Where’s Waldo: Matching People in Images of Crowds

Rahul Garg, Deva Ramanan, Steve Seitz ,Noah Snavely

(Presented by Deepti Ghadiyaram)

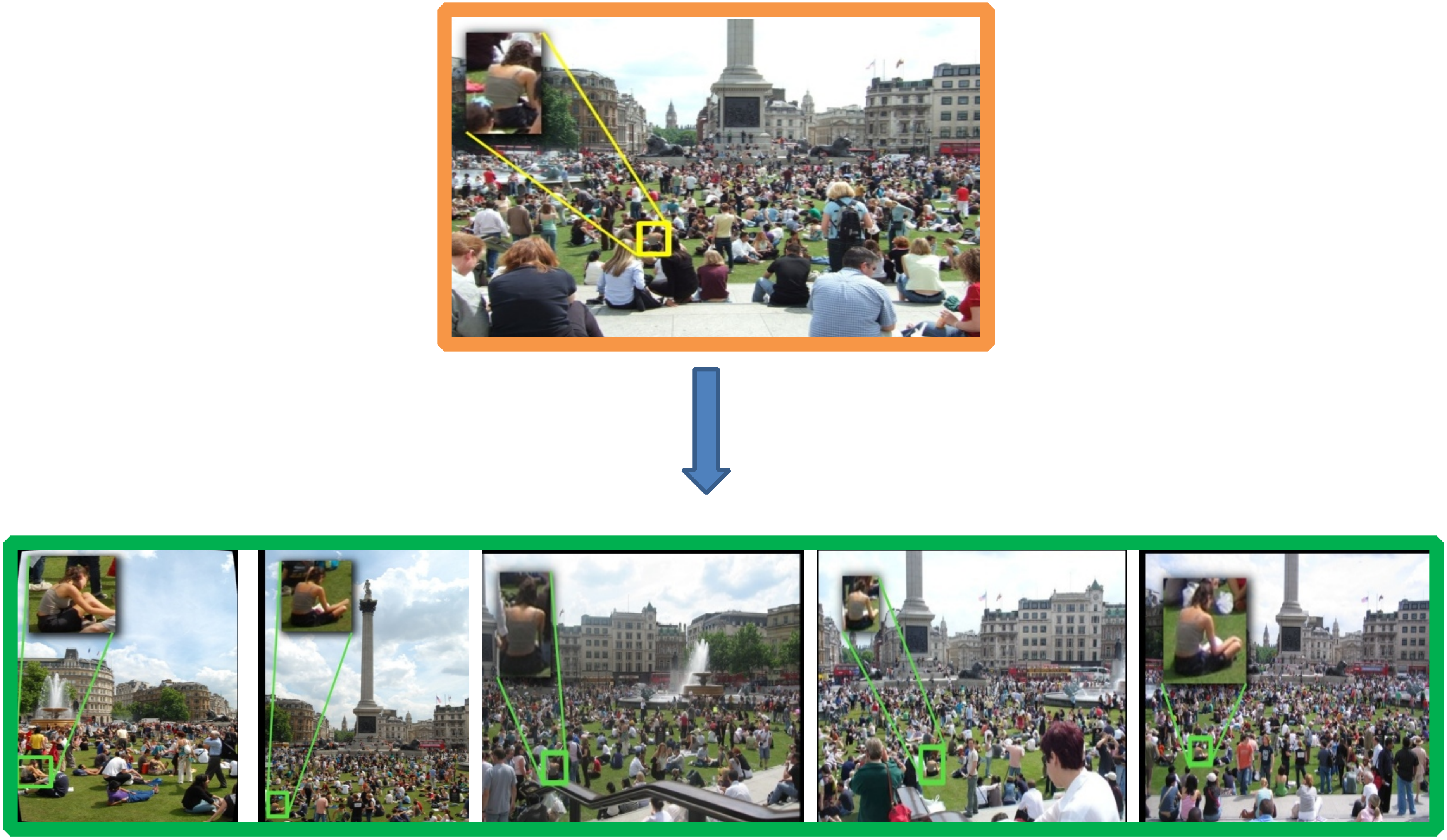
Motivation

{ all photos }

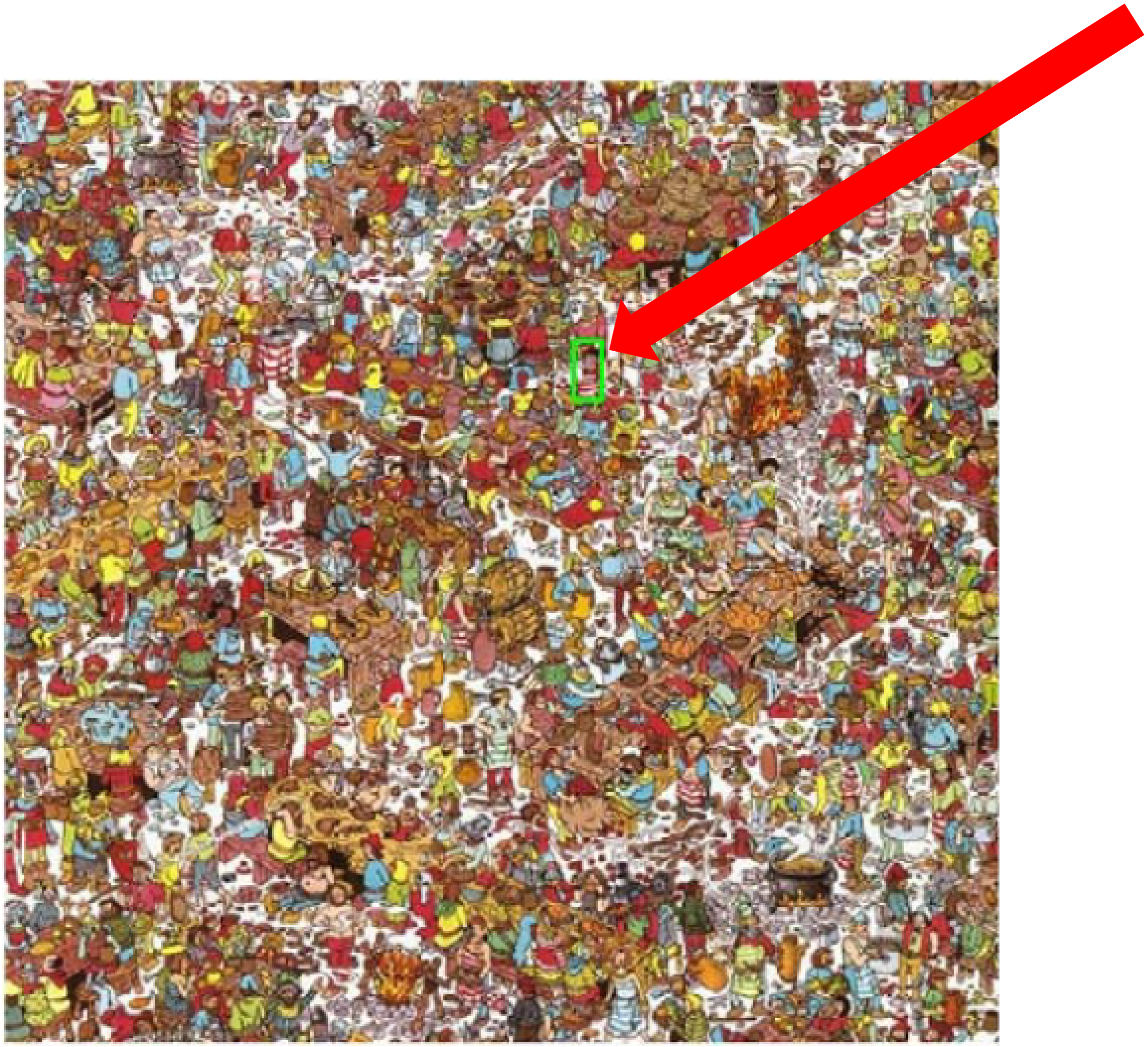
Motivation

Question – *How to browse such a collection and search for someone?*

Problem Definition



?



Applications

1. Photo browsing
2. Surveillance
3. Content based querying / search

   - Richer search experience.

CHALLENGES

Challenges


Pose Change

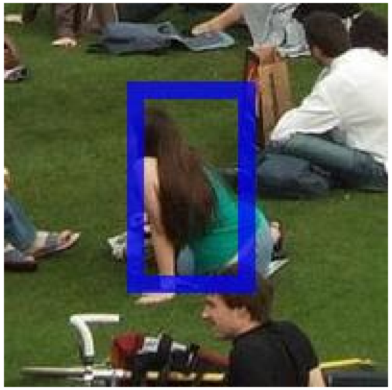

Severe occlusion


Low resolution

