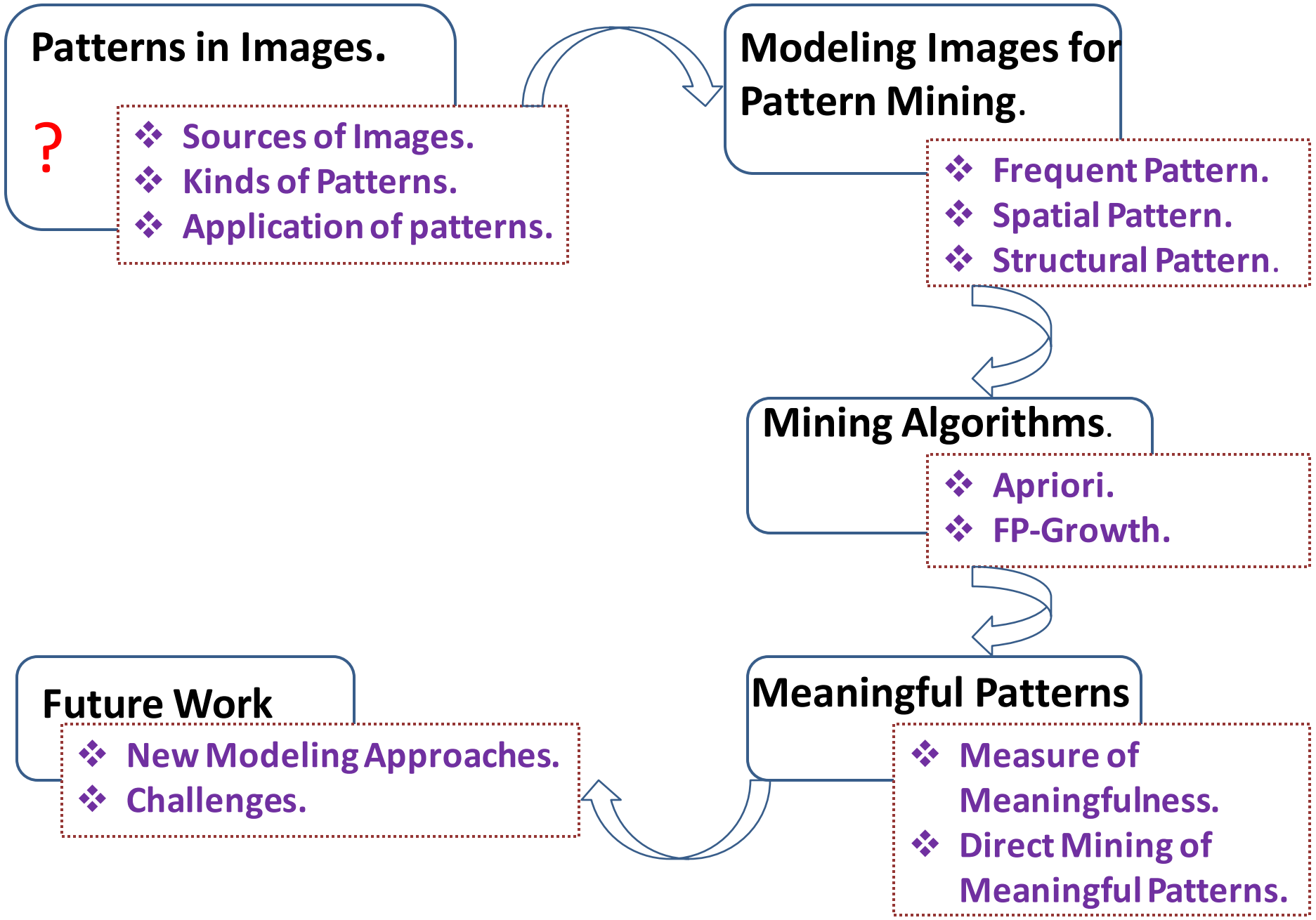
Upper_Bound(IG(C|α)) = Function(f) *[Cheng06]*

