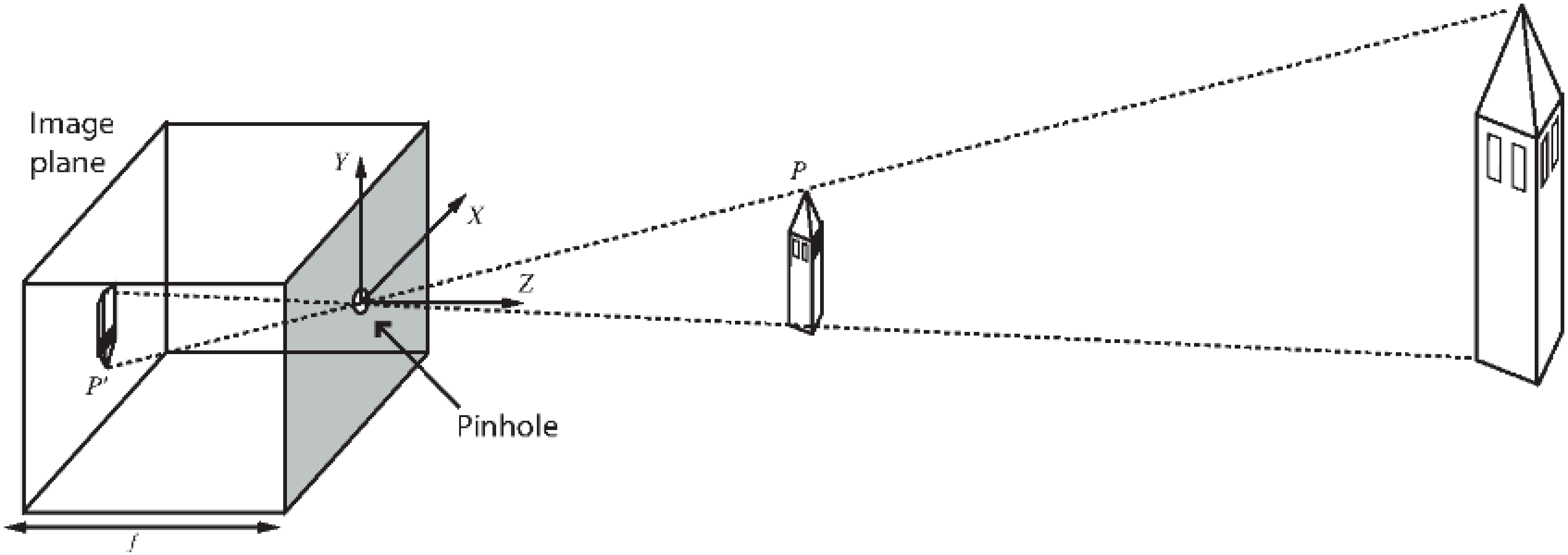
# Computer Vision

# Pinhole-camera model



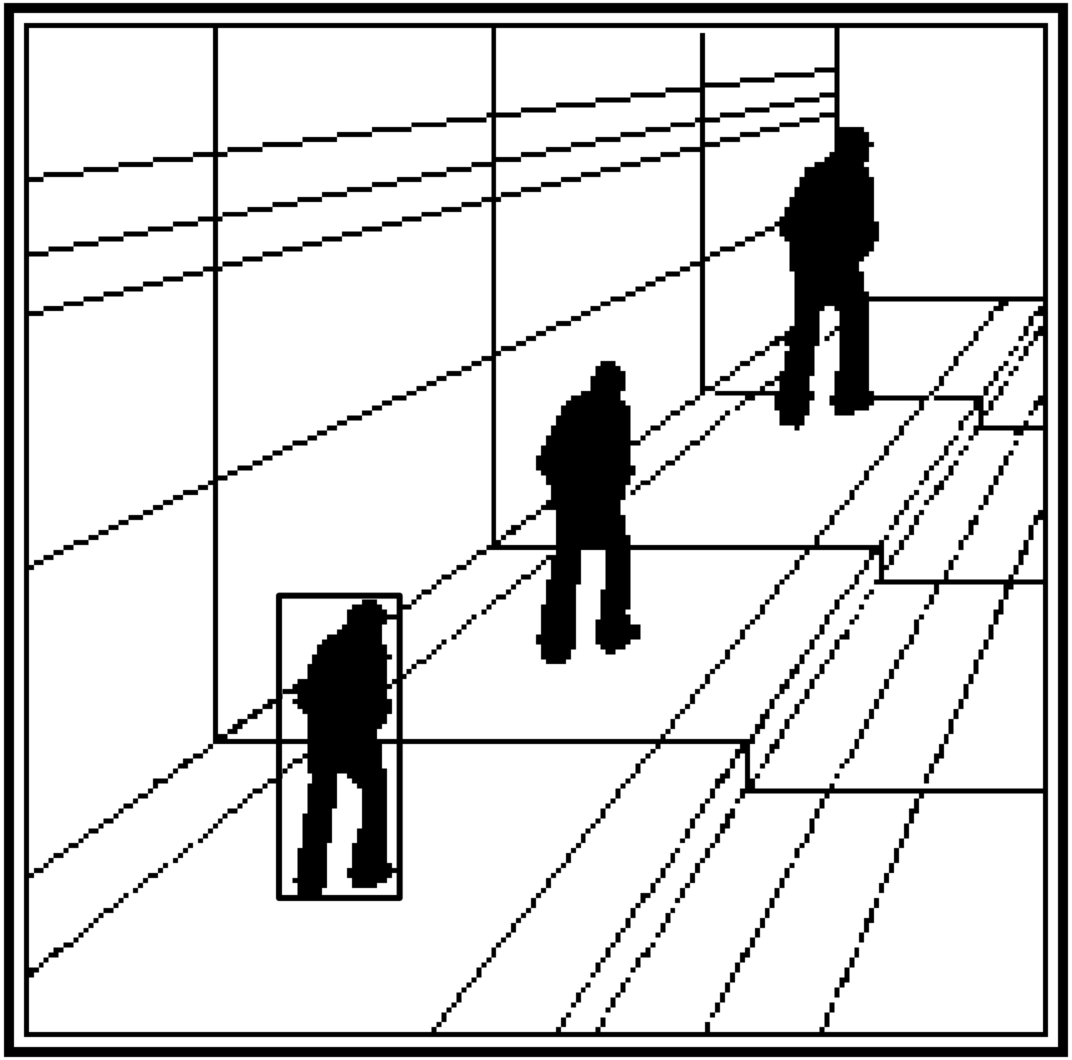

Point at $\lambda$ on line through $X_0\ Y_0\ Z_0$ in direction $U, V, W$ :

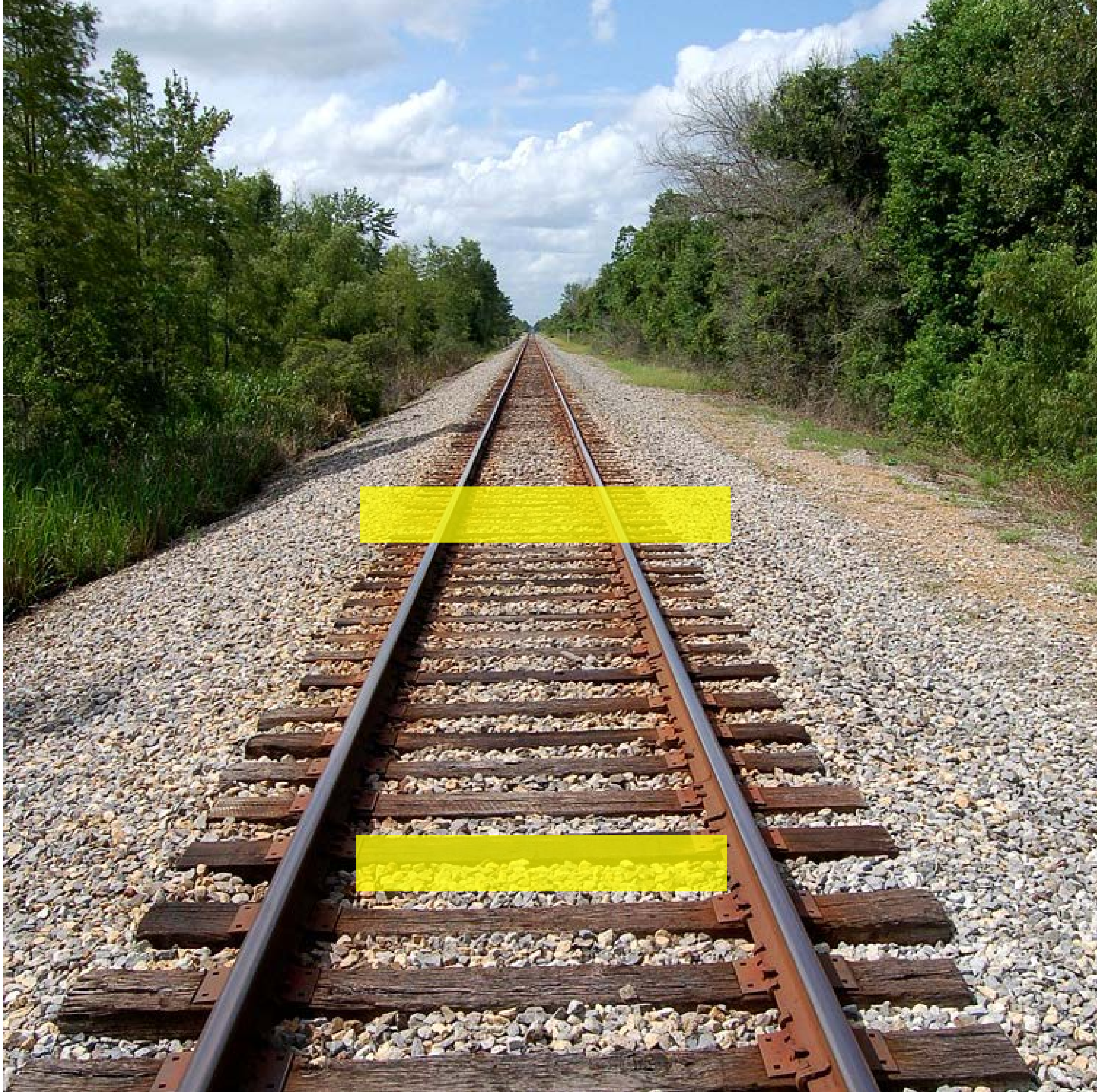
$$X_0 + \lambda U,\ \ Y_0 + \lambda V,\ \ Z_0 + \lambda W$$

Image projection $p$ : $\left( f \dfrac{X_0 + \lambda U}{Z_0 + \lambda W}, f \dfrac{Y_0 + \lambda V}{Z_0 + \lambda W} \right)$

$p$ at $\lambda = \infty$ : $(f\ U/W, f\ V/W)$ : **vanishing point**

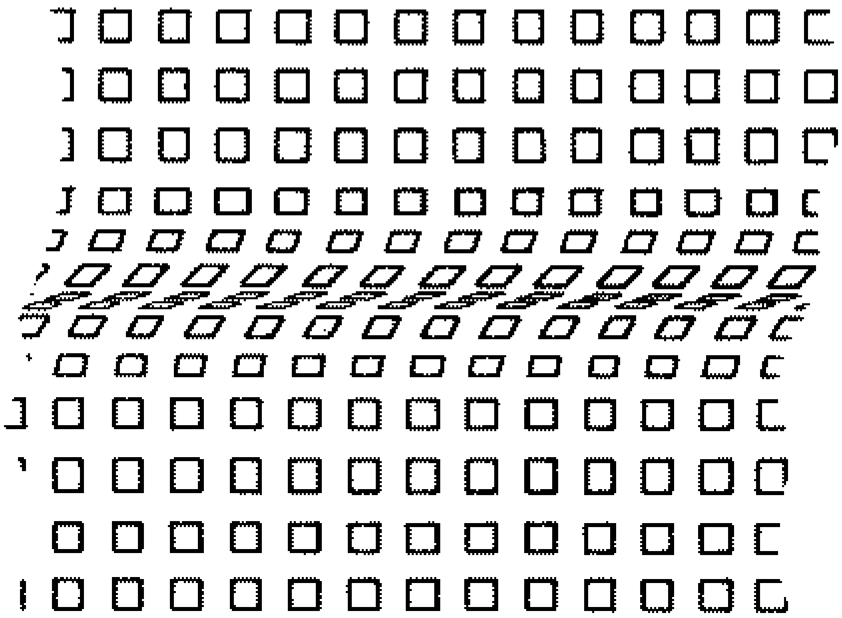
# Depth from ground-plane / horizon

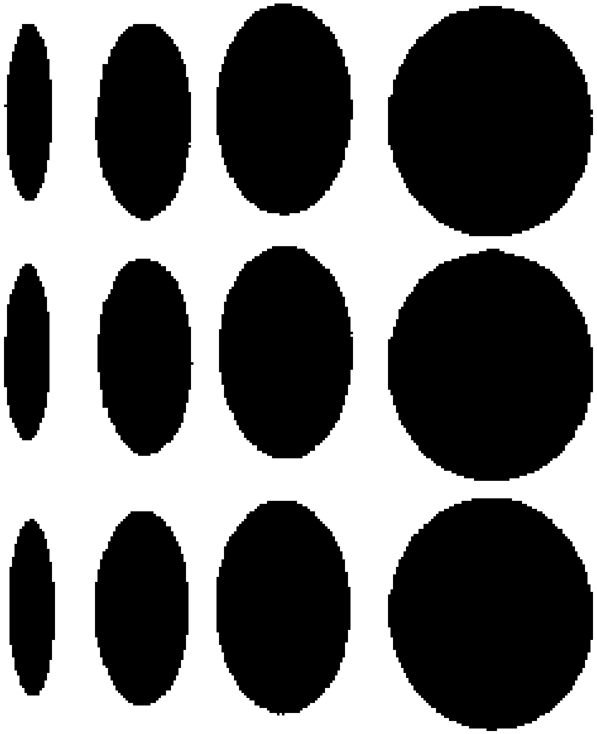




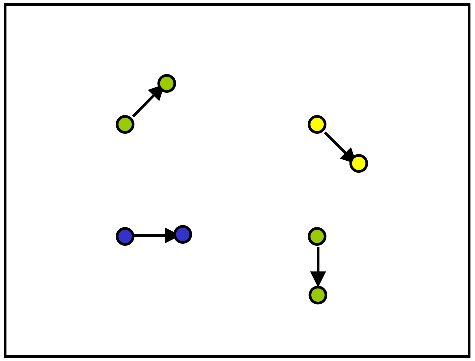
**Ponzo illusion**

# Shape from texture

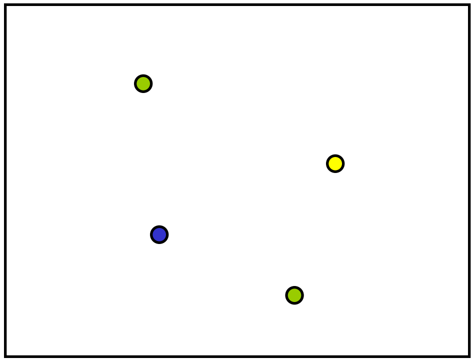




# Problem Definition: Optical flow



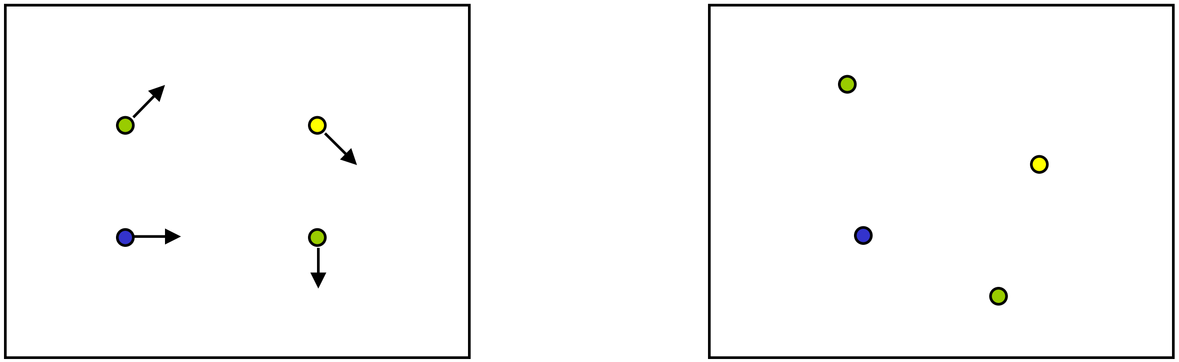
$I_t$



$I_{t+1}$

- How to estimate pixel motion from image $I_t$ to image $I_{t+1}$?
  - Solve pixel correspondence problem
    - given a pixel in $I_t$ look for nearby pixels of same colour in $I_{t+1}$

