Spatio-Temporal GrabCut Human Segmentation for Face and Pose Recovery

Antonio Hernández
Miguel Reyes
Sergio Escalera
Petia Radeva

Face and Pose Recovery







Outline

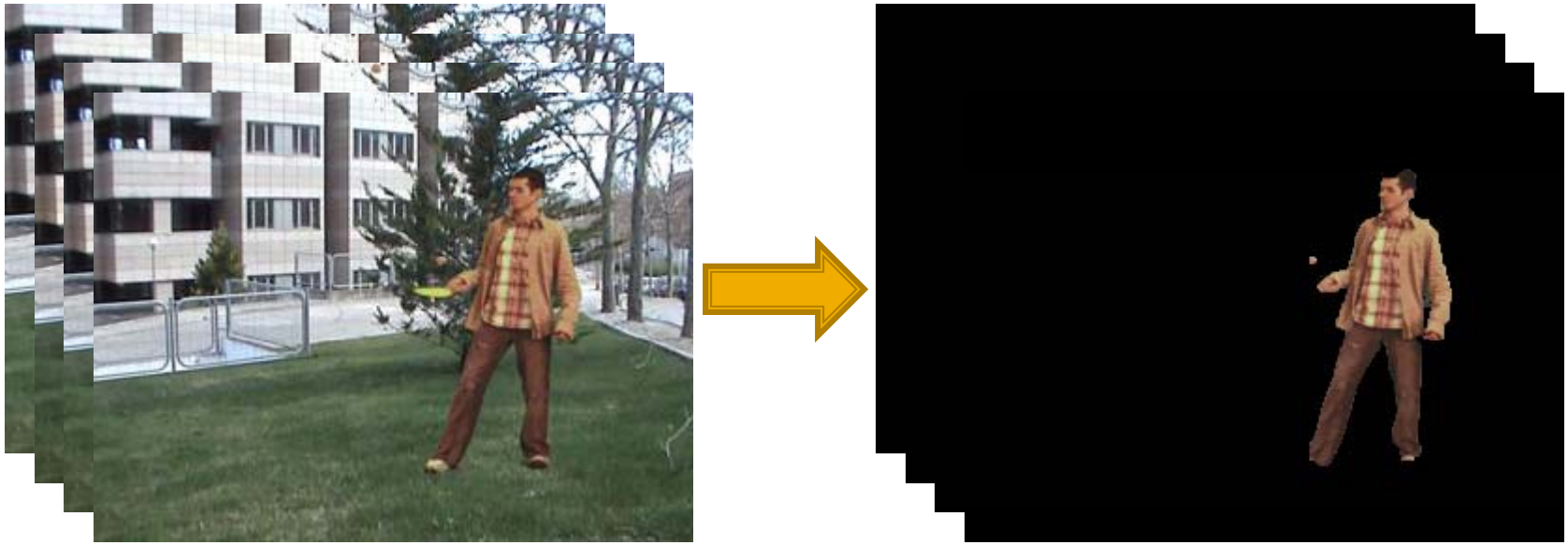
- Introduction
- Methodology
  - Spatio-Temporal GrabCut → Human segmentaton
  - Active Appearance models → Face fitting
  - Conditional Random Field → Limb recovery
- Validation
  - Existent public data sets
  - New Human Body Limb data set
- Conclusions and future work

Outline

- **Introduction**
- Methodology
  - Spatio-Temporal GrabCut → Human segmentaton
  - Active Appearance models → Face fitting
  - Conditional Random Field → Limb recovery
- Validation
  - Existent public data sets
  - New Human Body Limb data set
- Conclusions and future work

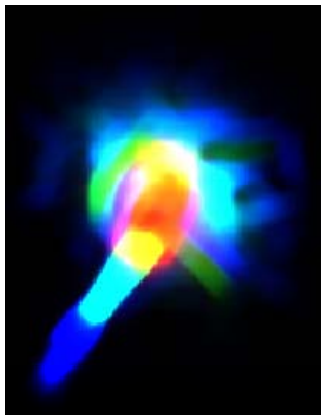
Introduction and motivations

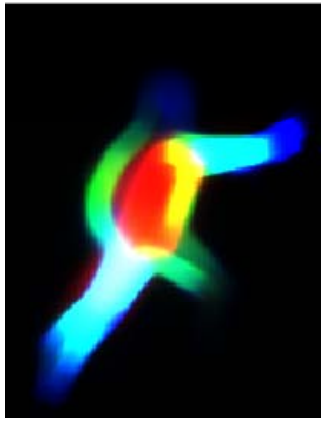
- Video oriented
- Human segmentation and pose recovery



