Concurrent Object Recognition and Segmentation by Graph Partitioning

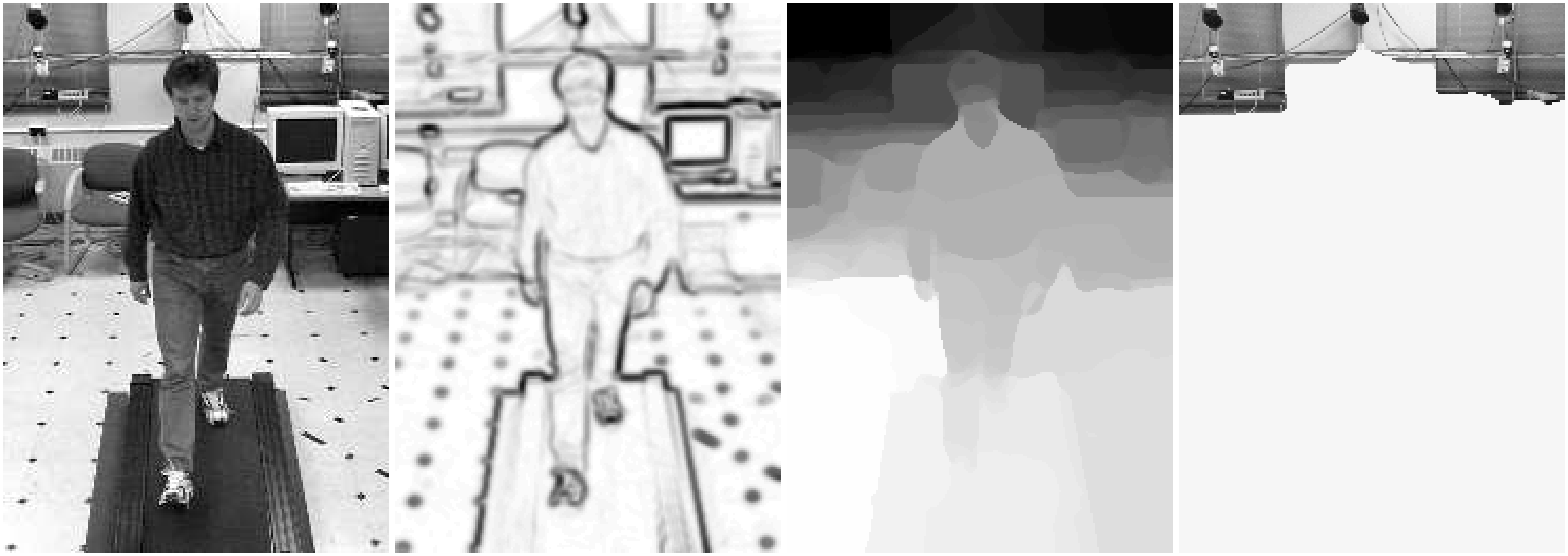
Stella X. Yu [1,2]   Ralph Gross [1]   Jianbo Shi [1]

Robotics Institute [1]
Carnegie Mellon University
Center for the Neural Basis of Cognition [2]

Why segmentation needs recognition?

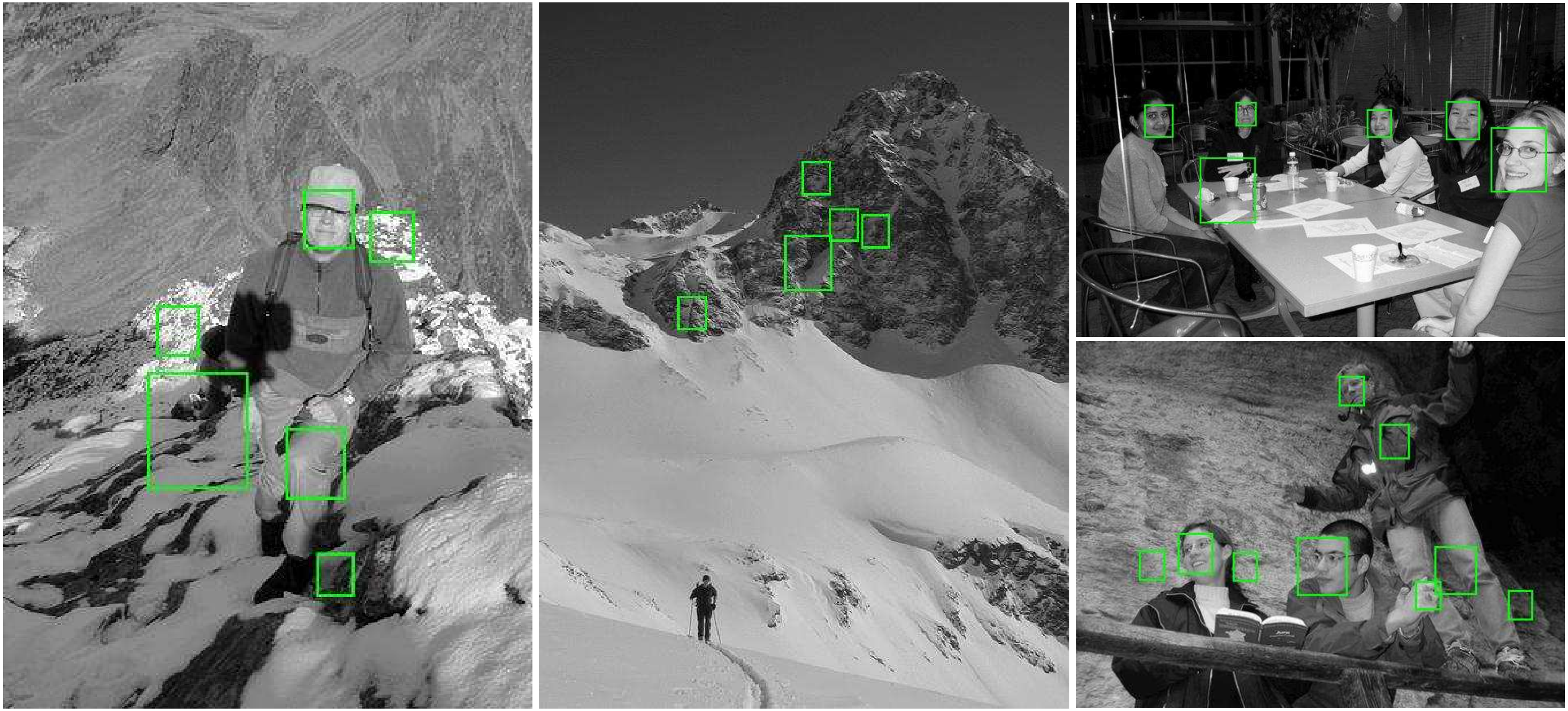
Why recognition needs segmentation?