Challenges (contd..)

Photos from 100s of users; different viewpoints

Different capture devices from different people.

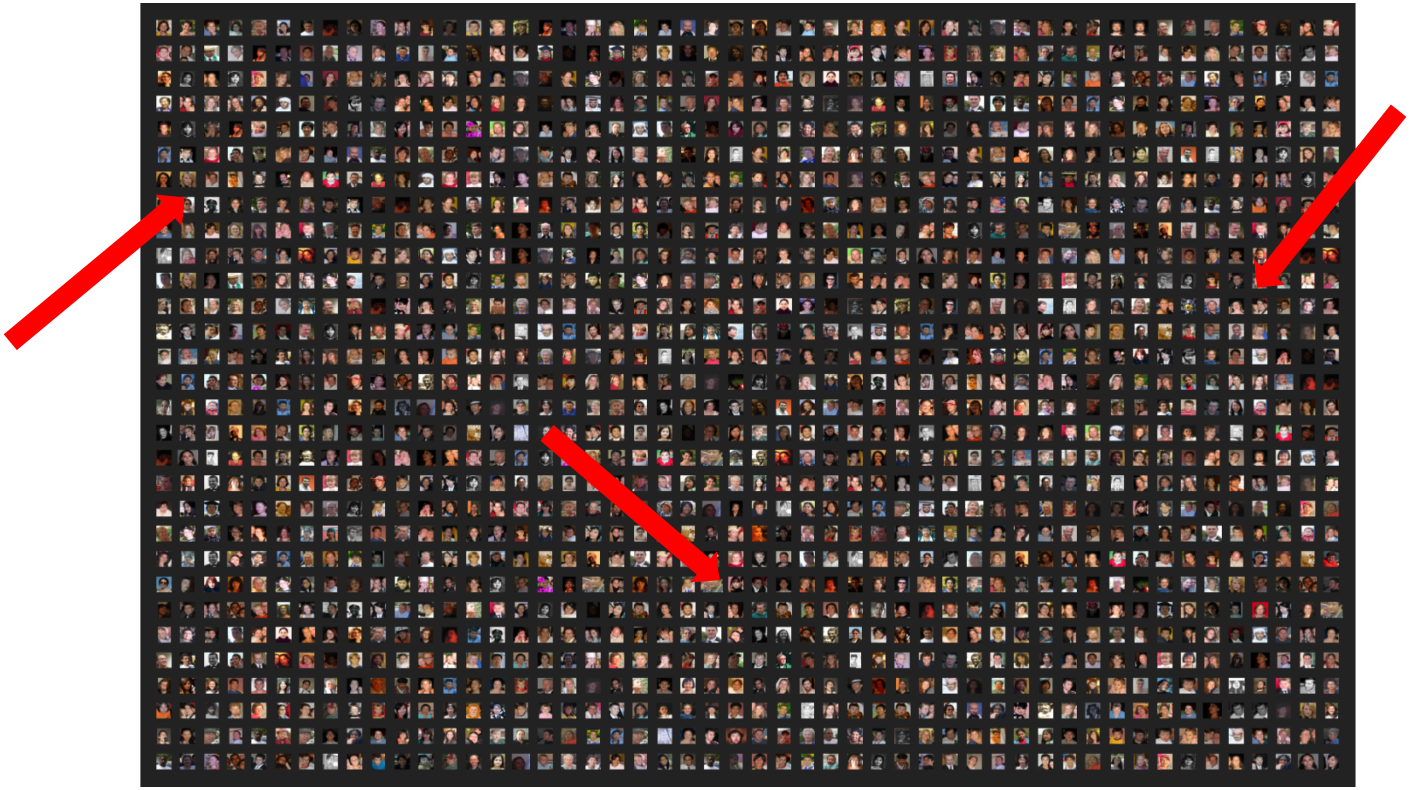


Matching 100s of people where even faces are not clearly visible.

http://homes.cs.washington.edu/~rahul/data/CVPR_supp/index.html

Challenges (contd..)

A particular “Waldo” appears in a small fraction of the entire collection.



http://homes.cs.washington.edu/~neeraj/projects/facesearch/#slides

Solution: Make Realistic Assumptions

1. People are relatively stationary over large intervals.

   Advantage? Multi View Stereo is applicable.