Introduction and motivations

- Human segmentation for human pose [1] and face recovery [2] assistance



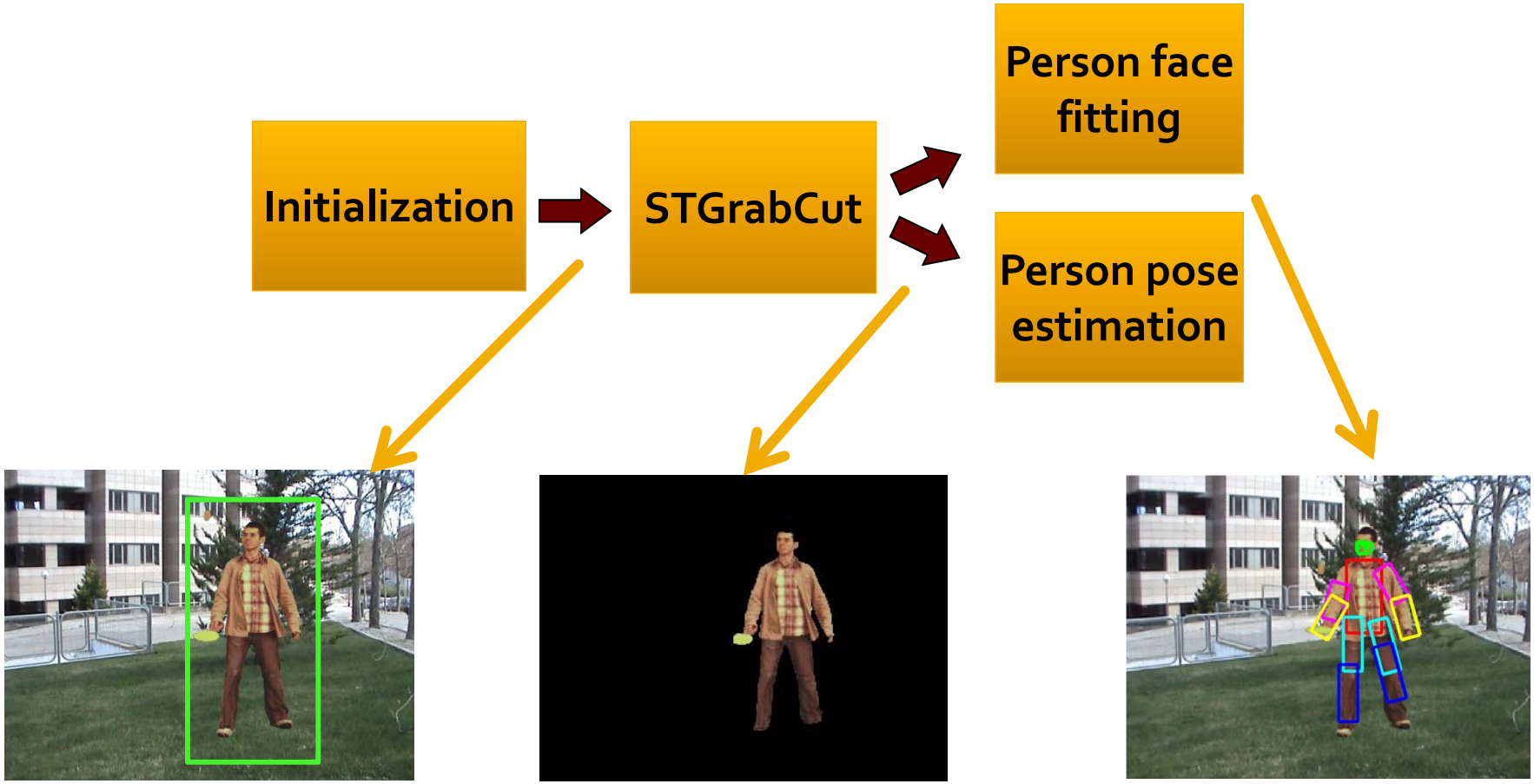


[1] Ramanan, D. "**Learning to Parse Images of Articulated Bodies** ", *In NIPS, 2006*.
[2] T. Cootes, J.Edwards and C. Taylor, "**Active Appearance models.**", *IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998*.

Outline

- Introduction
- **Methodology**
  - **Spatio-Temporal GrabCut** → **Human segmentaton**
  - **Active Appearance models** → **Face fitting**
  - **Conditional Random Field** → **Limb recovery**
- Validation
  - Existent public data sets
  - New Human Body Limb data set
- Conclusions and future work

