Scheduling Top-Down and Bottom-Up Processes

Song-Chun Zhu, Sinisa Todorovic, and Ales Leonardis

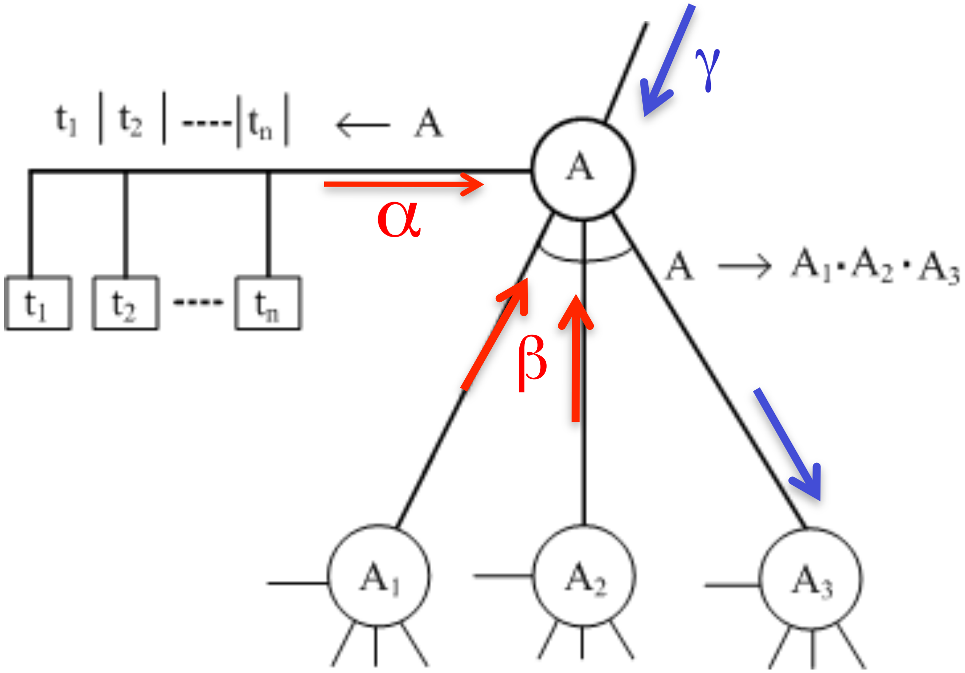
At CVPR, Providence, Rhode Island
June 16, 2012

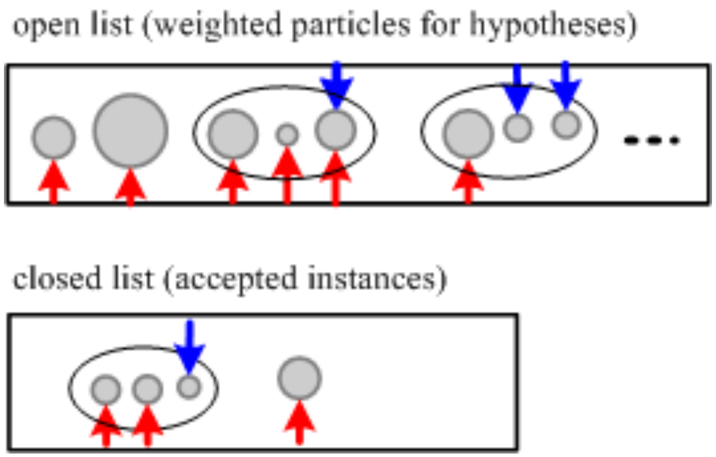
α, β and γ computing processes in AoG

The And-Or graph is a recursive structure.   So, consider a node A.

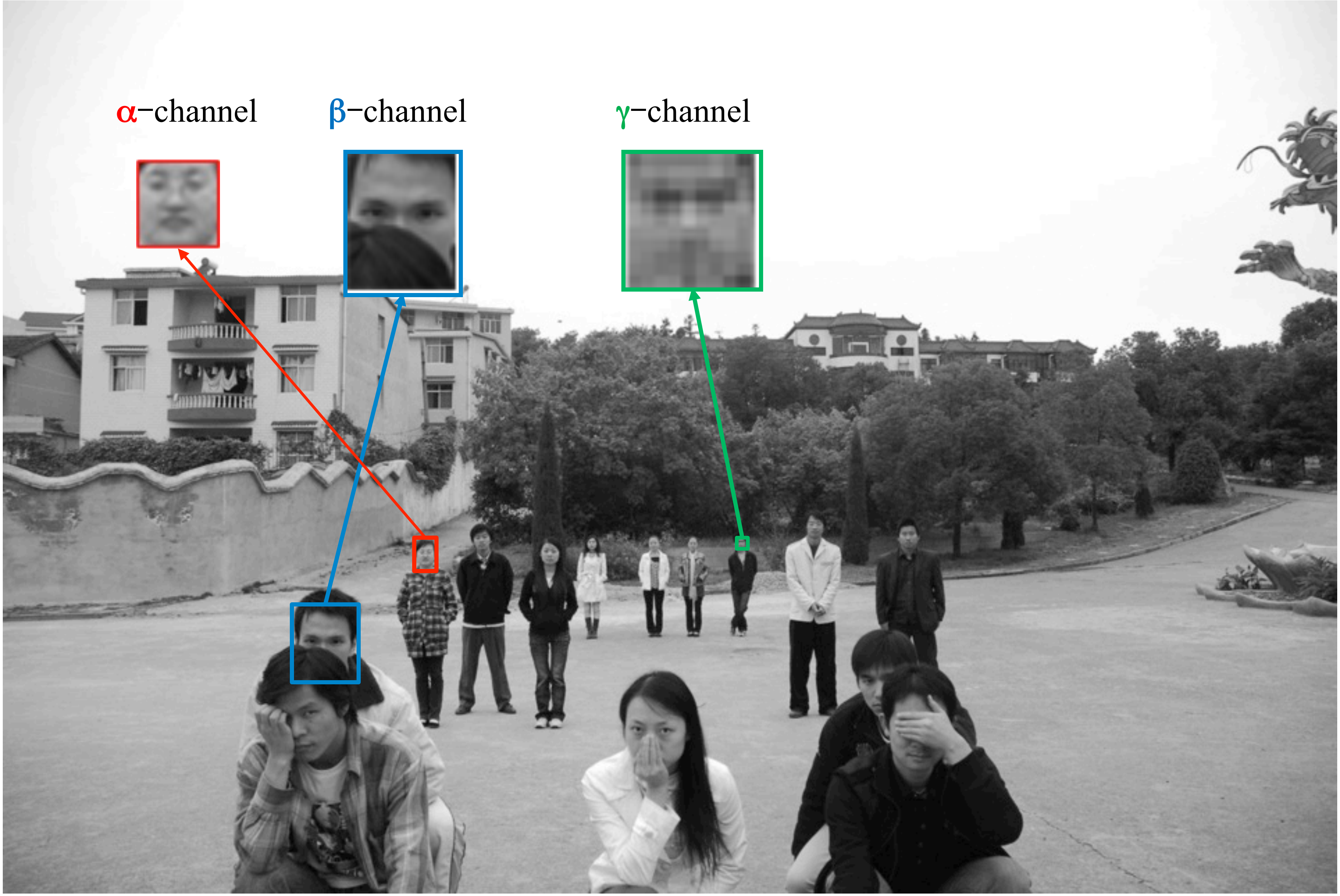
1. Node A at a coarse scale terminates to leaf nodes (ground)
2. Node A is connected to the root.

Estimate the $\alpha/\beta/\gamma$ channels when they are applicable ---an optimal scheduling problem


$t_1 | t_2 | ----|t_n| \leftarrow A$
γ
A
α
$A \rightarrow A_1 \cdot A_2 \cdot A_3$
t_1
t_2
----
t_n
β
A_1
A_2
A_3


open list (weighted particles for hypotheses)
...
closed list (accepted instances)

Example: Human Faces are Computed in 3 Channels



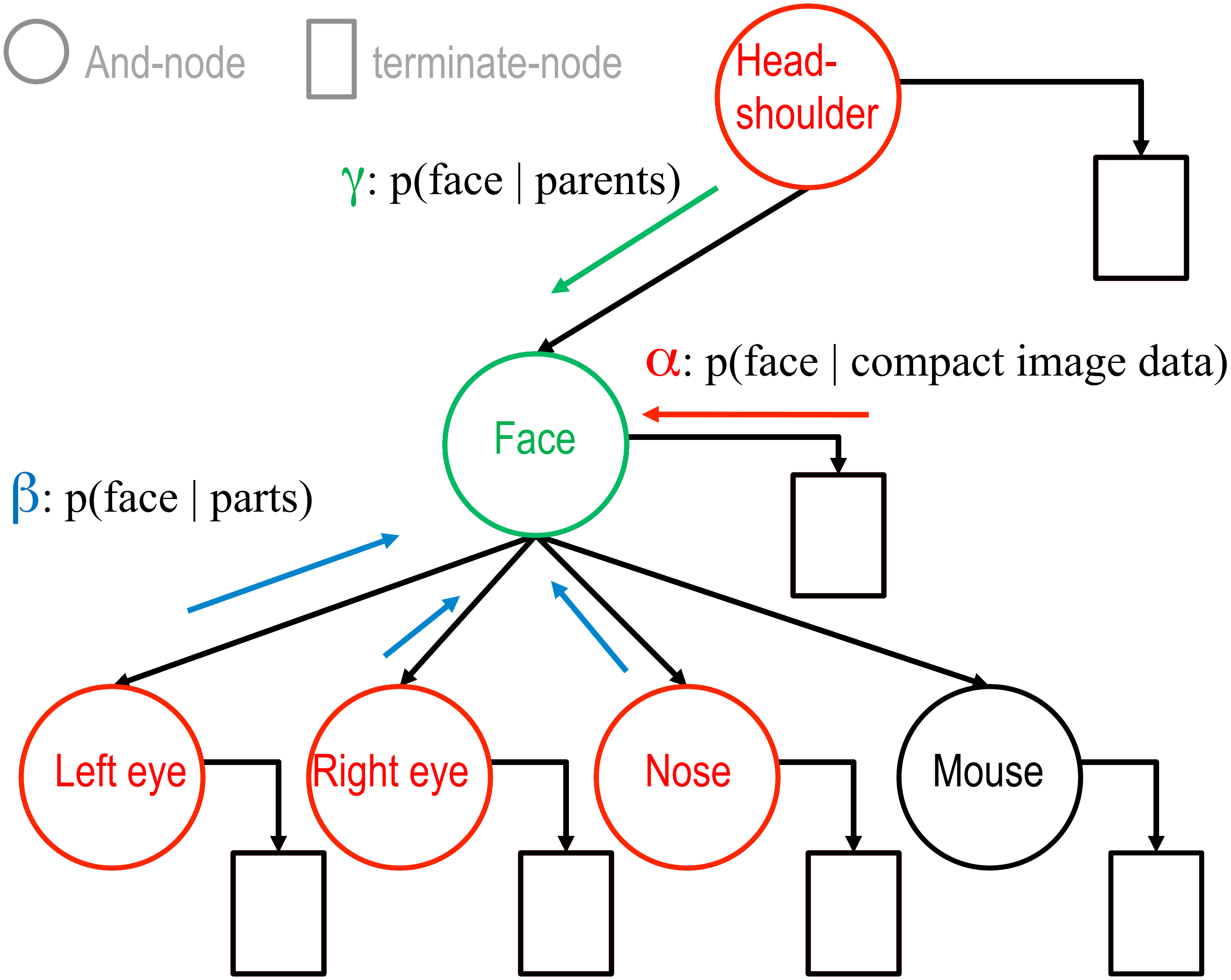
Human faces in real scenarios

α-channel

β-channel

γ-channel

Hierarchical modeling and α, β and γ computing

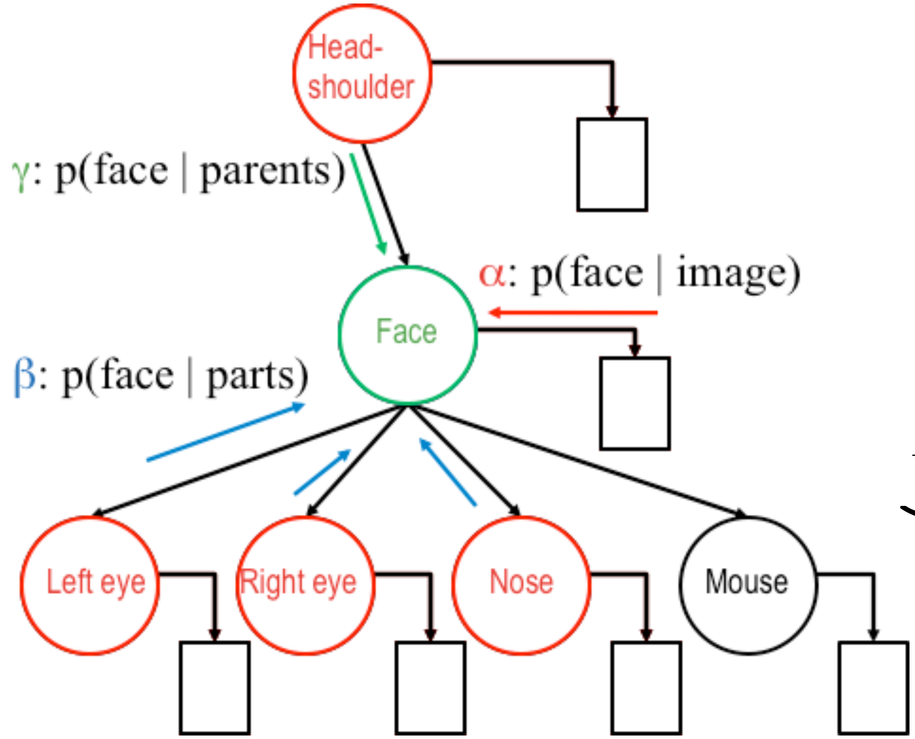


1. Each node has its own α, β and γ computing processes.

2. **How much does each channel contribute?**

Wu & Zhu IJCV11

Hierarchical modeling and α, β and γ computing



$$pg^{l*} = \arg\max_{pg} \Big[\underbrace{\log p(\wedge^l | \vee^l) + p(N) \sum_{i=1}^{N} \log p(X(\wedge_i^{l+}) | X(\wedge^l))}_{\text{parse graph connectivity}}$$

$$\underbrace{+ \ \log \frac{p(\Delta(t(\wedge^l))|t(\wedge^l))}{q(\Delta(t(\wedge^l)))}}_{\alpha \quad = \text{detector}}$$

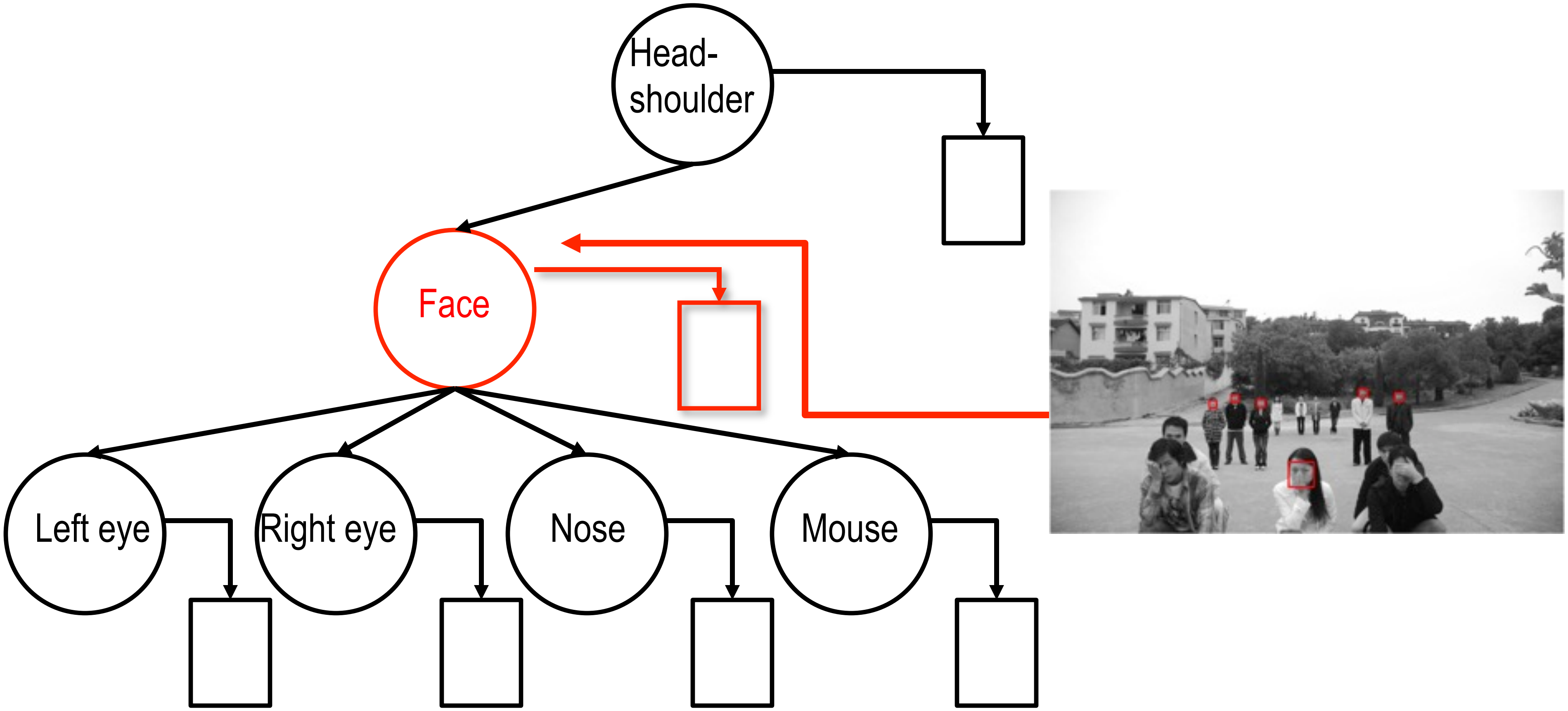
$$\underbrace{+ \ p(N)\Big[\sum_{i=1}^{N} \log \frac{p(\Delta(t(\wedge_i^{l+}))|t(\wedge_i^{l+}))}{q(\Delta(t(\wedge_i^{l+})))} + \sum_{i \neq j} \log p(X(\wedge_i^{l+}), X(\wedge_j^{l+}))\Big]}_{\beta \quad = \text{zoom-in}}$$

$$+ \ \underbrace{\log \frac{p(\Delta(t(\wedge^{l-}))|t(\wedge^{l-}))}{q(\Delta(t(\wedge^{l-}))} + \log p(X(\wedge^l)|X(\wedge^{l-}))}_{\gamma \quad = \text{zoom-out}}$$

Wu & Zhu IJCV11

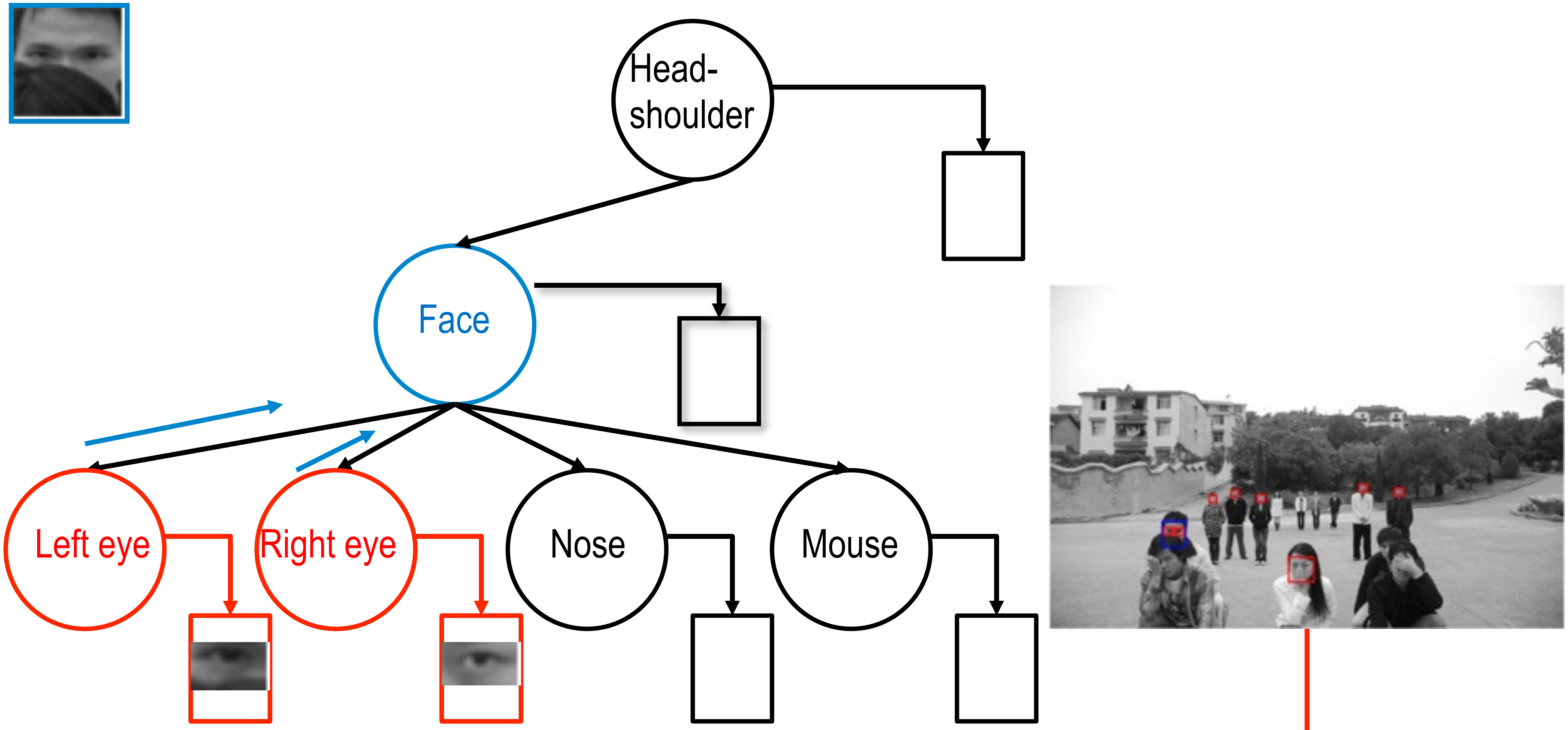
α processes for the face node

α-channels: p(face | compact image data)



β processes for the face node(when its α is off)

β-channels: p(face | parts), binding



Head-shoulder

Face

Left eye

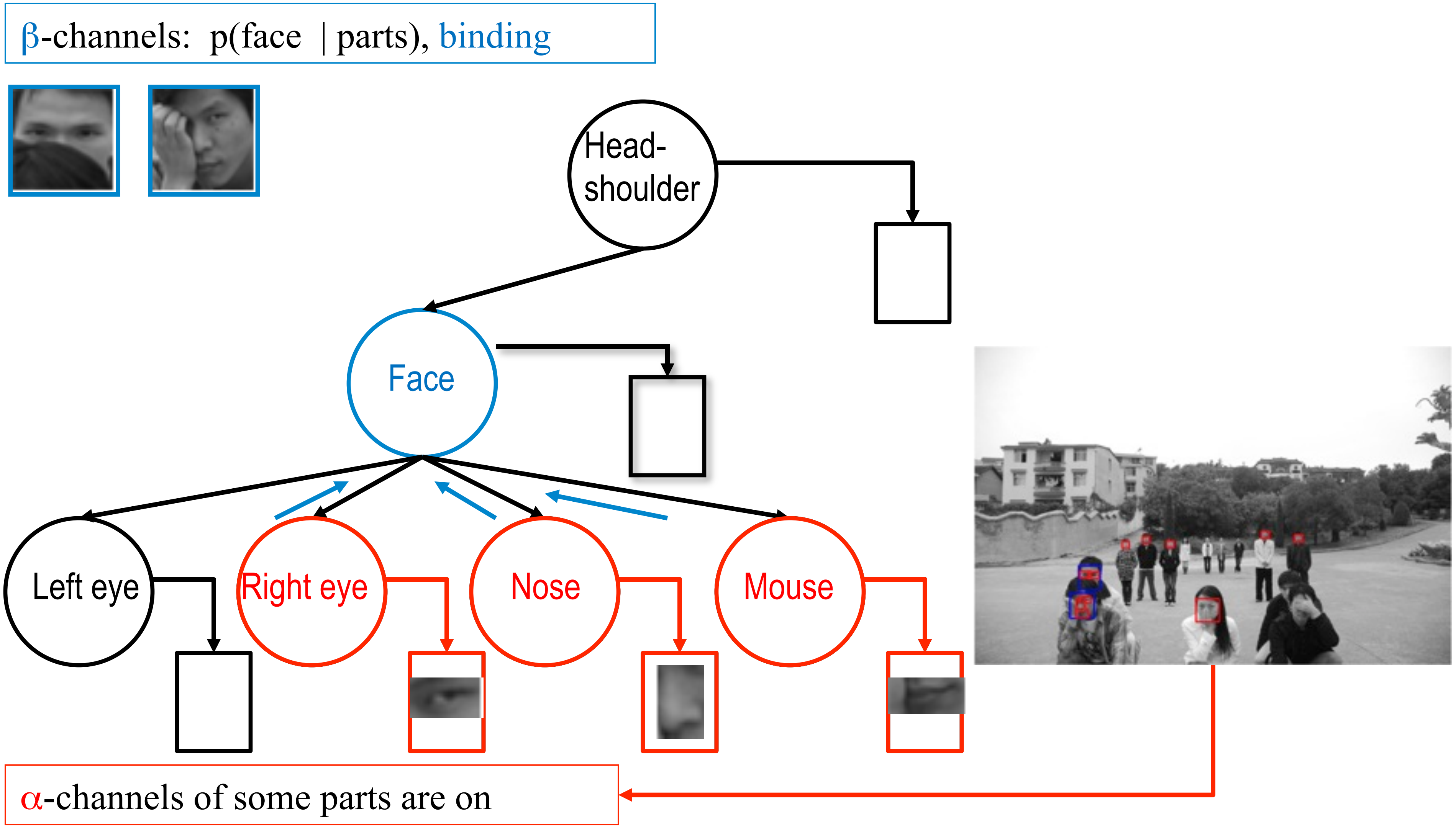
Right eye

Nose

Mouse

α-channels of some parts are on

β processes for the face node(when its α is off)

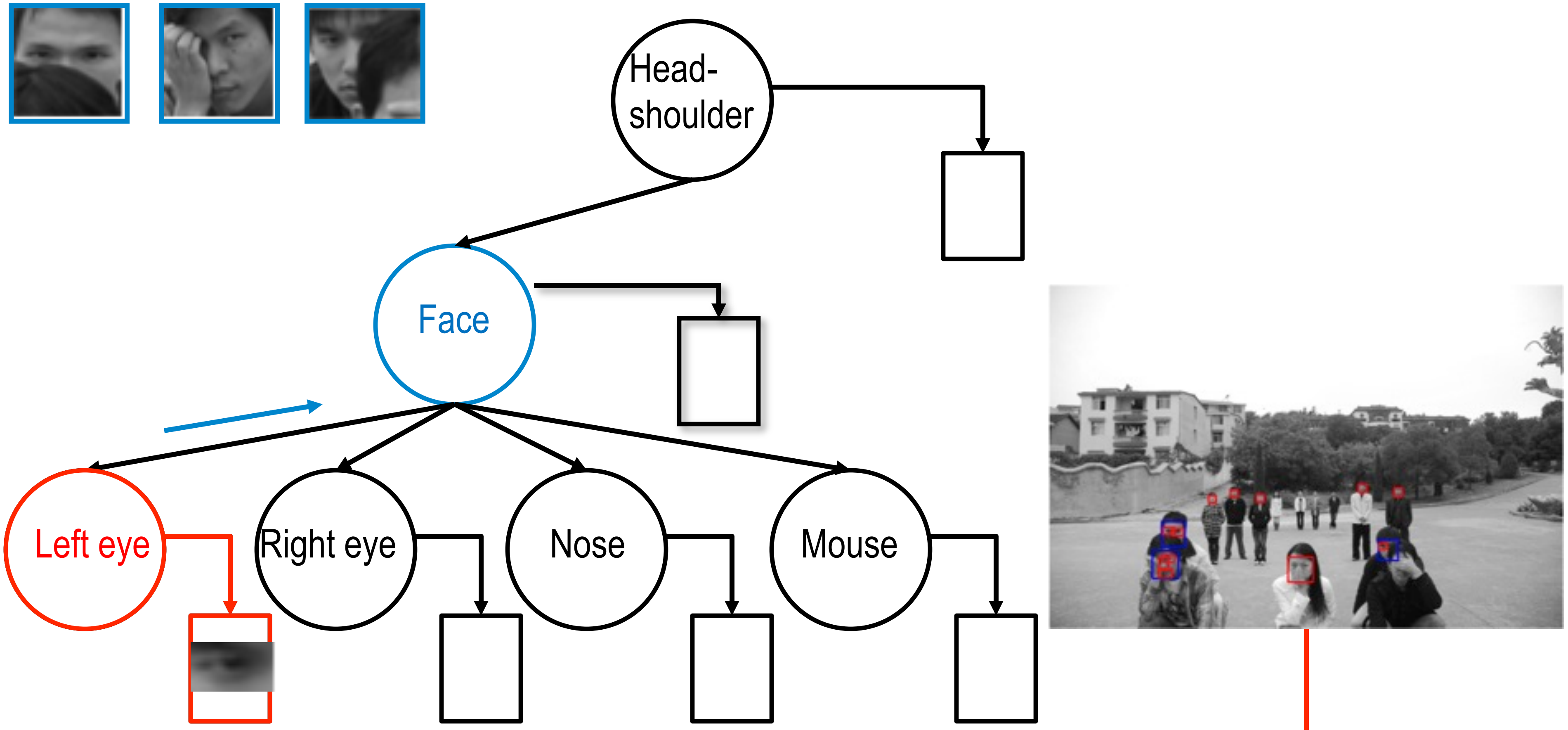


β-channels: p(face | parts), binding

α-channels of some parts are on

β processes for the face node(when its α is off)

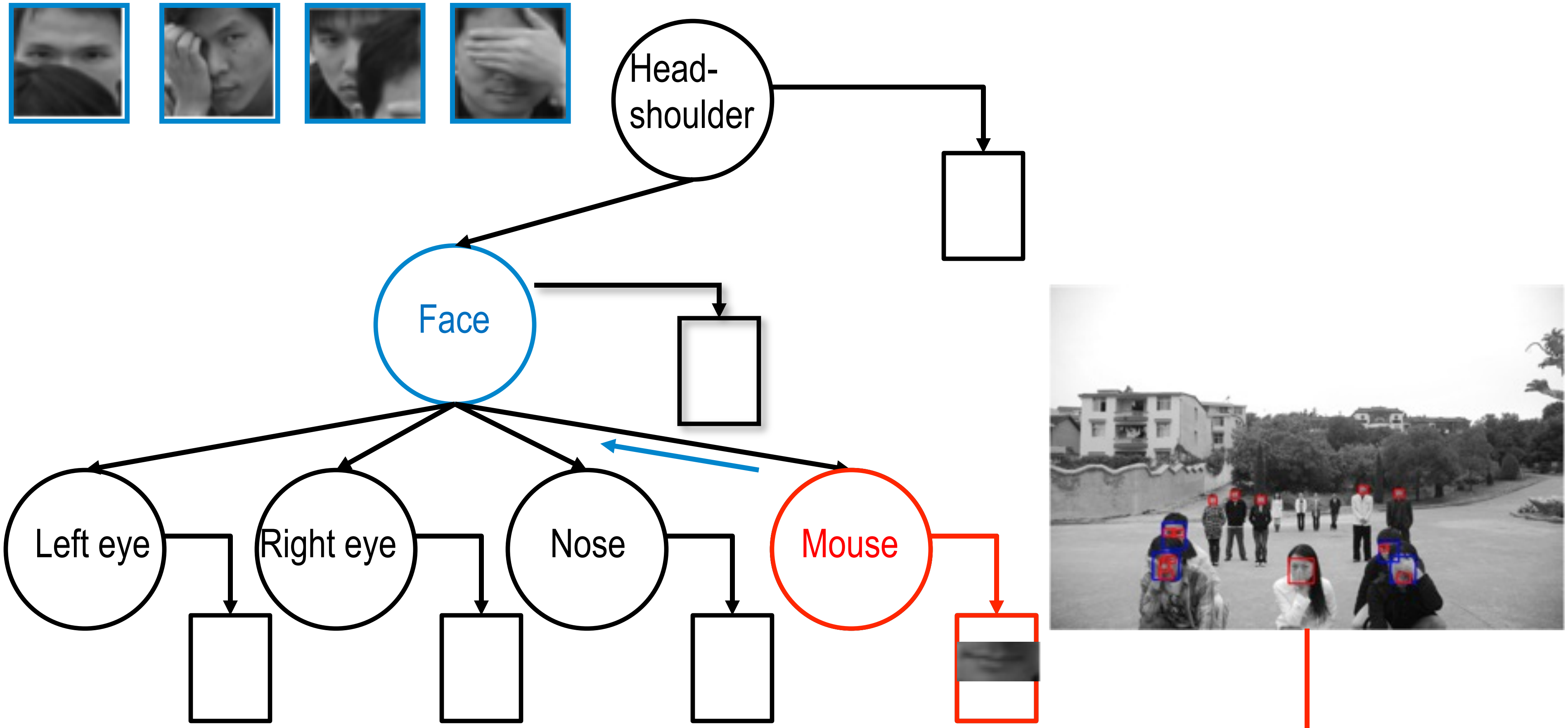
β-channels: p(face | parts), binding



α-channels of some parts are on

β processes for the face node(when its α is off)

β-channels: p(face | parts), binding

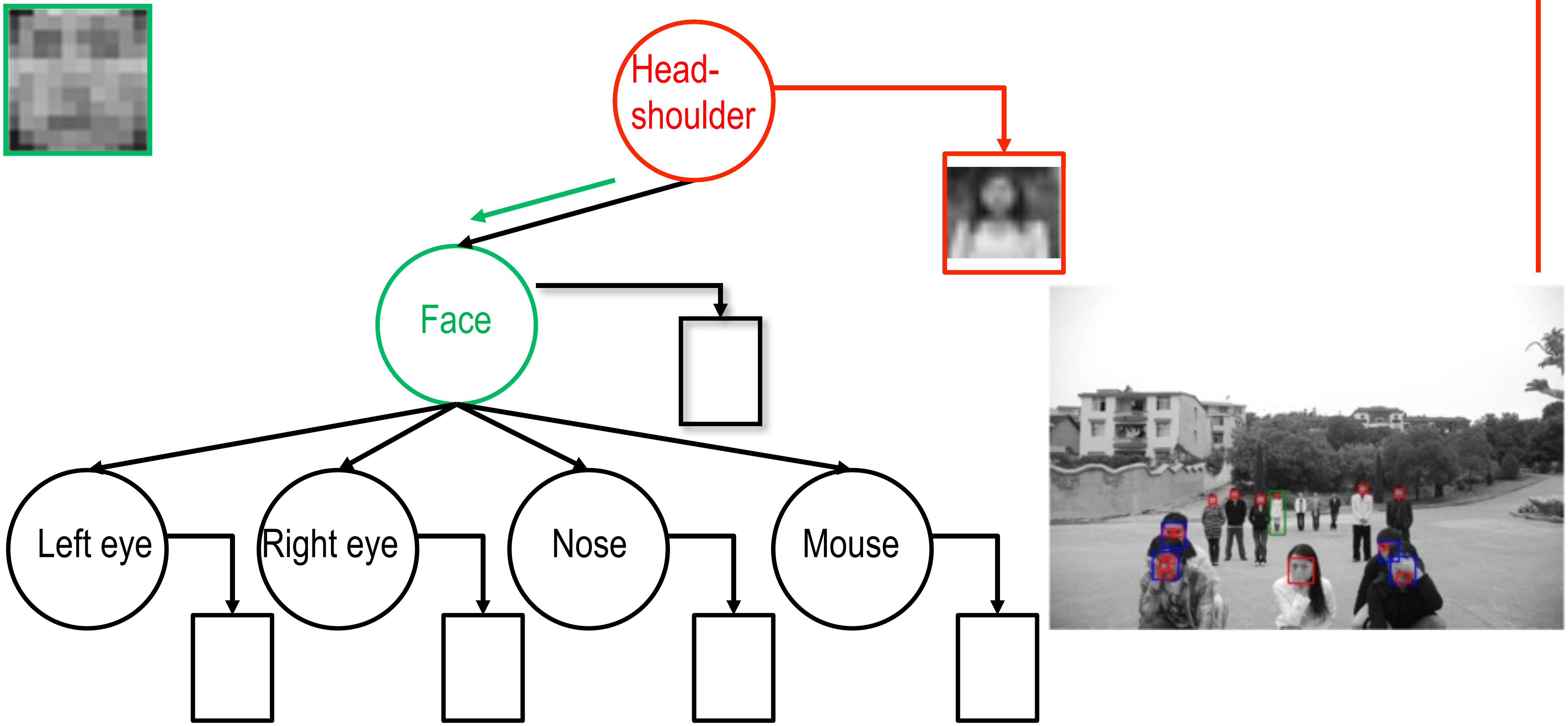


α-channels of some parts are on

γ processes for the face node(when it’s α and β is off)

γ-channels: p(face | parents), predicting

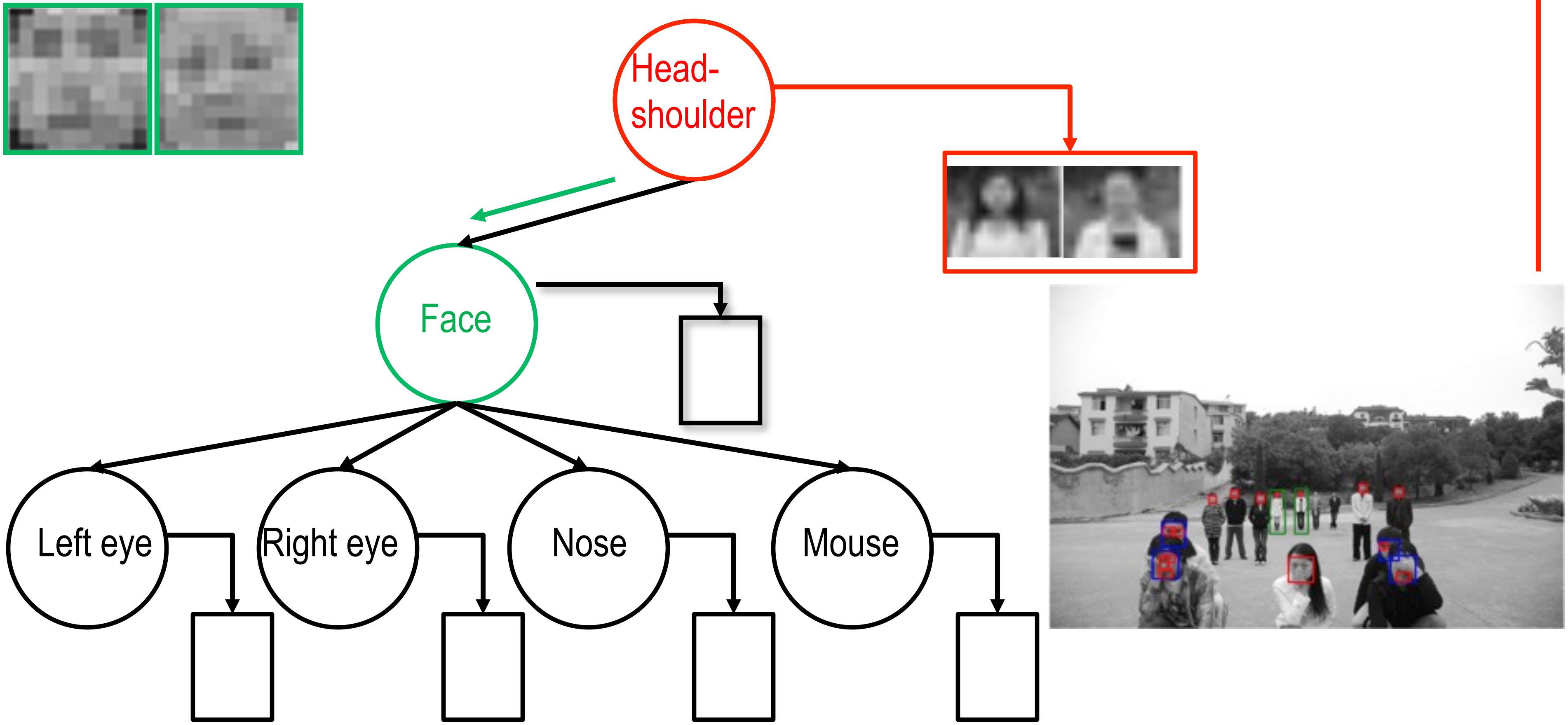
α-channels of some parents are on



γ processes for the face node(when it’ s α and β is off)