Image segmentation is often object-blind



1. Do not know which regions make up an object.
2. Easily miss object boundaries due to lighting and clutter.

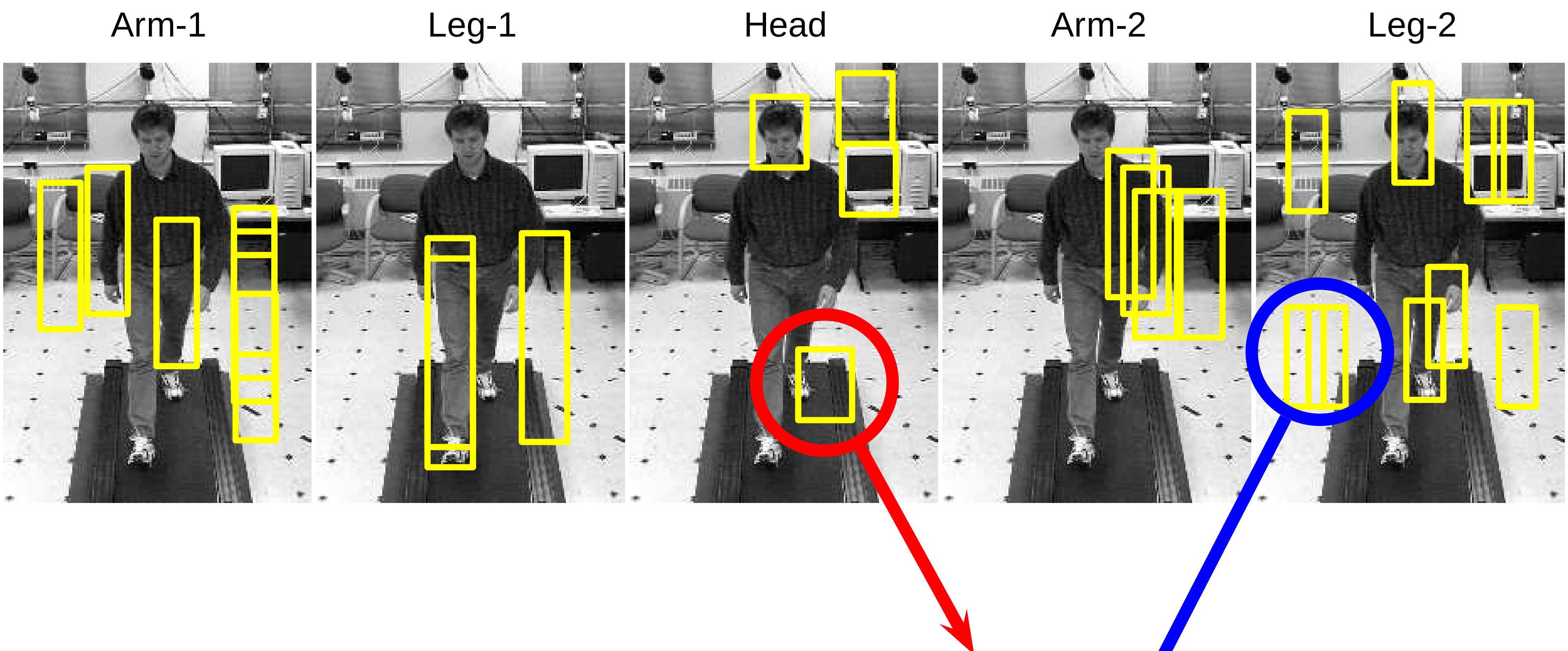
Object detection is often overwhelmed