Methodology



Methodology

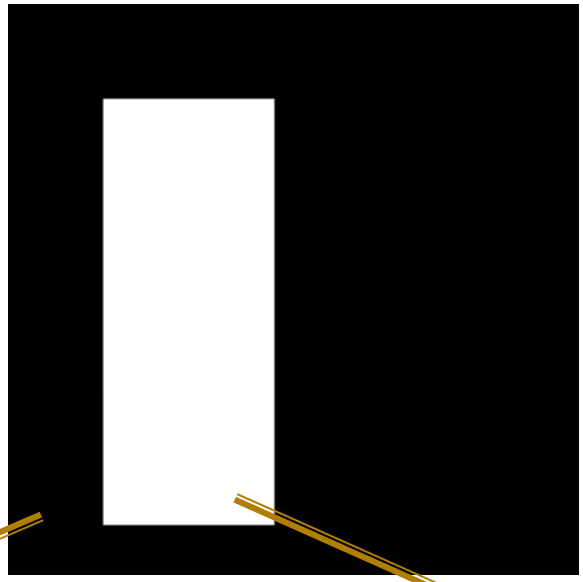
Seed initialization

STGrabCut

Person pose estimation

Person face fitting

- HOG-based [4] person detector

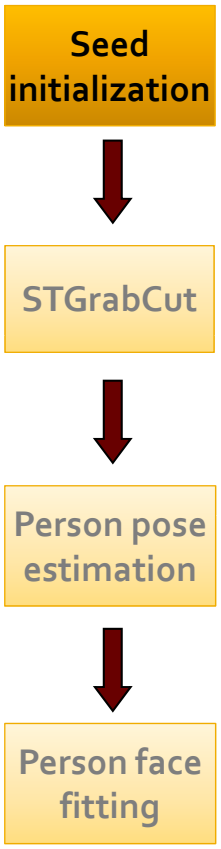


Background seeds

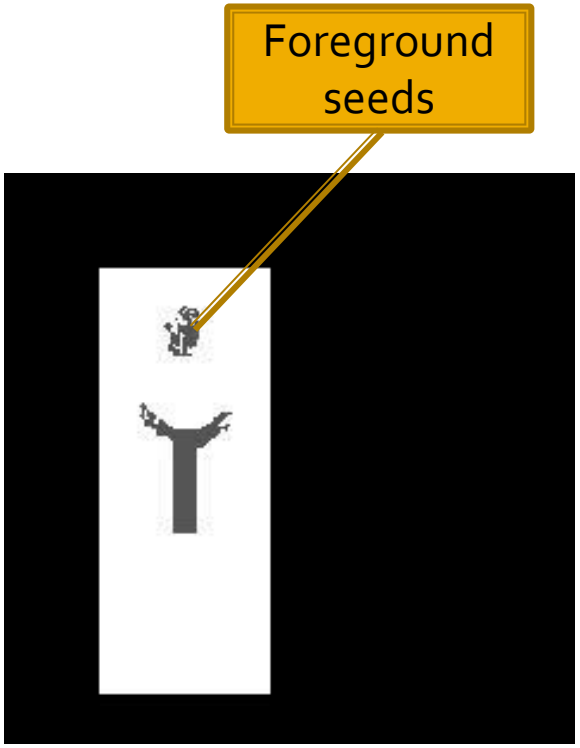
Unknown region

[4] Navneet Dalal and Bill Triggs, "**Histograms of Oriented Gradients for Human Detection**" , *In CVPR, 2005*.

Methodology



- Face detection [5]



[5] Paul Viola and Michael Jones, "**Robust real-time face detection**" , *International Journal of Computer Vision, 2004*.