γ-channels: p(face | parents), predicting

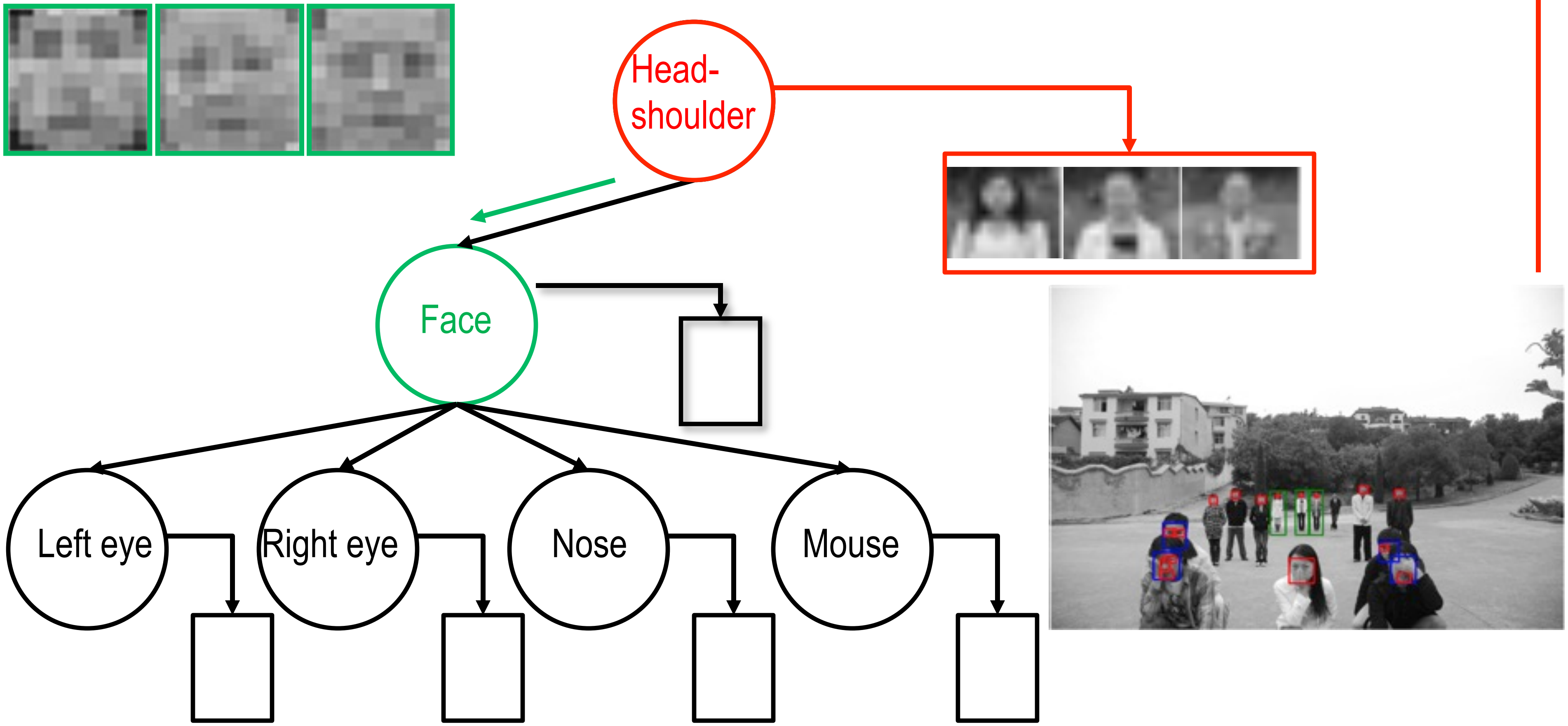
α-channels of some parents are on



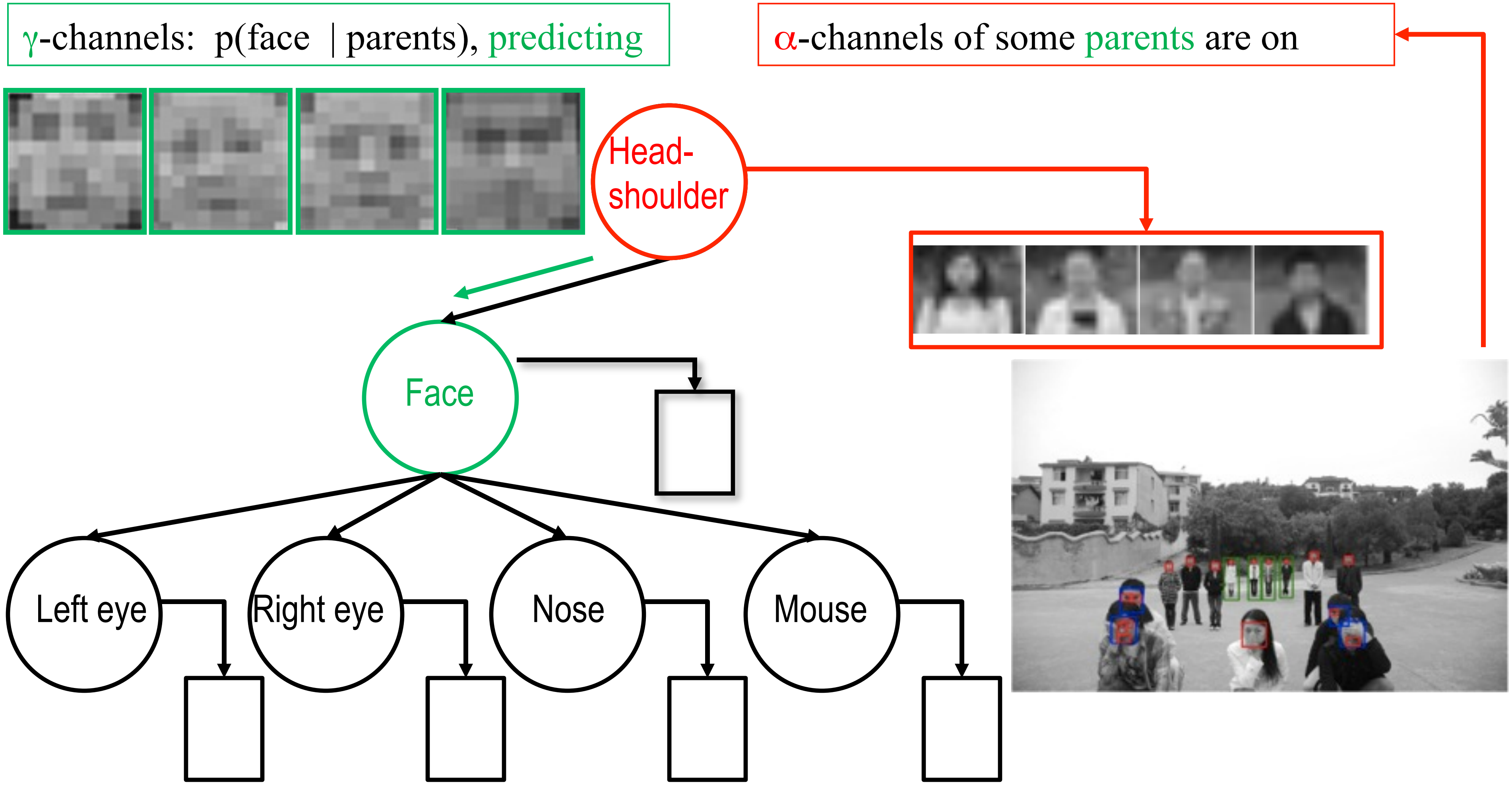
γ processes for the face node(when it’ s α and β is off)

γ-channels: p(face | parents), predicting

α-channels of some parents are on



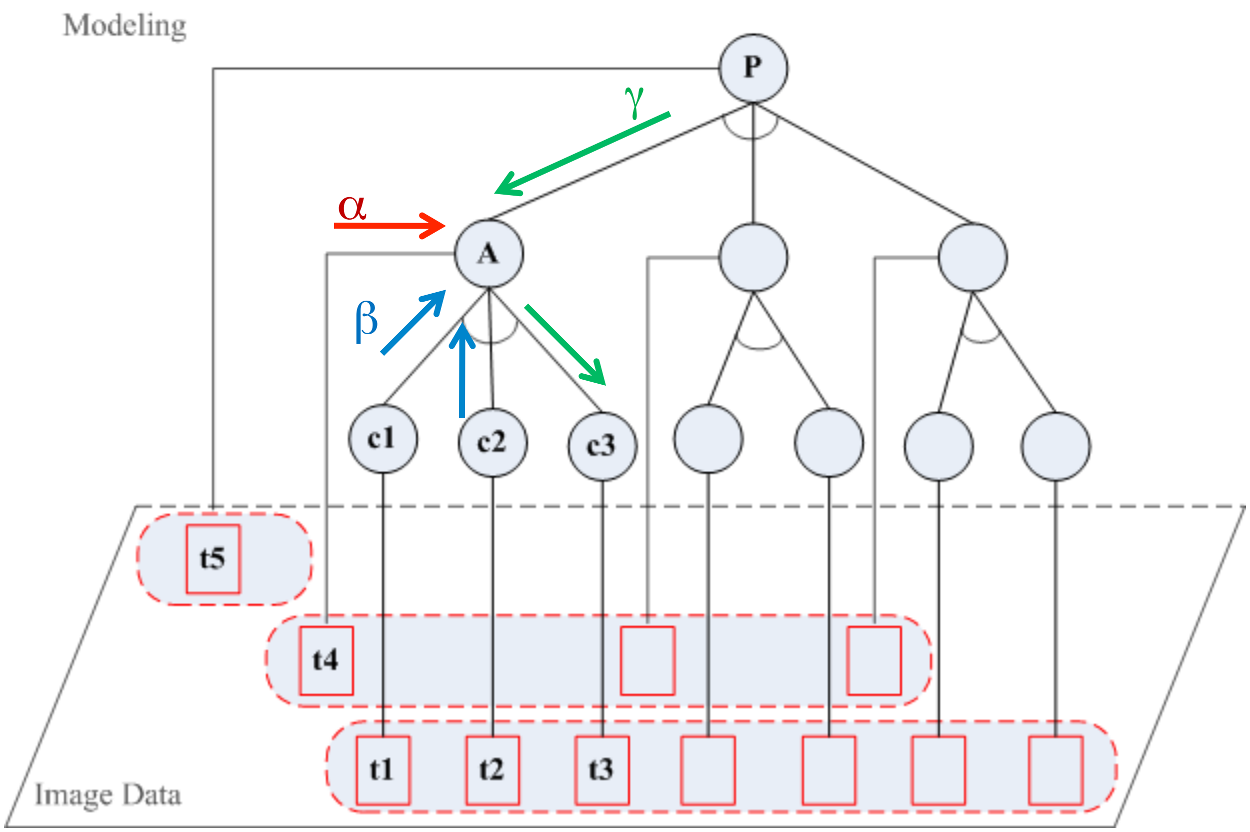
γ processes for the face node(when it’ s α and β is off)

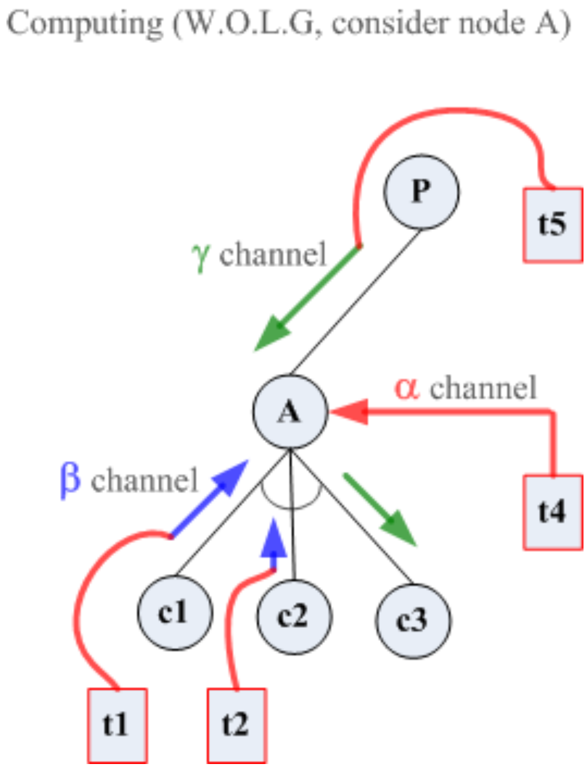


γ-channels: p(face | parents), predicting

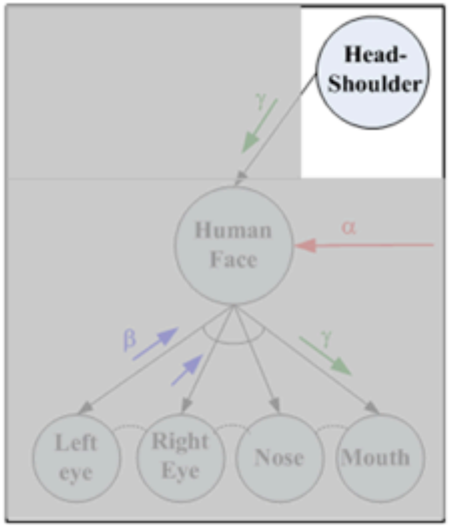
α-channels of some parents are on

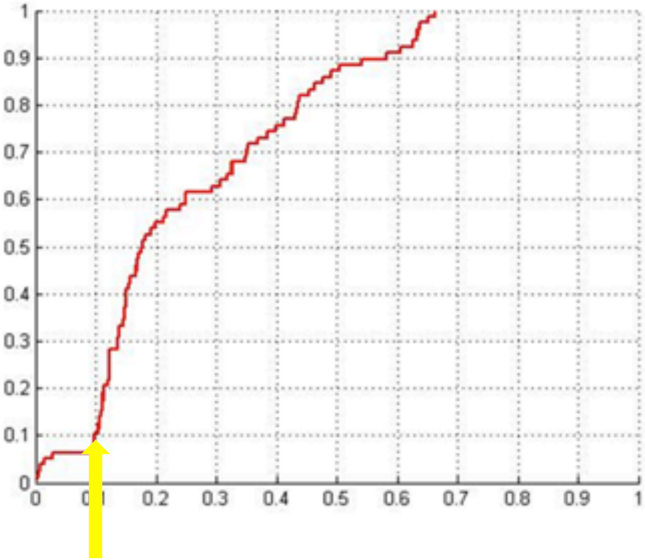
In general: recursive α, β and γ channels