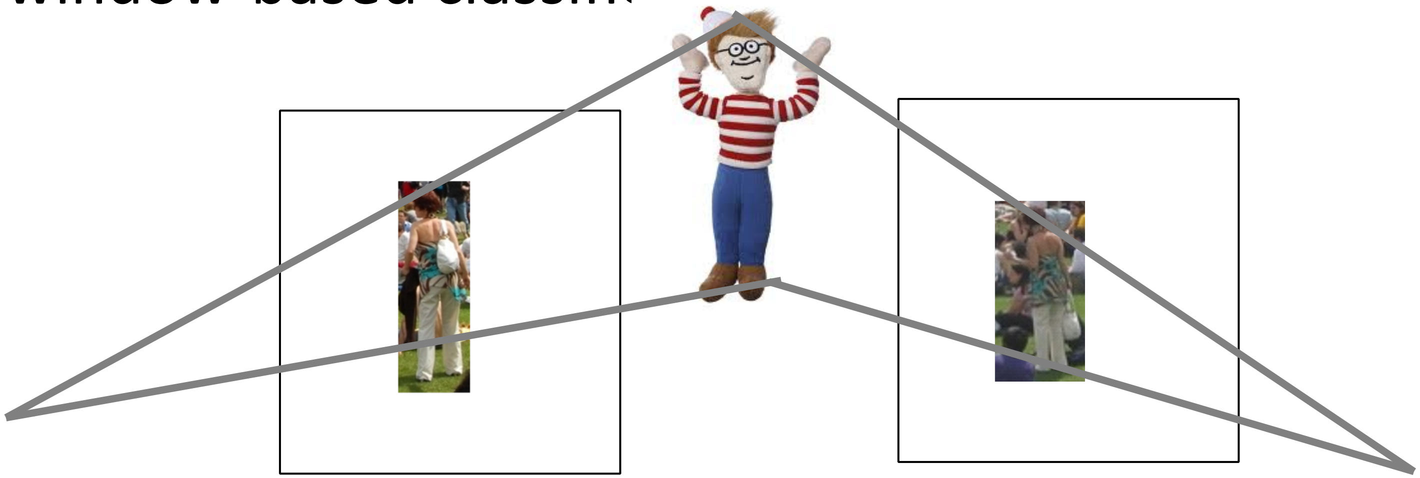
2. Images contain additional contextual information.
   - GPS tags, time stamps.
   - Social context.

   Advantage? Markov Random Field model is applicable.

MAIN CONTRIBUTIONS

Main contributions

1. Generalizing multi-view stereo to people-matching problem
   - NOT template matching
   - Use of a part-based appearance classifier instead of a window-based classifier



http://homes.cs.washington.edu/~rahul/data/CVPR2011/CVPR11-poster.pdf

1) Generalizing multi-view stereo to people-matching problem.

MVS	Waldo Problem
Photo consistency through NCC etc.	Appearance consistency through a part based classifier
3D Localization	3D Localization with **custom priors**
Smoothness in space via MRF	“Smoothness” over time and people via MRF

http://homes.cs.washington.edu/~rahul/data/CVPR2011/CVPR11-poster.pdf

Main contributions (contd..)

2) Exploiting contextual-cues via MRF

- Co-occurrence of people
- Timestamps.

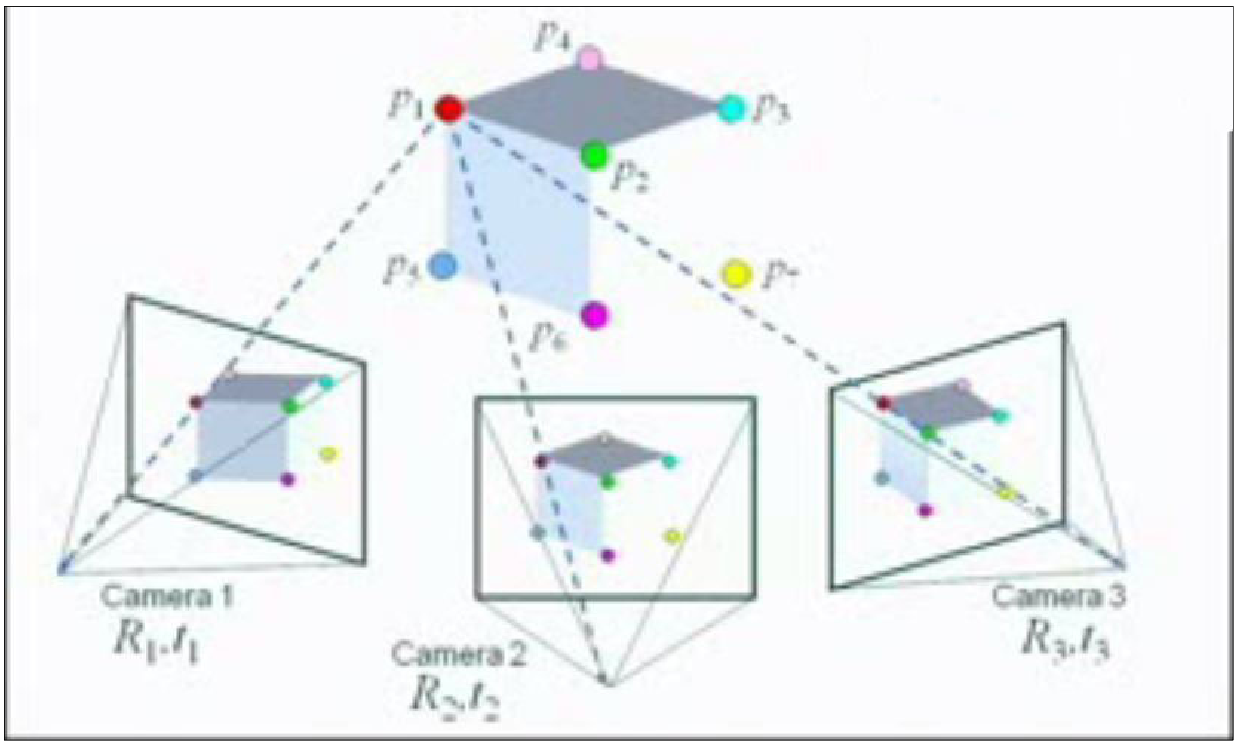
3) Making an extensively labeled dataset available.