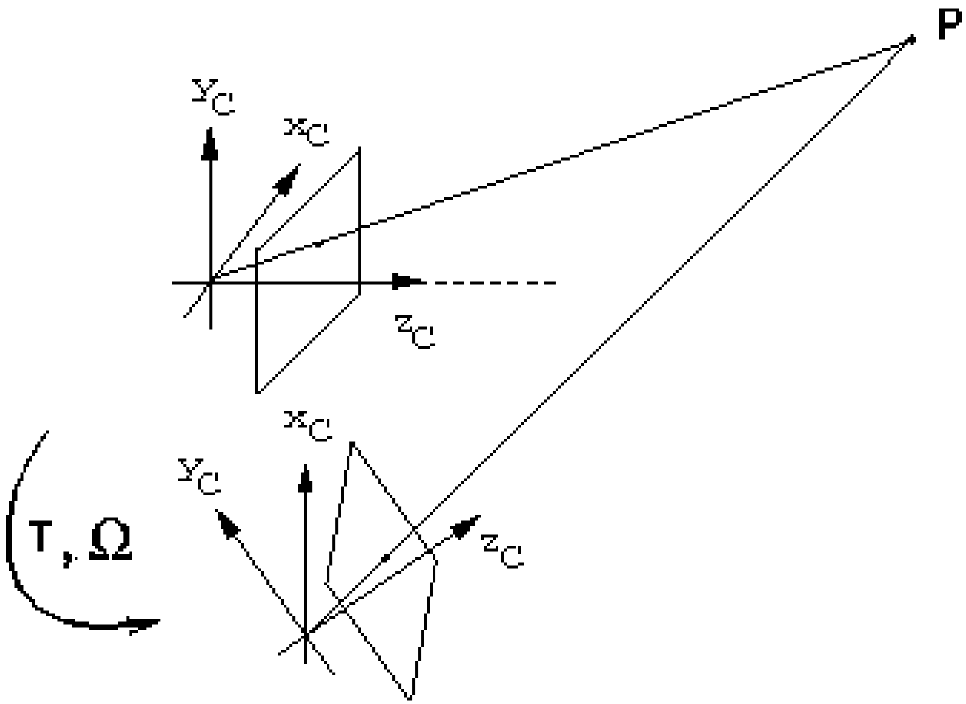
# Problem Definition: Optical flow



$I_t$ $I_{t+1}$

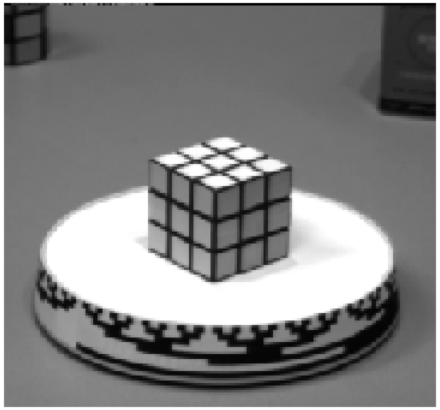
- Key assumptions
  - **color constancy**: pixel in $I_t$ looks the same in $I_{t+1}$
    - grayscale images : *brightness constancy*
  - **small motion**: points do not move very far

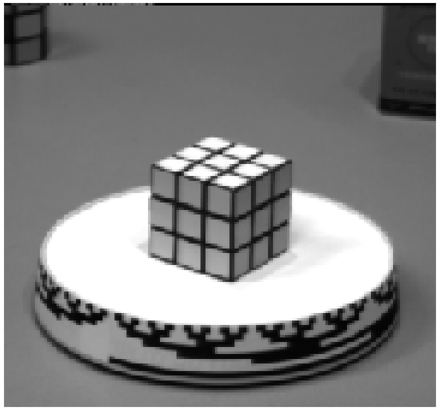
# Optical flow : Camera motion

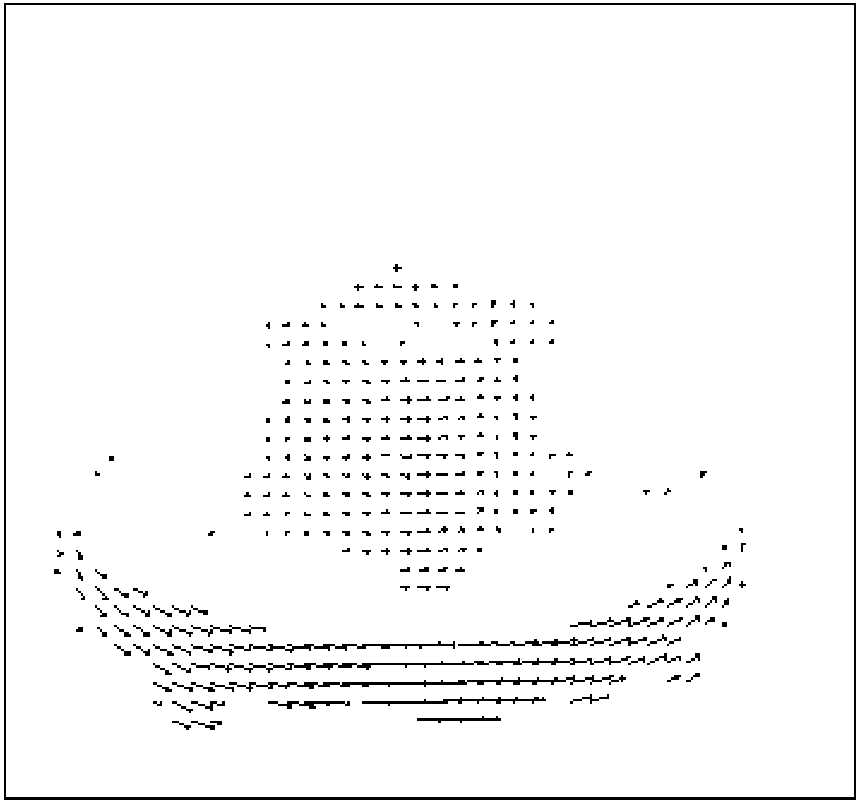



$$
\begin{aligned}
v_1(x, y) &= \left[-\frac{T_1}{Z(x,y)} - \omega_2 + \omega_3 y\right] - x\left[-\frac{T_3}{Z(x,y)} - \omega_1 y + \omega_2 x\right] \\
v_2(x, y) &= \left[-\frac{T_2}{Z(x,y)} + \omega_1 - \omega_3 x\right] - y\left[-\frac{T_3}{Z(x,y)} - \omega_1 y + \omega_2 x\right]
\end{aligned}
$$

# Optical flow : Object motion







Measurement of motion at every pixel

# Image Processing

# How to find correspondences?

- Regions of uniform colour are uninformative
- Edges : intensity change is max in one direction, not the other

  → uncertain along the edge
- Corners : intensity changes about equally in all directions

  → more easy to position

  (Harris corner detector)

# Image intensity change

Change of intensity for shift [*u,v*]:

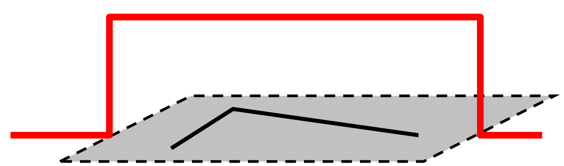
$$E(u,v) = \sum_{x,y} w(x,y)\left[I(x+u,y+v) - I(x,y)\right]^2$$

Window function

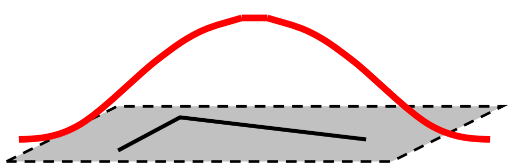
Shifted intensity

Intensity

window function $w(x,y)$ =



1 in window, 0 outside



Gaussian

[slide based on [BK Gunturk]]

# Eigenvalue Decomposition

For small u, v:

$$E(u,v) = \begin{bmatrix} u & v \end{bmatrix} \begin{bmatrix} A & C \\ C & B \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

Matrix written as:

$$M = \sum_{x,y} w(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

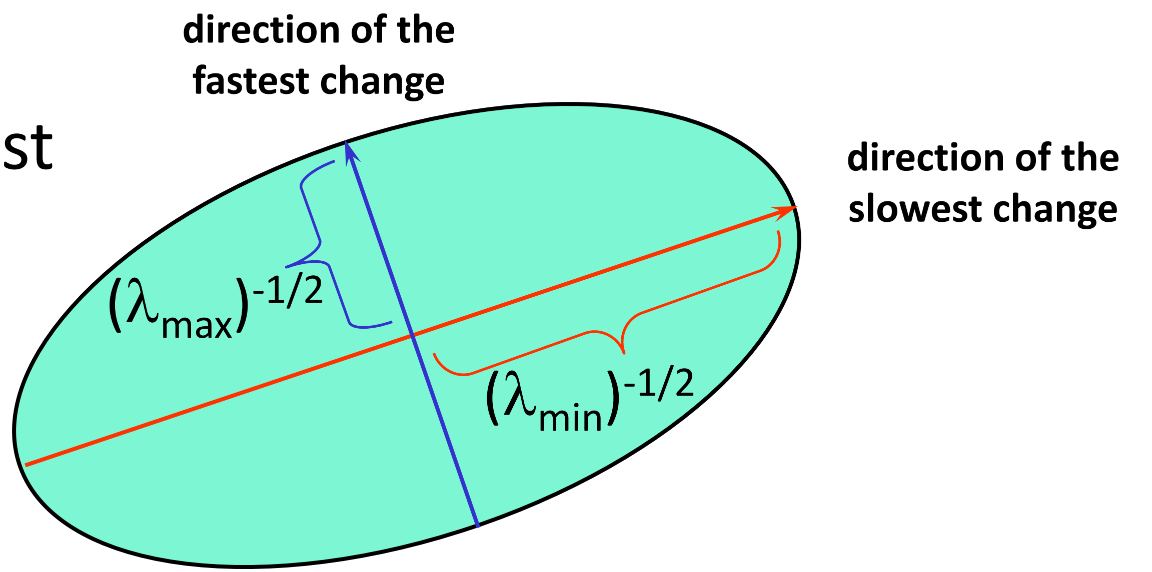
Eigenvalue analysis → directions of change

# Eigenvalue Decomposition

$$E(u,v) \cong [u,v] \; M \begin{bmatrix} u \\ v \end{bmatrix}$$

$\lambda_1, \lambda_2$ : eigenvalues of $M$

Ellipse $E(u,v)$ = const




Corner:

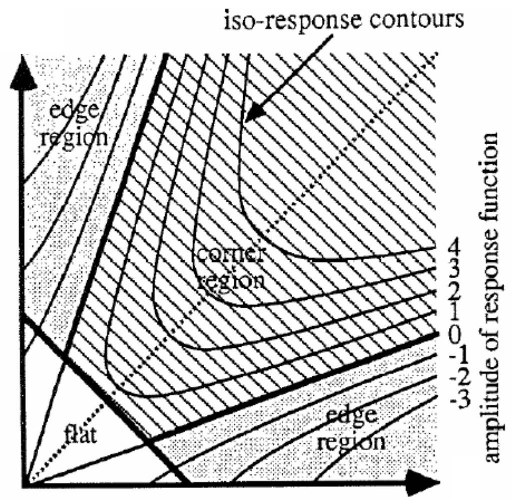
$\lambda_1 \sim= \lambda_2 > 0$

# Harris corner detector

Measure of corner response:

$$R = \det M - k\left(\operatorname{trace} M\right)^2$$

$$\det M = \lambda_1 \lambda_2$$
$$\operatorname{trace} M = \lambda_1 + \lambda_2$$




(*k* – empirical constant, *k* = 0.04-0.06)

Don't need to compute eigenvalues explicitly!

[slide based on [BK Gunturk]]

# Harris corner detector

1. Compute $x$ and $y$ derivatives of image

$$I_x = G_\sigma^x * I \quad I_y = G_\sigma^y * I$$

2. Compute products of derivatives at every pixel

$$I_{x2} = I_x.I_x \quad I_{y2} = I_y.I_y \quad I_{xy} = I_x.I_y$$

3. Compute the sums of the products of derivatives at each pixel

$$S_{x2} = G_{\sigma'} * I_{x2} \quad S_{y2} = G_{\sigma'} * I_{y2} \quad S_{xy} = G_{\sigma'} * I_{xy}$$

4. Define at each pixel $(x, y)$ the matrix

$$H(x,y) = \begin{bmatrix} S_{x2}(x,y) & S_{xy}(x,y) \\ S_{xy}(x,y) & S_{y2}(x,y) \end{bmatrix}$$

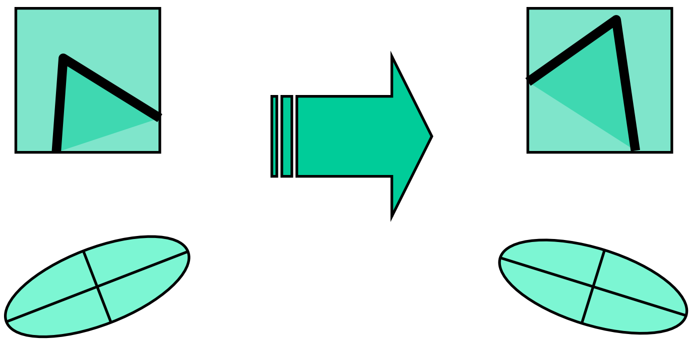
5. Compute the response of the detector at each pixel

$$R = Det(H) - k(Trace(H))^2$$

6. Threshold on value of R. Compute nonmax suppression.

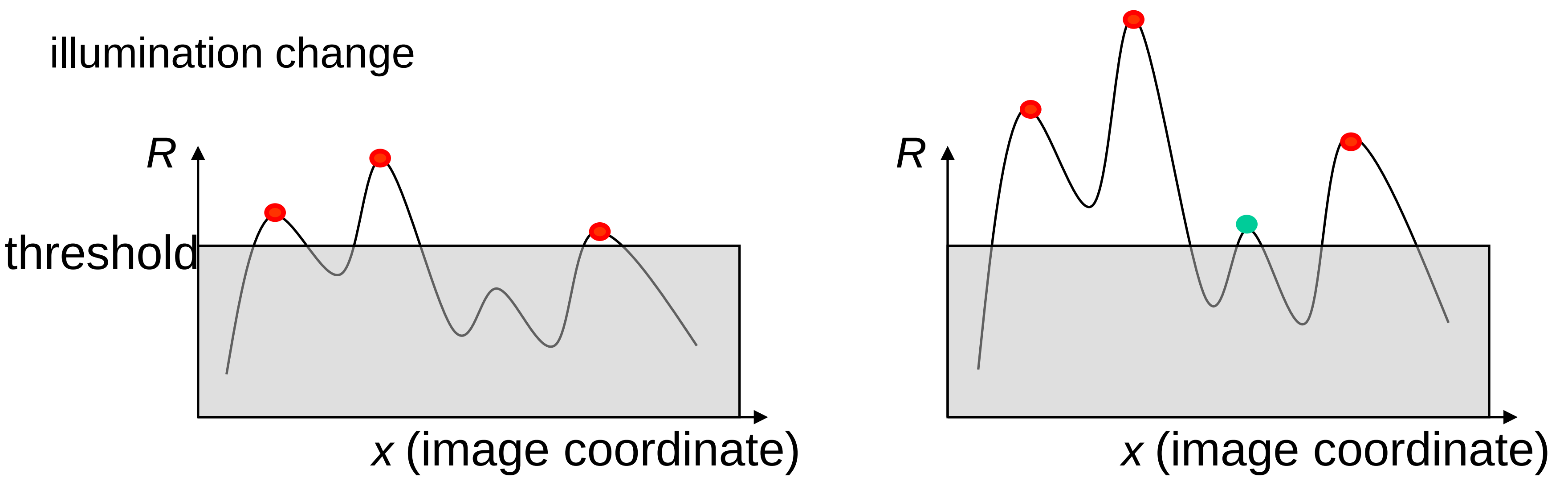
# Properties of Harris detector

- Rotation invariance
  [eigenvalues not affected]



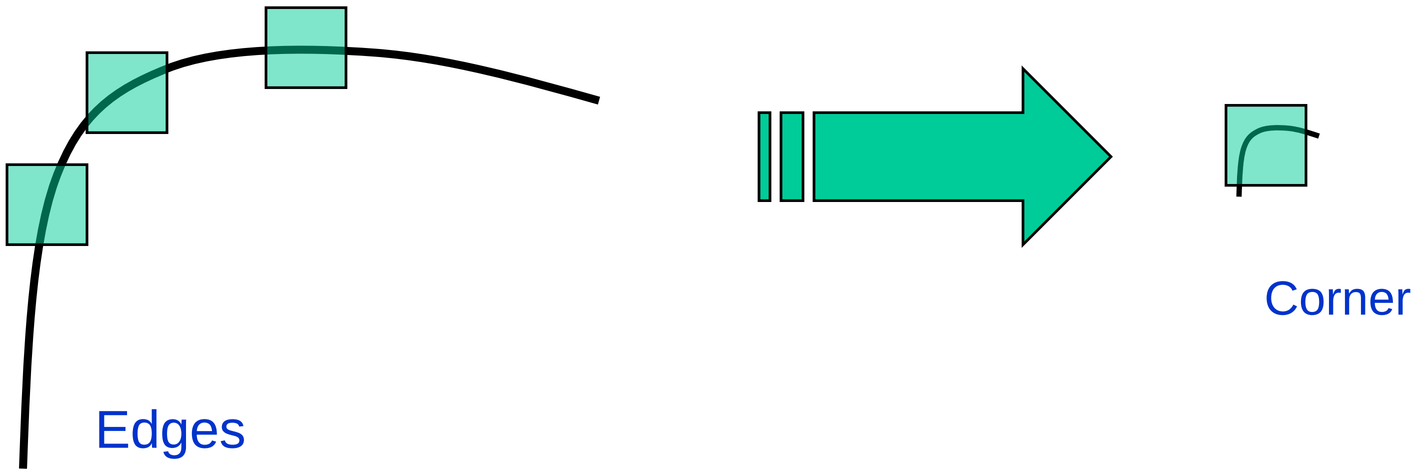
- Partly invariant to affine intensity change

illumination change



# Harris detector : Scale?

But: non-invariant to *image scale*!