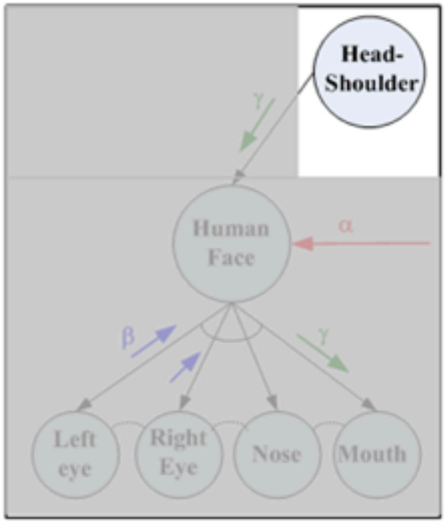


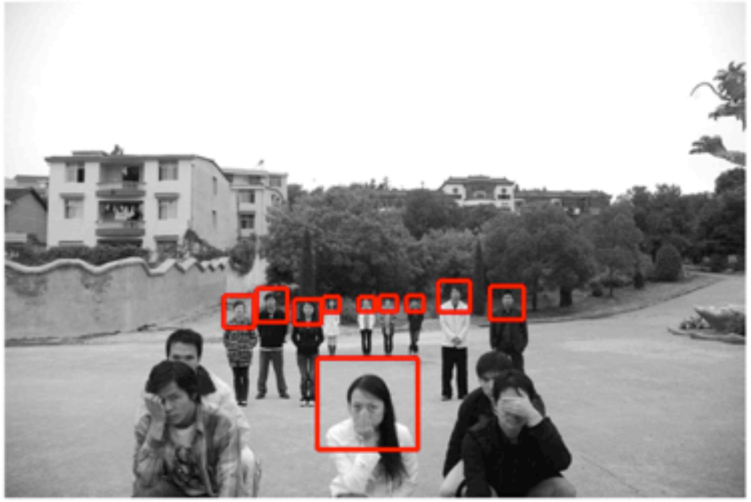
α−channel: head-shoulder

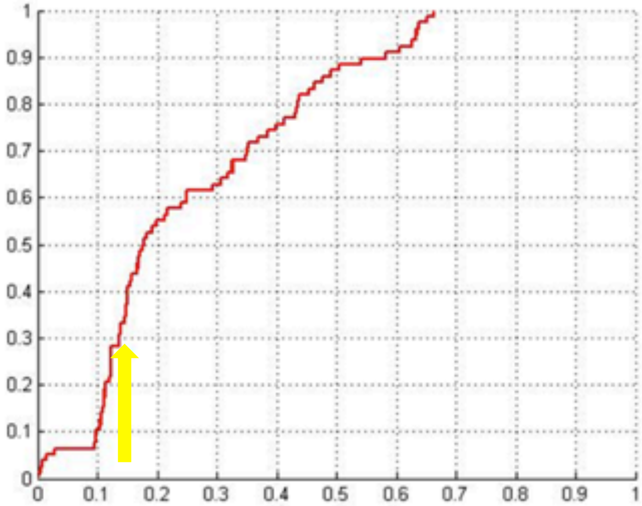


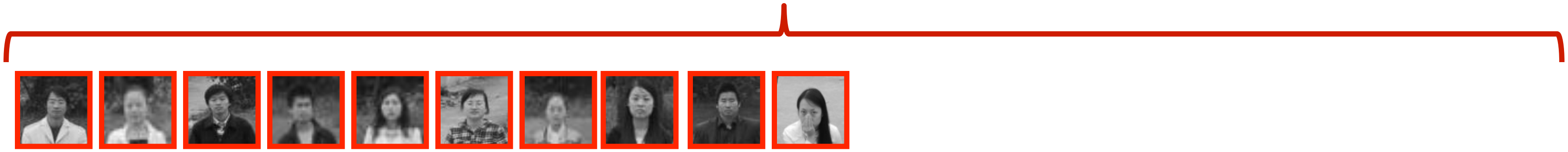


α−channel: head-shoulder

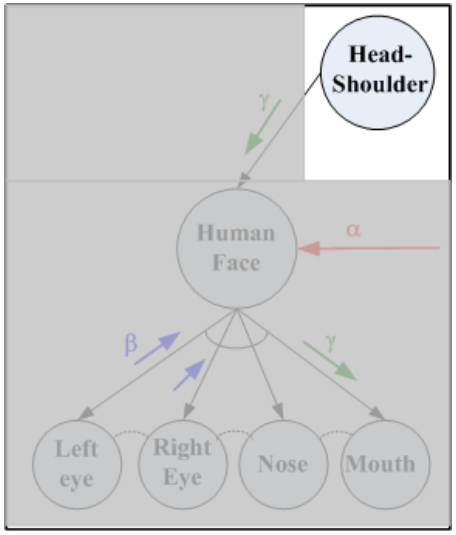


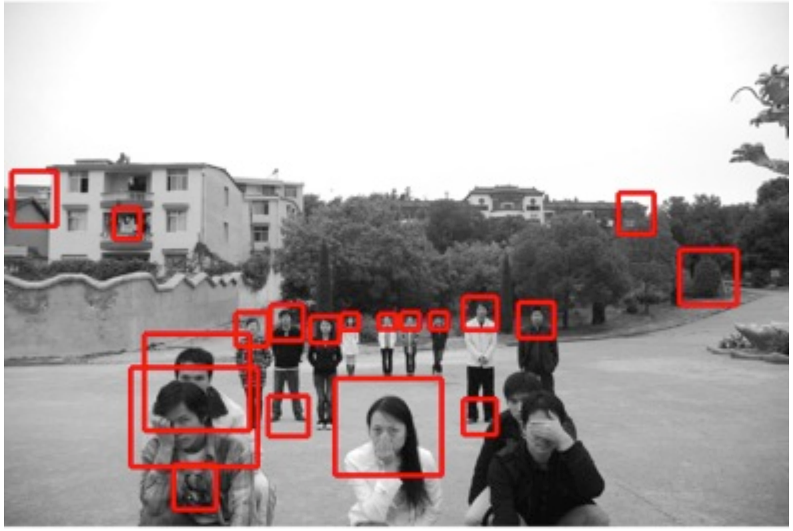


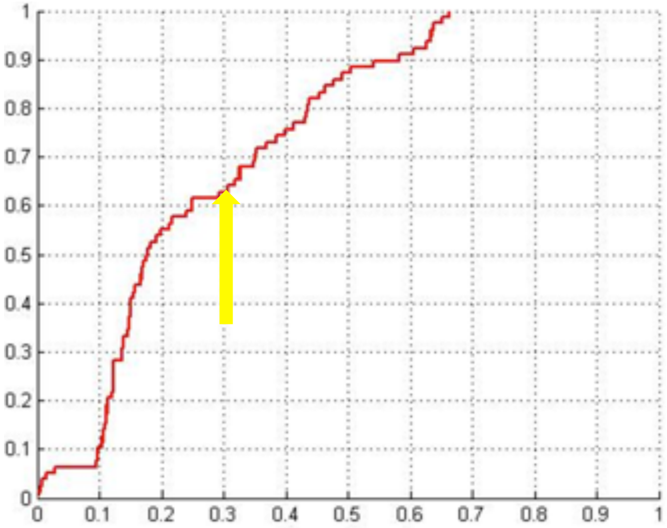


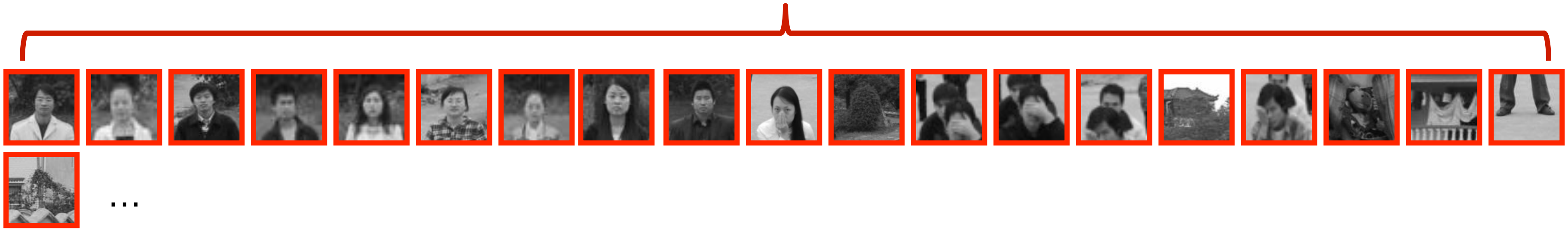


α−channel: head-shoulder



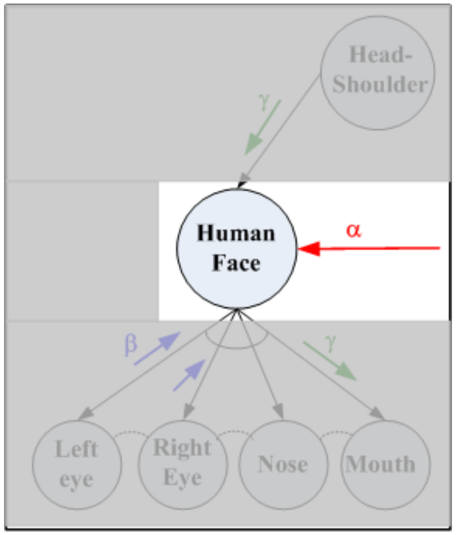


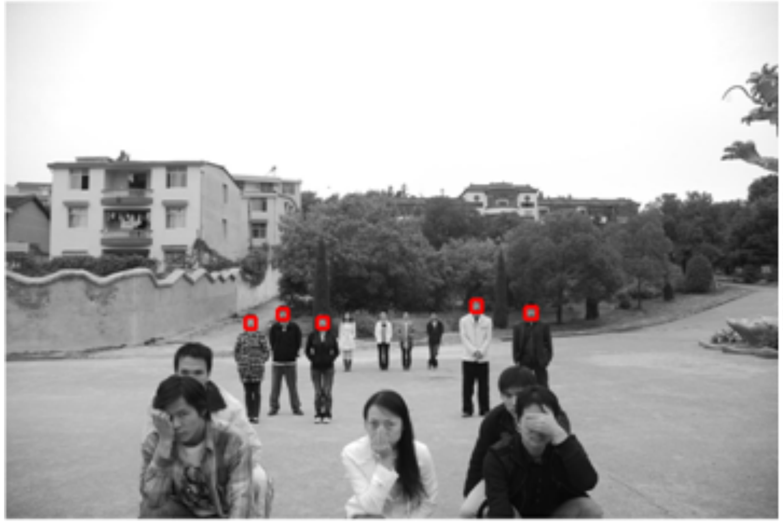


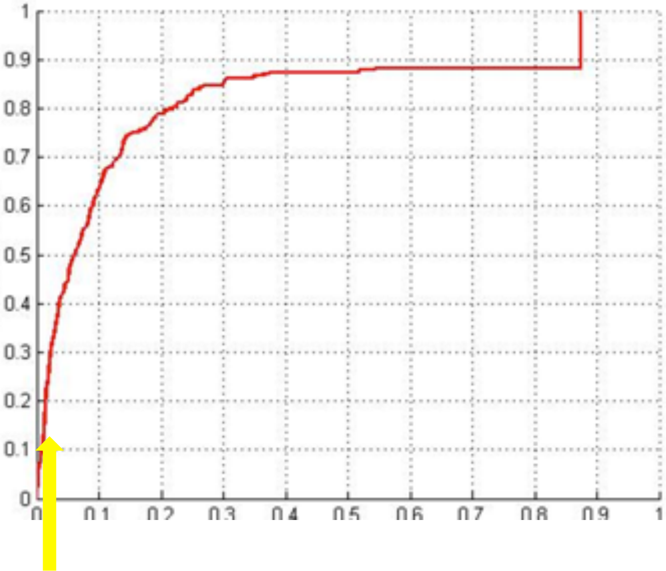


…

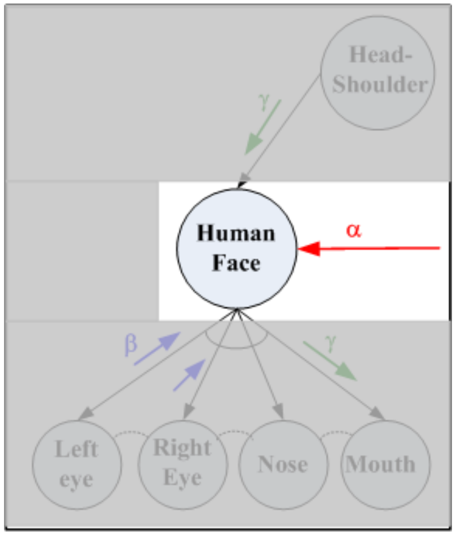
α−channel: face

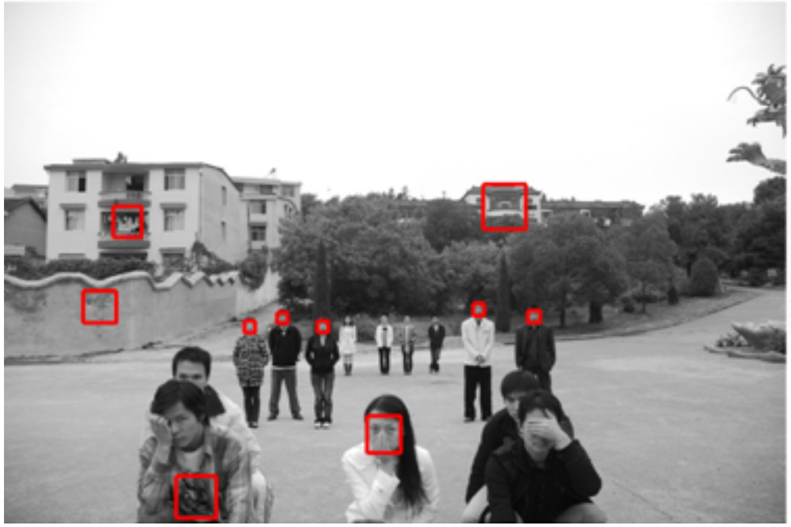


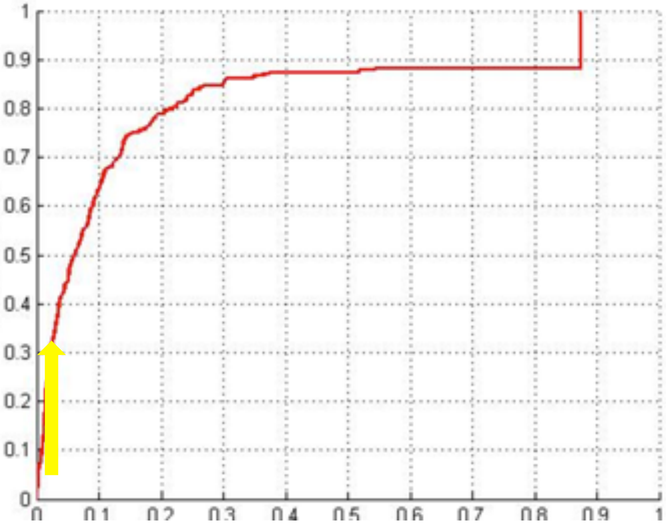


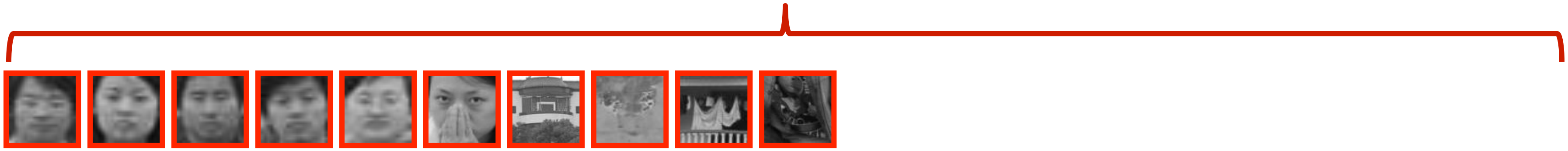


α−channel: face

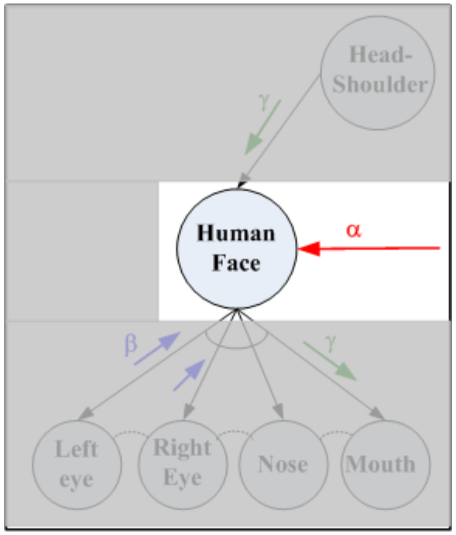


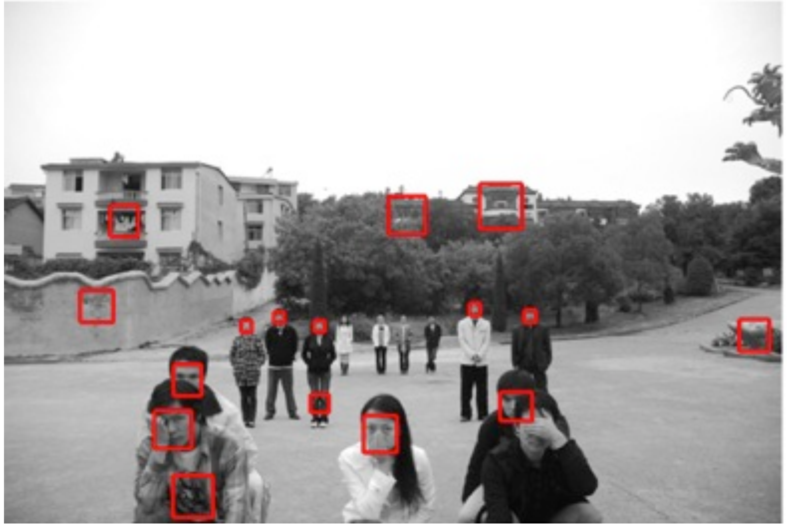


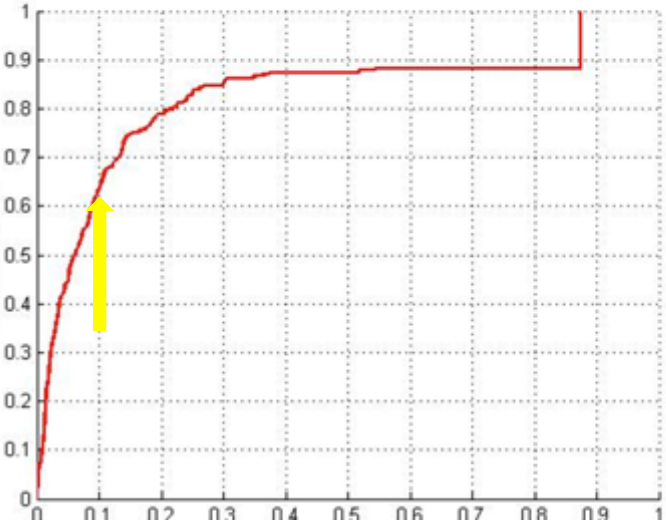


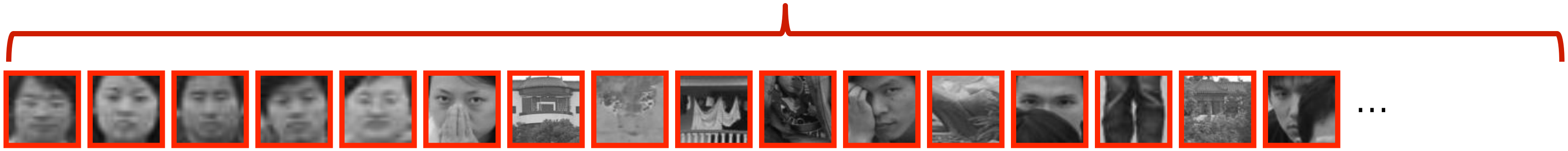


α−channel: face

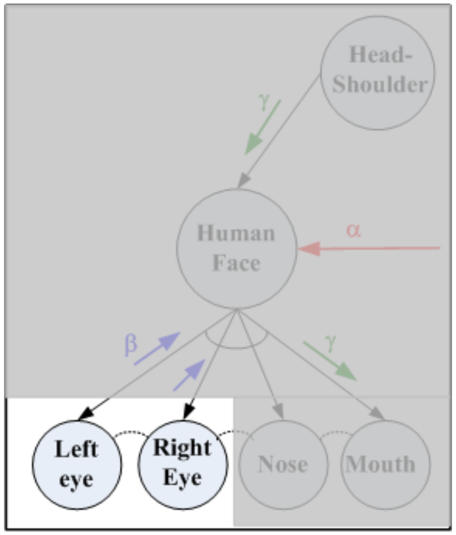


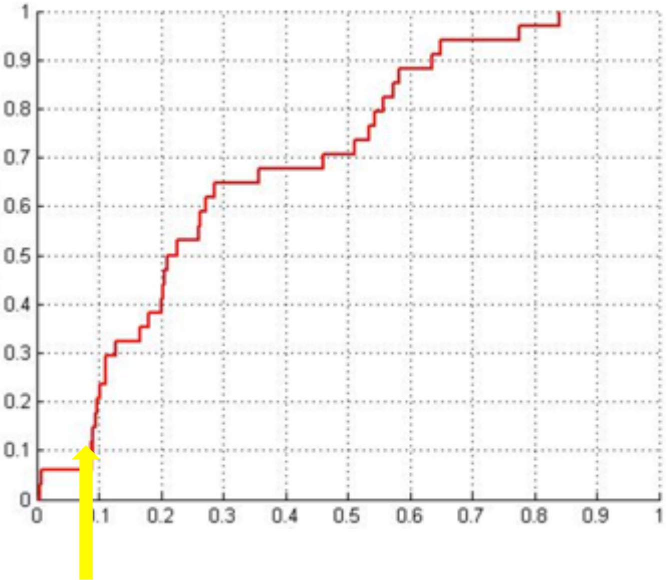


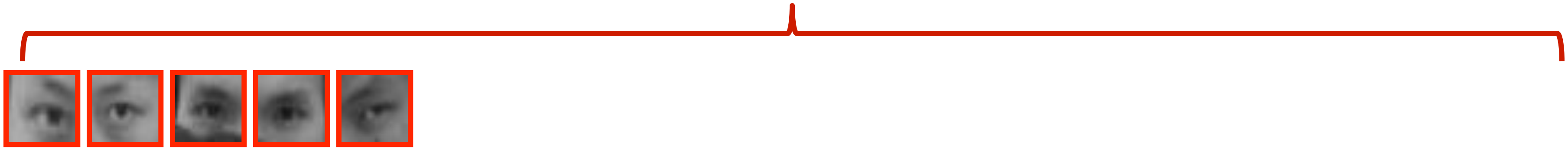




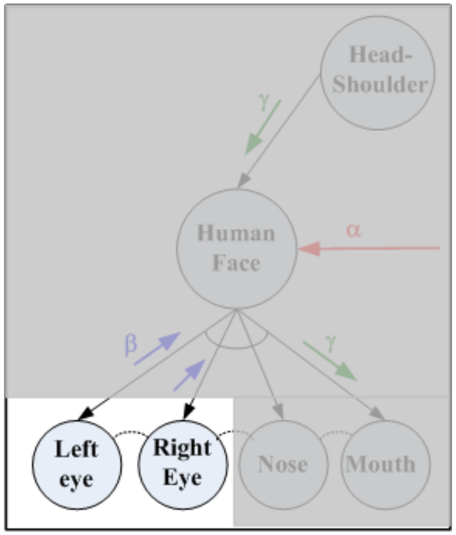
α−channel: eye

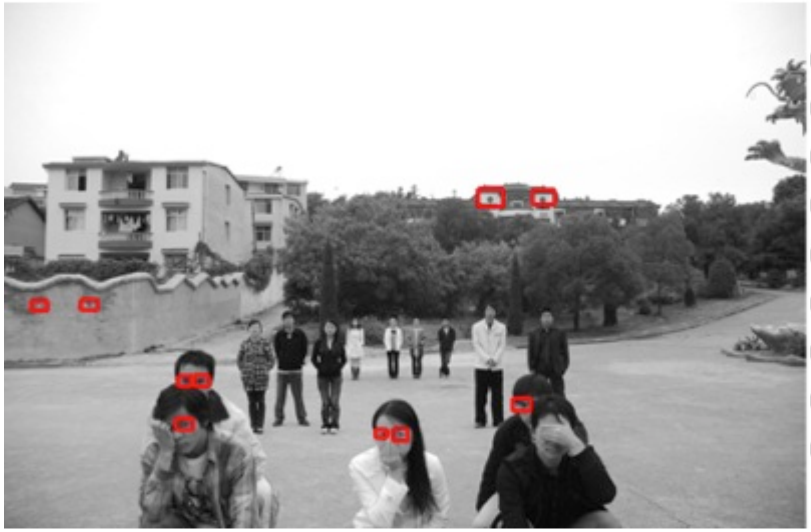


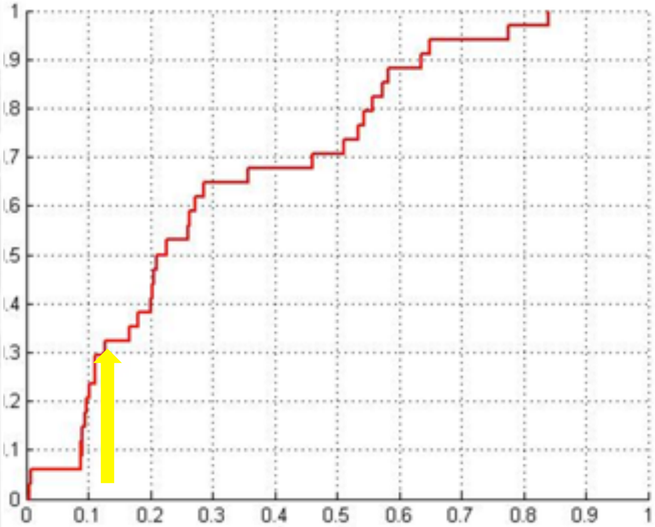


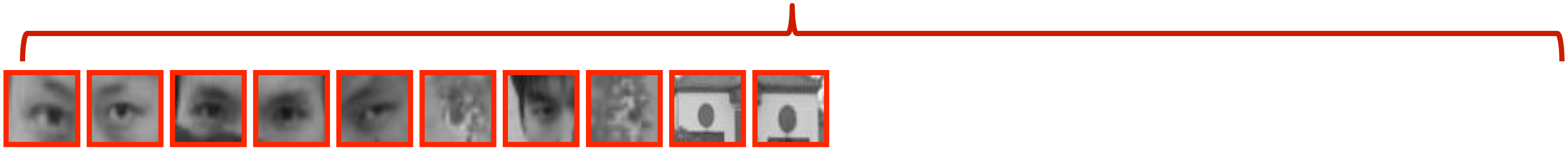


α−channel: eye

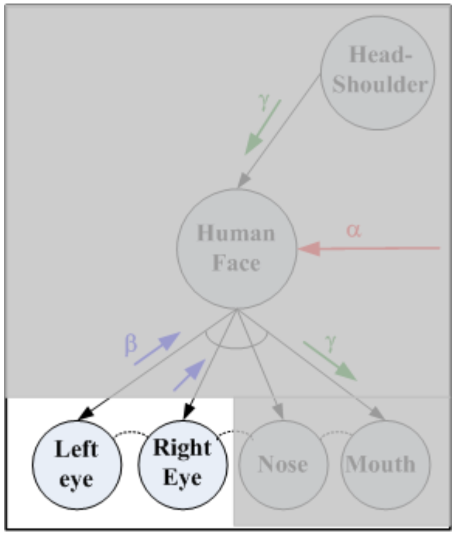


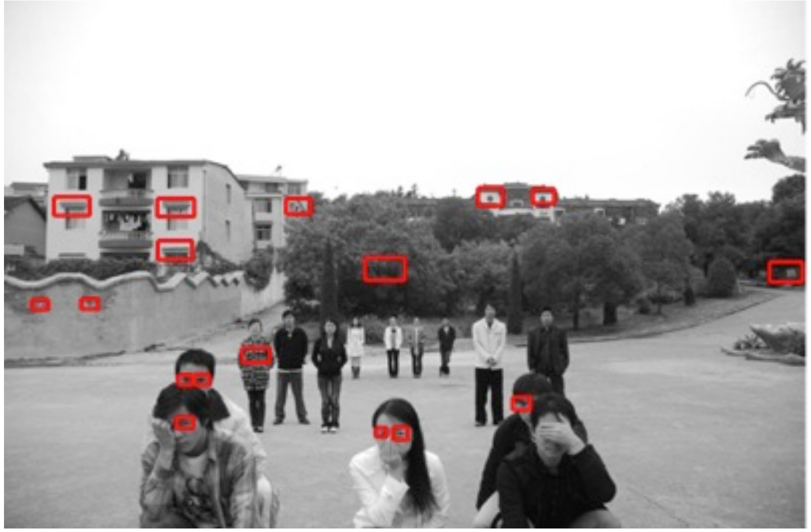


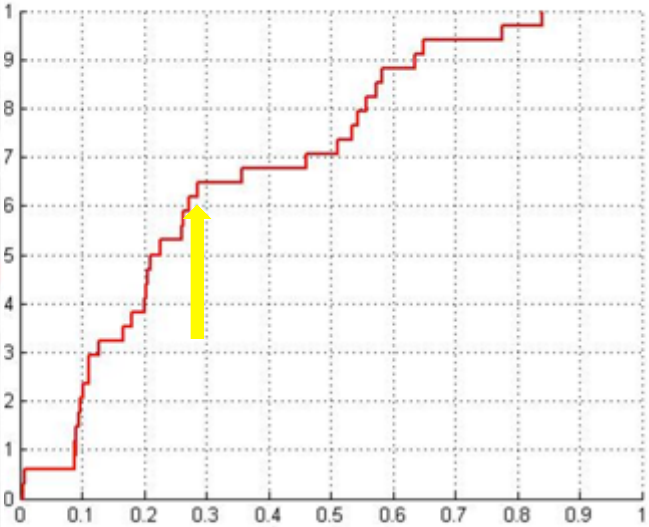


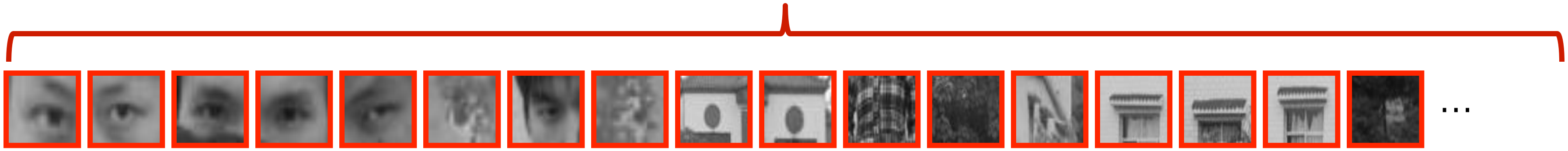


α−channel: eye



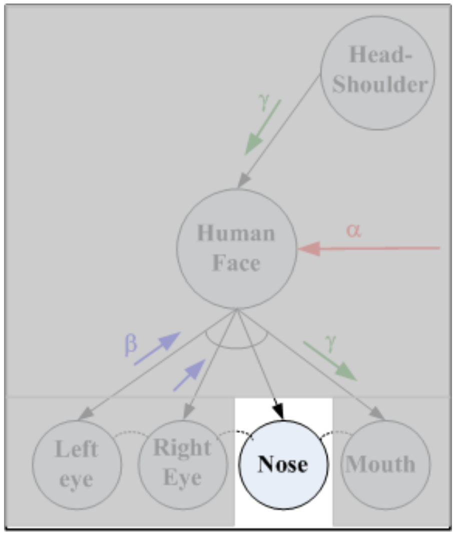


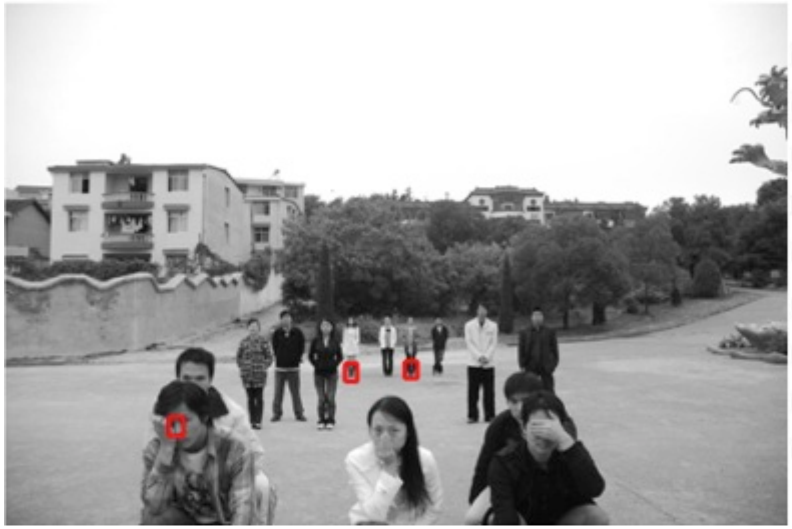


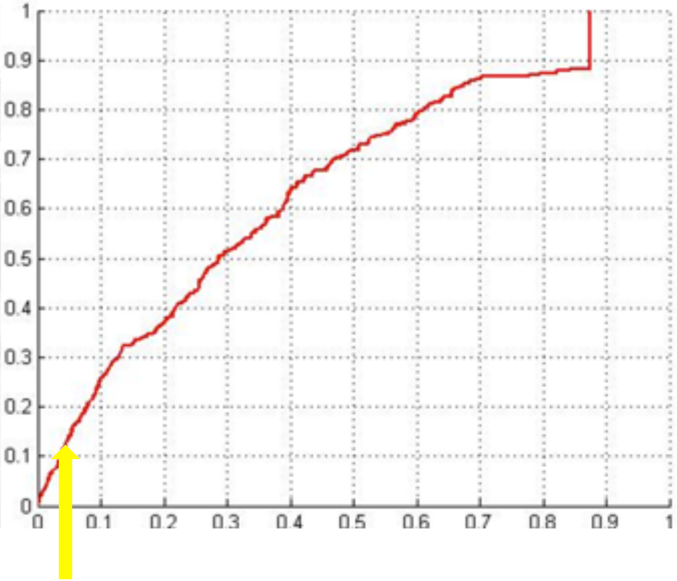


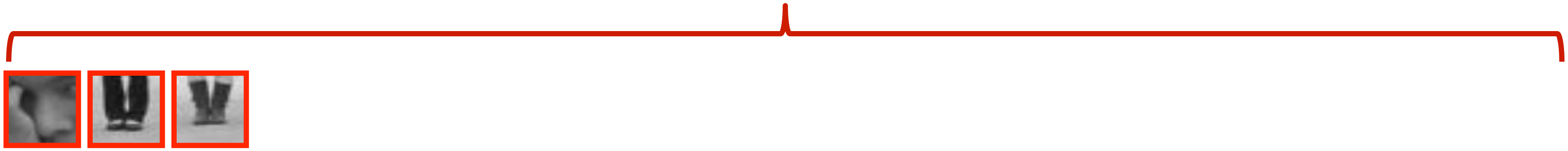
…

α−channel: nose

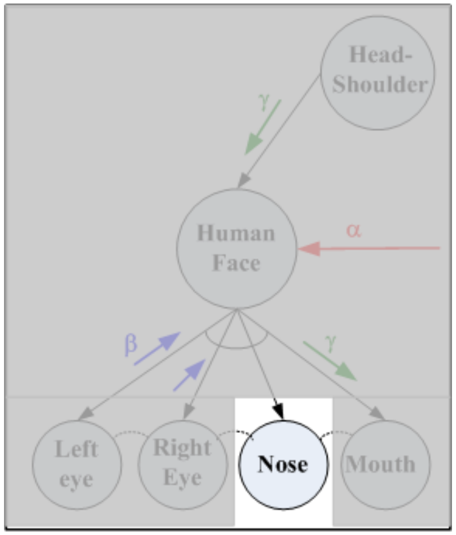


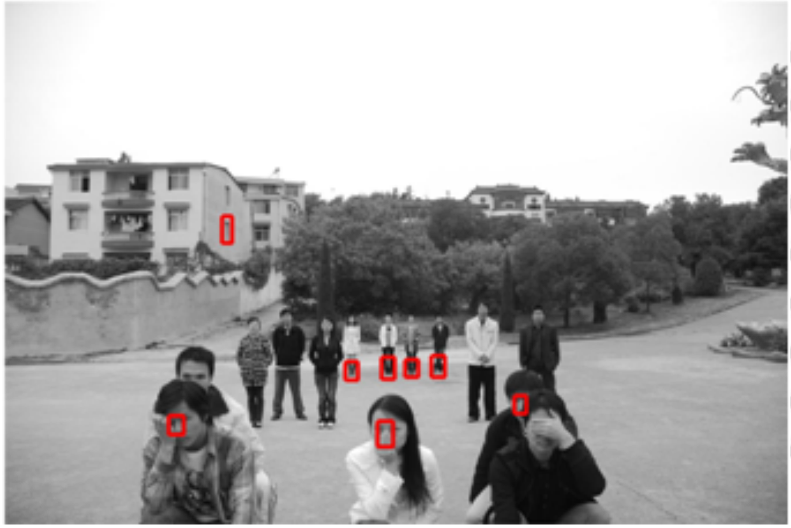


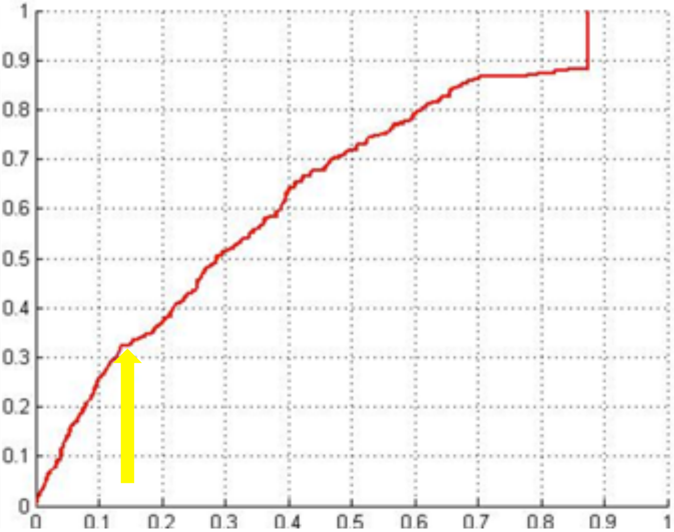


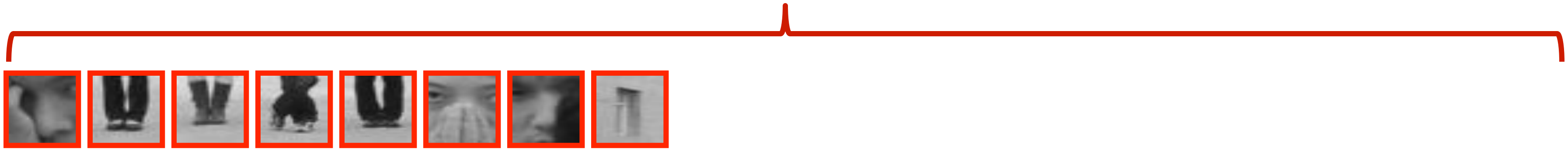


α−channel: nose

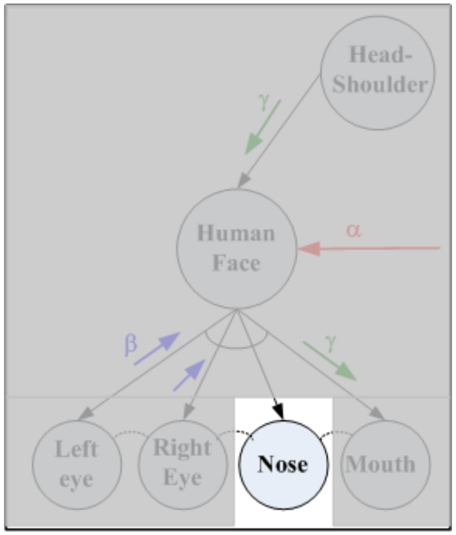


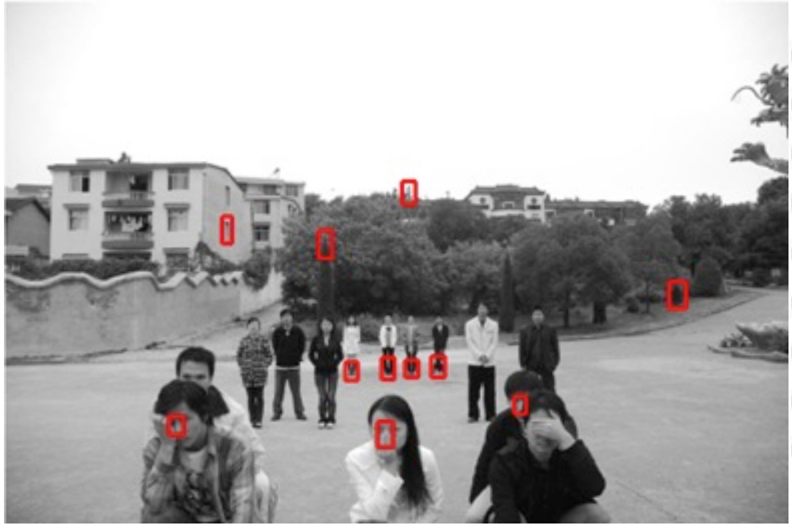


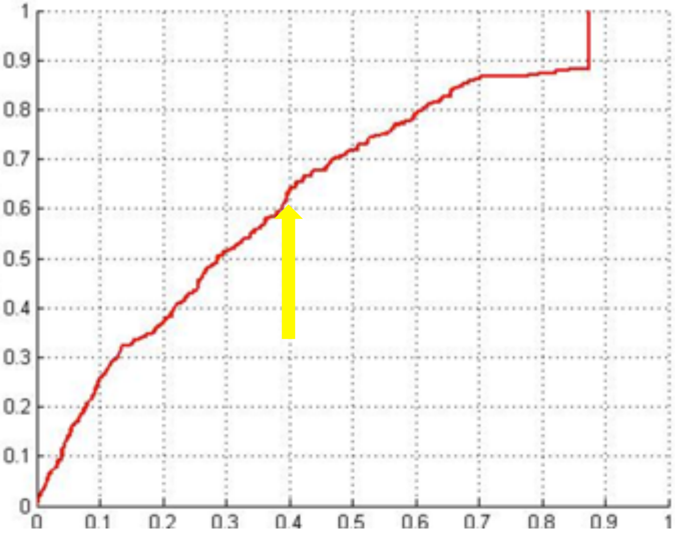


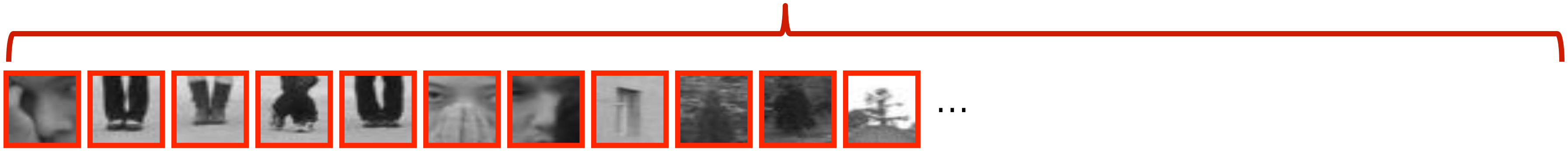


α−channel: nose

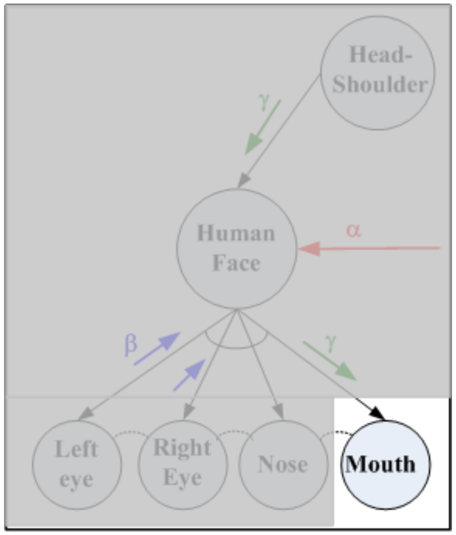


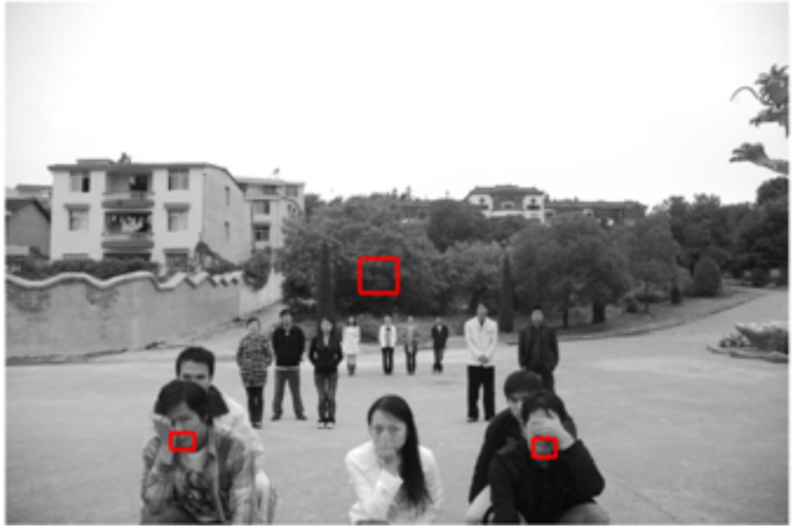


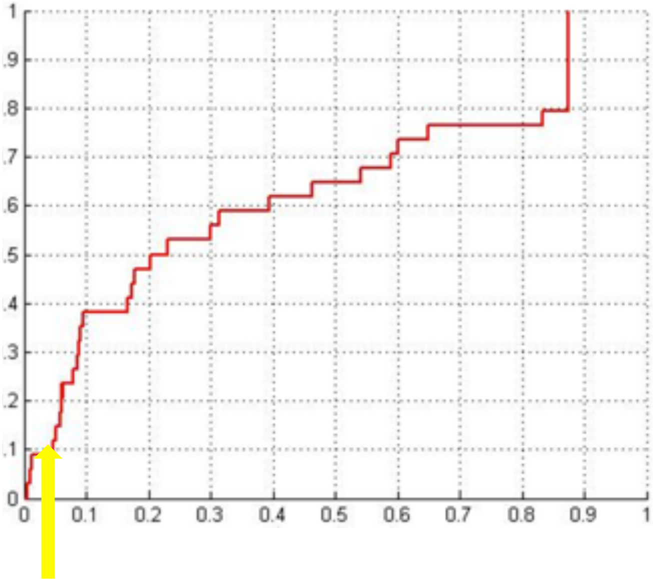


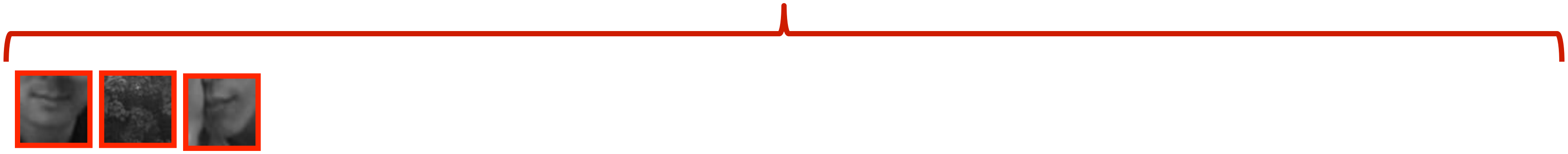


