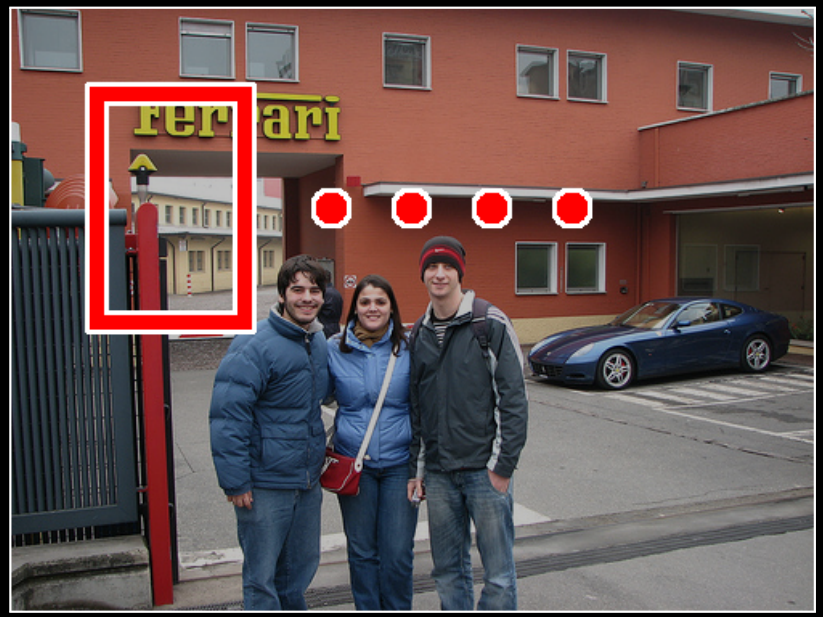
Object Detection and Segmentation from Joint Embedding of Parts and Pixels

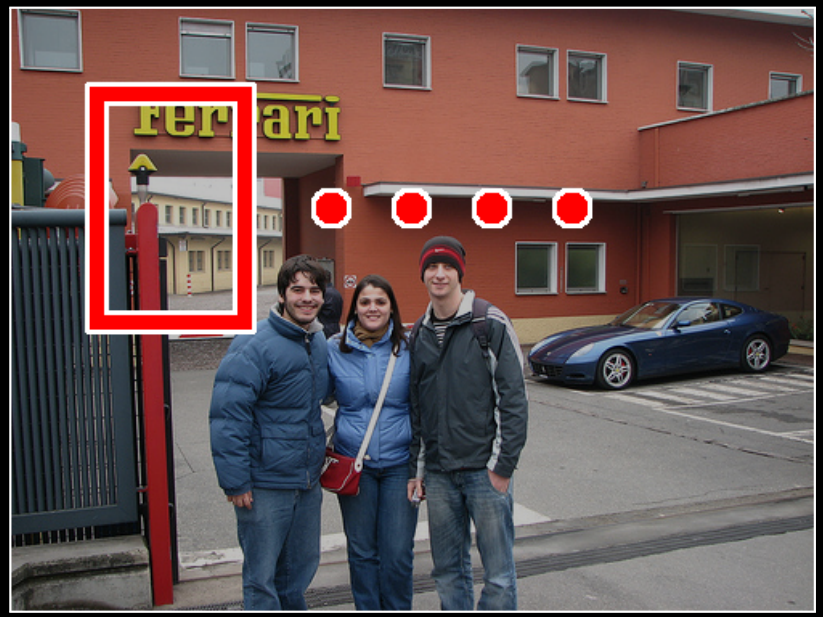
Michael Maire[1], Stella X. Yu[2], Pietro Perona[1]

[1]California Institute of Technology - Pasadena, CA 91125
[2]Boston College - Chestnut Hill, MA 02467

Segmentation



Detection



Segmentation

Detection

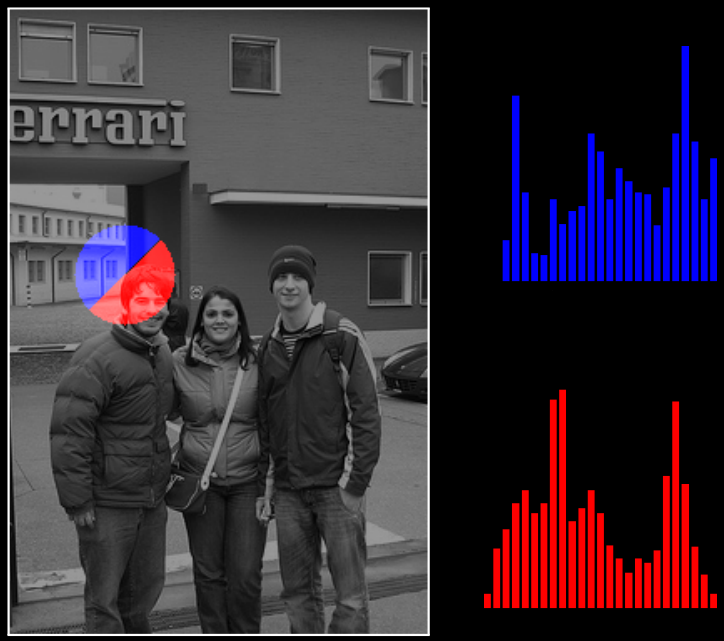
Perceptual Grouping Framework

Ingredients

Plug in state-of-the-art components:

Ingredients

Plug in state-of-the-art components:

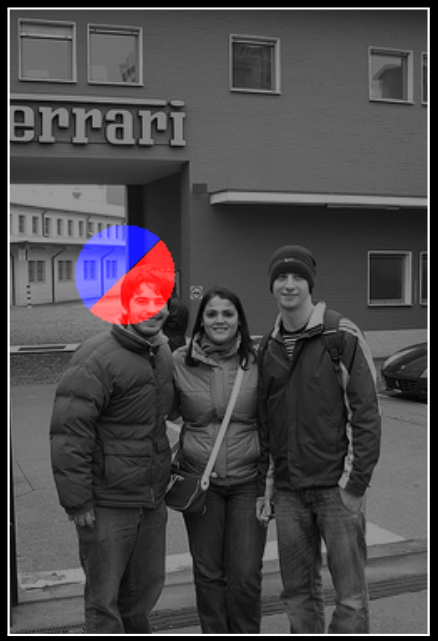


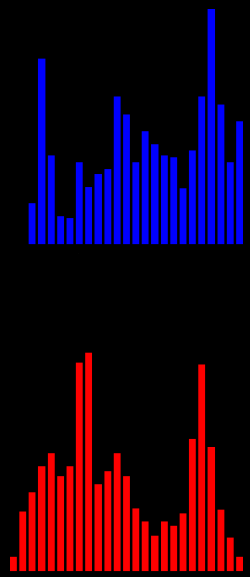
low-level cues:
color, texture, edges

[Arbeláez, Maire, Fowlkes, Malik, PAMI 2011]

Ingredients

Plug in state-of-the-art components:





low-level cues:
color, texture, edges

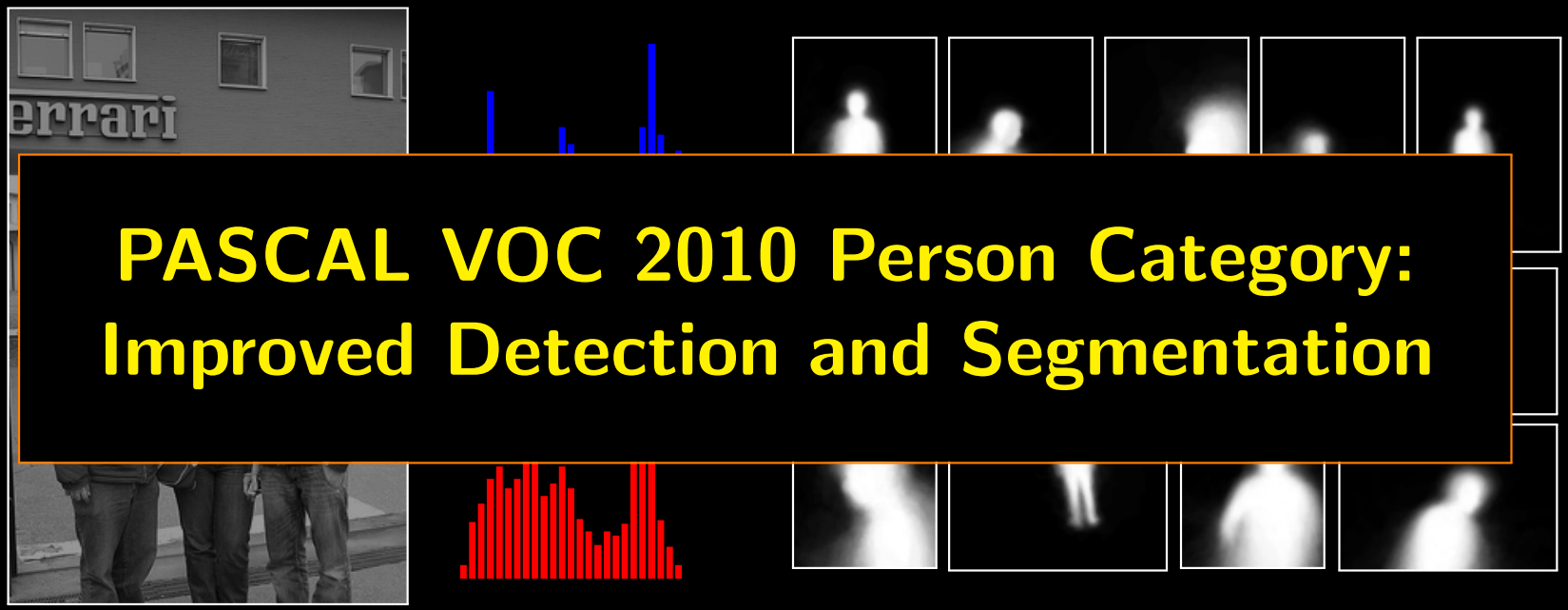
[Arbeláez, Maire, Fowlkes, Malik, PAMI 2011]

top-down parts:
poselets for person detection

[Bourdev, Maji, Brox, Malik, ECCV 2010]

Ingredients

Plug in state-of-the-art components:



PASCAL VOC 2010 Person Category: Improved Detection and Segmentation

low-level cues:
color, texture, edges

[Arbeláez, Maire, Fowlkes, Malik, PAMI 2011]

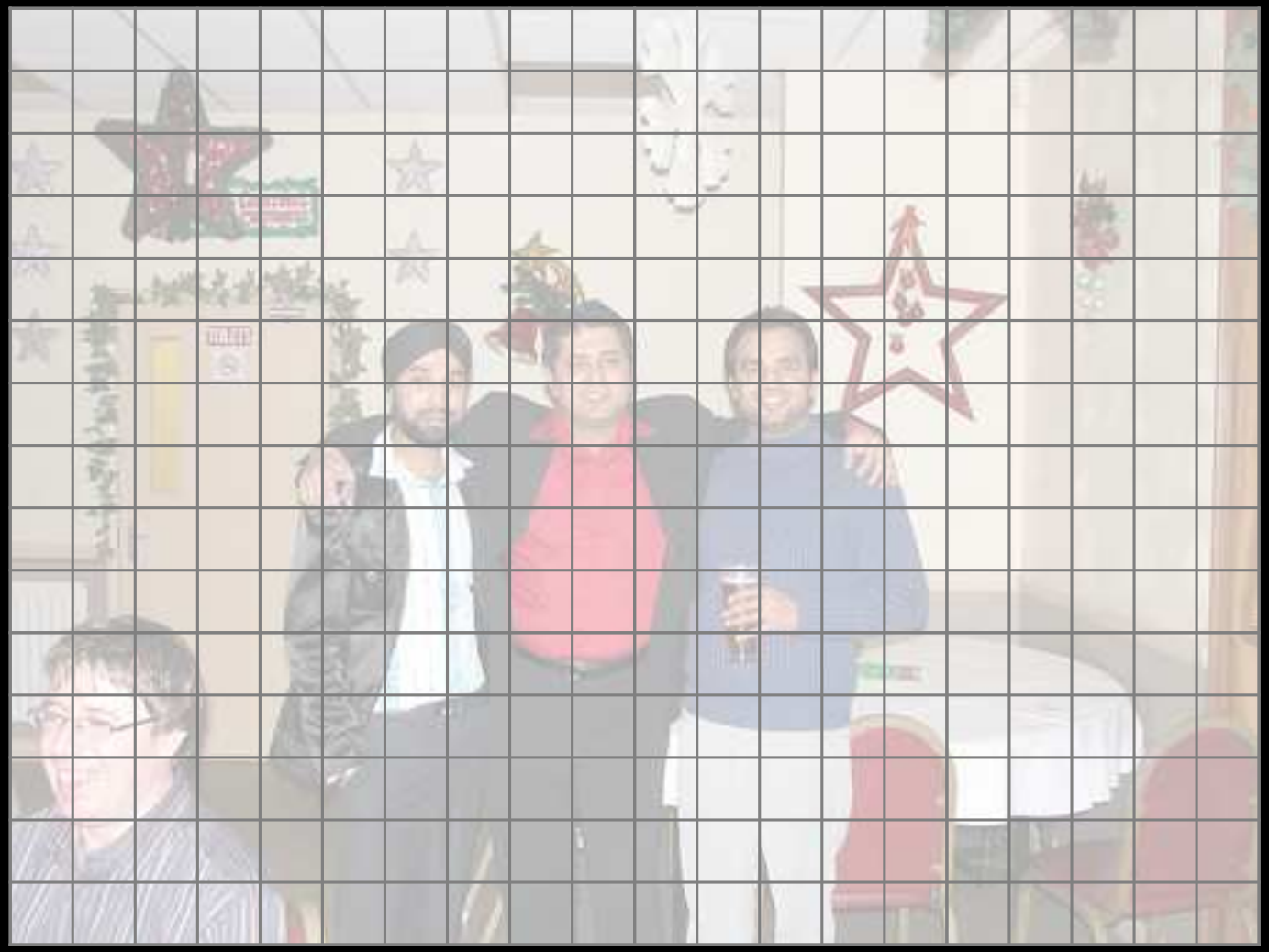
top-down parts:
poselets for person detection

[Bourdev, Maji, Brox, Malik, ECCV 2010]