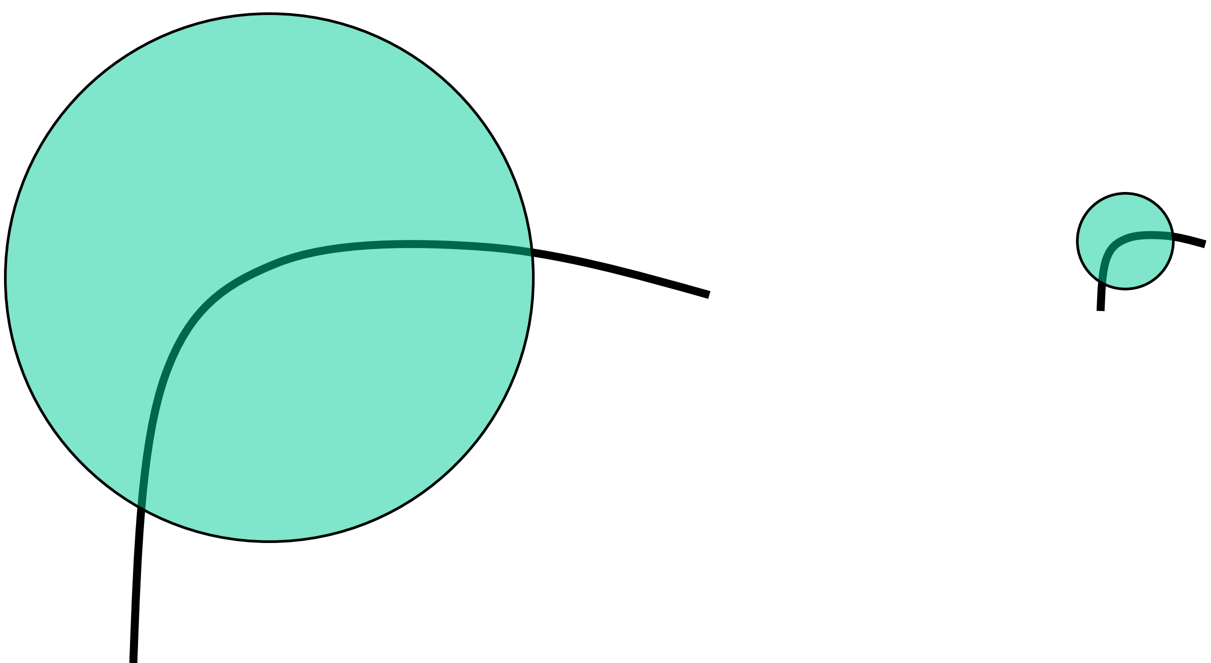

Q. Can we identify scale based on image properties?

# Scale Invariance

Construct neighbourhoods of different sizes

Regions of suitable scale will look same in both images

Identify scale at whiich DoG is maximal

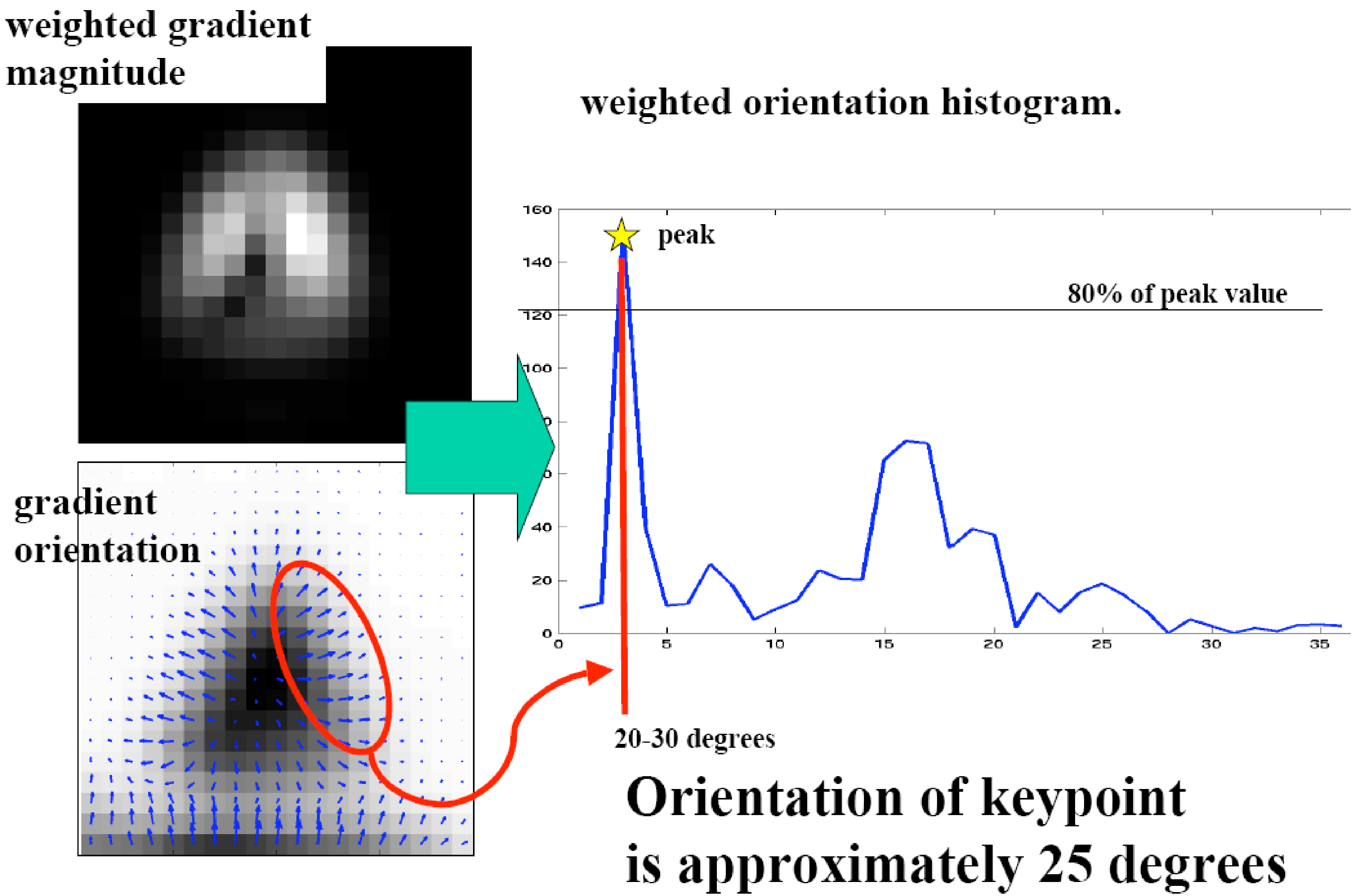


# Rotation Invariance

- All computations at scale with maximal DoG
- Find local gradients
- Find unique peak of orientation histogram

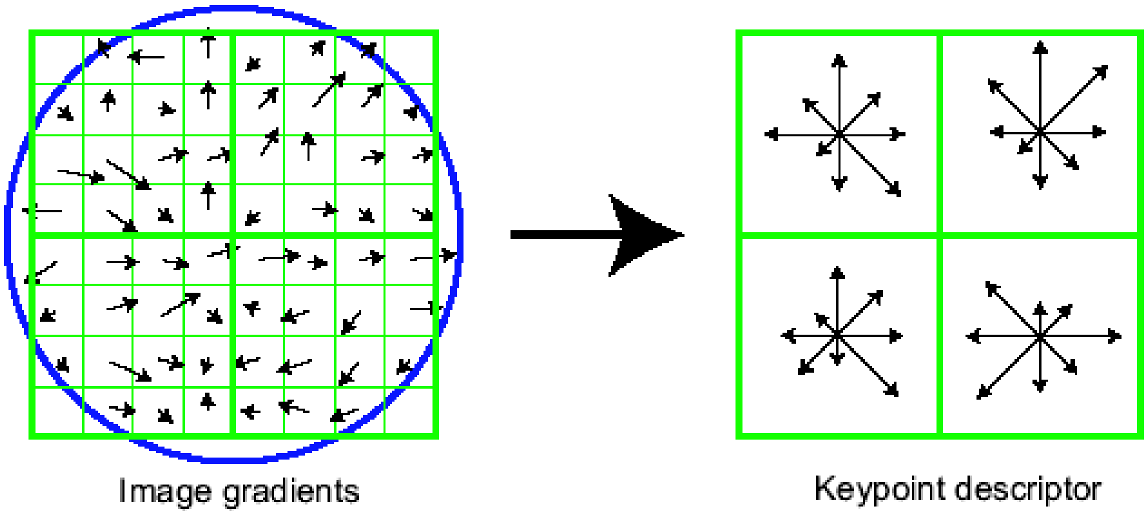
→ Describe in terms of local orientation

# Rotation Invariance



# SIFT: Scale-Independent Feature Transform

1. find “interest points”  where DoG has unique maximum
2. use scale at which DoG is maximum as scale for SIFT
3. local orientation = dominant gradient direction.
4. Construct grid at point using this scale and orientation; → invariant to scale and rotation.
5. Compute gradient orientation 8-histograms at each cell (=128-vector)



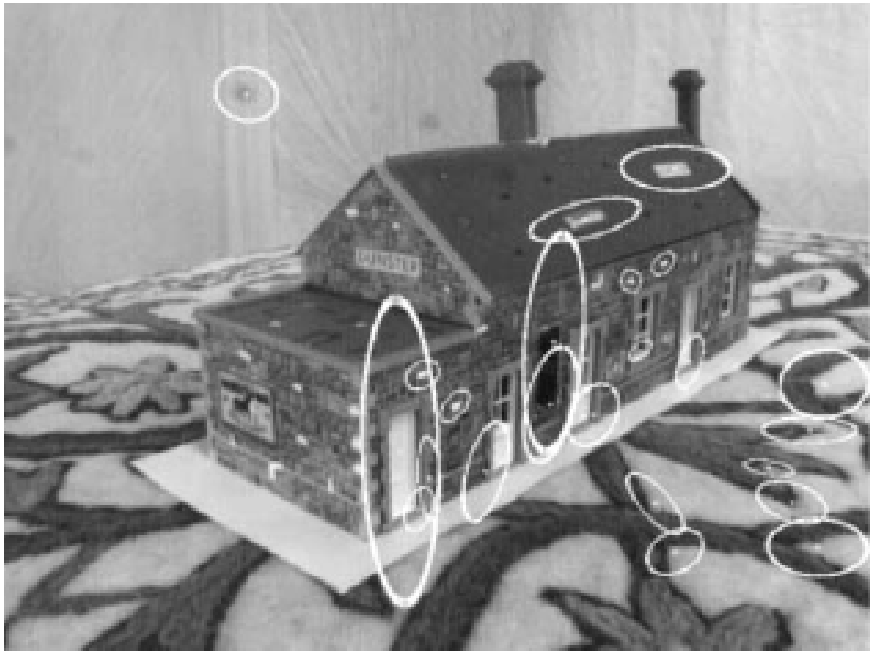
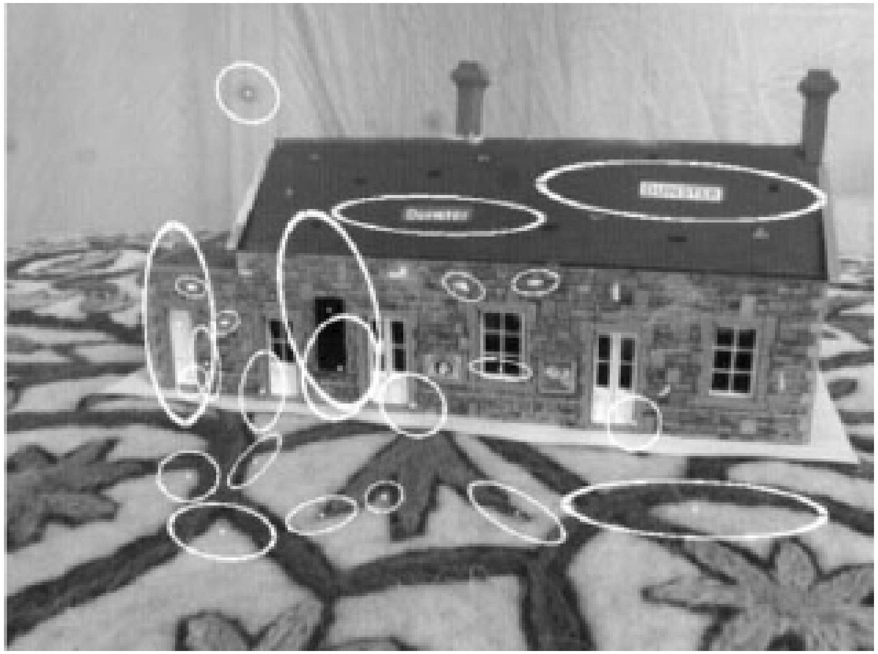

[Lowe 99, 04]

# SIFT Applications:  Correspondences

- Correspondence: Find matching locations in images from differing viewpoints
    - used for [stereo vision, optical flow]

Image: [Tuytelaars ECCV 06 tutorial]

# Robustness

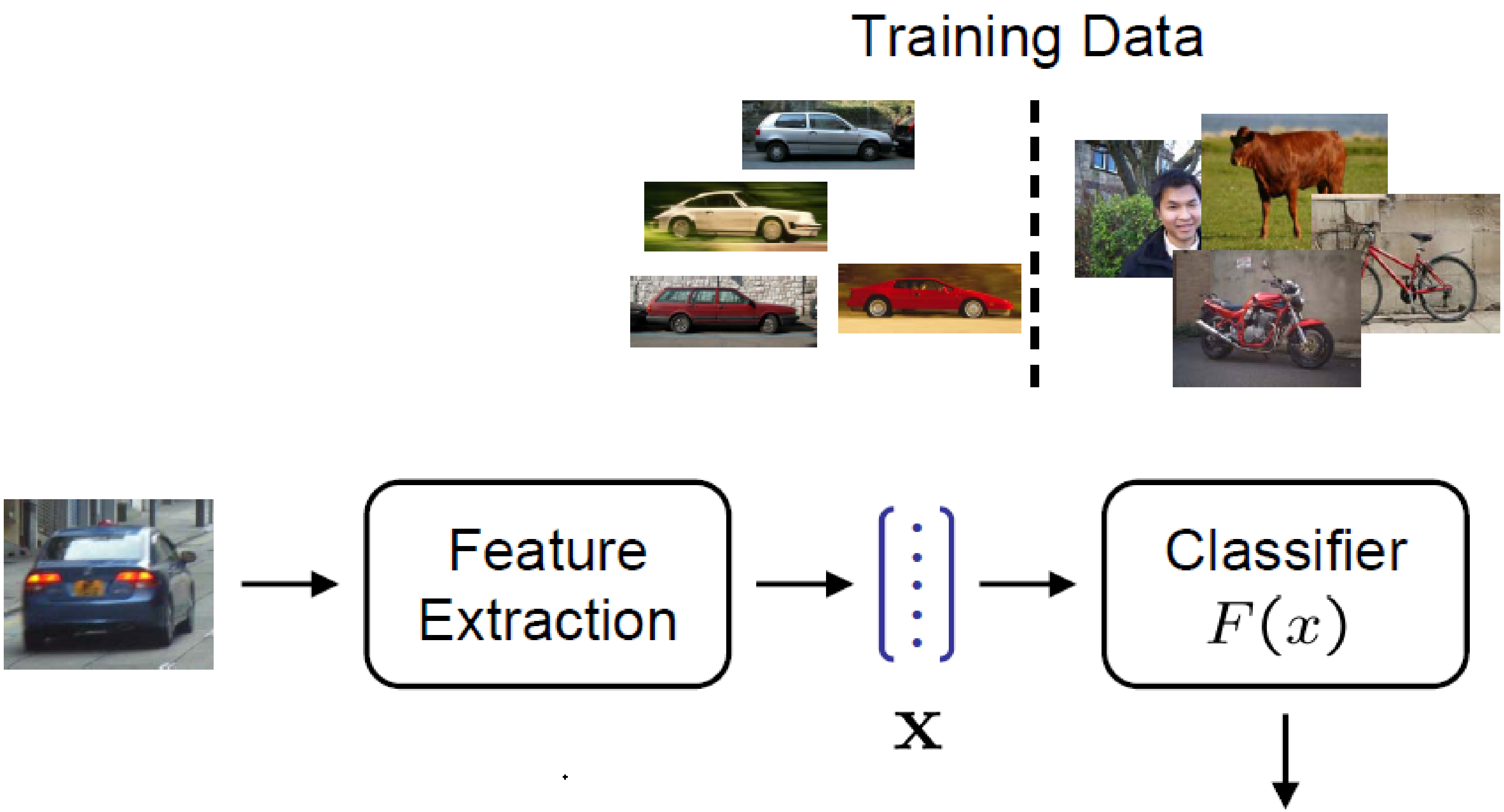
 

- Useful for finding correspondences
- Also for defining “signatures” for object categories→ bag of words approach

# Object Recognition

# Object Recognition

# Training Classifiers




Car/Non-car

$$P(c|\mathbf{x}) \propto F(\mathbf{x})$$

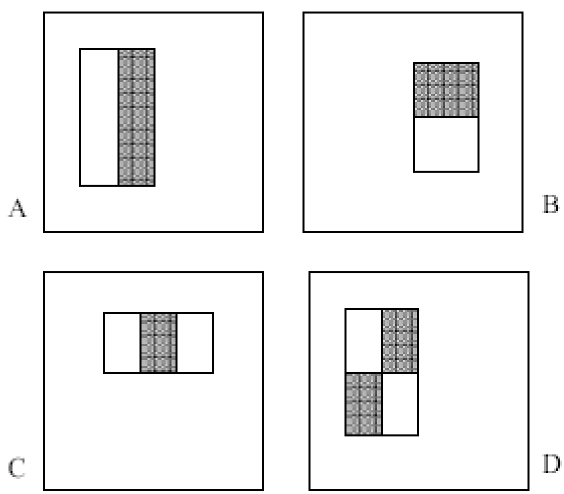
- Which features to use?
- How to design F()?