α−channel: mouth

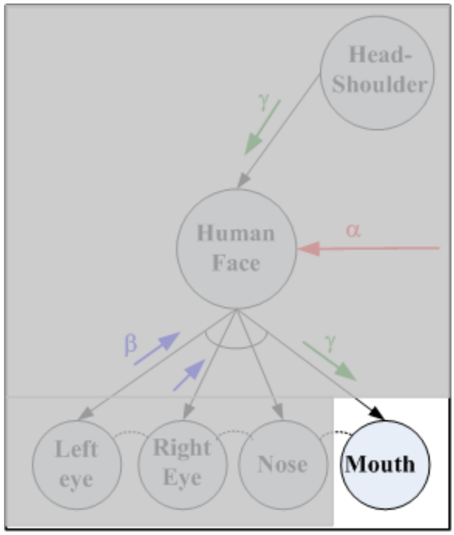


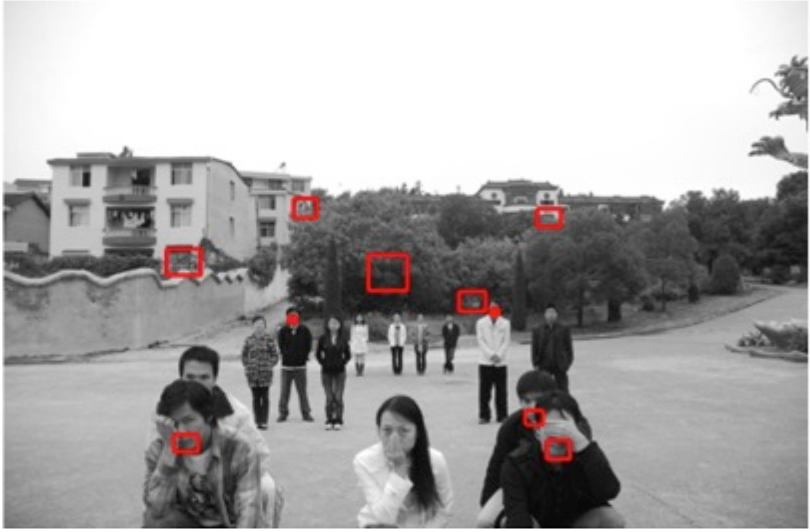


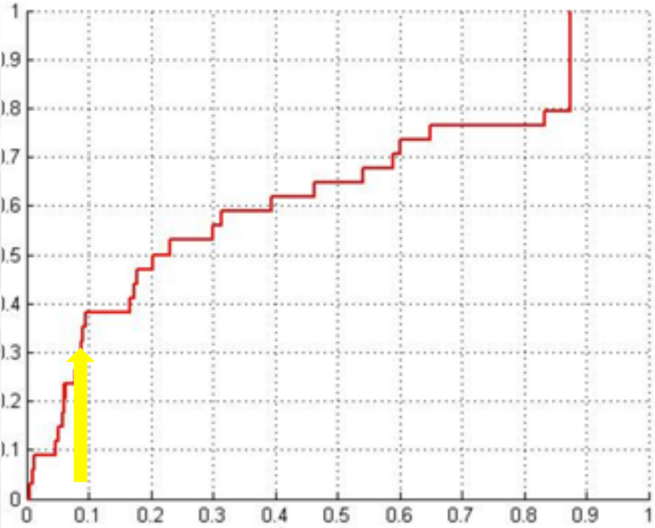


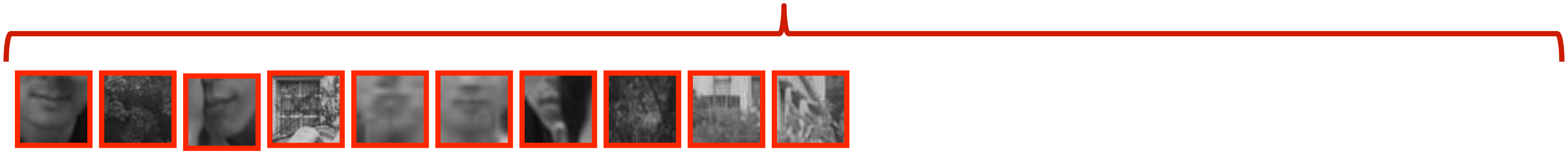


α−channel: mouth

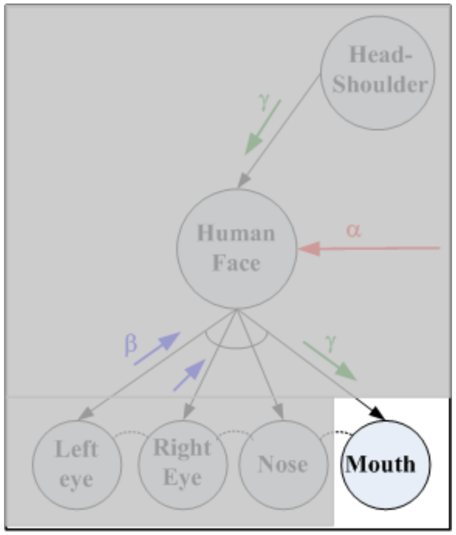


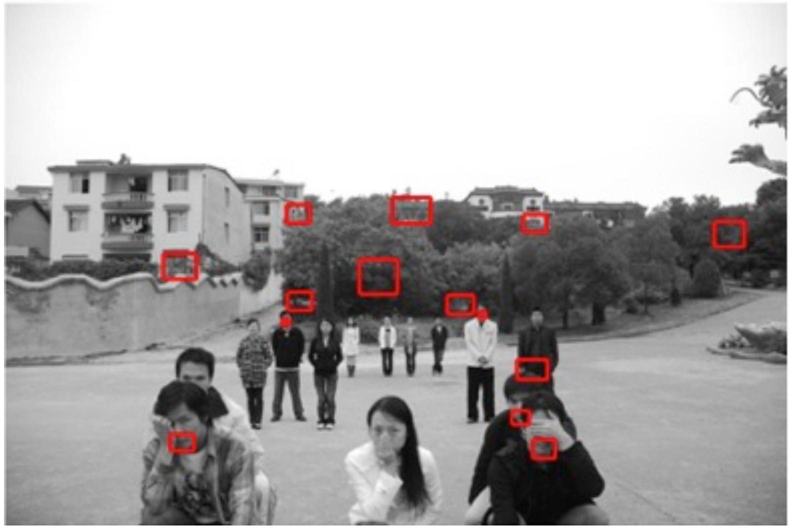


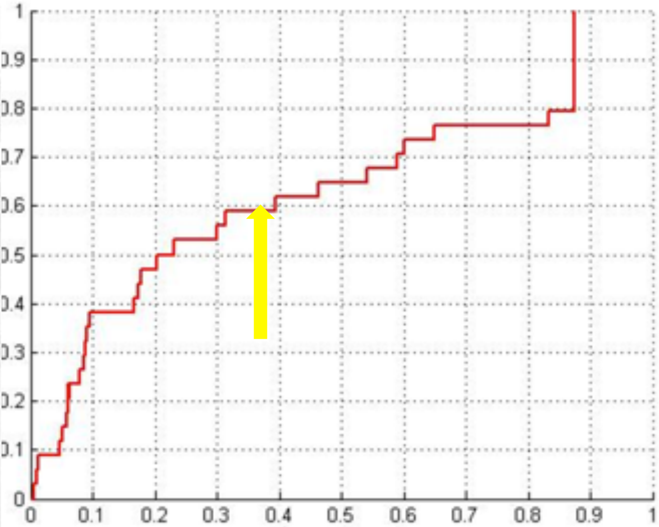


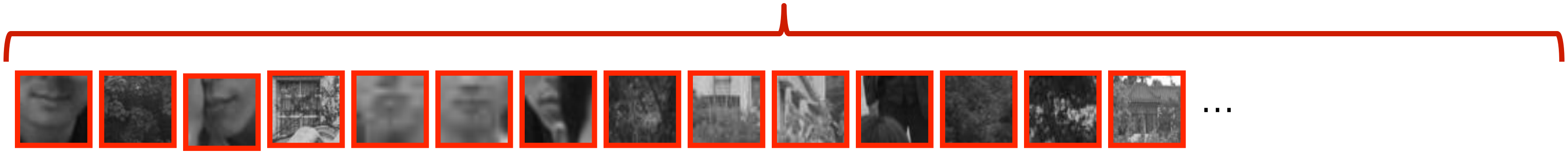


α−channel: mouth

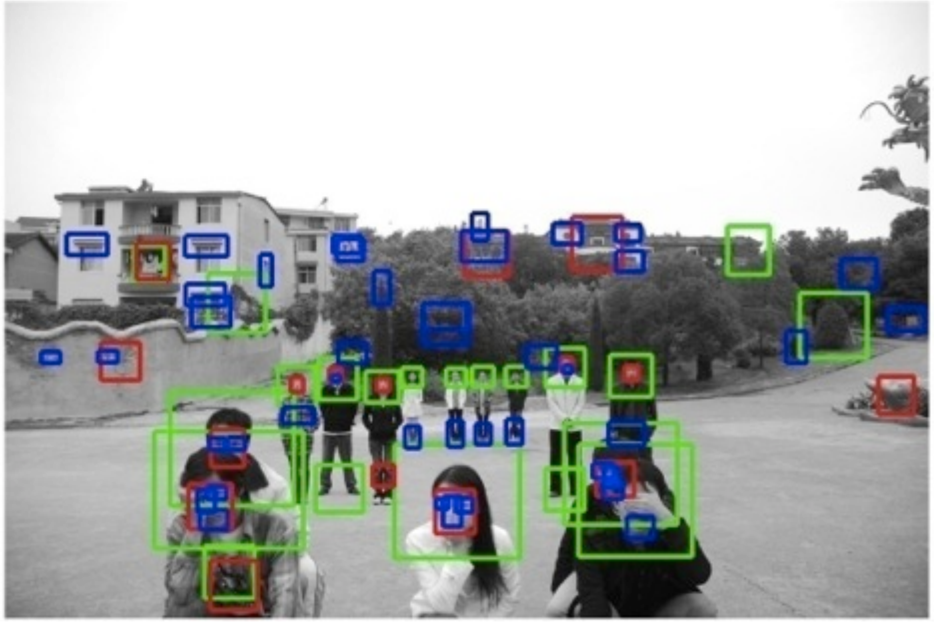






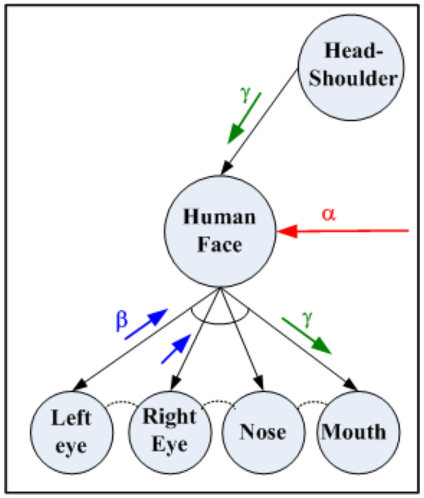


All α channels

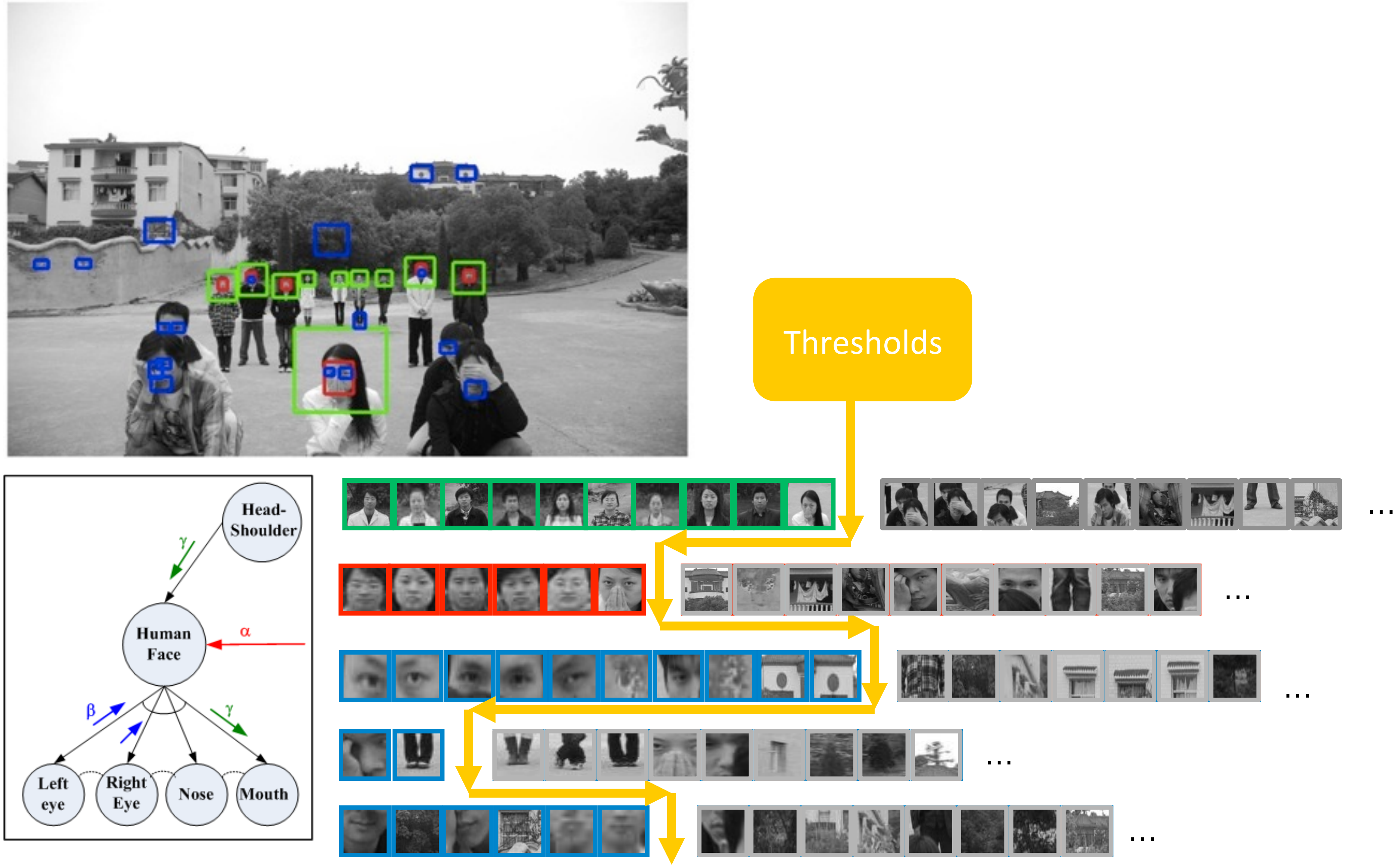


keep things and throw away stuffs by

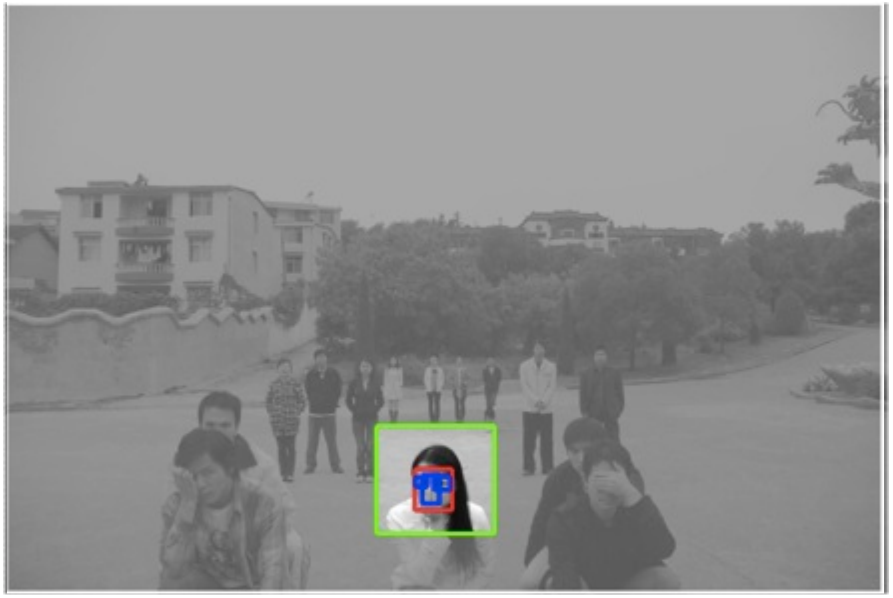
integrating α, β and γ channels

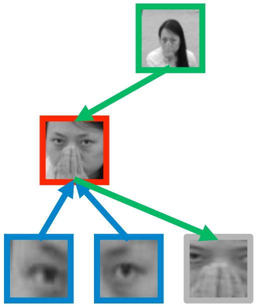


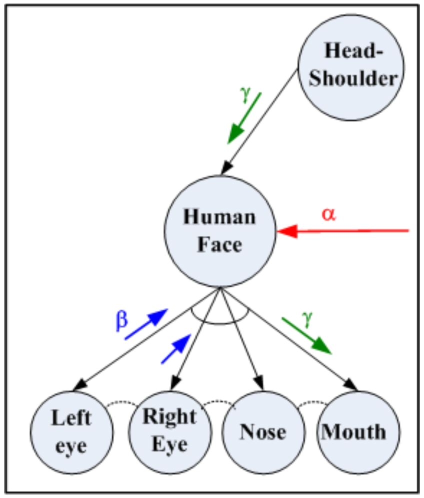
Integrating α, β and γ channels

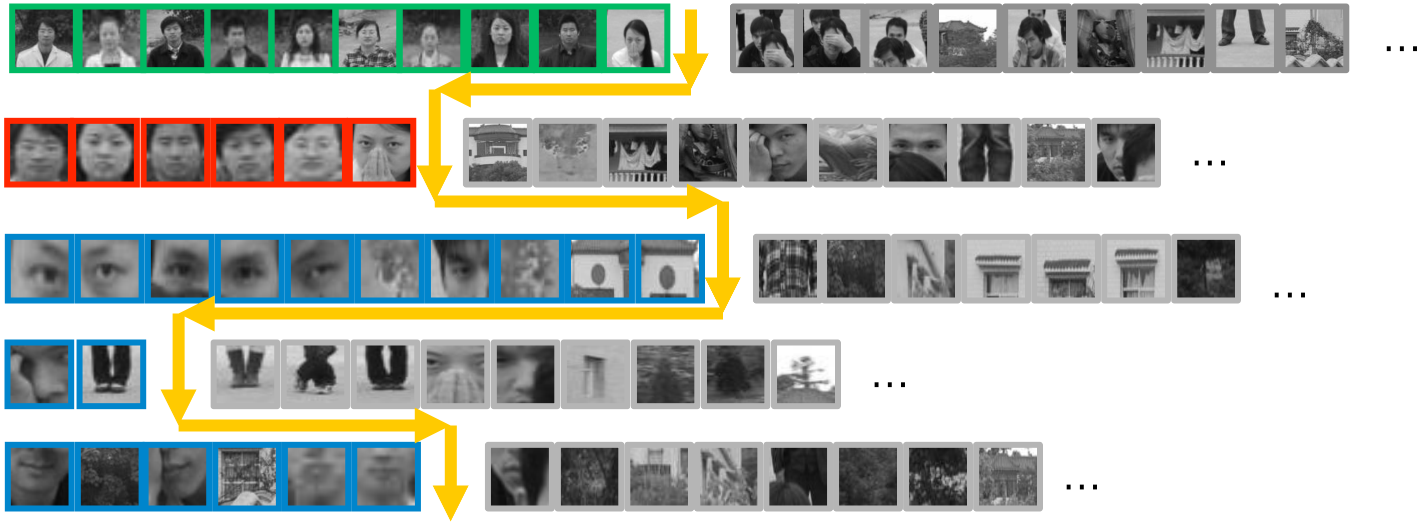


Integrating α, β and γ channels

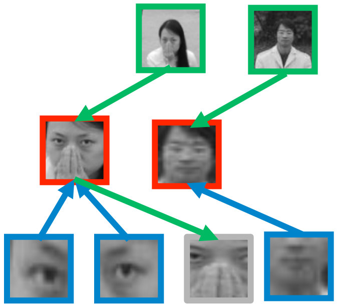


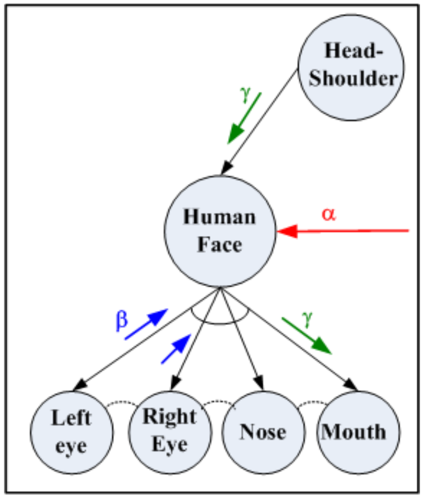


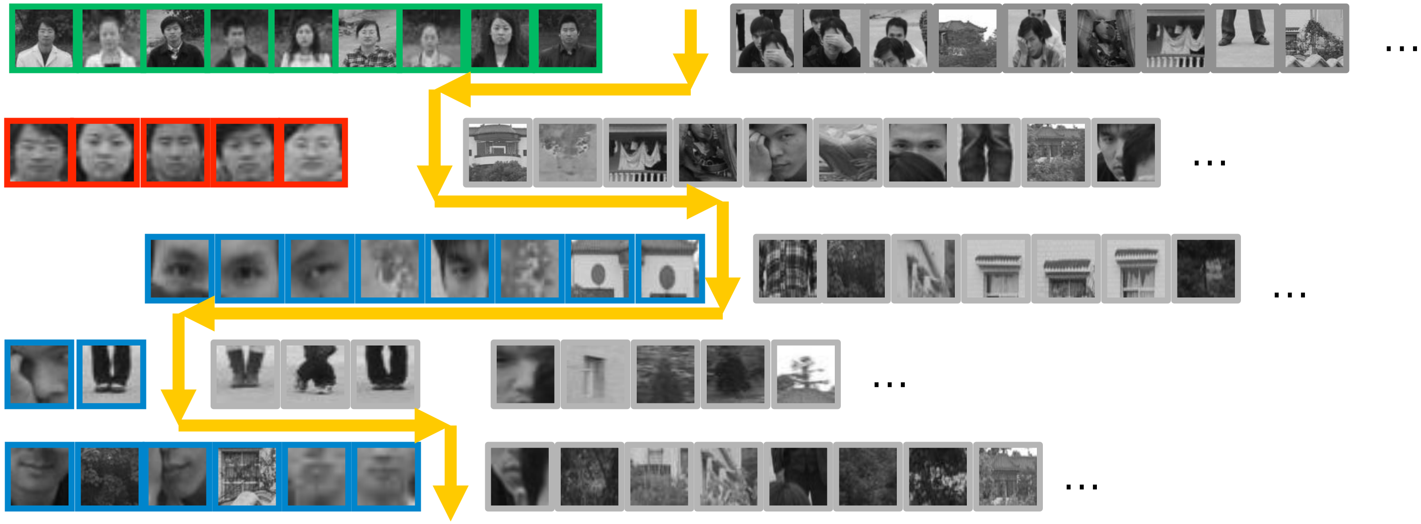




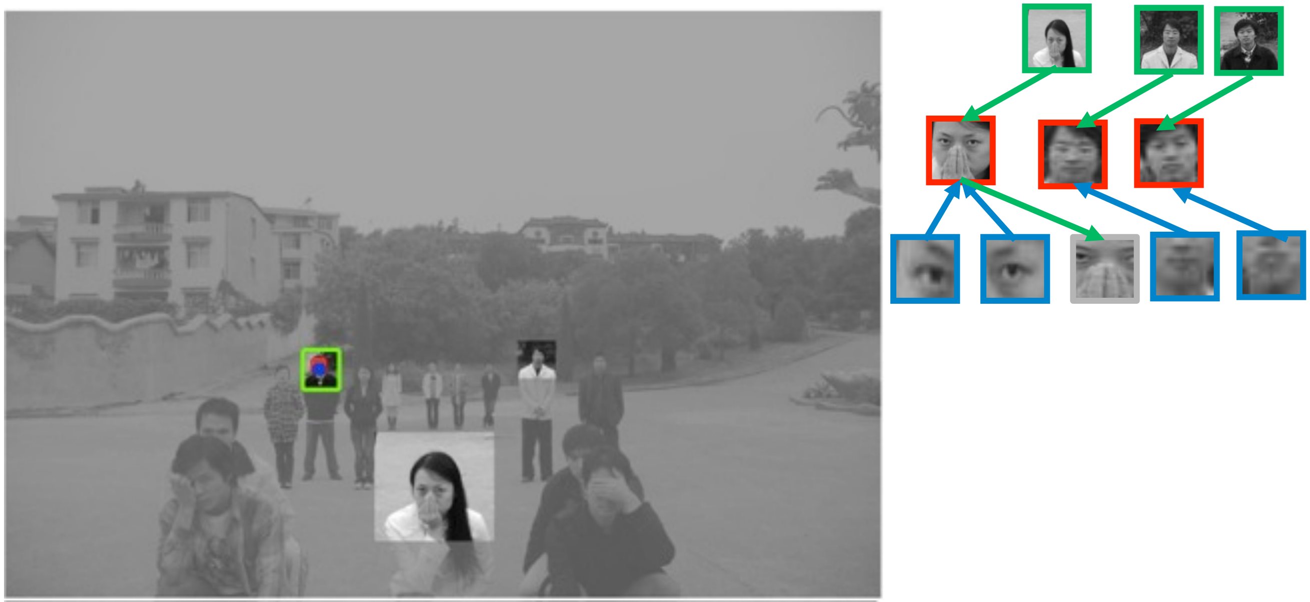
Integrating α, β and γ channels

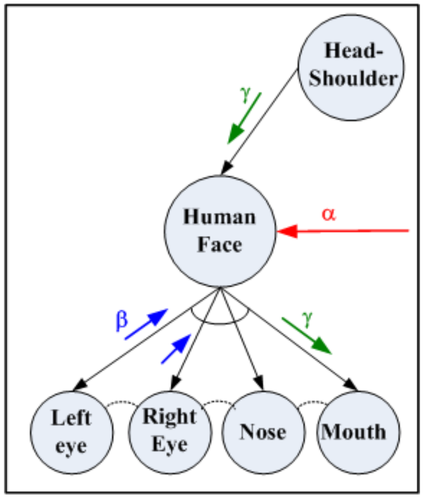


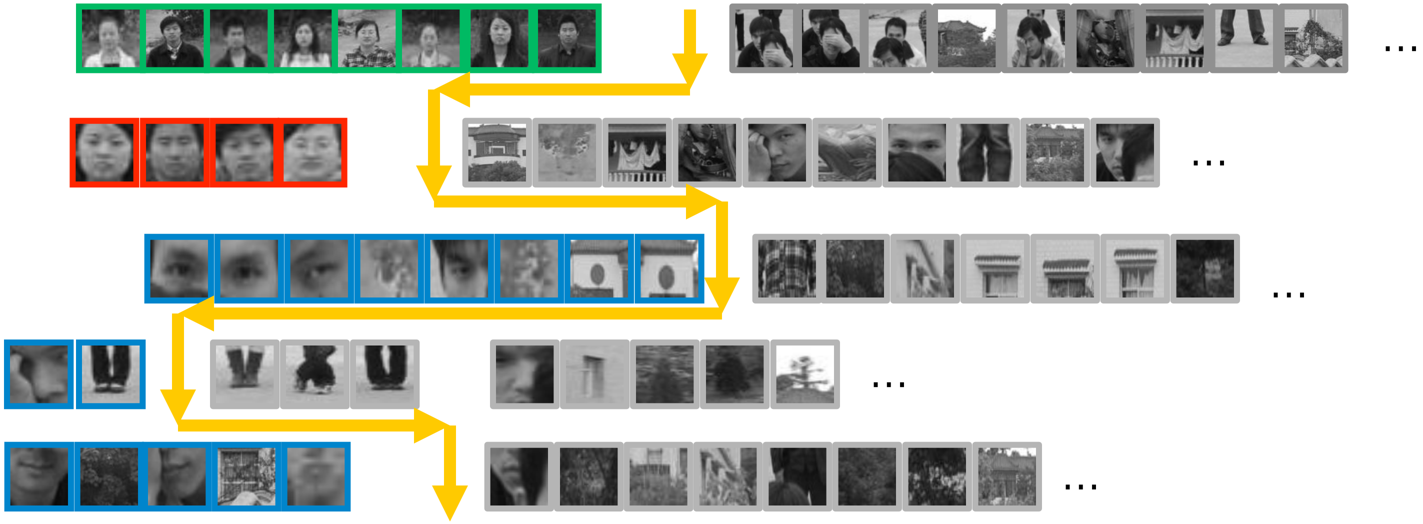




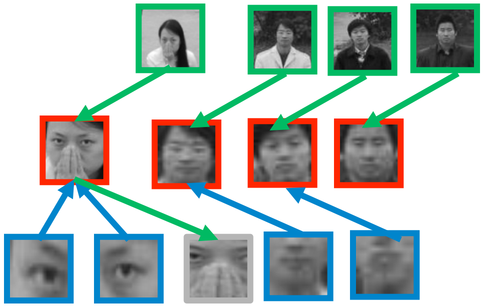
Integrating α, β and γ channels

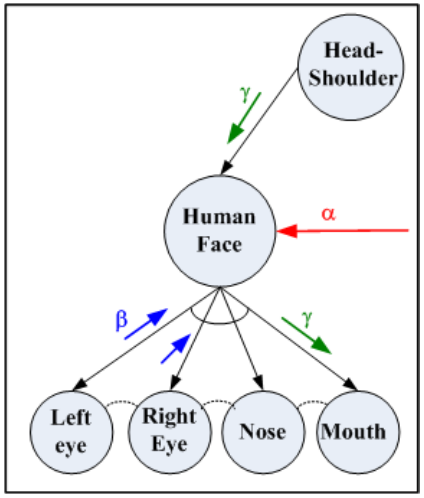


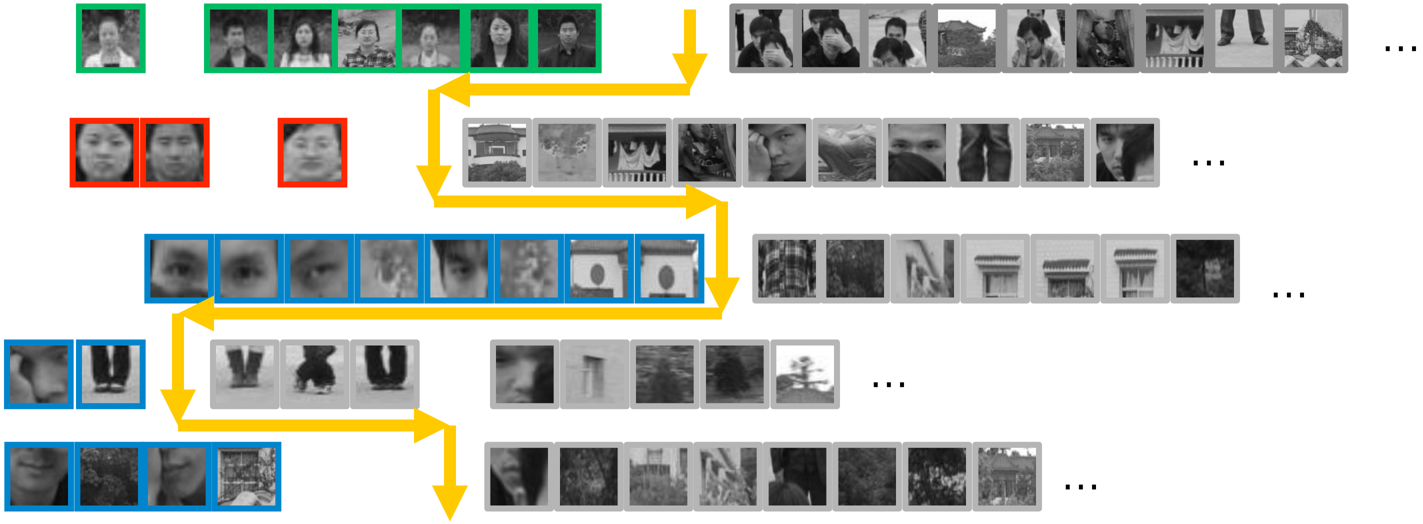




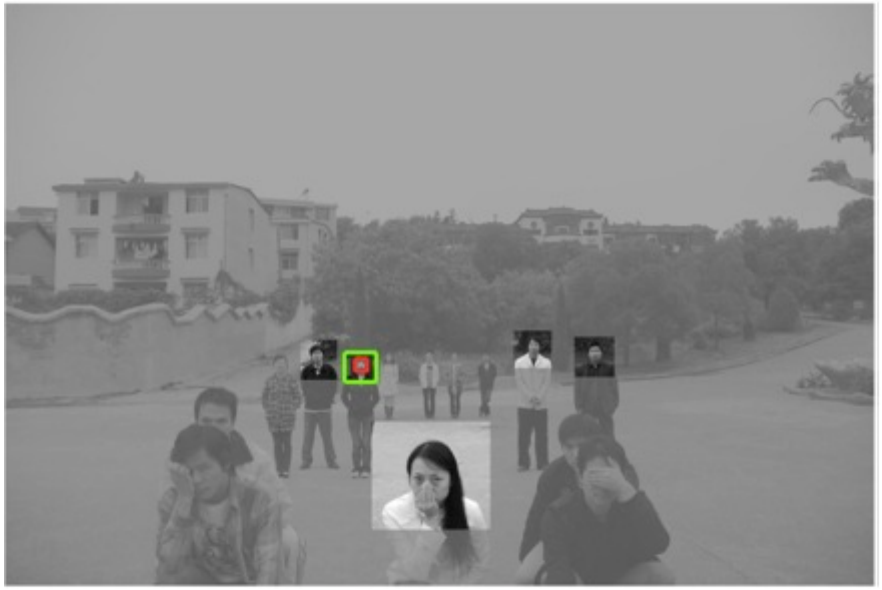
Integrating α, β and γ channels

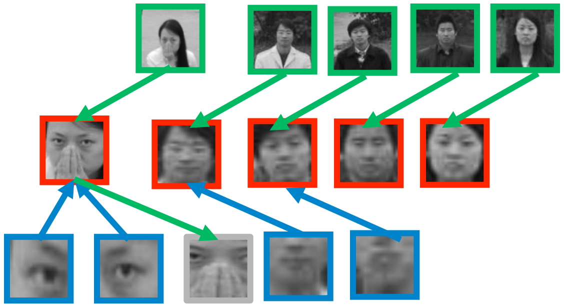


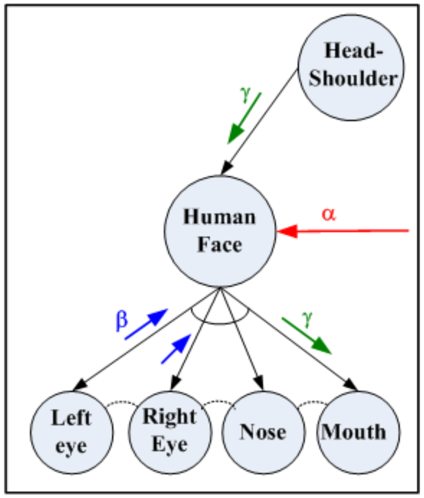


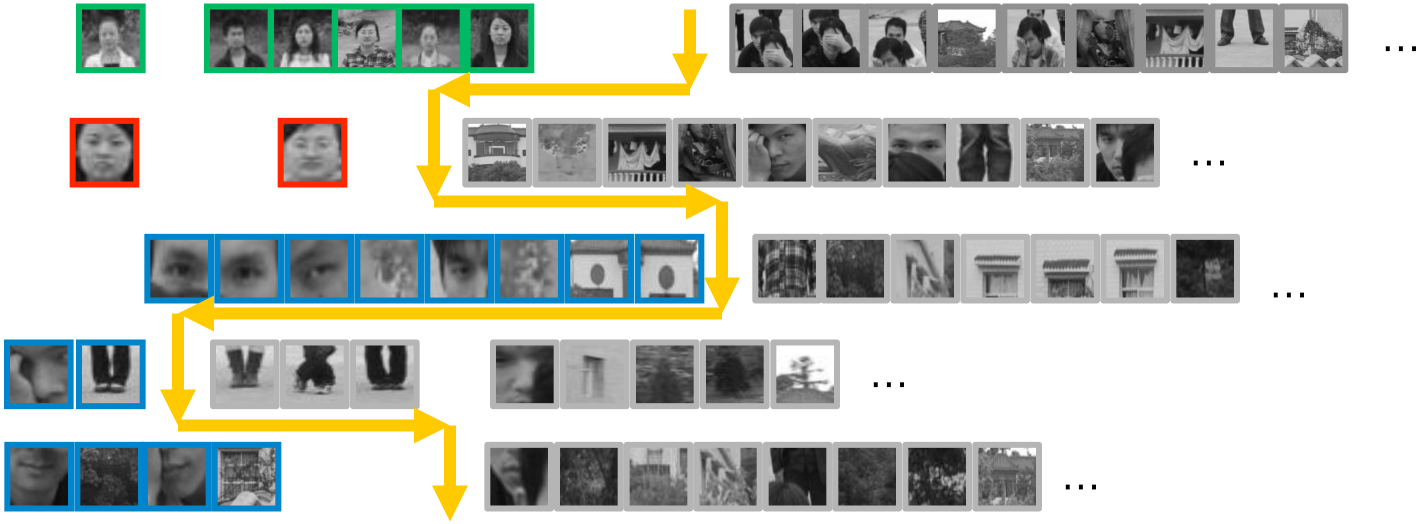


Integrating α, β and γ channels

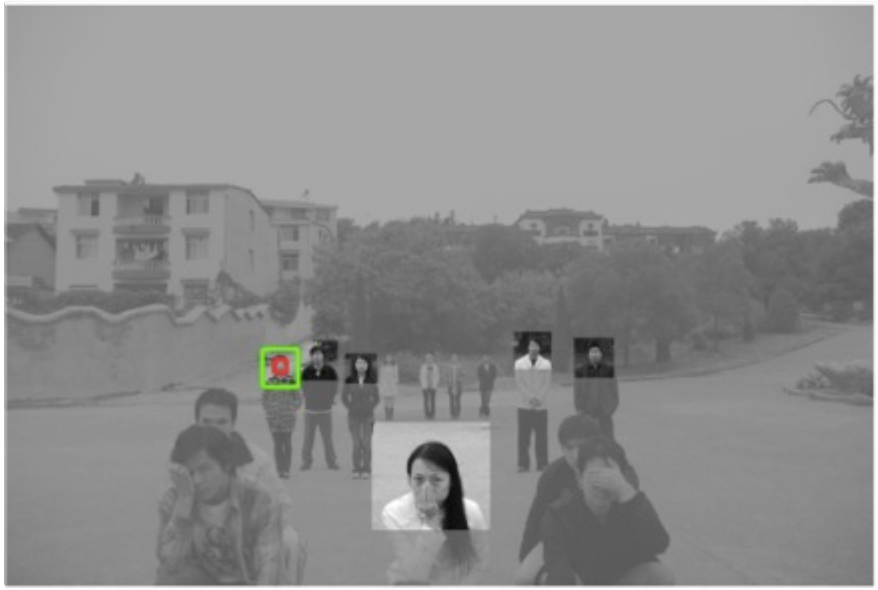


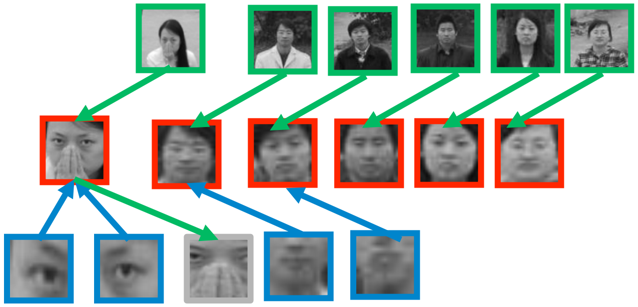


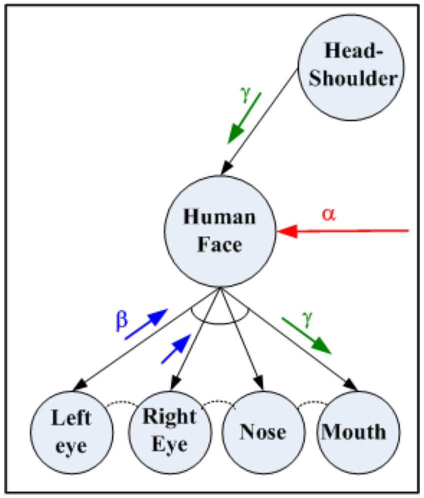


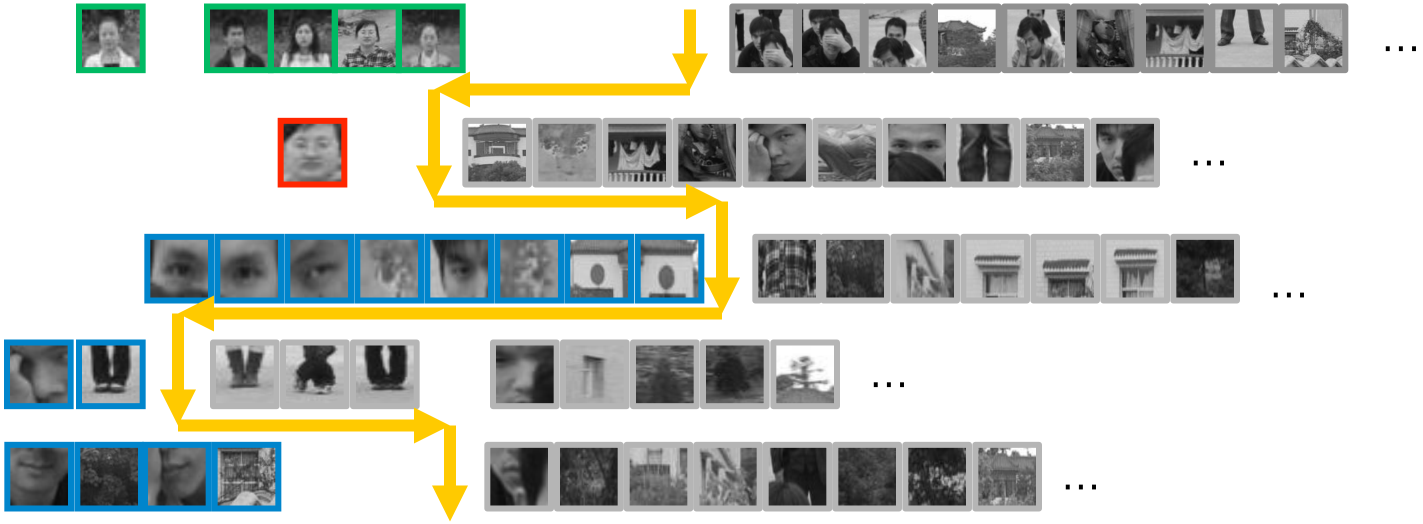


Integrating α, β and γ channels

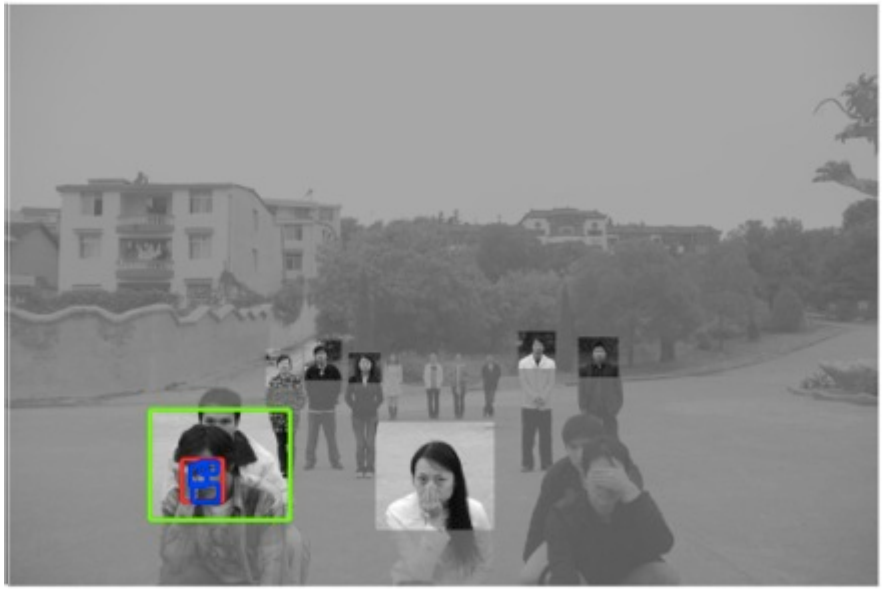


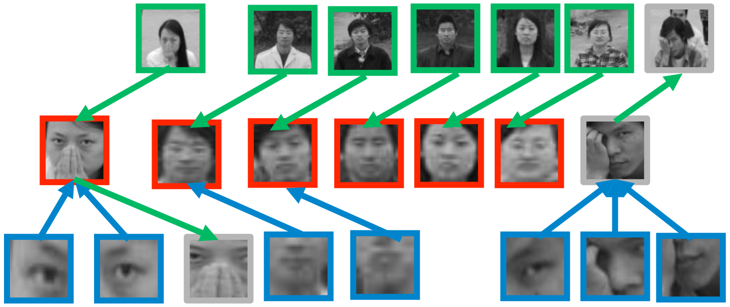


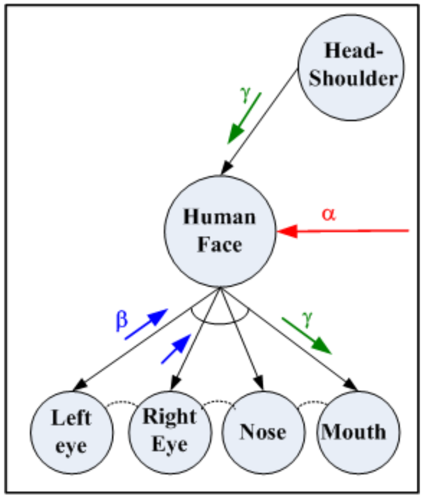


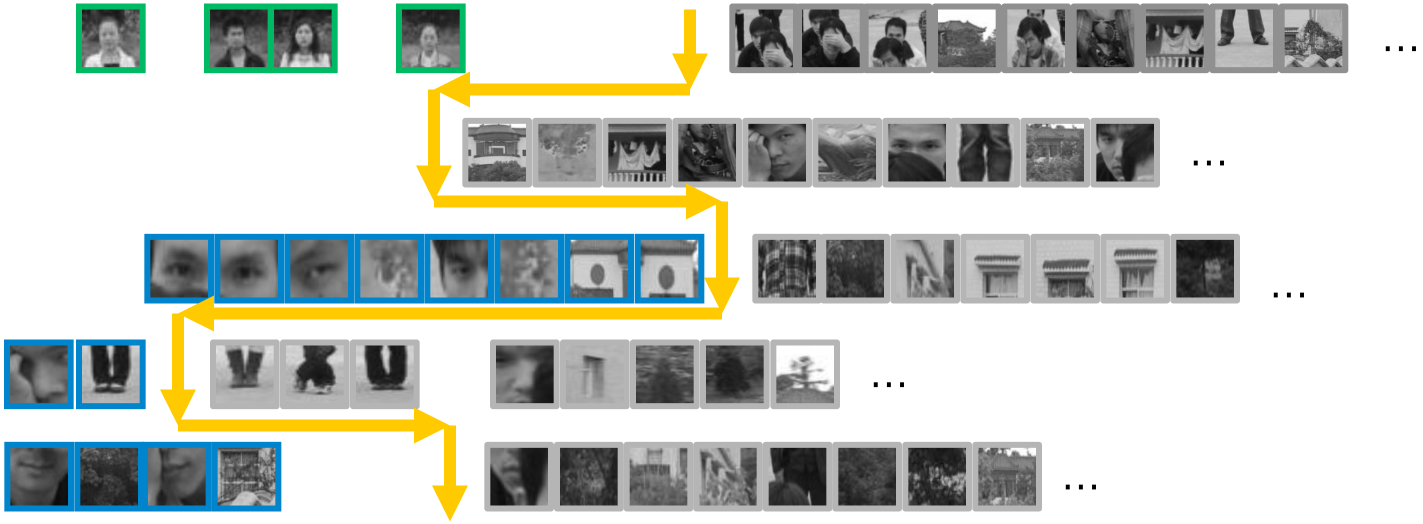


Integrating α, β and γ channels

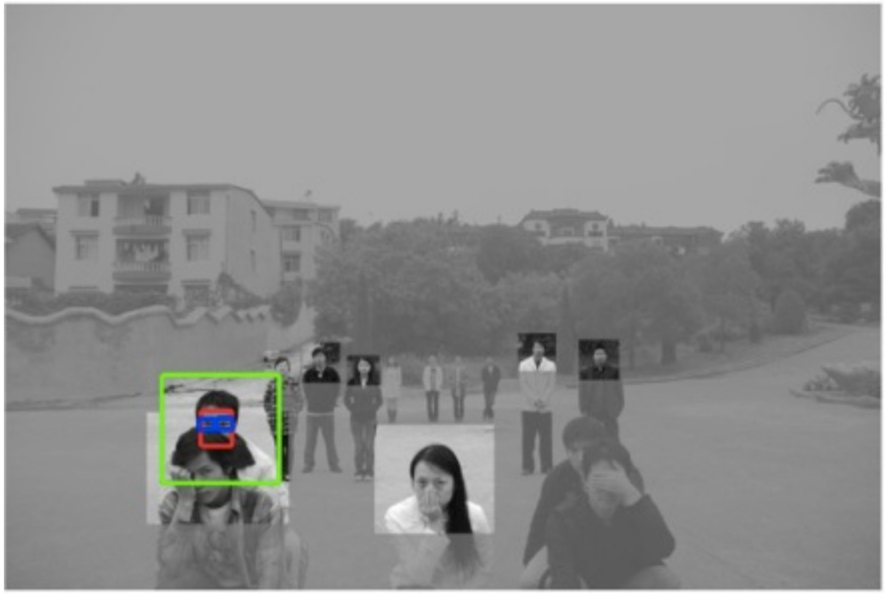


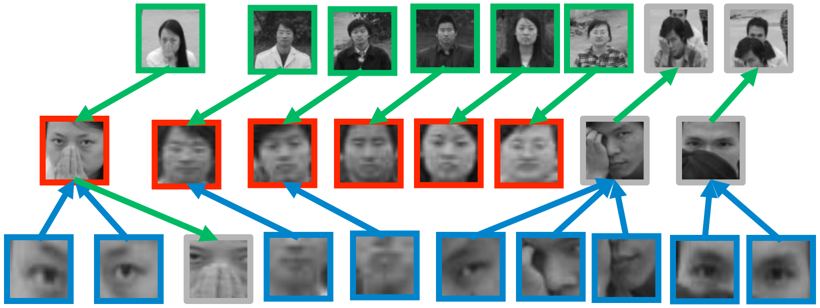


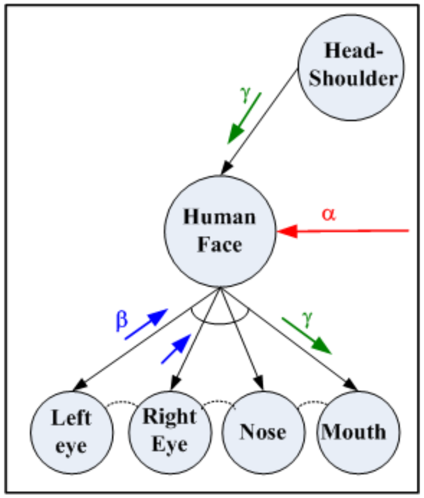


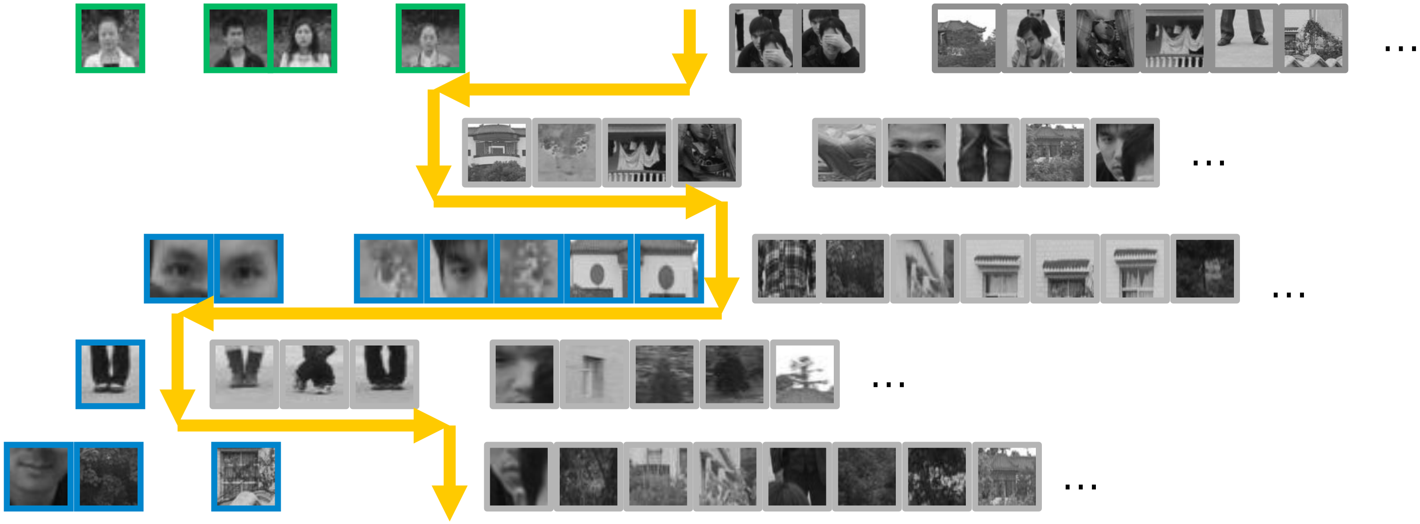


Integrating α, β and γ channels

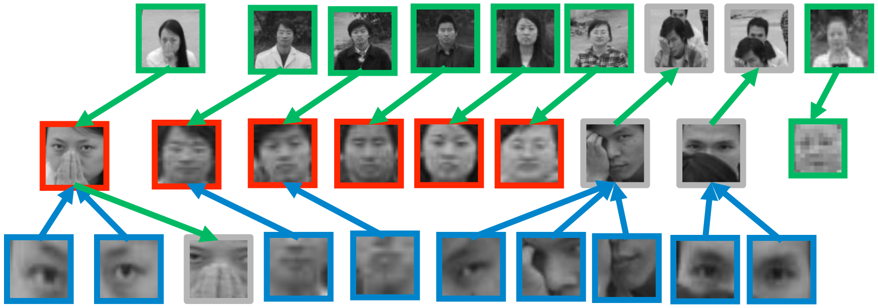