- Mine Patterns with *min_sup* = f*

$f^* = \arg\max_f(IG_{ub}(f) \le IG_0)$ where IG_0 is a user parameter

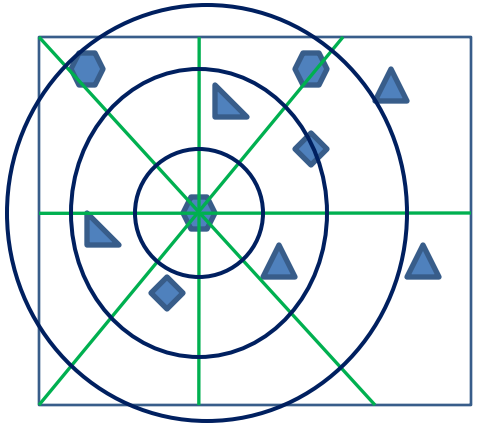
- Prune Search space *[Cheng07]*



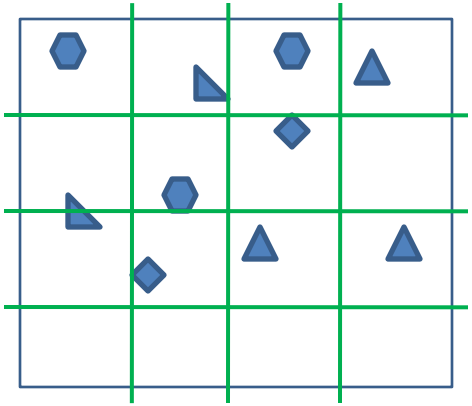

Patterns in Images.
?
Sources of Images.
Kinds of Patterns.
Application of patterns.
Modeling Images for Pattern Mining.
Frequent Pattern.
Spatial Pattern.
Structural Pattern.
Mining Algorithms.
Apriori.
FP-Growth.
Meaningful Patterns
Measure of Meaningfulness.
Direct Mining of Meaningful Patterns.
Future Work
New Modeling Approaches.
Challenges.

New Approaches of Modeling

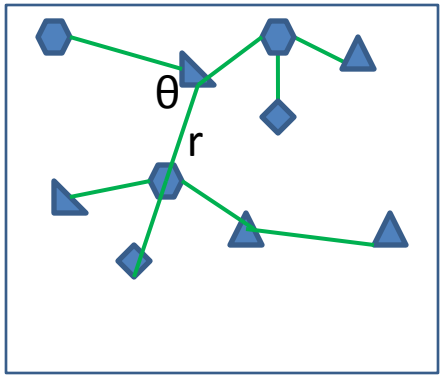
1. Polar Histogram Hashing.
2. Grid Hashing.
3. Graph (NN, MST)
   - Angle between nodes.
   - Edges labeled with relative distance.



1. Polar Histogram



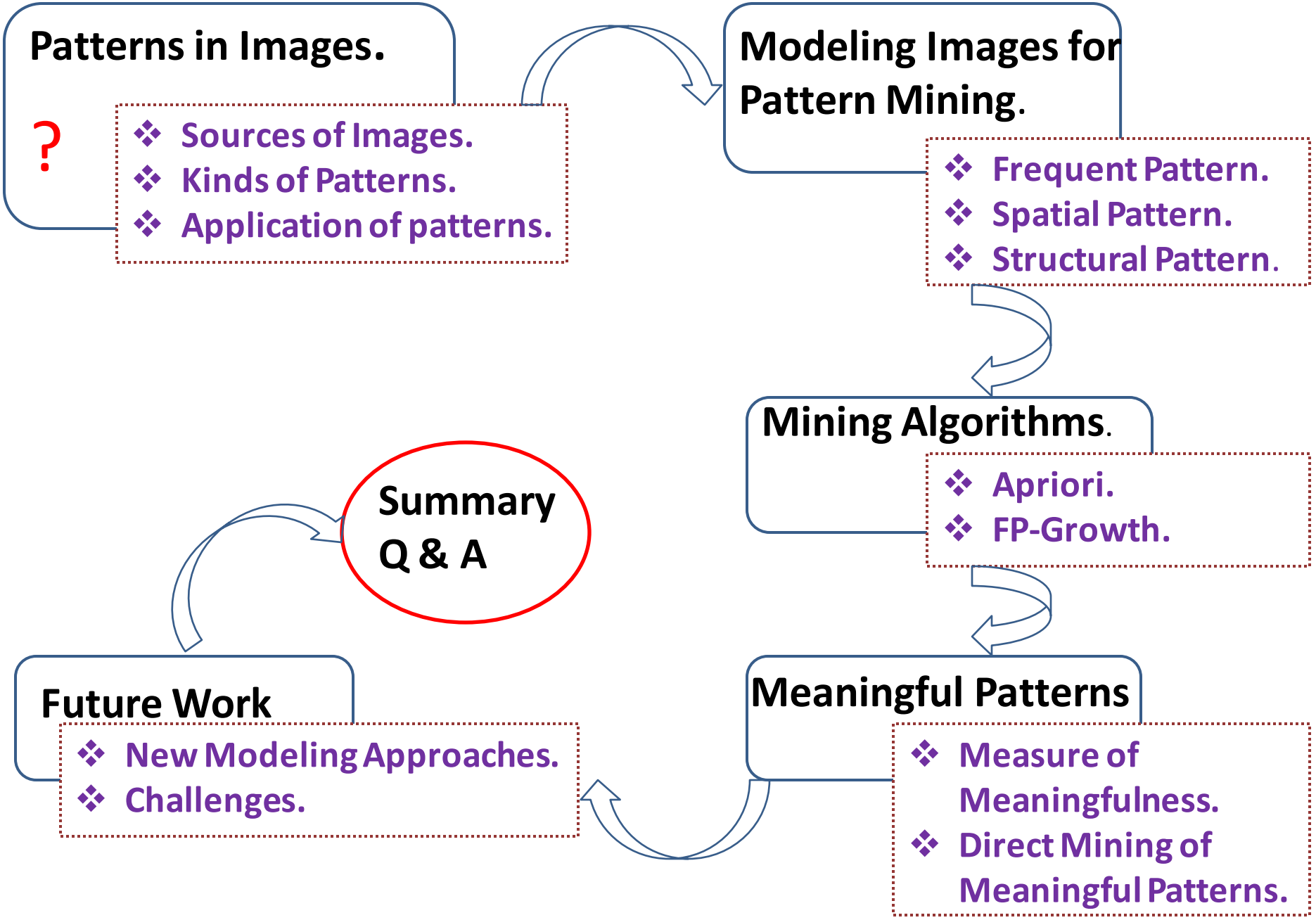
2. Grid (Multi-Resolution) [Zaine00]