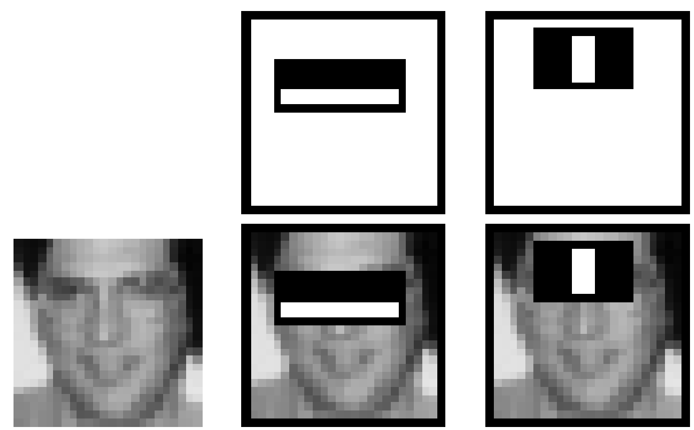
# Viola-Jones Face Detection

Fast detection, very slow training

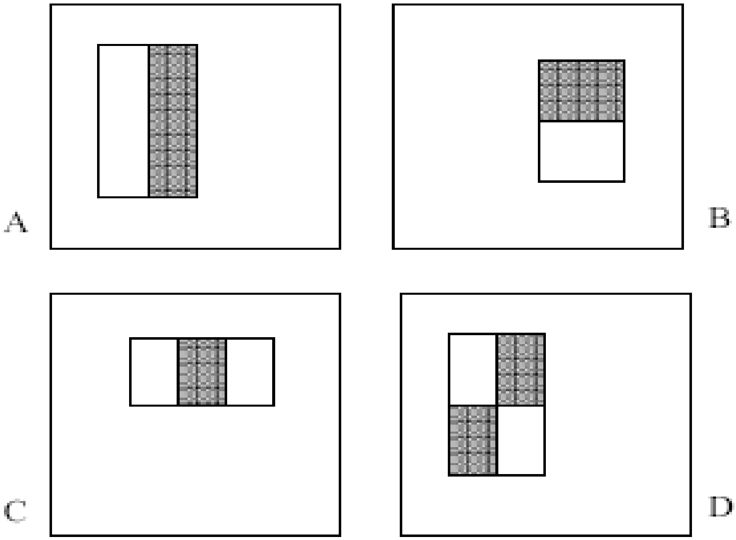
5000 face images, 3 million non-faces

Haar features - Four types.



# Integral Image




- Haar features : can compute in four integral data
  - Compute on sliding window
  - Large feature space -180K features
- classifier F: cascade Ada Boost

# Learning with many features

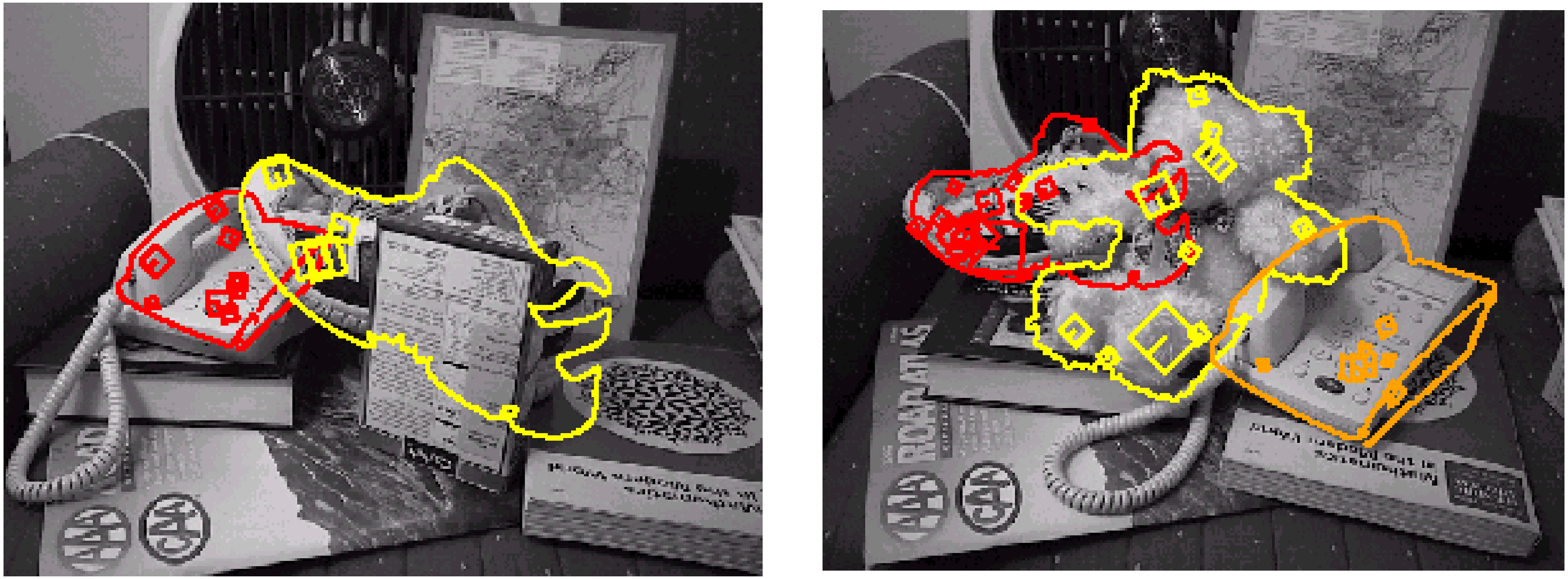
160,000 features : 5000 training data

Boosting:

- Learn a single simple classifier
- Look at where it makes errors
- Reweight the data so inputs with errors get higher weight in the learning process
- Learn $2^{nd}$ simple classifier on weighted data
- Combine $1^{st}$ and $2^{nd}$ classifier; weight the data accordingly
- Learn a $k^{th}$ classifier on the weighted data

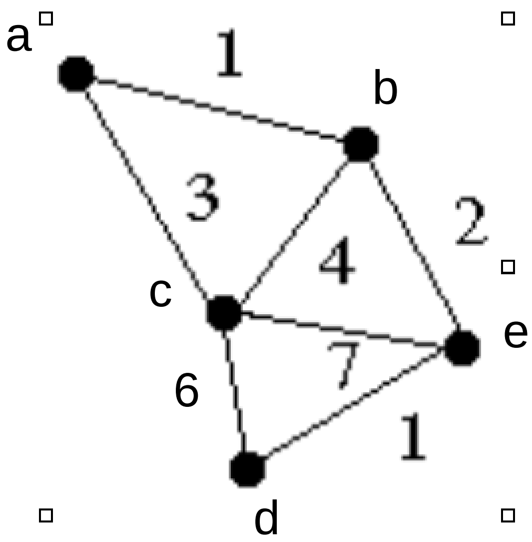
Final classifier : combination of all k classifiers

# Recognition under occlusion



# Segmentation

# Weighted Graphs and Their Representations



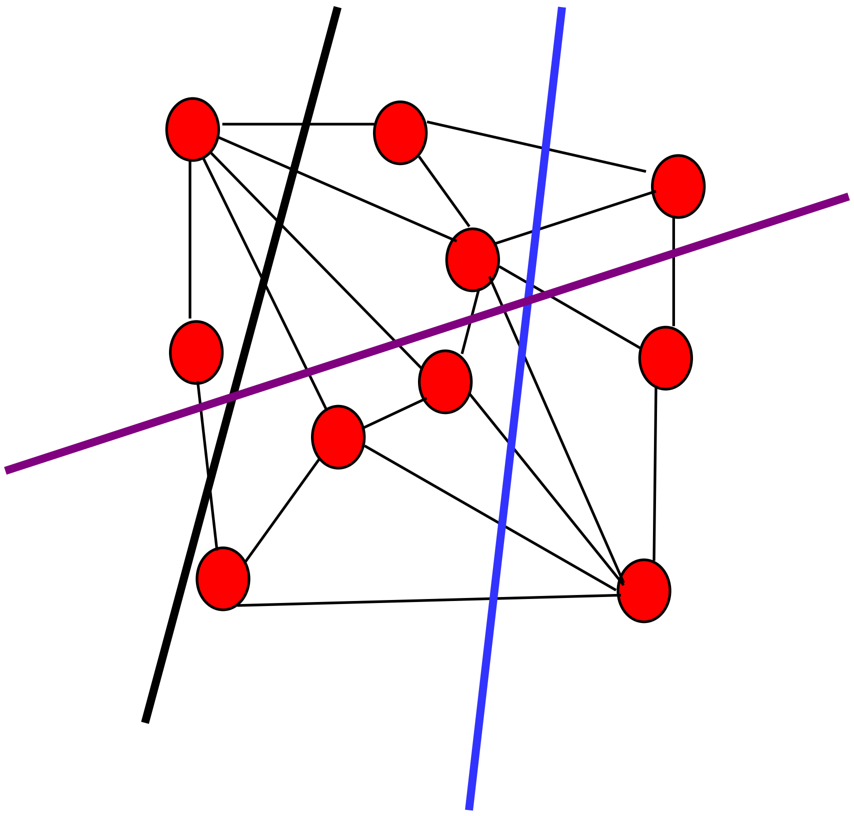

$$
\begin{bmatrix}
0 & 1 & 3 & \infty & \infty \\
1 & 0 & 4 & \infty & 2 \\
3 & 4 & 0 & 6 & 7 \\
\infty & \infty & 6 & 0 & 1 \\
\infty & 2 & 7 & 1 & 0
\end{bmatrix}
$$

Weight Matrix: W

# Superpixel Segmentation


[felzenszwalb huttenlocher 04]

# Minimum Cut

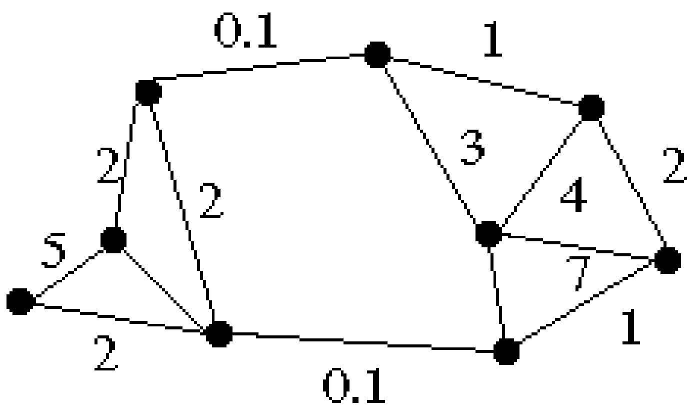


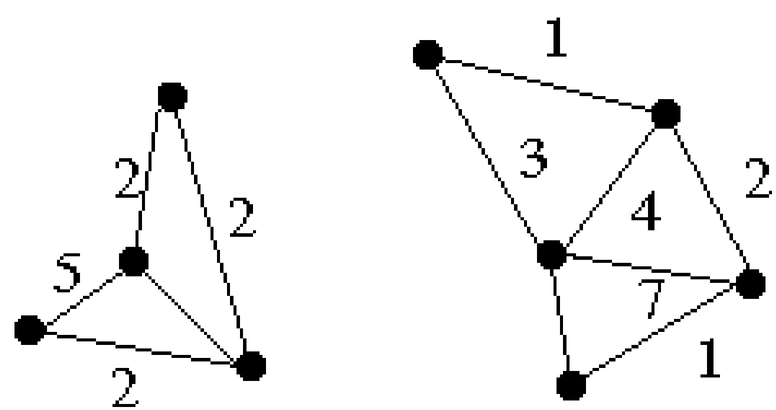
A cut of a graph $G$ is the set of edges $S$ such that removal of $S$ from $G$ disconnects $G$.

Minimum cut is the cut of minimum weight, where weight of cut <A,B> is given as

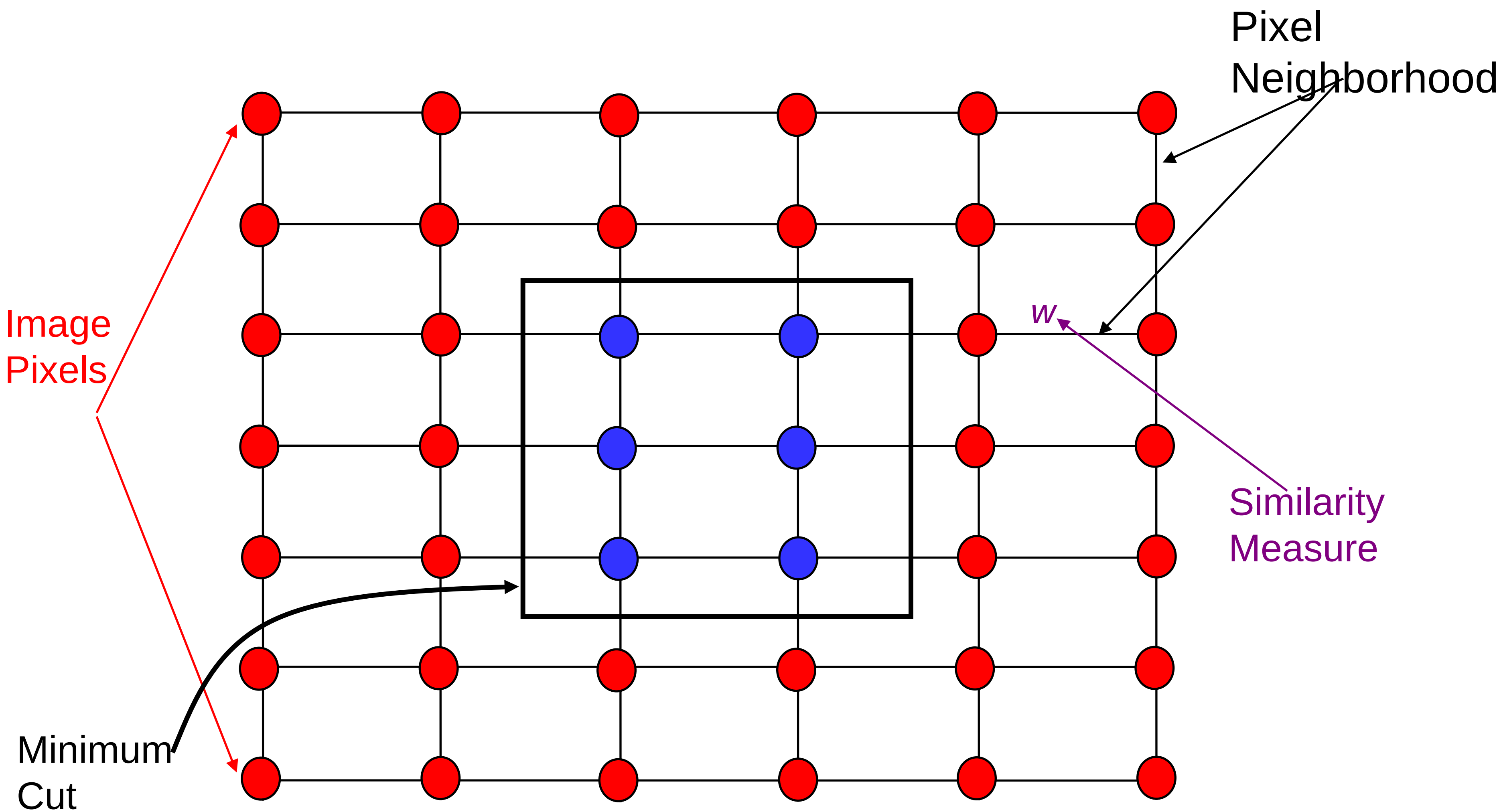
$$w(\langle A, B \rangle) = \sum_{x \in A, y \in B} w(x, y)$$