METHOD OVERVIEW

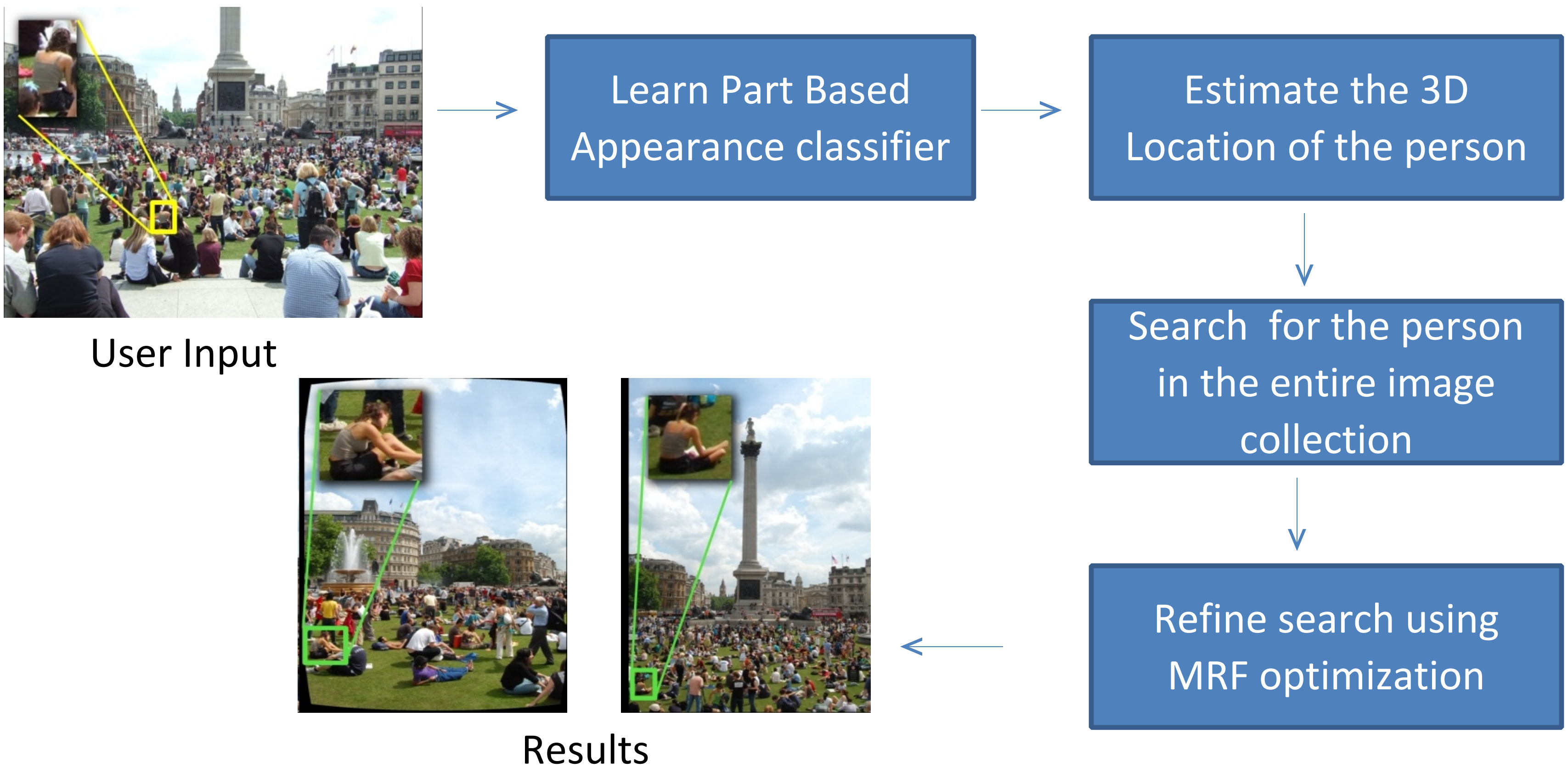
Step#0

Image Collection of an event → Register the Photo Collection using SFM

Structure From Motion



http://grail.cs.washington.edu/projects/cpc/


User Input
Learn Part Based
Appearance classifier
Estimate the 3D
Location of the person
Search for the person
in the entire image
collection
Refine search using
MRF optimization
Results

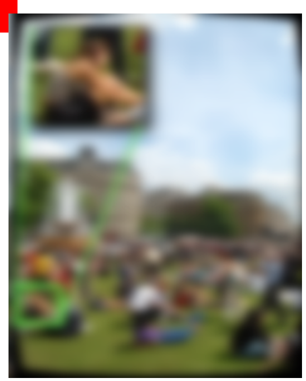
User Input

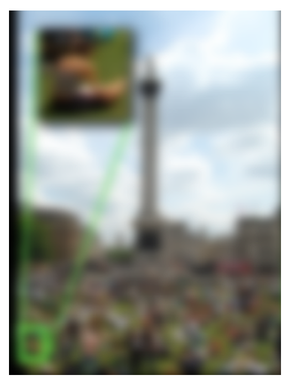
Learn Part Based Appearance classifier

Estimate the 3D Location of the person

Search for the person in the entire image collection

Refine search using MRF optimization

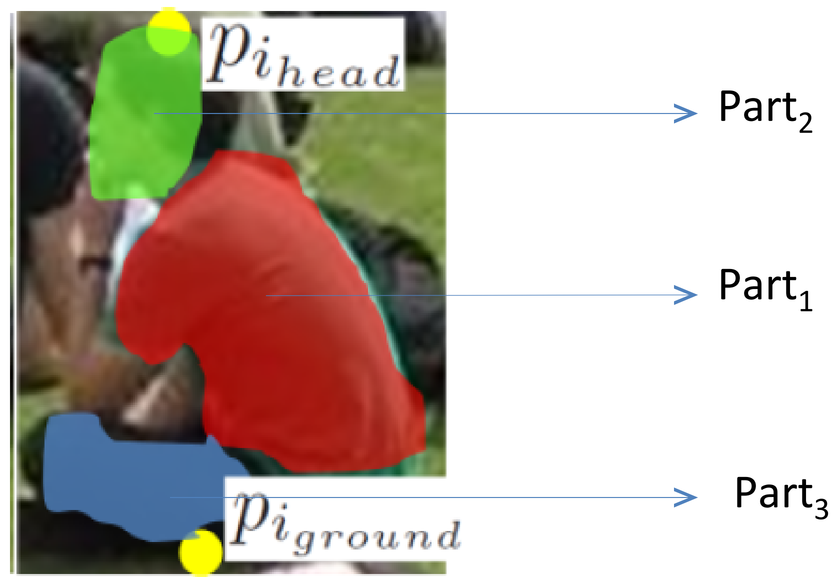




Results

User Input

- Input – Single instance of each person to be searched (p_i)



- Effective since the pose variation is implicitly captured.

