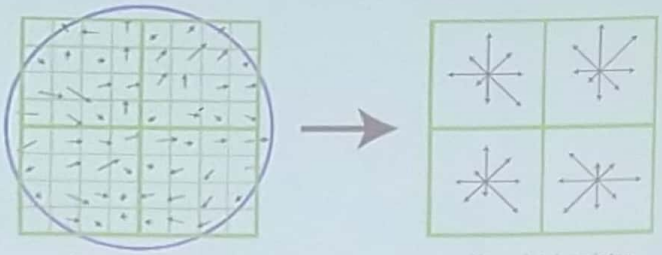
SIFT descriptor

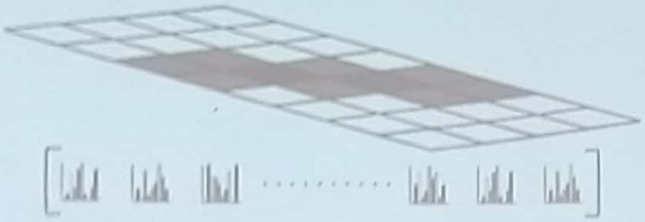
- The most popular gradient-based descriptor
- Typically used in combination with an interest point detector



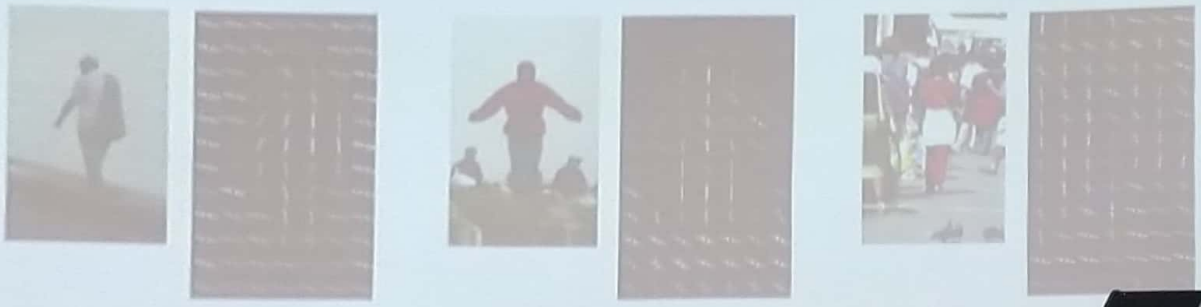
  - Region rescaled to a grid of 16x16 pixels
  - 4x4 regions = 16 histograms (concatenated)
  - Histograms: 8 orientation bins, gradients weighted by gradient magnitude
  - Final descriptor has 128 dimensions and is normalized to compensate for illumination differences

HOG Descriptor

- Concatenation of Blocks



- Visualization:



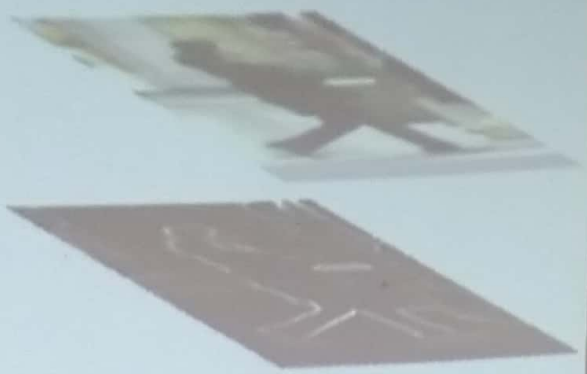
Descriptor

1. Compute gradients on an image region of 64x128 pixels

2. Compute histograms on 'cells' of typically 8x8 pixels (i.e. 8x16 cells)

3. Normalize histograms within overlapping blocks of cells (typically 2x2 cells, i.e. 7x15 blocks)

4. Concatenate histograms



Cell →

Block →

Overlap of Blocks →

Feature Sets

- Haar wavelets + SVM:
  - Papageorgiou & Poggio (2000)
  - Mohan et al (2001)
  - DePoortere et al (2002)
- Rectangular differential features + adaBoost:
  - Viola & Jones(2001)
- Parts based binary orientation position histogram + adaBoost:
  - Mikolajczk et al (2004)
- Edge templates + nearest neighbor:
  - Gavrila & Philomen (1999)
- Dynamic programming:
  - Felzenszwalb & Huttenlocher (2000),
  - Loffe & Forsyth (1999)
- Orientation histograms:
  - C.F. Freeman et al (1996)
  - Lowe(1999)
- Shape contexts:
  - Belongie et al (2002)
- PCA-SIFT:
  - Ke and Sukthankar (2004)