# Minimum Cut and Clustering



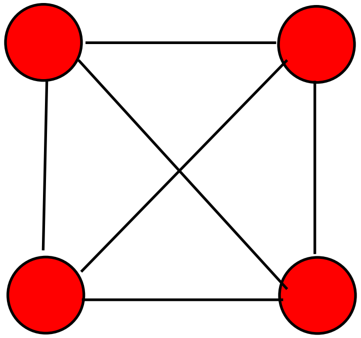



# Image Segmentation & Minimum Cut



# Minimum Cut

- There can be more than one minimum cut in a given graph

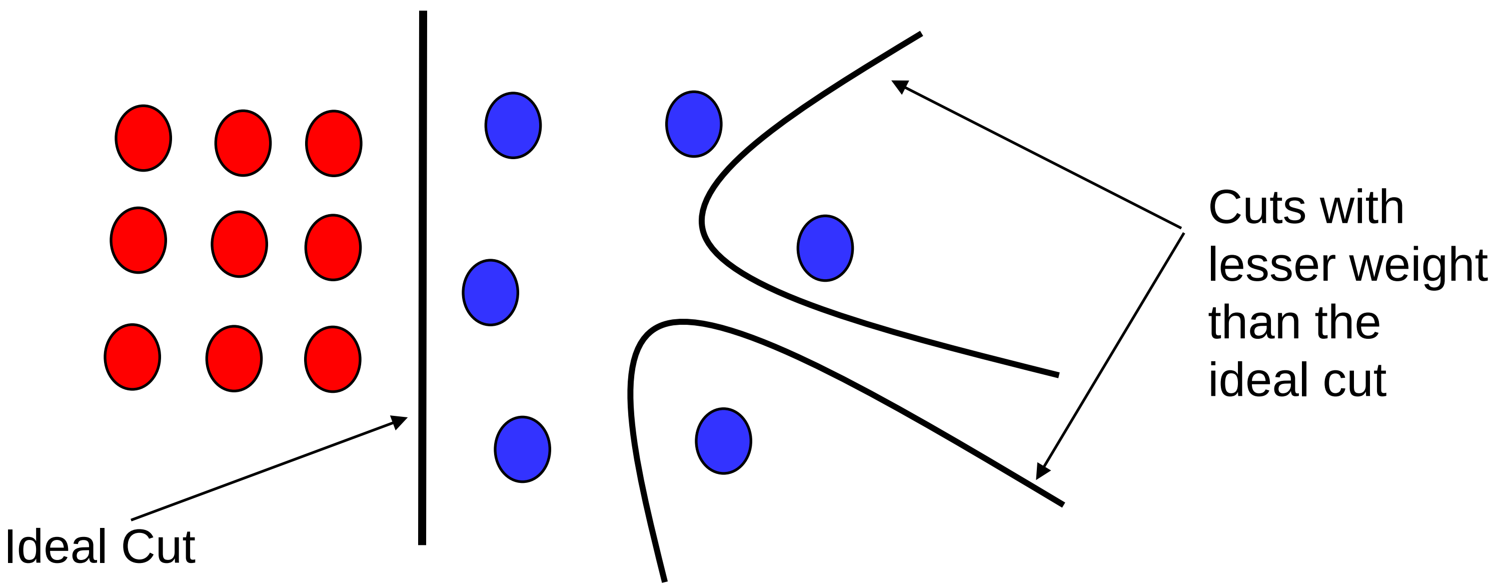


- All minimum cuts of a graph can be found in polynomial time[1].

[1] H. Nagamochi, K. Nishimura and T. Ibaraki, “Computing all small cuts in an undirected network. SIAM J. Discrete Math. 10 (1997) 469-481.

# Drawbacks of Minimum Cut

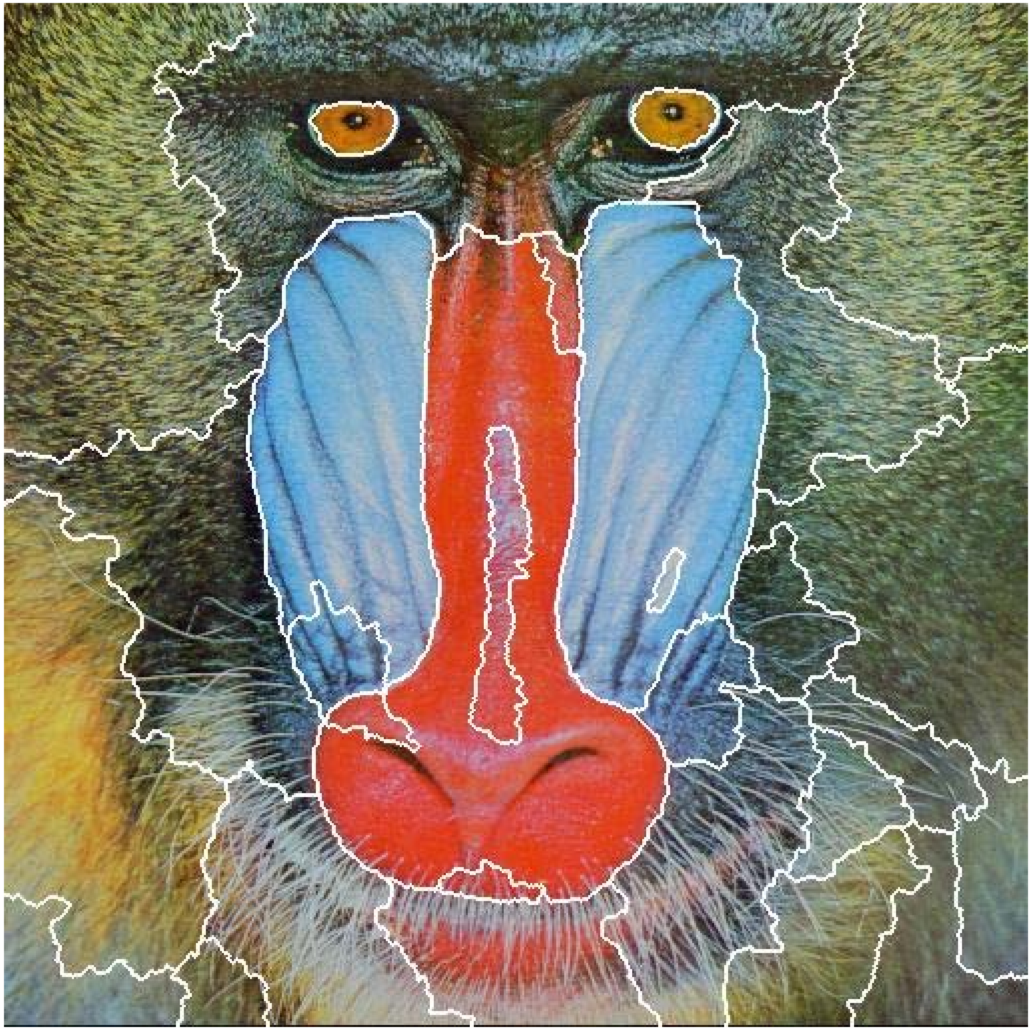
- Weight of cut is directly proportional to the number of edges in the cut.



# Image Segmentation

JSEG algorithm
[Deng & Manjunath 01]

Source code:



# Contour models

r

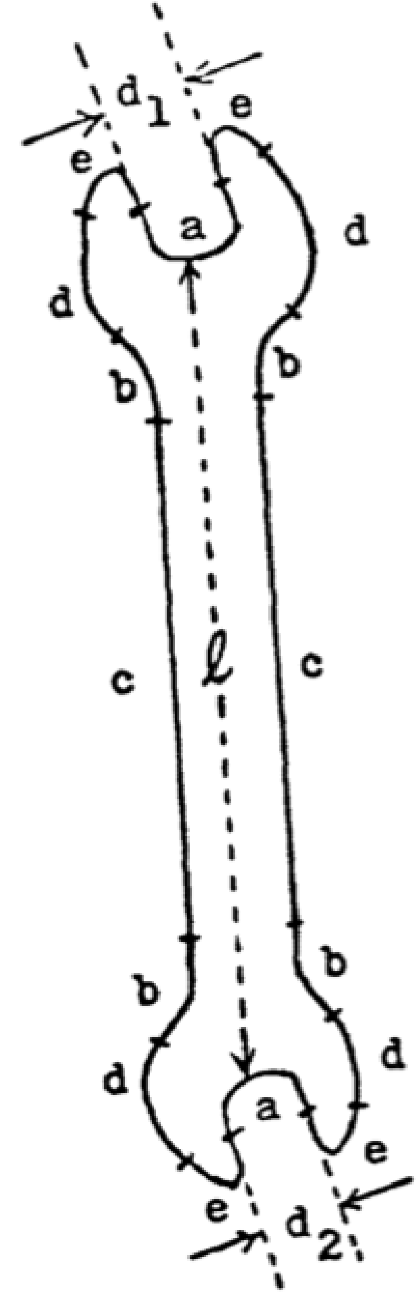
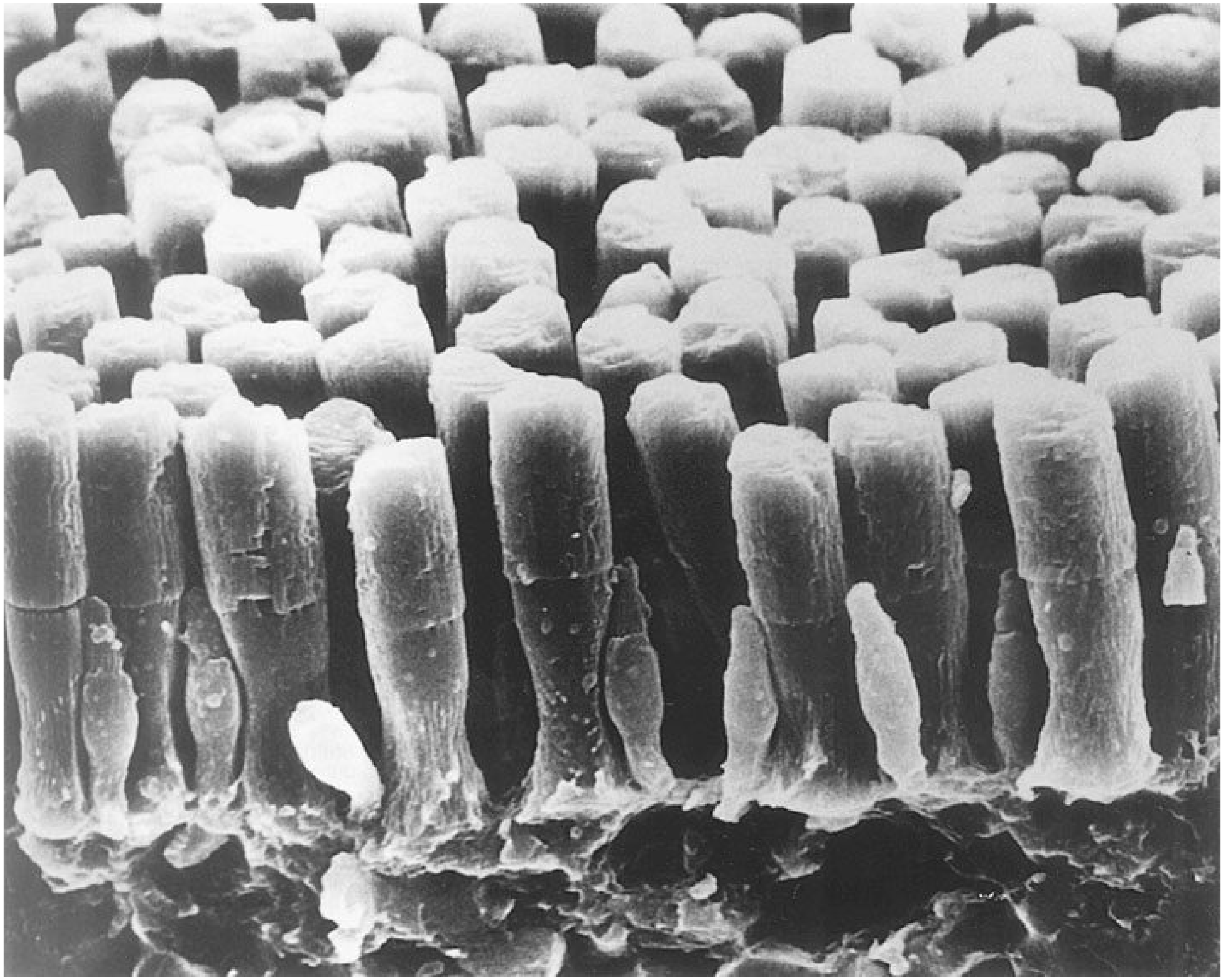


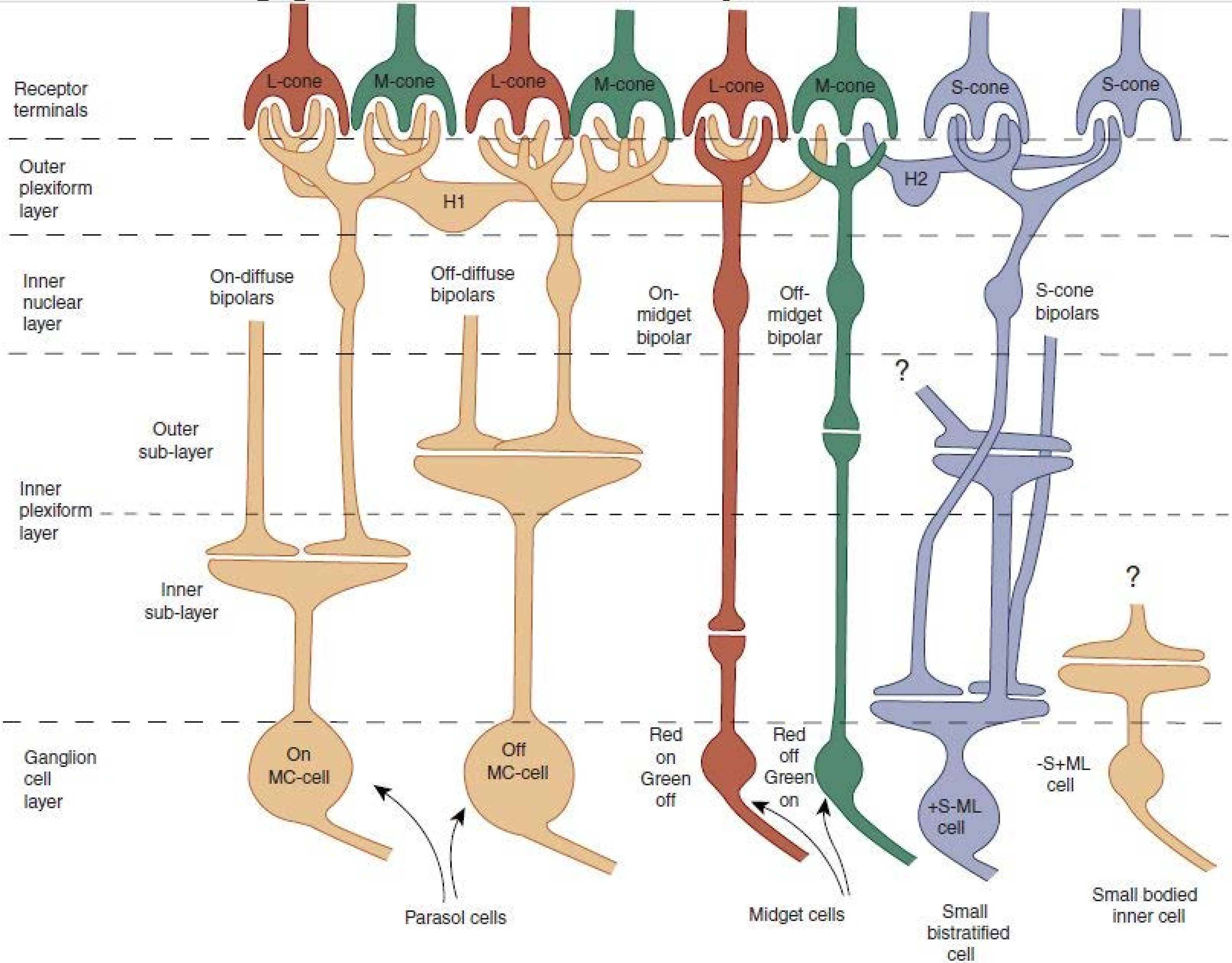

Fig. 8. I wrench and its boundary primitives.

