Social Media Computing

Lecture 3: Location and Image Data Processing

Lecturer: Aleksandr Farseev

E-mail: farseev@u.nus.edu

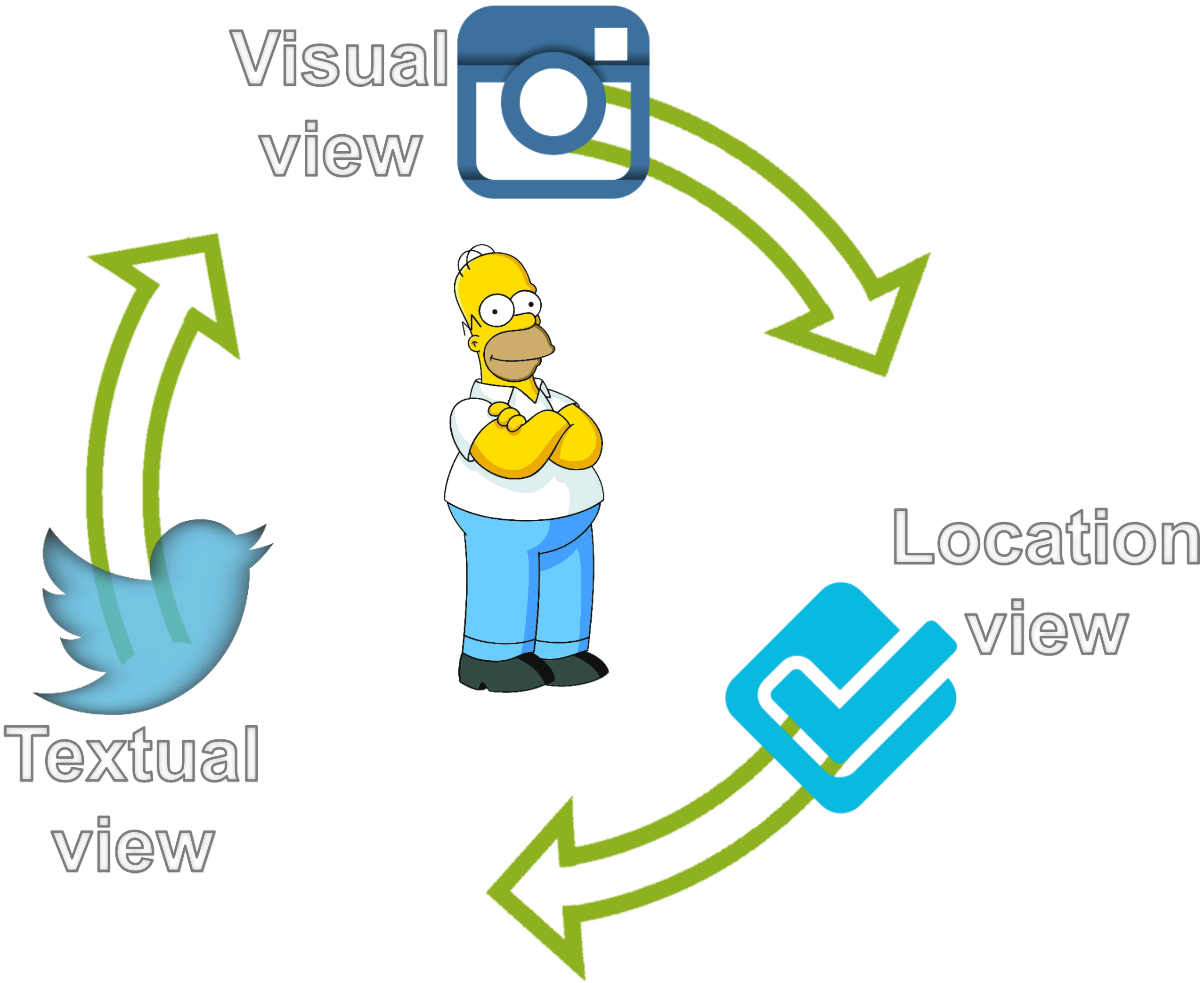
Slides: http://farseev.com/ainlfruct.html

Multiple sources describe user from multiple views

More than 50%
of online-active adults
use more than one social
network in their daily life*

*According Paw Research Internet Project's Social Media Update 2013 (www.pewinternet.org/fact-sheets/social-networking-fact-sheet/)

Multiple sources describe user from 360 °

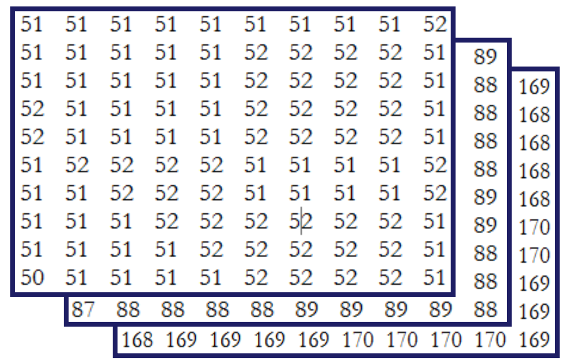


Contents

- Color Image Representations
- Advanced Image Representations
- Location Representation

How to represent images?

what we see



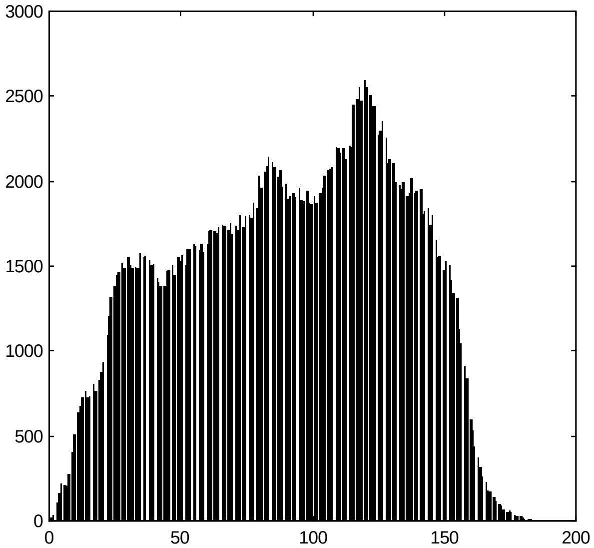
what computers see

Image Feature Extraction

- Simplest is as color histogram!!

Histogram Representation

- What is histogram?
  - The histogram function is defined over all possible intensity levels
  - For 8-bit representation, we have 256 levels or colors
  - For each intensity level, its value is equal to the number of the pixels with that intensity

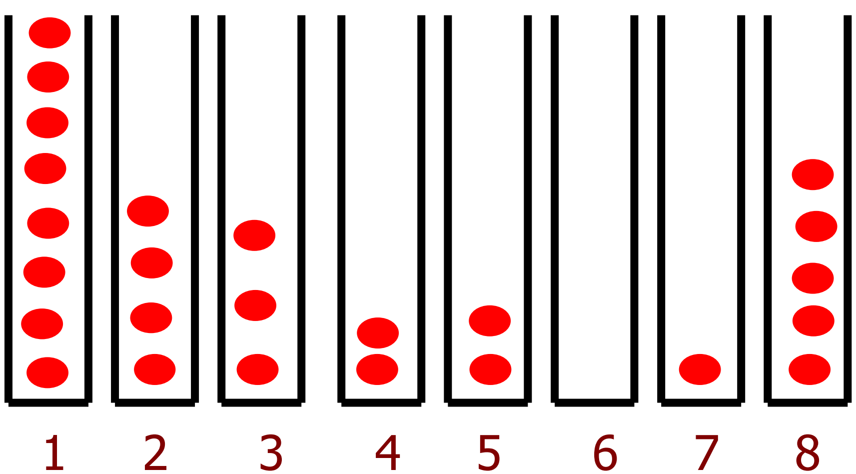


MATLAB function >imhist(x)

What is Histogram

- Example: Consider a 5x5 image with integer intensities in the range between of between 1 & 8, its histogram function $h(r_k)=n_k$ is:

1	8	4	3	4
1	1	1	7	8
8	8	3	3	1
2	2	1	5	2
1	1	8	5	2



Histogram Function:

$$h(r_1)=8$$
$$h(r_2)=4$$
$$h(r_3)=3$$
$$h(r_4)=3$$
$$h(r_5)=2$$
$$h(r_6)=0$$
$$h(r_7)=1$$
$$h(r_8)=5$$

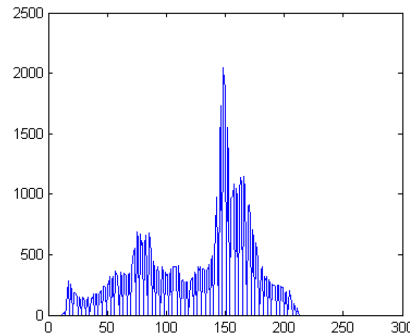
Normalized Histogram:

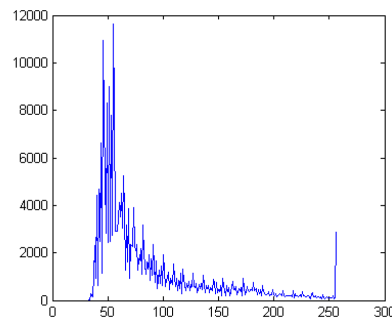
$$p(r_1)=8/25=0.32$$
$$p(r_2)=4/25=0.16$$
$$p(r_3)=3/25=0.12$$
$$p(r_4)=3/25=0.08$$
$$p(r_5)=2/25=0.08$$
$$p(r_6)=0/25=0.00$$
$$p(r_7)=1/25=0.04$$
$$p(r_8)=5/25=0.20$$

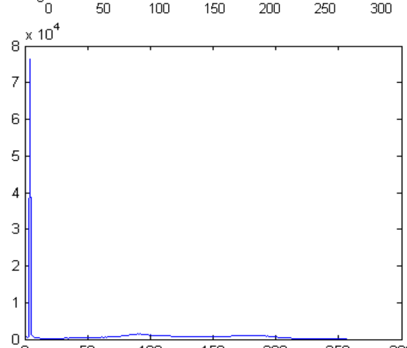
Examples of Image Histogram

Original image

Graph of the histogram function







Observation:

- Image intensity is skewed (not fully utilizing the full range of intensities)
- What can be done??

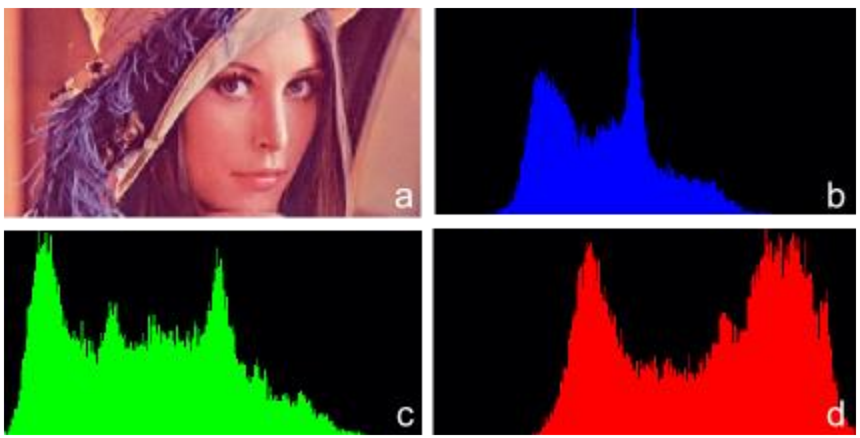
Color Histogram -1

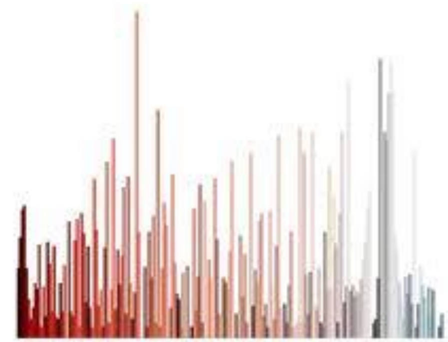
- Let image *I* be of dimension *p x q*
  - For ease in representation, need to quantize *p x q* potential colors into *m* colors (for m << p x q)
  - For pixel p = (x,y)∈ *I*, the color of pixel is denoted by *I*(p) = c_k

- Construction of Color Histogram
  - Extract color value for each pixel in image
  - Quantize color value into one of *m* quantization levels
  - Collect frequency of color values in each quantization level, where each bin corresponds to a color in the quantized color space

Color Histogram -2

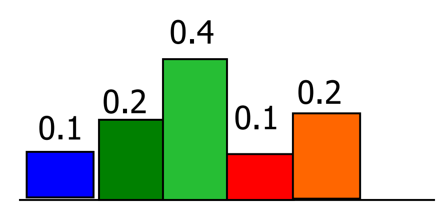
- Thus, image is represented as a color histogram H of size *m*
  - where H[i] gives # of pixels at intensity level I
- For example:





Into a single quantized histogram

- Normalize H to NH by dividing each entry by size of image p*q



Color Moment

- Let the set of pixel be:
  $I = [p_1, p_2, \ldots p_R]$, for a total of R=(p x q) pixels

- Represent color contents of image in terms of moments:

  1st Color moment (Mean): $\frac{1}{R}\sum_i X_i$

  2nd Color Moment about mean (Variance): $\frac{1}{R}\sum_i (X_i - \overline{X})^2$

- We can use these to model image contents
  - Advantages: Simple & efficient; Only one value for each representation
  - Disadvantage: Unable to model contents well
  - However, it can be effective at sub-image level, say sub-blocks

  HOW TO DO THIS??