Feature Descriptors

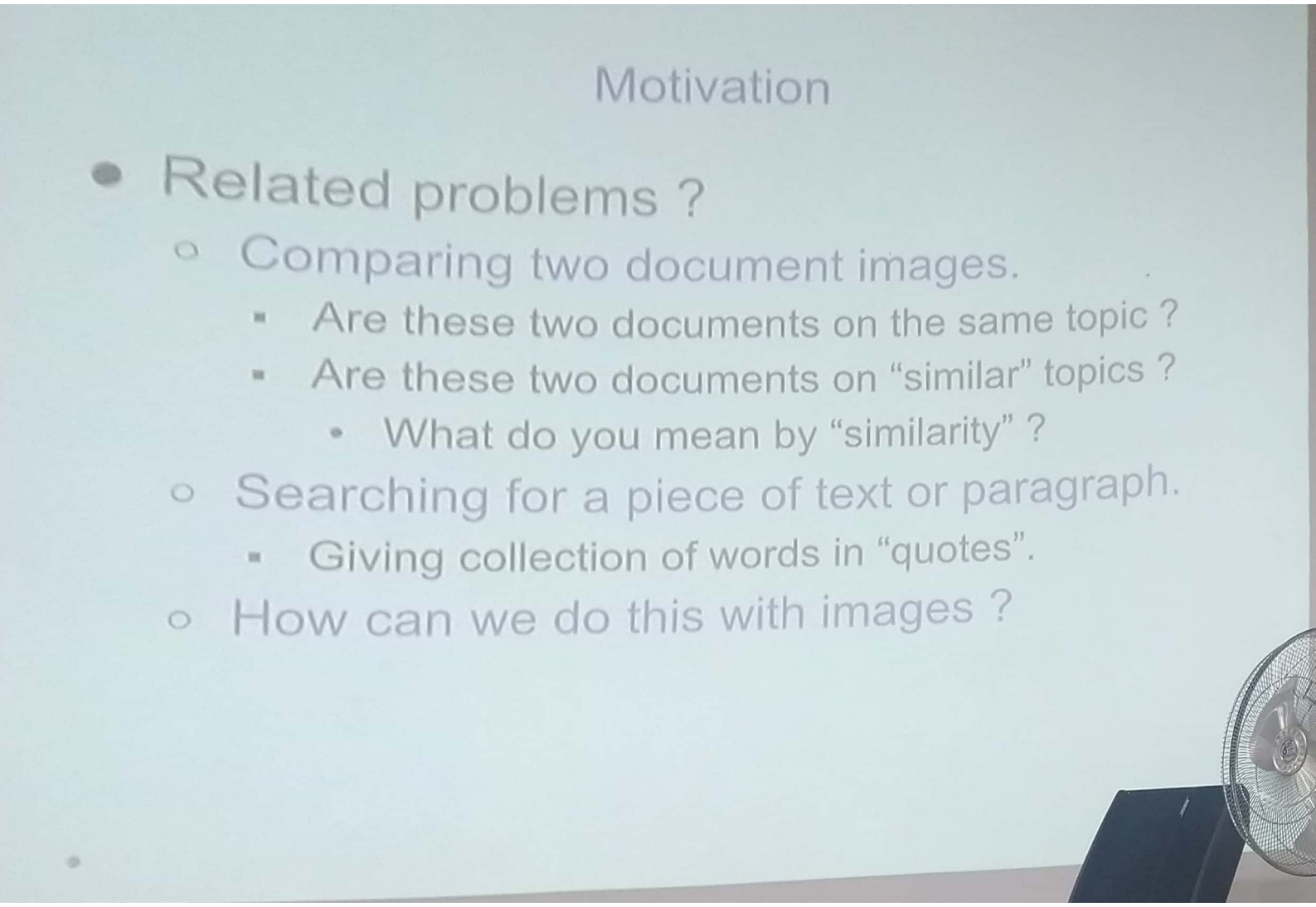
Local/Patch

- SIFT
- SURF
- FAST
- BRIEF
- ORB (≈FAST+BRIEF)
- GLOH

Global/Object

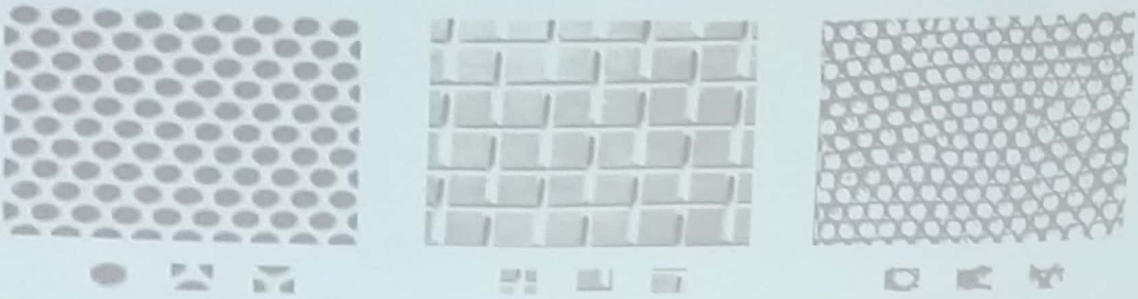
- HOG
- GIST
- Shape Context

Speeded Up Robust Features (SURF)
Oriented FAST and rotated BRIEF (ORB)
Gradient Location and Orientation Histogram (GLOH)
Binary Robust Independent Elementary Features (BRIEF)
Features from Accelerated Segment Test (FAST)