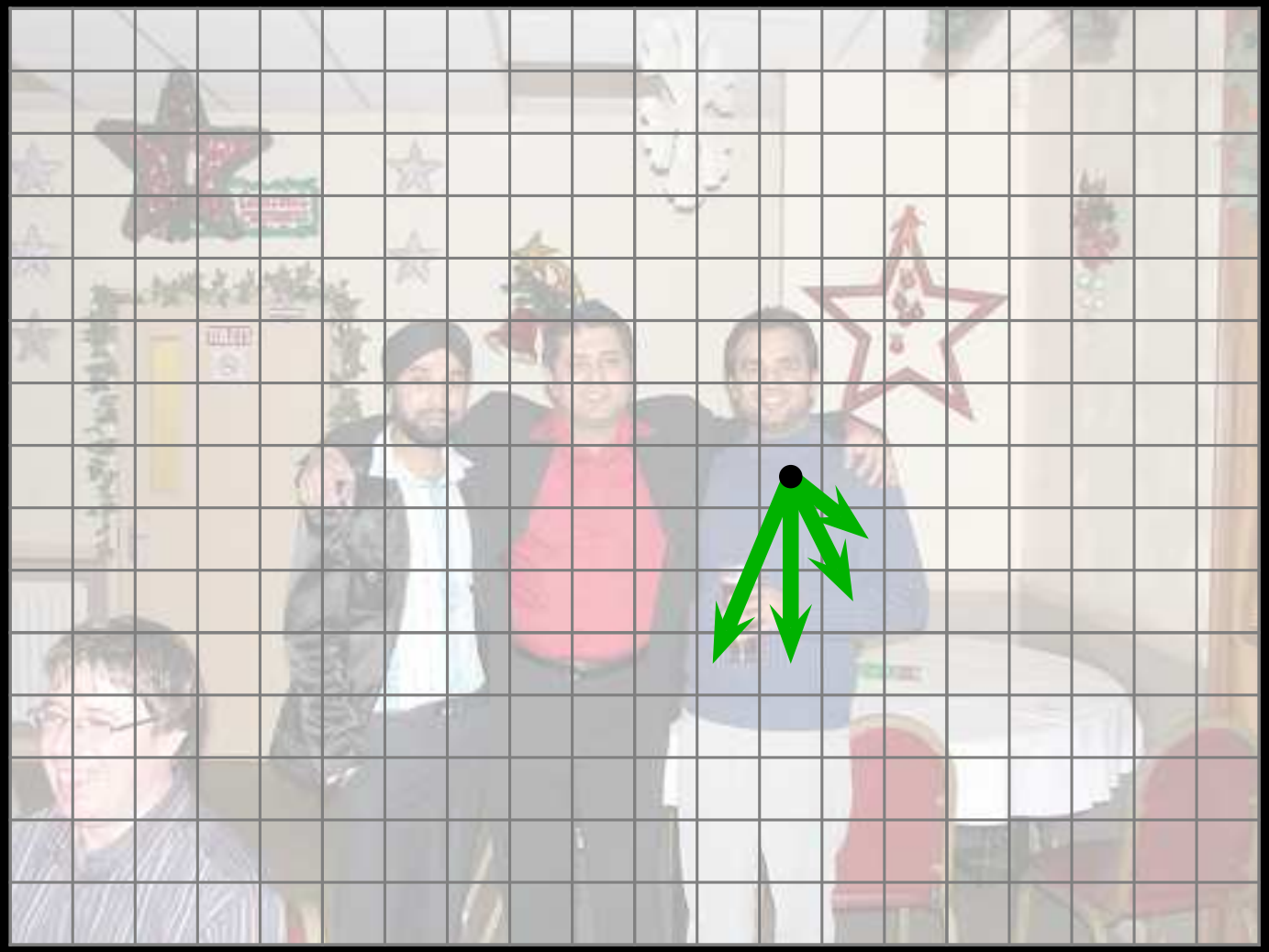
Grouping Relationships

Grouping Relationships

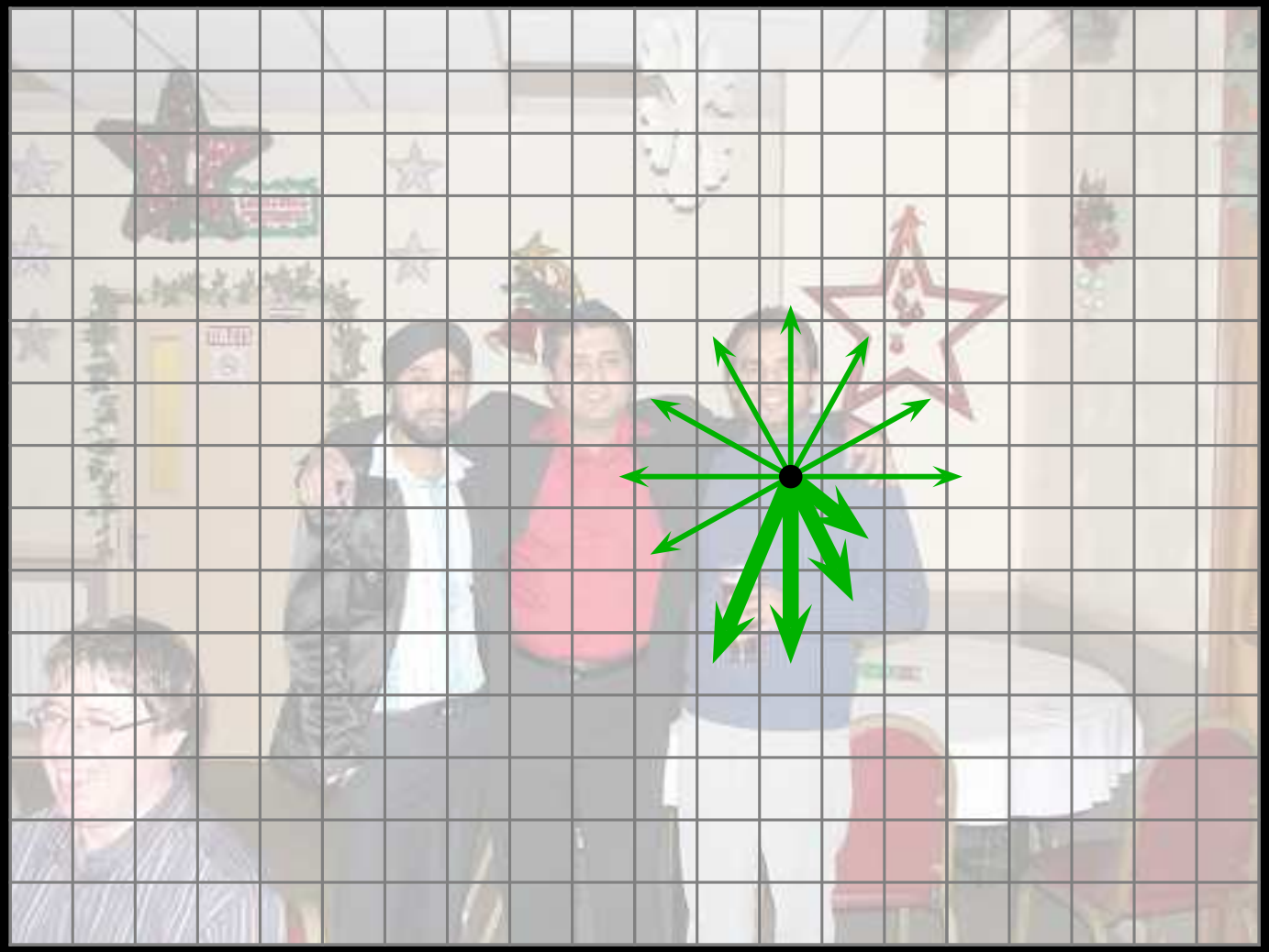
Pixel Affinity: Color, Texture Similarity



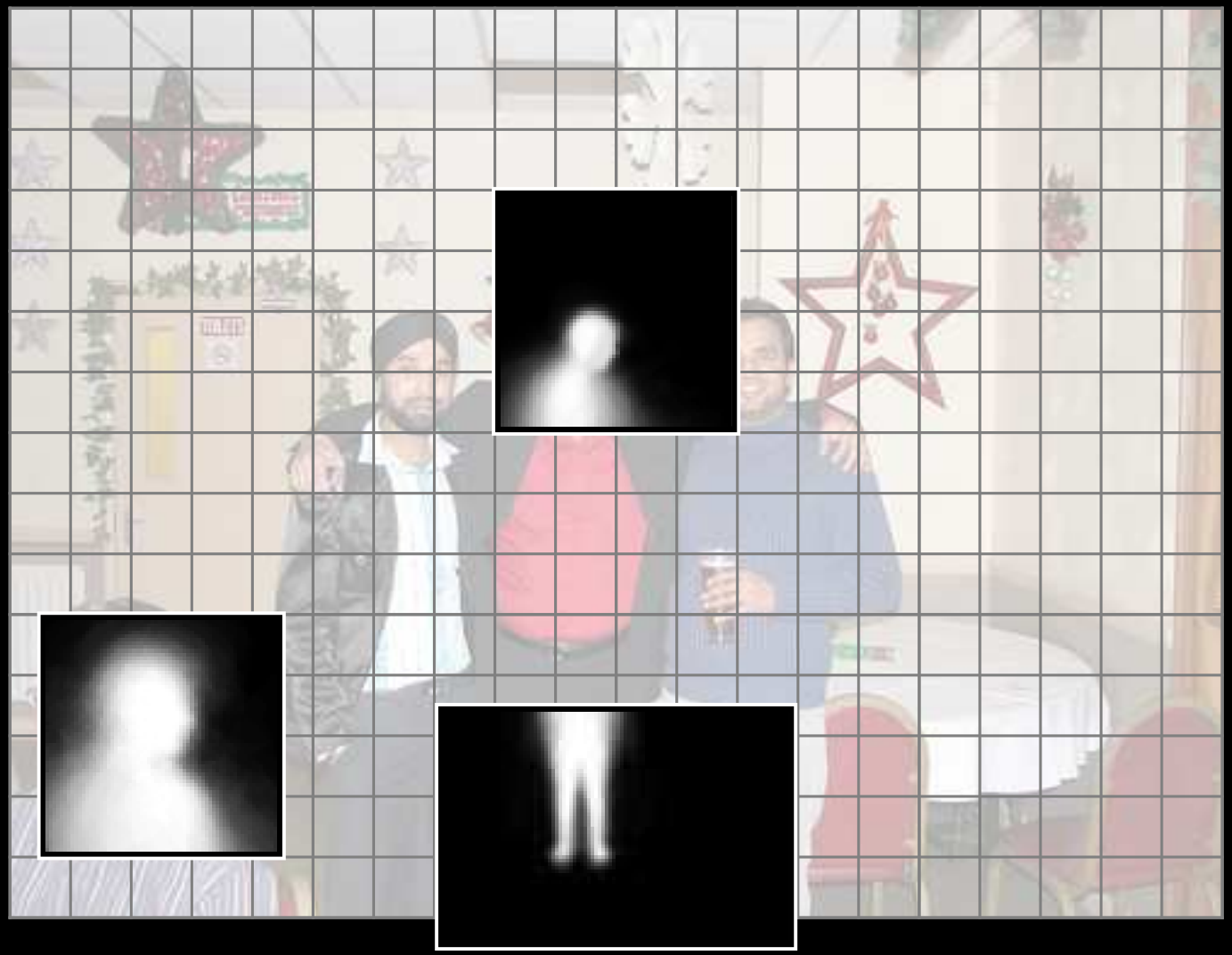
Pixel Affinity: Color, Texture Similarity



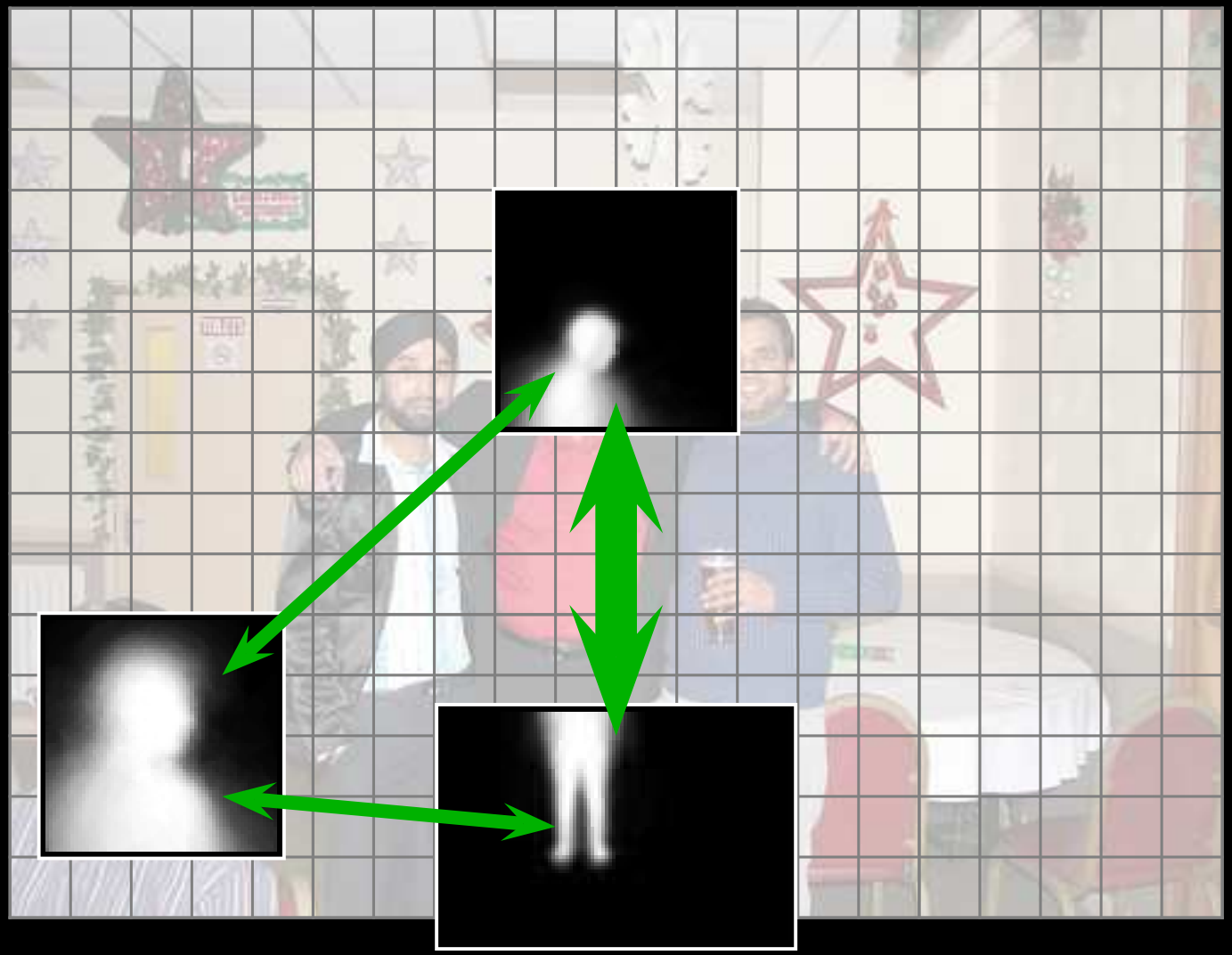
Pixel Affinity: Color, Texture Similarity