Part specific Color Model

Challenges:

- View point
- Scale
- Exposure
- Occlusion

1	←	R	G	B	RG	RB	GB	R^2	G^2	B^2
1	←	R	G	B	RG	RB	GB	R^2	G^2	B^2
-1	←	R	G	B	RG	RB	GB	R^2	G^2	B^2

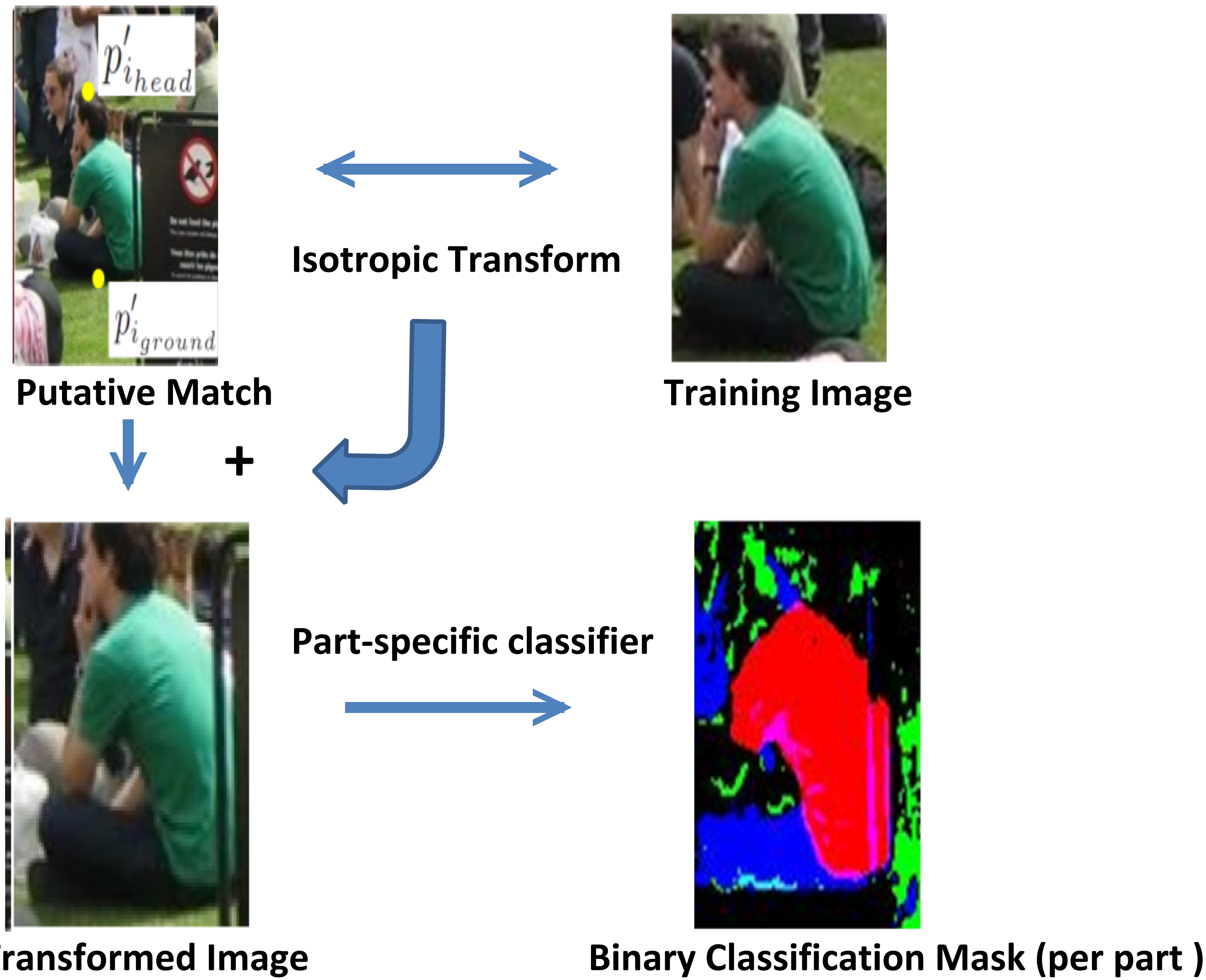
Y

X

⋮

$$w_{part} = \operatorname{argmin}_w \sum_j \log(1 + \exp(-y_j w^T x_j)).$$

Scoring a candidate match



Scoring a candidate match

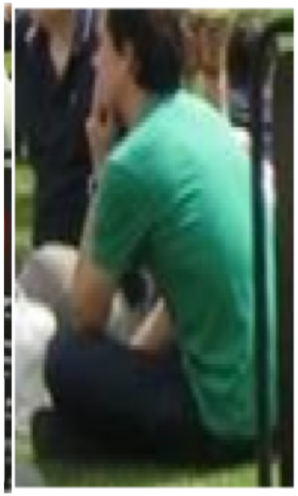
-

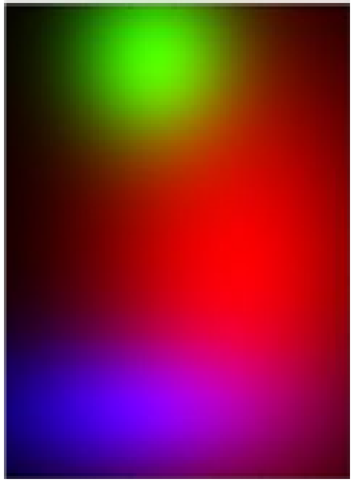
$\sum S_{part_k}$ Sum of the number of positively classified pixels inside a specific *part **k***

Summation accounts for occlusion.

Some parts are more discriminating than others.

$$\text{Score} = \begin{cases} 0 & \text{if } \sum S_{part_1} = 0 \\ \sum_{k=1}^{L} \sum S_{part_k} & \text{otherwise} \end{cases}$$





Discussion

- Very high dependence on the lighting conditions.
  - *Normalize* the RGB values in the appearance model?
  - HSV space or a *different color space*?

- Performance on a similarly dressed crowd images.

Eg: Convocation ceremony.
  - Requires additional cues beyond appearance.

- *Face detection* during appearance modeling (when applicable)
- *Soft threshold* on the appearance score rather than a hard threshold as it is now.