(Schneiderman, 02): vasc.ri.cmu.edu/demos/faceindex

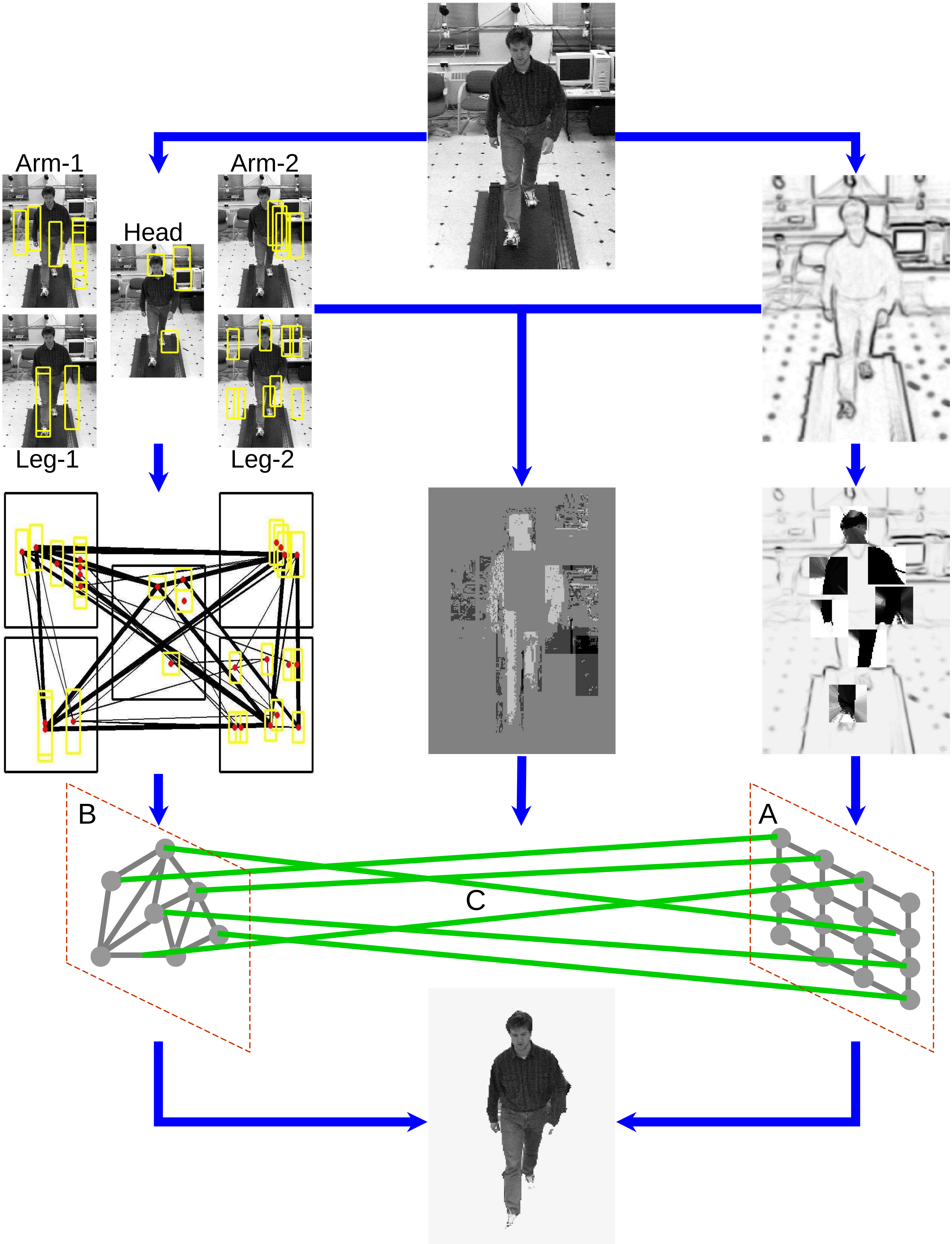
1. Tradeoff between false positives and detection rate.
2. Constraints in reducing false detection: increase in classifier complexity and training size.