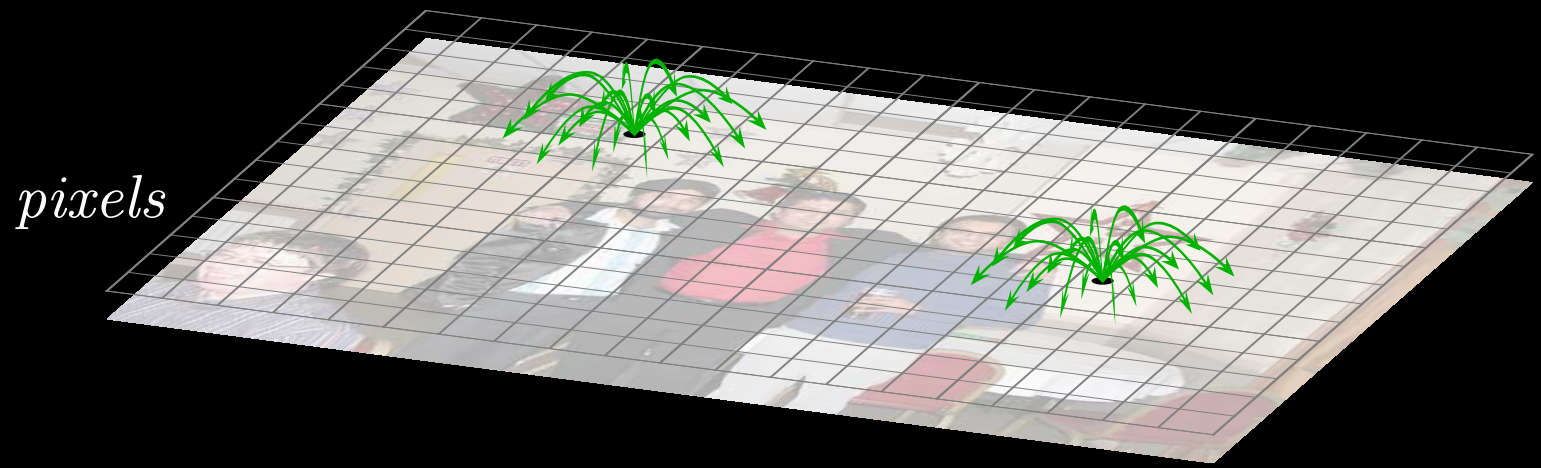


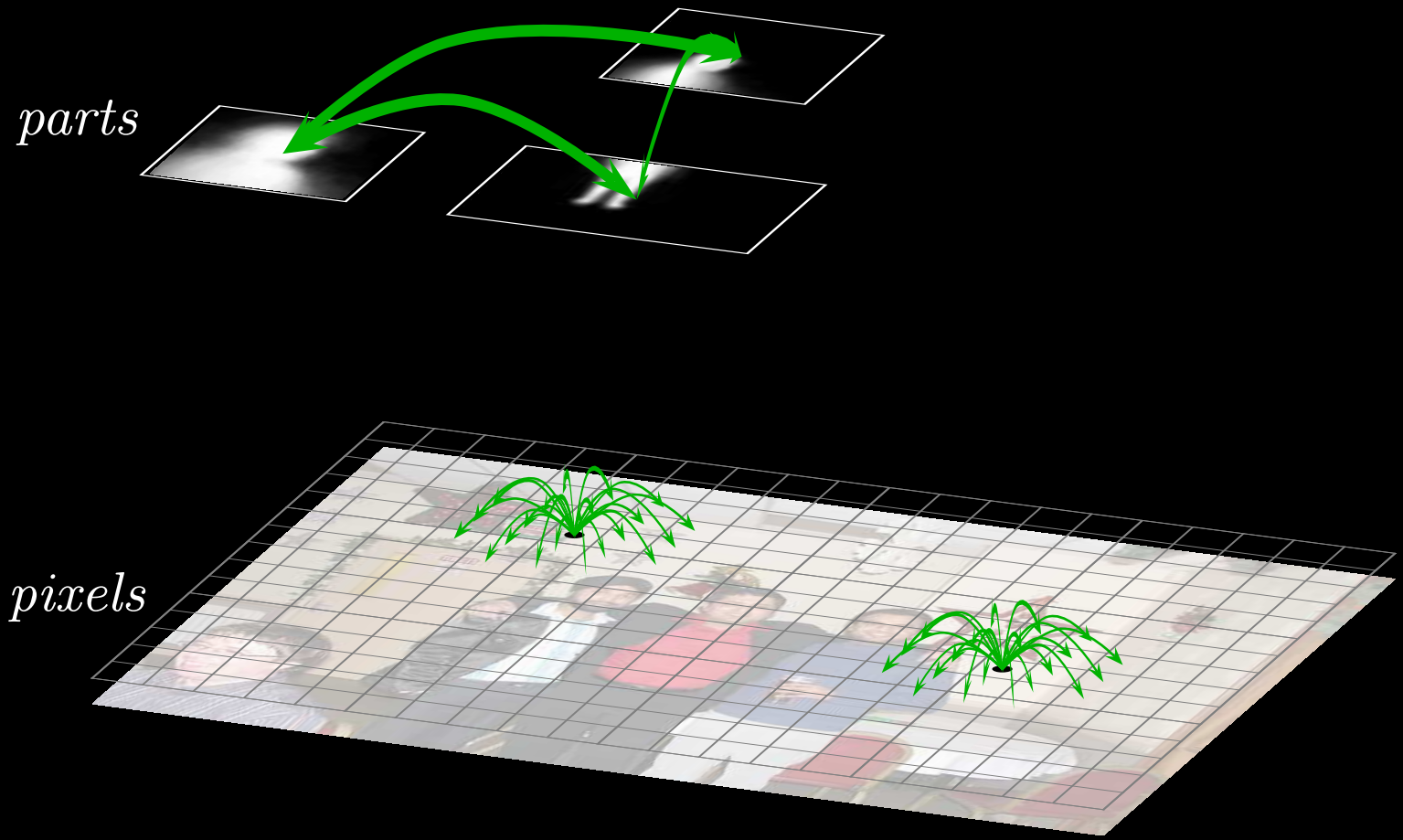
Part Affinity: Geometric Compatibility

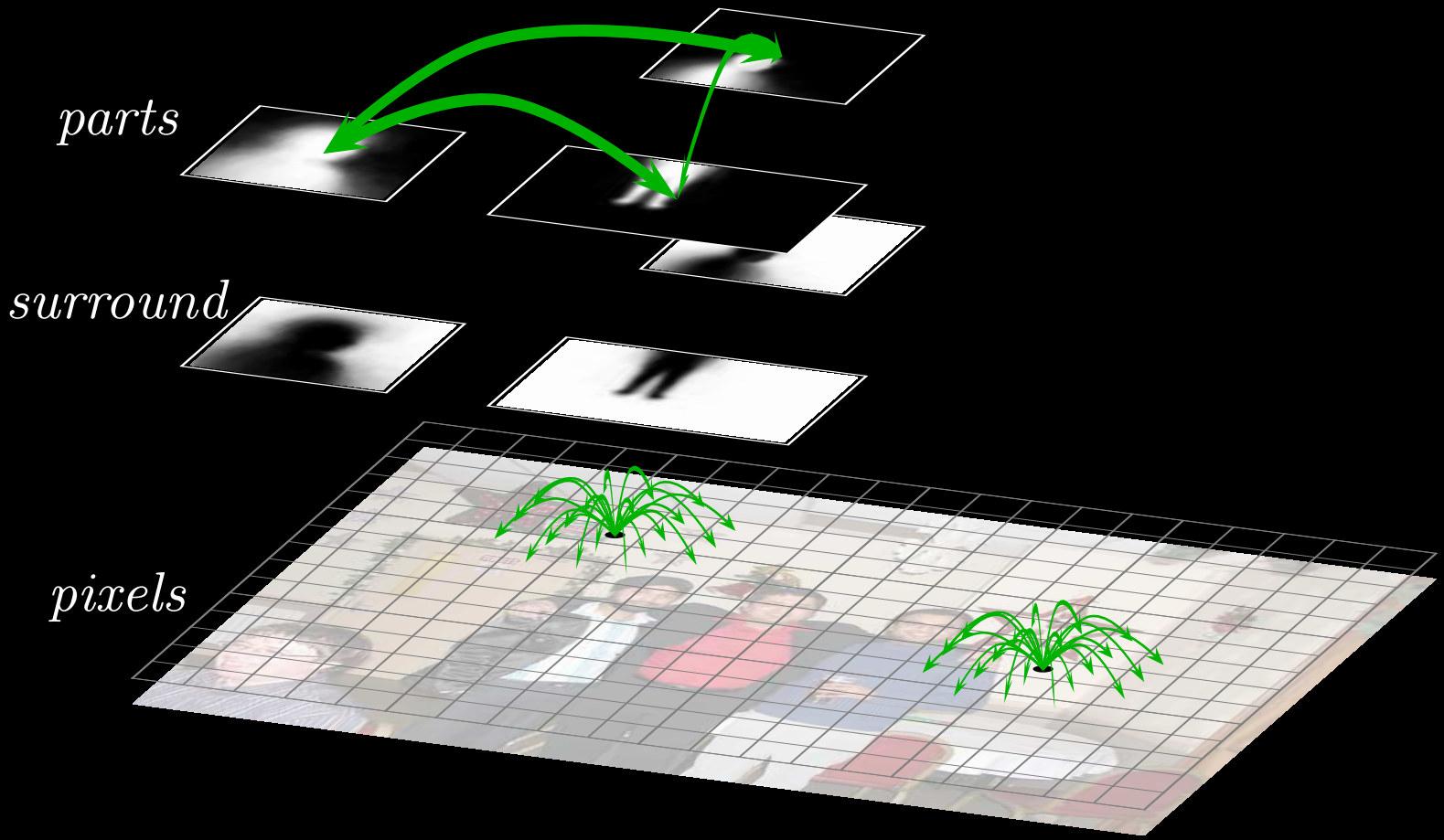


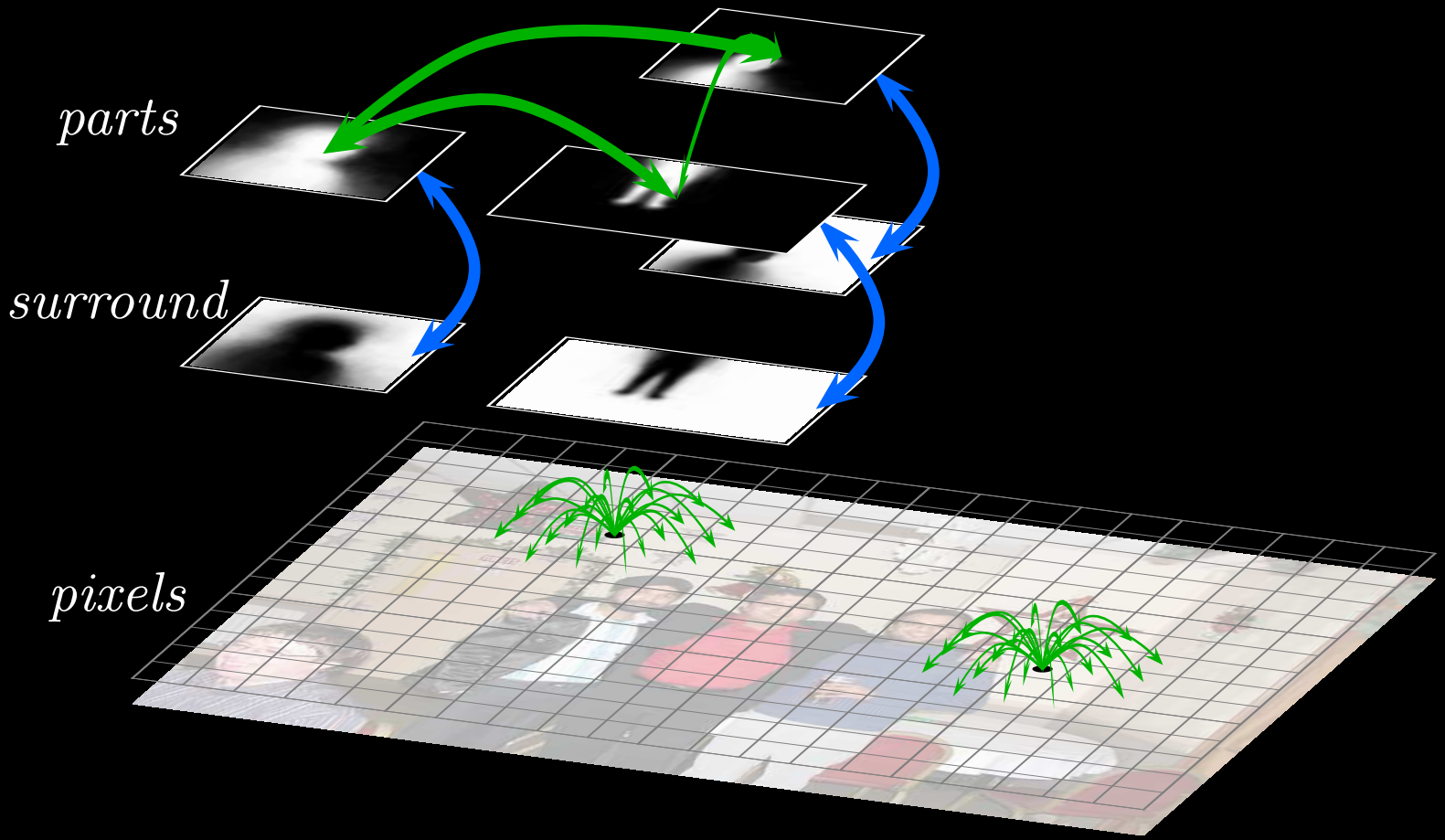
Part Affinity: Geometric Compatibility

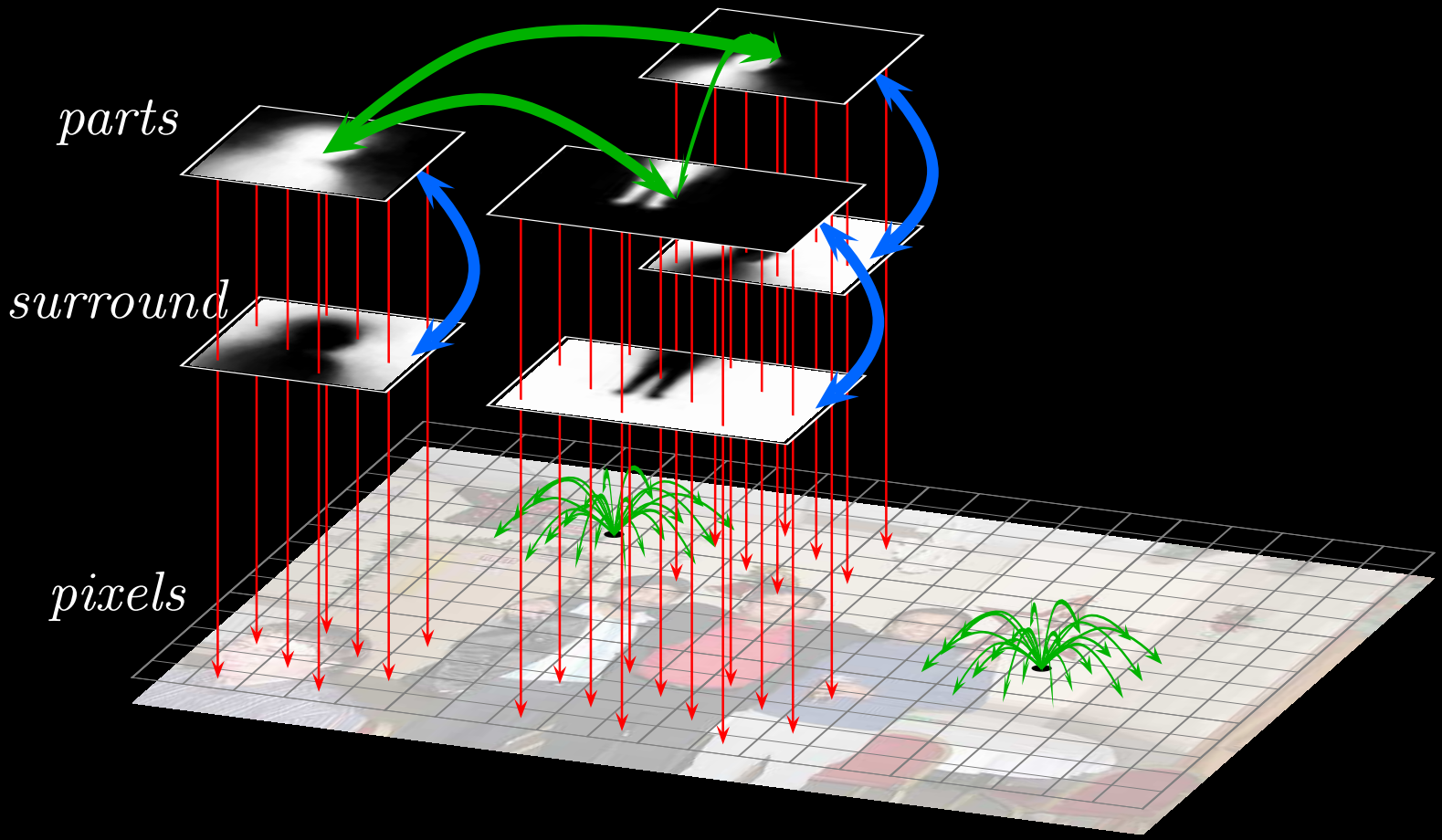


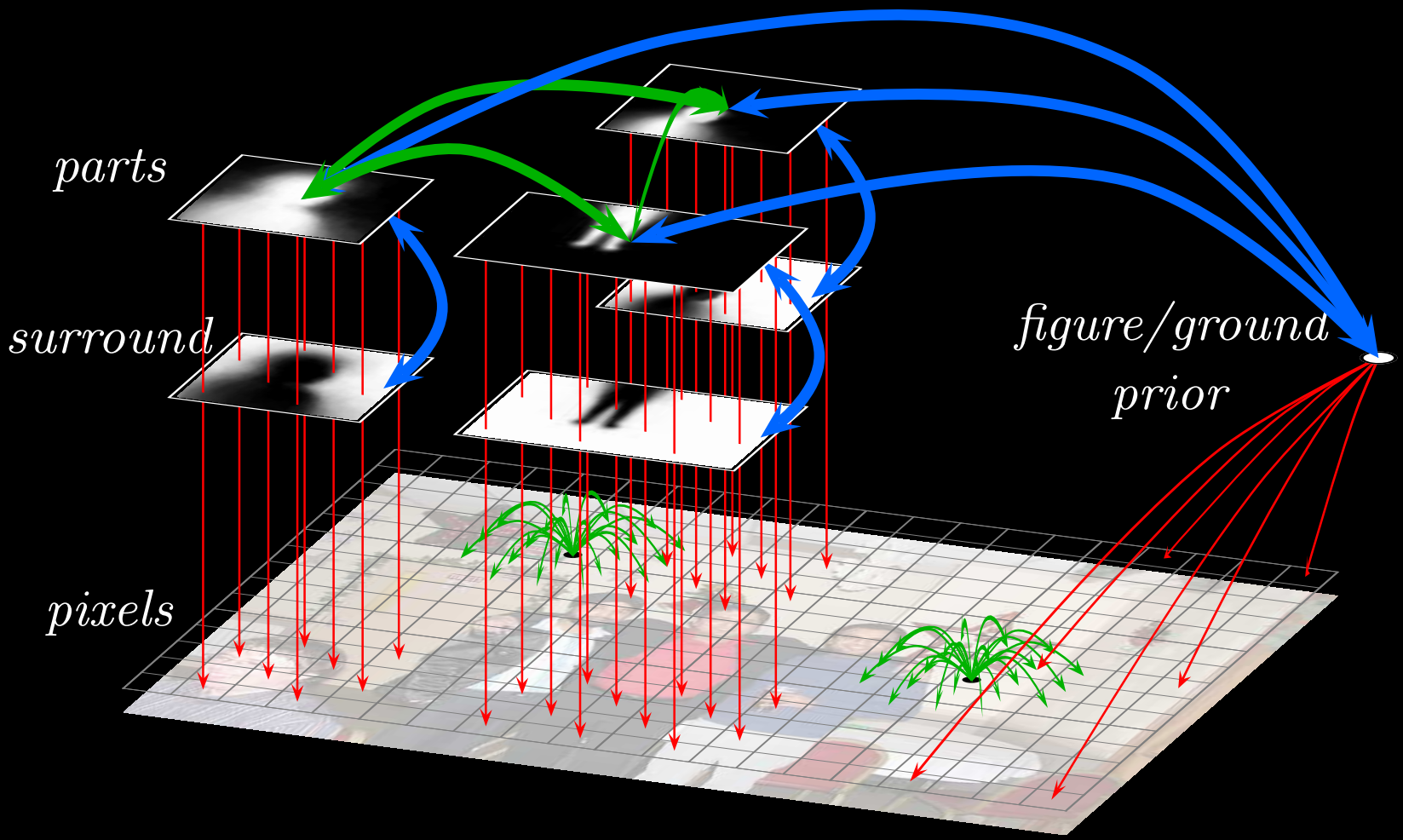

pixels

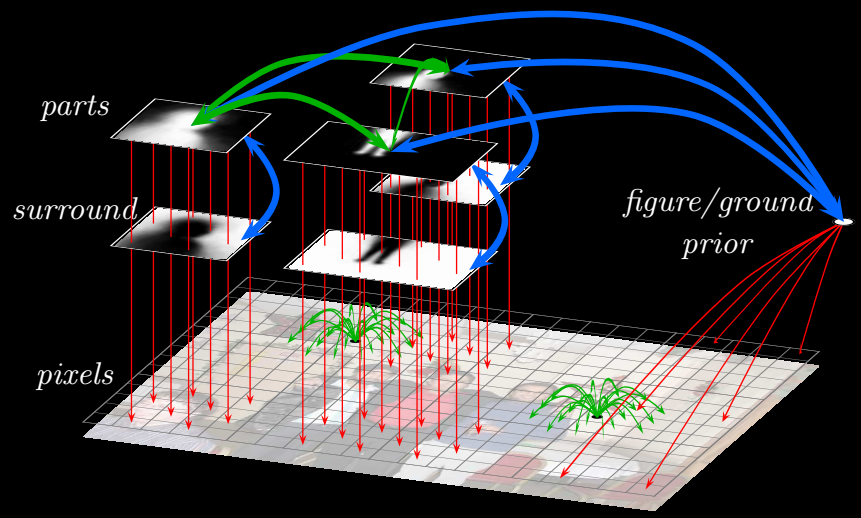

parts
pixels


parts
surround
pixels


parts
surround
pixels


parts
surround
pixels


parts
surround
pixels
figure/ground
prior

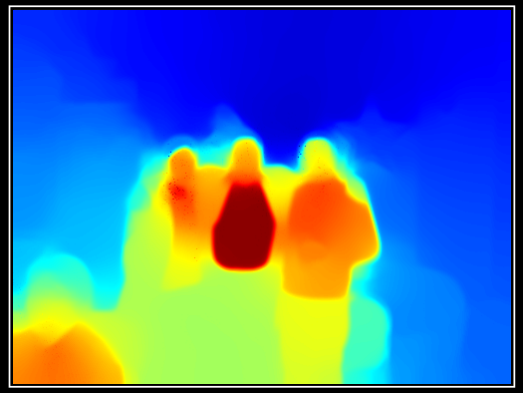


⇓

Angular Embedding

⇓ ⇓ ⇓

objects



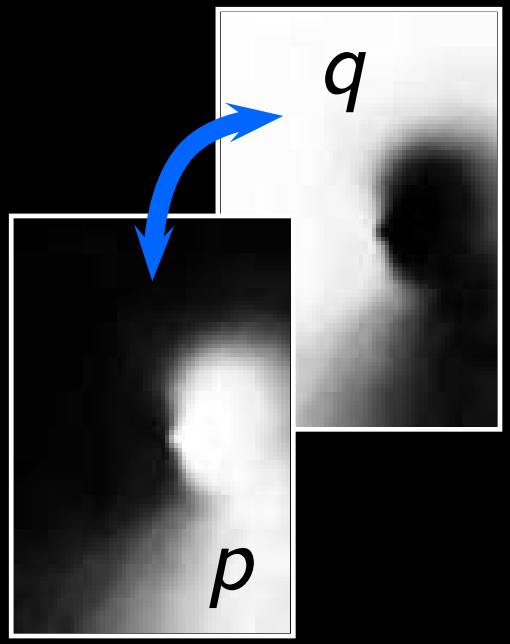
figure/ground

segmentation

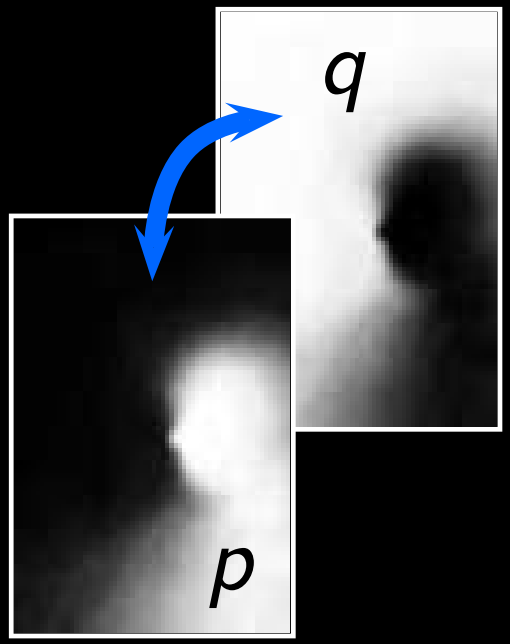
Angular Embedding

Angular Embedding

Angular Embedding



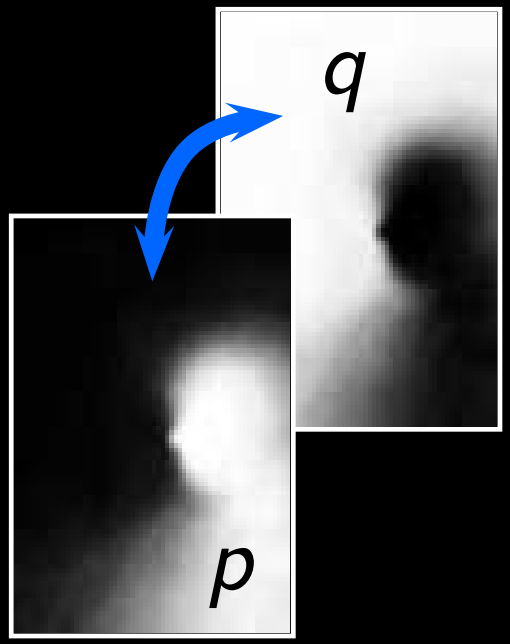