User Input

Learn Part Based Appearance classifier

Estimate the 3D Location of the person

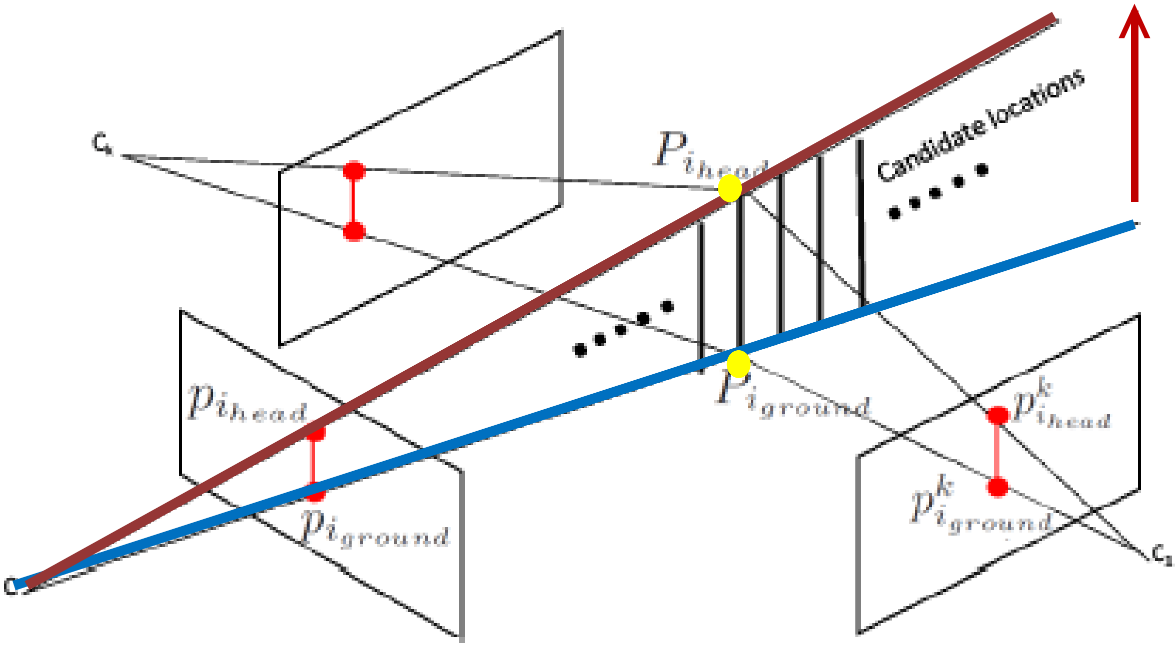
Search for the person in the entire image collection

Refine search using MRF optimization

Results

3D Localization

Assumption: Orientation of the person is along the vertical.



- **Searching in 1-D for $P_{i_{ground}}$**


Learn Part Based Appearance classifier
Estimate the 3D Location of the person
Search for the person in the entire image collection
Refine search using MRF optimization
User Input
Results

3D Localization (contd..)

- For each candidate pair in $(P_{i_{head}}\ P_{i_{ground}})$

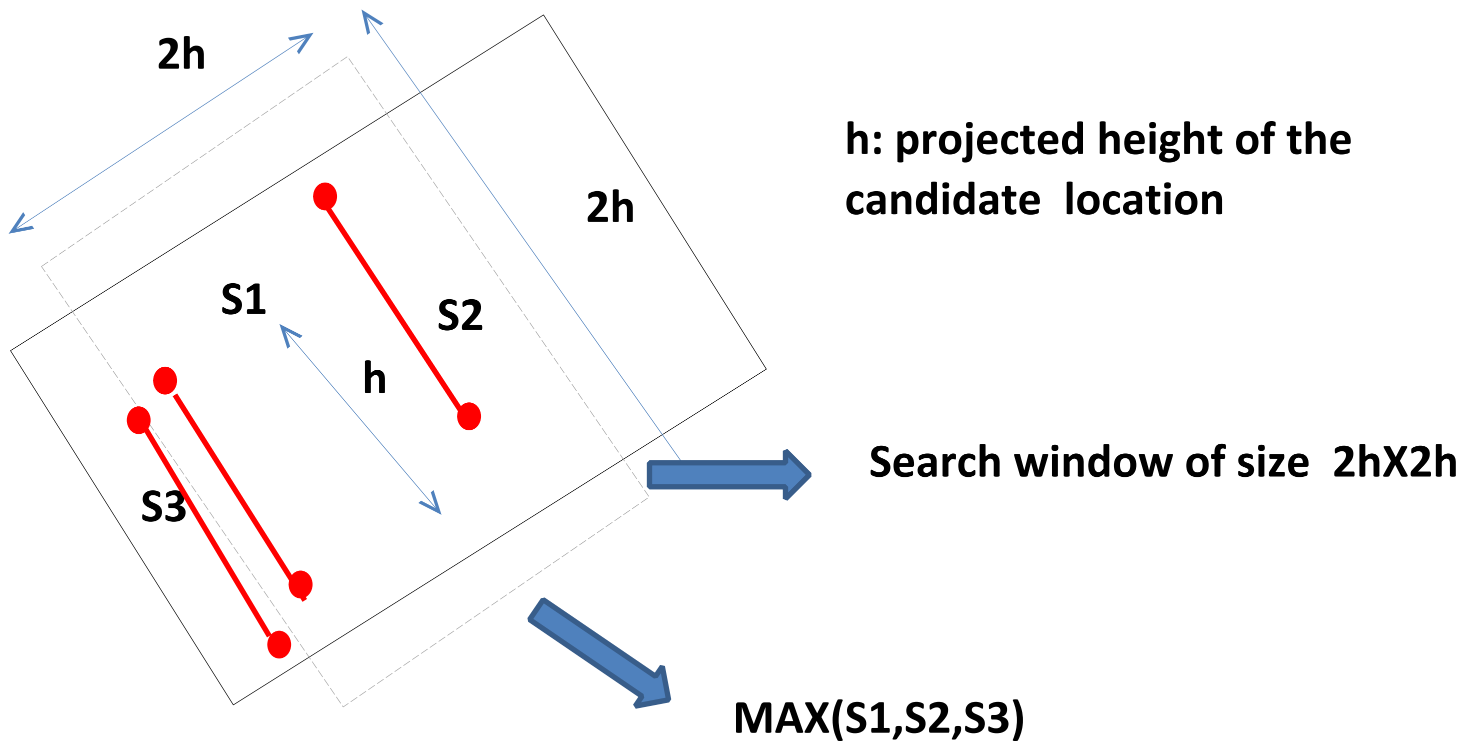
Project it into all the images ***(timestamp constrained)***