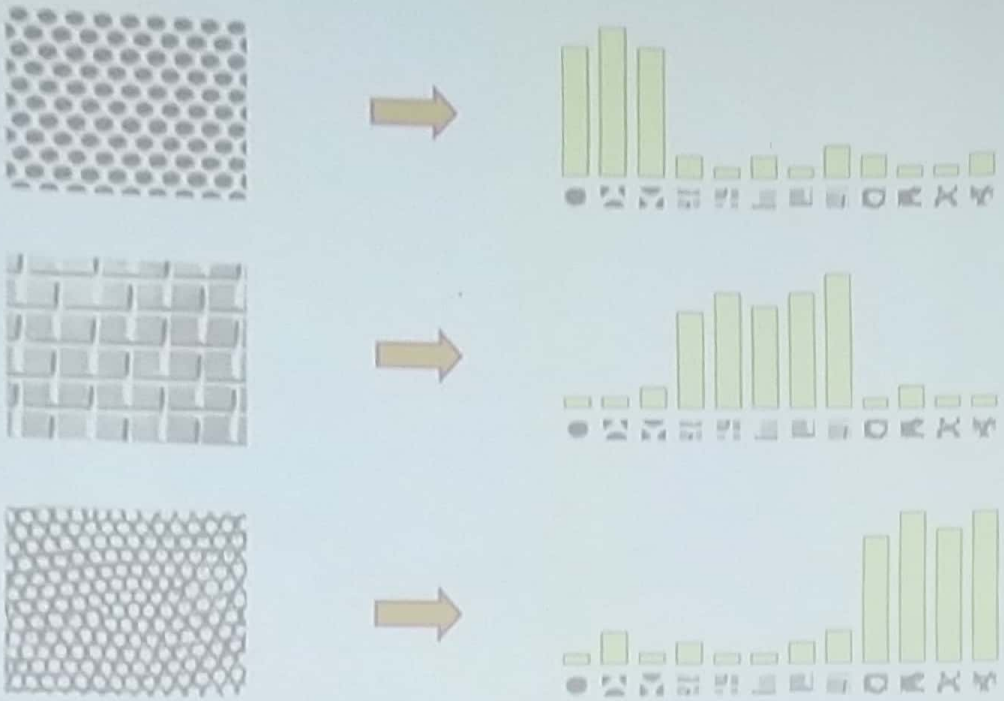
Origin 1: Texture Recognition

- A texture is characterized by the repetition of basic elements called textons.
- For stochastic textures, the repetition patterns are not regular.



Julesz, 1981; Cula and Dana, 2001; Leung and Malik, 2001; Mori, Belongie and Malik, 2001; Schmid 2001, Varma and Zisserman 2002, 2003; Lazebnik, Schmid and Ponce 2003.

Origin 1: Texture Recognition



Julesz, 1981; Cula and Dana, 2001; Leung and Malik, 2001; Mori, Belongie and Malik, 2001; Schmid 2001, Varma and Zisserman 2002, 2003; Lazebnik, Schmid and Ponce 2003.

Functions for comparing histograms

- L1 distance

$$D(h_1, h_2) = \sum_{i=1}^{N} | h_1(i) - h_2(i) |$$

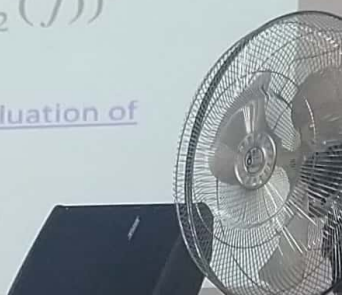
- χ^2 distance

$$D(h_1, h_2) = \sum_{i=1}^{N} \frac{(h_1(i) - h_2(i))^2}{h_1(i) + h_2(i)}$$

- Quadratic distance (*cross-bin*)

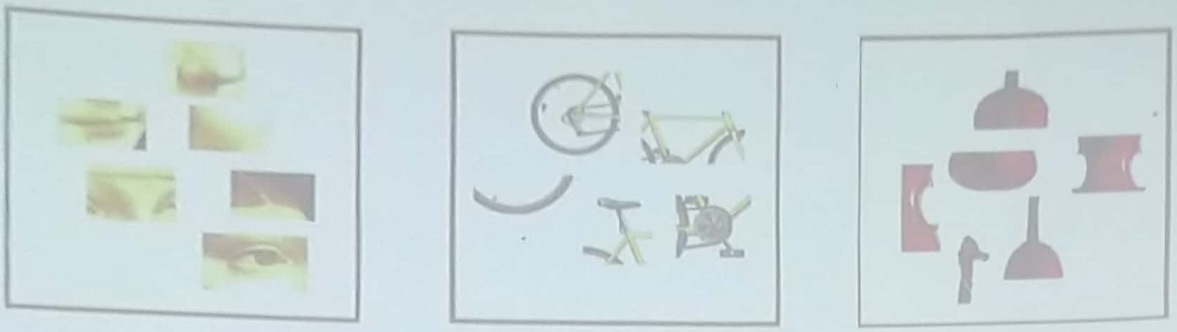
$$D(h_1, h_2) = \sum_{i,j} A_{ij} (h_1(i) - h_2(j))^2$$

Jan Puzicha, Yossi Rubner, Carlo Tomasi, Joachim M. Buhmann: Empirical Evaluation of Dissimilarity Measures for Color and Texture. ICCV 1999

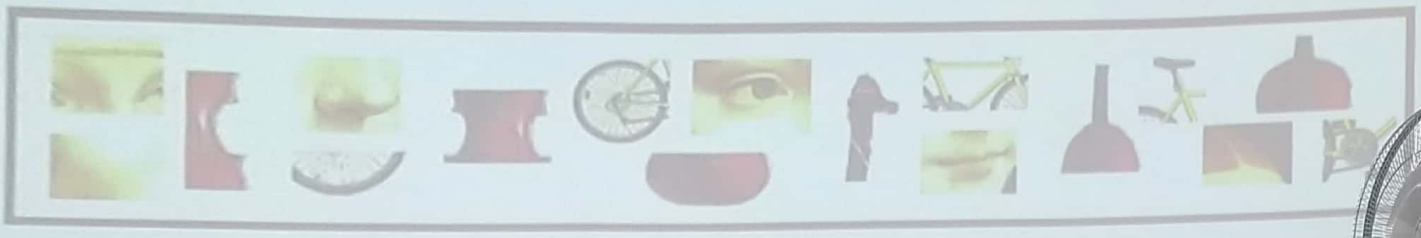


Bag of Features for Image Classification

- Extract Features



- Learn "visual vocabulary"