Angular Embedding



Given:

- Relative ordering $\Theta(\cdot, \cdot)$
- Confidence on relationships $C(\cdot, \cdot)$

Angular Embedding



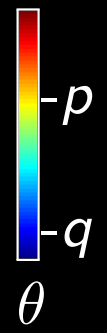
Given:

- Relative ordering $\Theta(\cdot,\cdot)$
- Confidence on relationships $C(\cdot,\cdot)$

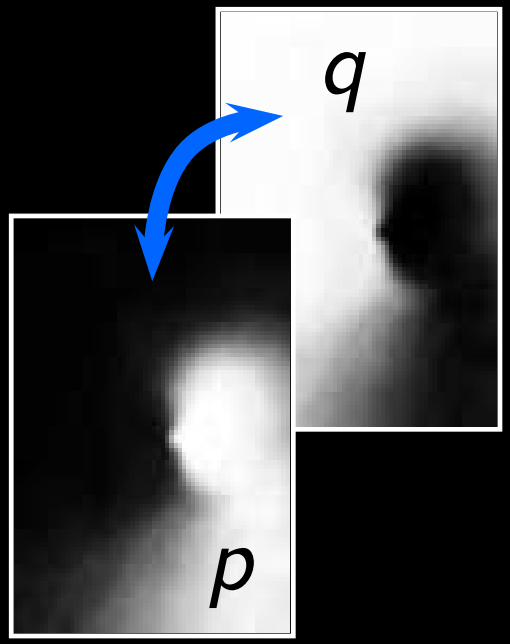
Compute:

- Global ordering $\theta(\cdot)$
- Embed into unit circle:

$$p \longrightarrow z(p) = e^{i\theta(p)}$$



Angular Embedding



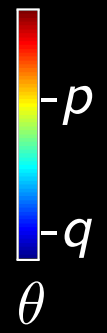
Given:

- Relative ordering $\Theta(\cdot,\cdot)$
- Confidence on relationships $C(\cdot,\cdot)$

Compute:

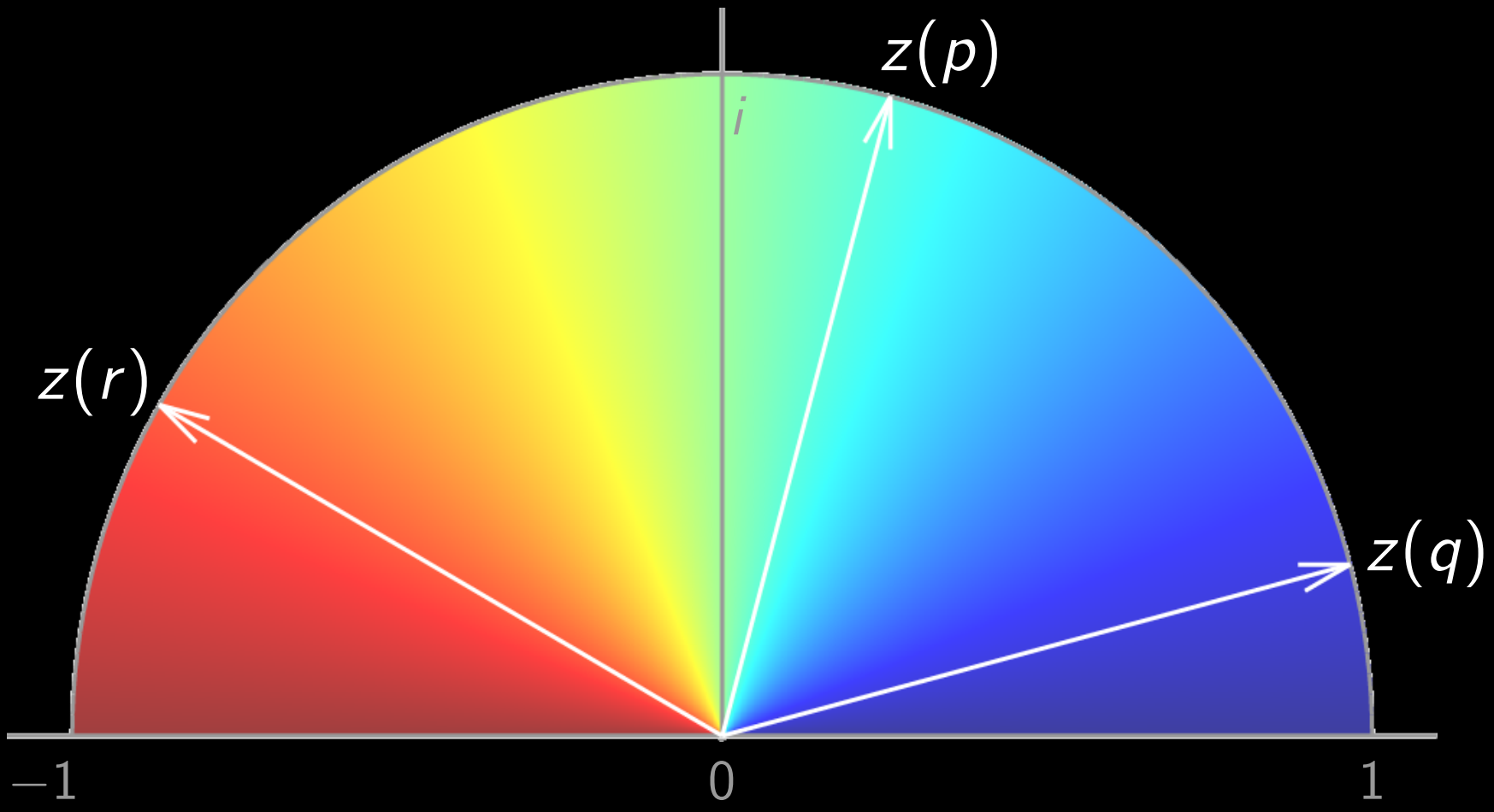
- Global ordering $\theta(\cdot)$
- Embed into unit circle:

$$p \longrightarrow z(p) = e^{i\theta(p)}$$

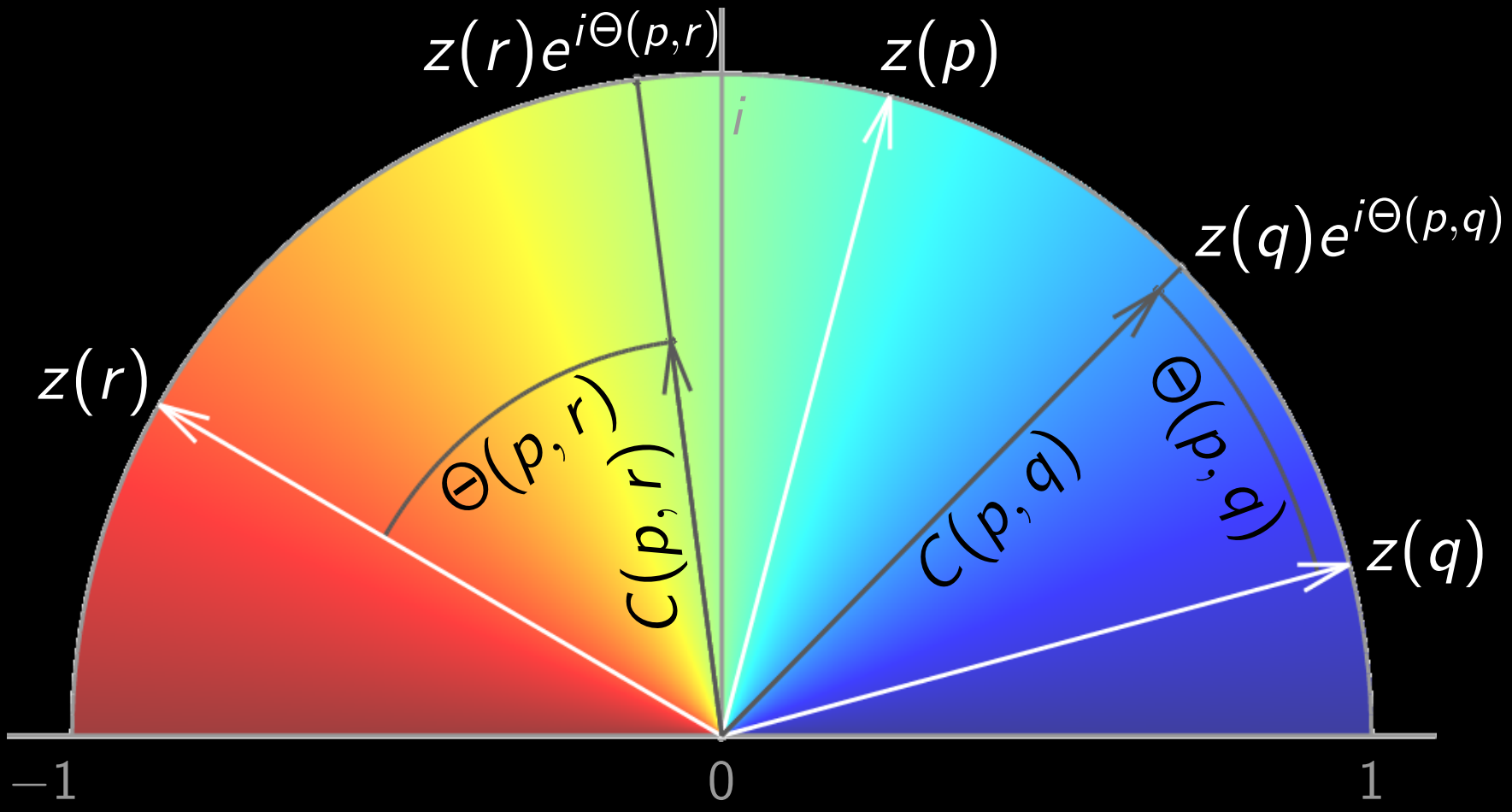


Subject to:

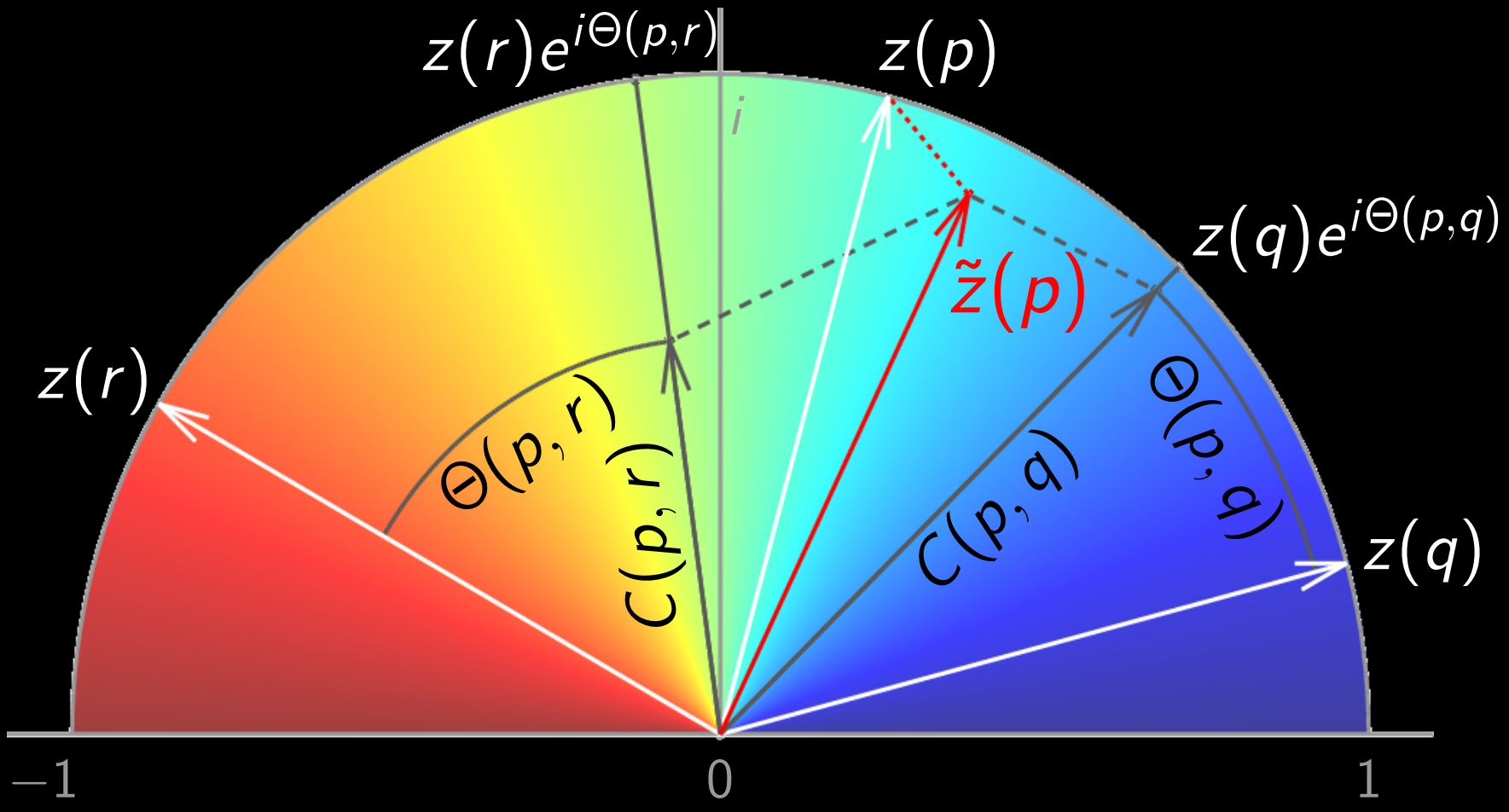
- Linear constraints on embedding solution in columns of U



[Yu, PAMI 2011]

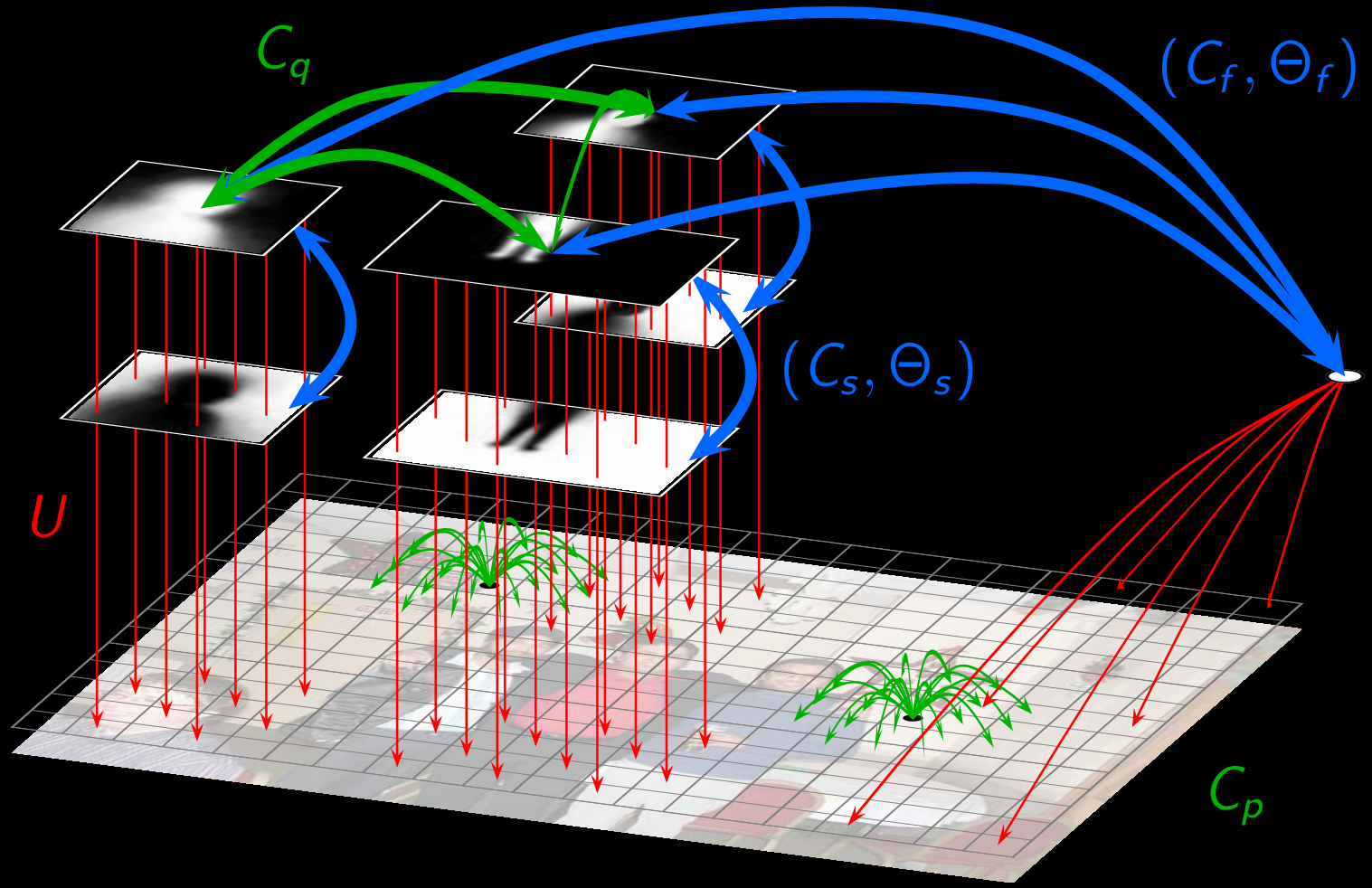

$z(r)e^{i\Theta(p,r)}$
$z(p)$
i
$z(q)e^{i\Theta(p,q)}$
$z(r)$
$\Theta(p,r)$
$C(p,r)$
$C(p,q)$
$\Theta(p,q)$
$z(q)$
-1
0
1

[Yu, PAMI 2011]



minimize: $\varepsilon = \sum_p \frac{\sum_q C(p,q)}{\sum_{p,q} C(p,q)} \cdot |z(p) - \tilde{z}(p)|^2$

[Yu, PAMI 2011]


C_q
(C_f, Θ_f)
(C_s, Θ_s)
U
C_p

$$
C = \begin{array}{c}
\begin{array}{cccc} \overbrace{}^{pixels} & \overbrace{}^{parts} & \overbrace{}^{surround} & \overbrace{}^{prior} \end{array} \\
\begin{bmatrix}
C_p & 0 & 0 & 0 \\
0 & \alpha \cdot C_q & \beta \cdot C_s & \gamma \cdot C_f \\
0 & \beta \cdot C_s^T & 0 & 0 \\
0 & \gamma \cdot C_f^T & 0 & 0
\end{bmatrix}
\end{array}
$$

$$
\Theta = \Sigma^{-1} \begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & \Theta_s & \Theta_f \\
0 & -\Theta_s^T & 0 & 0 \\
0 & -\Theta_f^T & 0 & 0
\end{bmatrix}
$$

Angular Embedding

Relax to generalized eigenproblem $QPQz = \lambda z$:

$$P = D^{-1}W$$
$$Q = I - D^{-1}U(U^T D^{-1}U)^{-1}U^T$$

with D and W defined as:

$$D = Diag(C1_n)$$
$$W = C \bullet e^{i\Theta}$$

Eigenvectors $\{z_0, z_1, ..., z_{m-1}\}$ embed pixels and parts into $\mathbb{C}^m$

Angular Embedding

$\angle z_0$ encodes global ordering

$z_1, z_2, \ldots, z_{m-1}$ encode grouping

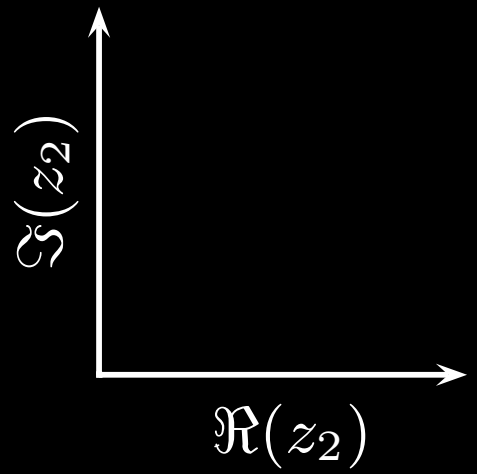
Angular Embedding

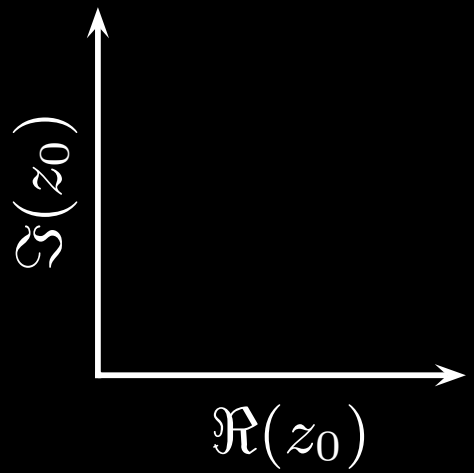
$\angle z_0$ encodes global ordering

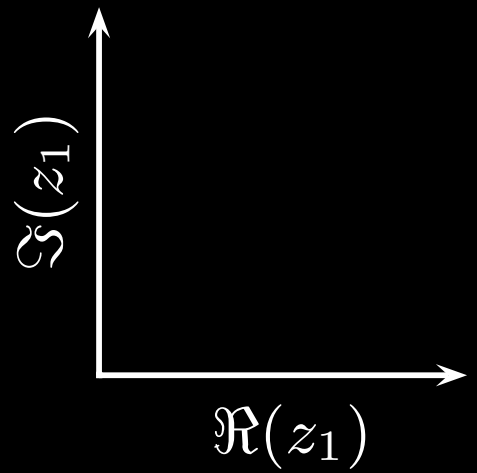
$z_1, z_2, \ldots, z_{m-1}$ encode grouping

if $\Theta = 0 \quad \Rightarrow \quad$ Normalized Cuts
(grouping without ordering)