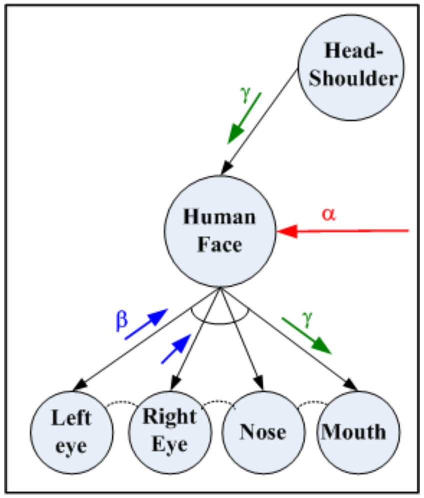


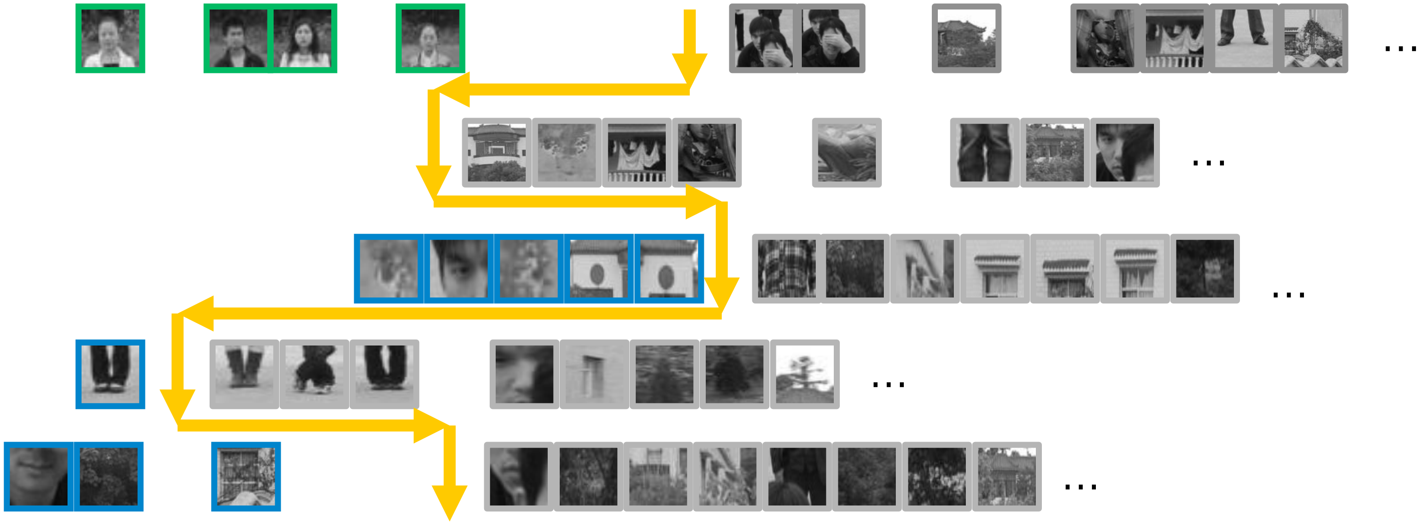




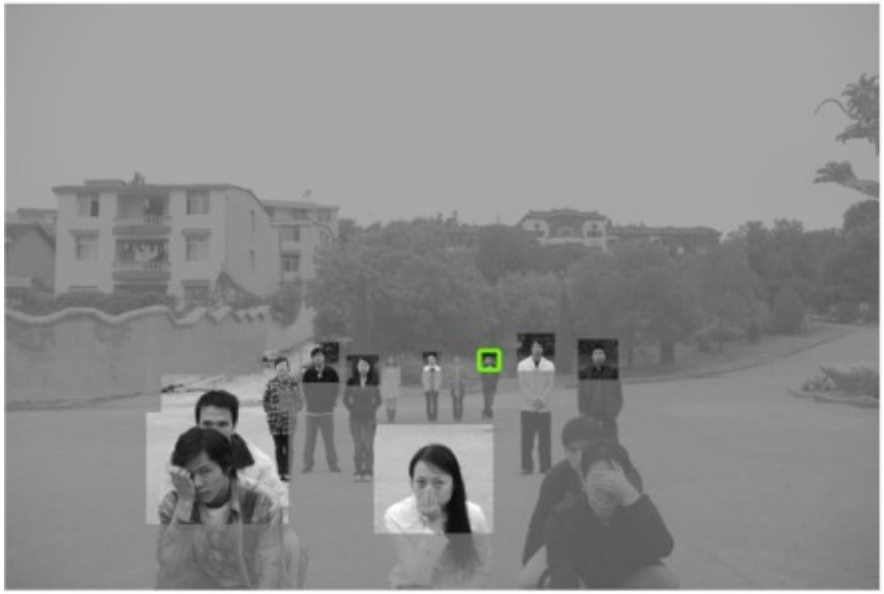
Integrating α, β and γ channels

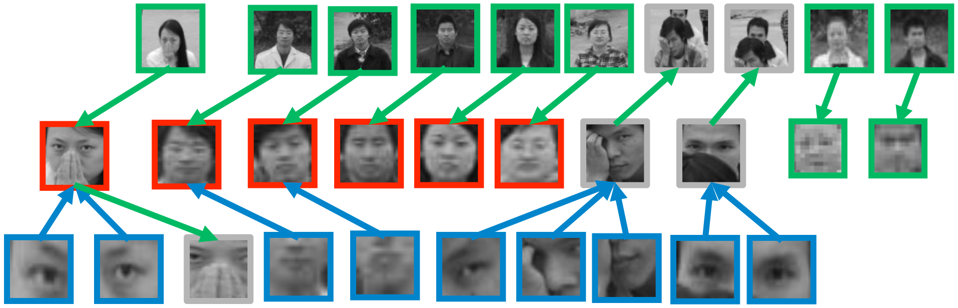


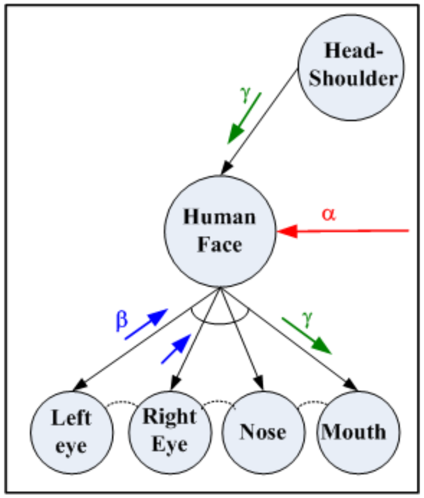


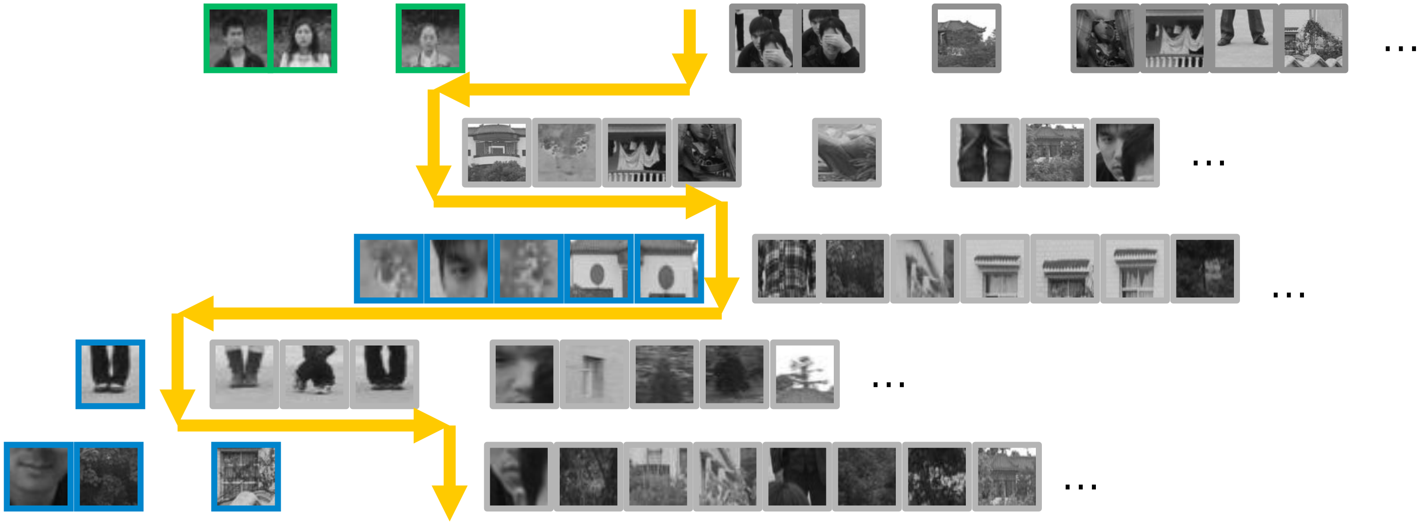


Integrating α, β and γ channels

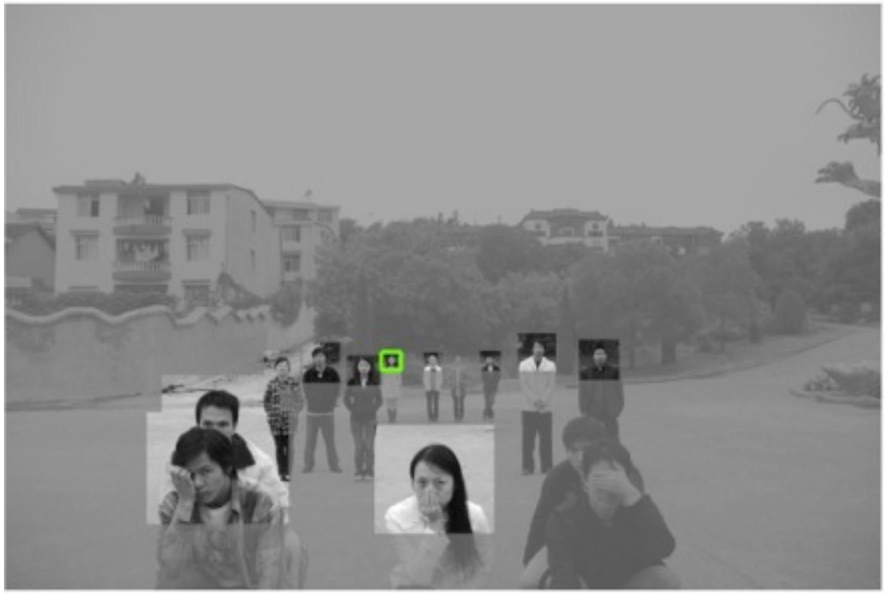


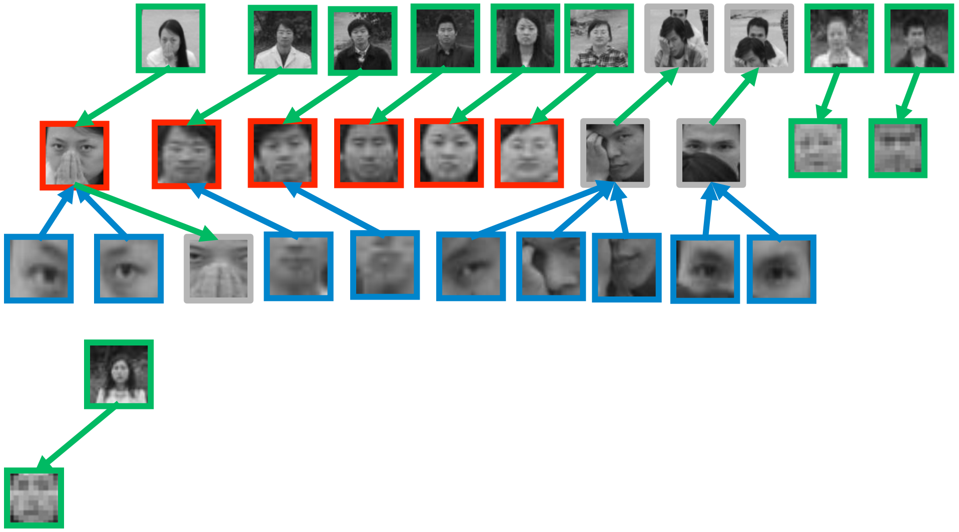


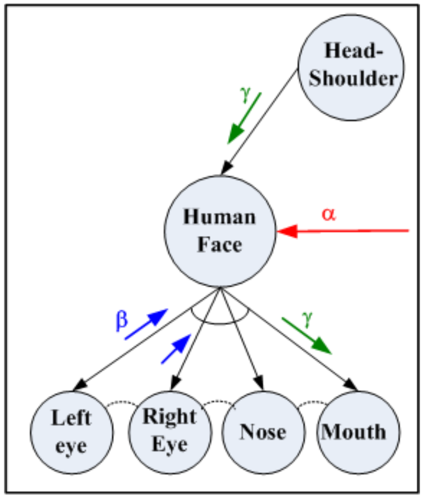


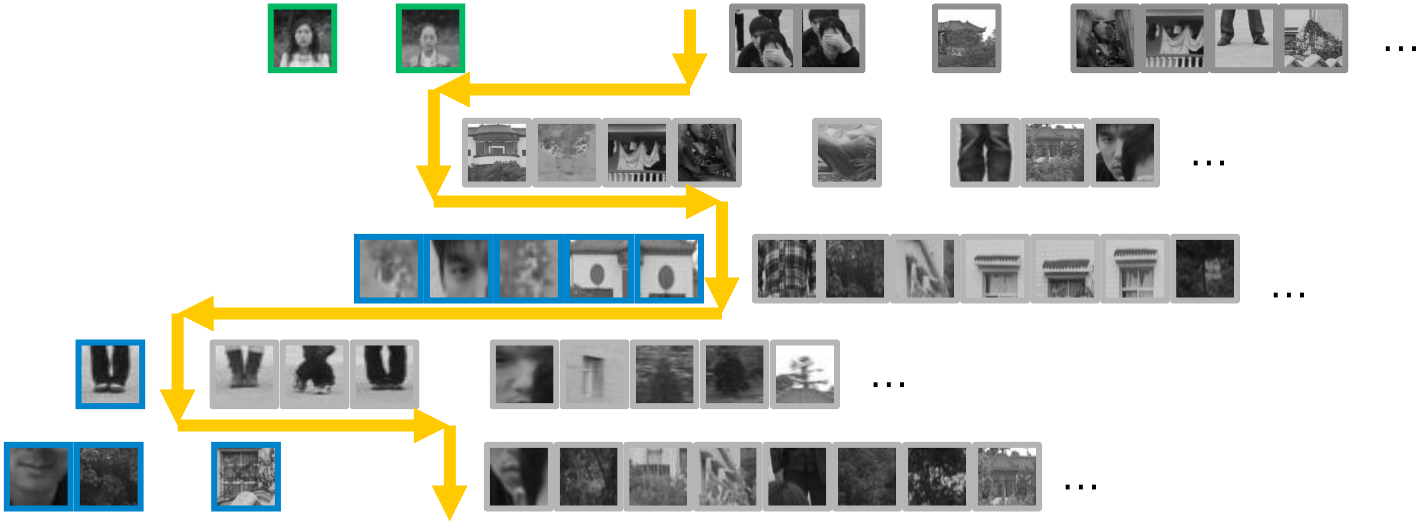


Integrating α, β and γ channels

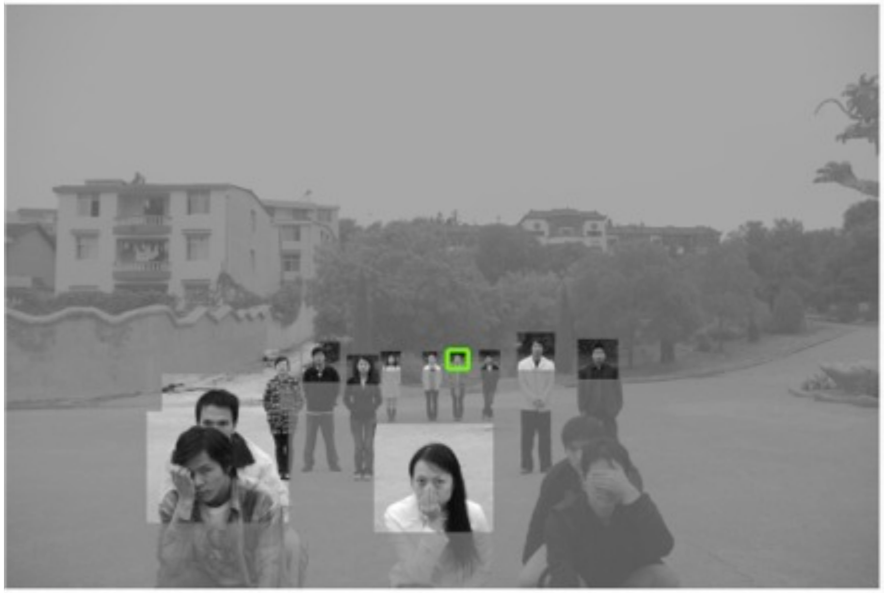


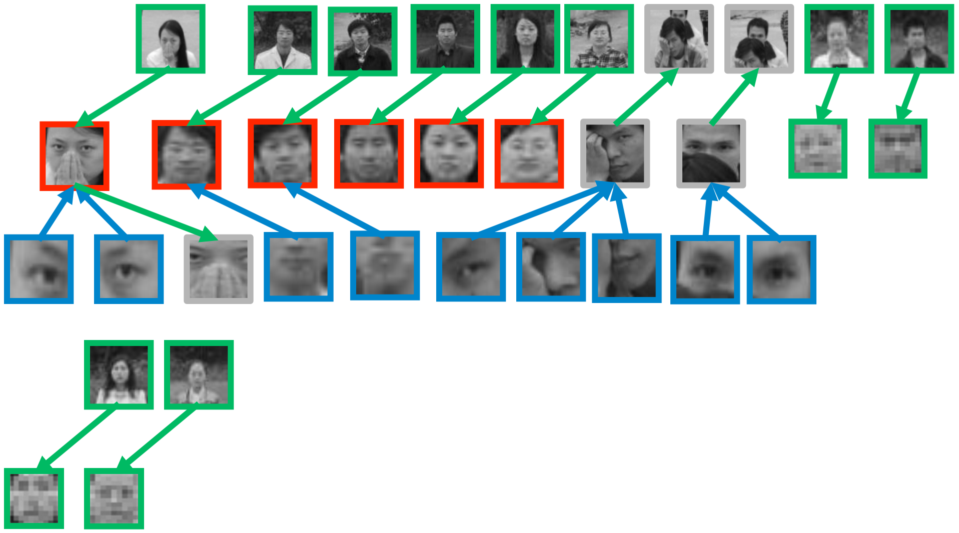


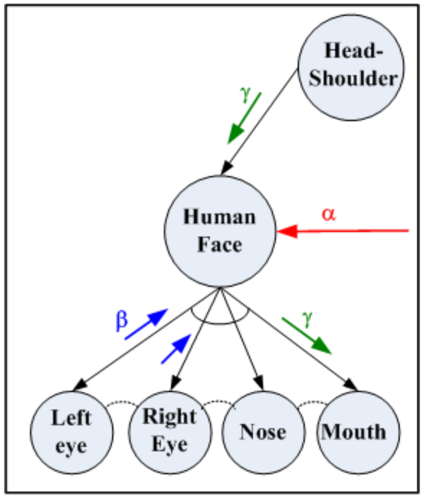


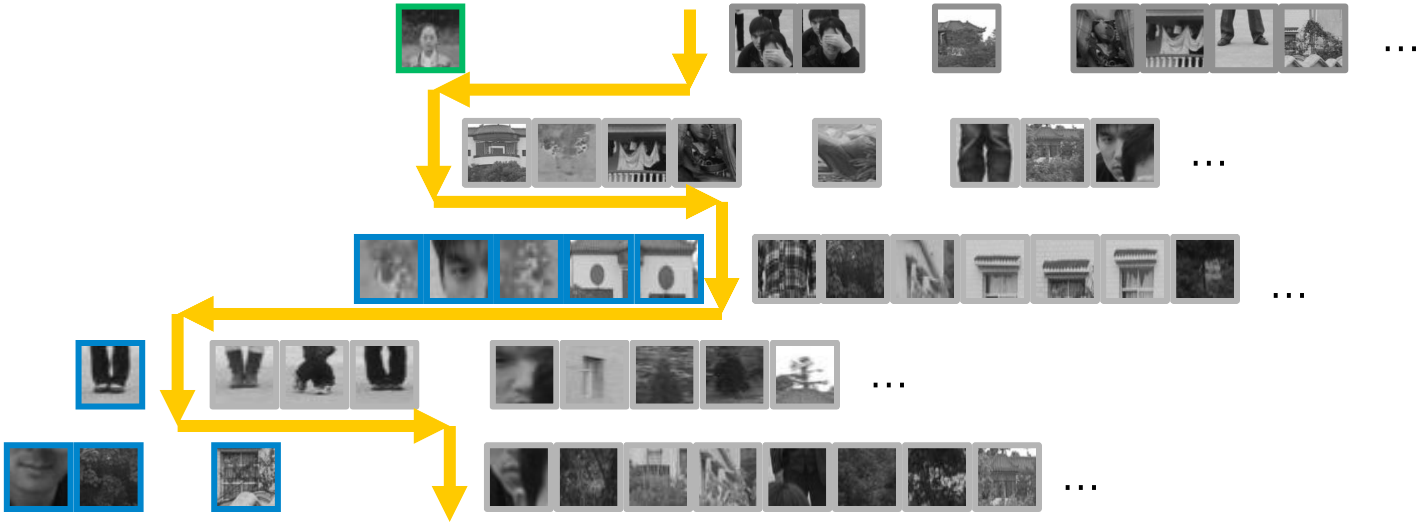


Integrating α, β and γ channels

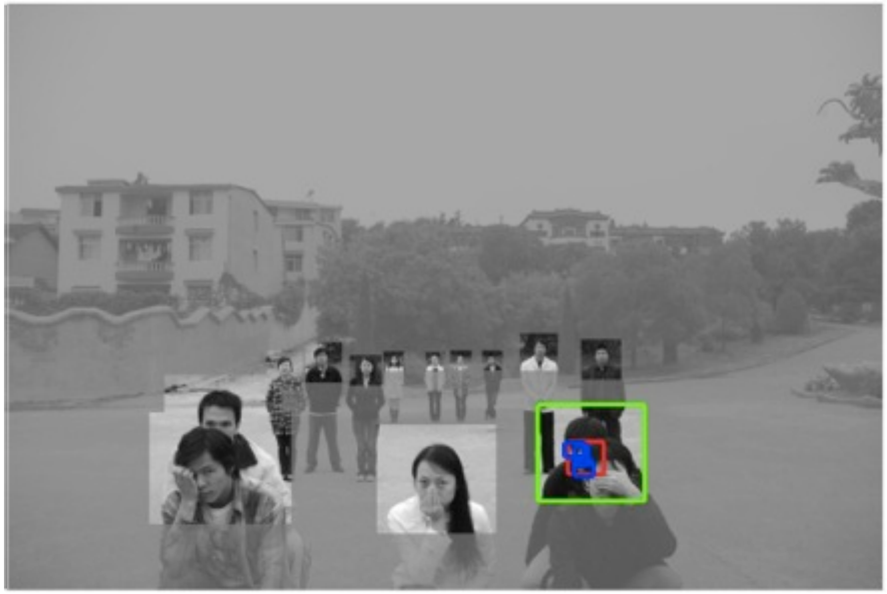


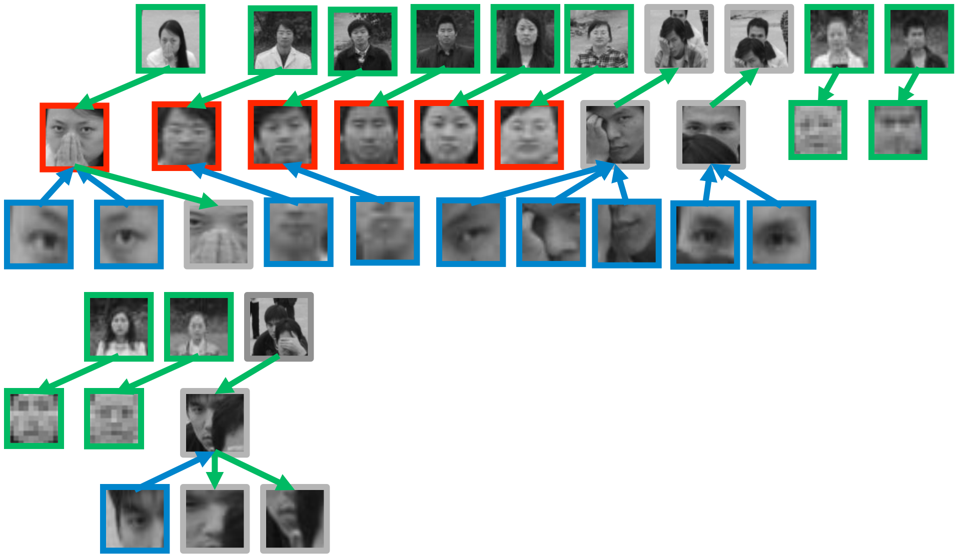


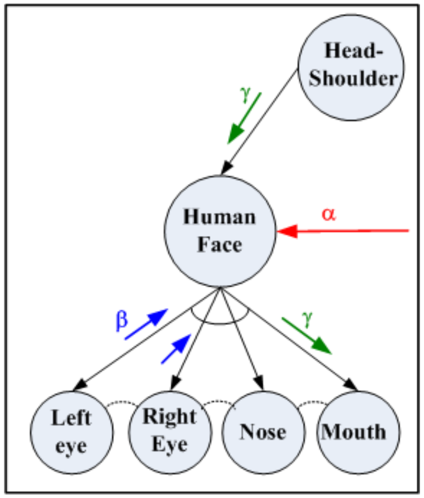


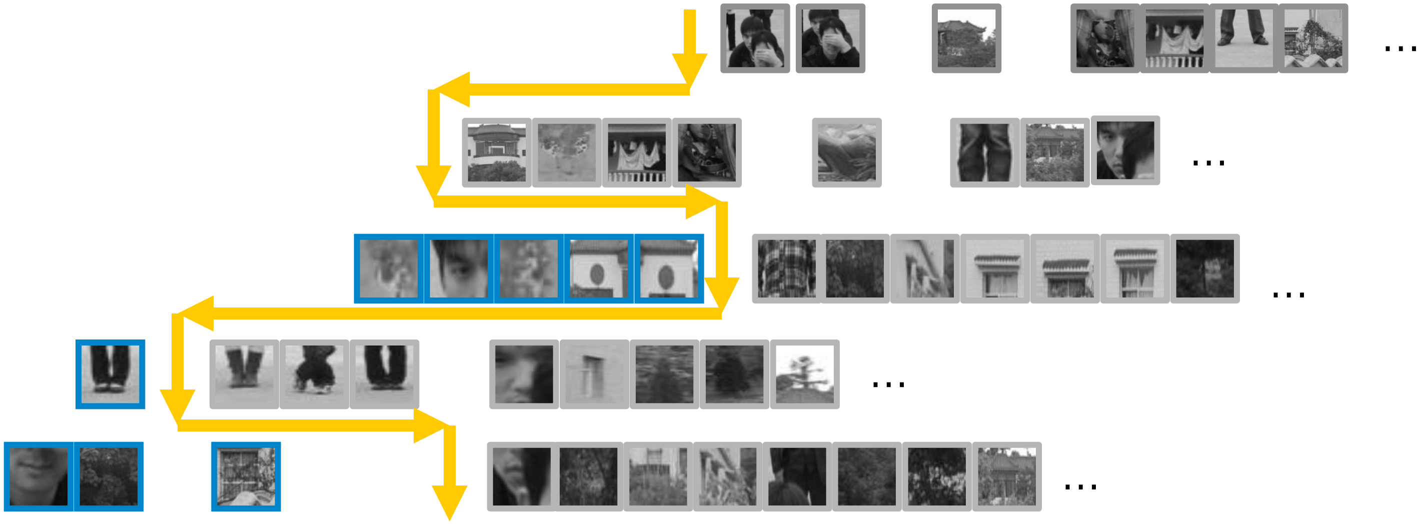


Integrating α, β and γ channels

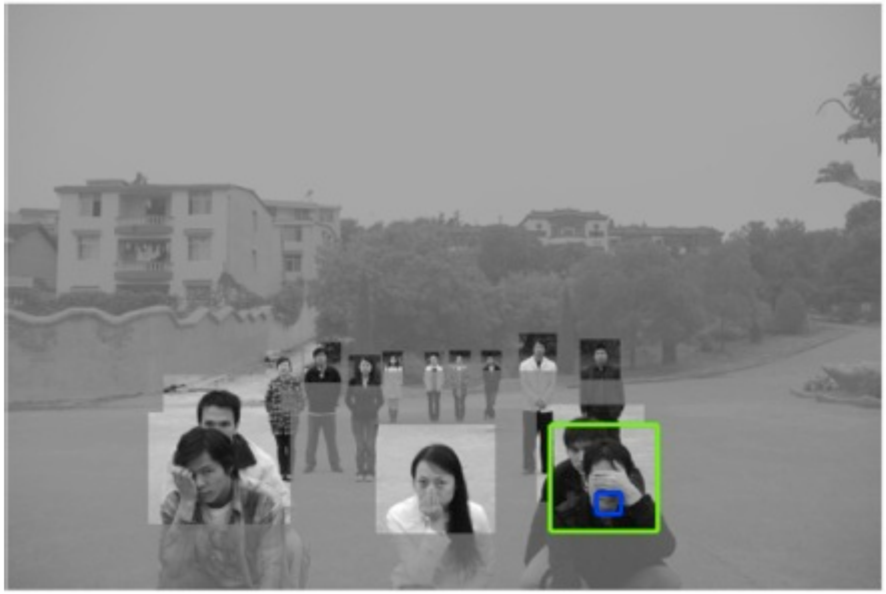


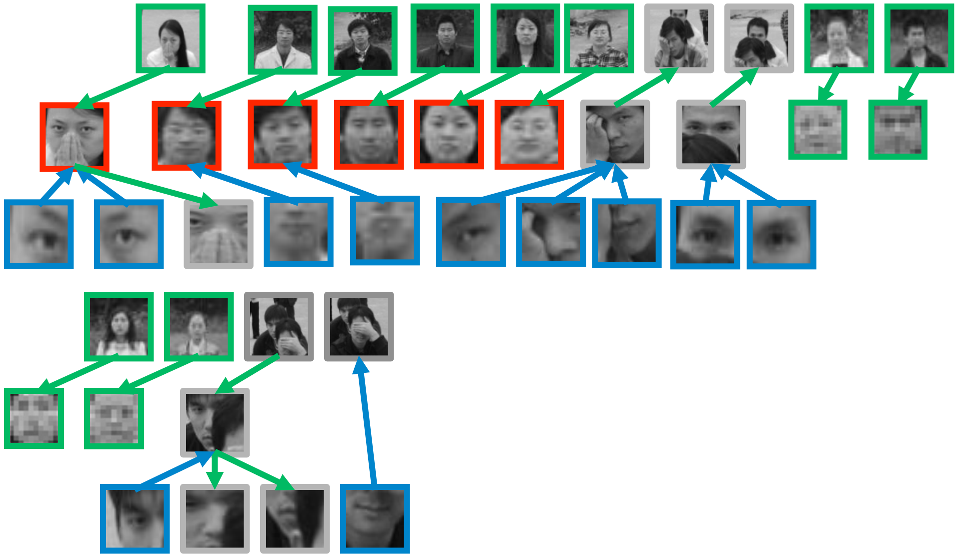


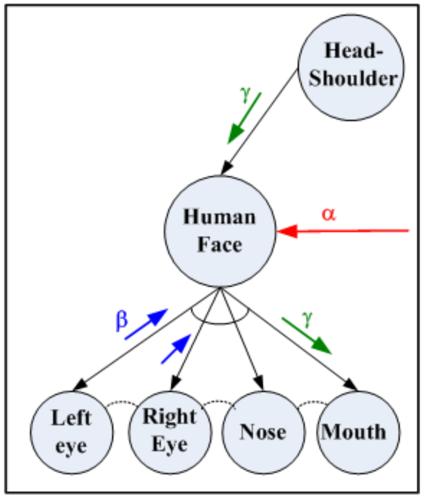


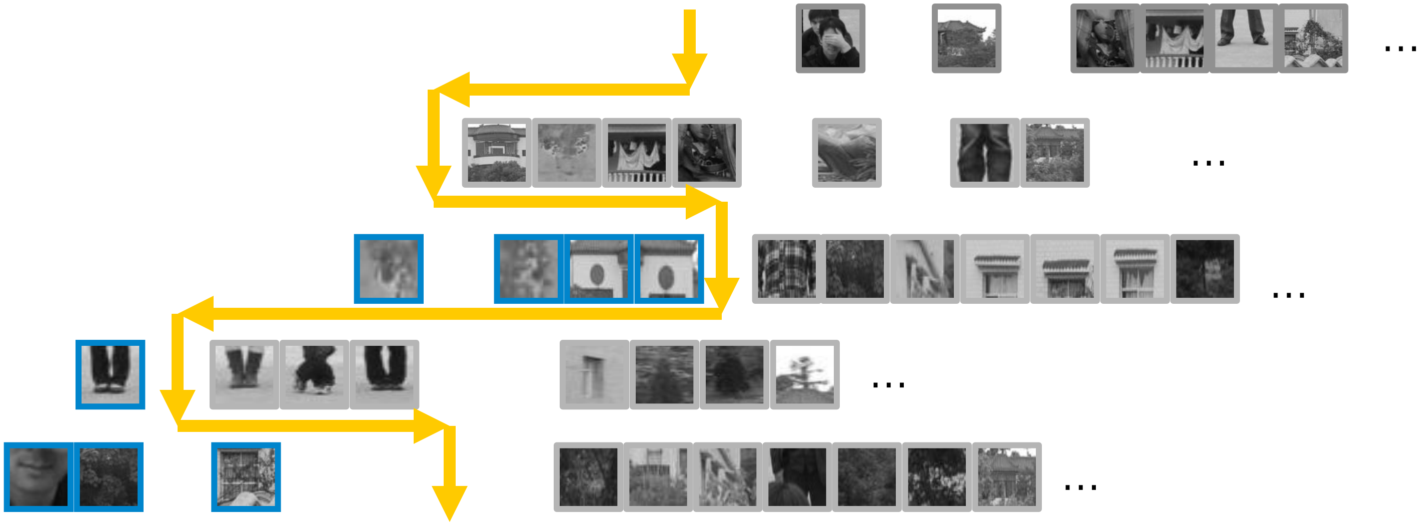


Integrating α, β and γ channels

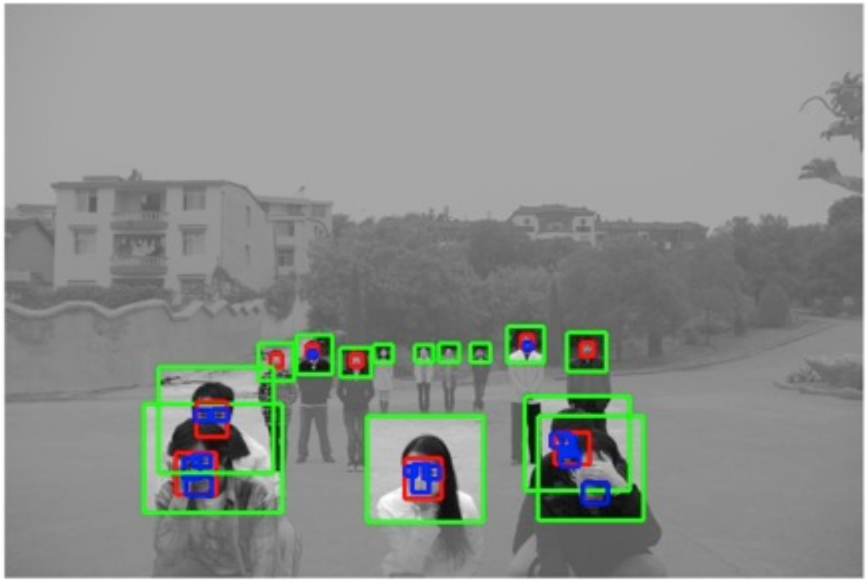


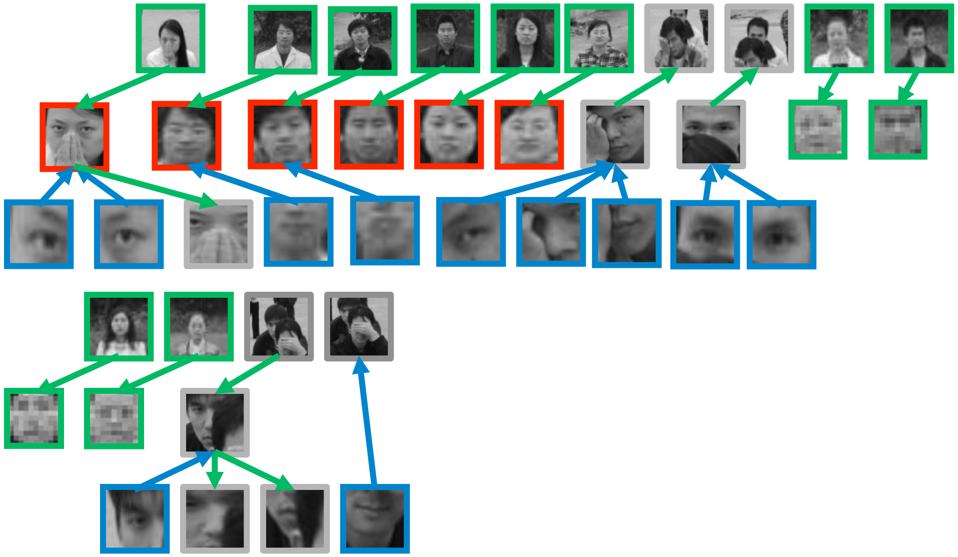


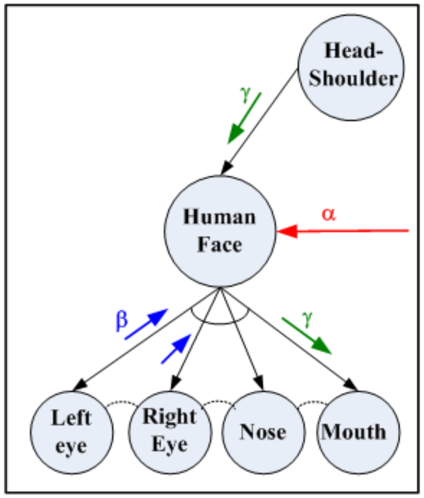


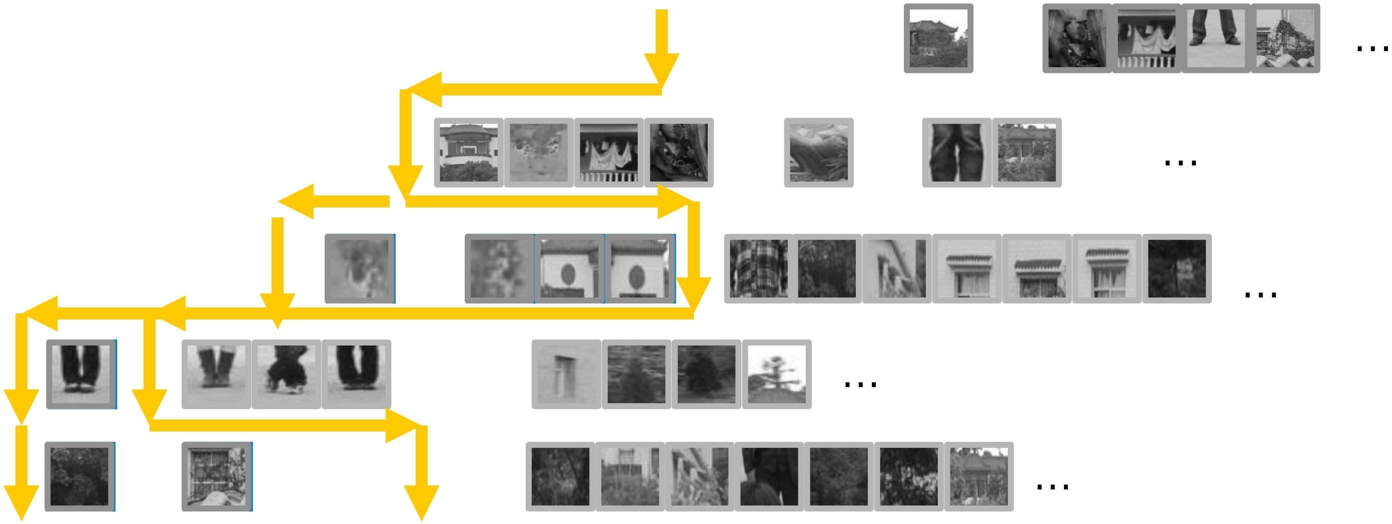


Integrating α, β and γ channels



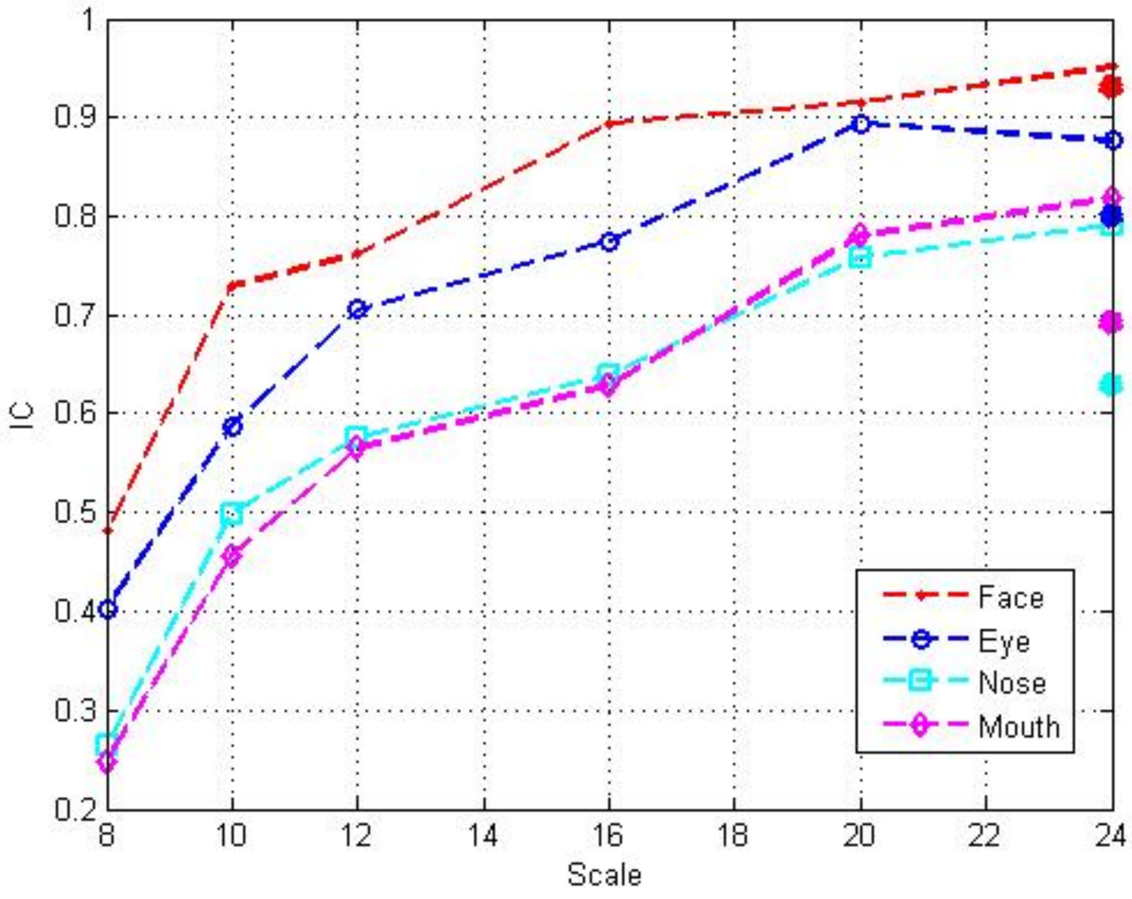






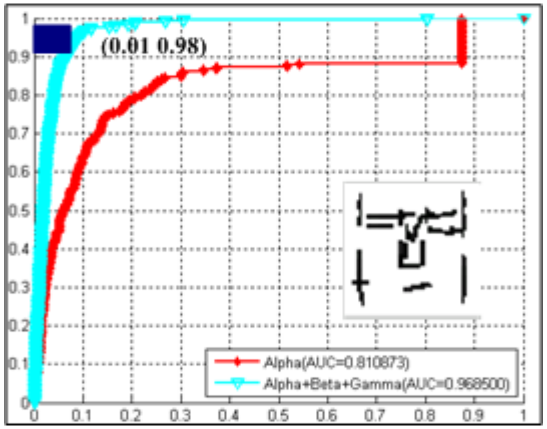
Information contribution

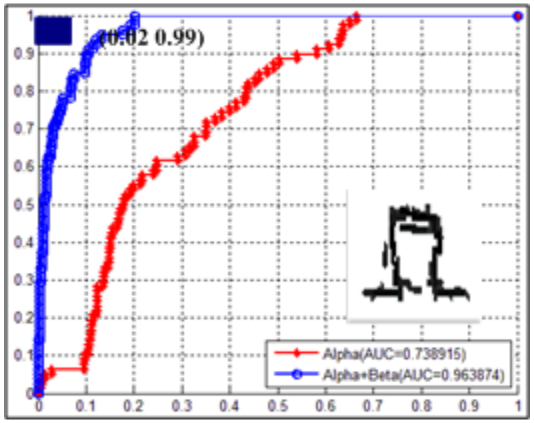
α channels

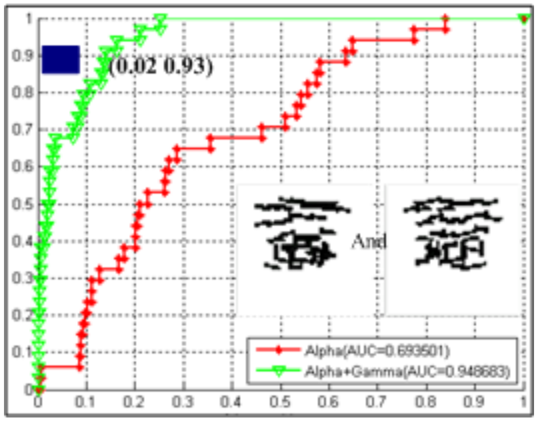


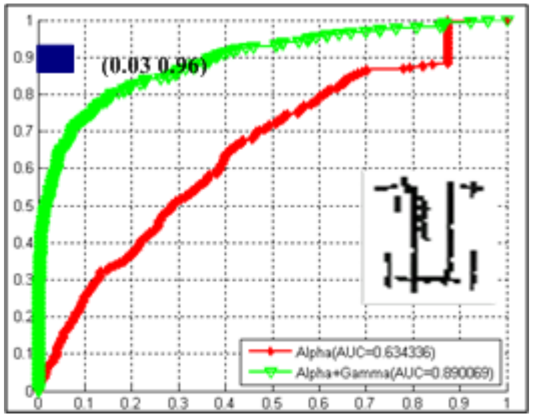
Performance improvement

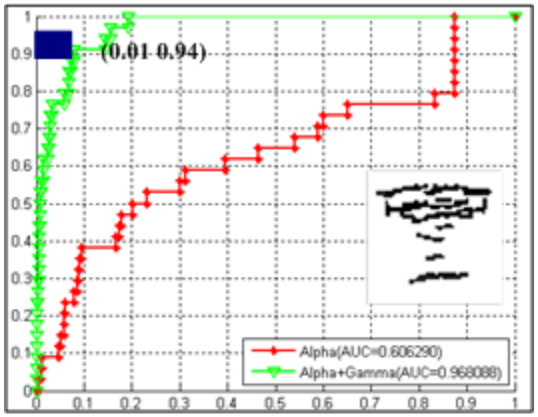
red for α, blue for α+β, green for α+γ, cyan for α+β+γ channels









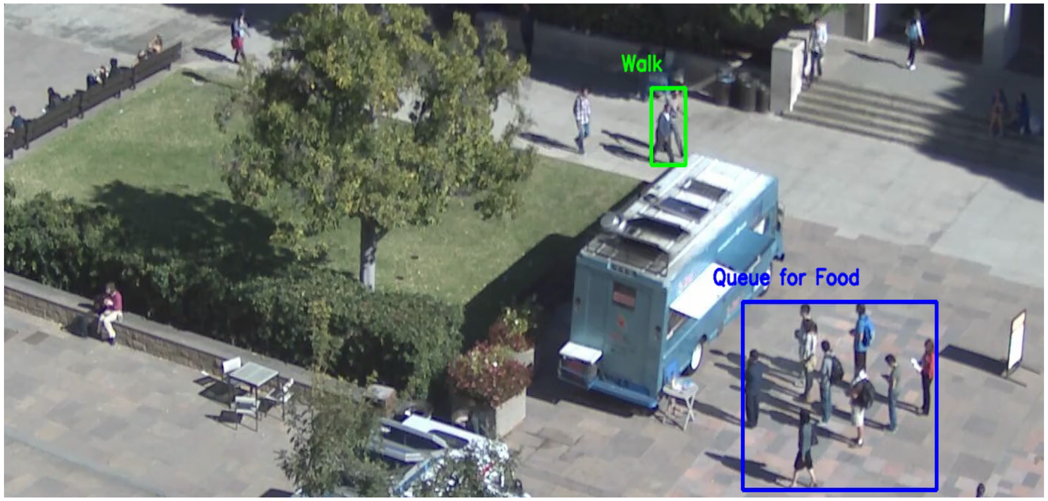


Cost-Sensitive and Goal-Guided Inference in AoG

Song-Chun Zhu, Sinisa Todorovic, and Ales Leonardis

At CVPR, Providence, Rhode Island
June 16, 2012

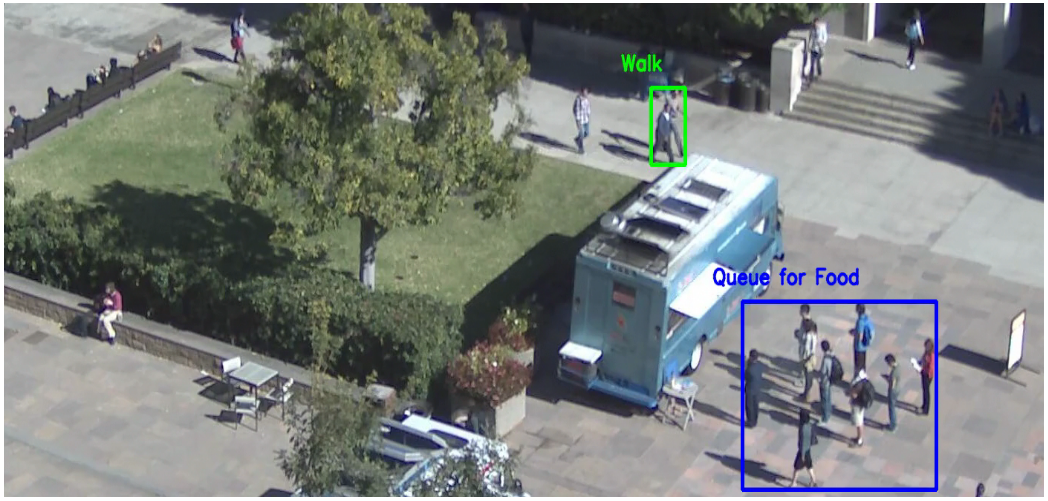
Problem



Given a high-resolution and long video, showing

- Spatially large scenes,
- Many people engaged in a wide range of
  - individual actions
  - group activities

Problem

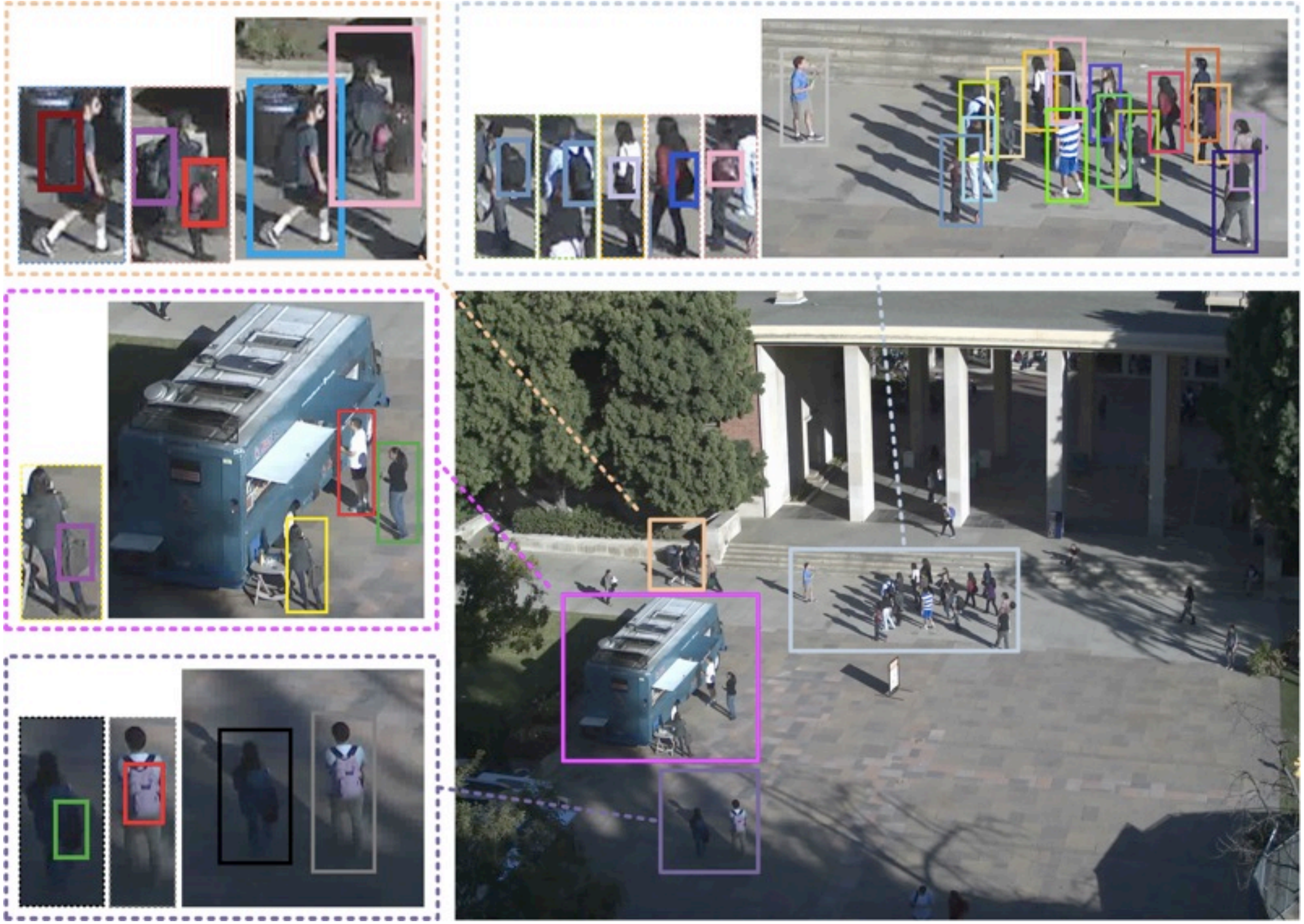


Goal:

Answer WHAT, WHERE, and WHEN queries about

- individual actions
- group activities

Multiscale Event Understanding

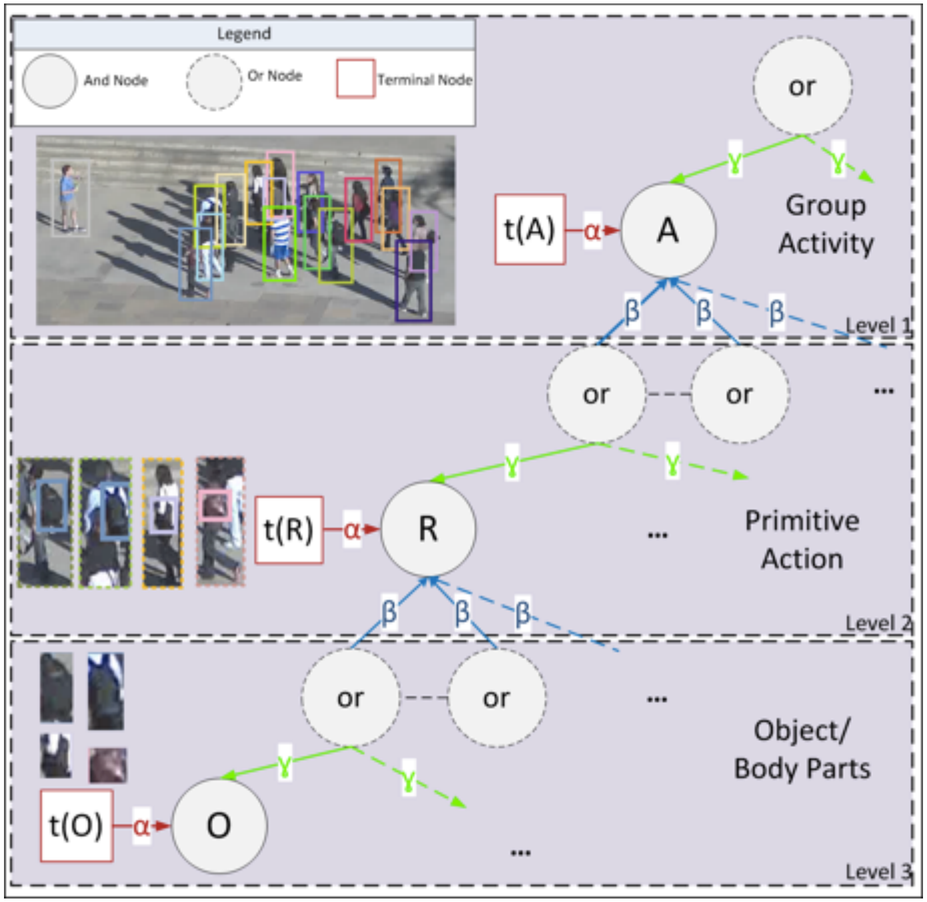


A Principled Framework Needed for

- **Goal-driven** zooming-in or zooming-out,

- **Cost-sensitive** inference via
  - – Scheduling of the bottom-up and top-down inference
  - – Any-time inference -- adapt to a given time budget

And-Or Graph and Parse Graph



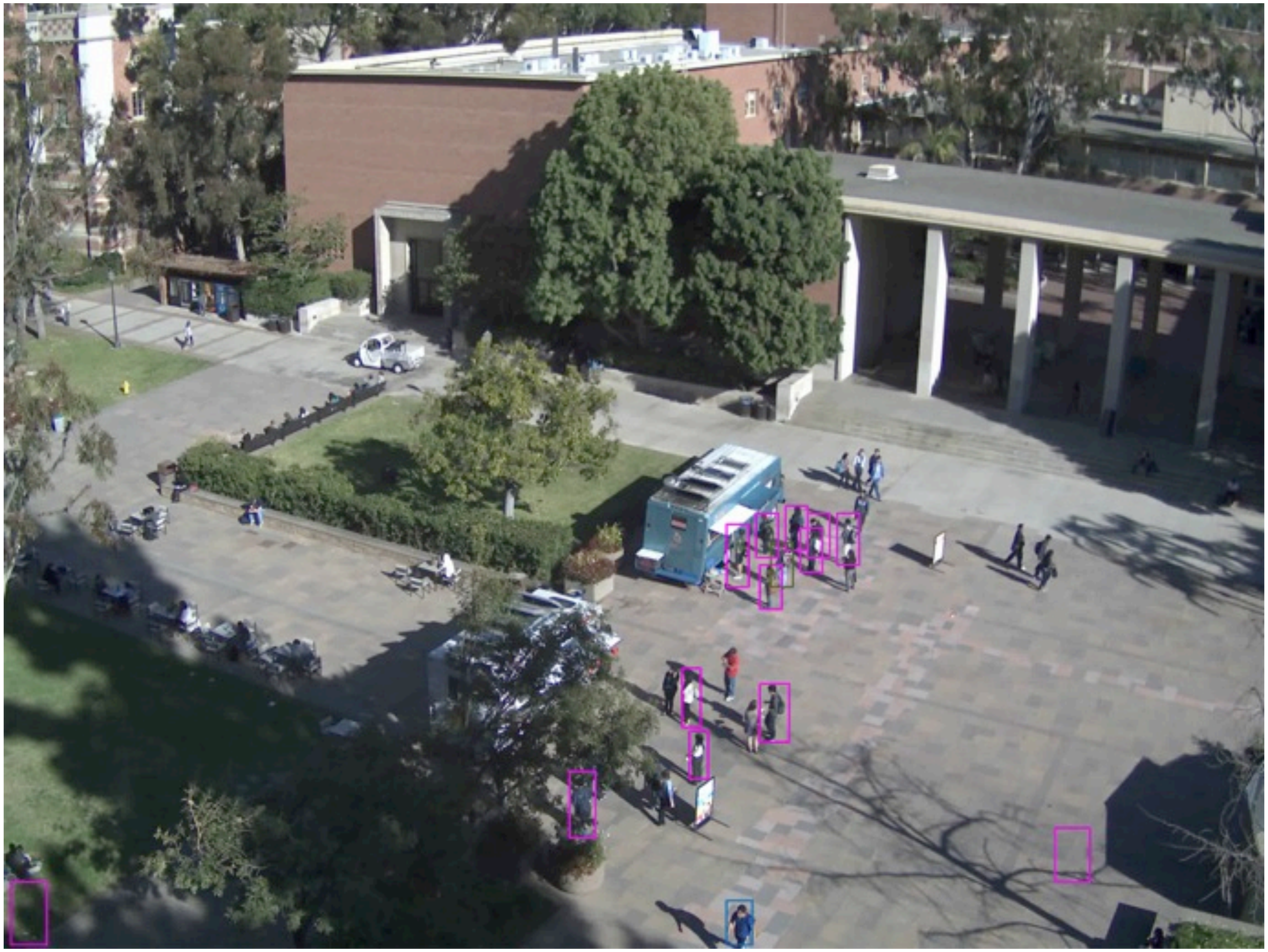
$$pg^{l*} = \arg\max_{pg} \Big[\underbrace{\log p(\wedge^l|\vee^l) + p(N)\sum_{i=1}^{N} \log p(X(\wedge_i^{l+})|X(\wedge^l))}_{\text{parse graph connectivity}}$$

$$\underbrace{+ \ \log \frac{p(\Delta(t(\wedge^l))|t(\wedge^l))}{q(\Delta(t(\wedge^l)))}}_{\alpha \ \text{= detector}}$$

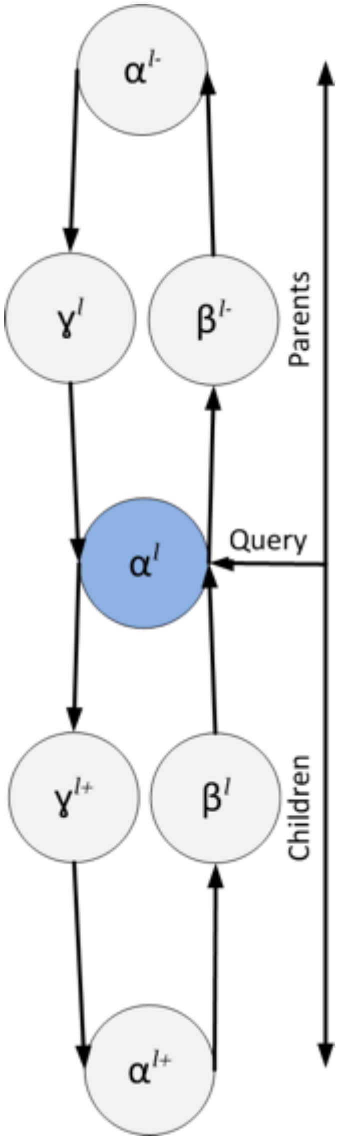
$$\underbrace{+\, p(N)\Big[\sum_{i=1}^{N} \log \frac{p(\Delta(t(\wedge_i^{l+}))|t(\wedge_i^{l+}))}{q(\Delta(t(\wedge_i^{l+})))} + \sum_{i\neq j} \log p(X(\wedge_i^{l+}), X(\wedge_j^{l+}))\Big]}_{\beta \ \text{= zoom-in}}$$

$$+ \ \underbrace{\log \frac{p(\Delta(t(\wedge^{l-}))|t(\wedge^{l-}))}{q(\Delta(t(\wedge^{l-})))} + \log p(X(\wedge^l)|X(\wedge^{l-}))}_{\gamma \ \text{= zoom-out}}$$

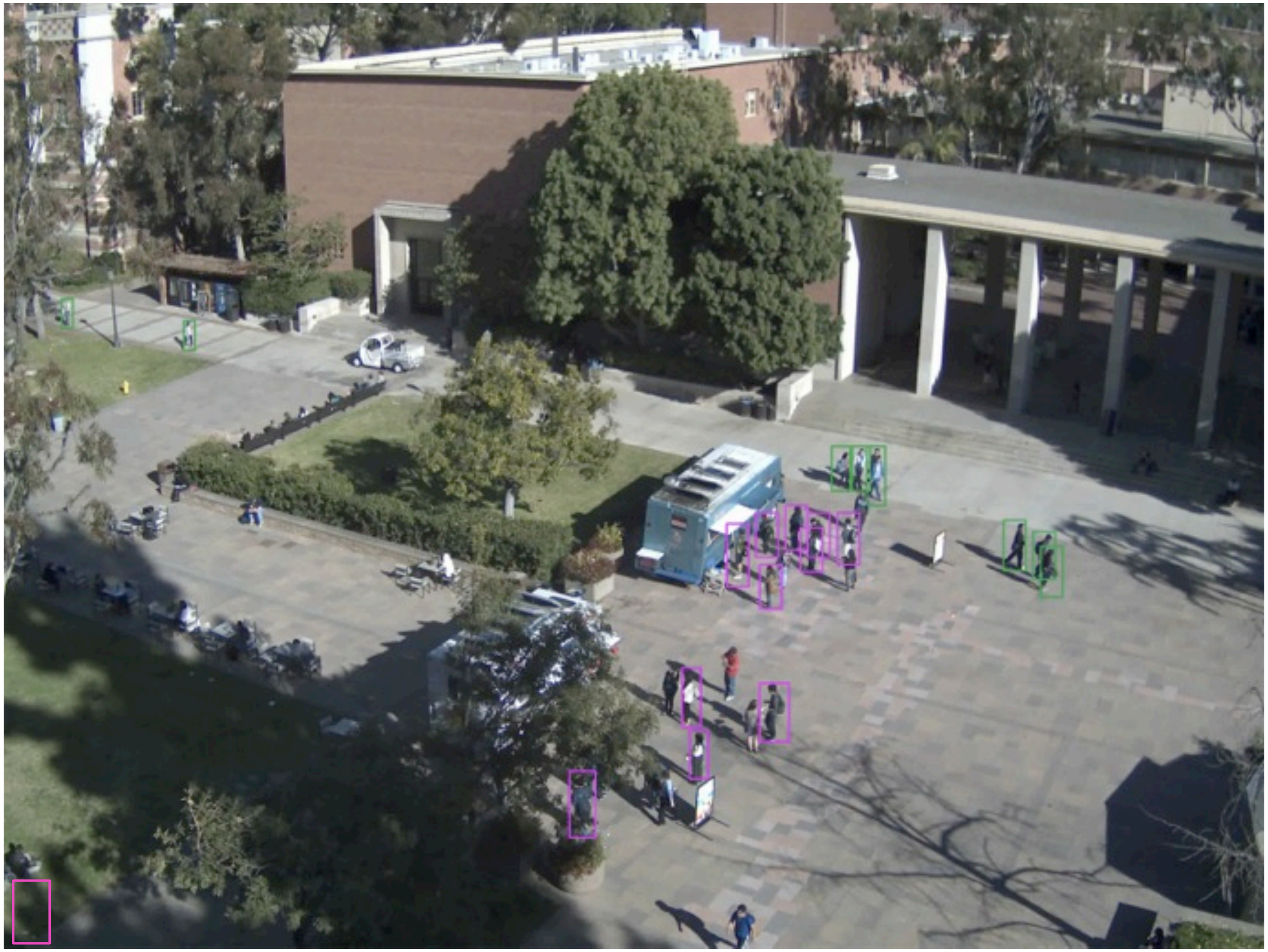
Inference -- Alpha of Individual Action



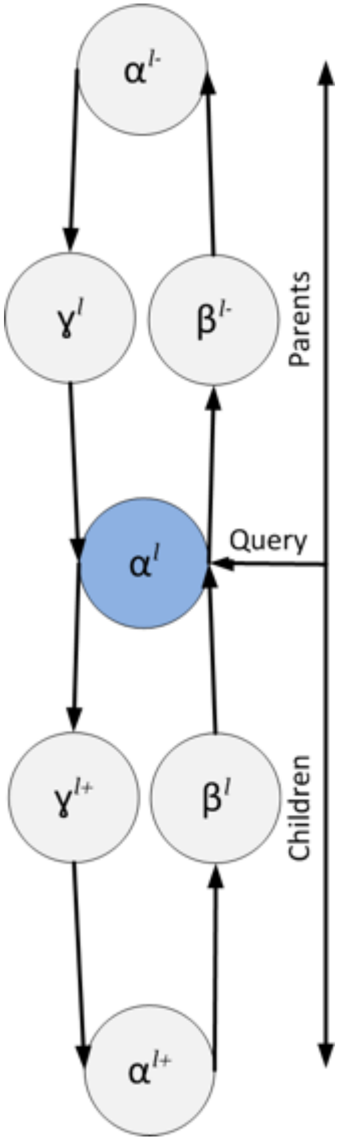
🟪 Waiting



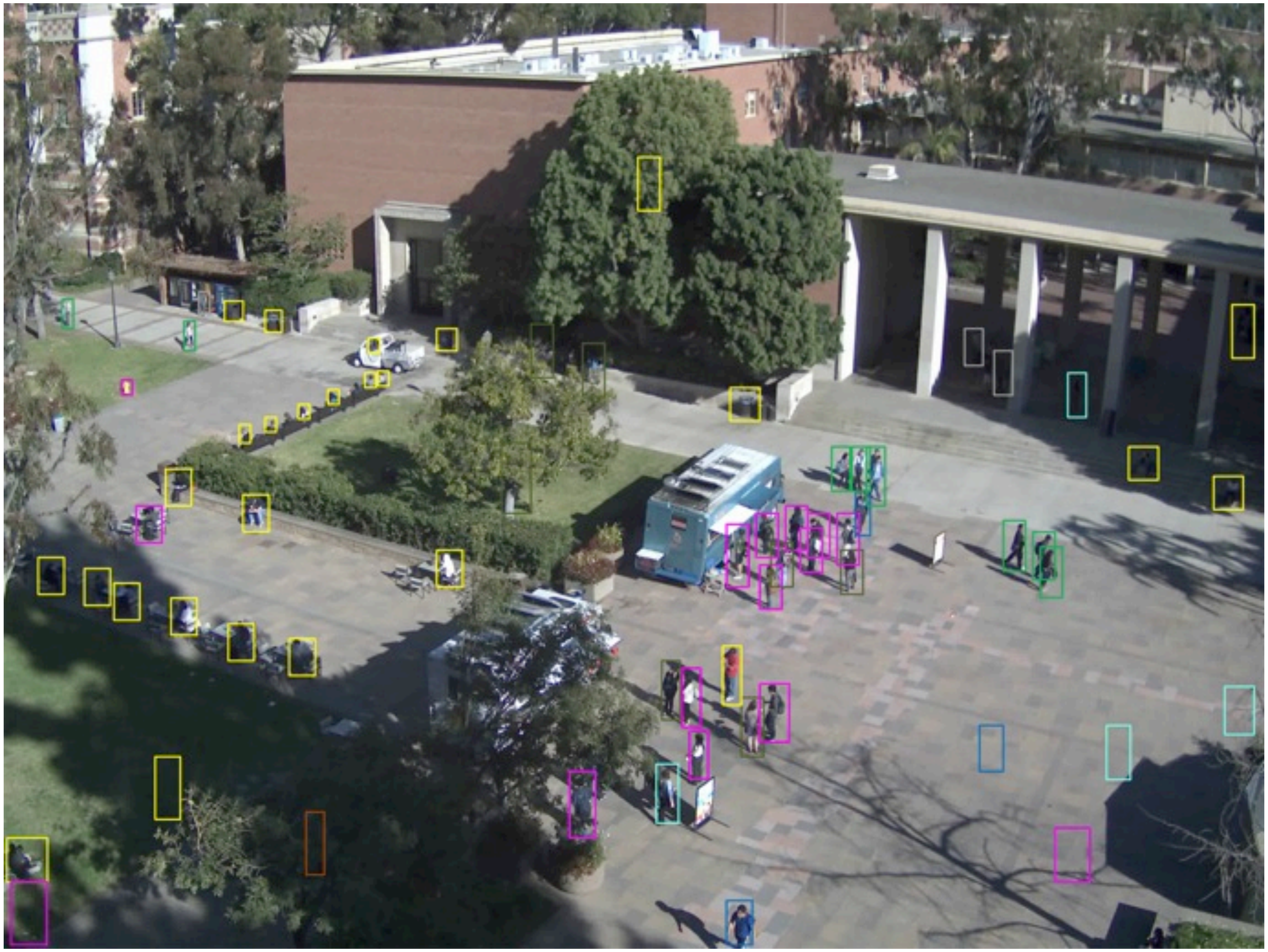
Inference -- Alpha of Individual Action



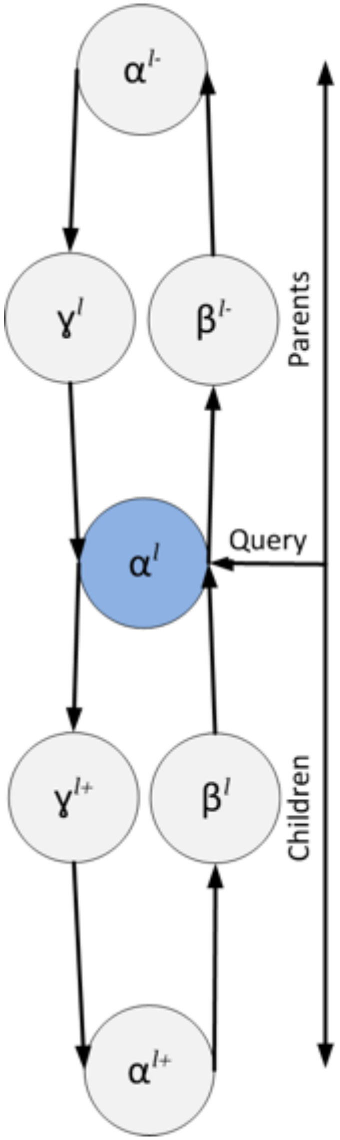
🟪 Waiting 🟩 Walking



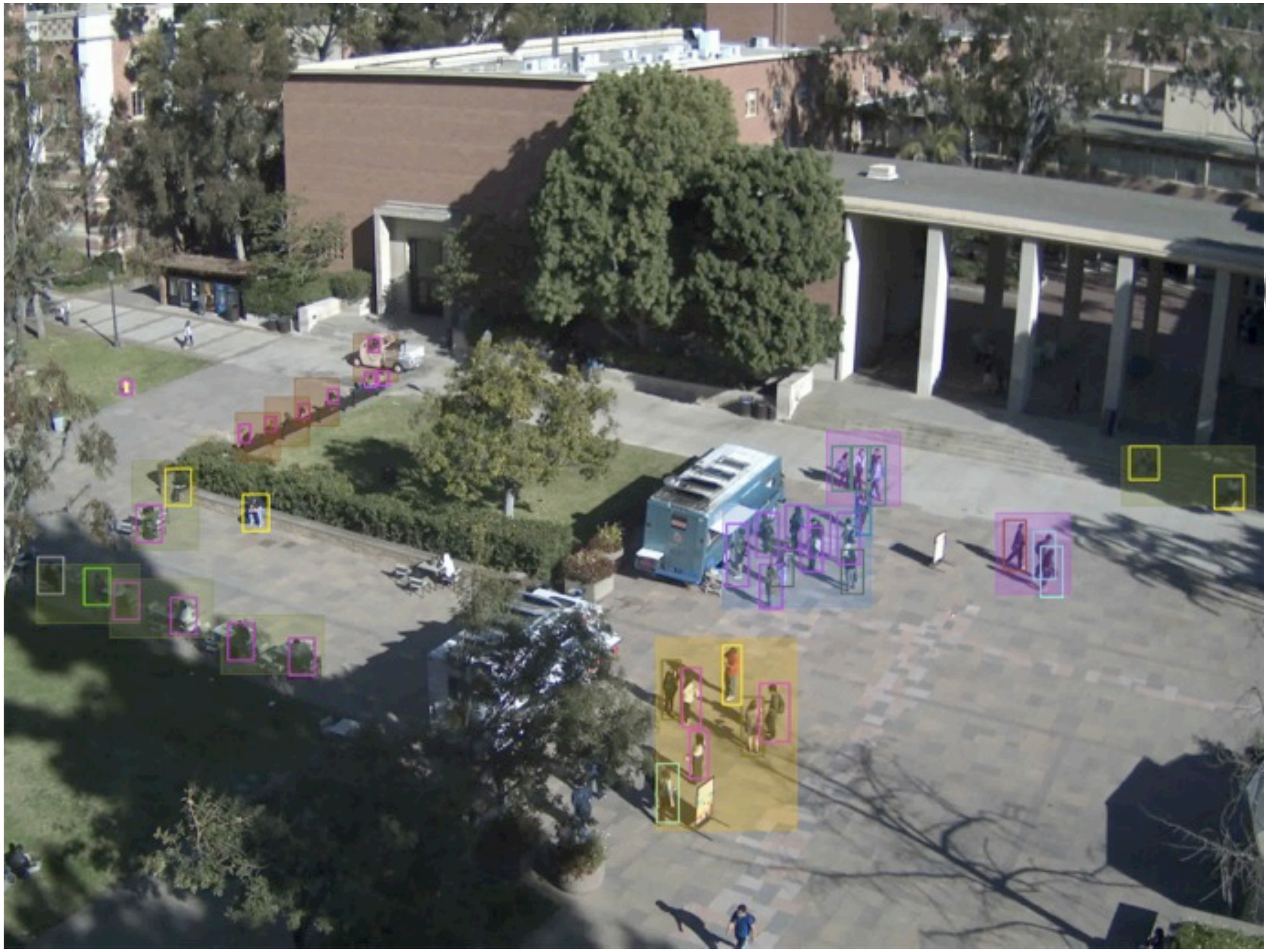
Inference -- Alpha of Individual Action



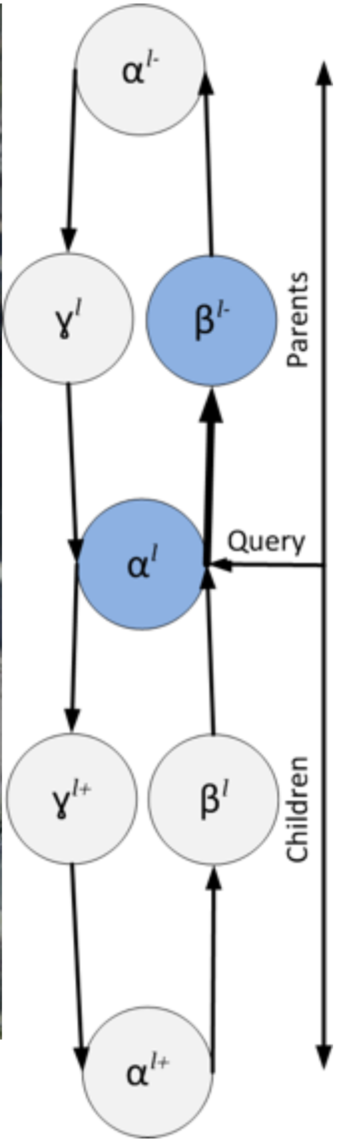
🟪 Waiting 🟩 Walking 🟨 🟥 ⬜



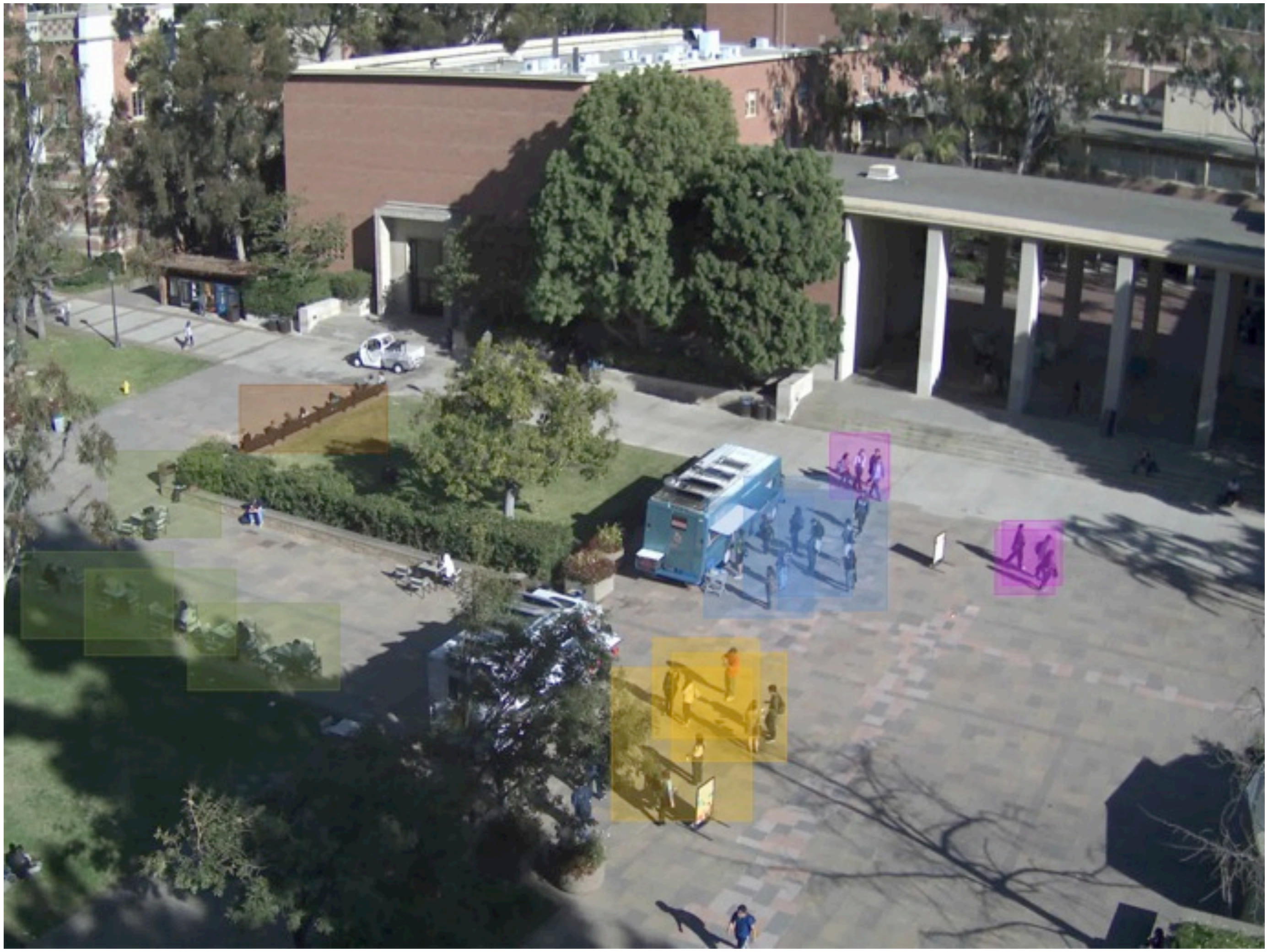
Inference -- Bottom Up (Beta)



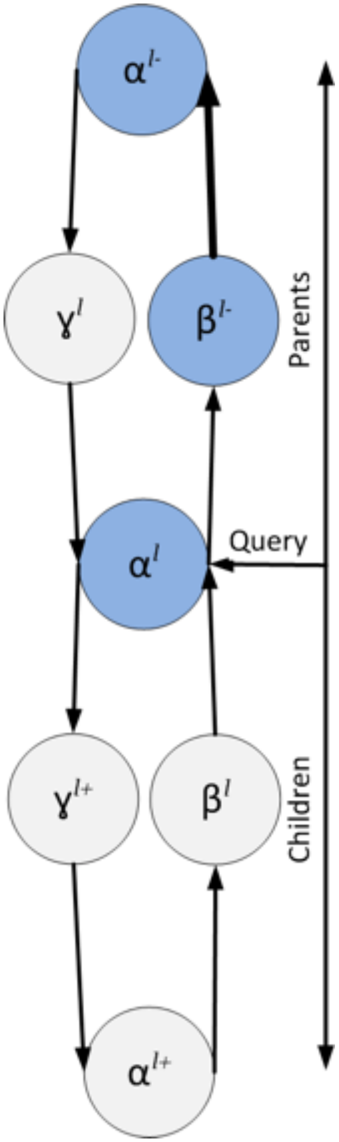
🟪 Group Walking 🟩 Group Discuss 🟨 🟥 🟦



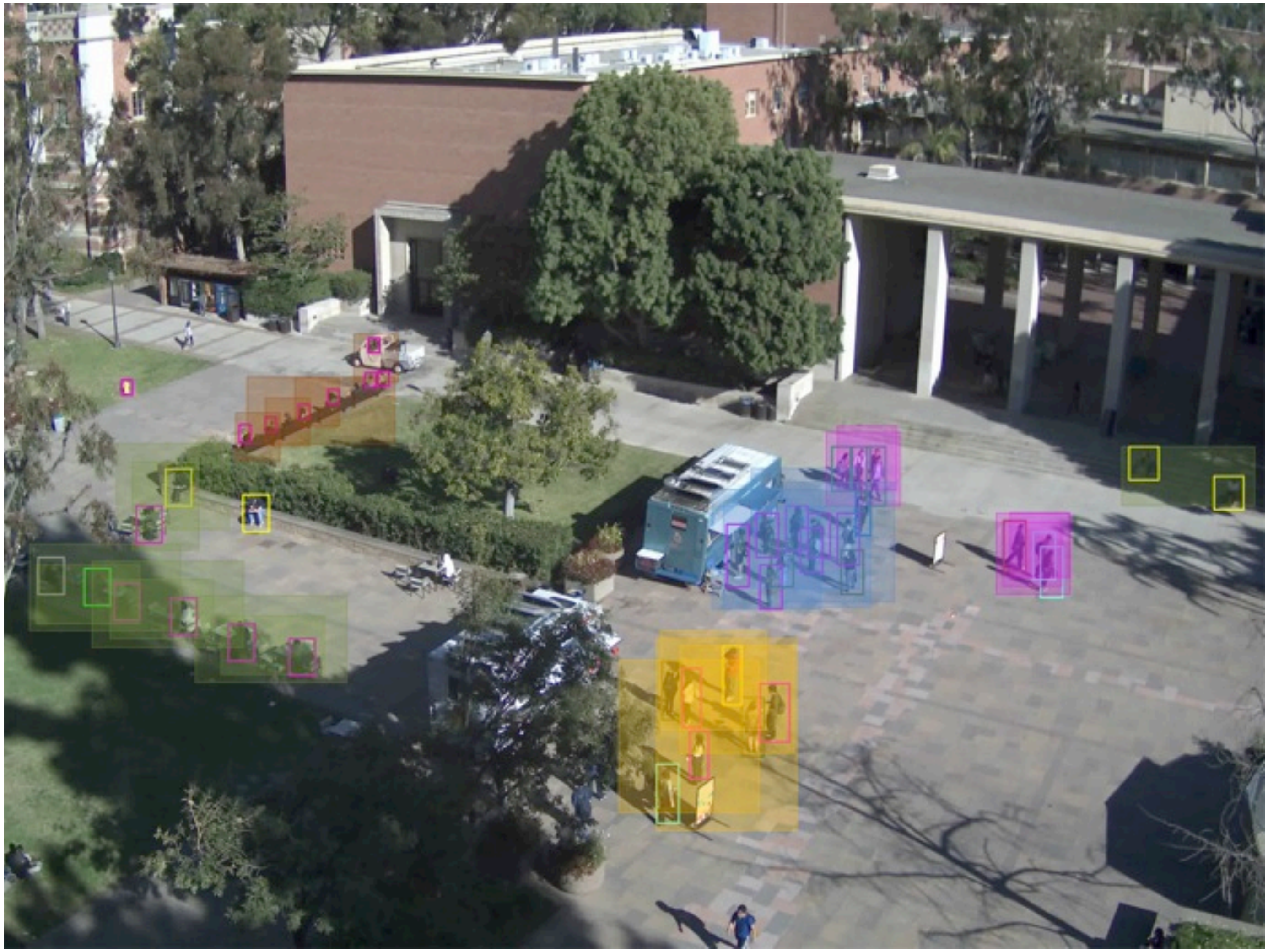
Inference – Alpha of Group Activities

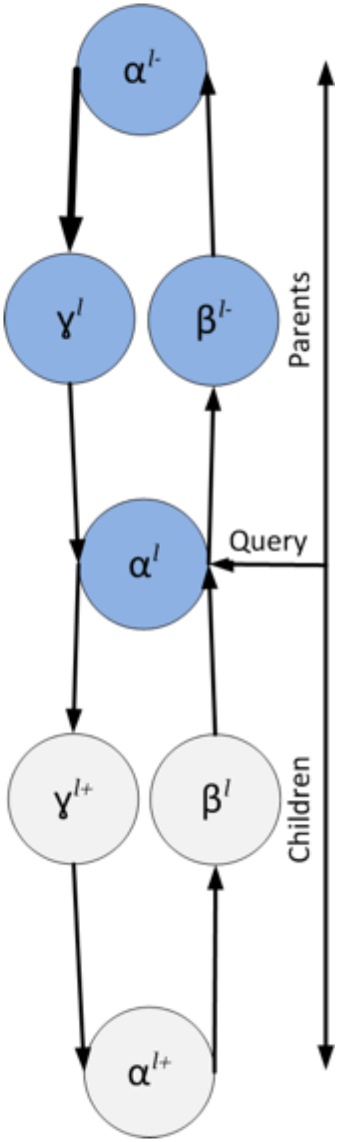


🟪 Group Walking 🟩 Group Discuss 🟨 🟥 🟦



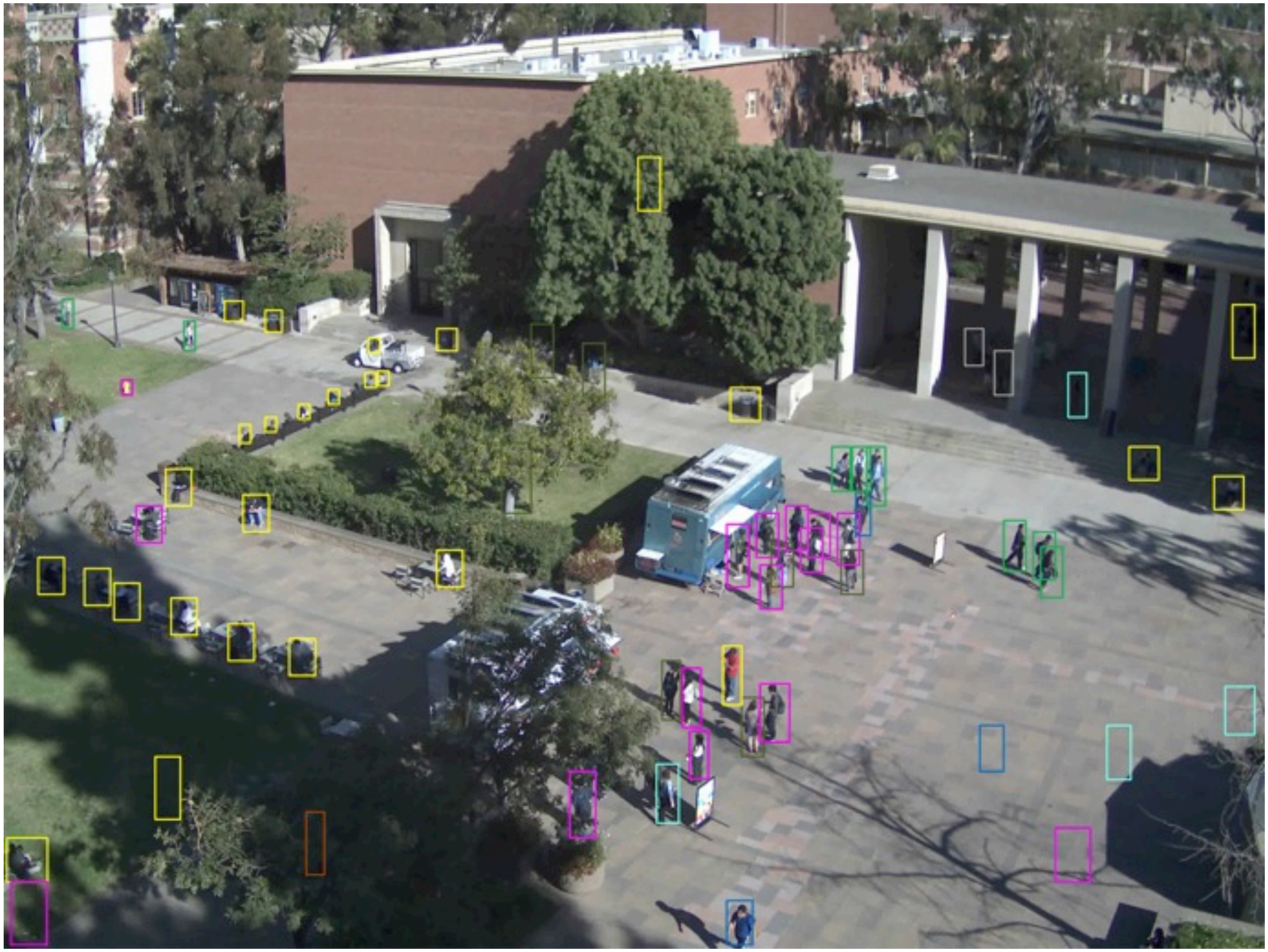
Inference -- Top Down (Gamma)



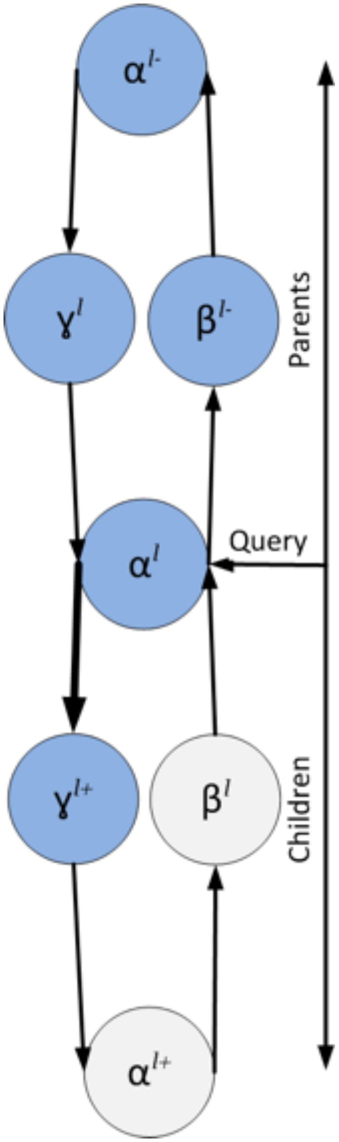


🟪 Group Walking 🟩 Group Discuss 🟨 🟥 🟦

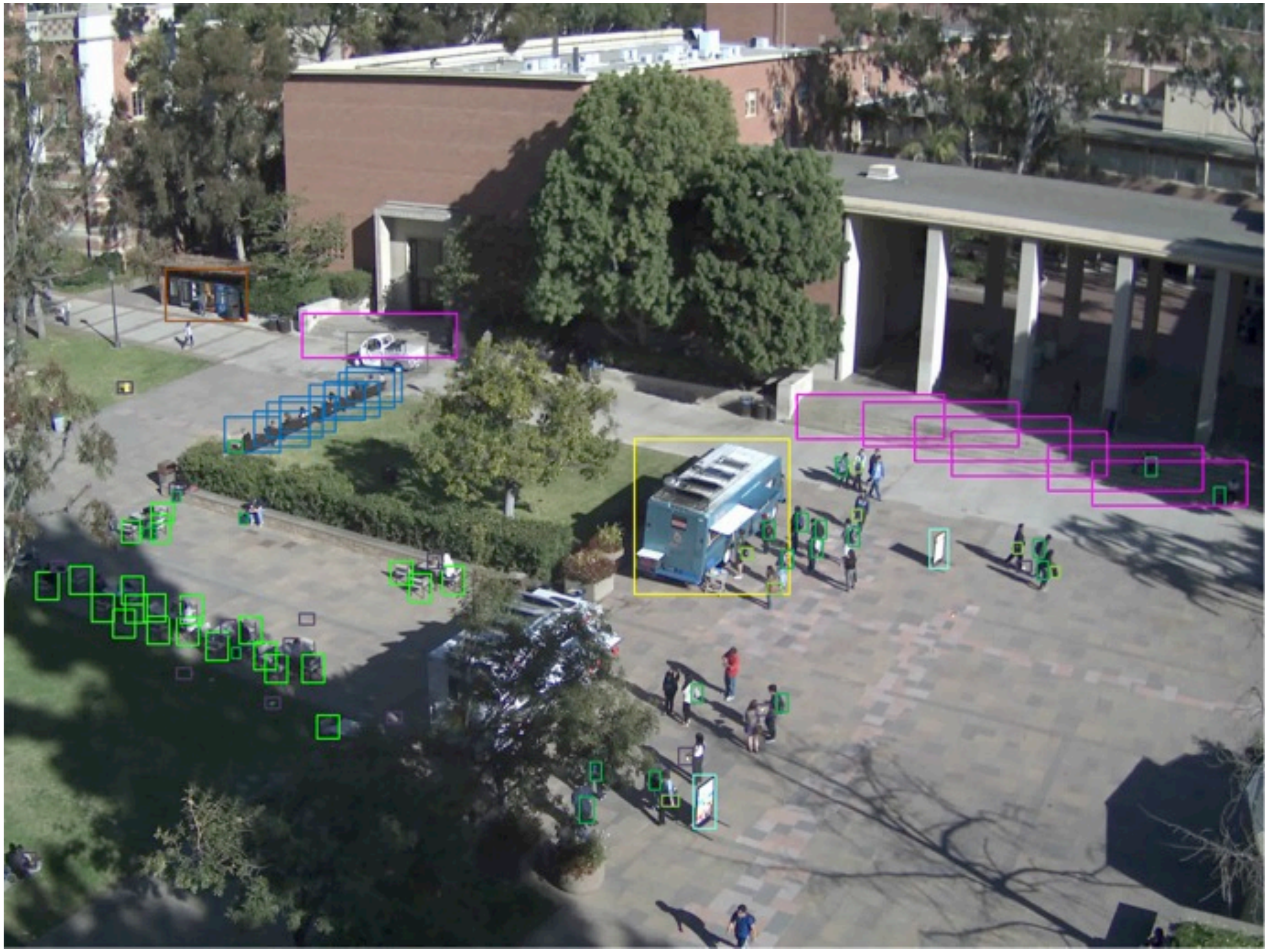
Inference -- Top Down (Gamma)