3. MST [*Petrakis97]*

Challenges

- Recurrent objects. *[Zaine00, Pan05]*
- Spatial extent of objects. *[Silva05, Selim03, Ma99]*
- Scalability.
- Transaction-less mining.
- Pattern modeling: visualization and storage.
- Application: clustering and indexing.


Patterns in Images.
?
Sources of Images.
Kinds of Patterns.
Application of patterns.
Modeling Images for Pattern Mining.
Frequent Pattern.
Spatial Pattern.
Structural Pattern.
Mining Algorithms.
Apriori.
FP-Growth.
Meaningful Patterns
Measure of Meaningfulness.
Direct Mining of Meaningful Patterns.
Future Work
New Modeling Approaches.
Challenges.
Summary
Q & A

Summary

- Patterns in images are important.
- Modeling of images for patterns is challenging.
- Use of conventional approach for mining.
- Meaningful measure for cost effective and efficient mining.
- Need of new approaches for more précised mining.

Work Till Date

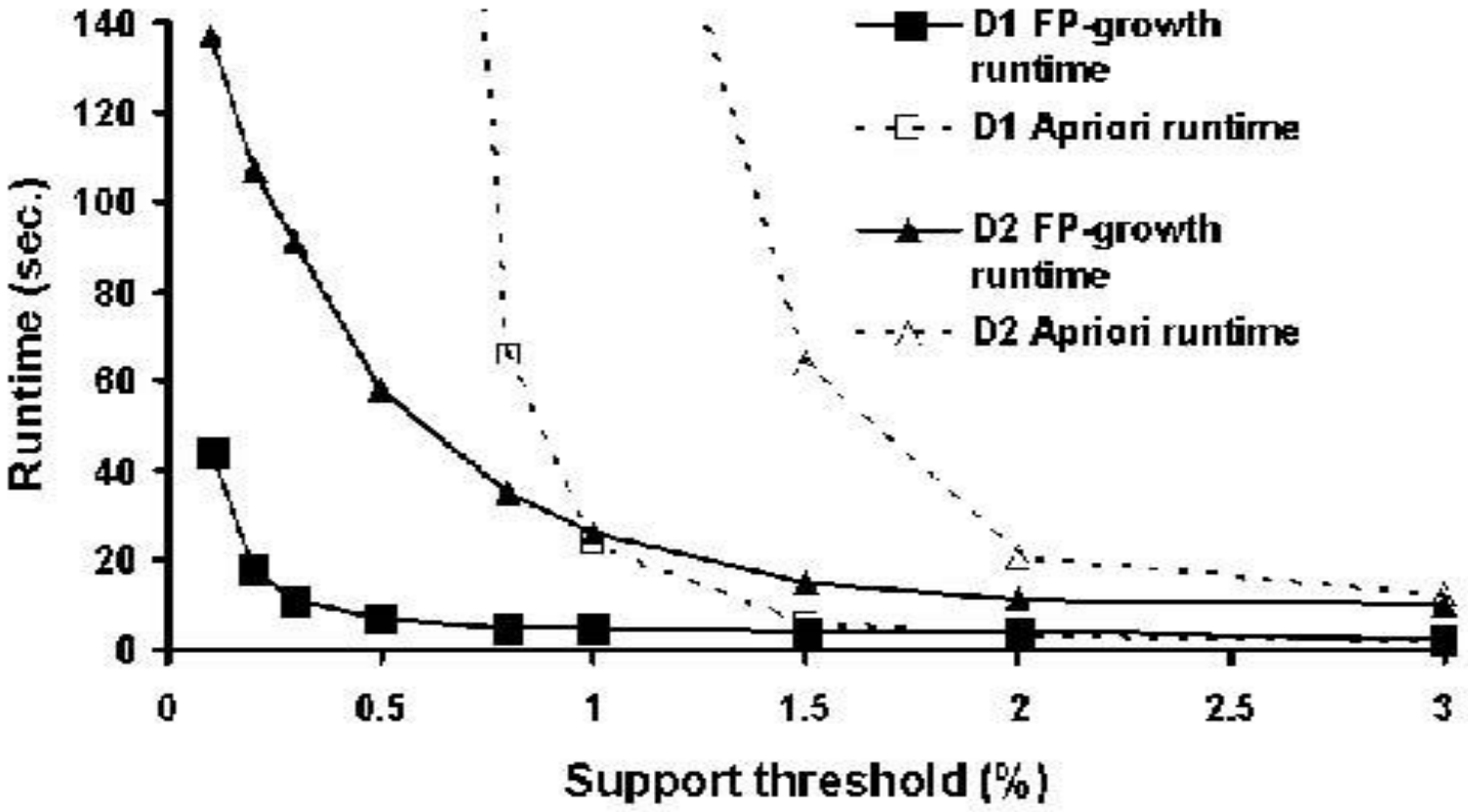
- Querying significant sub-regions in large image databases.
  - Split image into tiles.
  - Global feature descriptor for each tile.
  - Log odd score for matching tiles.