Color Coherence Vector (CCV) -1

- Problems of color histogram rep
  - Easy to find 2 different images with identical color histogram
  - As it does not model local and location info

Exactly same color distribution & similar shape

- Need to take spatial info into consideration when utilizing colors:
  - Color Coherence Vector (CCV) representation

- CCV
  - A simple and elegant extension to color histogram
  - Not just count colors, but also check adjacency
  - Essentially form 2 color histograms – one where colors form sufficiently large regions, while the other for isolated colors

CCV Representation -2

- Example:
  - Define sufficiently large region as those > 5 pixels

2	1	2	2	1	1
2	2	1	2	1	1
2	1	3	2	1	1
2	2	2	1	3	3
2	2	1	1	3	3
2	2	1	1	3	3

→

2	1	2	2	1	1
2	2	1	2	1	1
2	1	3	2	1	1
2	2	2	1	3	3
2	2	1	1	3	3
2	2	1	1	3	3

→

Region	A	B	C	D	E
Color	2	1	3	1	3
Size	15	3	1	11	6

Color	1	2	3
Hα	11	15	6
Hβ	3	0	1

- Treats H_α and H_β separately

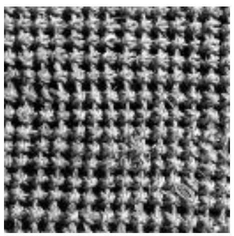
- Similarity measure:
  - Give higher weight to Hα, as it tends to correspond more to objects

$$\mathrm{Sim}(Q, D) = \mu\, \mathrm{Sim}(Q_\alpha, D_\alpha) + (1- \mu)\, \mathrm{Sim}(Q_\beta, D_\beta)$$

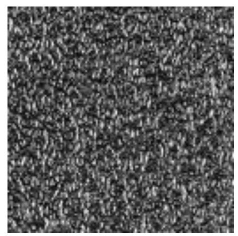
for $\mu > 0.5$

Texture Representation

- What is texture?
  - Something that repeats with variation
  - Must separate what repeats and what stays the same
  - Model as repeated trials of a random process



Fabric



Metal

Leaves

Flowers

- Tamura representation: classifies textures based on psychology studies
  - Coarseness
  - Contrast
  - Directionality
  - Linelikeness
  - Regularity
  - Roughness
- Consider simple realization of Tamura features
  - May be simplified as distributions of edges or directions

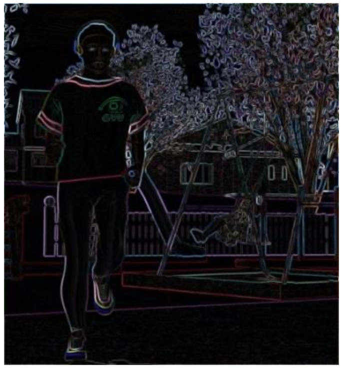
Edge Representation -1

- Spatial Domain Edge-based texture histogram
  - To extract an edge-map for the image, the image is first converted to luminance *Y* (via *Y* = 0.299*R*+0.587*G*+0.114*B*)
  - A *Sobel edge operator* is applied to the *Y* -image by sliding the following 3×3 weighting matrices (*convolution masks*) over the image and applying (*) it on each sub segment A.

$$d_x = \begin{array}{|c|c|c|} \hline -1 & 0 & 1 \\ \hline -2 & 0 & 2 \\ \hline -1 & 0 & 1 \\ \hline \end{array} \ *A \qquad d_y = \begin{array}{|c|c|c|} \hline 1 & 2 & 1 \\ \hline 0 & 0 & 0 \\ \hline -1 & -2 & -1 \\ \hline \end{array} \ *A$$

  - The edge magnitude *D* and the edge gradient ϕ are given by:

$$D = \sqrt{d_x^2 + d_y^2}, \quad \phi = \arctan \frac{d_y}{d_x}$$

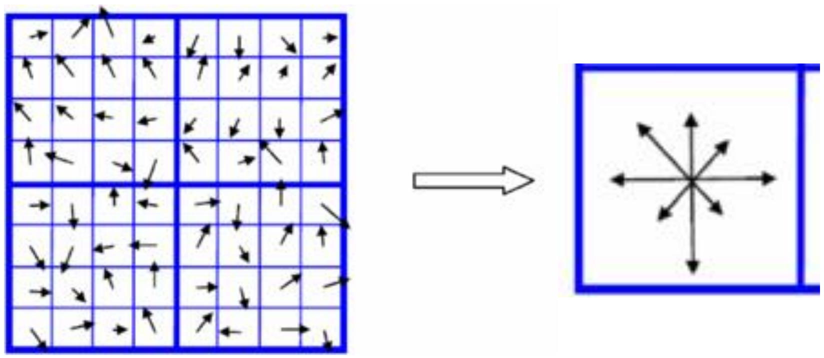


Edge Representation -2

- Represent texture of image as 1 or 2 histograms:

Edge histogram

- Quantize the edge direction Φ into 8 directions:



- Setup $H(\Phi)$ (with 8 dimension)

Magnitude histogram

- Quantize the magnitude D into, say 16 values
- Setup $H(D)$, with 16 dimension.

- Edge Histogram is normally used

Segmented Image Representation

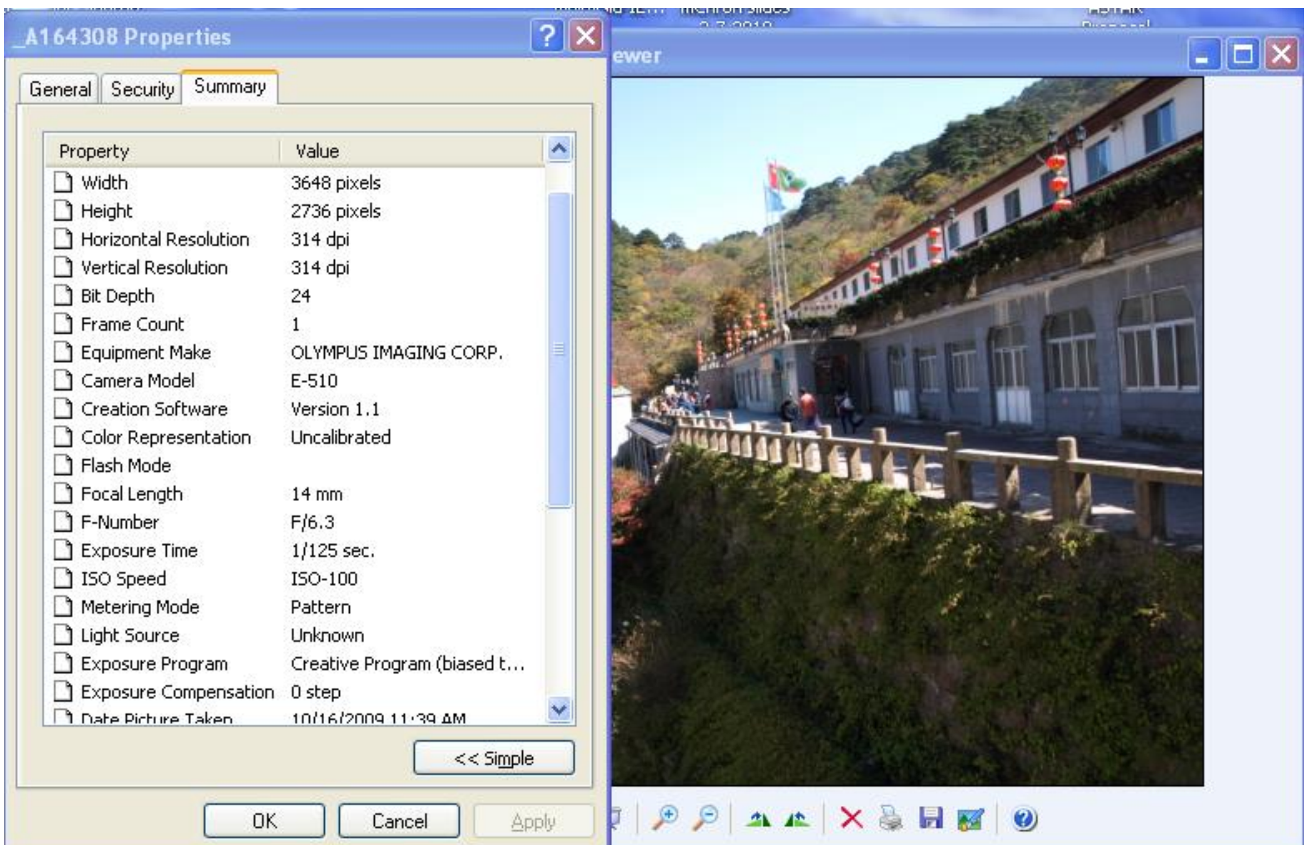
- Problems with global image representation – can't handle layout and object level matching very well
- One simple remedy: use segmented image (example, 4x4):

(1,1)	(1,2)	(1,3)	(1,4)
(2,1)	(2,2)	(2,3)	(2,4)
(3,1)	(3,2)	(3,3)	(3,4)
(4,1)	(4,2)	(4,3)	(4,4)

- Compute histograms for individual window
- Match at sub-window level between Q and D:
  - o between corresponding sub-windows or
  - o between all possible pairs of sub-windows
  - o May give higher weights to central sub-windows
- Pros: able to capture some local information
- Cons: more expensive, may have mis-alignment problem

Metadata of Images

- Cameras store image metadata as "EXIF tags"
  - EXIF (**Exchangeable image file format**)
  - Timestamp, focal length, shutter speed, aperture, etc
  - Keywords can be embedded in images



Metadata of Images -2

- Other form of metadata: semantic tags (or concepts)
  - Supply manually by users
  - Reasonable thru social tagging

- With metadata, we can perform advanced analysis:
  - Use existing set of semantic tags
  - Automatic keyword generation (leveraging on EXIF info)
  - Camera knows *when* a picture was taken…
  - A GPS tracker knows *where* you were…
  - EXIF knows the conditions that picture was taken
  - Your calendar (or phone) knows *what* you were doing…

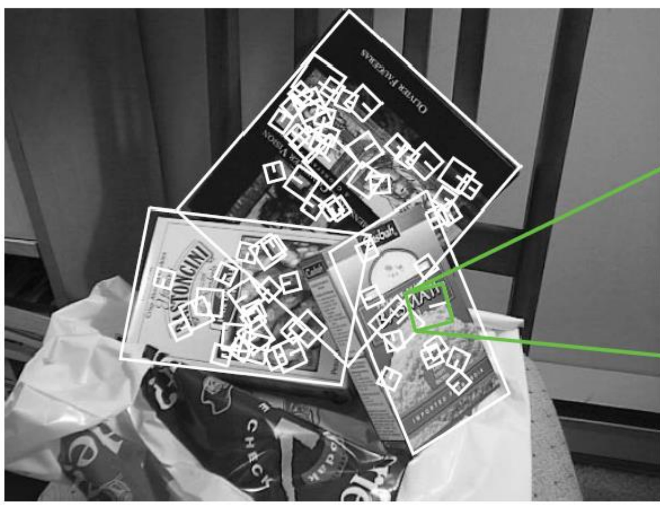
  - Combine these together into a list of keywords