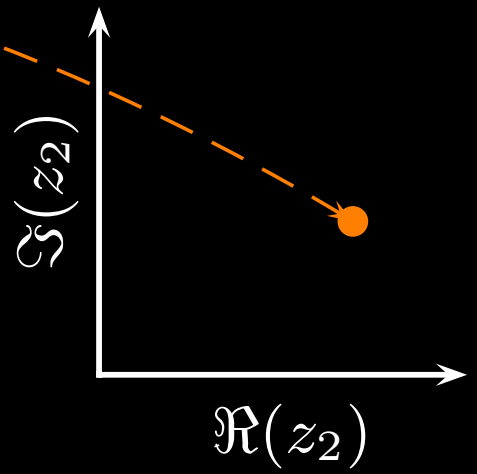
Decoding Eigenvectors: Object Detection

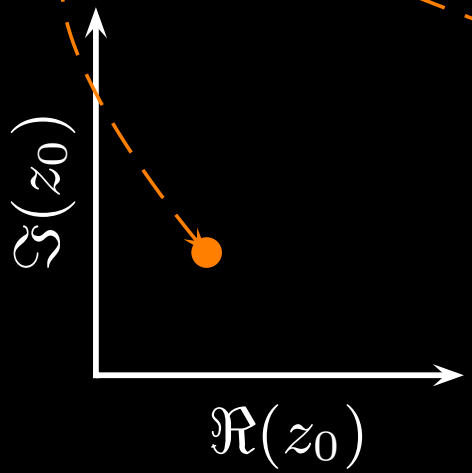


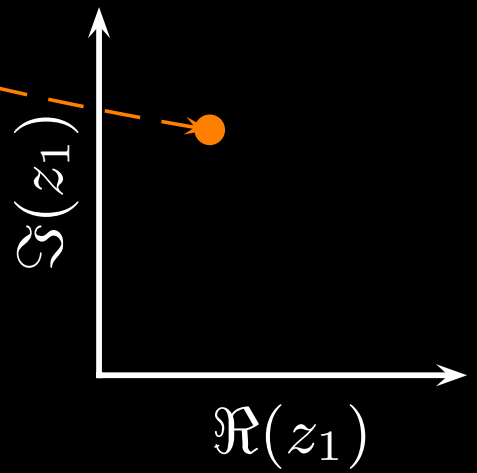




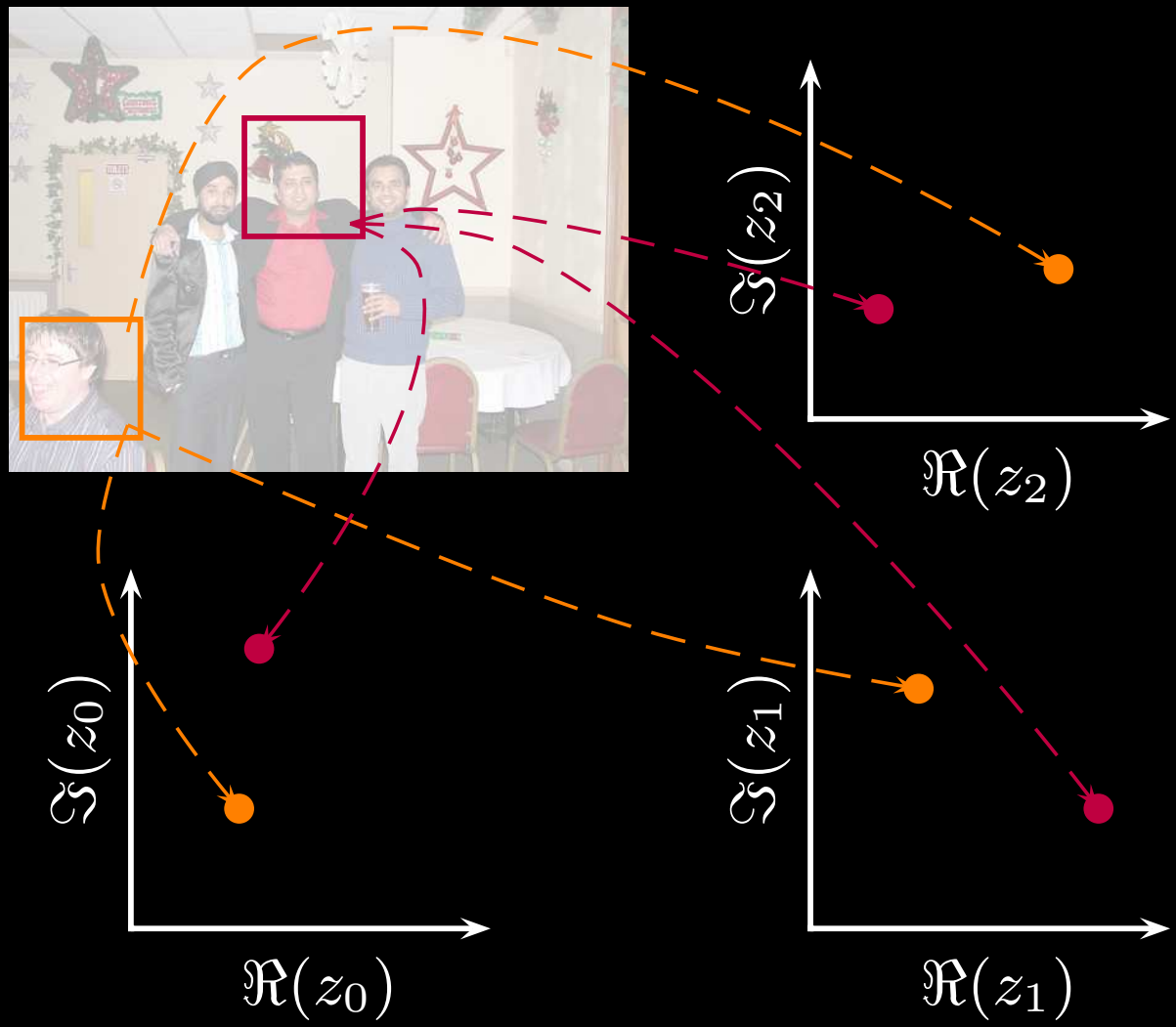
Decoding Eigenvectors: Object Detection



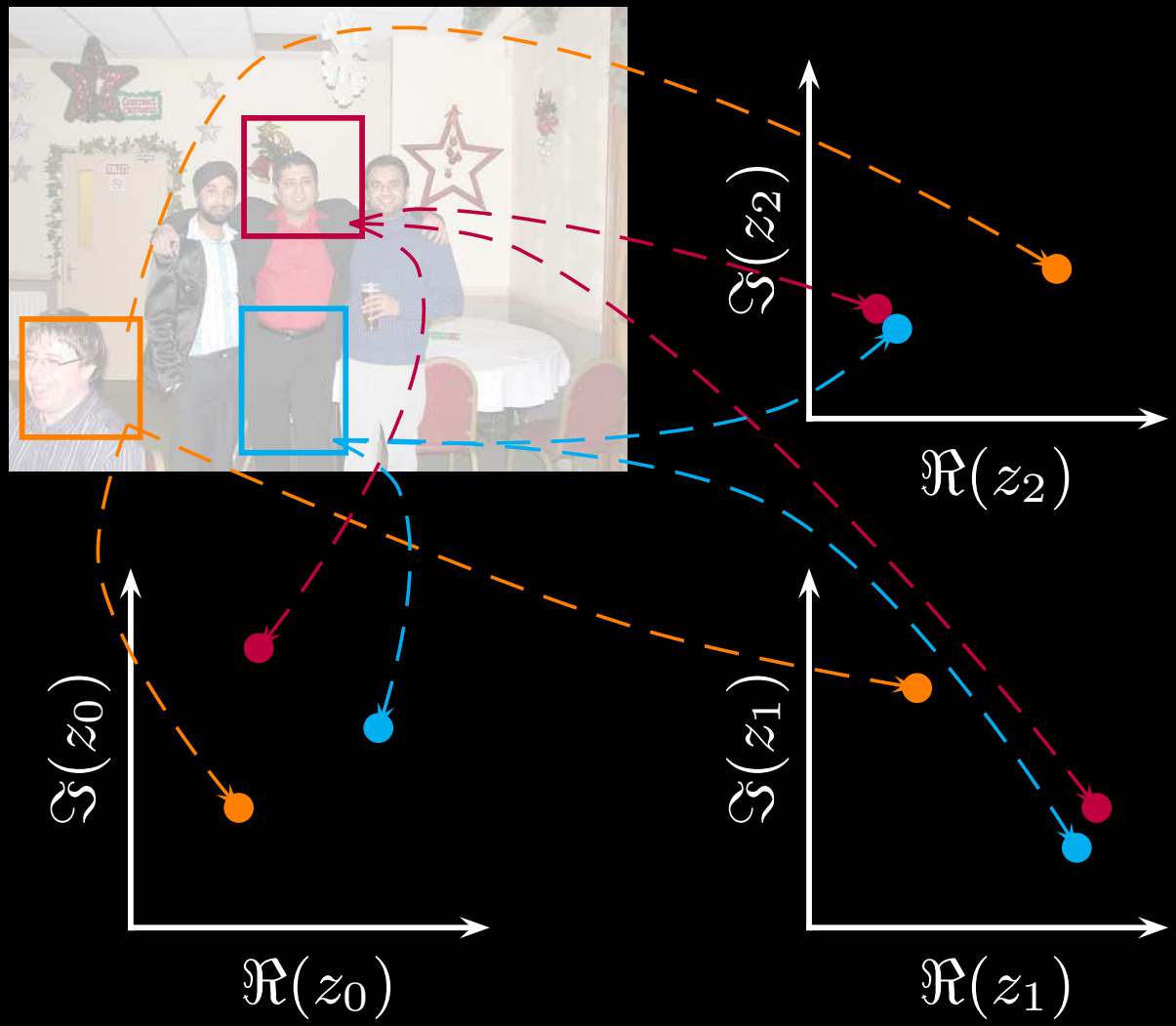




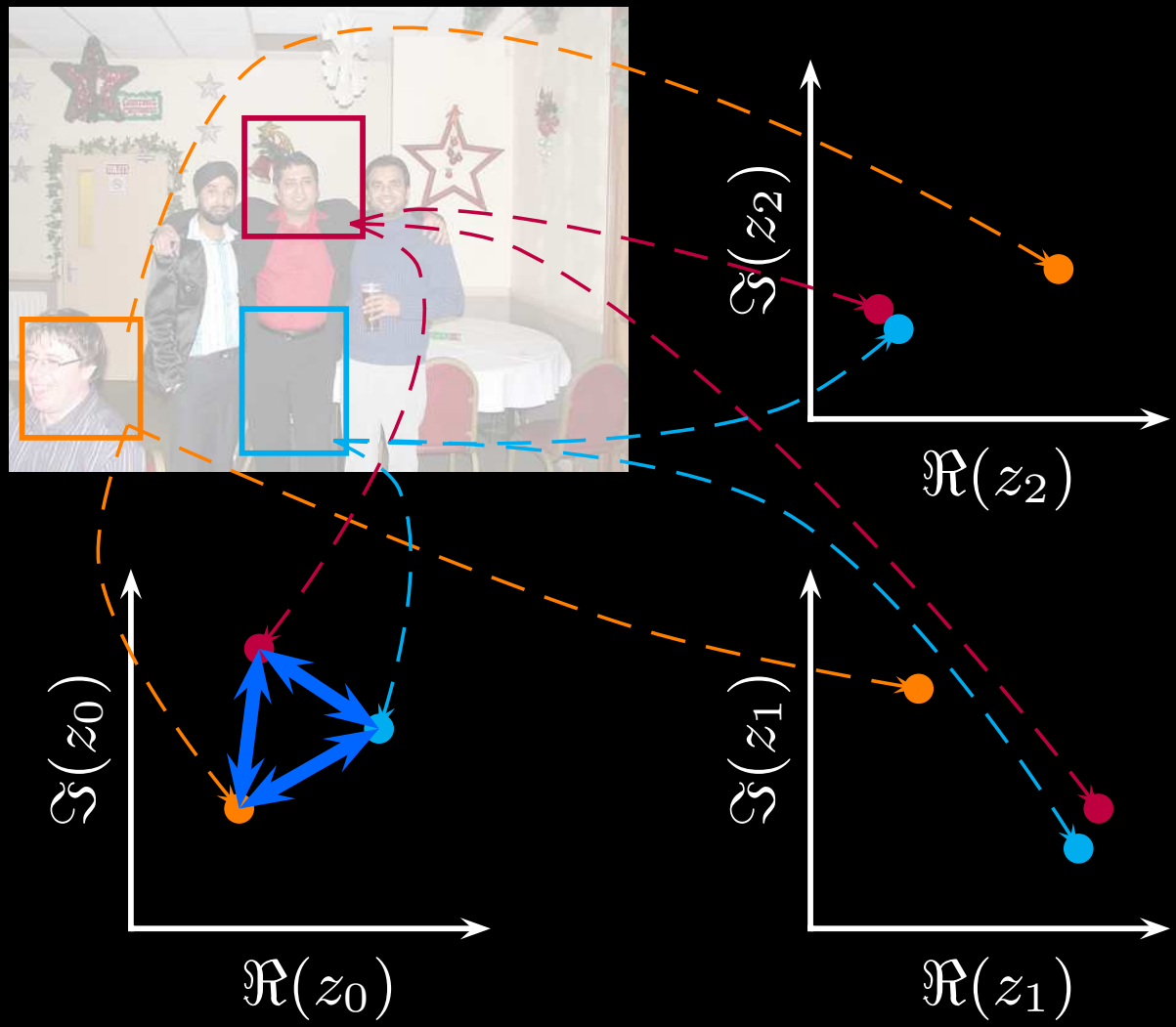
Decoding Eigenvectors: Object Detection



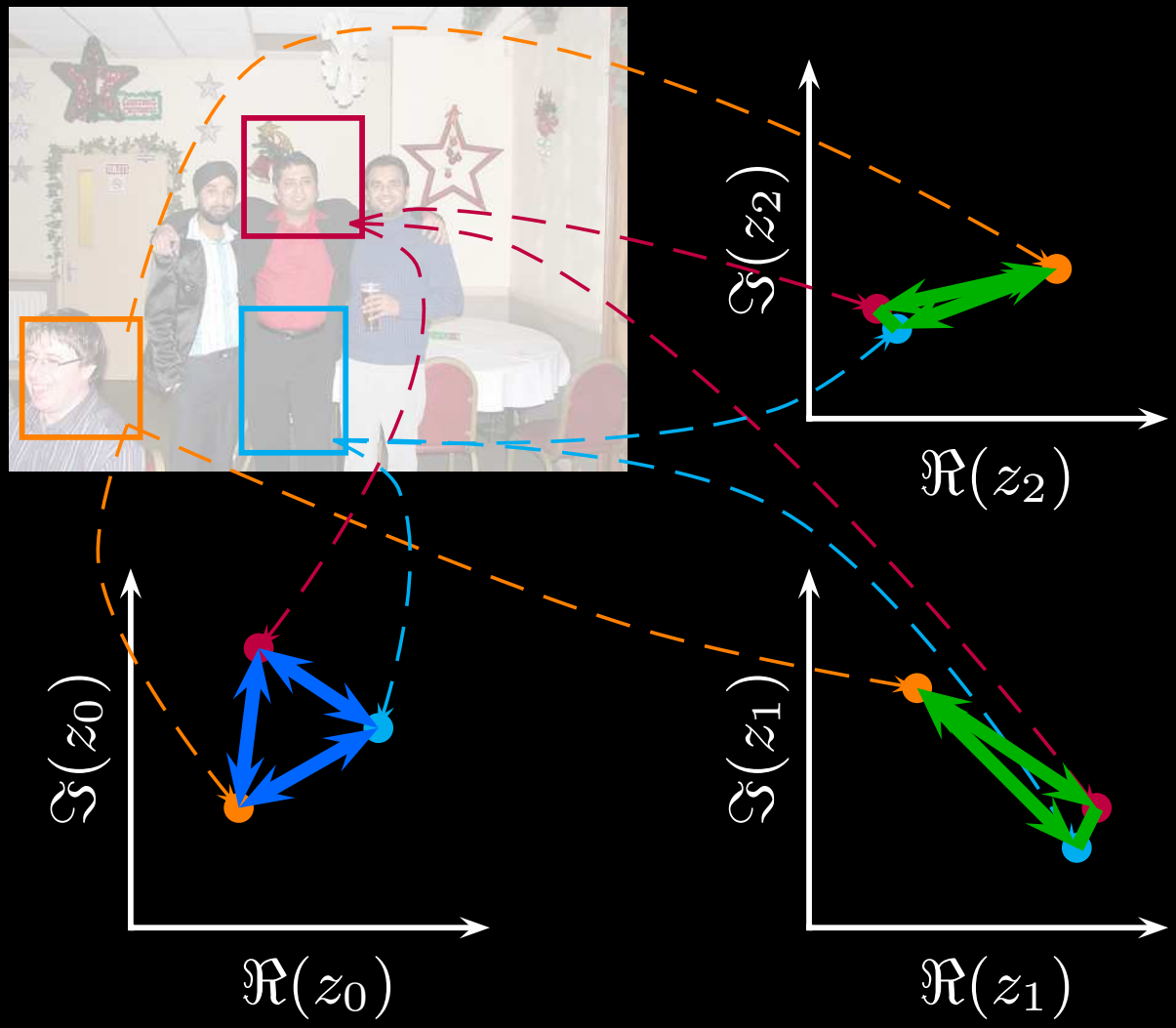
Decoding Eigenvectors: Object Detection



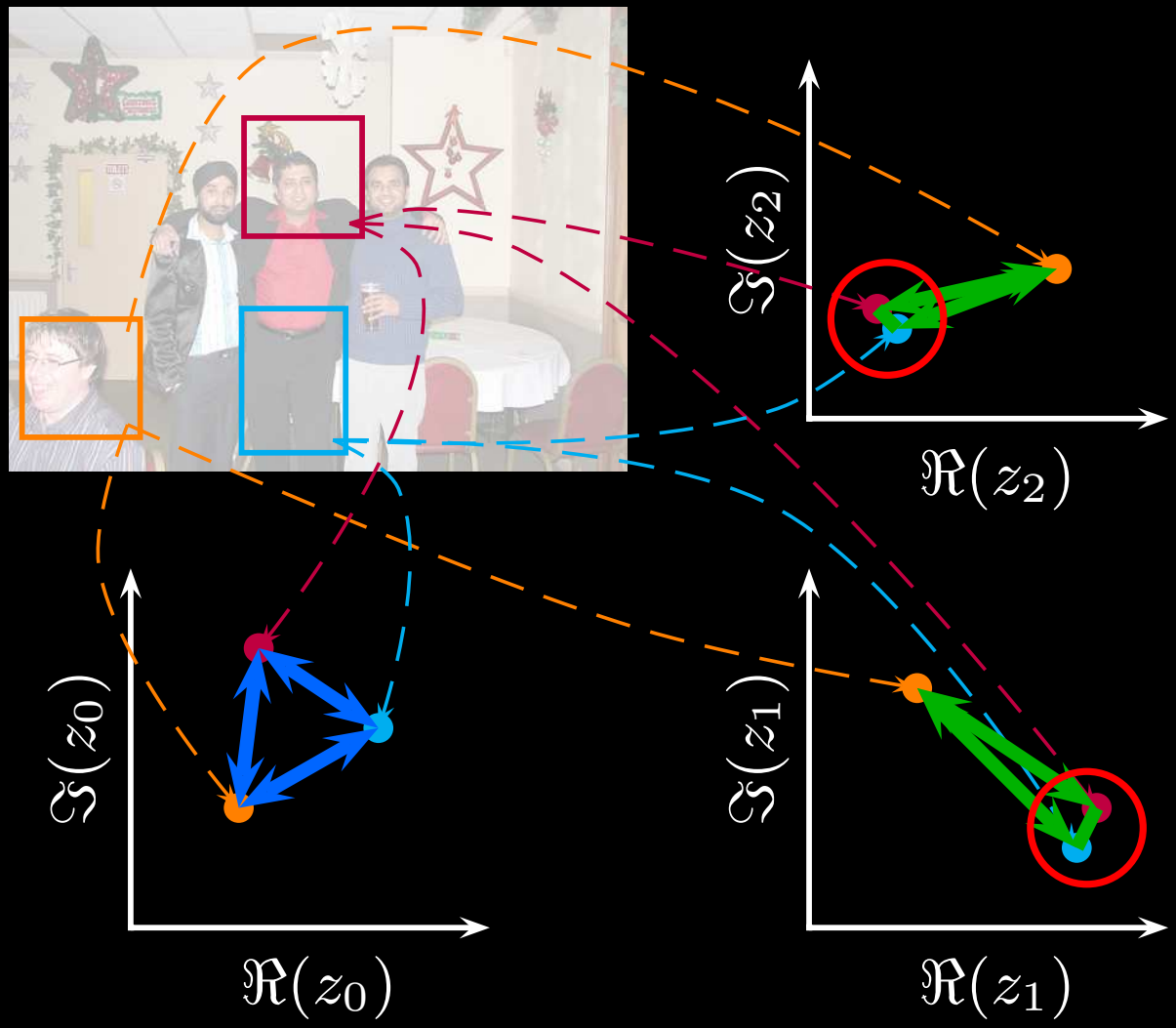
Decoding Eigenvectors: Object Detection



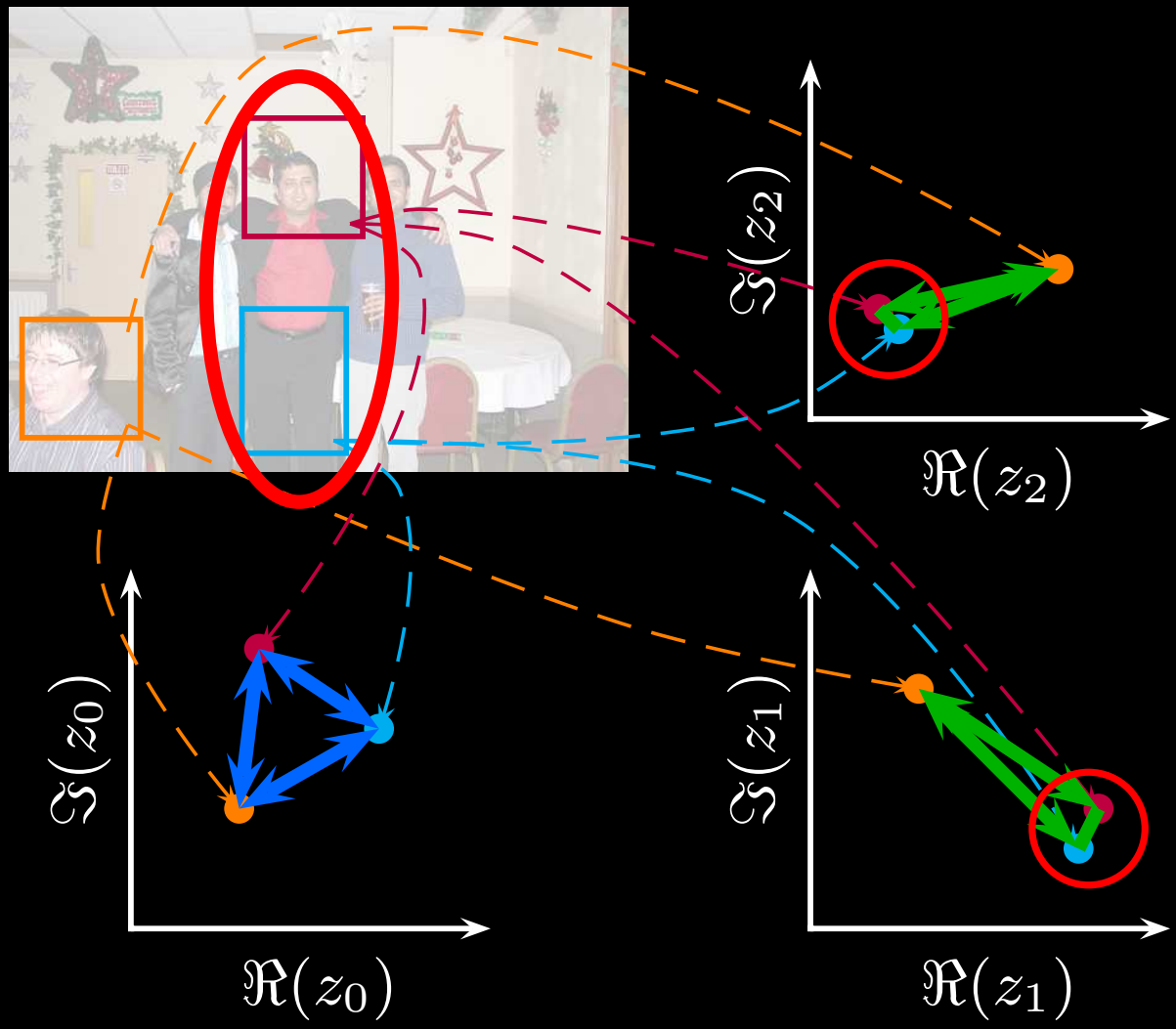
Decoding Eigenvectors: Object Detection



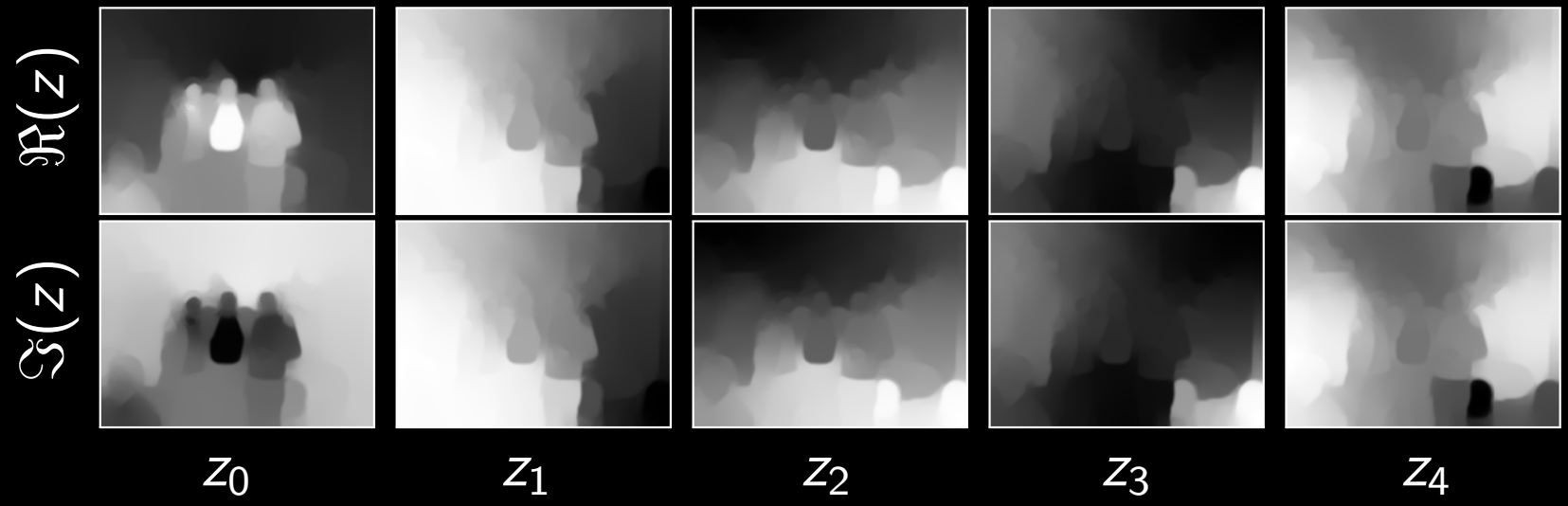