  - Tiled query image overlapped with DB image for all orientations.
  - For each overlap, max scoring sub-region by dynamic programming.
  - Threshold algorithm for scaling.

- Efficient computation of statistical significance of query results in databases. [singh08]
  - Histogram binned into constant number of buckets.
  - Cascading Convolution.

Q & A

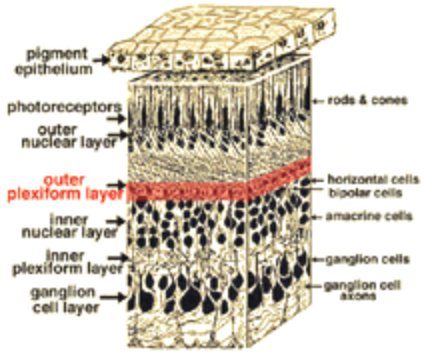
FP-growth vs. Apriori: Scalability With the Support Threshold

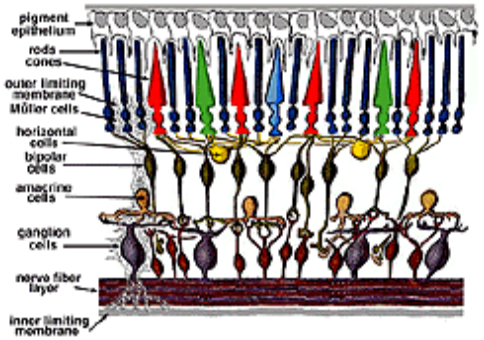
Data set (D1) T25I20D10K / Data set (D2) T25I20D100K



[Han00]

Active Researchers	Conferences
❖ Jiawei Han(UIUC) ❖ Mong Li Lee (NUS) ❖ Selim Aksoy (Bilkent) ❖ Ying Wu (Northwestern) ❖ Luc Van Gool (ETH) ❖ J. T. Lee (NTU, Taiwan)	❖ SIGMOD ❖ VLDB ❖ KDD ❖ ICDE ❖ SDM ❖ ICDM ❖ CIVR ❖ CVPR





1. Isophotal magnitude
2. Isophotal area
3. Semi-major axis
4. Semi-minor axis
5. Eccentricity
6. Sky Variance

1. Thematic Mapper
   - Blue
   - Green
   - Red
   - Reflective – Infrared
   - Mid – Infrared
   - Thermal – Infrared
   - Thermal Infrared
   - Mid – Infrared2
2. SPOT
3. TIFF
4. AVHRR