Difference Between Features & Words

Words

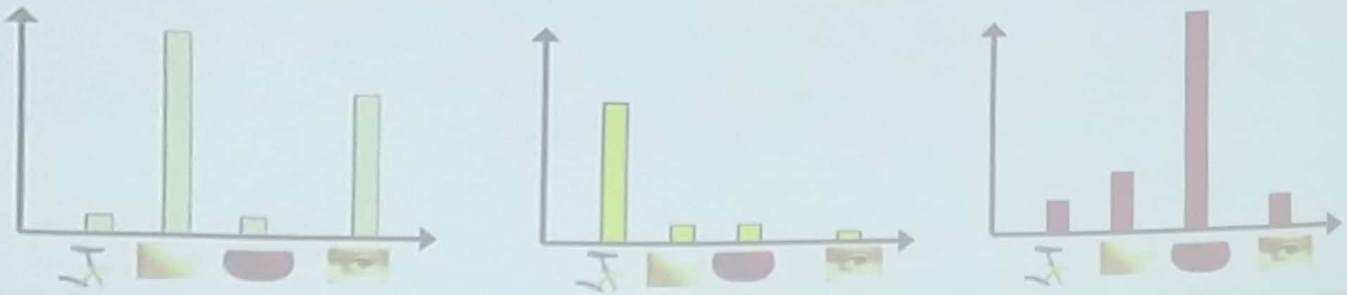
- Dictionary/Vocab.
- Meaning.
- Finite/Precise.
- Language known.

Features

- Dictionary/Vocab ?
- Meaning ?
- Finite/Precise ?
- What Language ?

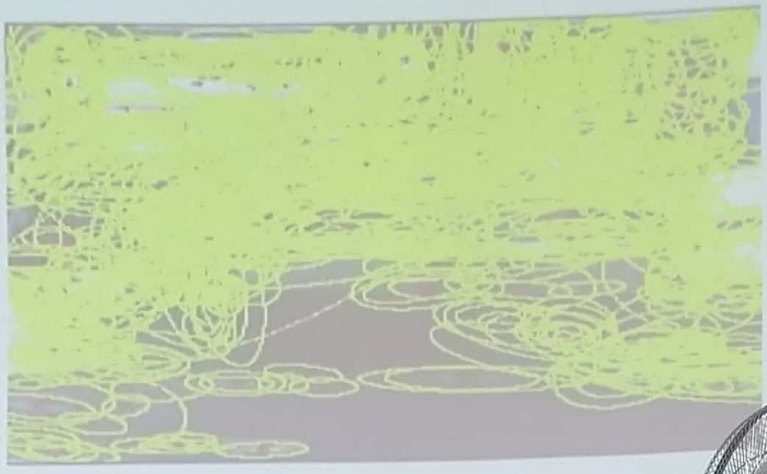
Bag of Features for Image Classification

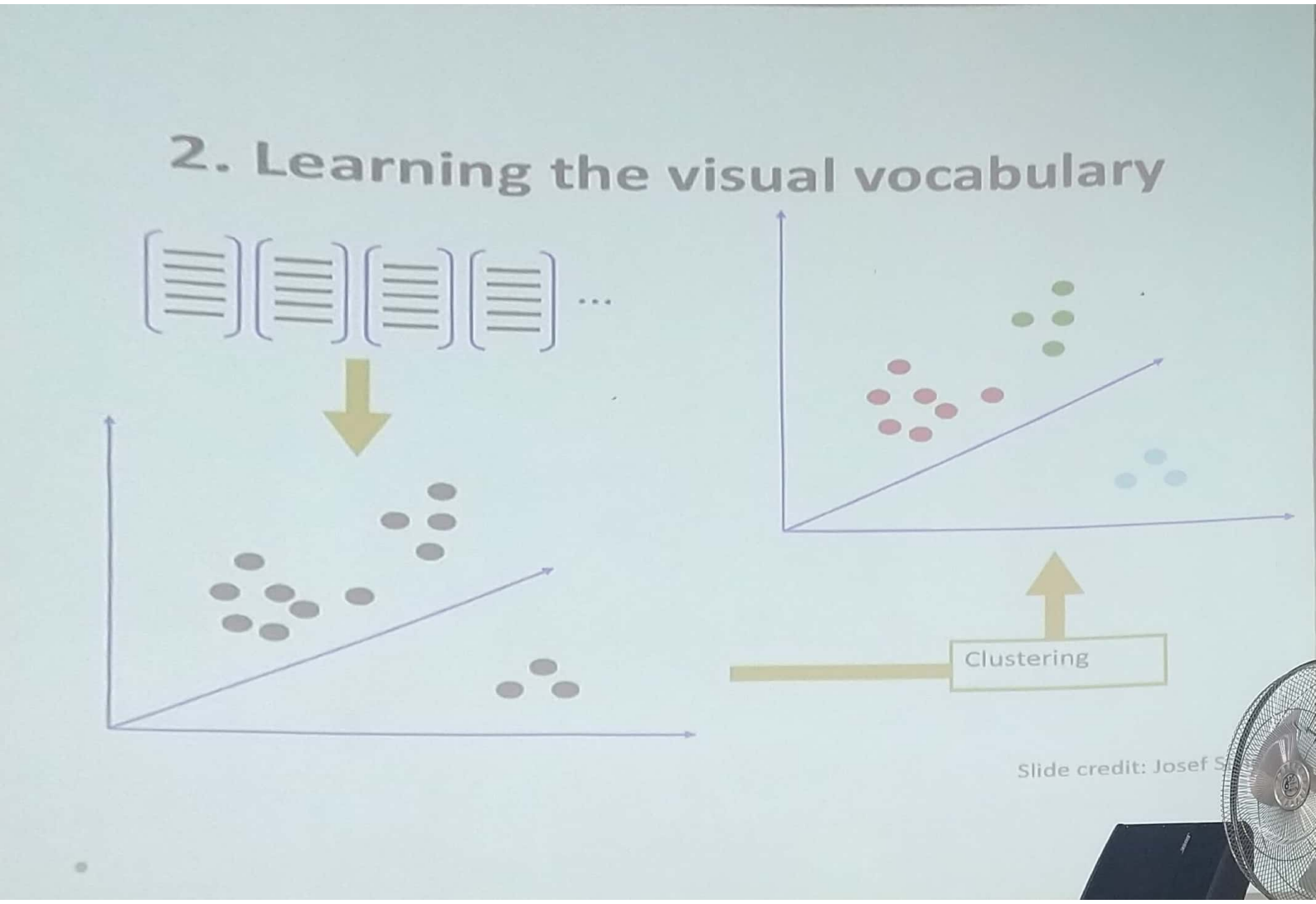
- Extract Features
- Learn "visual vocabulary"
- Quantize features using visual vocabulary
- Represent images by frequencies of visual words