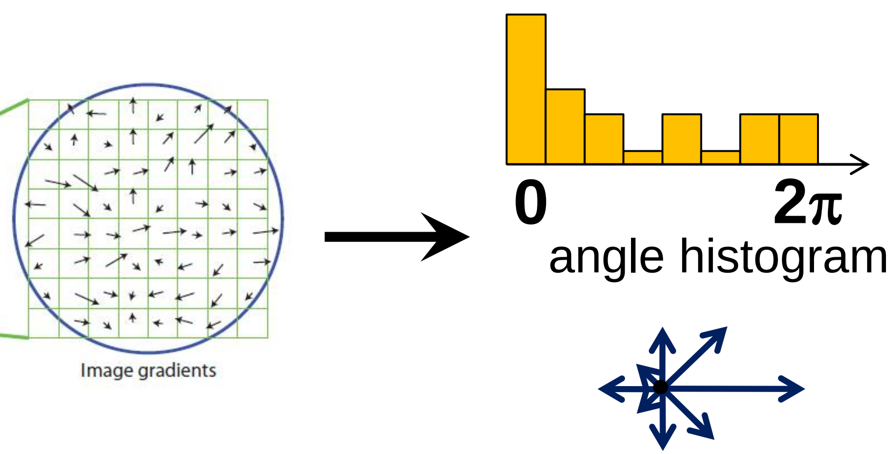
Contents

- Color Image Representations
- Advanced Image Representations
- Location Representation

Scale Invariant Feature Transform (SIFT) descriptor -1

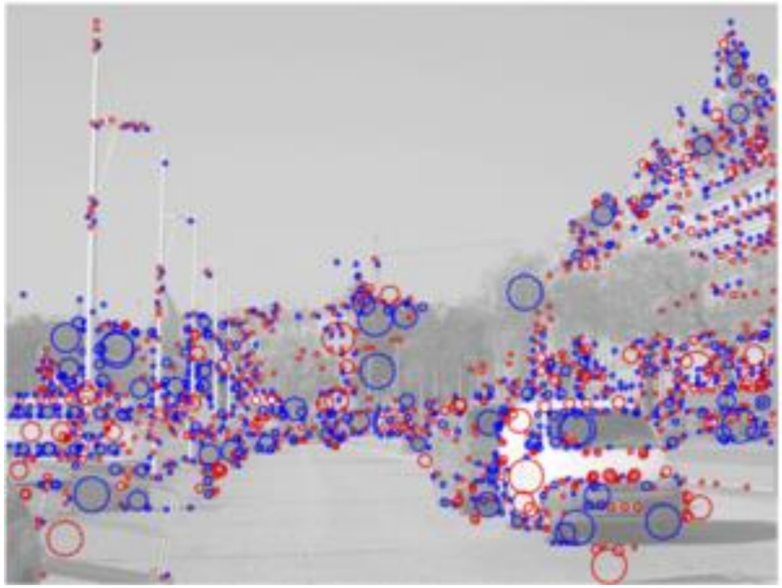
- Basic idea: use edge orientation representation
  - Obtain interest points from scale-space extrema of differences-of-Gaussians (DoG)
  - Take 16x16 square window around detected interest point
  - Compute edge orientation for each pixel
  - Throw out weak edges (threshold gradient magnitude)
  - Create histogram of surviving edge orientations





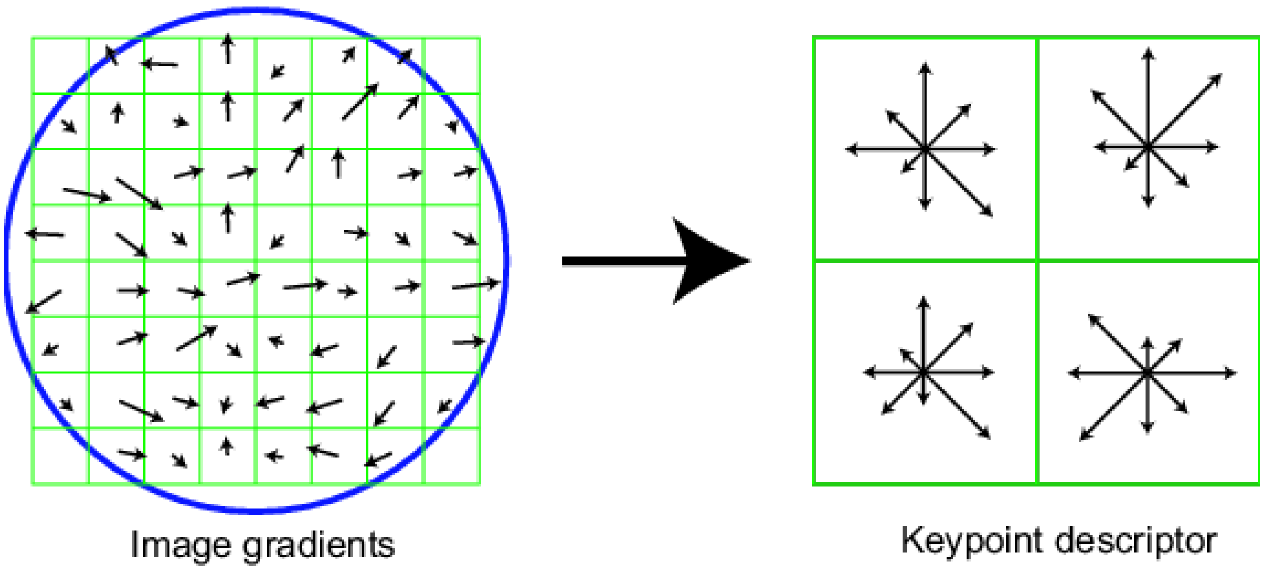
http://www.scholarpedia.org/article/Scale_Invariant_Feature_Transf

Detected Interest Points

Scale Invariant Feature Transform (SIFT) descriptor -2

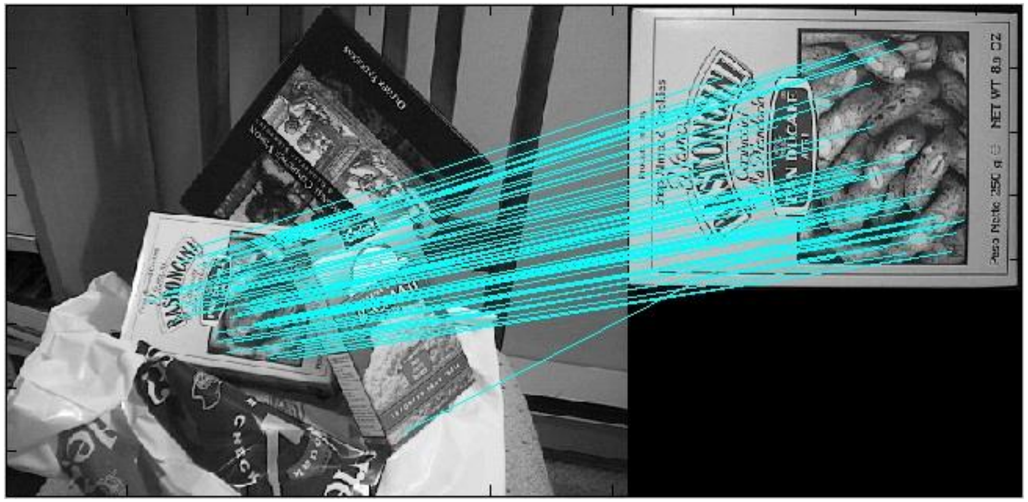
- A popular descriptor:
  - Divide the 16x16 window into a 4x4 grid of cells (we show the 2x2 case below for simplicity)
  - Compute an orientation histogram for each cell
  - 16 cells X 8 orientations = 128 dimensional descriptor



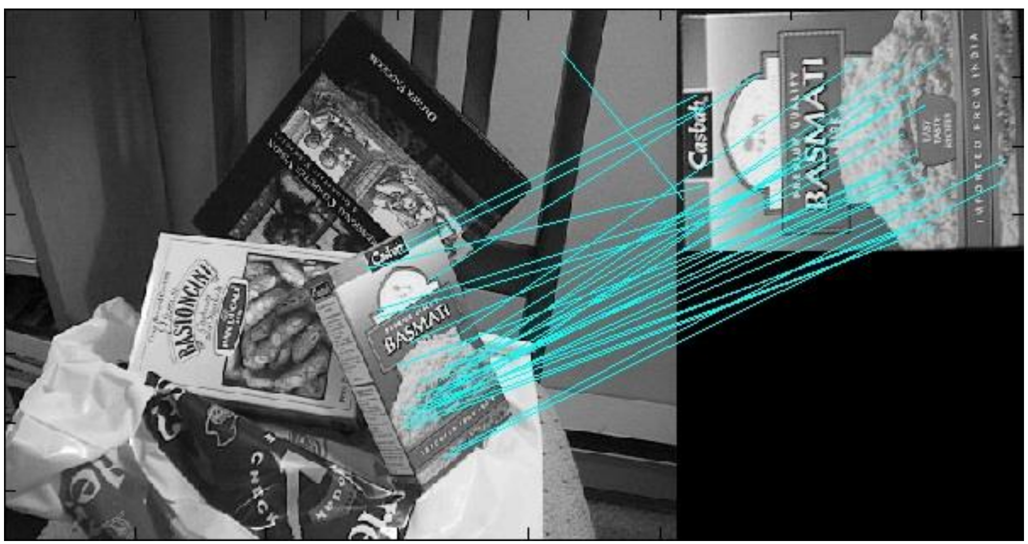
Scale Invariant Feature Transform (SIFT) descriptor -3

- Invariant to
  - Scale
  - Rotation
- Partially invariant to
  - Illumination changes
  - Camera viewpoint
  - Occlusion, clutter

Examples of SIFT matching



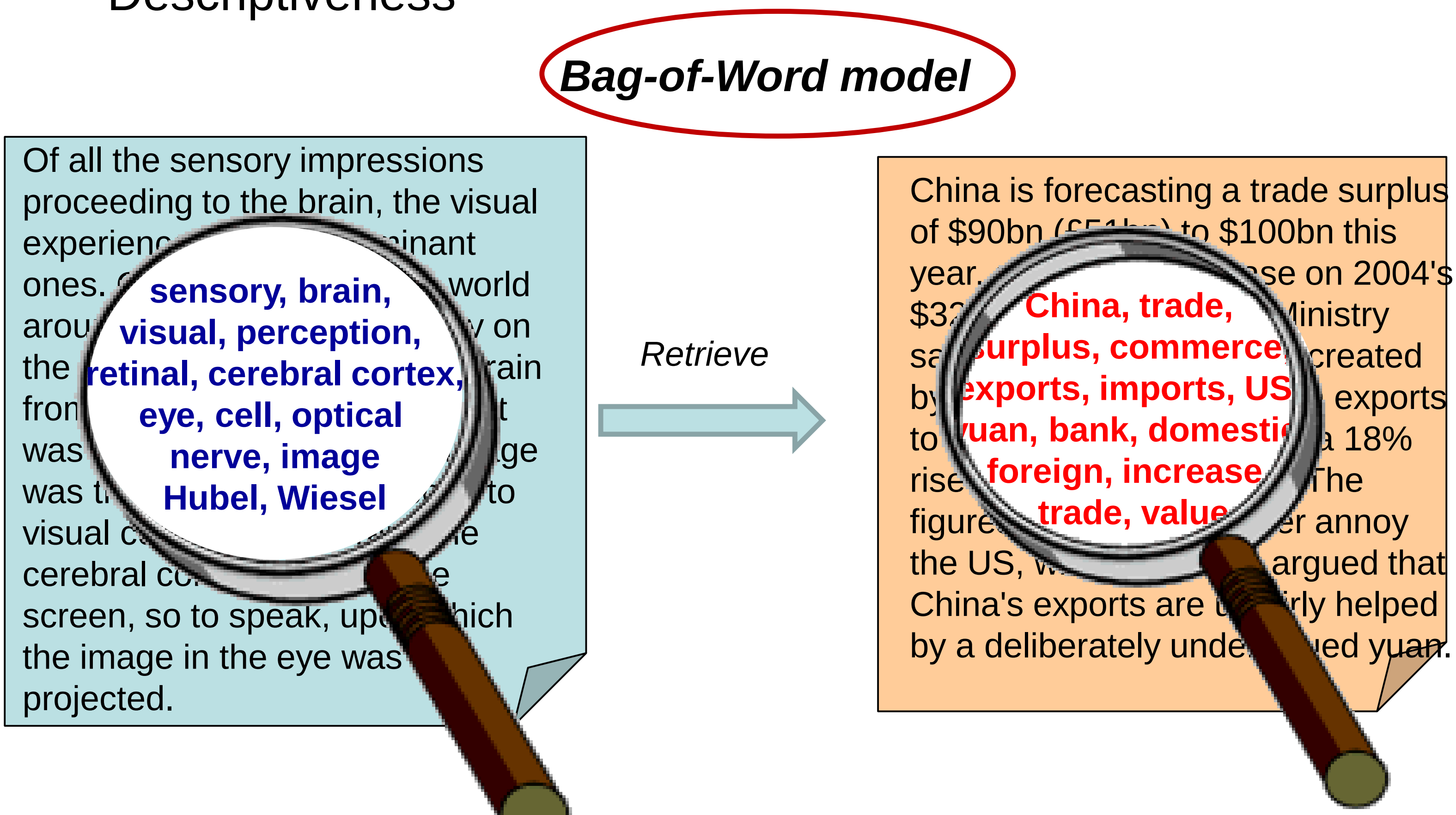
80 matches



34 matches

Overall Representation: as Bag of Visual Words -1

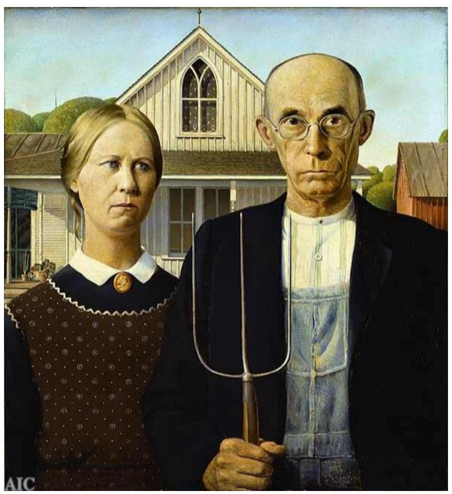
- Text Words in Information Retrieval (IR)
  - Compactness
  - Descriptiveness



Overall Representation: as Bag of Visual Words -2

- Can images be represented as Bag-of-Visual Words?