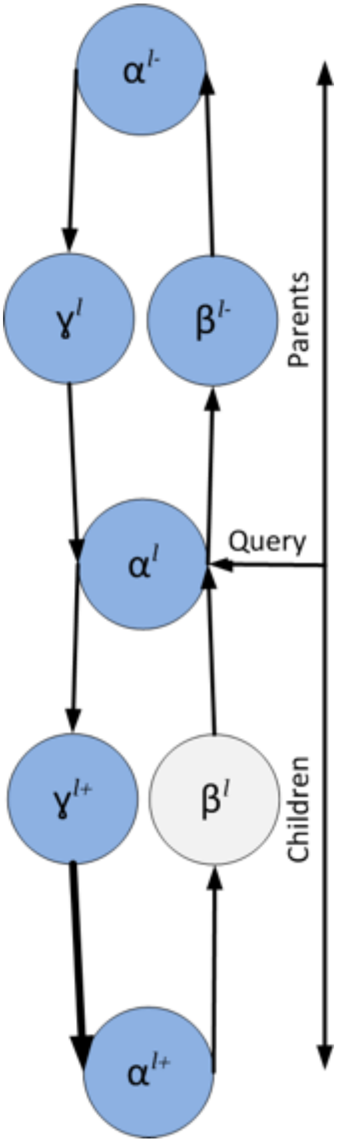
🟪 Waiting 🟩 Walking 🟨 🟥 ⬜



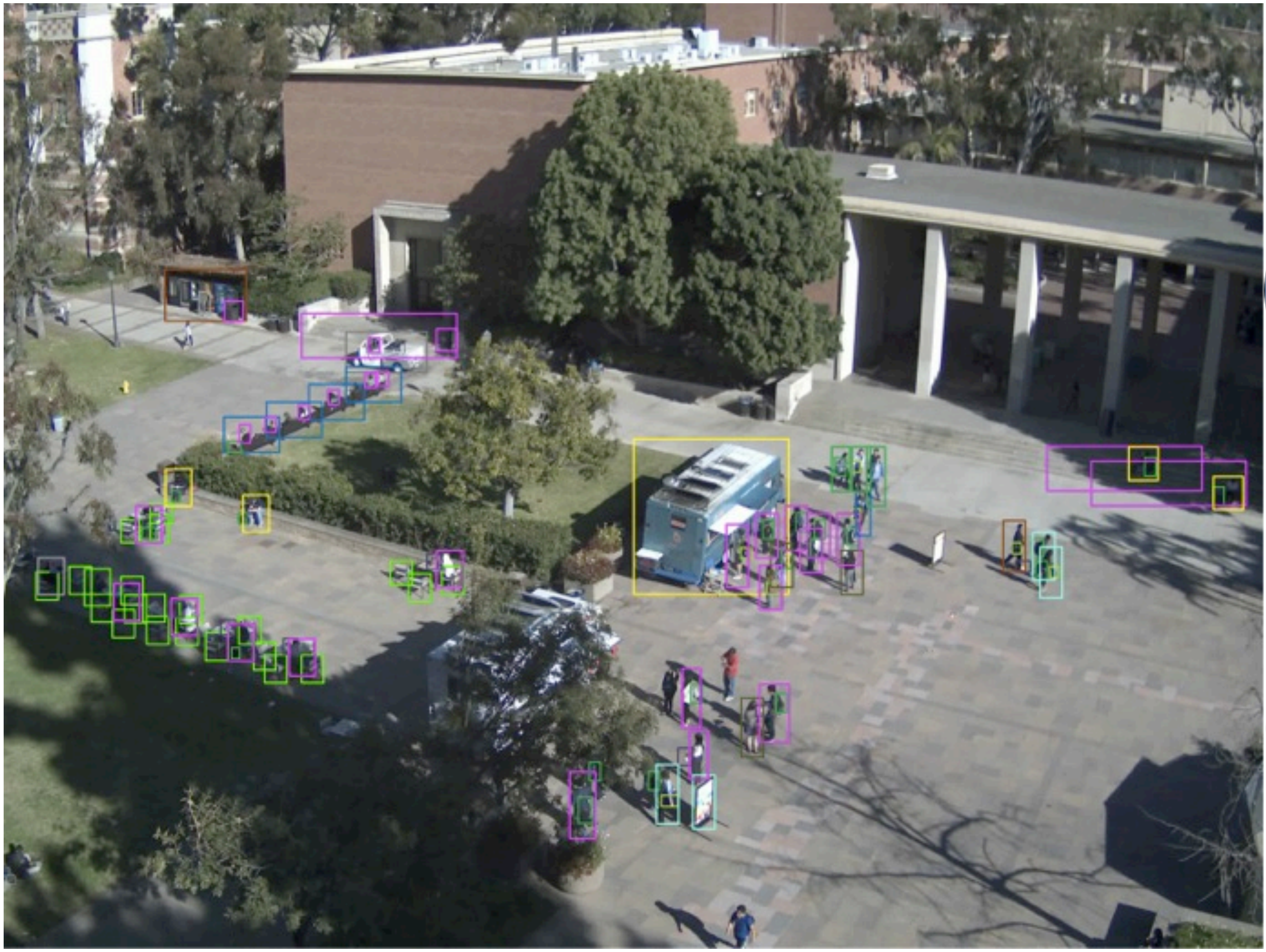
Inference -- Alpha of Objects



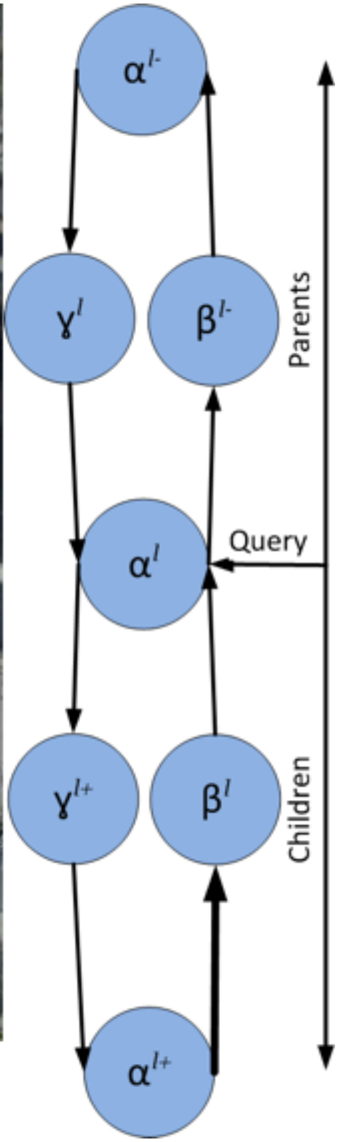
🟪 Stairs 🟩 Chairs 🟨 Food bus 🟥 ⬛



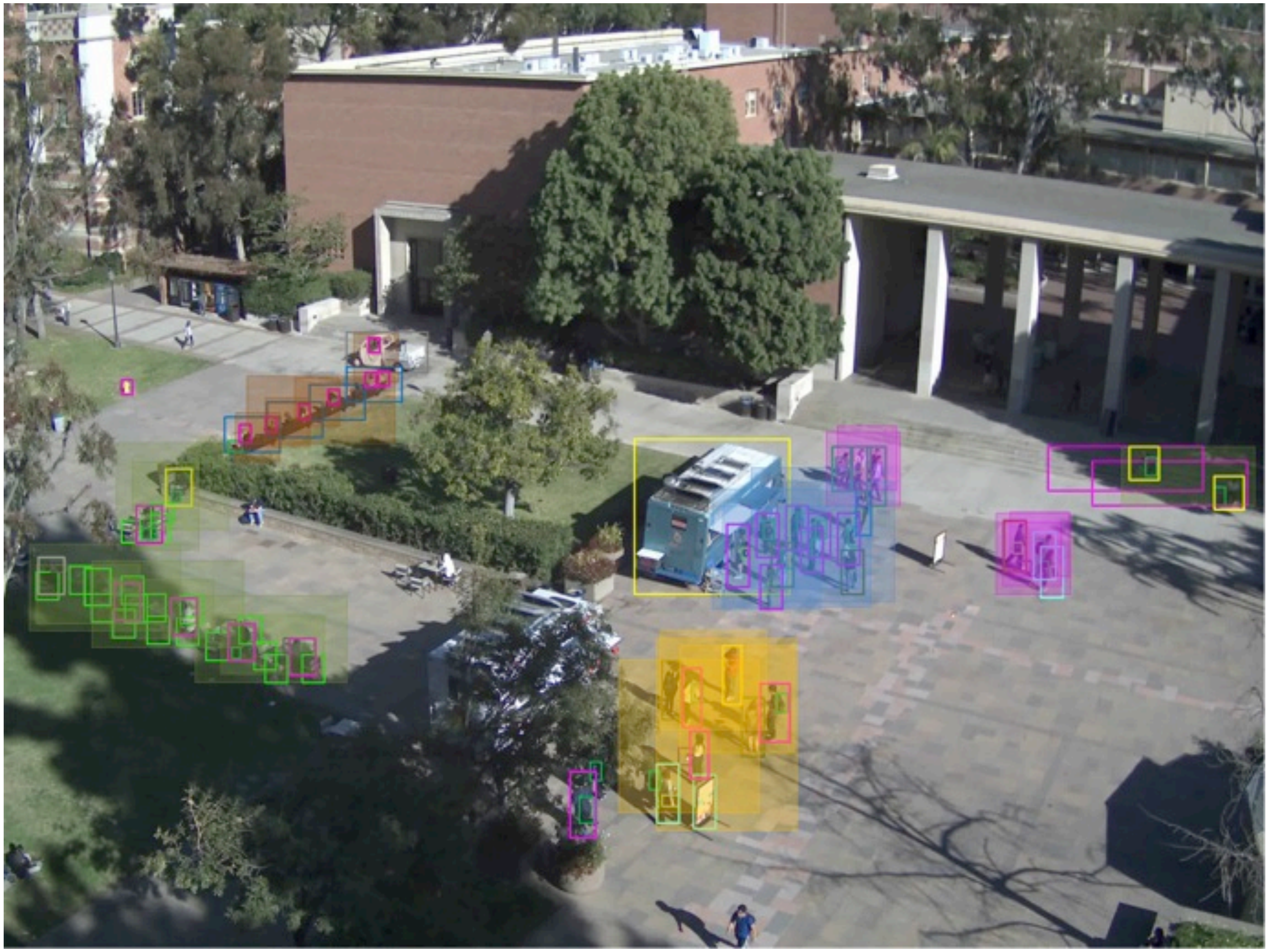
Inference -- Bottom Up (Beta)



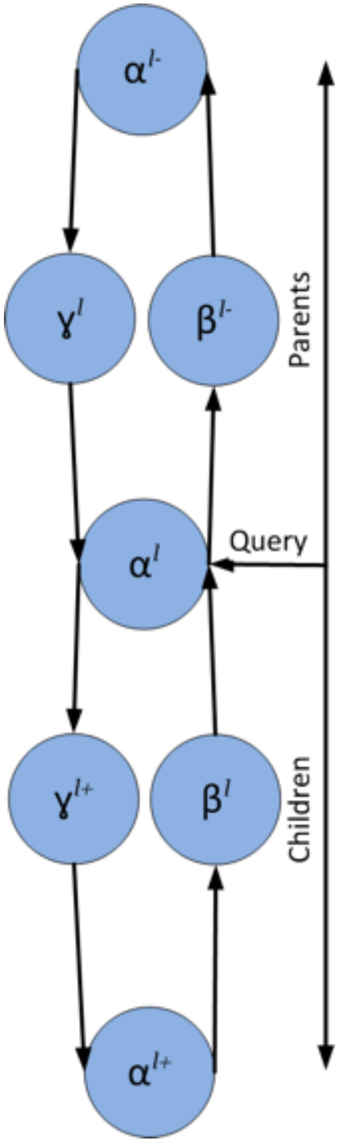
🟪 Waiting 🟩 Walking 🟨 🟥 ▨



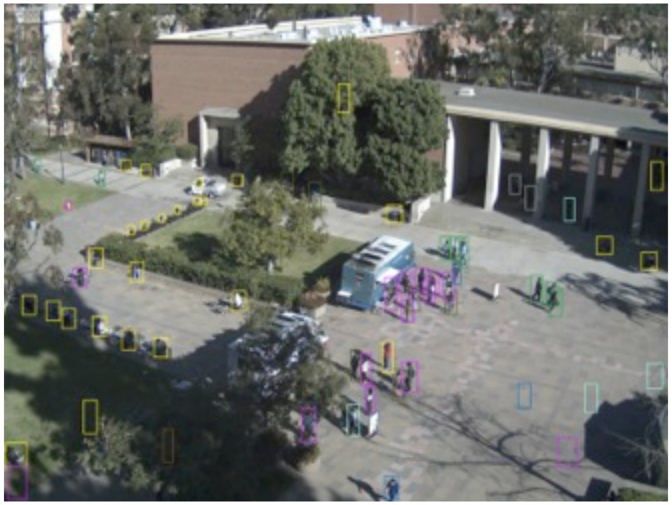
Inference



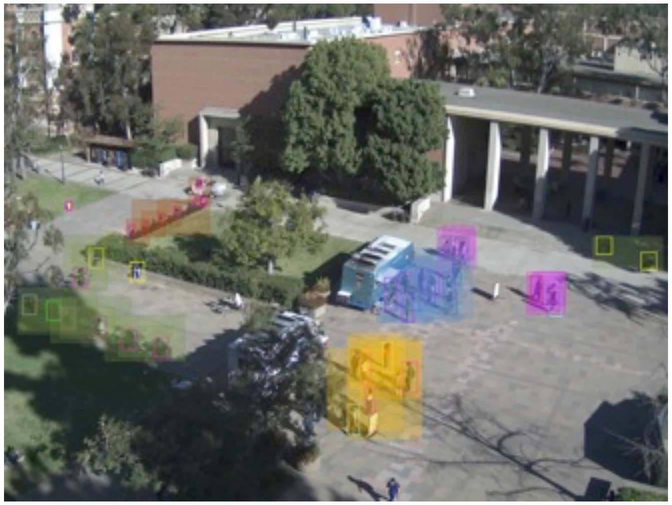
🟪 Waiting 🟩 Walking 🟨 🟥 ◼



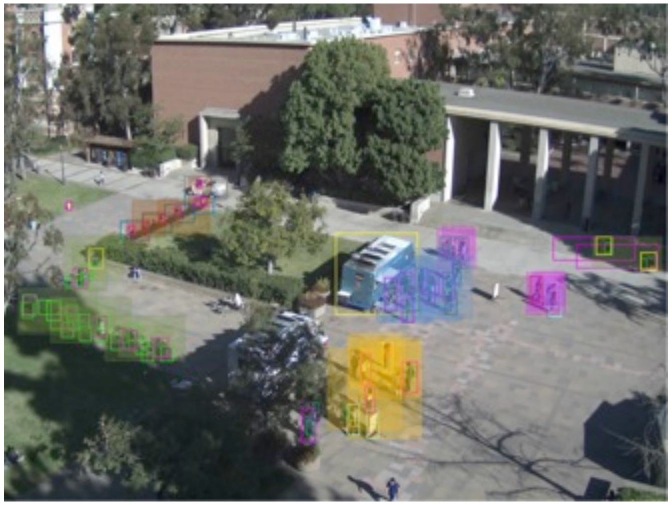
Queries about Individual Actions



Only Alpha



Zoom-out



Zoom-in and Zoom-out

Scheduling

$$\mathbb{V}_{\pi_{\mathcal{Q}}}^{(\tau)}(s) = \mathbb{E}[G(s,a)] + \sum_{s'} \mathbb{T}(s, \pi_{\mathcal{Q}}(s,\tau), s') \cdot \mathbb{V}_{\pi_{\mathcal{Q}}}^{(\tau-1)}(s'), \quad \tau = 1, ..., \mathbb{B}$$

- States (s) =
  - Query,
  - All previous actions,
  - Their rewards until this point
- Actions (a) =
  - Run detector and terminate (Alpha)
  - Bottom-up (Beta)
  - Top-down (Gamma)

Policies, Given a Query (Q)

1. Alpha(Q)

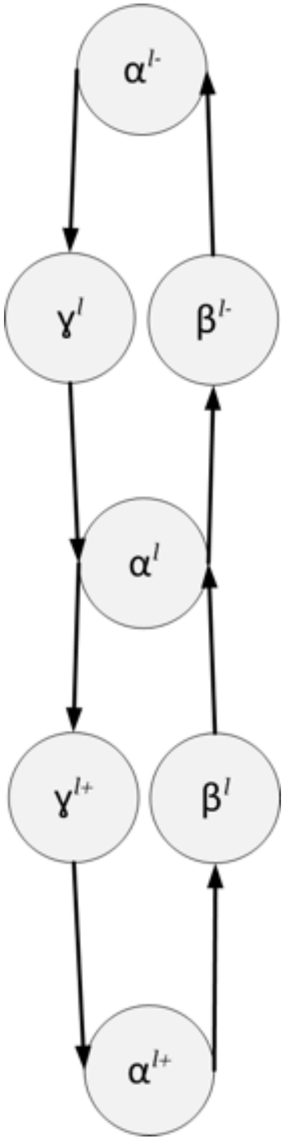
2. Alpha(Q), Bottom-up, Alpha(Context), Top-down

3. Alpha(Q), Top-down, Alpha(Details), Bottom-up

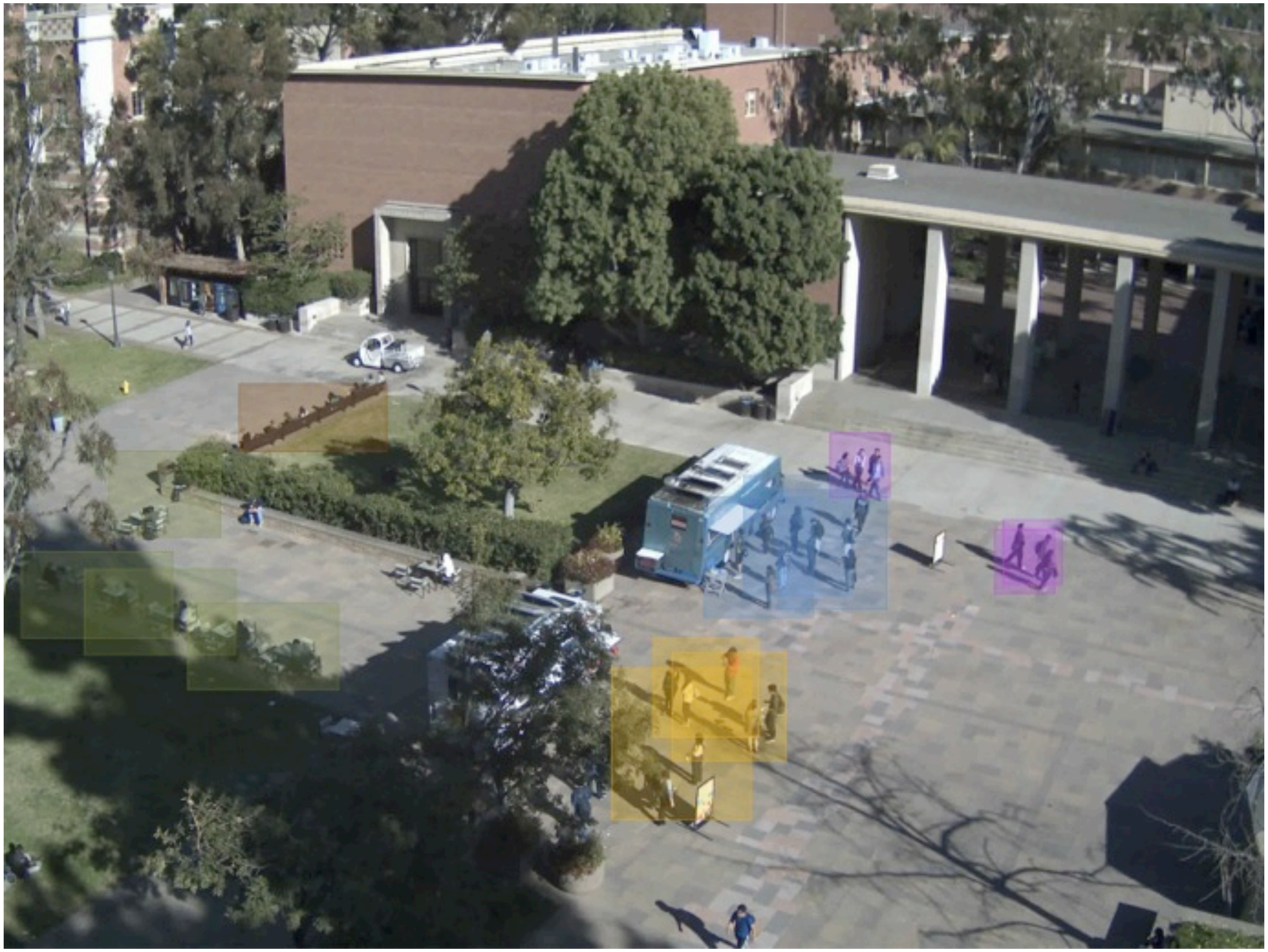
4. Alpha(Q), Bottom-up, Alpha(Context), Top-down, Alpha(Details), Bottom-up