1. Feature extraction

- Regular grid
  - Vogel & Schiele, 2003
  - Fei-Fei & Perona, 2005
- Interest point detector
  - Csurka et al. 2004
  - Fei-Fei & Perona, 2005
  - Sivic et al. 2005





K-means clustering

- Want to minimize sum of squared Euclidean distances between points x_i and their nearest cluster centers m_k

$$D(X, M) = \sum_{\text{cluster } k} \sum_{\substack{\text{point } i \text{ in} \\ \text{cluster } k}} (x_i - m_k)^2$$

- Algorithm:
- Randomly initialize K cluster centers
- Iterate until convergence:
  - Assign each data point to the nearest center
  - Recompute each cluster center as the mean of all points assigned to it

From clustering to vector quantization

- Clustering is a common method for learning a visual vocabulary or codebook
  - Unsupervised learning process
  - Each cluster center produced by k-means becomes a codevector
  - Codebook can be learned on separate training set
  - Provided the training set is sufficiently representative, the codebook will be "universal"

- The codebook is used for quantizing features
  - A *vector quantizer* takes a feature vector and maps it to the index of the nearest codevector in a codebook
  - Codebook = visual vocabulary
  - Codevector = visual word

Difference between Features & Words

Features

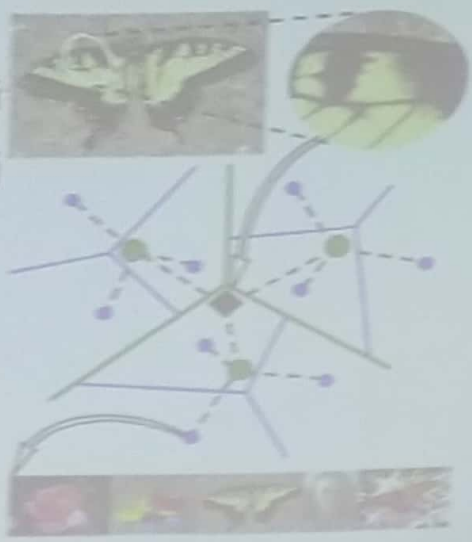
- Dictionary/Vocab ?
- Meaning ?
- Finite/Precise ?
- What language ?
- Complicated!
- Training subjective!

Words

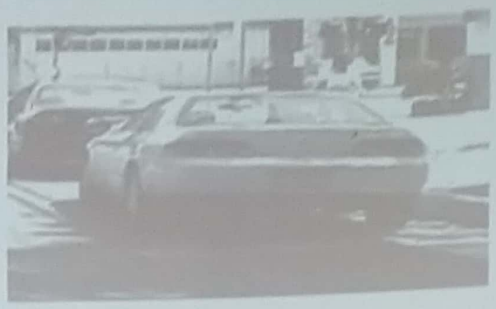
- Dictionary/Vocab.
- Meaning.
- Finite/Precise.
- Language known.
- Simple.
- Language-objective