# Human Vision

# Human vision

Rods
and
Cones
(salamander)





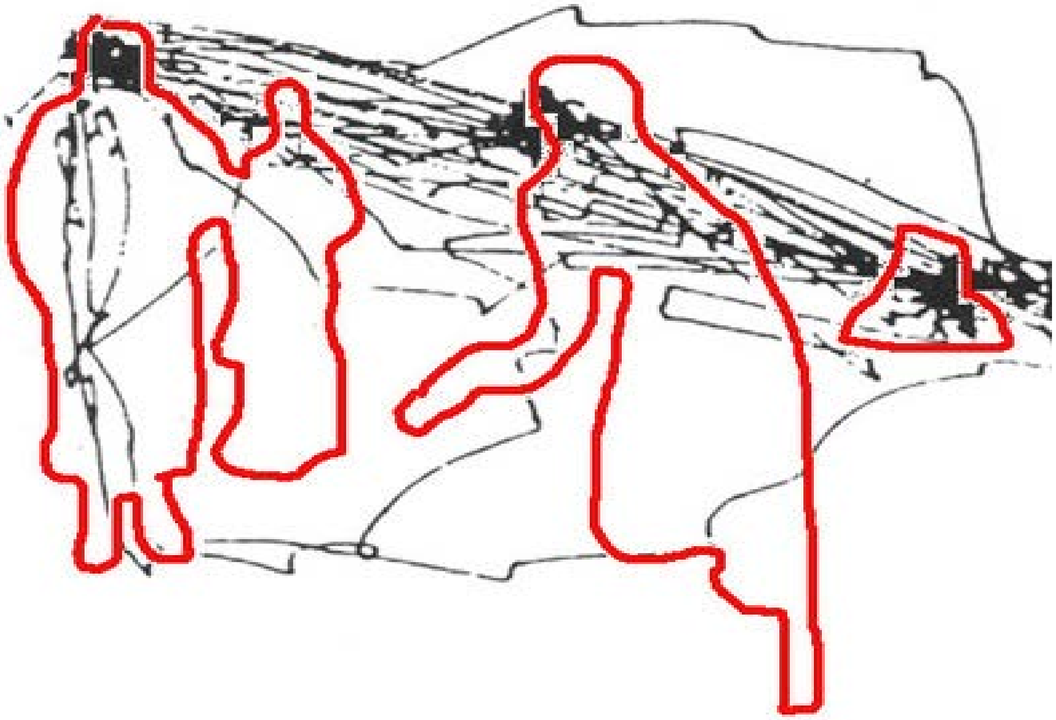
L-cone
M-cone
L-cone
M-cone
L-cone
M-cone
S-cone
S-cone
Receptor terminals
Outer plexiform layer
H1
H2
Inner nuclear layer
On-diffuse bipolars
Off-diffuse bipolars
On-midget bipolar
Off-midget bipolar
S-cone bipolars
?
Outer sub-layer
Inner plexiform layer
Inner sub-layer
?
Ganglion cell layer
On MC-cell
Off MC-cell
Red on Green off
Red off Green on
+S-ML cell
-S+ML cell
Parasol cells
Midget cells
Small bistratified cell
Small bodied inner cell

# Vision and the Task

Eye movement data from [Yarbus 1967]

*The unexpected visitor* by Ilya Repin

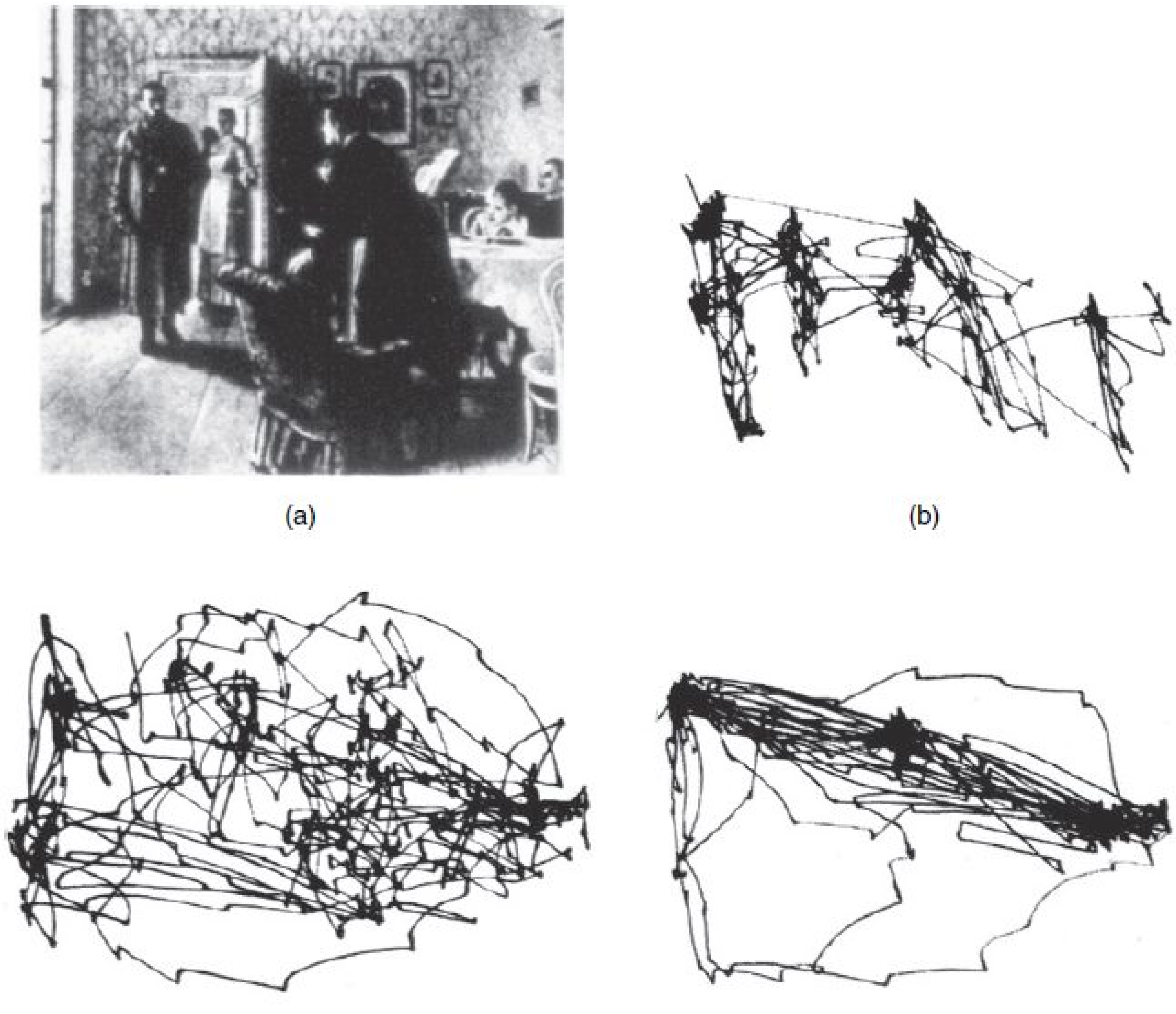
# Eye Movements and Information



**Task**:

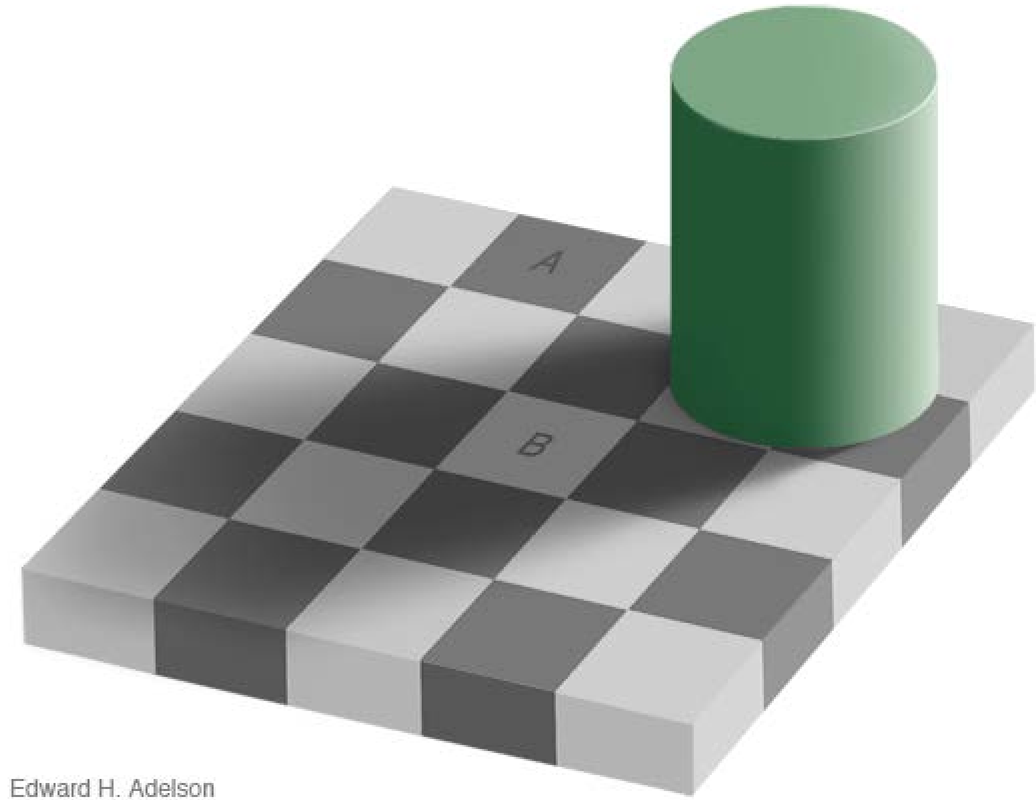
how long has the visitor been away?

# Attention: Top-Down

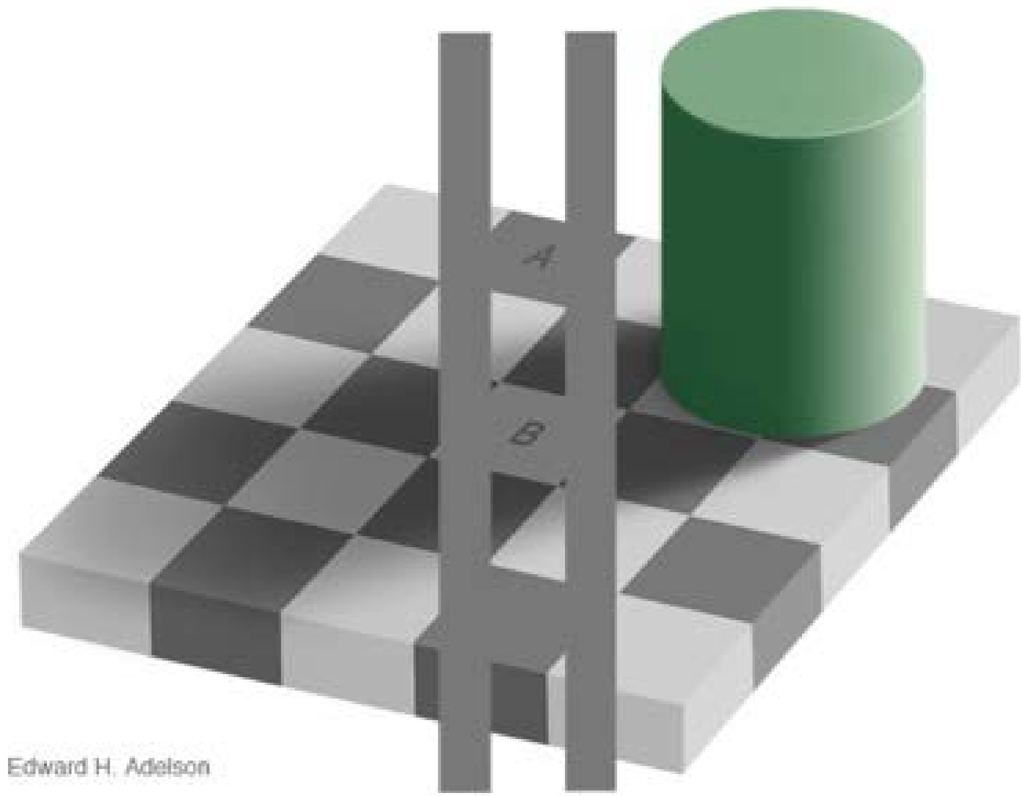

(a) (b) (c) (d)