Image → **Bag of 'visual words'**

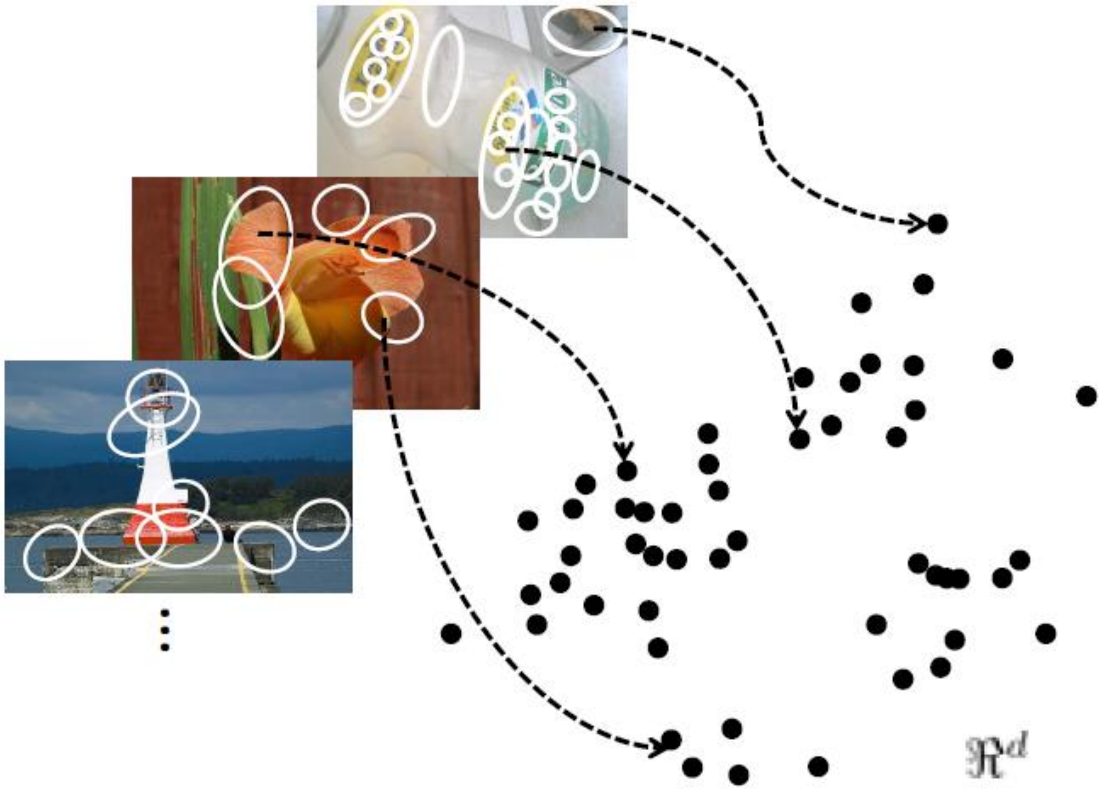


?

- Idea: quantize SIFT descriptors of all training images to extract representative visual words!

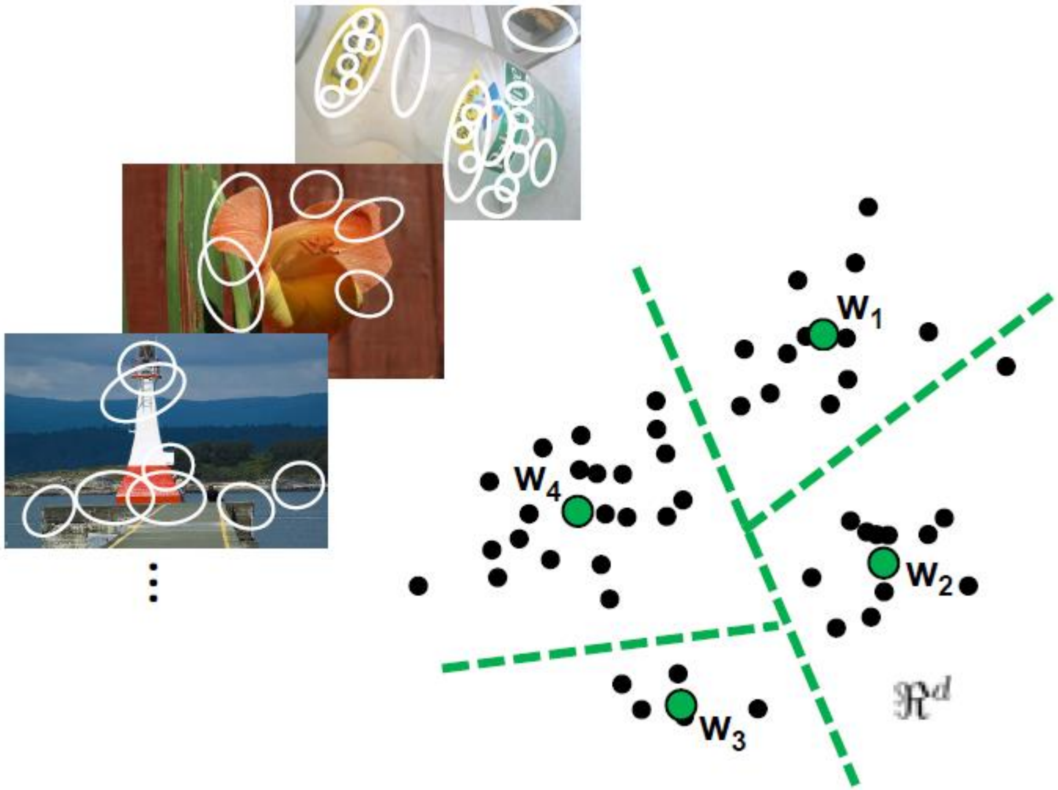
Overall Representation: as Bag of Visual Words -3

Step 1: Extract interest points of all training images



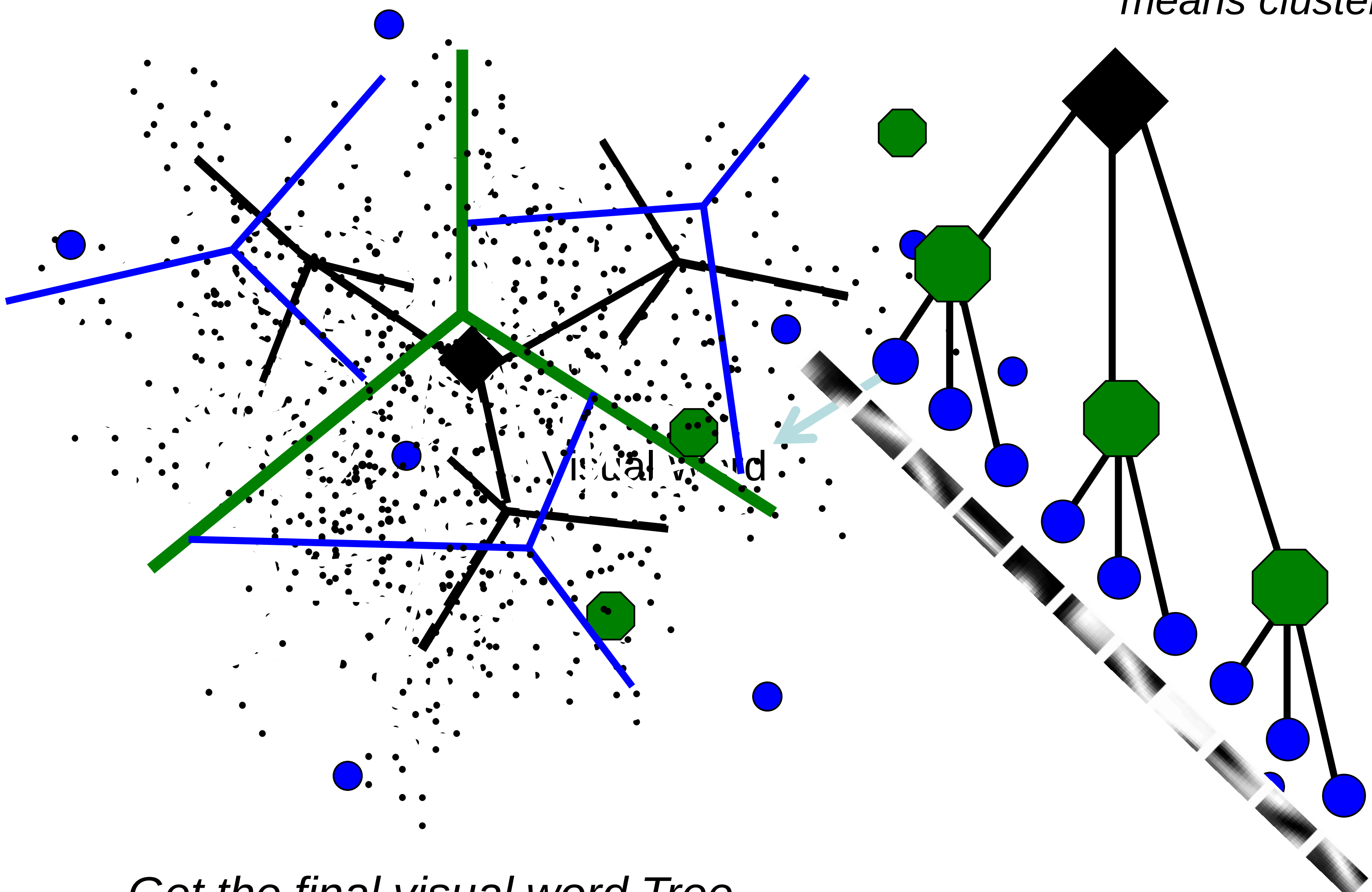
Overall Representation: as Bag of Visual Words -4

Step 2: Features are clustered to quantize the space into a discrete number of visual words.