Characteristics of false positives



1. Lack of high-level part label compatibility.
2. Lack of low-level image feature saliency.

Overview of our object segmentation



Representation

Graph: $\mathsf{G} = (\mathsf{V}, \mathsf{E}, W)$ = (nodes, edges, weights)

Node set: $\mathsf{V} = \mathsf{V}_{\mathrm{pixels}} \cup \mathsf{V}_{\mathrm{patches}}$

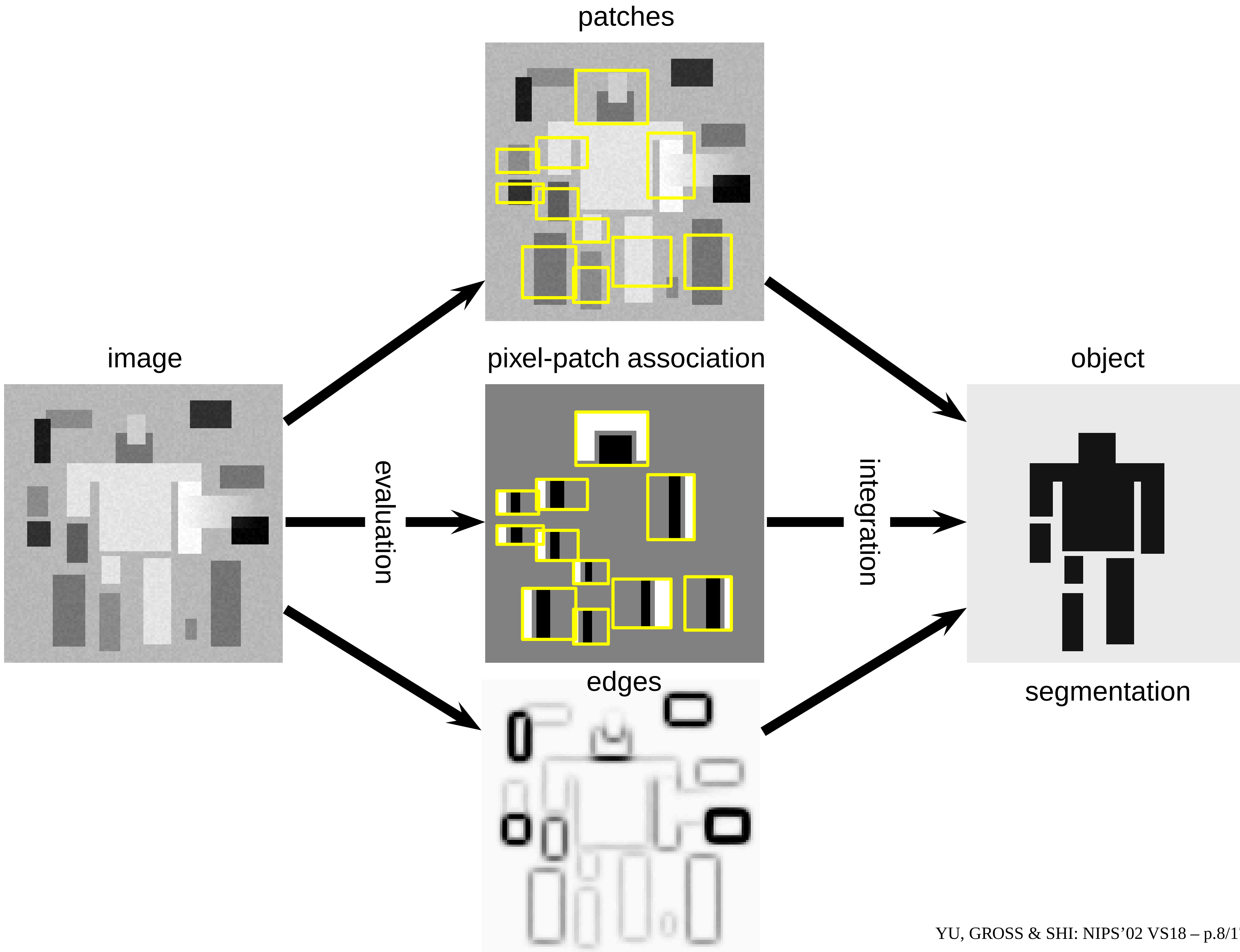
Edge set: $\mathsf{E} = \mathsf{E}_{\mathrm{pixel-pixel}} \cup \mathsf{E}_{\mathrm{patch-patch}} \cup \mathsf{E}_{\mathrm{pixel-patch}}$

Weights: $W = \begin{bmatrix} A & C^T \\ C & B \end{bmatrix}$

A: pixel-pixel similarity matrix

B: patch-patch compatibility matrix

C: pixel-patch association matrix


patches
image
pixel-patch association
object
evaluation
integration
edges
segmentation