[Yarbus 6
Ilya Repi

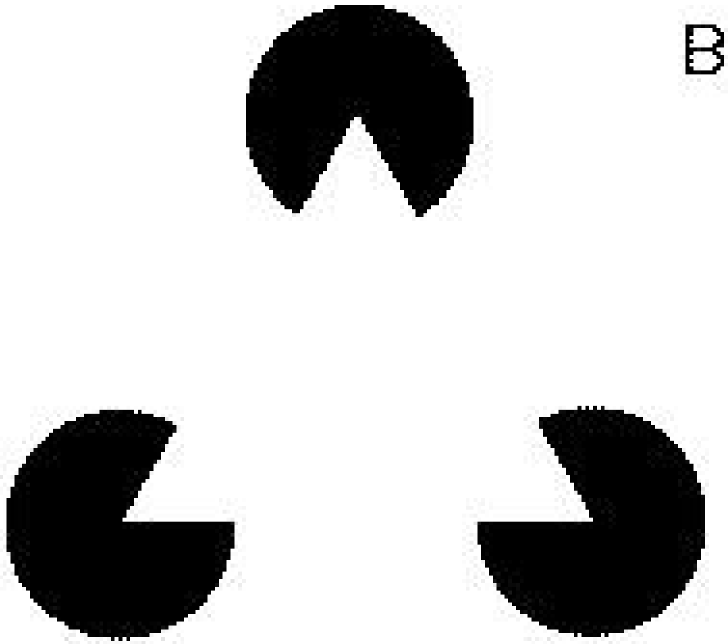
# Colours

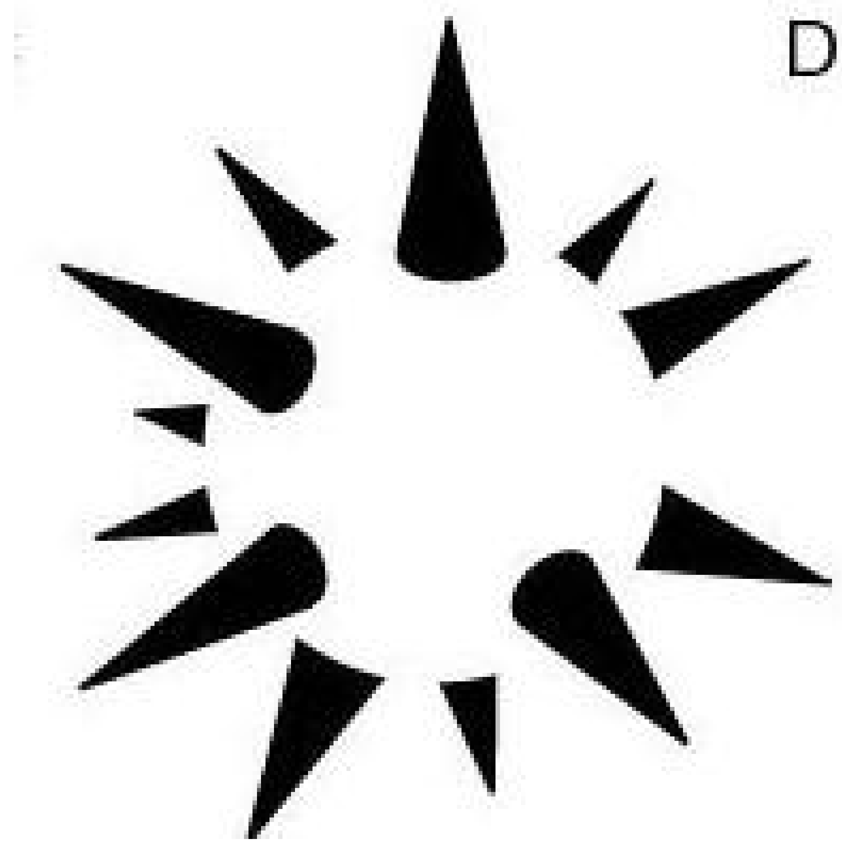


# Colours

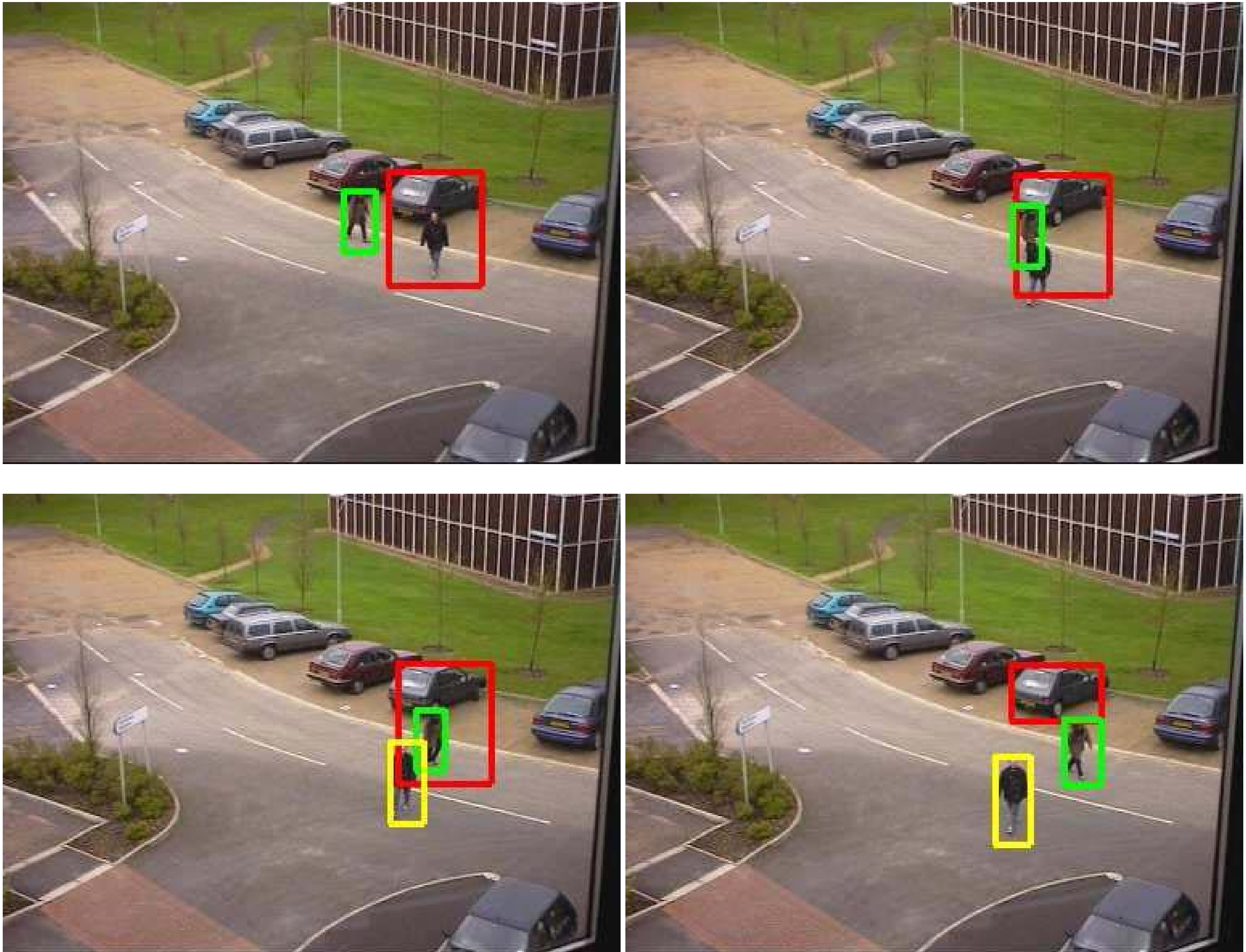


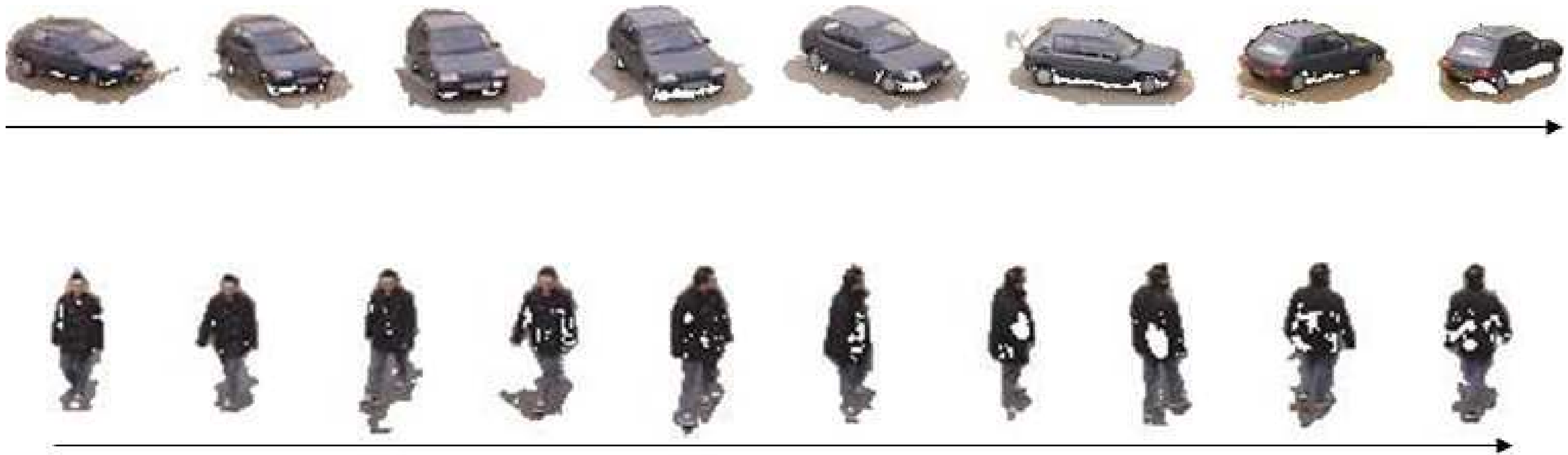
# Reification

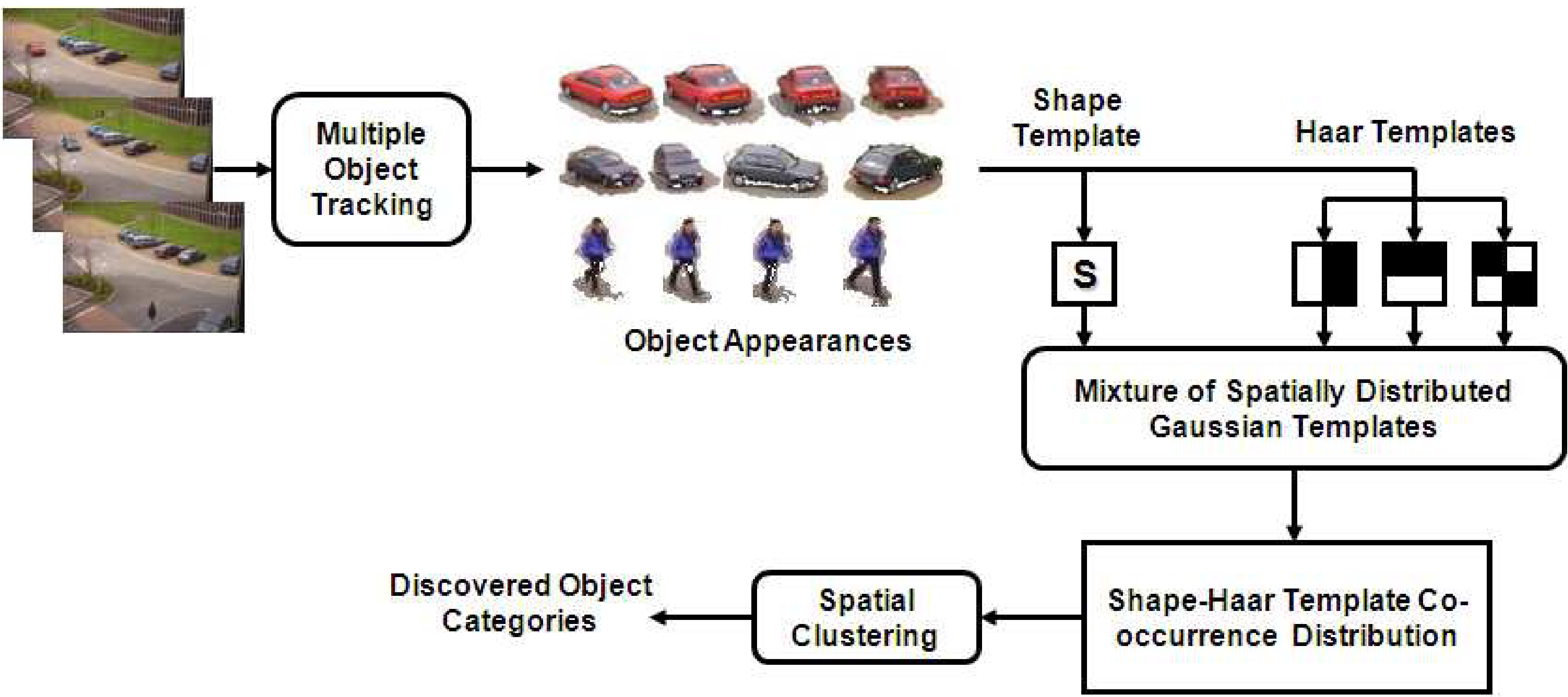


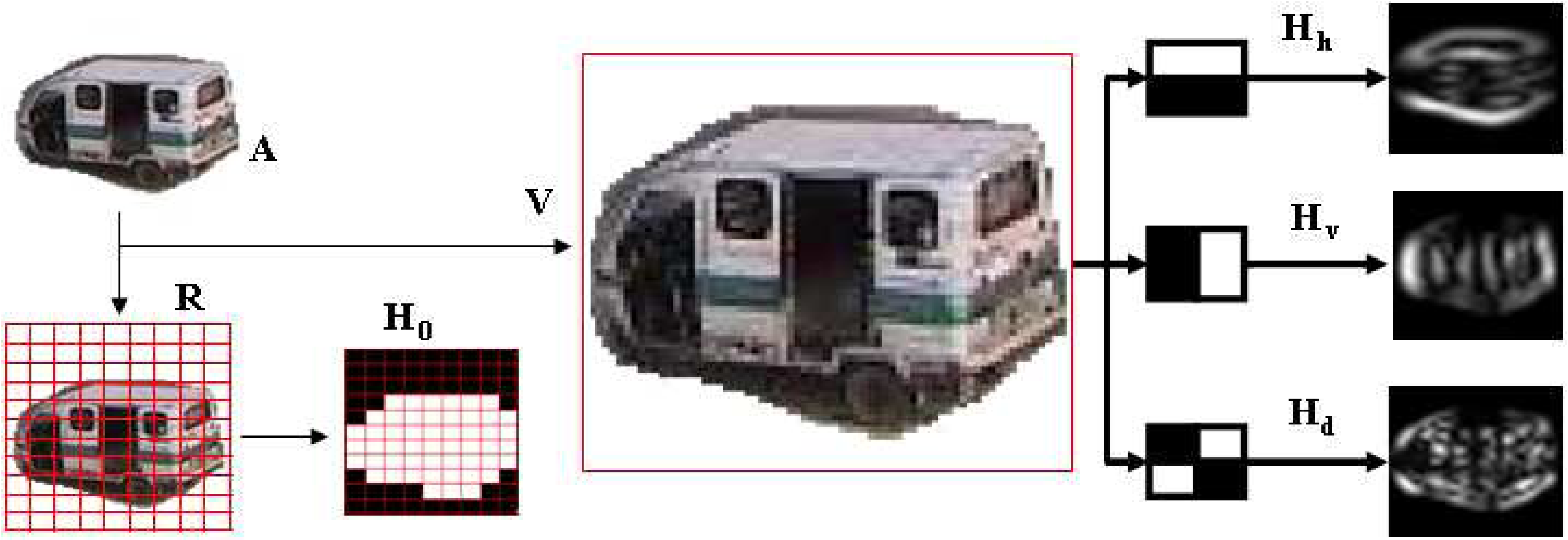
# Unsupervised Modeling and Mapping to Language



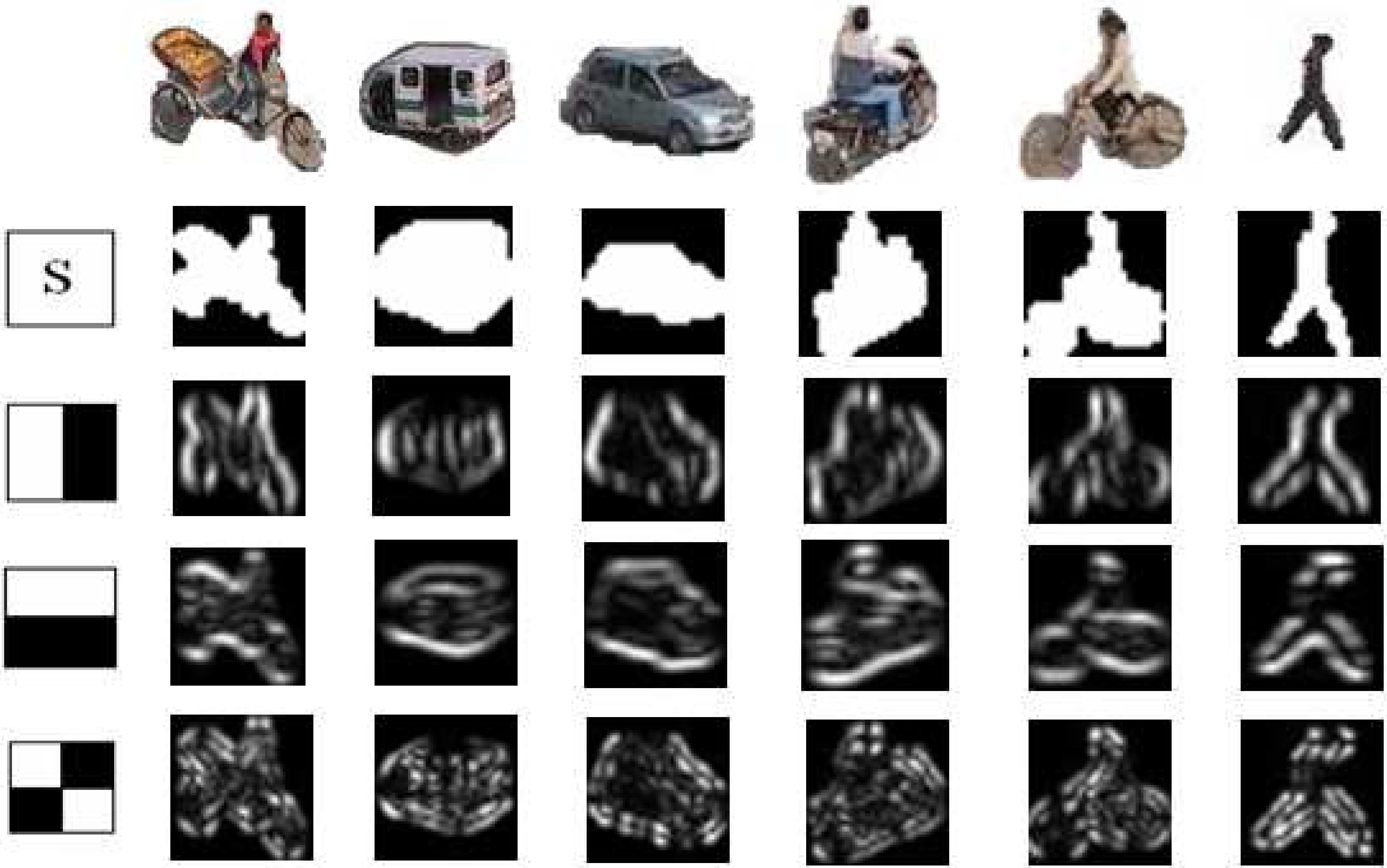




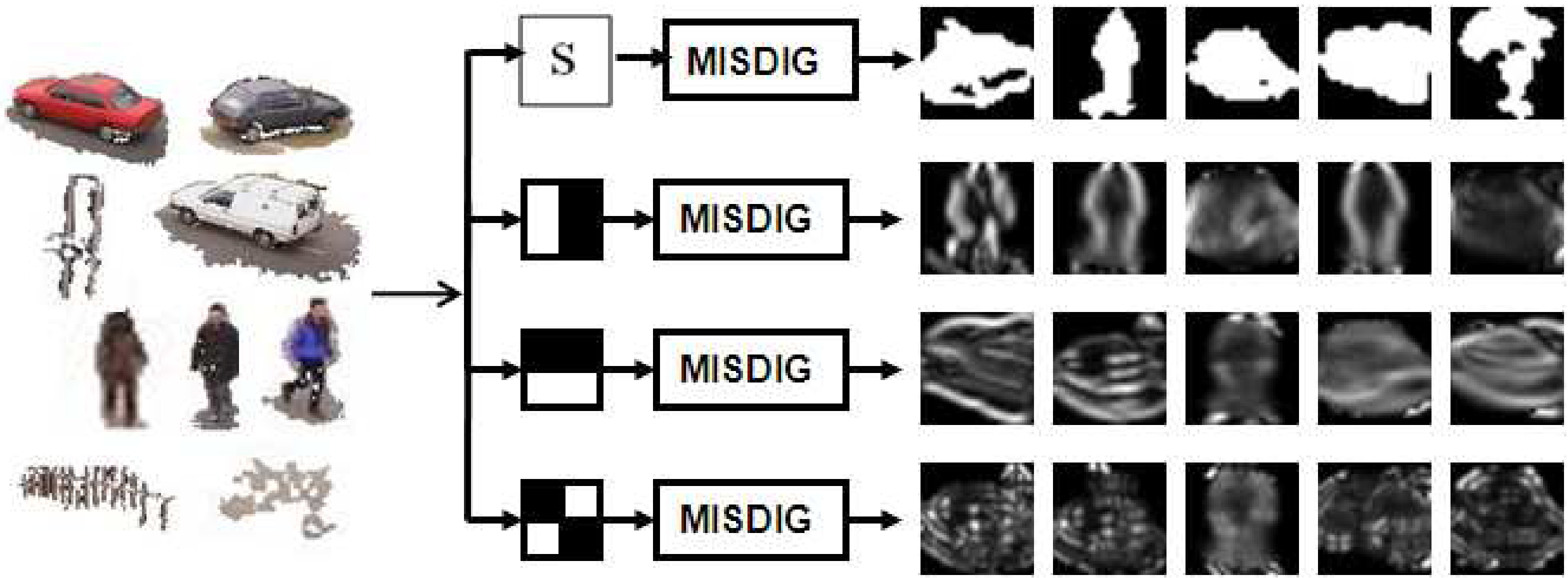
Multiple Object Tracking
Object Appearances
Shape Template
Haar Templates
S
Mixture of Spatially Distributed Gaussian Templates
Shape-Haar Template Co-occurrence Distribution
Spatial Clustering
Discovered Object Categories