Overall Representation: as Bag of Visual Words -5

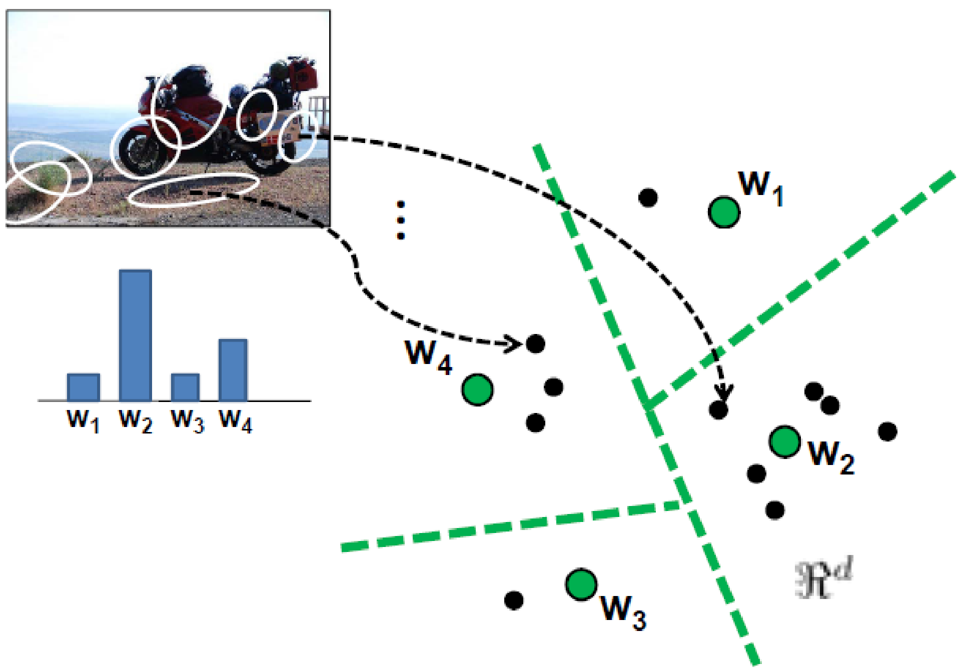
Hierarchical K-means clustering



Get the final visual word Tree

Overall Representation: as Bag of Visual Words -6

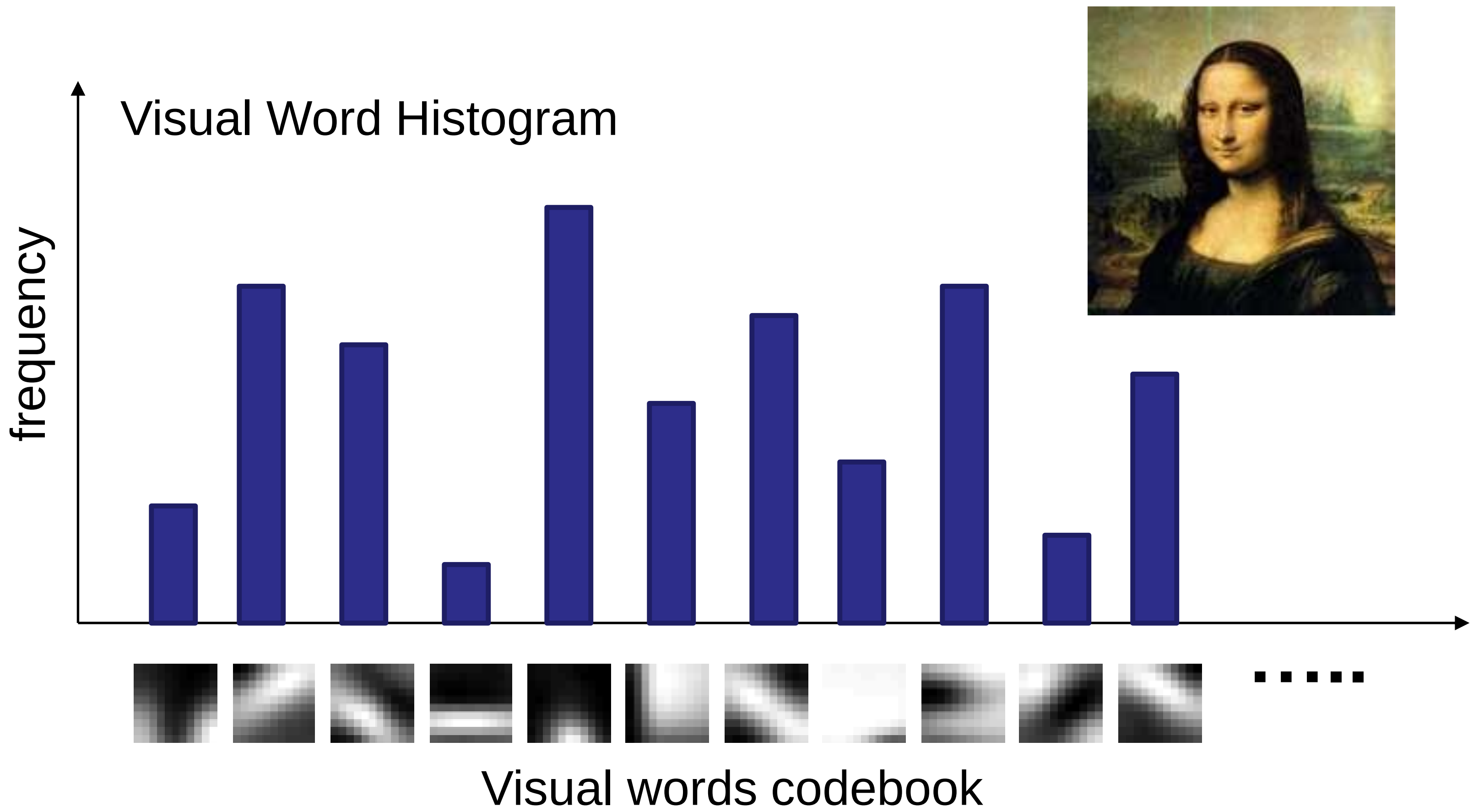
Step 3: Summarize (represent) each image as histogram of visual words



and use as basis for matching and retrieval!

Overall Representation: as Bag of Visual Words -7

- Another example:



What do we mean by “Concept Recognition”

Verification:
Is that a statue of rabbit?

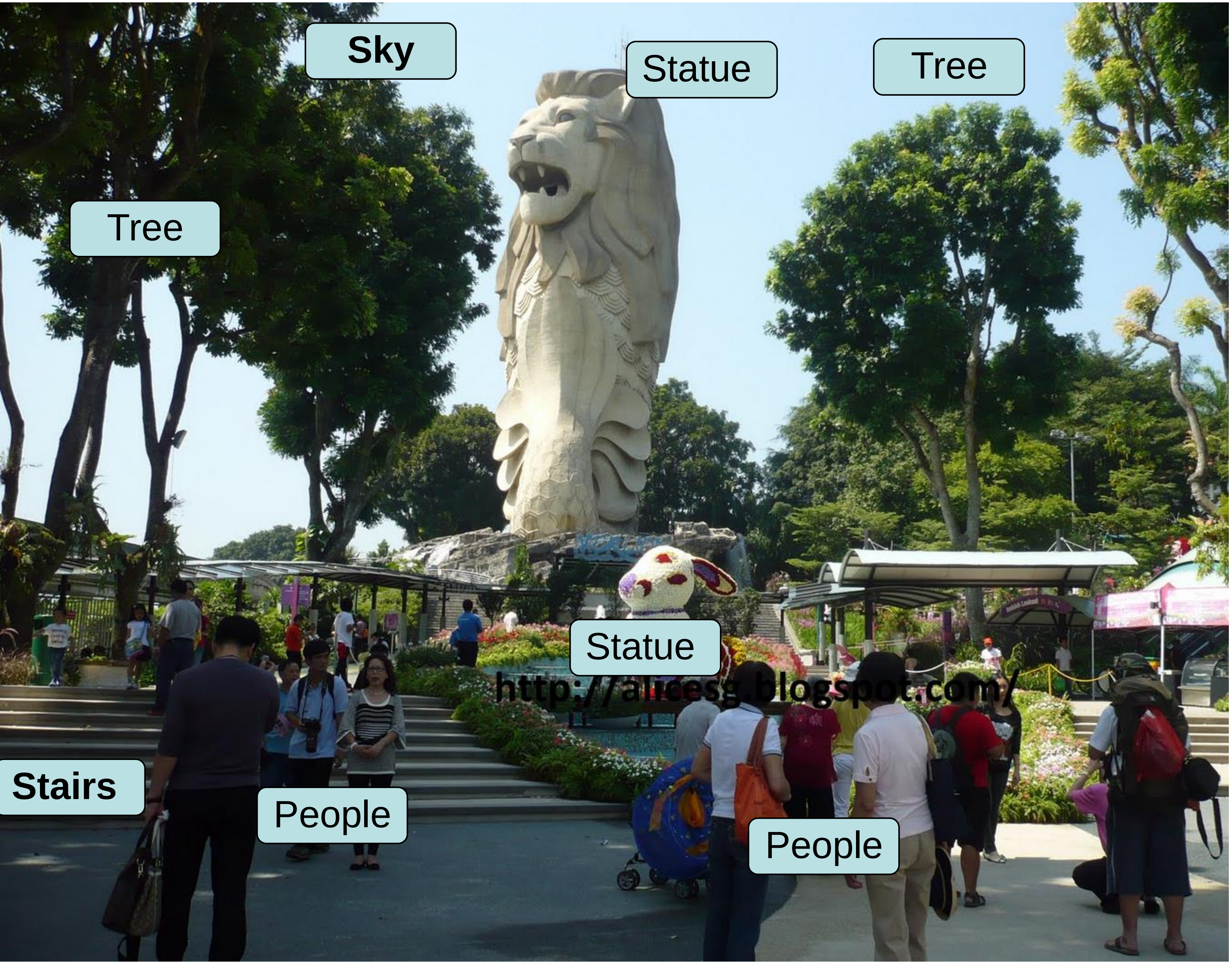
What does Concept Recognition involve?

Detection:
Are there trees?

What does Concept Recognition involve?

Identification:
Is that the merlion, Singaore's landmark?

What does Concept Recognition involve?



Object Categorization

What does Concept Recognition involve?

Scene
and Context
Categorization


~10,000 to 30,000

Concept Recognition: Challenges

- View point variation
- Illumination
- Occlusion
- Scale
- Deformation
- Background clutter