5. Alpha(Q), Top-down, Alpha(Details), Bottom-up Alpha(Context), Top-down

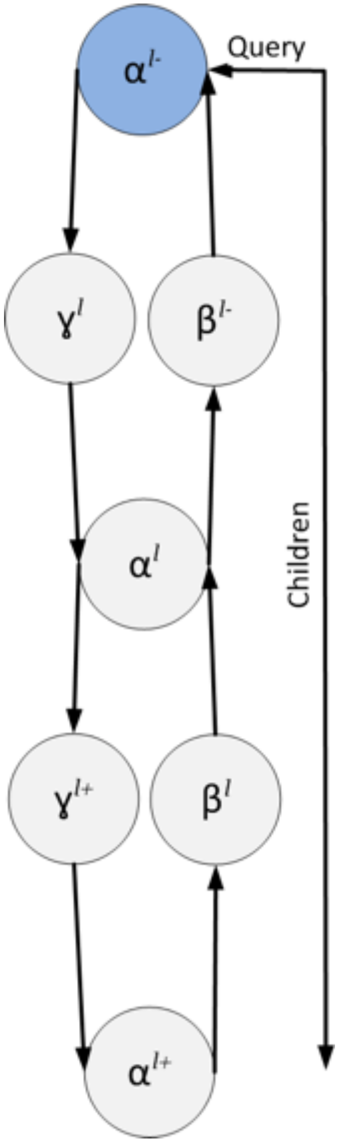
Group Query



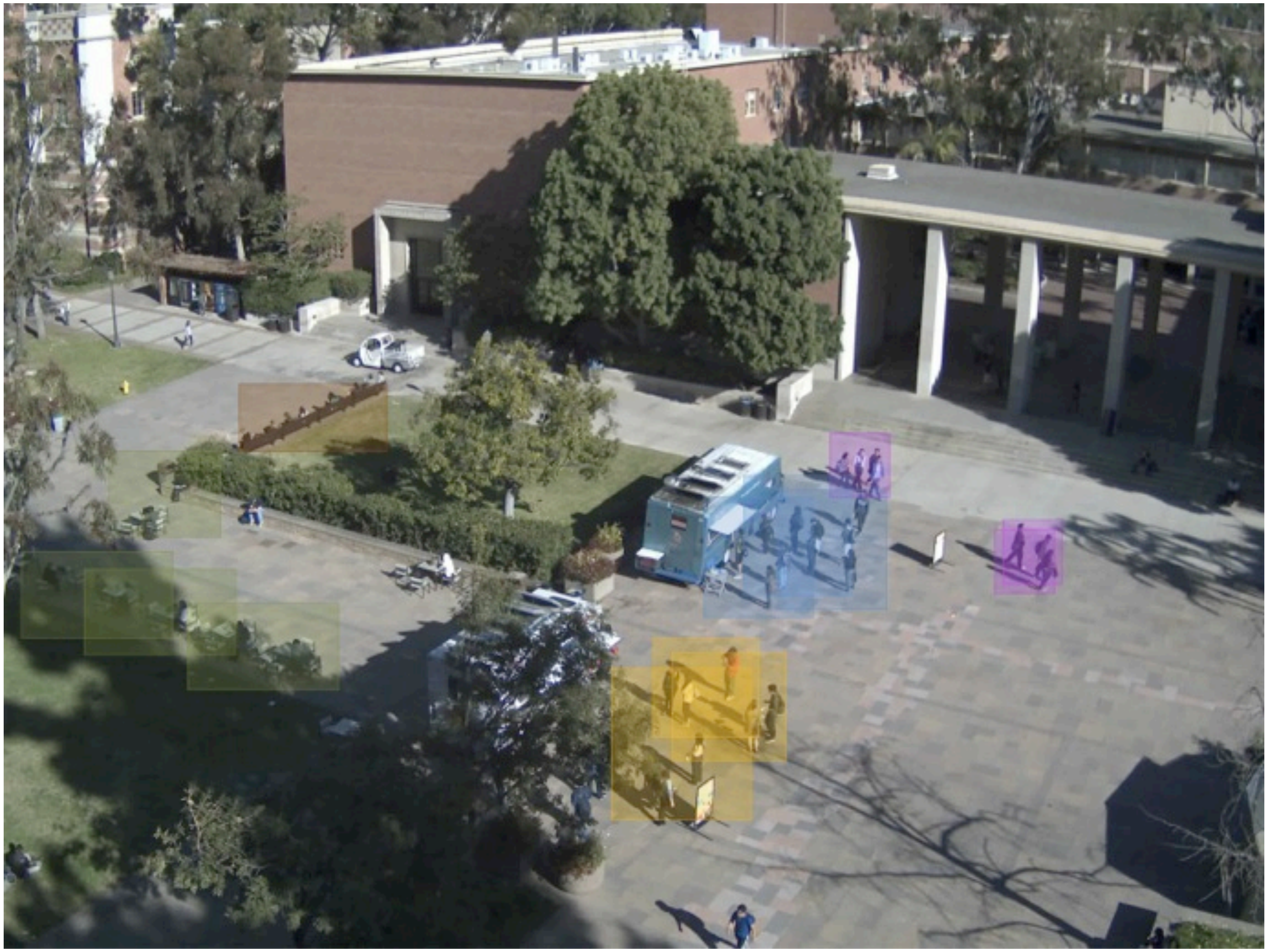
Group Query Policy #1



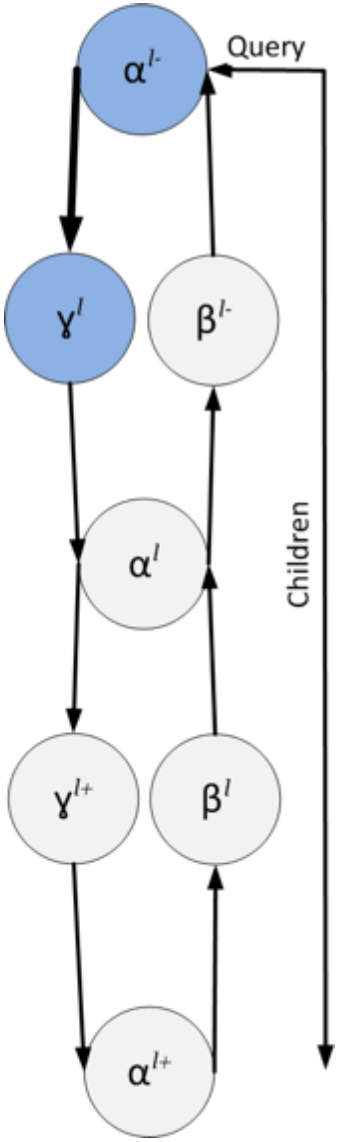
🟪 Group Walking 🟩 Group Discuss 🟨 🟥 🟦



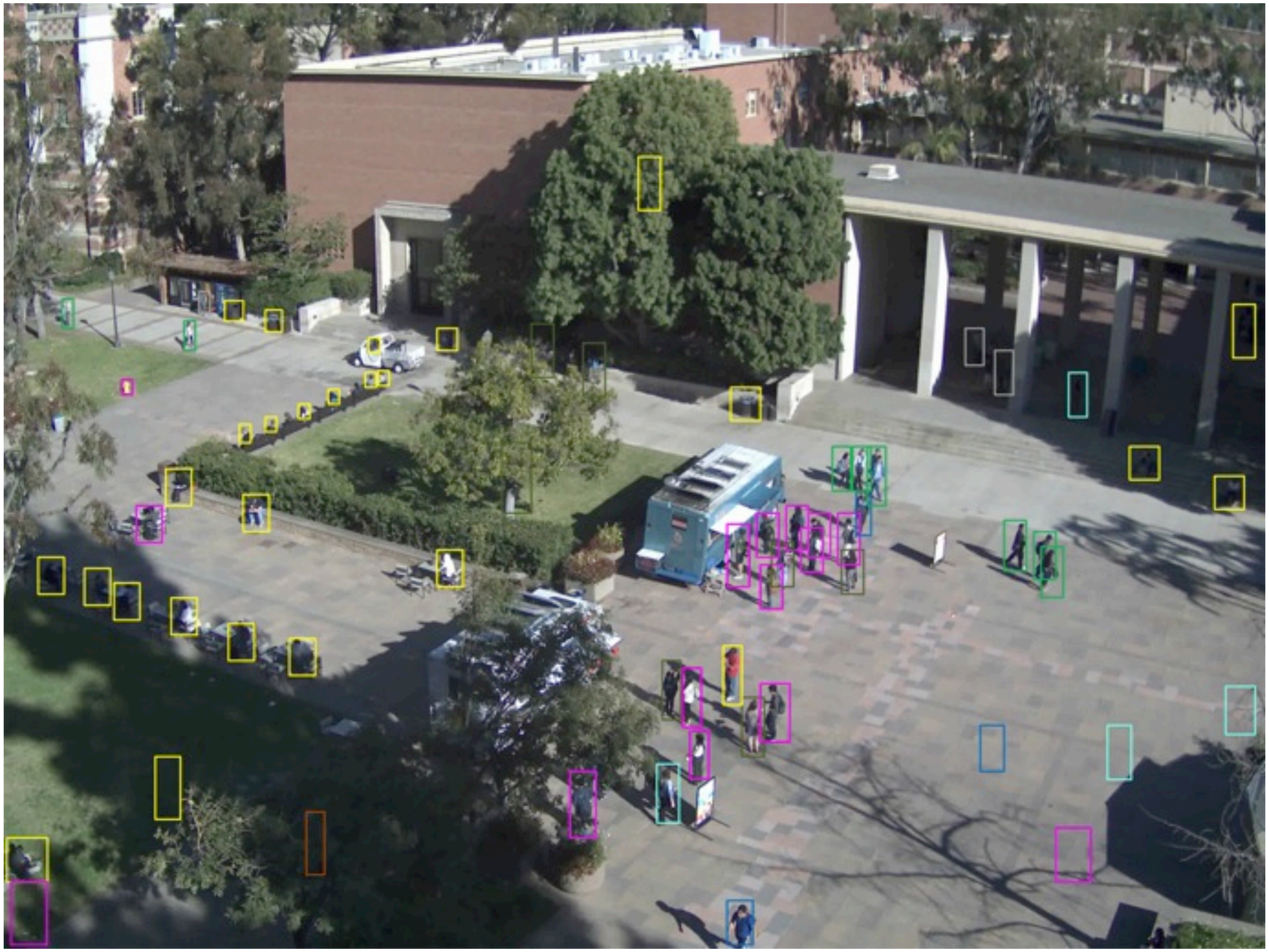
Group Query Policy #1



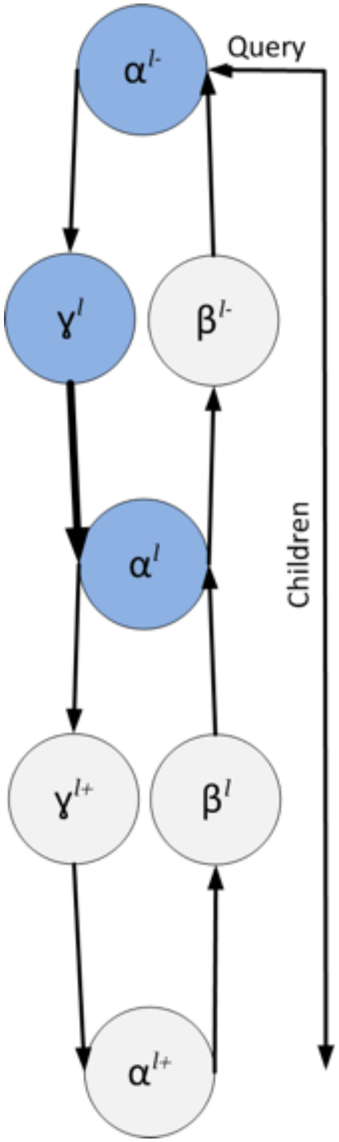
🟪 Group Walking 🟩 Group Discuss 🟨 🟥 🟦



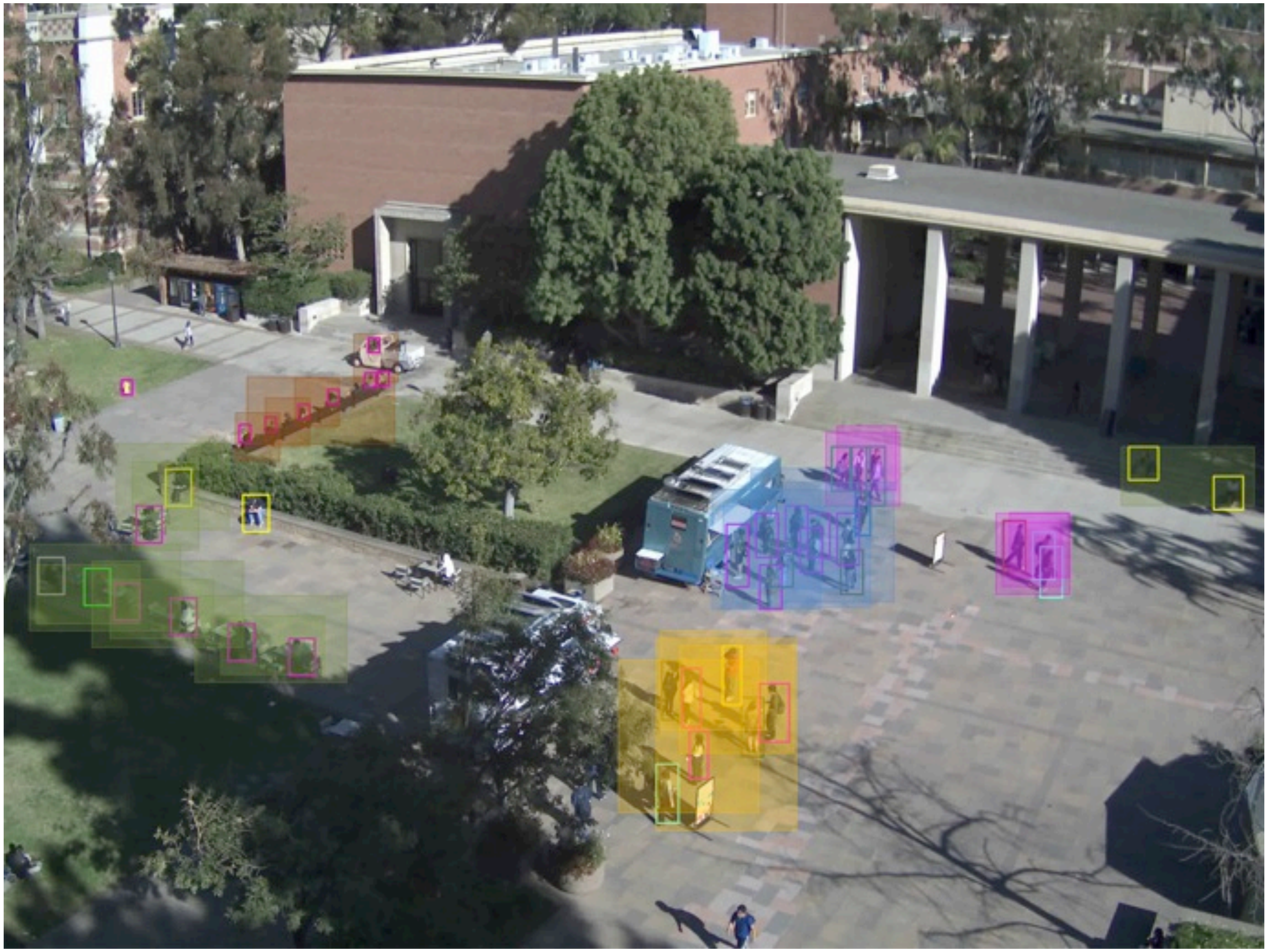
Group Query Policy #1



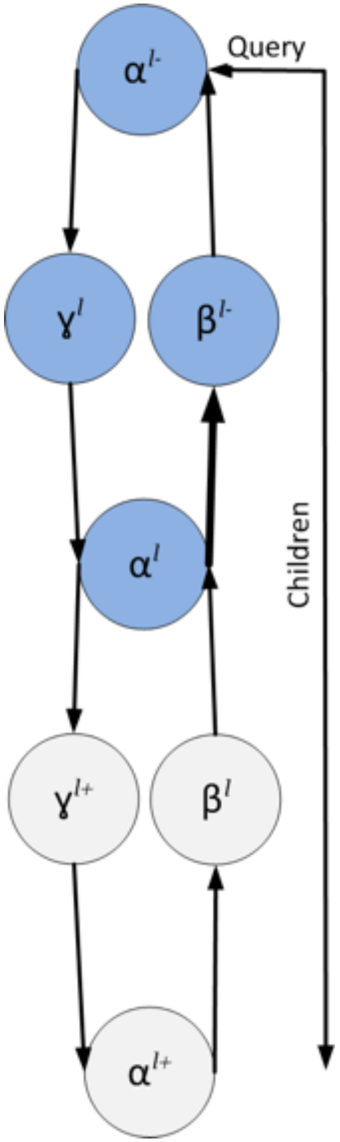
🟪 Waiting 🟩 Walking 🟨 🟥 ■



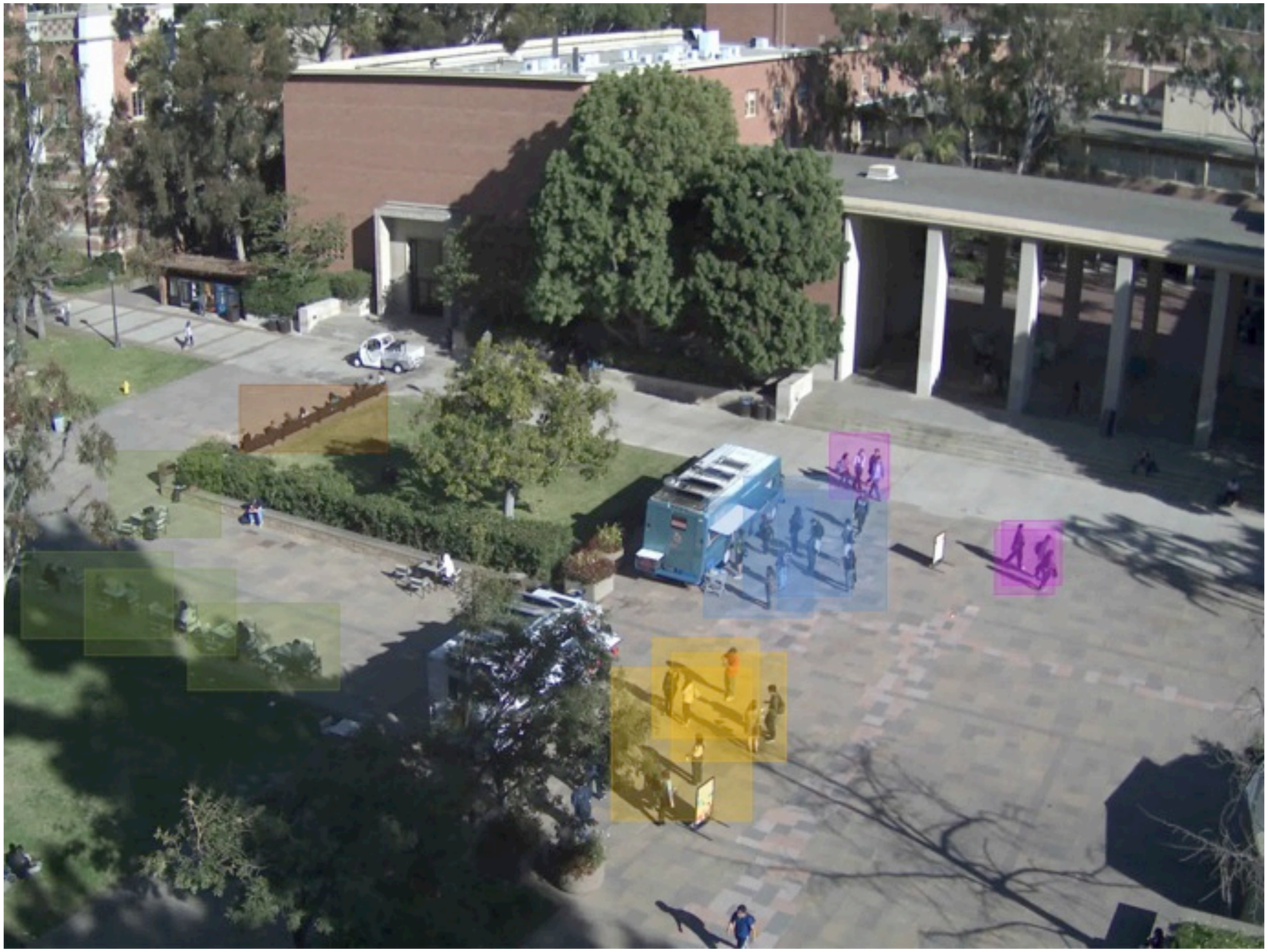
Group Query Policy #1

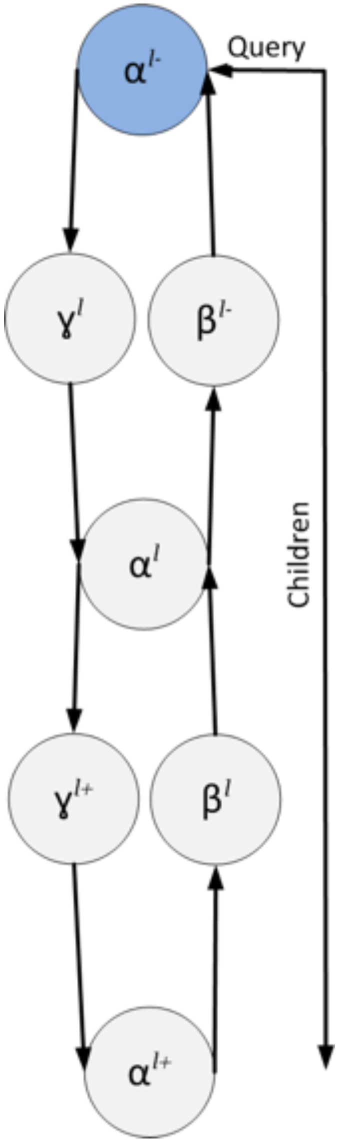


🟪 Group Walking 🟩 Group Discuss 🟨 🟥 🟦



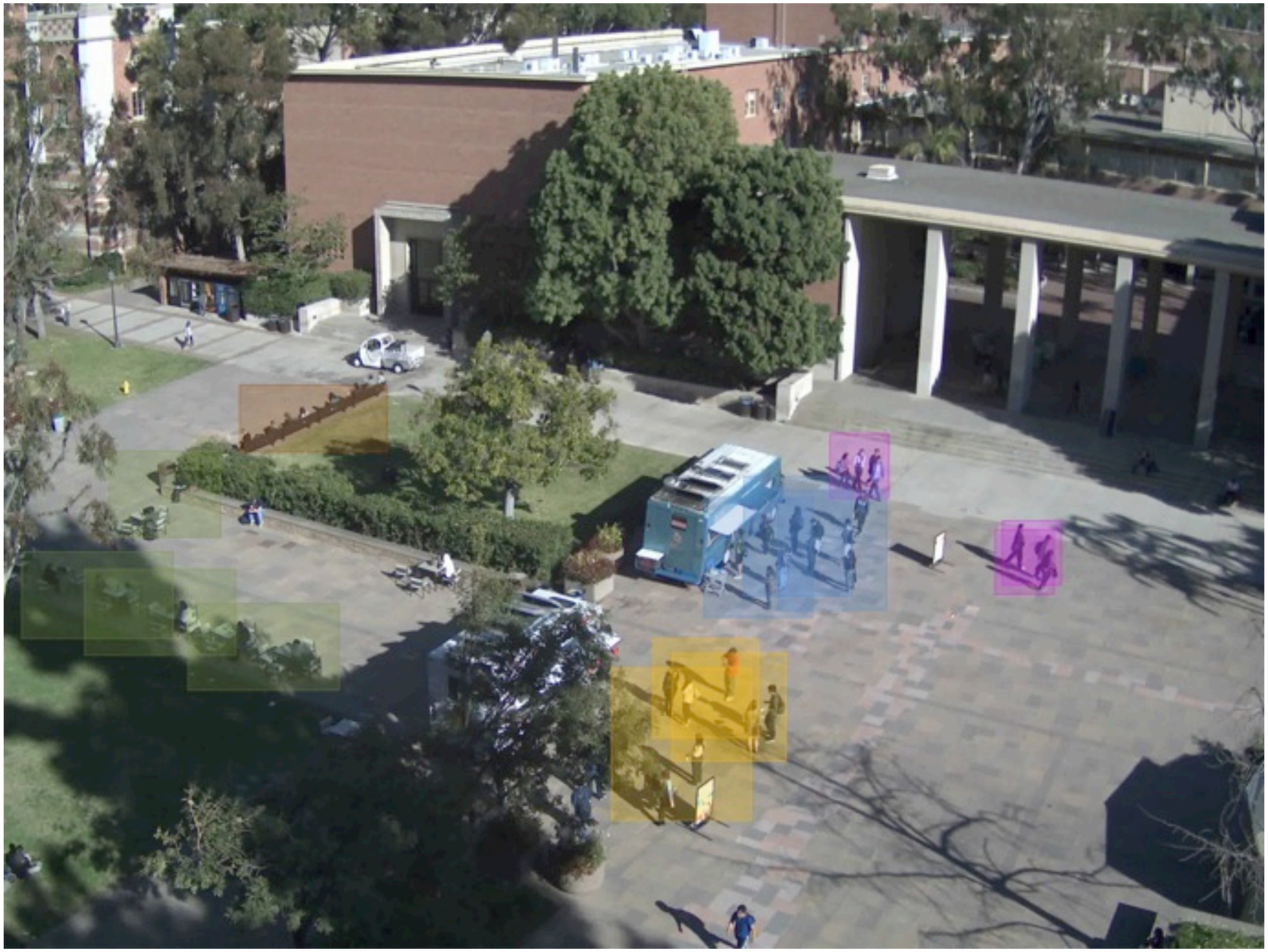
Group Query Policy #2



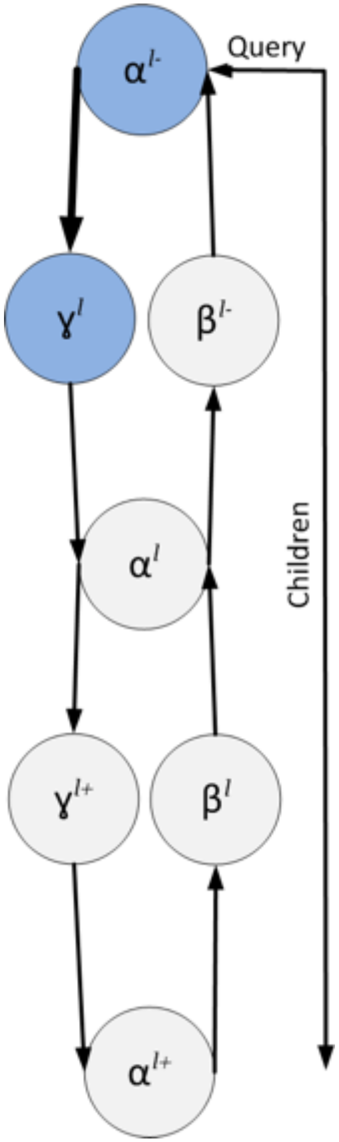


🟪 Group Walking 🟩 Group Discuss 🟨 🟥 🟦

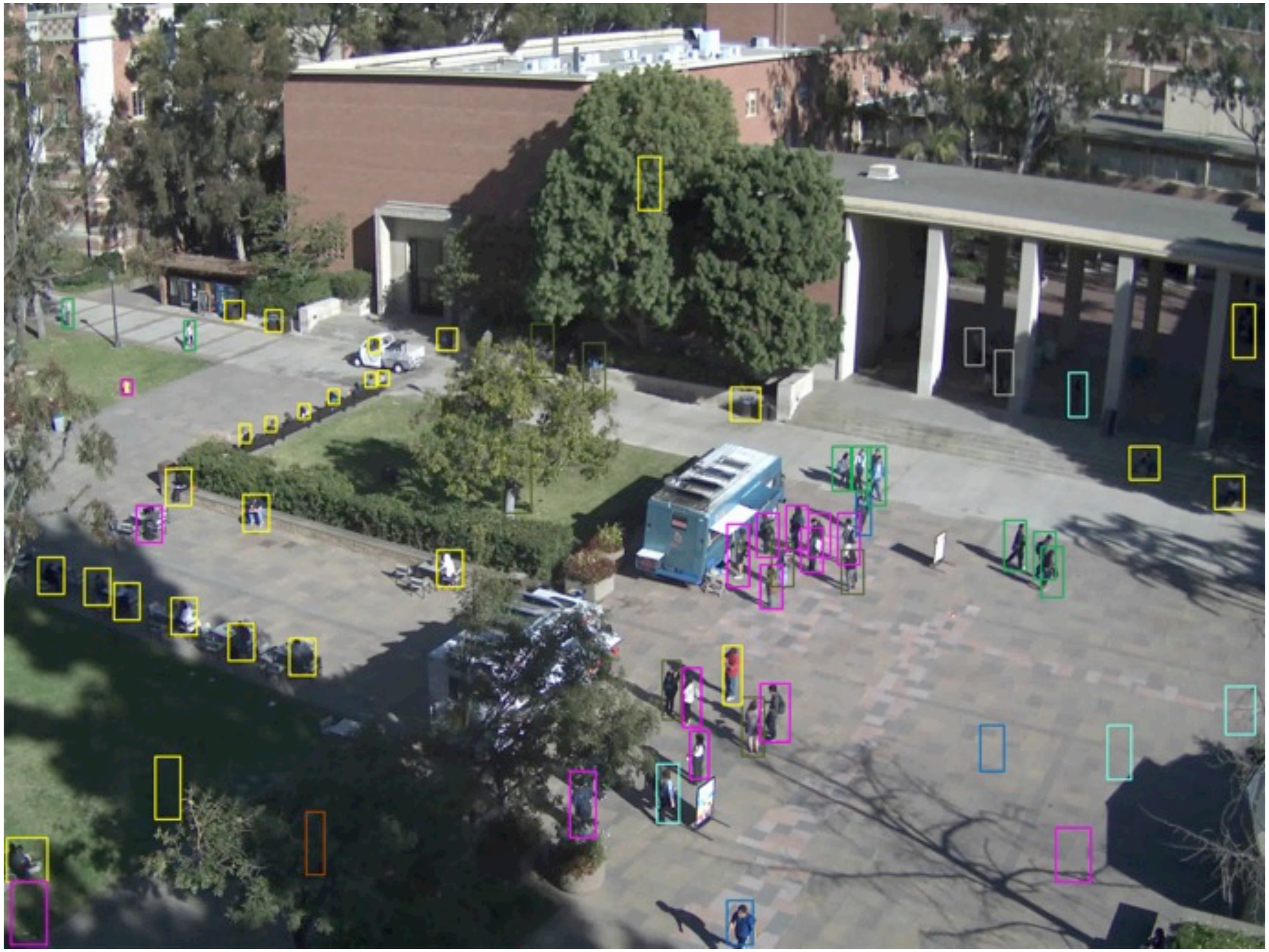
Group Query Policy #2

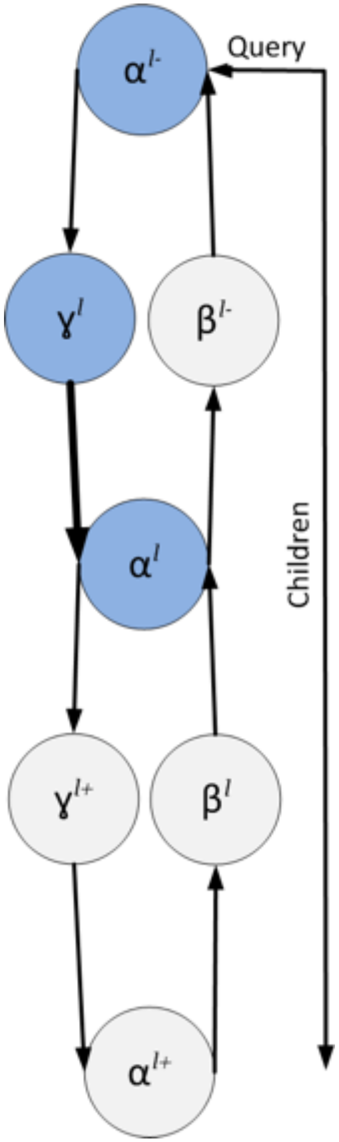


🟪 Group Walking 🟩 Group Discuss 🟨 🟥 🟦

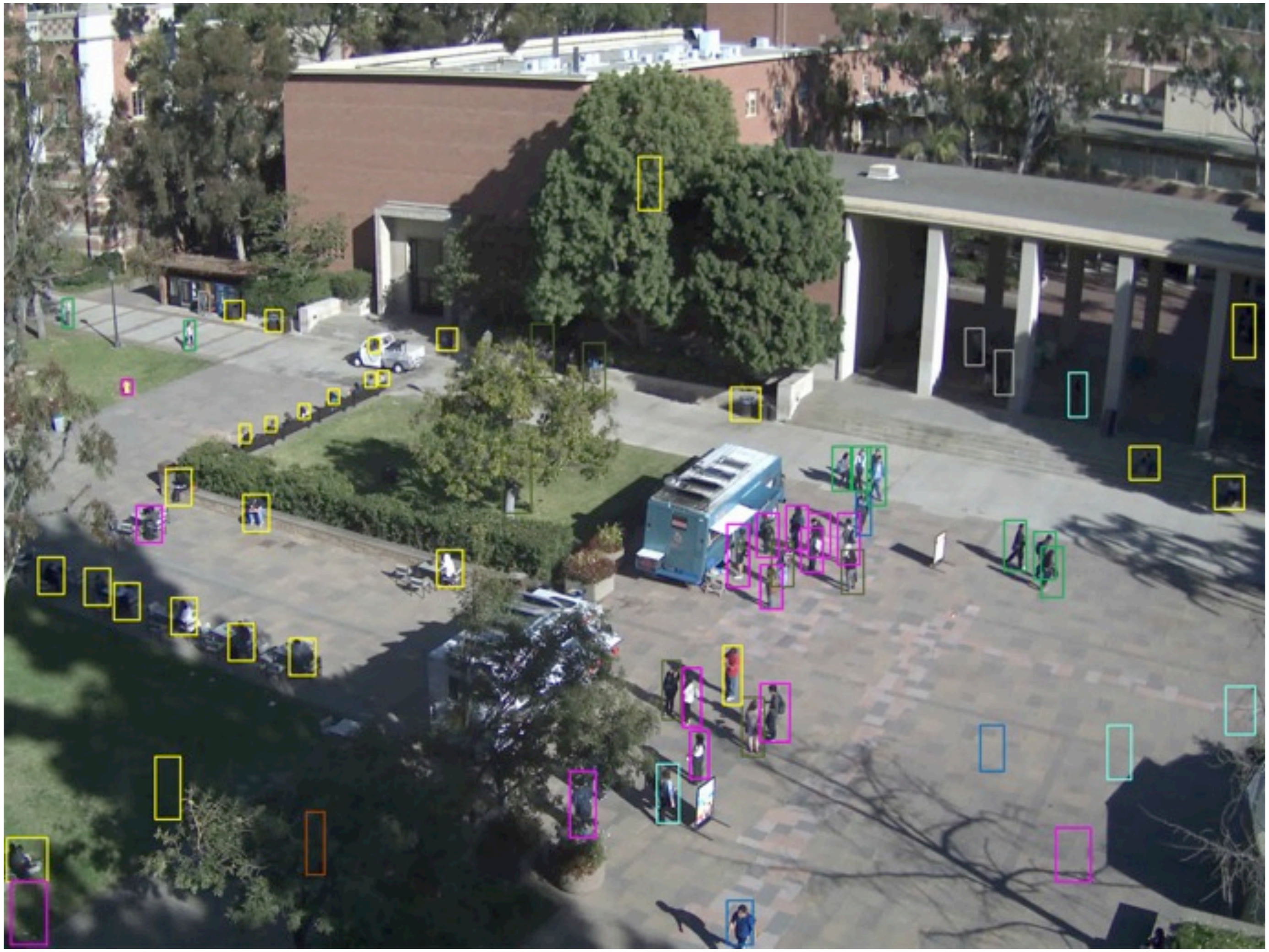


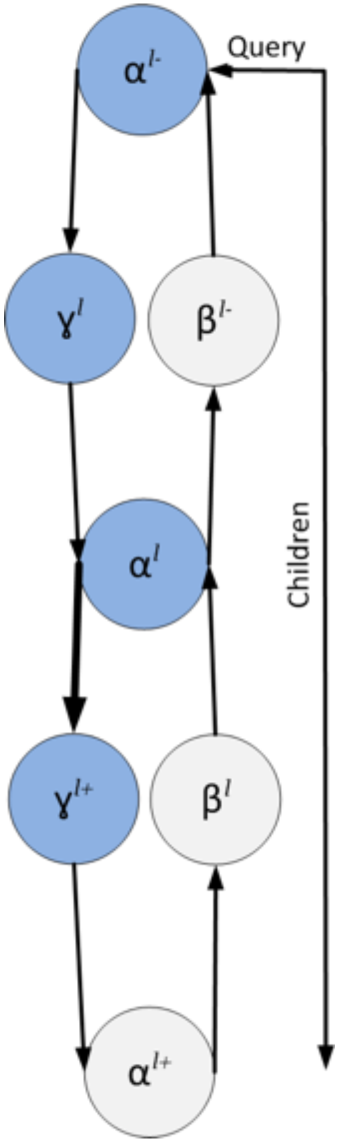
Group Query Policy #2



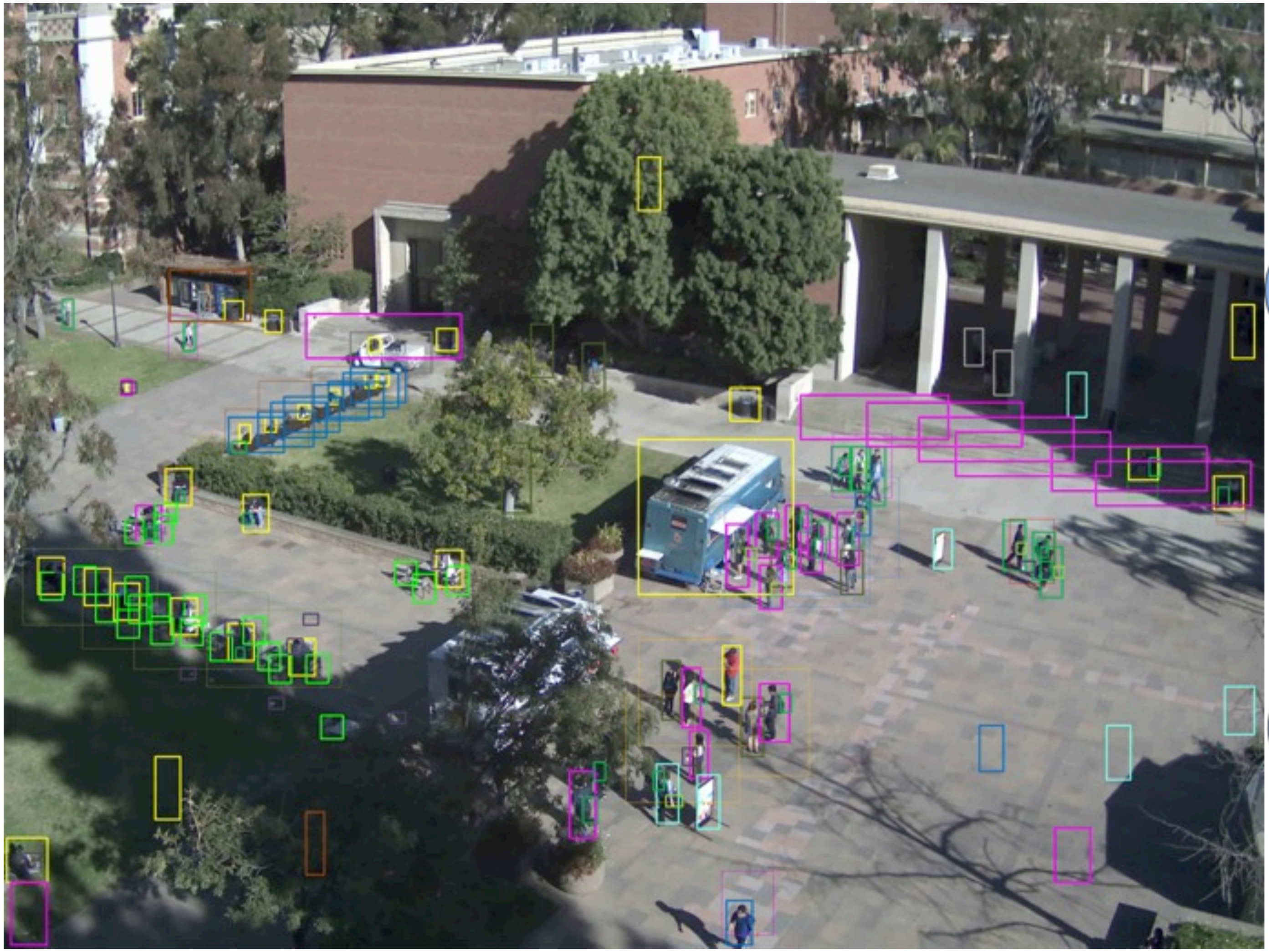


Group Query Policy #2

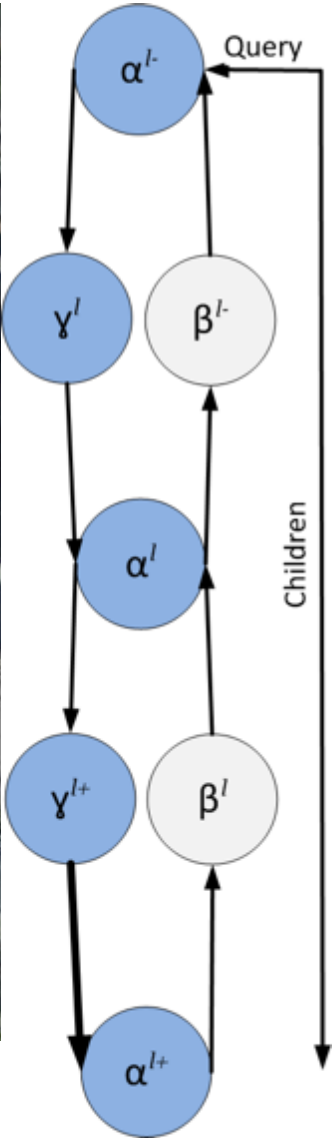




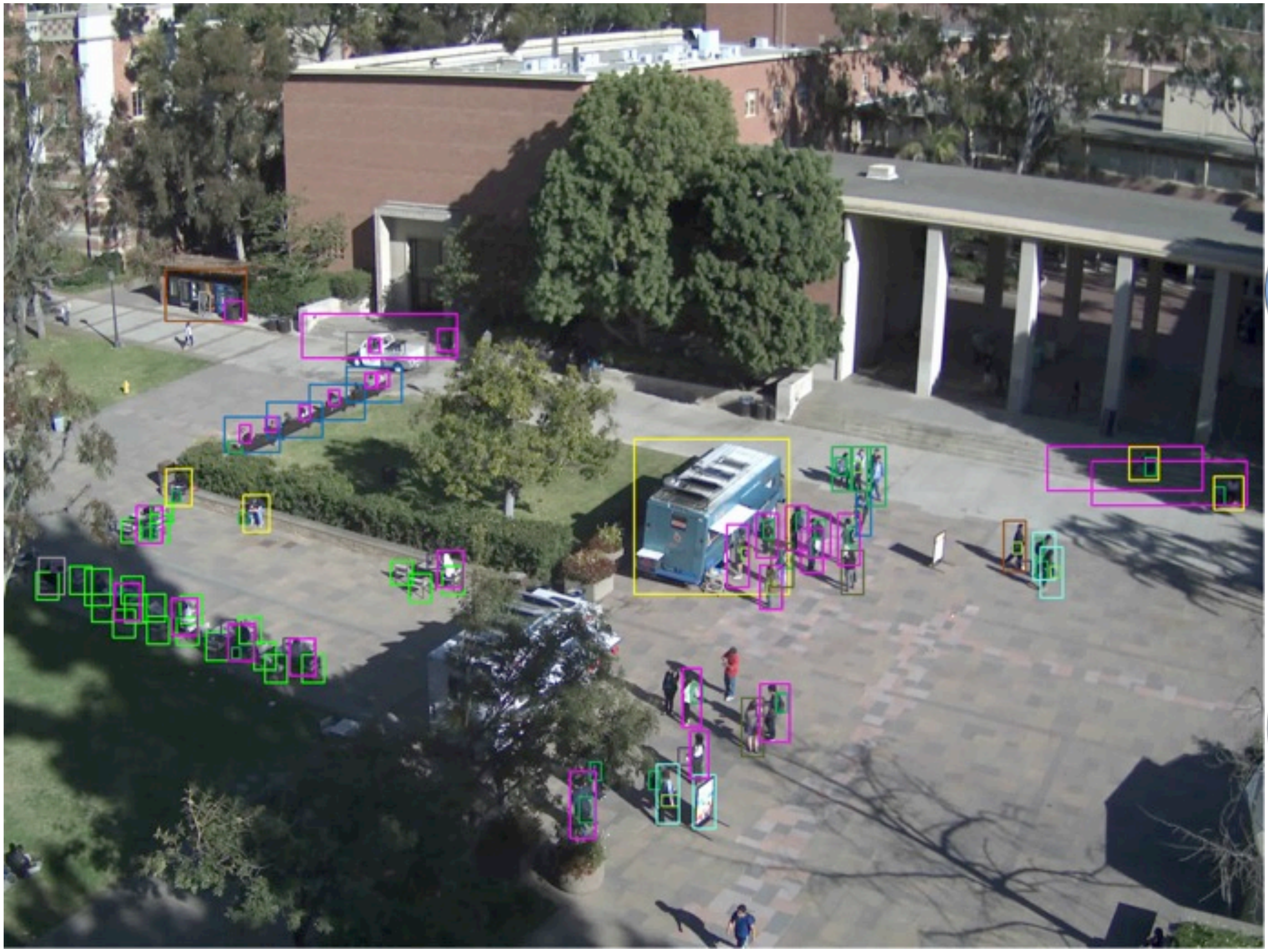
Group Query Policy #2



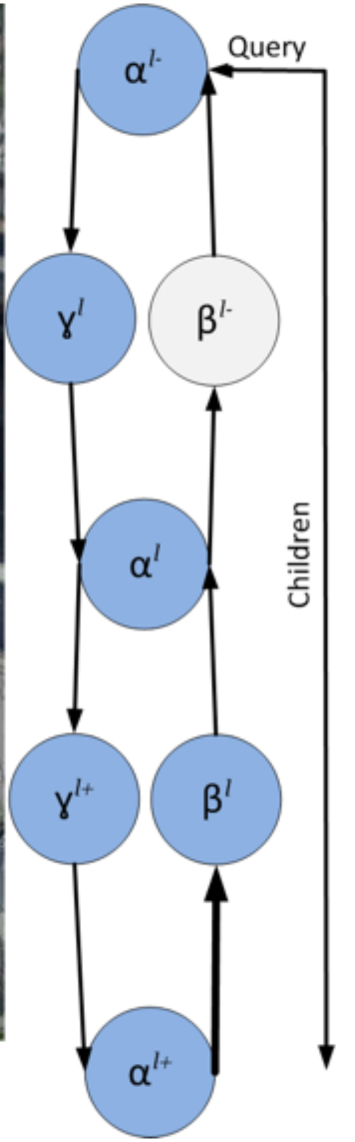
🟪 Stairs 🟩 Chairs 🟨 Food bus 🟥 ◼



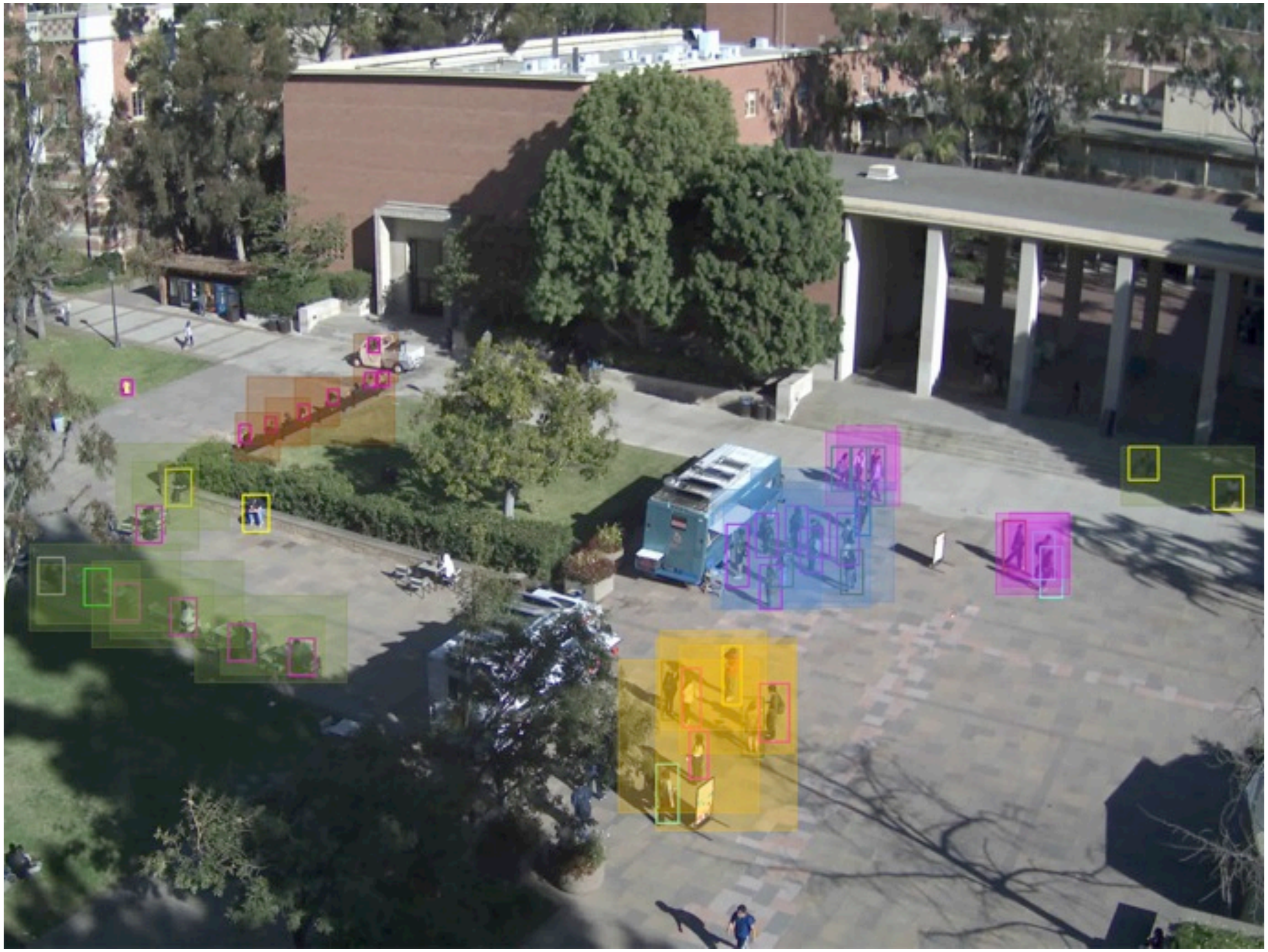
Group Query Policy #2



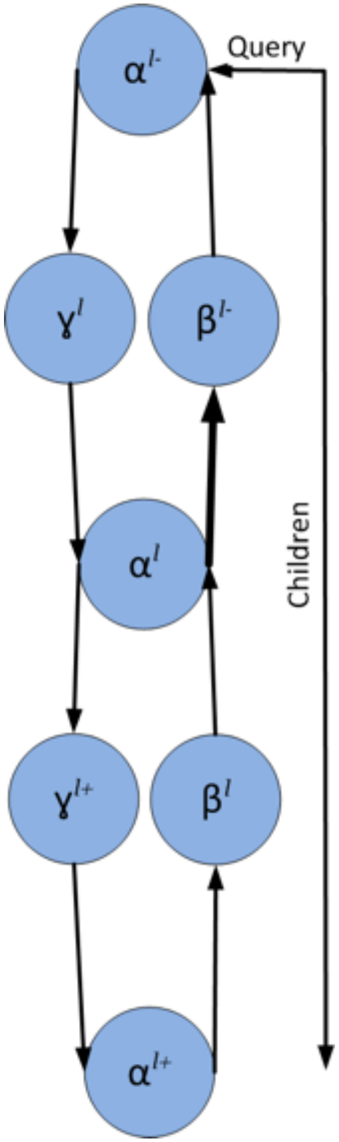
🟪 Waiting 🟩 Walking 🟨 🟥 ⬛



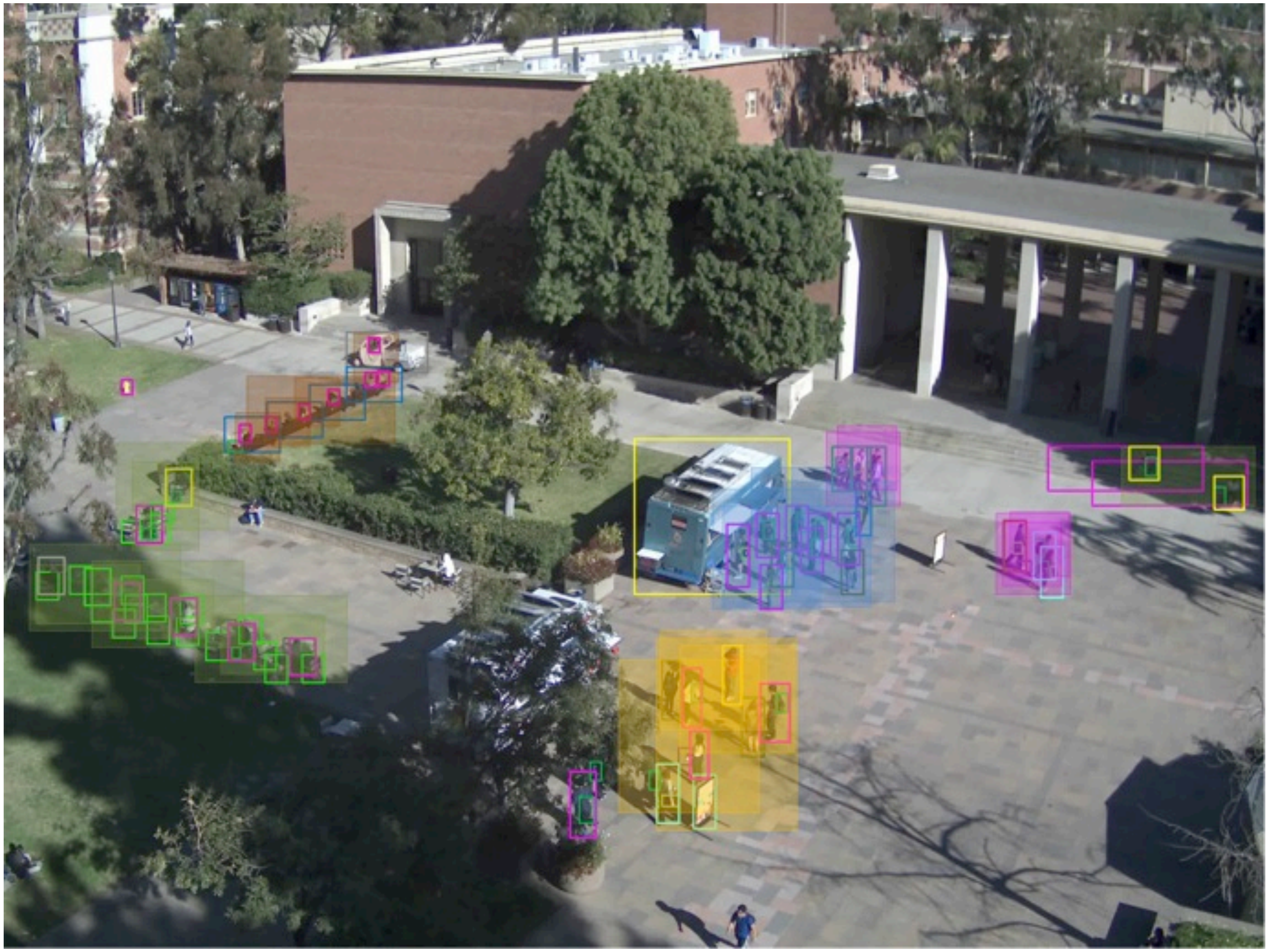
Group Query Policy #2



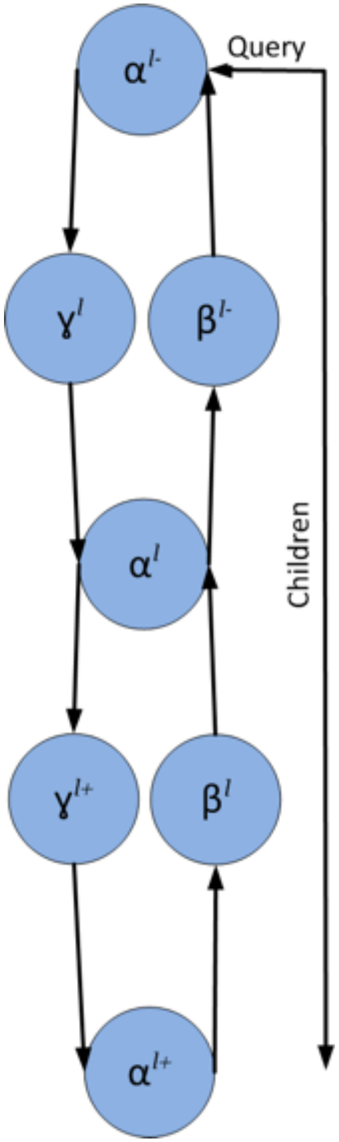
🟪 Group Walking 🟩 Group Discuss 🟨 🟥 🟦



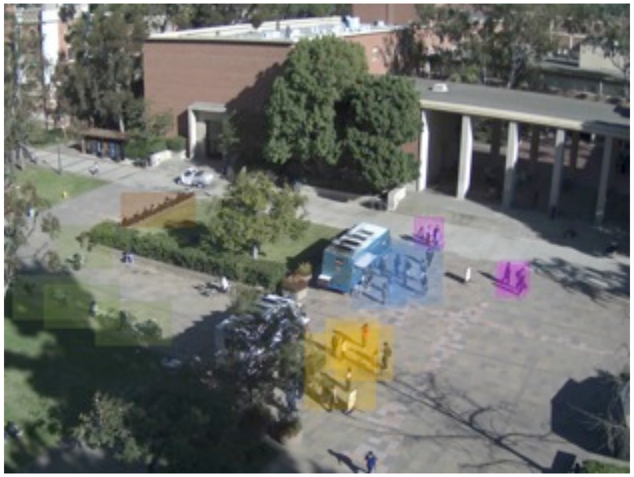
Group Query Policy #2



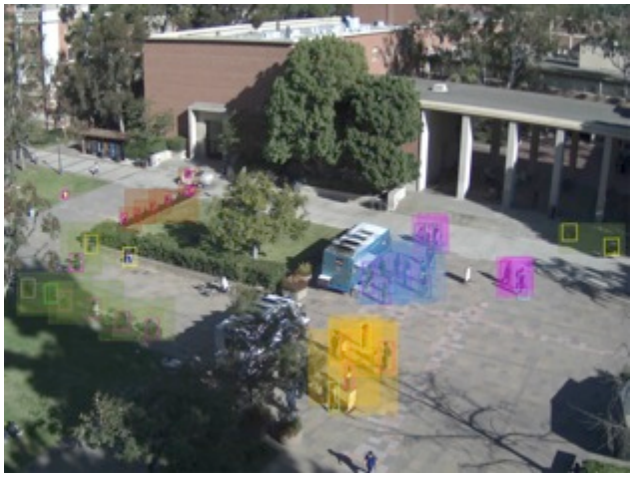
🟪 Group Walking 🟩 Group Discuss 🟨 🟥 🟦



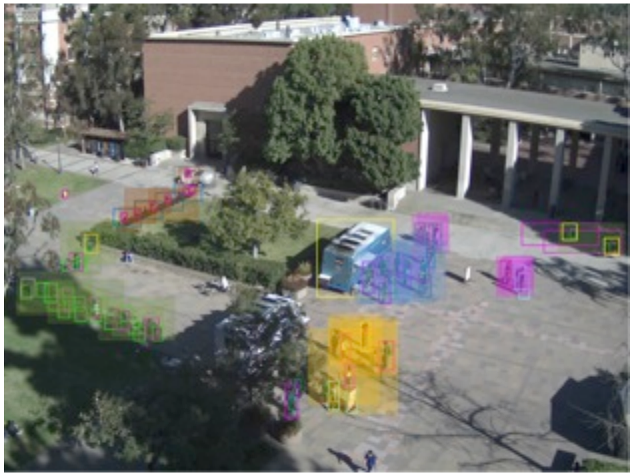
Group Query



Alpha

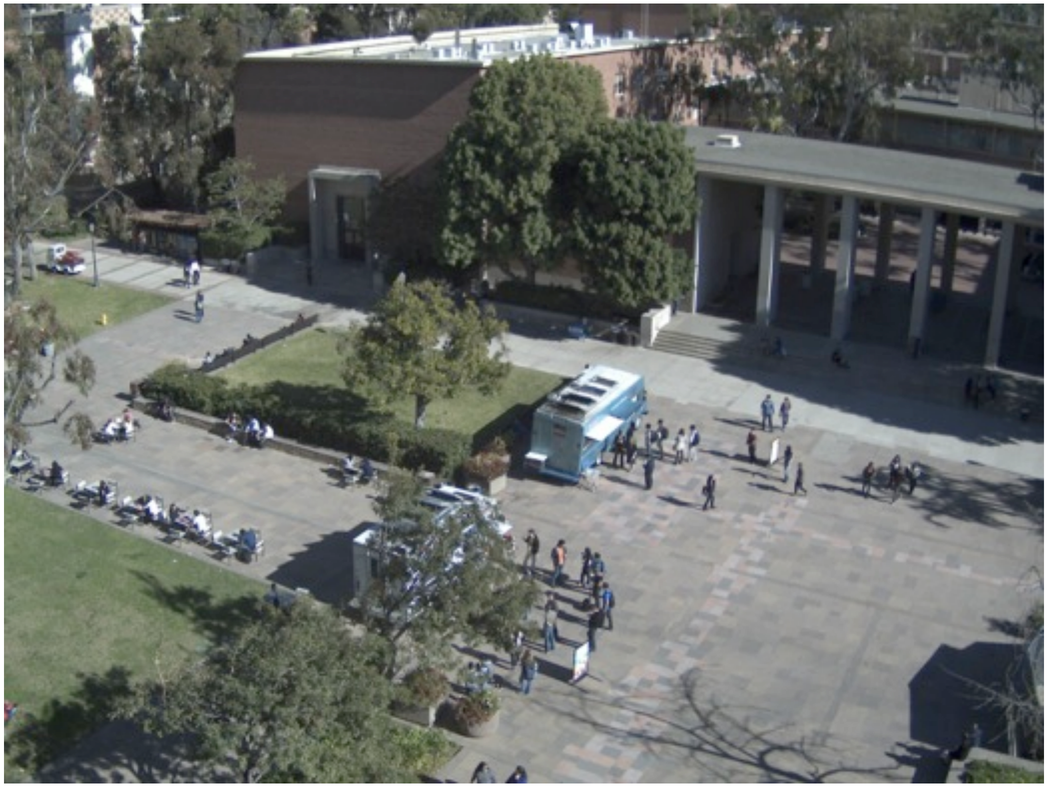


Policy #1



Policy #2

New Dataset



- Footage: 106min
- Frames: 12 fps
- Res.: 2560x1920 pixels
- Portable acquisition system

Benchmark datasets:

- single actor
- single group
- single action

Domain Knowledge

- 5 Group Activities:
  - Walking together, Queuing, Campus tour, ...
- 10 Individual Actions:
  - Walking, Sitting, Riding a bike, …
- 16 Objects:
  - Food truck, Vending machine, Bike, Backpack, ...