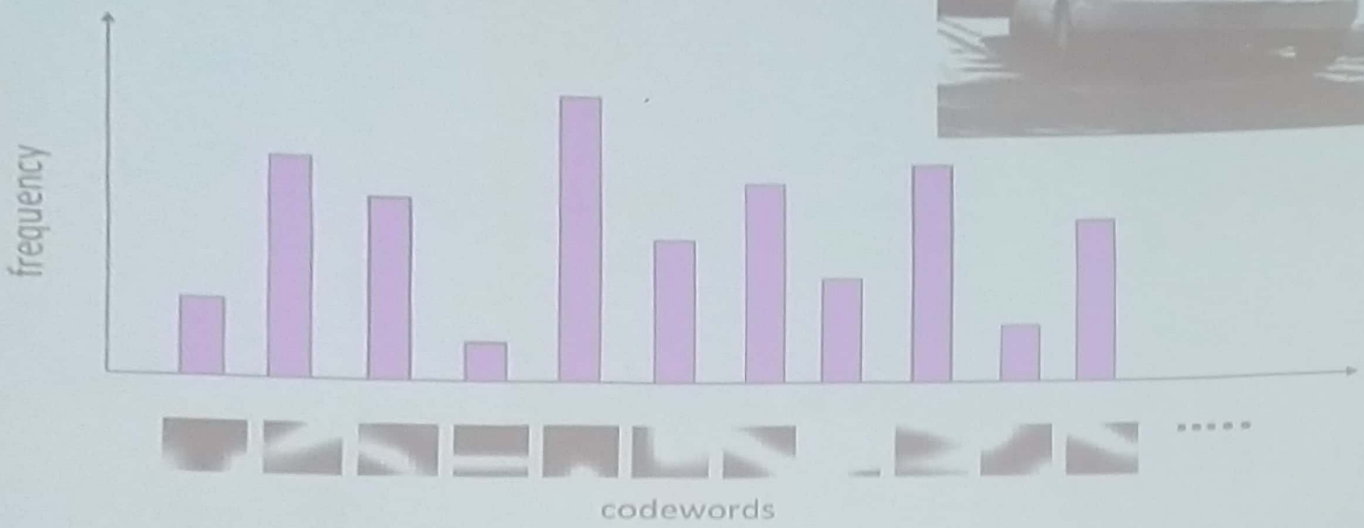
Visual vocabularies: Issues

- How to choose vocabulary size?
  - Too small: visual words not representative of all patches
  - Too large: quantization arti
- Generative or discriminati
- Computational efficiency
  - Vocabulary trees (Nister & Stewenius, 2006)



3. Image representation



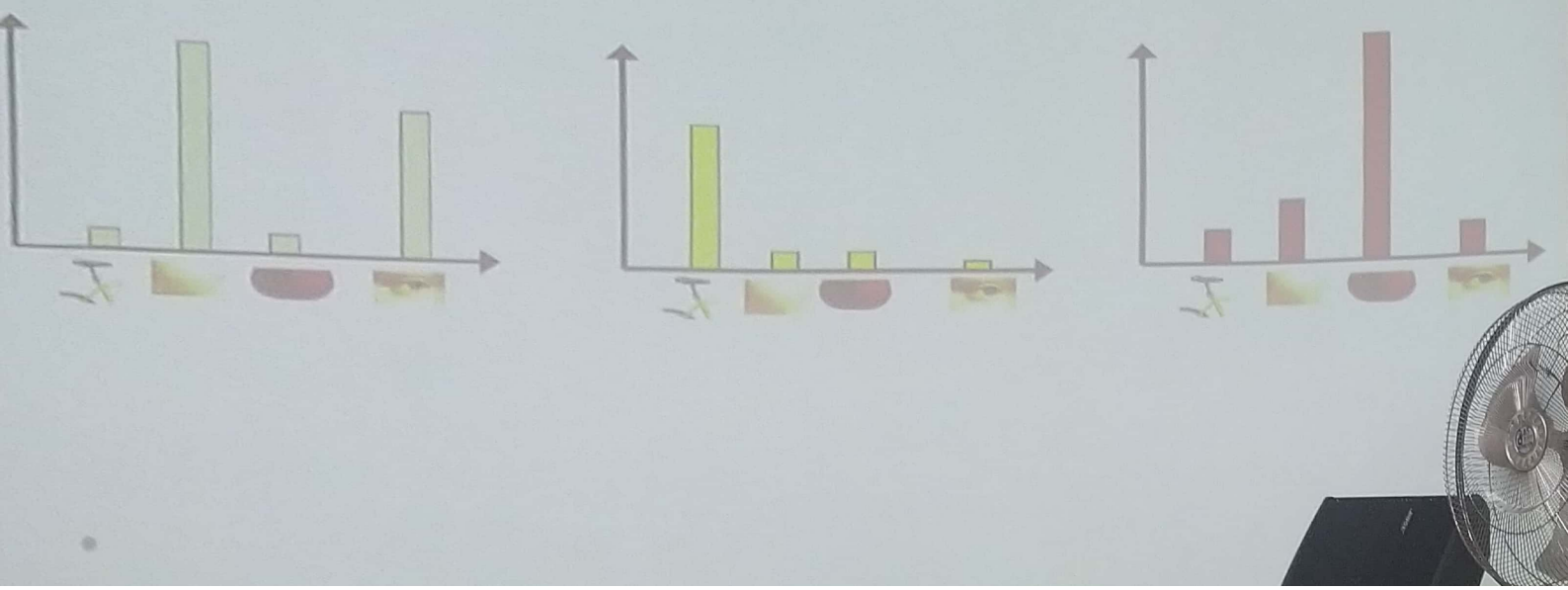


Problems we can solve (attempt) ?

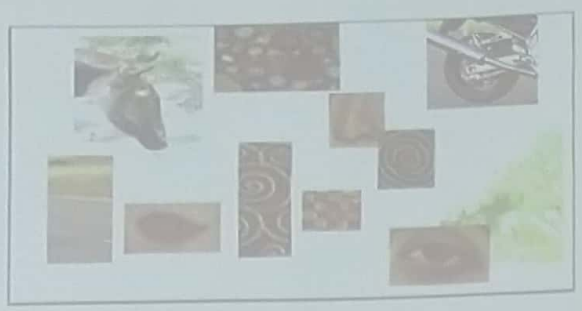
- Classification: Do these two images (visual documents) belong to the same subject ?
- Recognition: Which images contain chairs ?
- If images are documents, what are videos ?
- Actions are sequence of visual words organized in time.
- How to get better (spatial) representation (to enforce "structure" in documents/images) ?
- Search!