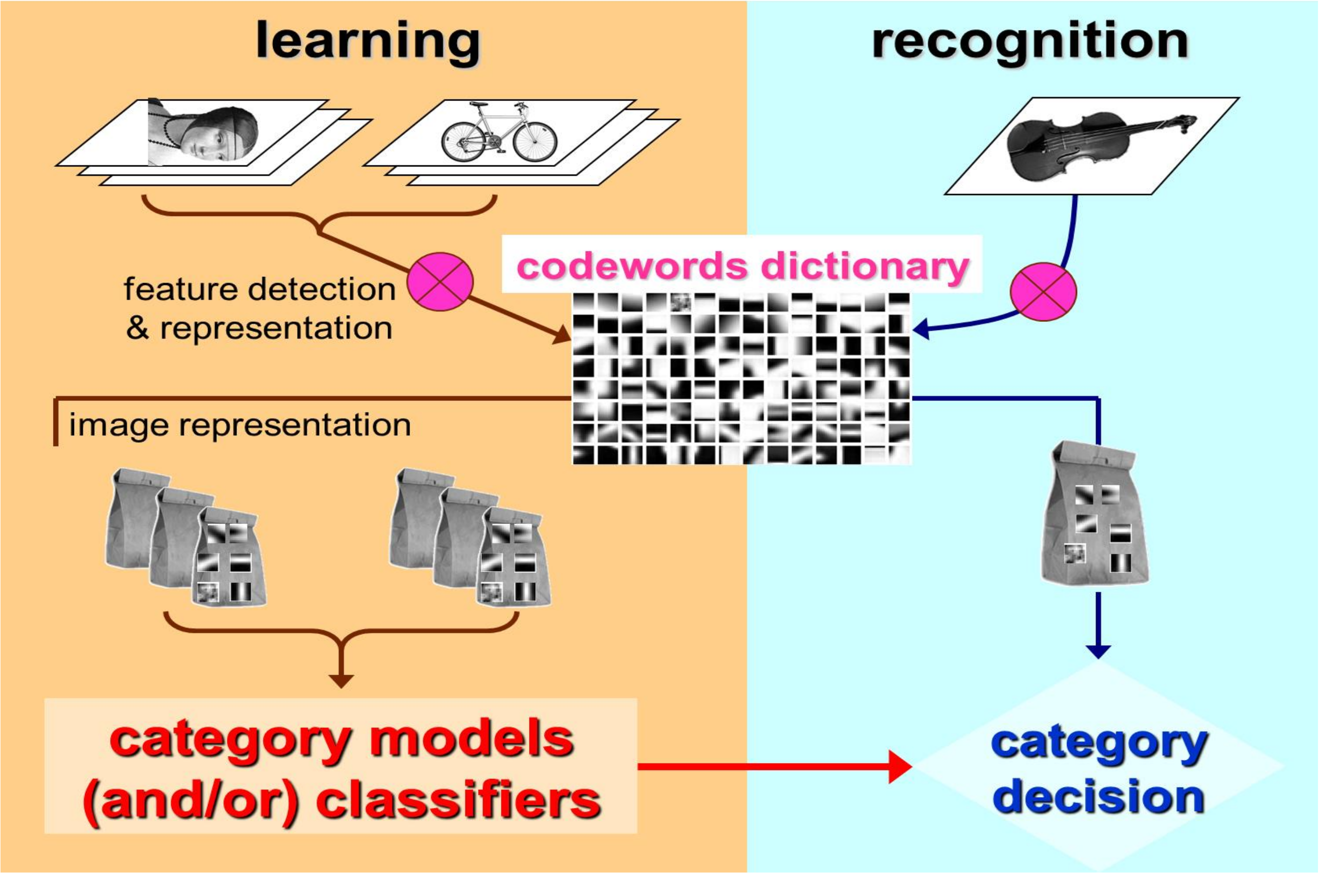




Concept Recognition: Bag-of-Word Model

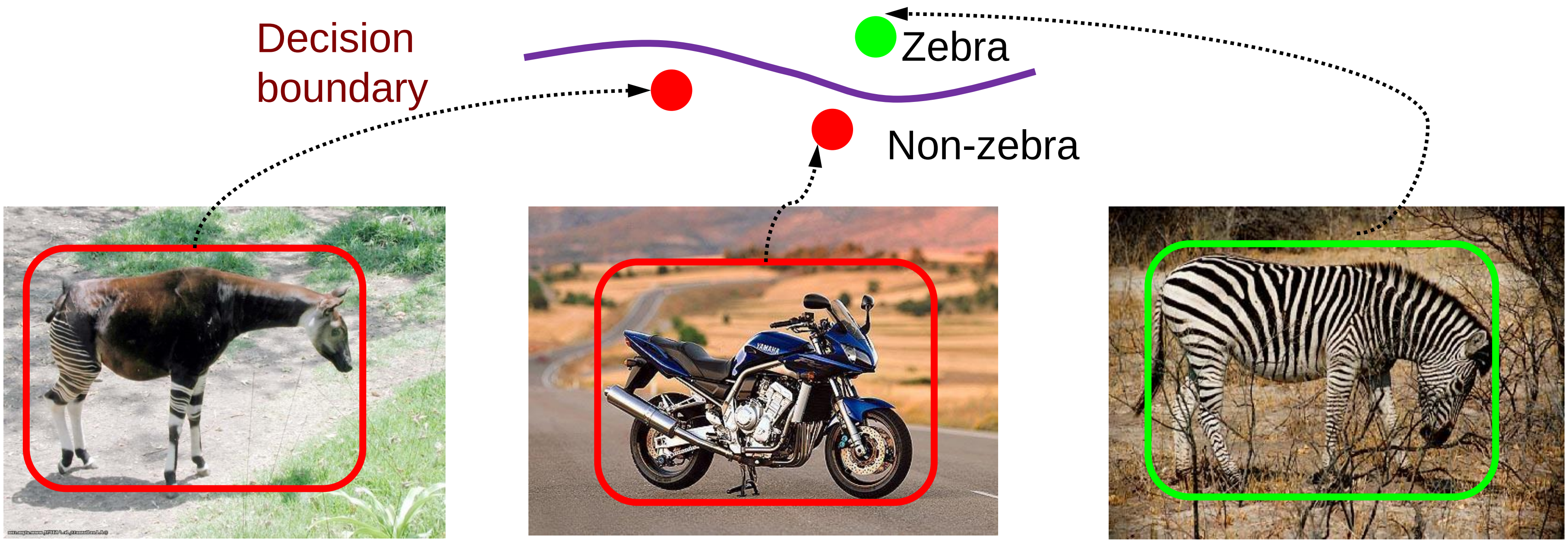
BASIC IDEA:

- Representative set of images in each category is collected
- An image is represented by a collection of "visual words"
- Object categories are modeled by the distributions of these visual words



Concept Recognition: Discriminative Model

- Object detection and recognition is formulated as a classification problem. The image is partitioned into a set of overlapping windows, and a decision is taken at each window about if it contains a target object or not.
- Each window is represented by a large number of features that encode info such as boundaries, textures, color, spatial structure.
- The classification function, that maps an image window into a binary decision, is learnt using methods such as SVMs or neural networks



Contents

- Color Image Representations
- Advanced Image Representations
- Location Representation

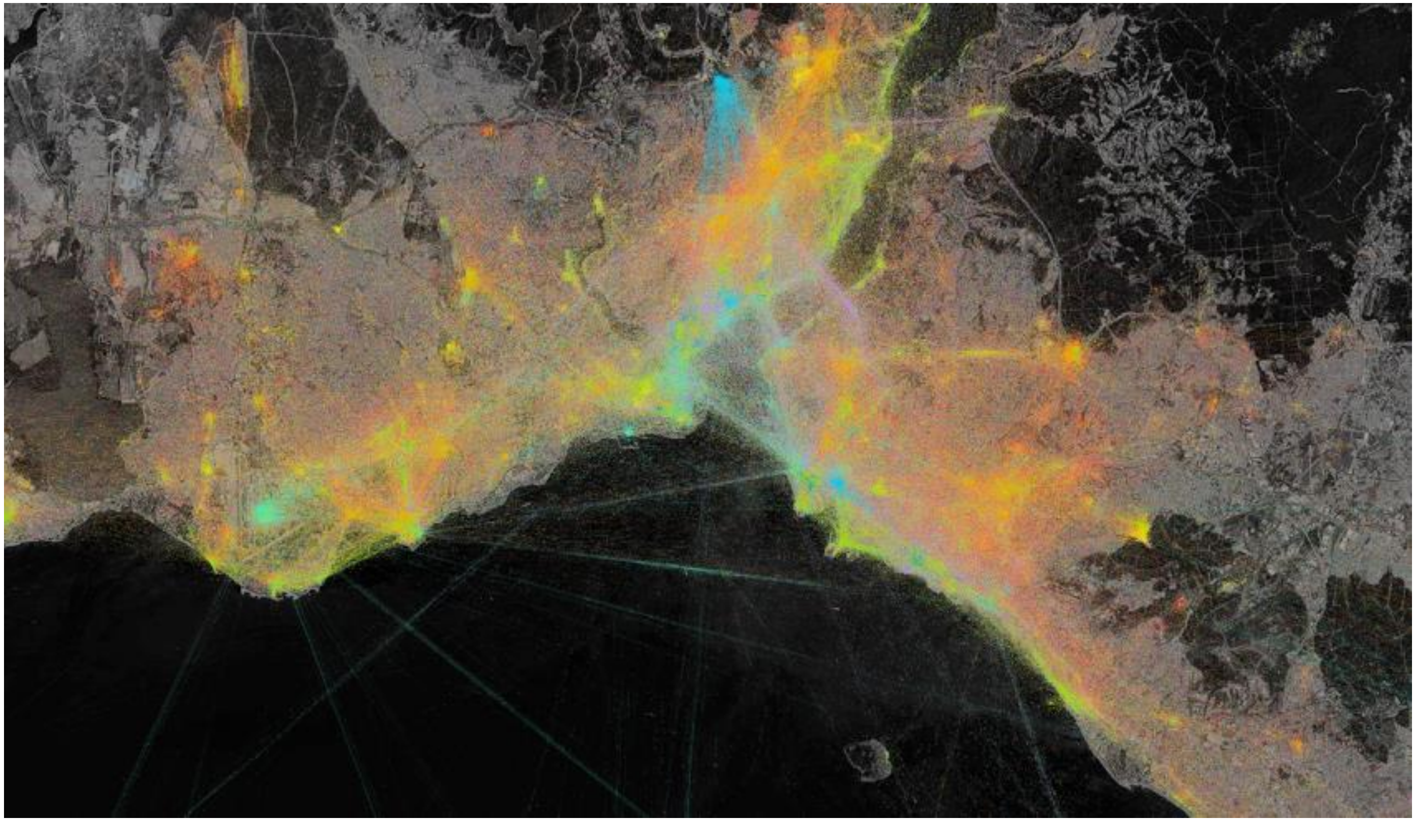
Location-Based Social Networks

- People want to share their geographic position with their friends.

Foursquare: Main Player in LBSN market

- After five years of intense crowdsourcing that generated billions of check-ins, Foursquare is evolving to become the search engine of the city.

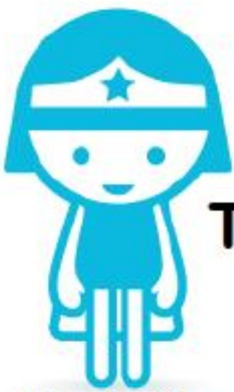
Sensing the City



- Istabul: https://www.youtube.com/watch?v=pnkD7OnvCgY
- London: https://www.youtube.com/watch?v=gsXs5TEPzRM
- Tokyo: https://www.youtube.com/watch?v=jtwzADysoMQ