Decoding Eigenvectors: Object Detection



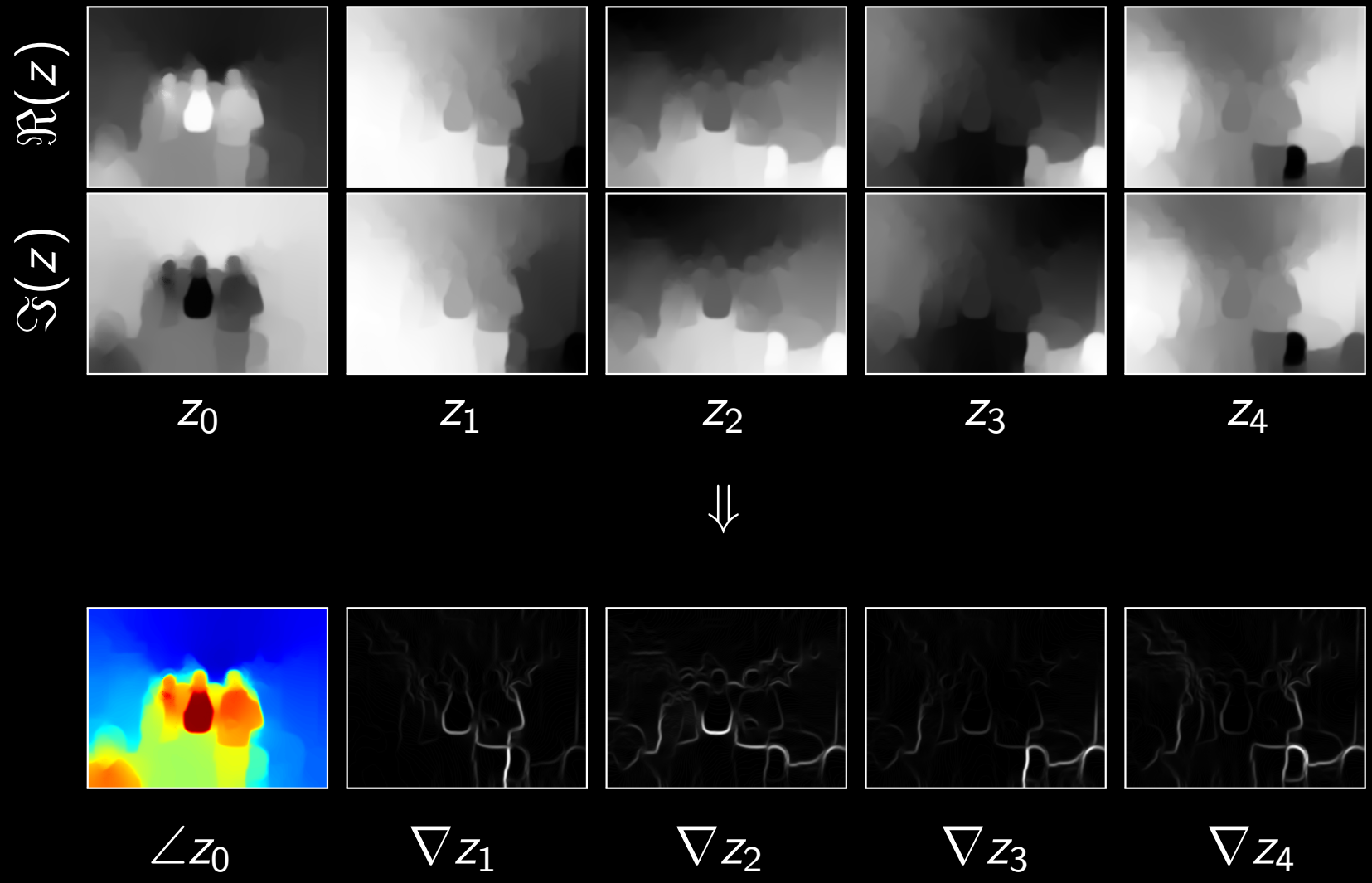
Decoding Eigenvectors: Object Detection



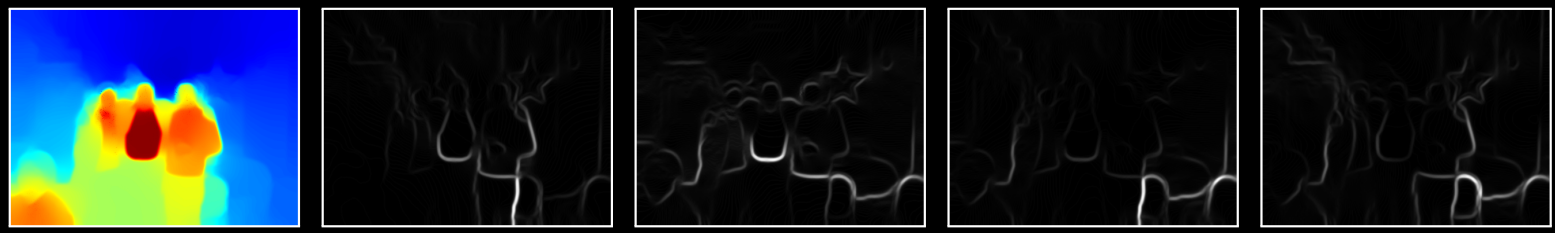
Decoding Eigenvectors: Figure/Ground



Decoding Eigenvectors: Figure/Ground



Decoding Eigenvectors: Segmentation



$\angle z_0$ ∇z_1 ∇z_2 ∇z_3 ∇z_4

Figure/Ground

Hierarchical Segmentation

[Arbeláez, Maire, Fowlkes, Malik, PAMI 2011]

Decoding Eigenvectors: Object Segmentation

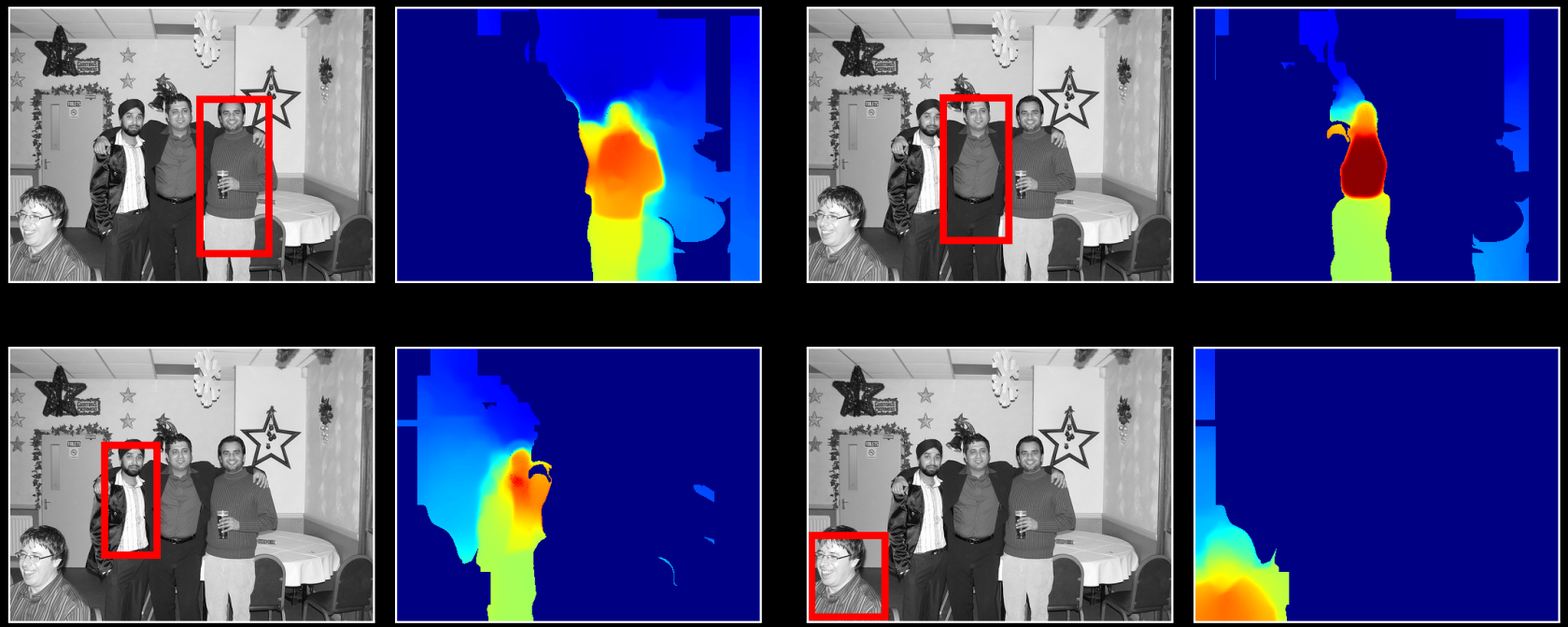
Assign pixels p_k to objects Q_i via parts q_j:

$$p_k \rightarrow \underset{Q_i}{\operatorname{argmin}} \left\{ \min_{q_j \in Q_i} \{ Dist(p_k, q_j) \} \right\}$$