Image classification

- Given the bag-of-features representations of images from different classes, how do we learn a model for distinguishing them?

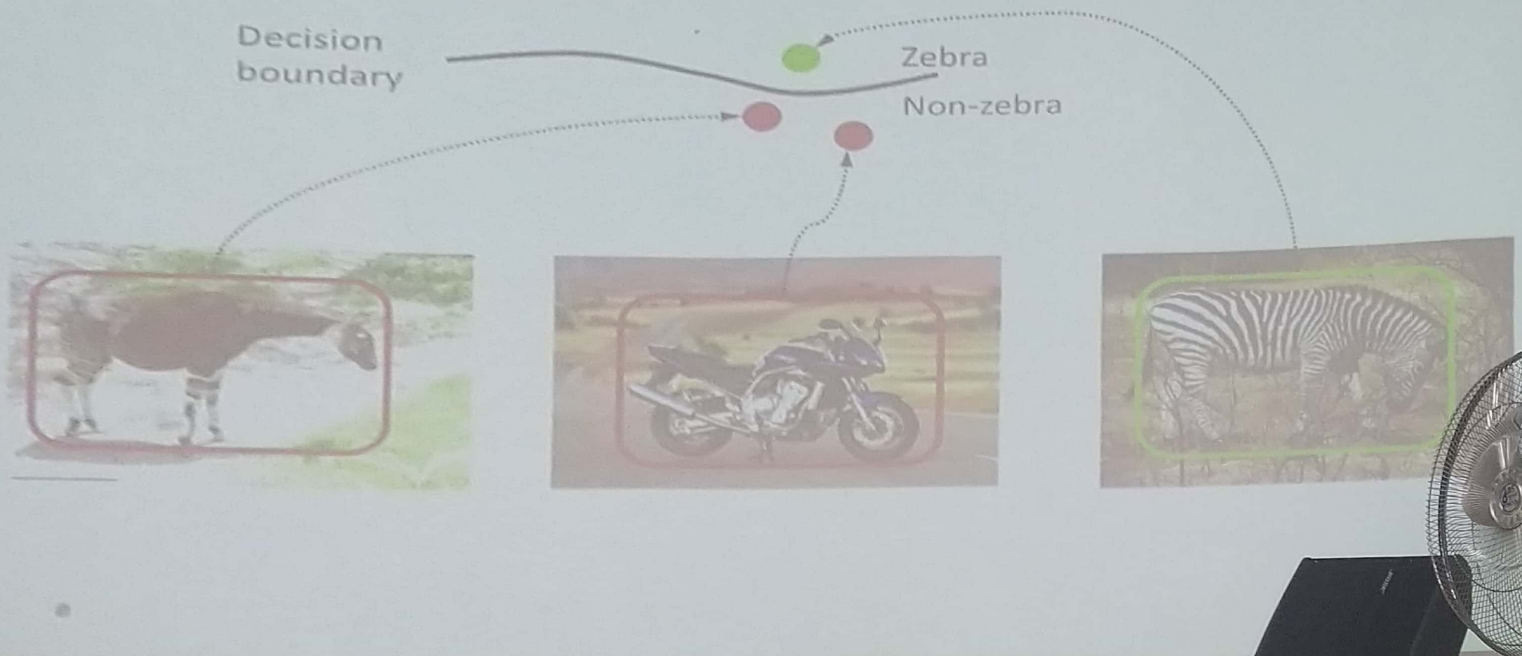


Discriminative and generative methods for bags of features



Discriminative methods

- Learn a decision rule (classifier) assigning bag-of-features representations of images to different classes

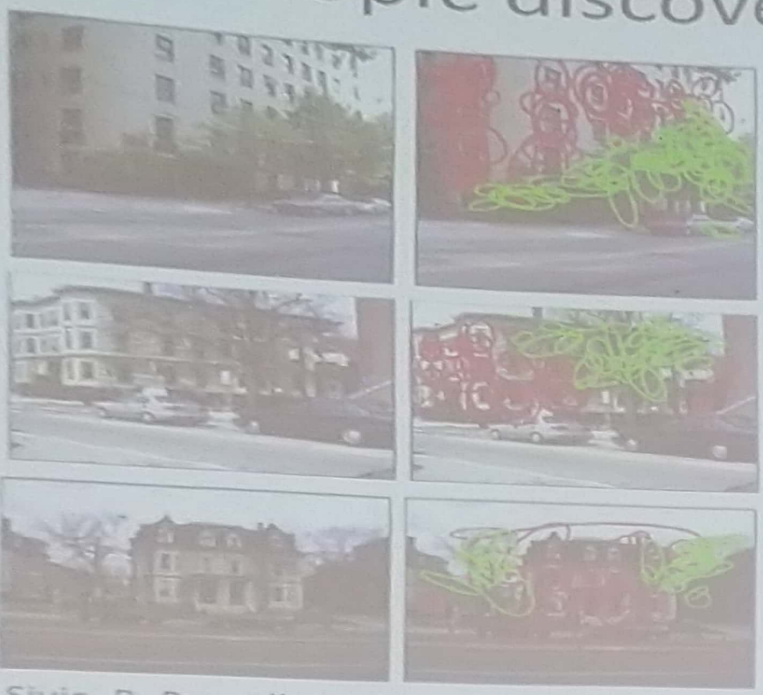


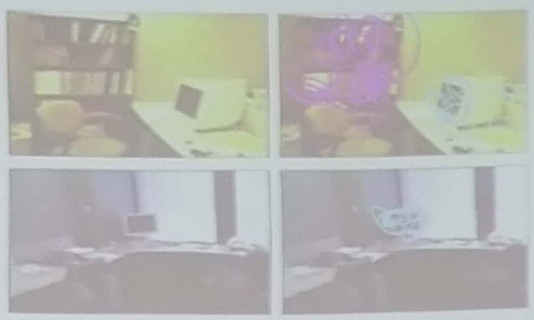
K-Nearest Neighbors

- For a new point, find the k closest points from training data
- Labels of the k points "vote" to classify
- Works well provided there is lots of data and the distance function is good