The Multi-Dimensional Check-In

Social

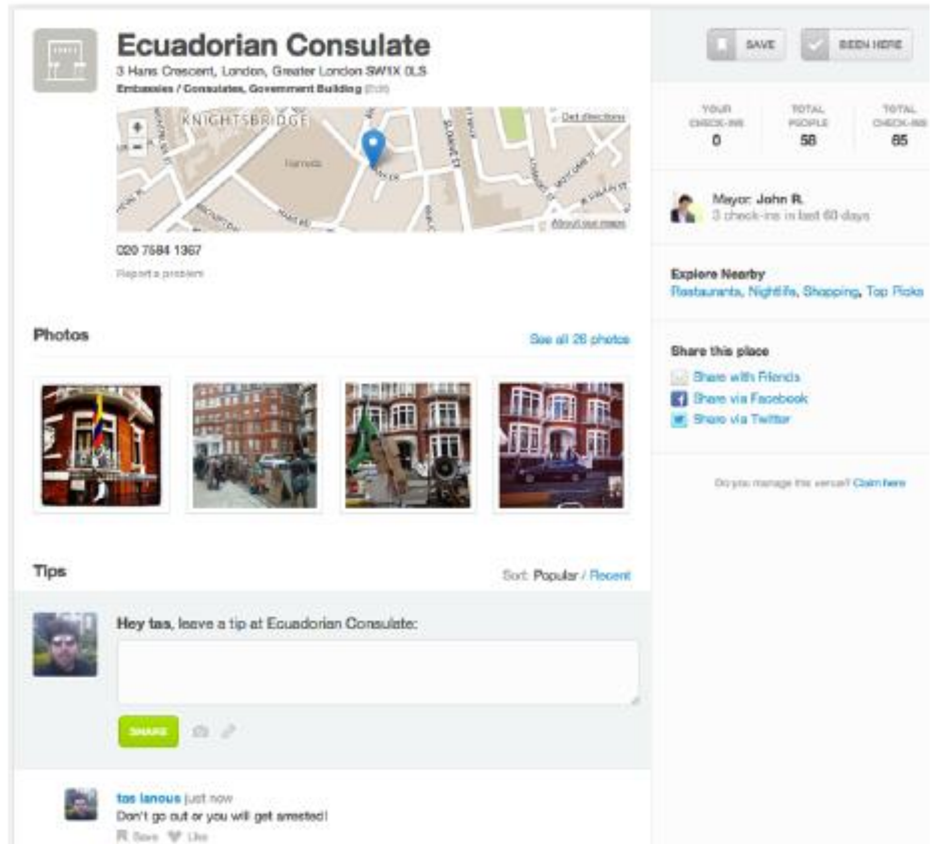


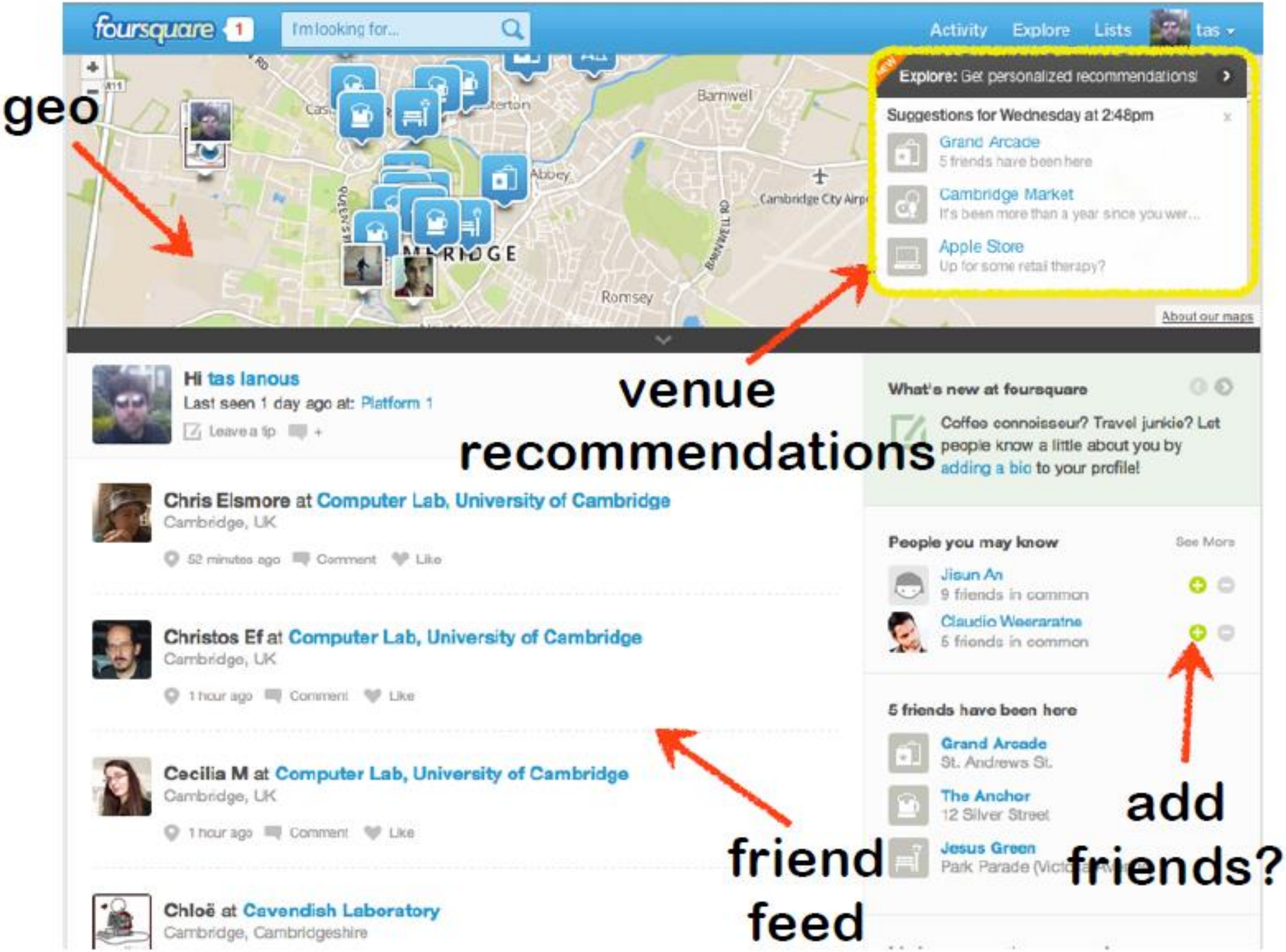
Geo

Temporal

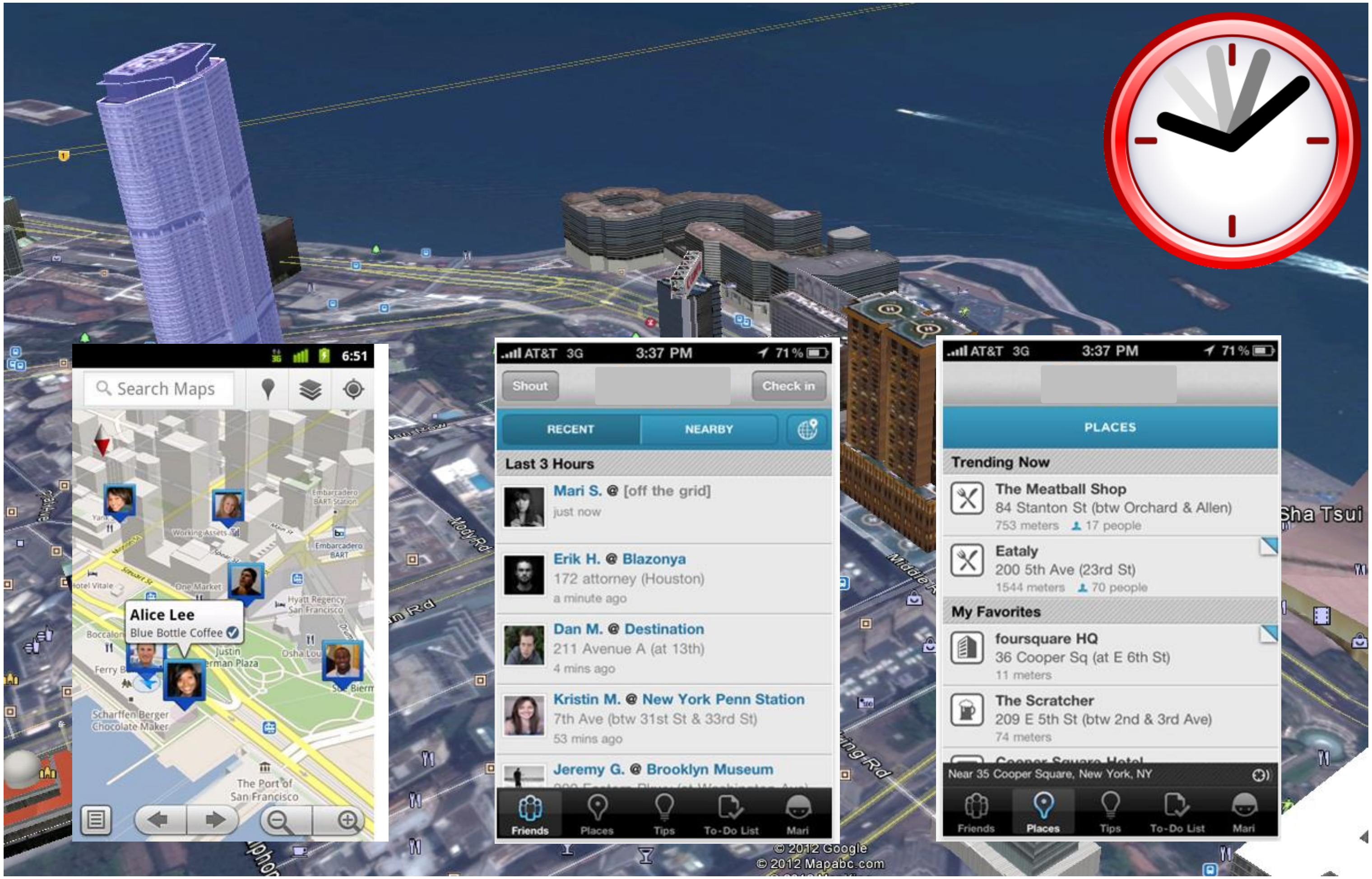
Mobile

Linguistic
Multi-Media
Game Features

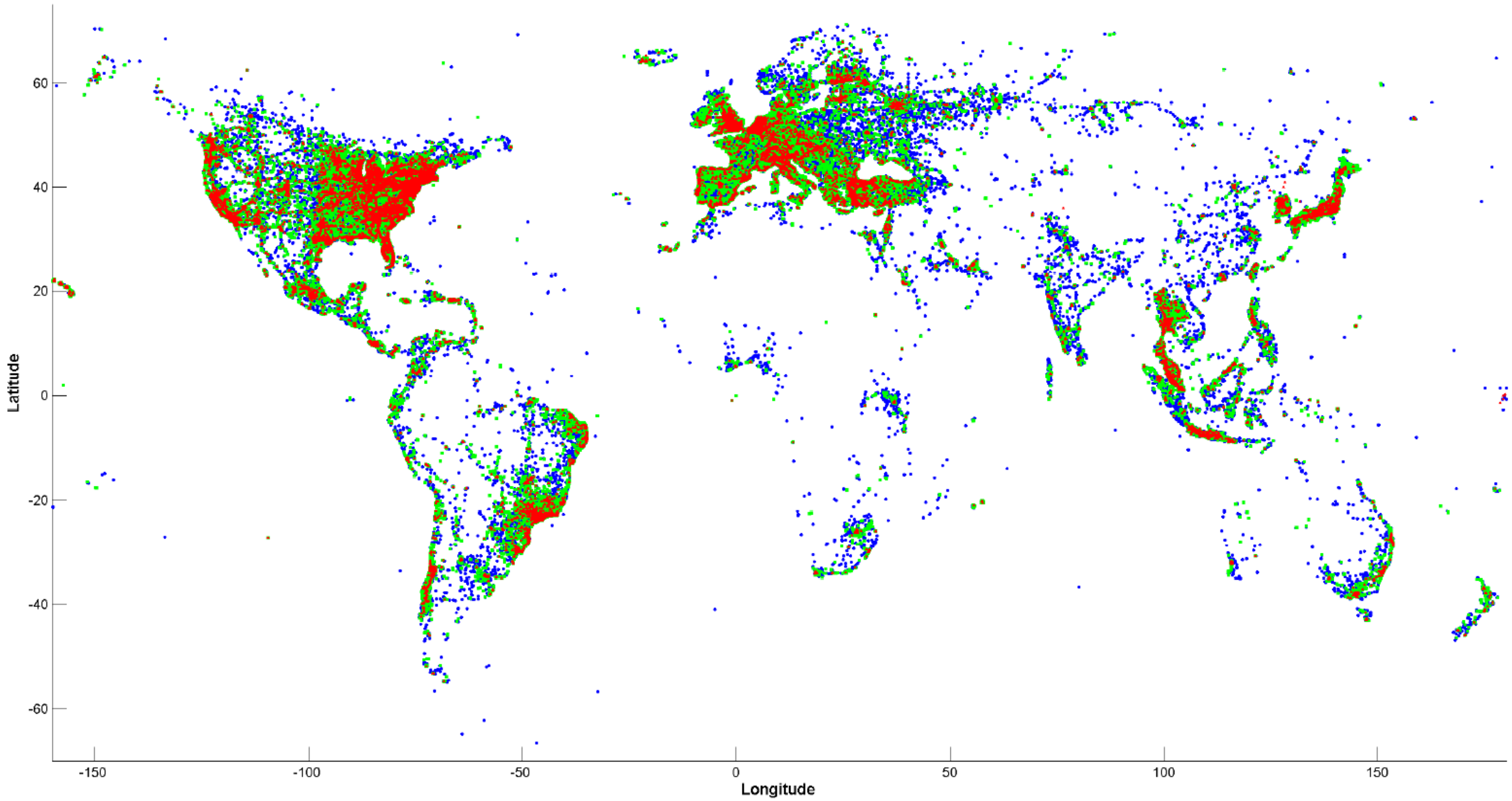



foursquare
I'm looking for...
Activity
Explore
Lists
tas
Barnwell
Abbey
Romsey
Cambridge City Airp
Explore: Get personalized recommendations!
Suggestions for Wednesday at 2:48pm
Grand Arcade
5 friends have been here
Cambridge Market
It's been more than a year since you wer...
Apple Store
Up for some retail therapy?
About our maps
Hi tas lanous
Last seen 1 day ago at: Platform 1
Leave a tip
Chris Elsmore at Computer Lab, University of Cambridge
Cambridge, UK
52 minutes ago
Comment
Like
Christos Ef at Computer Lab, University of Cambridge
Cambridge, UK
1 hour ago
Comment
Like
Cecilia M at Computer Lab, University of Cambridge
Cambridge, UK
1 hour ago
Comment
Like
Chloë at Cavendish Laboratory
Cambridge, Cambridgeshire
What's new at foursquare
Coffee connoisseur? Travel junkie? Let people know a little about you by adding a bio to your profile!
People you may know
See More
Jisun An
9 friends in common
Claudio Weeraratne
5 friends in common
5 friends have been here
Grand Arcade
St. Andrews St.
The Anchor
12 Silver Street
Jesus Green
geo
venue recommendations
friend feed
add friends?

Examples of 4sq users' activities



Large scale: Global User Mobility Analysis



Global distribution of sampled Foursquare venues. Colors represent the popularity of venues with “red”: # check-ins > 100, “green”: 50 ≤ # check-ins ≤ 100 and “blue”: 10 ≤ # check-ins < 50.