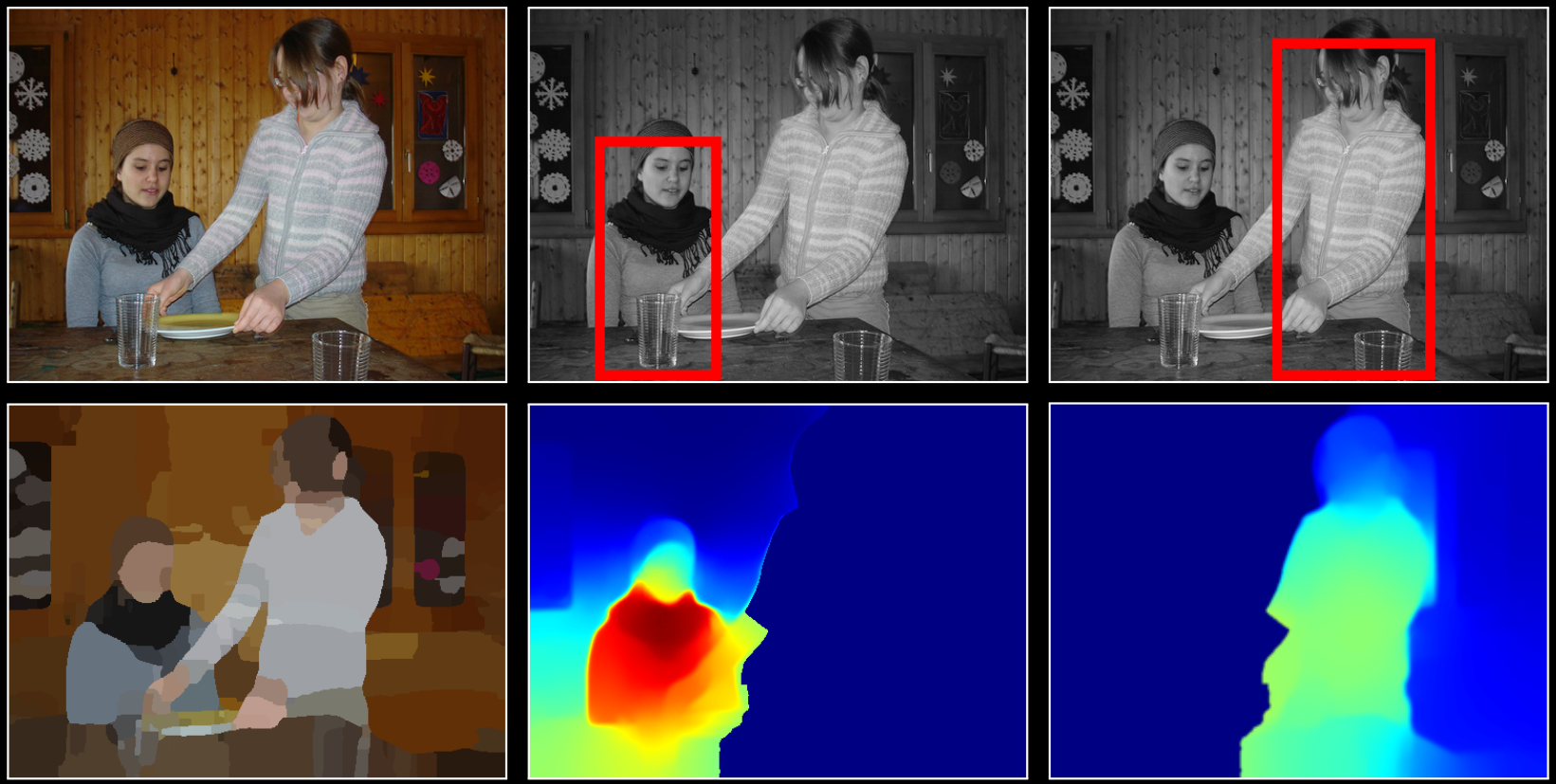
Decoding Eigenvectors: Object Segmentation

Assign pixels p_k to objects Q_i via parts q_j:

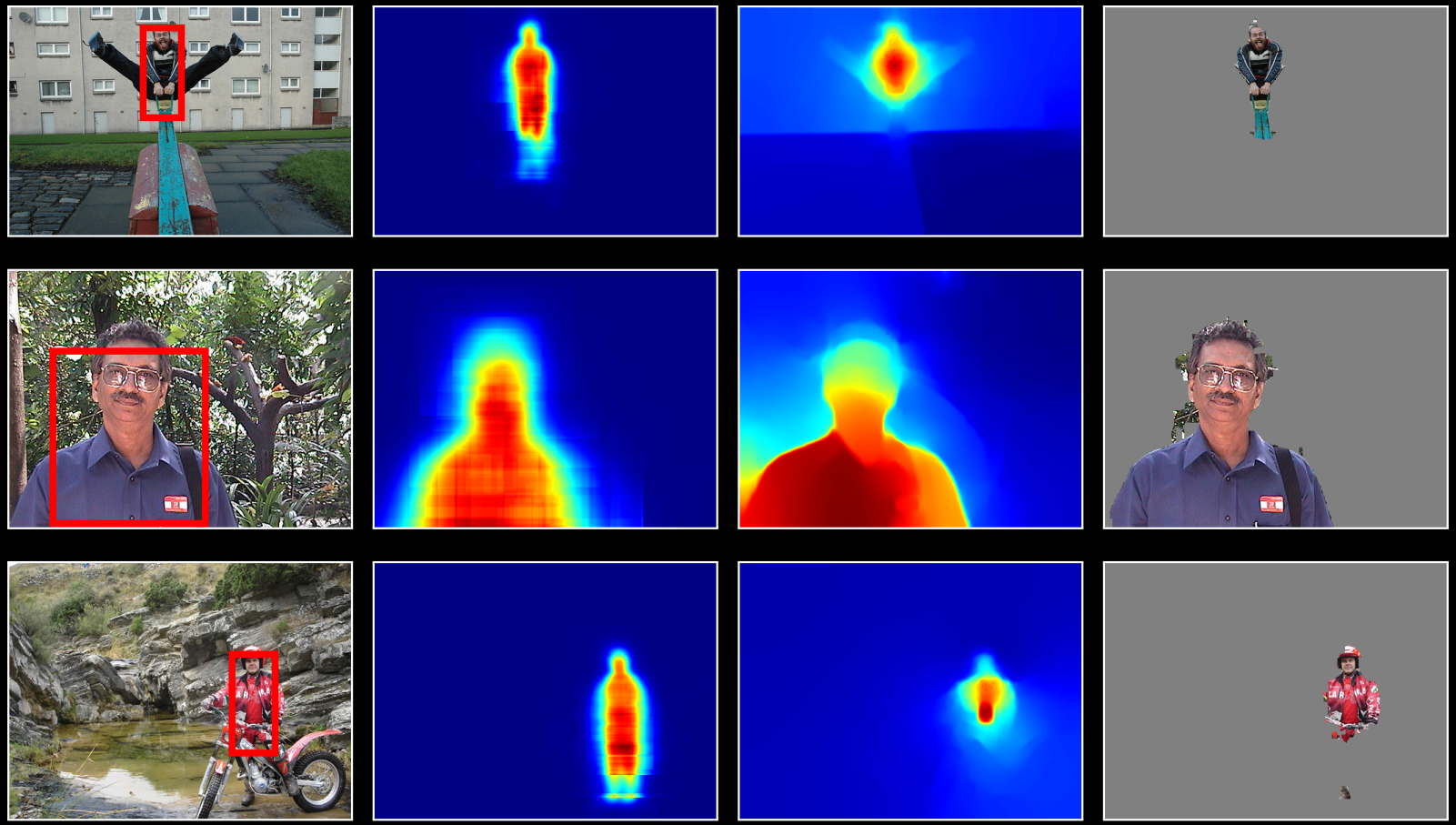
$$p_k \longrightarrow \underset{Q_i}{\operatorname{argmin}} \left\{ \min_{q_j \in Q_i} \{ Dist(p_k, q_j) \} \right\}$$



Decoding Eigenvectors

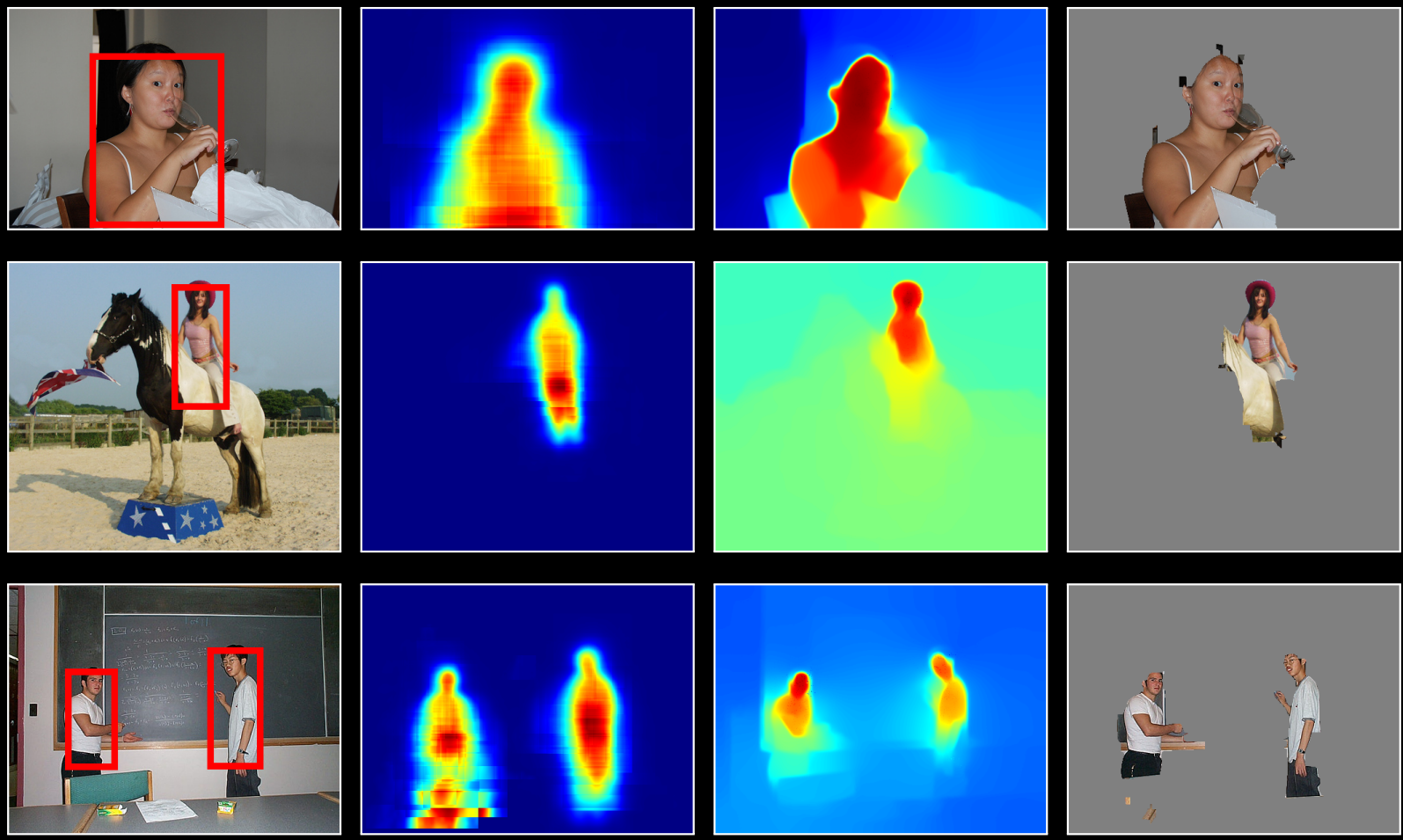


Results: PASCAL 2010 Person Category



Detections Poselet Mask F/G Mask Segmentation

Results: PASCAL 2010 Person Category



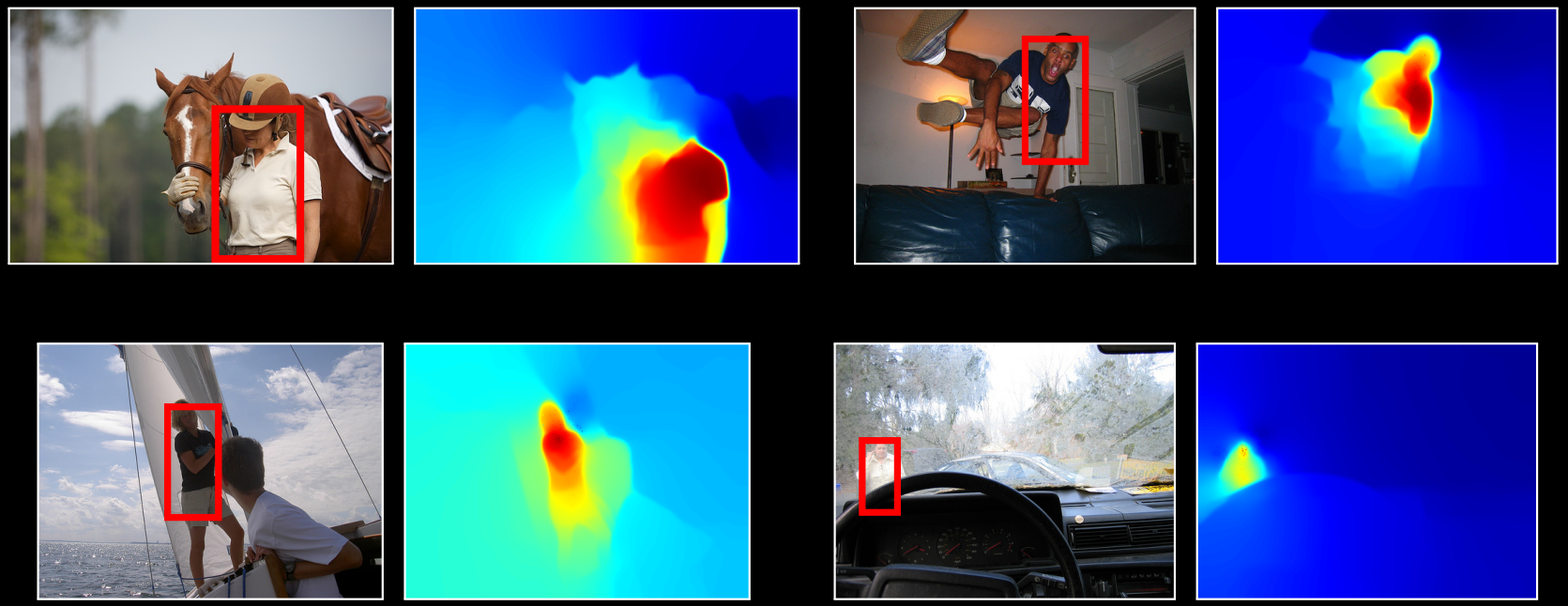
Detections Poselet Mask F/G Mask Segmentation

Results: PASCAL 2010 Person Category

- Segmentation task score: 41.1 (35.5 for poselet baseline)

Results: PASCAL 2010 Person Category

- Segmentation task score: 41.1 (35.5 for poselet baseline)
- 11% relative improvement due to better detection



Summary

- Simultaneous segmentation and detection:
  - Part detectors $\longrightarrow$ figure pop-out, object grouping
  - Color, texture $\longrightarrow$ pixel grouping

Summary

- Simultaneous segmentation and detection:
  - Part detectors $\longrightarrow$ figure pop-out, object grouping
  - Color, texture $\longrightarrow$ pixel grouping
- Graph:
  - Parts and pixels as nodes
  - Links encode multiple relationship types

Summary

- Simultaneous segmentation and detection:
  - Part detectors $\longrightarrow$ figure pop-out, object grouping
  - Color, texture $\longrightarrow$ pixel grouping
- Graph:
  - Parts and pixels as nodes
  - Links encode multiple relationship types
- Embedding: graph nodes $\longrightarrow \mathbb{C}^m$

Summary

- Simultaneous segmentation and detection:
  - Part detectors $\longrightarrow$ figure pop-out, object grouping
  - Color, texture $\longrightarrow$ pixel grouping
- Graph:
  - Parts and pixels as nodes
  - Links encode multiple relationship types
- Embedding: graph nodes $\longrightarrow \mathbb{C}^m$
- Decode:
  - Figure/ground
  - Image segmentation
  - Detected objects
  - Segmentation of each object instance

Summary

- Simultaneous segmentation and detection:
  - Part detectors $\longrightarrow$ figure pop-out, object grouping
  - Color, texture $\longrightarrow$ pixel grouping
- Graph:
  - Parts and pixels as nodes
  - Links encode multiple relationship types
- Embedding: graph nodes $\longrightarrow \mathbb{C}^m$
- Decode:
  - Figure/ground
  - Image segmentation
  - Detected objects
  - Segmentation of each object instance
- Better person detection and segmentation on PASCAL