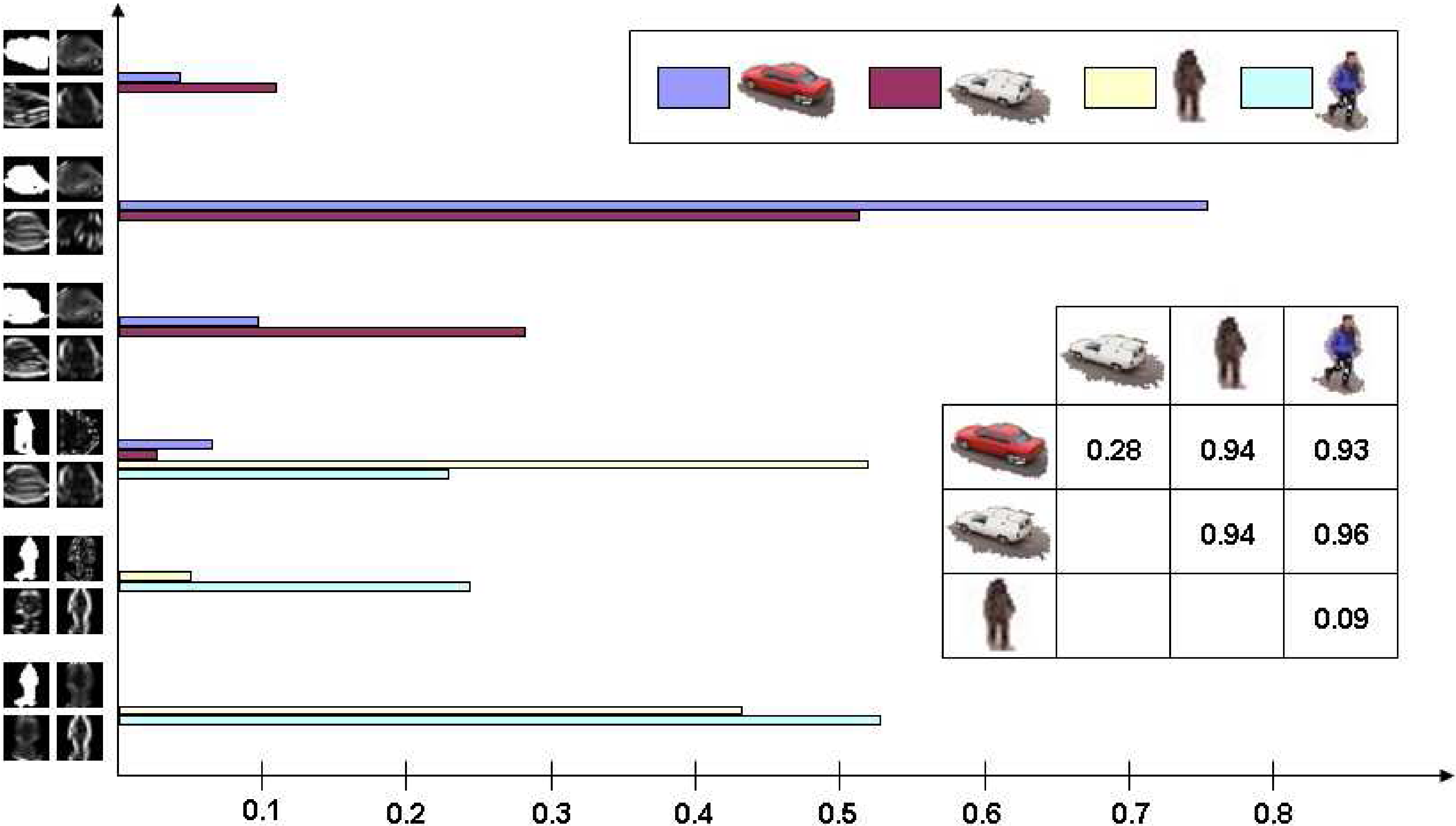


A
V
R
H0
Hh
Hv
Hd


S

# Feature Discovery




0.28
0.94
0.93
0.94
0.96
0.09
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8

# Does concept content drive language acquisition?

image sequence

foreground

tracked blobs = agents

bag of visual words (PHOW) + k-means clustering

30 clusters → 7 groundtruth categories

C1:

C10:

C16:

C19:

C29:

# Test on novel video

1. original video

2. novel video A

3. novel video B

foreground blobs
of 13 agents

8 labels
correct

5 incorrect

# Instructions for Narrators

This is a traffic video.

Watch the video for 40 seconds.

Next you will have to describe the same 40 sec of the video.

In Hindi, describe the objects, people, vehicles and what they are doing. [*]

Now, you will continue to describe the full video of around 4.5 minutes.

*: at this point, about 15% of subjects given some feedback – e.g.
“focus on events on the video and not generalities”

| | Sentence | Interval |
|---|---|---|
| S1 | ek bAik gayI abhI | 1158 -1224 |
| | One bike go+past now. | |
| | A bike went now. | |
| S2 | sAiD me.n sAikal rikshA pe ek ADamI caDhA | 1216-1382 |
| | Side [on] one cycle rickshaw [on] one man climb+past | |
| | A man climbed on a cycle rickshaw on the side (of the scene). | |
| S3 | * sAikal bAik Aye jA rahe hai.N. | 1239 -1354 |
| | Bicycles bikes come+pp go+pp are. | |
| | Bicycles, bikes are coming and going. | |

# Traffic video

frame
1200

frame
1255

frame
1300

frame
1350

# Probability Computation

Label-concept joint probability

$$J(l,c) = \frac{1}{T * |S|} * \sum_{t=1}^{T} \sum_{s \in S} P(c|s,t) * P(l|s,t)$$
$$= \frac{1}{T} * \sum_{t=1}^{T} P(c|t) * P(l|t)$$

Marginalized probabilities – label ; concept

$$P(c) = \frac{1}{T * |S|} * \sum_{t=1}^{T} \sum_{s \in S} P(c|s,t)$$
$$P(l) = \frac{f(l)}{\sum_l f(l)}$$

# Association measures

Conditional probability $$P(l|c) = J(l,c)/P(c)$$

Mutual Information
$$MI(l,c) = J(l,c) * \log(\frac{J(l,c)}{P(c) * P(l)})$$

Dominance-Weighted Joint probability

$$DJ(l,c) = w_s(l,c) * J(l,c)$$

$$w(l,c) = \frac{1}{|C|-1} \sum_{x \neq c} (NJ(l,c) - NJ(l,x))$$

$$w_s(l,c) = \frac{w(l,c) - min_x\{w(l,x)\}}{max_x\{w(l,x)\} - min_x\{w(l,x)\}}$$

# Results

# With word boundary knowledge

| ( $P_w$, CP / MI , T+, G+, A+, obj, ALL) | | | | |
|---|---|---|---|---|
| **Concept (c)** | **CP** | | **MI** | |
| | $l$ | $M(l, c)$ | $l$ | $M(l, c)$ |
| TEMPO | Tempo | 4.46 | kAr | 7.41 |
| | kAr | 4.33 | bAik | 7.34 |
| | pe | 4.25 | Tempo | 6.54 |
| BICYCLE | sAikal | 1.95 | sAikal | 1.34 |
| | ek sAikal | 1.14 | ek sAikal | 0.96 |
| | moTarsAikal | 0.79 | gais silinDar | 0.53 |
| MOTORCYCLE | pe | 8.60 | pe | 12.88 |
| | bAik | 7.12 | bAik | 11.64 |
| | Tempo | 6.56 | skUTar | 8.99 |
| TRUCK | Trak | 17.29 | Trak | 15.01 |
| | ek Trak | 10.67 | ek Trak | 9.91 |
| | pe | 3.24 | tIn sAikalwAle | 2.37 |
| HUMAN | saD.ak | 7.50 | saD.ak | 27.90 |
| | krOs | 6.68 | krOs | 20.76 |
| | roD | 6.54 | roD | 18.19 |
| CAR | kAr | 7.76 | kAr | 9.30 |
| | ek kAr | 4.89 | ek kAr | 6.61 |
| | gADI | 3.99 | gADI | 4.38 |

Top3 word
(k = 1 to 4)

# Without word boundary knowledge (syllable concatenation)

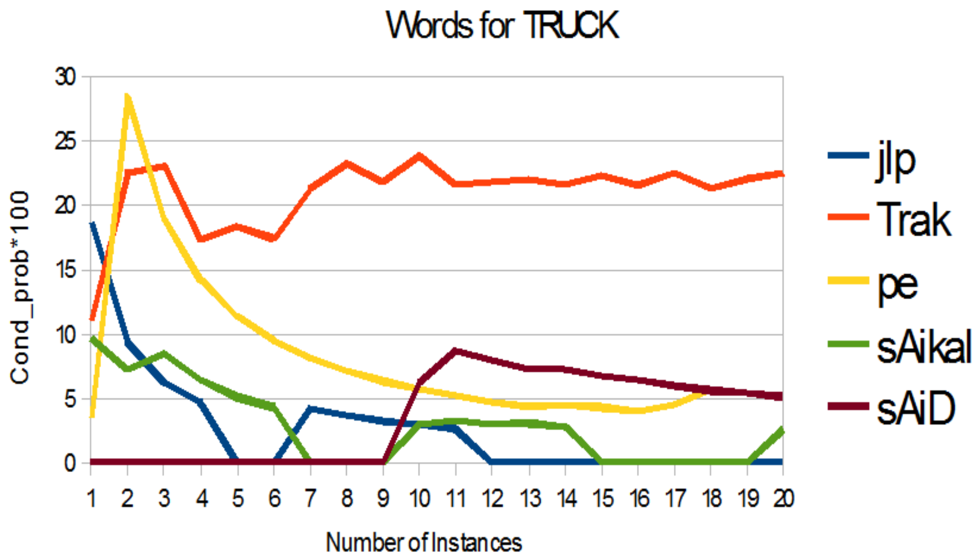
| ( $P_s$, CP / MI, T+, G+, A+, obj, ALL) | | | | |
|---|---|---|---|---|
| **Concept (c)** | **CP** | | **MI** | |
| | $l$ | $M(l, c)$ | $l$ | $M(l, c)$ |
| TEMPO | ik | 12.23 | ik | 29.68 |
| | jAr | 9.23 | jAr | 21.35 |
| | kal | 8.76 | kal | 19.70 |
| BICYCLE | sAikal | 2.90 | sAikal | 2.81 |
| | jAr | 1.62 | eksAi | 1.60 |
| | eksAi | 1.30 | ksAik | 1.60 |
| MOTORCYCLE | ik | 19.09 | ik | 39.35 |
| | D | 15.08 | D | 28.61 |
| | jAr | 13.43 | Tar | 26.42 |
| TRUCK | Trak | 19.23 | Trak | 22.55 |
| | ekTrak | 11.83 | ekTrak | 14.70 |
| | jAr | 10.20 | jAr | 9.42 |
| HUMAN | hAhai | 14.37 | hAhai | 62.35 |
| | D | 13.85 | D | 53.54 |
| | jAr | 10.86 | wAlA | 46.14 |
| CAR | ekkAr | 5.15 | ekkAr | 9.38 |
| | jAr | 4.51 | gADI | 6.12 |
| | rahlhai | 4.33 | rahlhai | 5.05 |

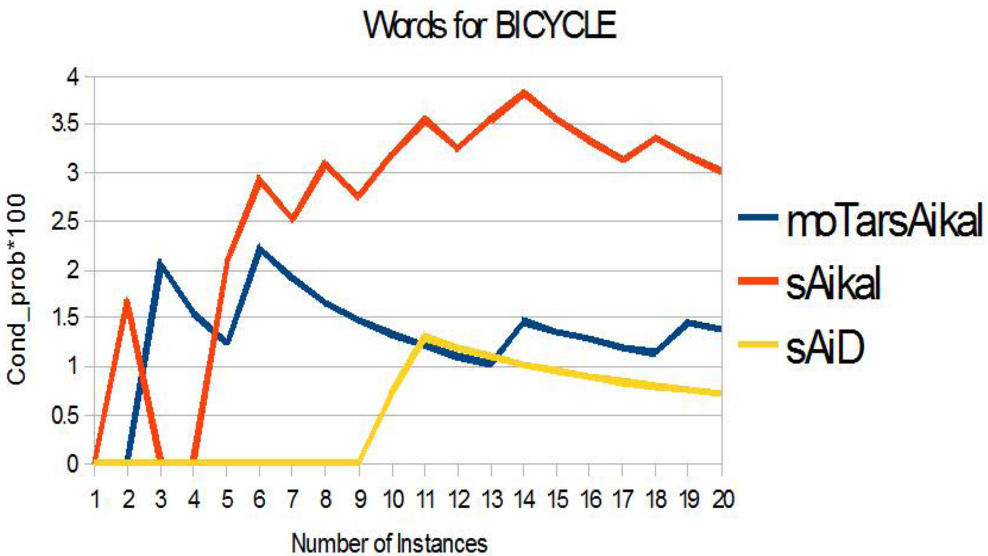
Top3 word
(k = 1 to 4)

# Top words (1000) removed

| ( W, M*, T+, G+, A+, obj, ALL) | | | | | | |
|---|---|---|---|---|---|---|
| **Concept (c)** | **DJ** | | **CP** | | **MI** | |
| | $l$ | $M(l, c)$ | $l$ | $M(l, c)$ | $l$ | $M(l, c)$ |
| TEMPO | bAik | 0.42 | Tempo | 4.46 | mArutI | 2.24 |
| | kAr | 0.40 | kAr | 4.33 | bAik | 1.92 |
| | piilii | 0.36 | pe | 4.25 | kAr | 1.72 |
| BICYCLE | sAikal | 0.02 | sAikal | 1.95 | silinDar | 0.16 |
| | uspar | 0.02 | moTarsAikal | 0.79 | I.njin | 0.14 |
| | I.njin | 0.02 | pe | 0.63 | Dilaks | 0.14 |
| MOTORCYCLE | skUTar | 0.76 | pe | 8.60 | bAik | 4.65 |
| | bAik | 0.70 | bAik | 7.12 | pe | 4.43 |
| | a.ndar | 0.54 | Tempo | 6.56 | skUTar | 4.30 |
| TRUCK | Trak | 0.41 | Trak | 17.29 | Trak | 6.22 |
| | Ta.Nkar | 0.21 | pe | 3.24 | peTrol | 1.47 |
| | peTrol | 0.16 | sAikal | 2.84 | Ta.Nkar | 1.13 |
| HUMAN | saD.ak | 2.74 | saD.ak | 7.50 | saD.ak | 12.51 |
| | biThAke | 2.12 | krOs | 6.68 | krOs | 7.06 |
| | rikshAwAlA | 1.84 | roD | 6.54 | biThAke | 5.81 |
| CAR | kAr | 0.40 | kAr | 7.76 | kAr | 3.61 |
| | camcamAtI | 0.23 | gADI | 3.99 | gADI | 1.46 |
| | mahAshay | 0.22 | nikalii | 2.81 | nikalii | 1.33 |

# Confidence gain




The winning word begins dominating from after 5-8 linguistic narrative exposures

# 30 clusters (no merging)

| Cluster | l | CP | l | MI |
|---|---|---|---|---|
| C0 (T) | kAr | 4.98 | bAik | 2.4 |
| | bAik | 4.96 | moTarsAikal | 2.13 |
| | Tempo | 4.5 | kAr | 2.01 |
| C8 (M) | bAik | 14.22 | skUTar | 4.57 |
| | skUTar | 12.66 | bAik | 3.7 |
| | pe | 12.53 | pe | 2.27 |
| C15 (B) | sAikalwAle | 8.82 | sAikal | 3.64 |
| | sAikal | 7.12 | sAikalwAle | 3.63 |
| | dAe.N | 6.85 | dAe.N | 1.67 |
| C19 (C) | kAr | 8.27 | kAr | 3.78 |
| | gADI | 4.05 | gADI | 1.4 |
| | nikalii | 2.85 | nikalii | 1.27 |
| C22 (M) | roD | 6.82 | roD | 0.41 |
| | pe | 2.68 | khAll | 0.3 |
| | skUTar | 1.92 | laDkI | 0.29 |
| C25 (T) | Tempo | 18.33 | Tempo | 5.62 |
| | pe | 11.75 | mUD | 3.05 |
| | sAikal | 6.87 | pe | 2.75 |
| C28 (B) | Tempo | 12.36 | Tempo | 3.02 |
| | sAikal | 8.48 | sAikalwAle | 3.01 |
| | sAikalwAle | 6.27 | mUD | 1.83 |
| C29 (L) | Trak | 26.4 | Trak | 4.83 |
| | pe | 8.02 | sAmAn | 1.51 |
| | sAmAn | 5.95 | Ore.nj | 1.25 |