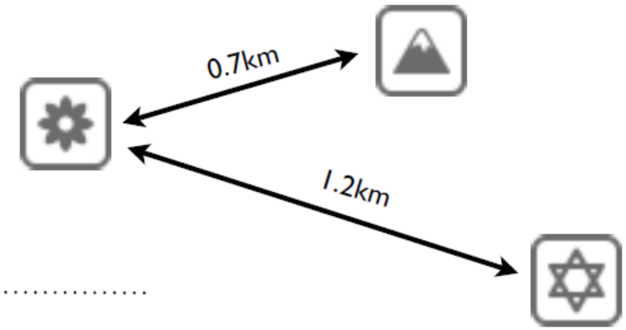
Location Data Representation -1

User Specific features

- historic visits
- friend check-ins
- preferred venue types

Geographic

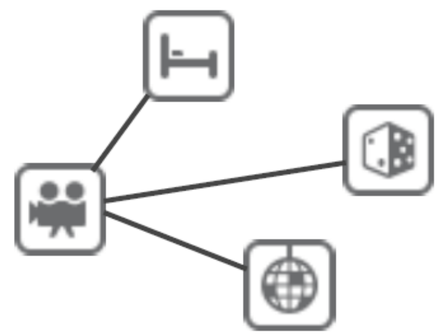
- distance and rank-distance



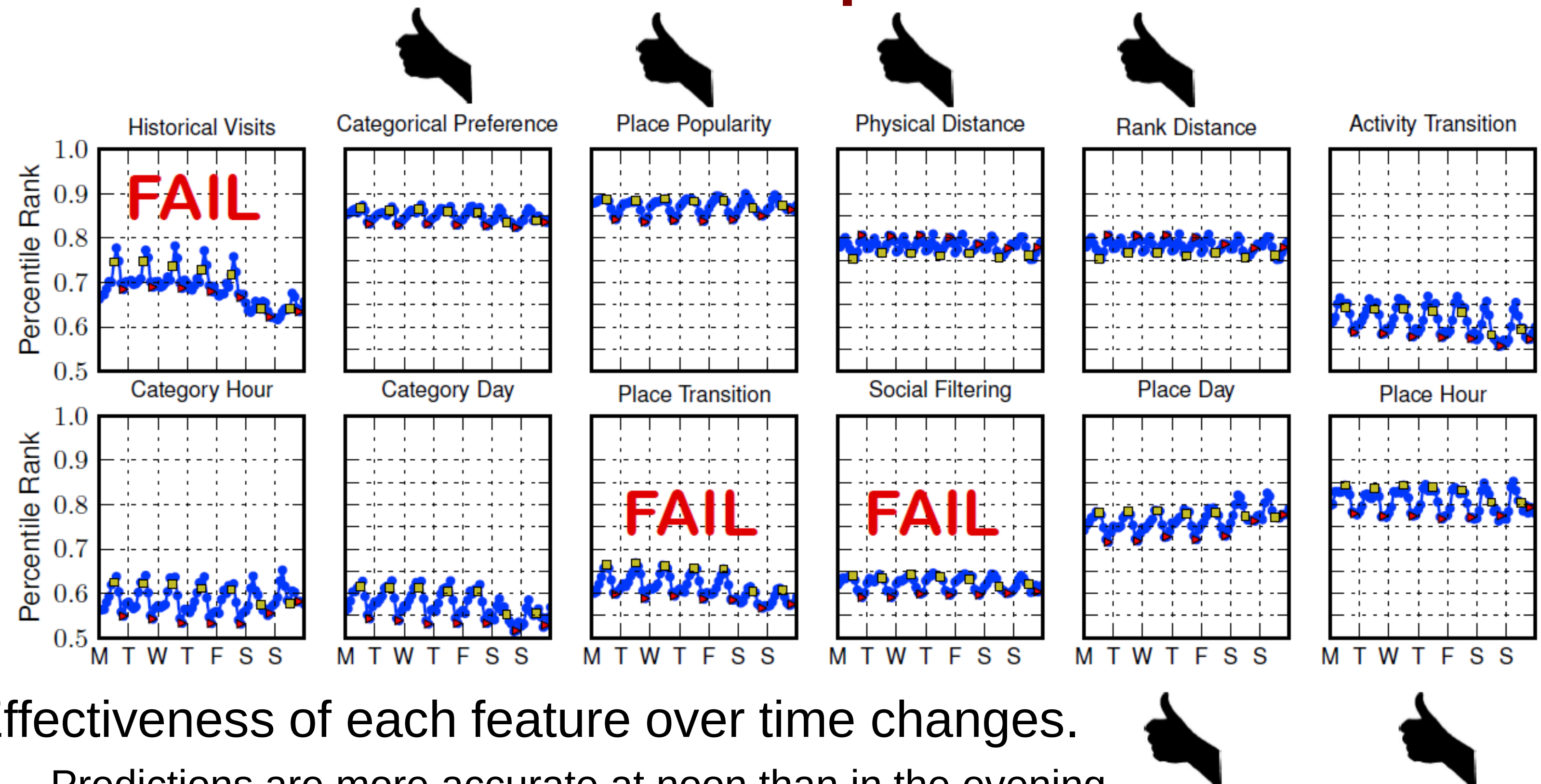
Temporal

- trending places (hour/day)
- trending place types (eg. cinema at nights)

Place Network



Location Data Representation -2

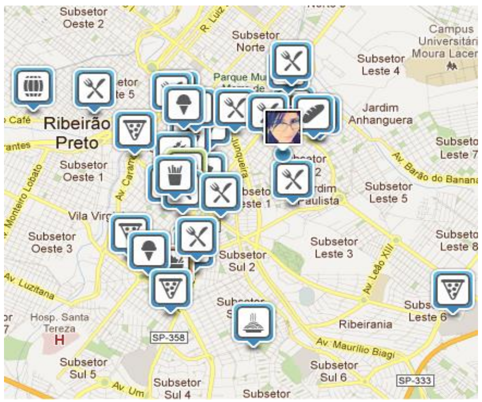


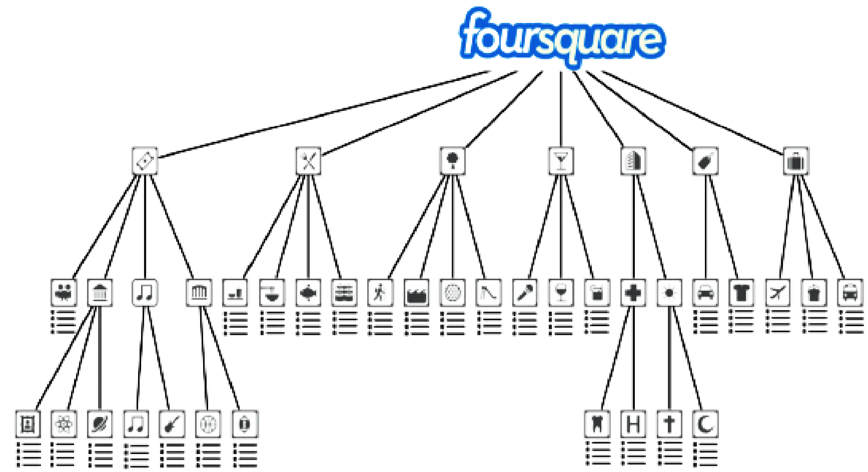
- Effectiveness of each feature over time changes.
  - Predictions are more accurate at noon than in the evening
  - Predictions for Physical & Rank distances reverse -- users cover shorter distance at night
  - Predictions for Historical Visits & Place Transition drop significantly over weekends
  - whereas Categorical Preference, Place Popularity & distance based features are more stable

*A Noulas S Scellato, N Lathia & C Mascolo (2012). Mining User Mobility Features for Next Place Prediction in Location-based Services. IEEE Int'l Conf. on Data Mining (ICDM), Dec 2012.

Location Data Representation -3

Map all Foursquare check – ins to Foursquare categories from category hierarchy.



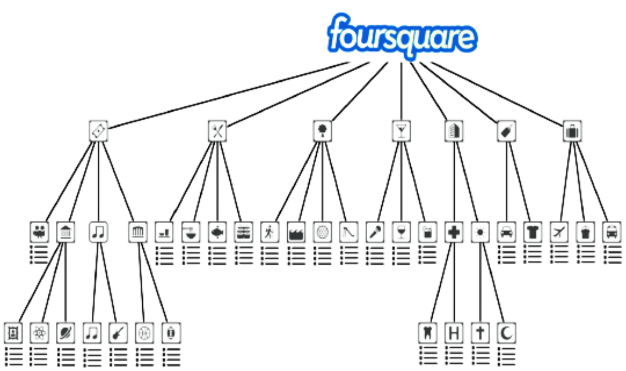


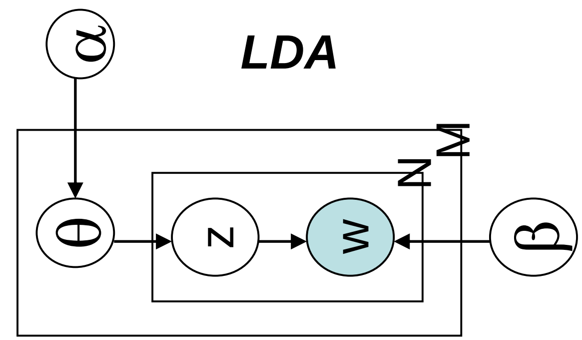
For case when user performed check-ins in two restaurants and airport but did not perform check-ins in other venues:

	$Category_1$	...	$Category_{restaurant}$	...	$Category_{airport}$	...	$Category_n$
U_1	0	0	2	0	1	0	0
...	*	*	*	*	*	*	*
U_n	*	*	*	*	*	*	*

Location Data Representation -4

1. Map all Foursquare check – ins to Foursquare venue categories from category hierarchy.
2. Form user – related documents, containing venue categories of every check-in
3. Apply LDA on it represent as distribution among n latent topics, where Users – documents, words – Foursquare venue categories



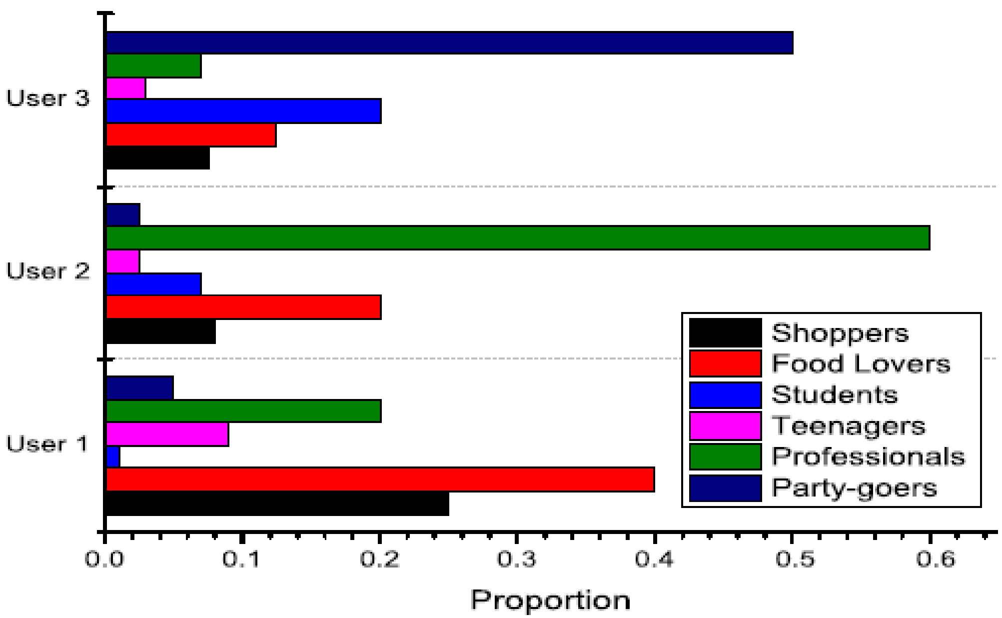


	LDA_1	LDA_2	LDA_3	…	LDA_n
U_1	0.05	**0.4**	**0.1**	0.35	0.1
…	*	*	*	*	*
U_n	*	*	*	*	*

Location Data Representation -5

Category distribution among LDA topics

ID	Categories	LDA Topics
T1	Malay Res-t, Mall, University, Indian Res-t, Aisian Res-t	Food Lovers
T2	Cafe, Airport, Hotel, Coffee Shop, Chinese Res-t	Travelers (Business)
T3	Nightclub, Mall, Food Court, Trade School, Res-t, Coffee Shop	Party Goers
T4	Home, Office, Build., Neighbor-d, Gov. Build., Factory	Family Guys (Youth)
T5	University (Collage), Gym, Airport, Hotel, Fitness Club	Students
T6	Train St., Apartment, Mall, High School, Bus St.	Teenagers (Youth)



LDA word distribution over 6 topics for collected Foursquare check-ins.

Every venue category is considered as a word, each Foursquare user - as a document

Summary

- Data from different sources describe users from multiple aspects – must incorporate different data modalities
- Images could be represented in different ways:
  - Color Histograms (consider just colors)
  - CCV vectors (consider colours and it's mutual position)
  - Textures (consider repeated patterns)
  - Edges (consider edges)
  - Visual words (consider scale invariant (SIFT) features)
  - Concepts (consider high level image concepts)
  - Meta Information
- Locations are not just locations (lon, lat) but also Location Semantics (venue categories)
- Locations could be represented in different ways:
  - Venue categories distribution
  - Latent topics
  - Mobility features (Spatial – Temporal aspect)

Next Lesson

- **Introduction to Classification**

Assignment -1

- DATASET:
- http://lms.comp.nus.edu.sg/research/NUS-MULTISOURCE.htm
- DESCRIPTION OF THE DATA IS IN PAPER*.
- Please, ask any questions during the conference and after: farseev@u.nus.edu

*Aleksandr Farseev, Liqiang Nie, Mohammad Akbari, and Tat-Seng Chua. 2015. ***Harvesting Multiple Sources for User Profile Learning: a Big Data Study*** In Proceedings of the 5th ACM on International Conference on Multimedia Retrieval (ICMR '15).

Assignment -2

- All slides and will be here:
  - http://farseev.com/ainlfruct.htm
- Recommended software to use:
  - KNIME (No programming required) https://www.knime.org/
  - Python and it's Machine Learning Support
  - Any other language you like. Just make it work ;)