- Get $(p^k{}_{i_{head}}\ p^k{}_{i_{ground}})$

Score the projection using appearance model S_i

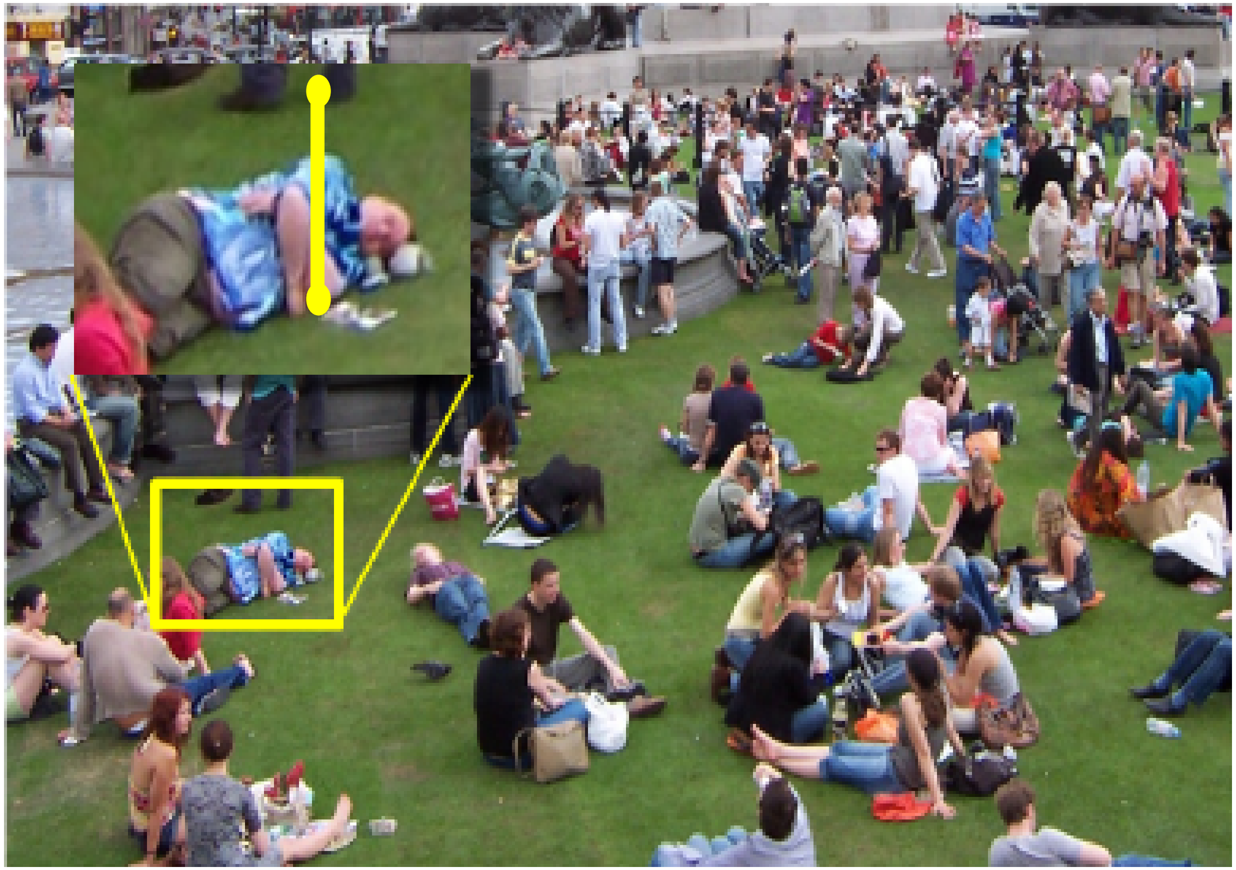
$$\sum_{I_k \in A} max(S_i(p^k_{i_{head}}, p^k_{i_{ground}}) - thresh, 0)$$

Wiggle search



The score is multiplied by height and ground priors.

When orientation of the person is not vertical.