Computing pixel-pixel similarity A

$A(1,3) \approx 1$

$A(1,2) \approx 0$

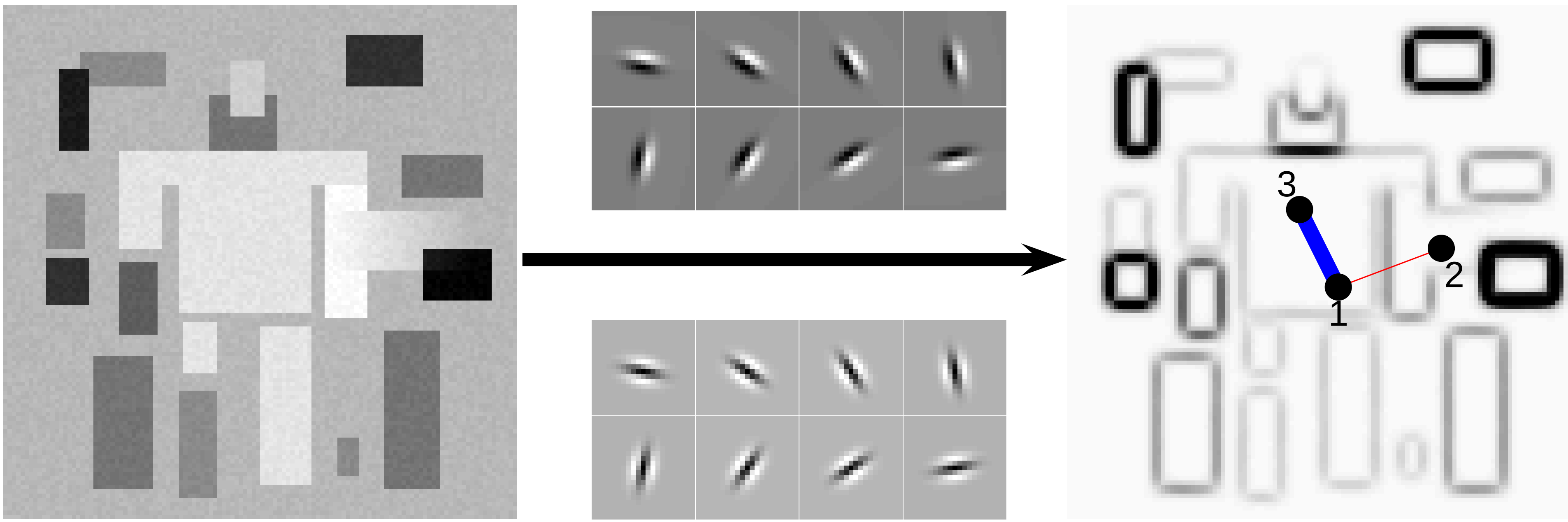


image filters edge magnitudes

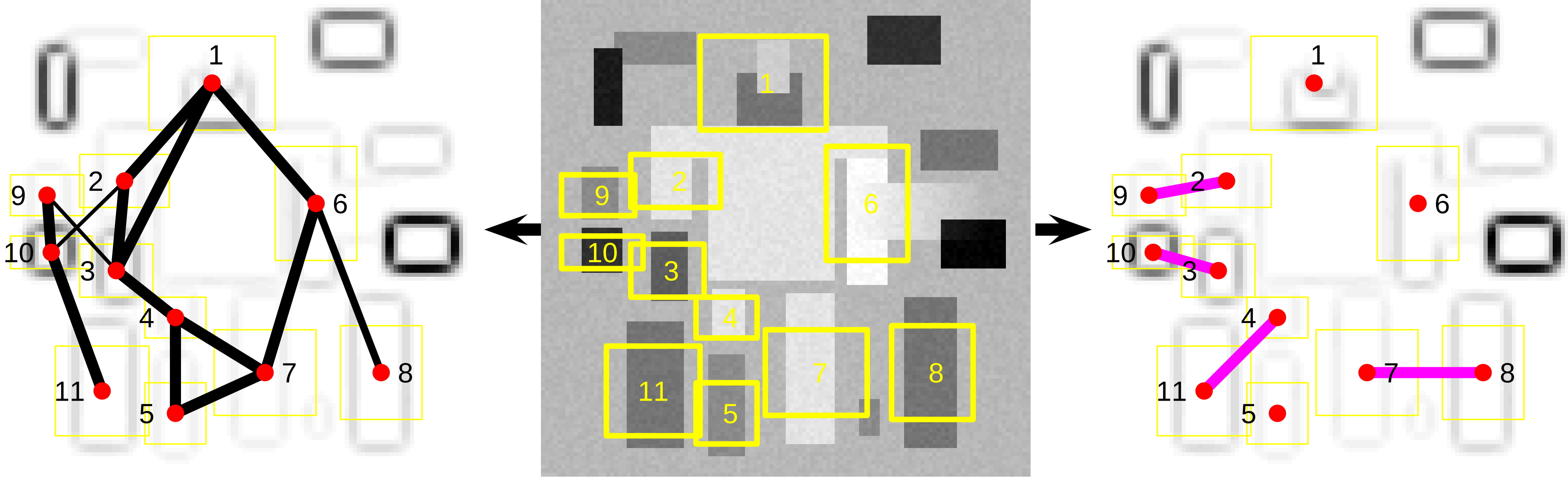
$$A(i,j) = \exp\left(-\frac{1}{2\sigma_e^2} \cdot \left[\frac{\max_{t\in(0,1)} OE(\underline{i} + t\cdot \underline{j})}{\max_k OE(\underline{k})} \right]^2 \right), \quad \underline{k} = \text{location of } k.$$