Source:

Topic discovery in images





J. Sivic, B. Russell, A. Efros, A. Zisserman, B. Freeman, Discovering Objects and their Loc
in Images, *ICCV 2005*

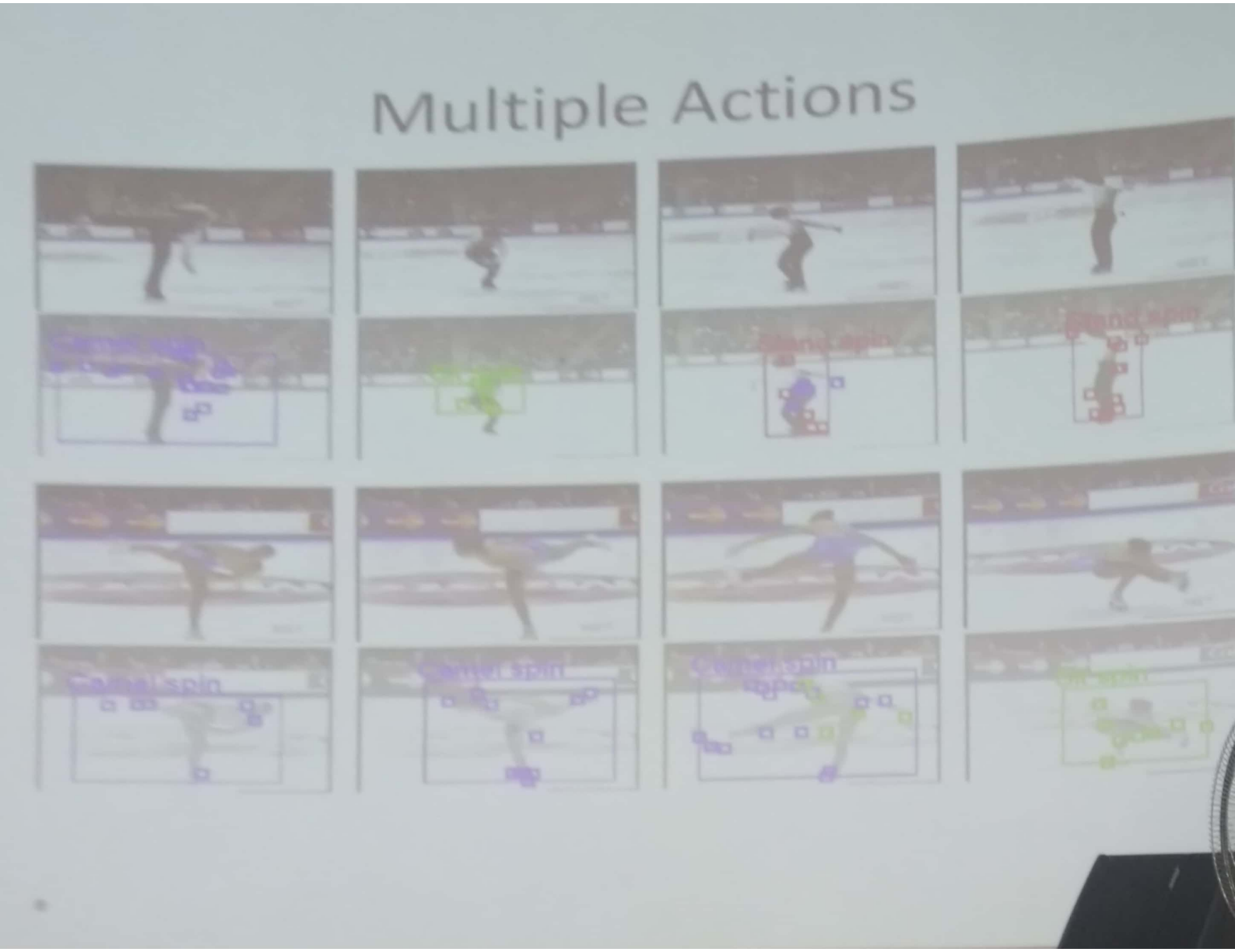


Image and Document differences

Document

- Single scale.
- Words have order in document!
- 1D space.
- Google!

Image

- Multiple scales!
- Features are *loosely* coupled.
- 2D space.
- TinEye!

Spatial pyramid representation

- Extension of a bag of features
- Locally orderless representation at several levels of resolution

Lazebnik, Schmid & Ponce (CVPR 2006)

Slide: S. Laz

Image Search!

- What kind of searches possible ?
  - We will see at least 2 types.
- Why is image search important ?
  - Copyright problems / attribution !
  - Words might not be enough. (Find me images of dresses / goggles worn by …. in ….. movie)
  - What is the name of the monument in this image ?