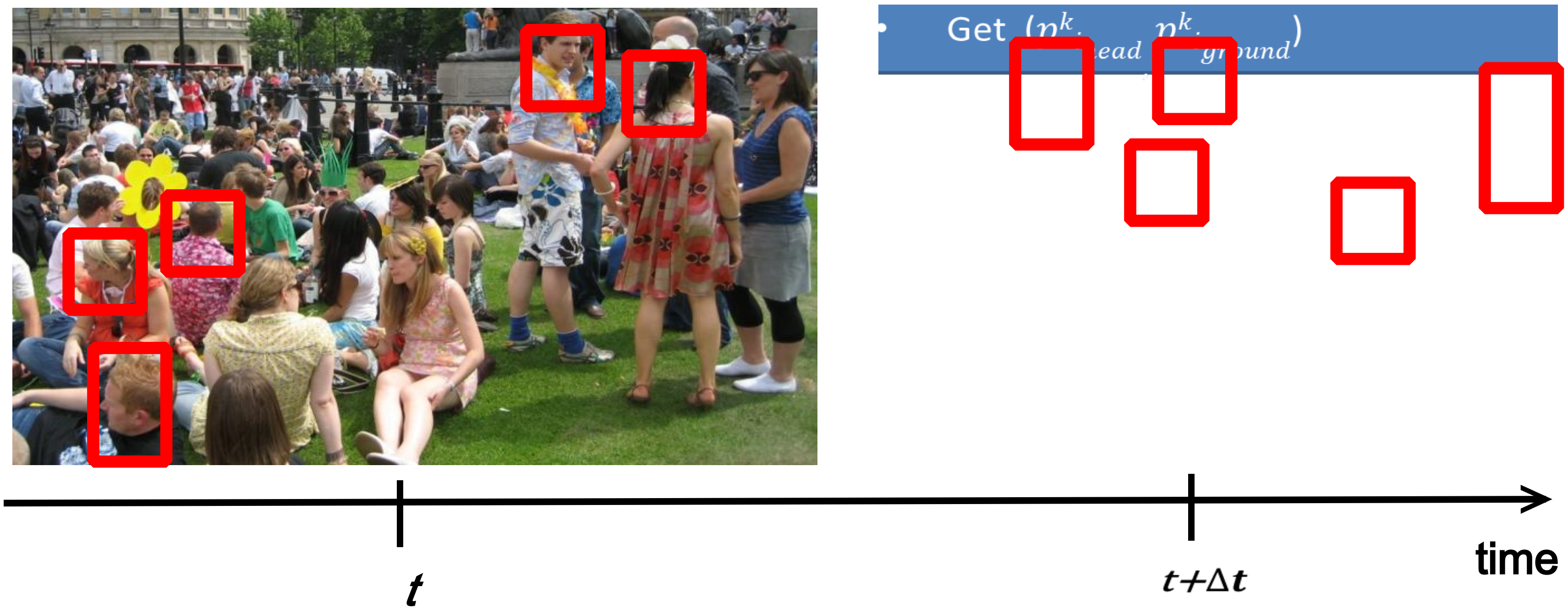


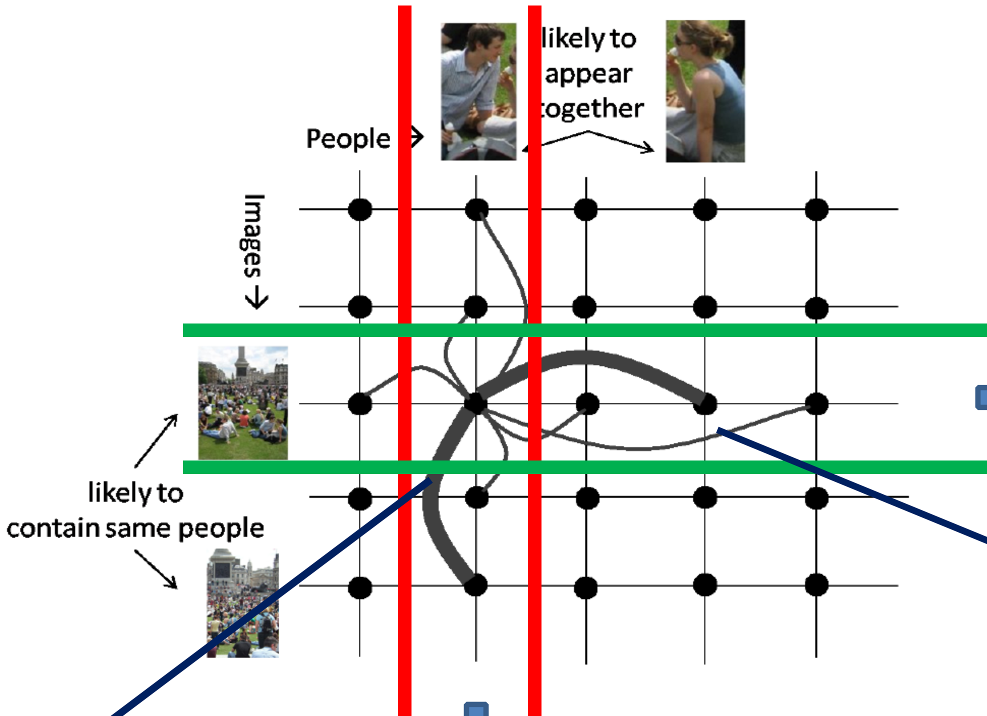
$P_{i_{head}}$ should be marked at a sitting height (Sitting prior)


User Input
Learn Part Based Appearance classifier
Estimate the 3D Location of the person
Search for the person in the entire image collection
Refine search using MRF optimization
Results

Contextual Cues

1. People appear ***together*** with the same group of people.
2. Images which are ***nearby in time*** are likely to contain





People who are likely to appear together in an image

People Affinity $\alpha_p\,(p_i, p_{i'})$

Image Affinity $\alpha_I\,(I_j, Ij_{'})$

A person in a set of similar images.

Minimize Objective Function

$$E(\pounds) = \sum_{ij} U(l_{ij}) + \sum_{ij}\sum_{i'j'} \phi(l_{ij}, l_{i'j'})$$

Unary Potential

Pairwise potentials

Discussion

- For the MRF model to be applicable, is every person, in every image, every time?
  - (OR) Is every person in the training image identified?

- Cues hallucinate the person when not present if other people with high affinities with that person are detected in the image.
  - Wont the appearance score be zero for this missing person?

User Input

Learn Part Based Appearance classifier

Estimate the 3D Location of the person

Search for the person in the entire image collection

Refine search using MRF optimization

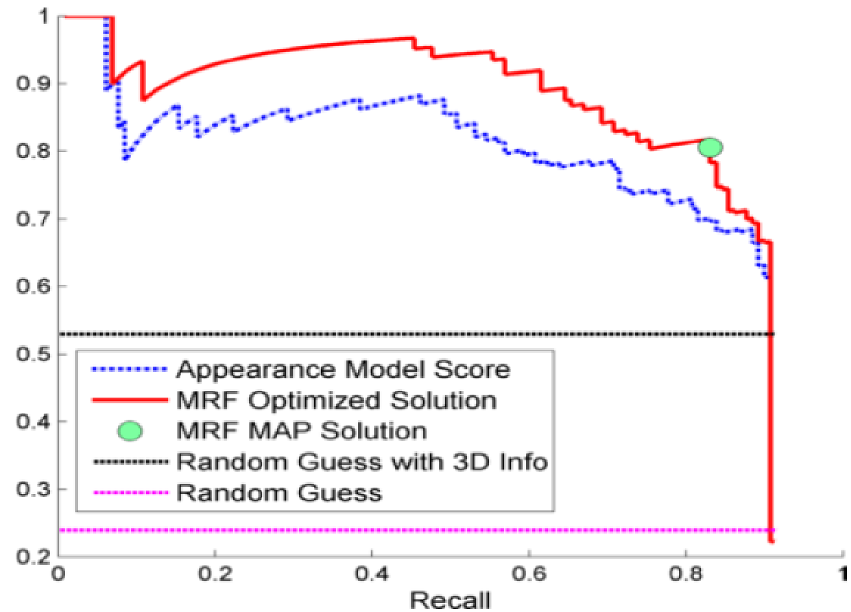
Results

Datasets & Ground Truth Data

- **Dataset#1**
  - **34 photos ; single photographer ; Trafalgar Square ; single day.**
- **Dataset#2**
  - **282 photos ; 89 different photographers ; Trafalgar Square ; single day.**
- **Dataset#3**
  - **45 photos from 19 different users taken ; Hackday ; over two days. (Indoor)**
- **Ground truth labeling**
  - **Manually labeled with assistance from geometry**
  - **Does not follow the contextual cues.**

Results – Dataset#1

Results of individual people



Precision-Recall curves

Results – Dataset#2

Illustrating failure to identify matches

- Torso (Red) not distinct from the background.
- Blue – too many colors.

Extensions

- Relaxing each of the assumptions made.
  - Allow large motion of people.
- Track people’s movement through the scene.
- More powerful and accurate appearance models.
- Larger image datasets.

Understanding Images of Groups of People