Computing patch compatibility and competition

$B(1,2) \approx 1$

$B(10,5) \approx 0$

7 and 8 cannot both be parts of the object

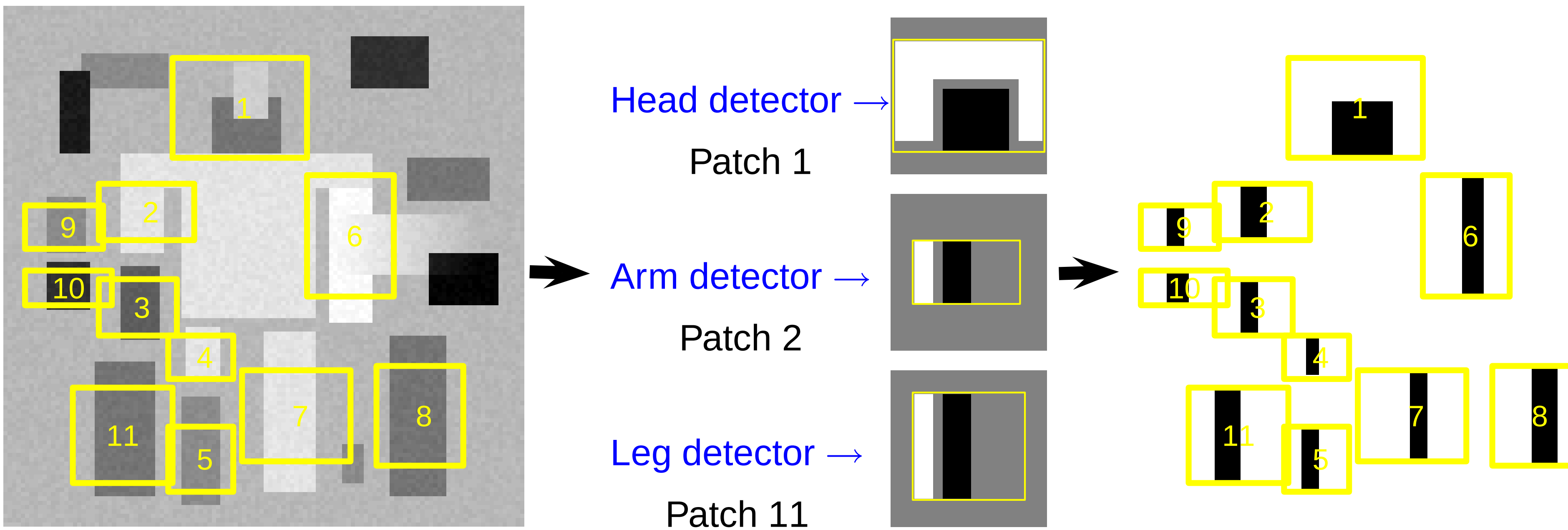


compatibility | patches | competition

$B(p,q)$ is small if p, q form rare configurations for part labels $\acute{p}$, $\acute{q}$:

$$B(p,q) = \exp\left(-\frac{1}{2}(\underline{p}-\underline{q}-\mu_{\acute{p}\acute{q}})^T \Sigma_{\acute{p}\acute{q}}^{-1}(\underline{p}-\underline{q}-\mu_{\acute{p}\acute{q}})\right), \quad \underline{p} = \text{location of } p.$$

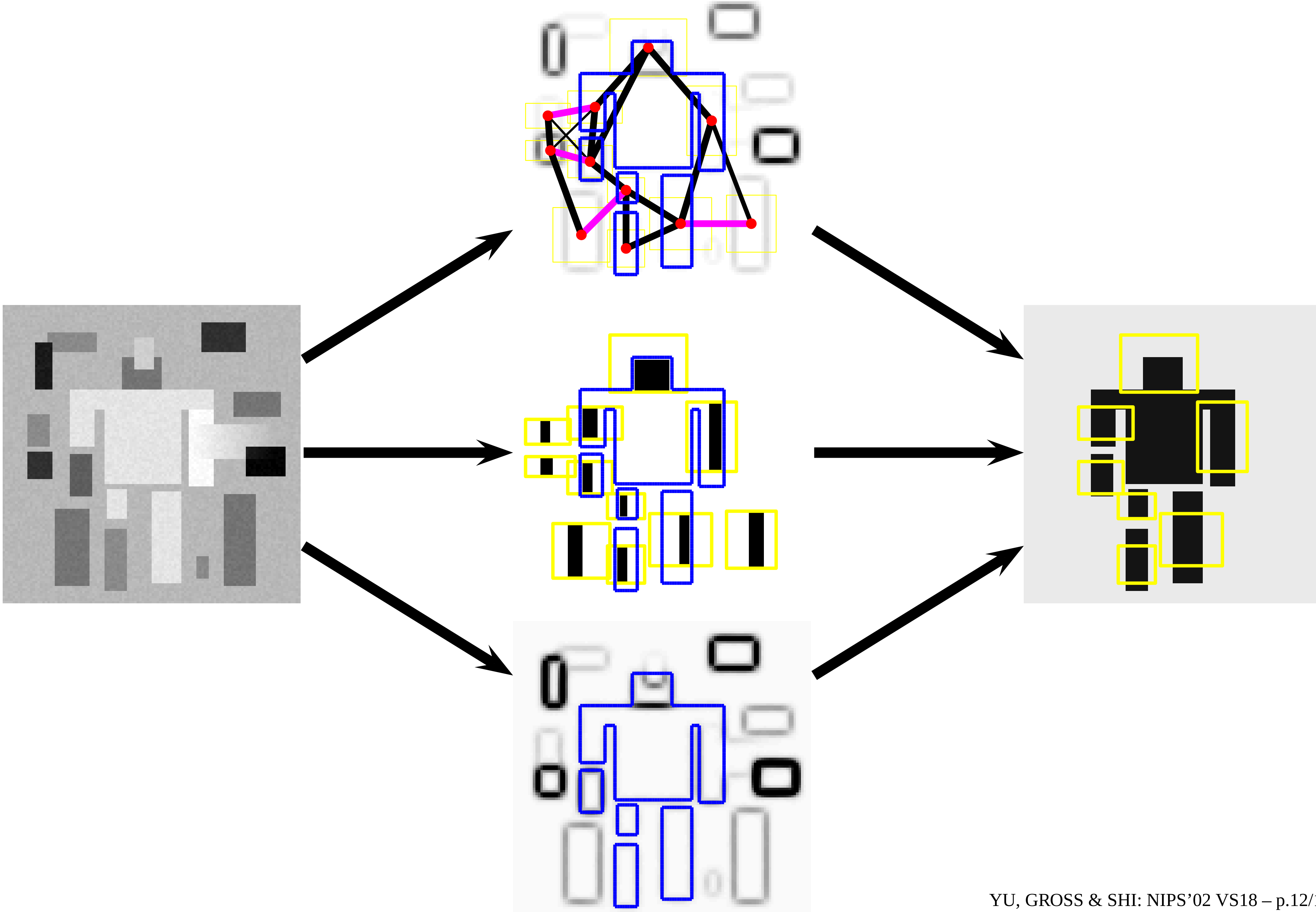
Computing pixel-patch association C



patches expected local segmentation association

$$C(i,p) = \begin{cases} 1, & \text{if } i \text{ is an object pixel of patch } p \\ 0, & \text{otherwise} \end{cases}$$

Find low-cost cuts subject to patch competition



Encoding graph cuts

Segmentation: V $=$ V$_1 \cup$ V$_2$ = object nodes $\cup$ the rest.

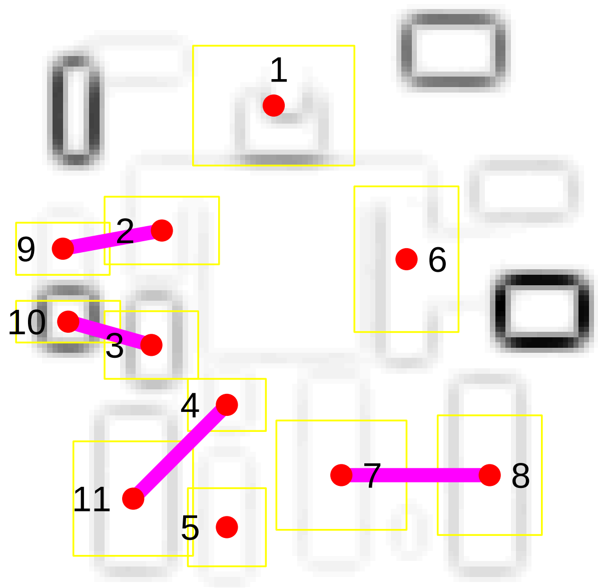
Indicators: $X = [X_1, X_2]$ = [is-object, is-nonobject].

Degree: $D =$ diag$(W \cdot 1)$.

Cuts criterion: $\max \text{NCuts}(X) = \frac{X_1^T W X_1}{X_1^T D X_1} + \frac{X_2^T W X_2}{X_2^T D X_2}.$

(Shi and Malik, 97)

Encoding patch competition



Competing nodes: pairs of patches of the same label.

$$S = N + \{\{2,9\},\{3,10\},\{4,11\},\{7,8\},\{1,12\}\}$$

e.g. $X_1(N+2) + X_1(N+9) = 1.$

N = total number of pixels

Exclusion condition: one winner only

$$\sum_{k \in S_m} X_1(k) = 1,\ m = 1 : |S|.$$

S_m = a set of nodes in competition.

Computational solution

Change of variable:

$$x = X_1 - \frac{X_1^T D X_1}{1^T D 1},$$

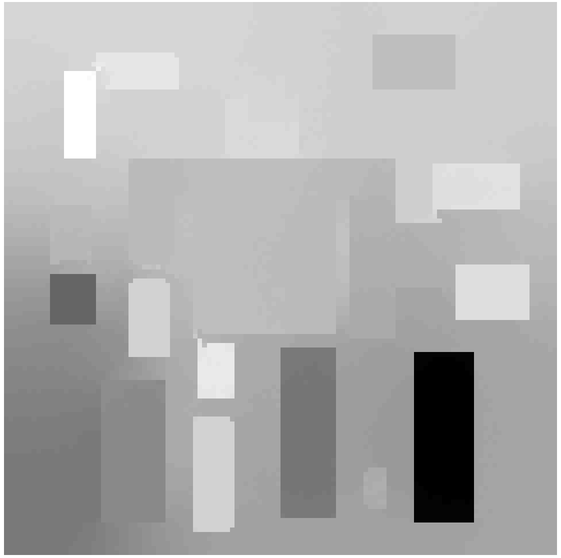
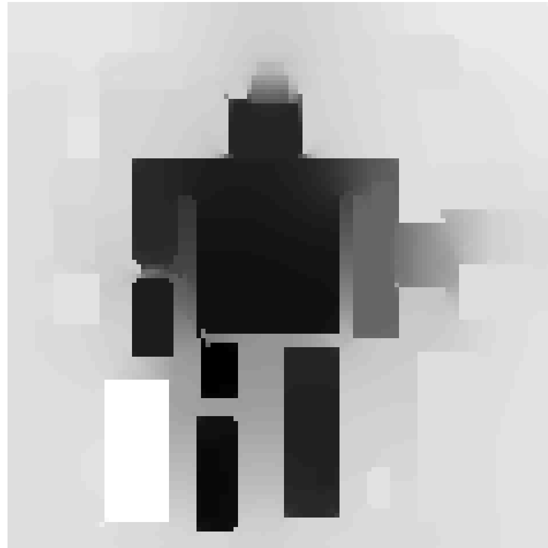
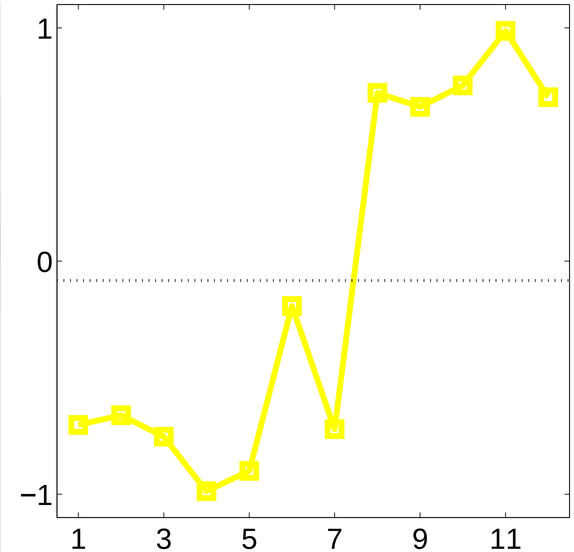
we have constrained eigenvalue problem:

$$x^* = \arg\max \frac{x^T W x}{x^T D x}, \quad \text{subject to } L^T x = 0.$$

Eigensolution in the relaxed continuous domain:

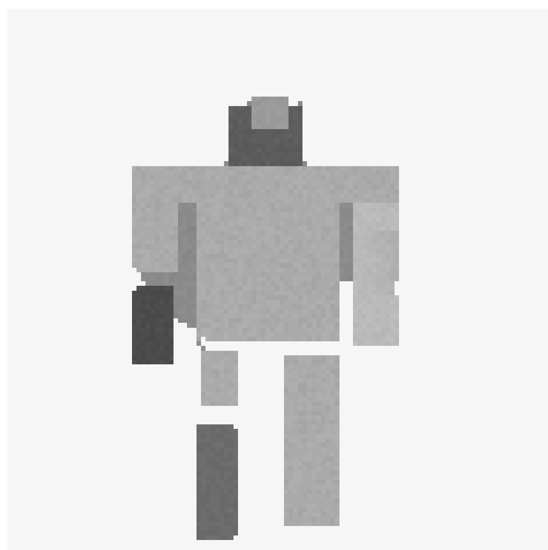
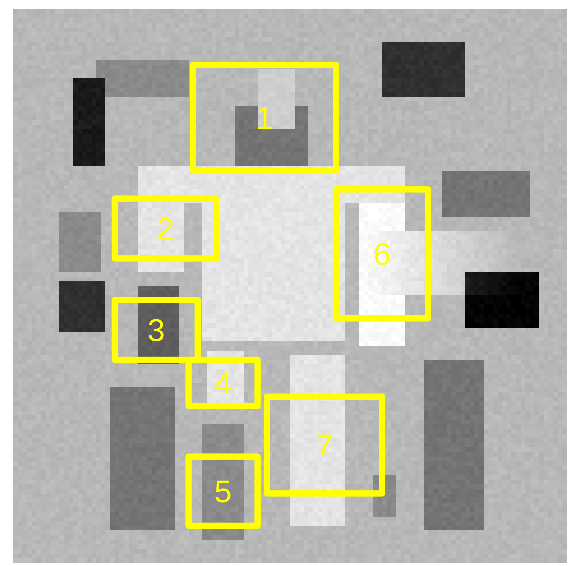
$$\begin{aligned} Q D^{-1} W x^* &= \lambda x^*, \\ Q &= I - D^{-1} L (L^T D^{-1} L)^{-1} L^T. \end{aligned}$$

Results I

segmentation alone

segmentation-recognition

44 seconds

17 seconds

Results II

segmentation alone: 68s

segmentation-recognition: 58s