Methodology

Seed initialization

↓

STGrabCut

↓

Person pose estimation

↓

Person face fitting

- GrabCut segmentation

GrabCut

↓

Pixel-level segmentation refinement

↓

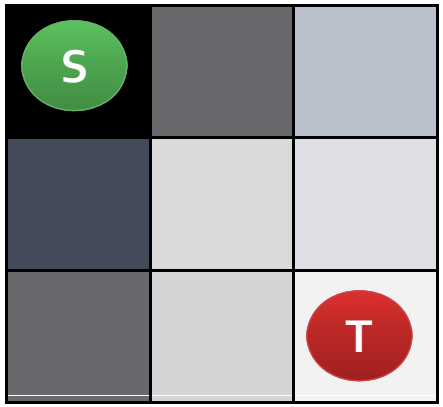
GrabCut

↓

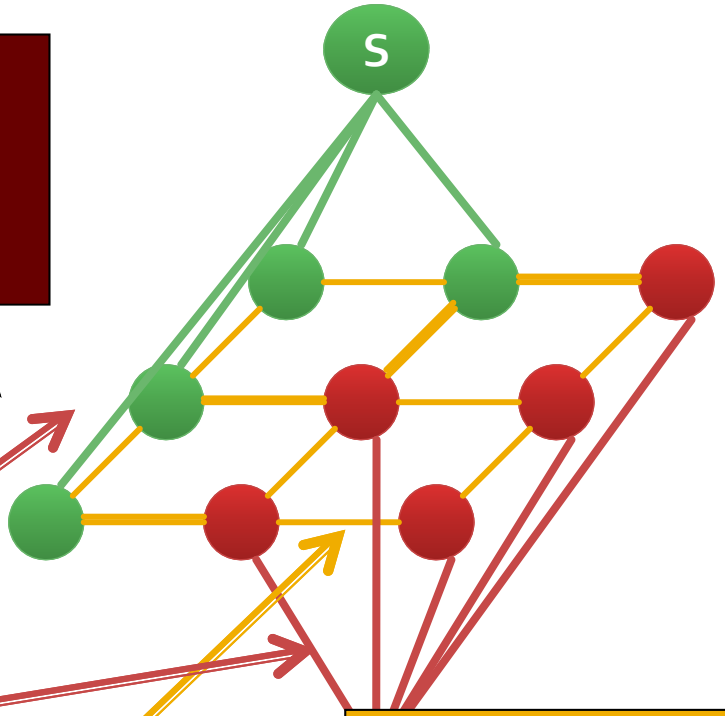
Seed init for next frame

GrabCut [3]

- Toy example



Log-likelihood of GMM models (BGD,FGD) over RGB



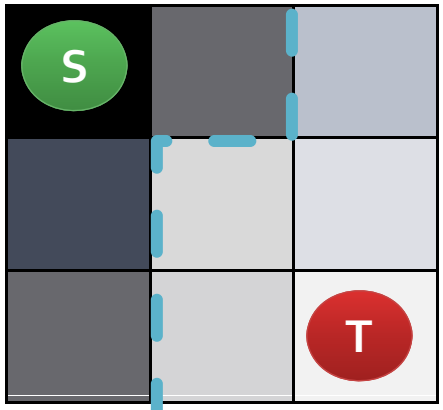
Pixel differences based on RGB Euclidean distance

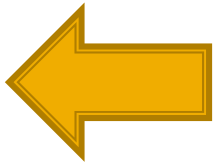
$$\mathbf{E}(\underline{\alpha}, \mathbf{k}, \underline{\theta}, \mathbf{z}) = U(\underline{\alpha}, \mathbf{k}, \underline{\theta}, \mathbf{z}) + V(\underline{\alpha}, \mathbf{z})$$

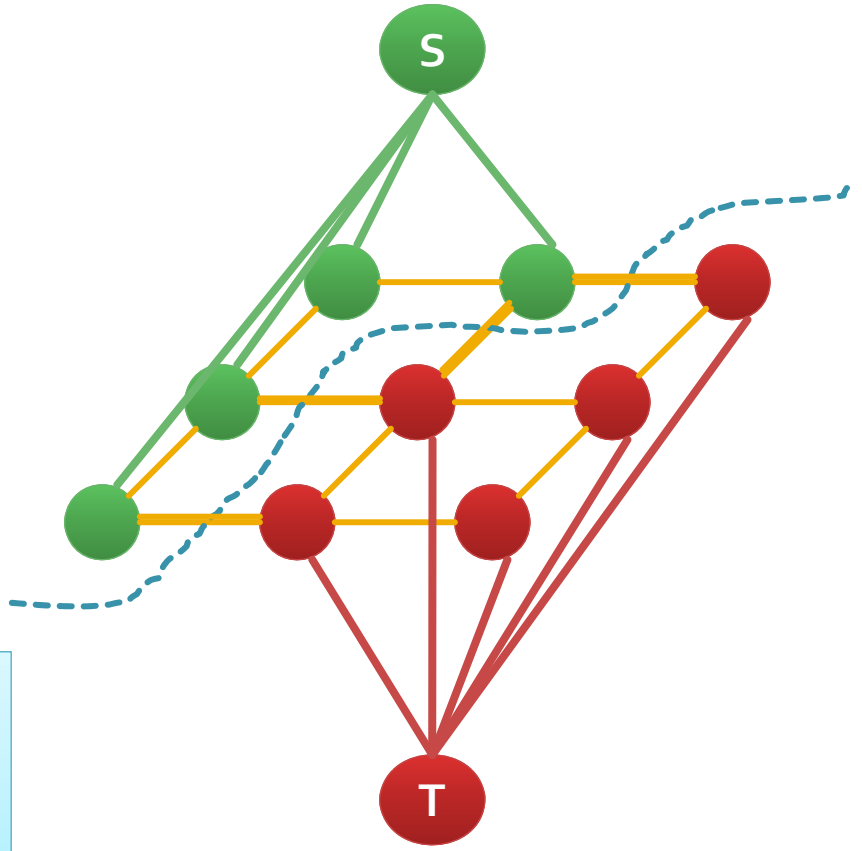
[3] C Rother, V Kolmogorov, A Blake. "**Grabcut: Interactive foreground extraction using iterated graph cuts**" , *ACM Transactions on Graphics, 2004*.

GrabCut [3]

- Toy example







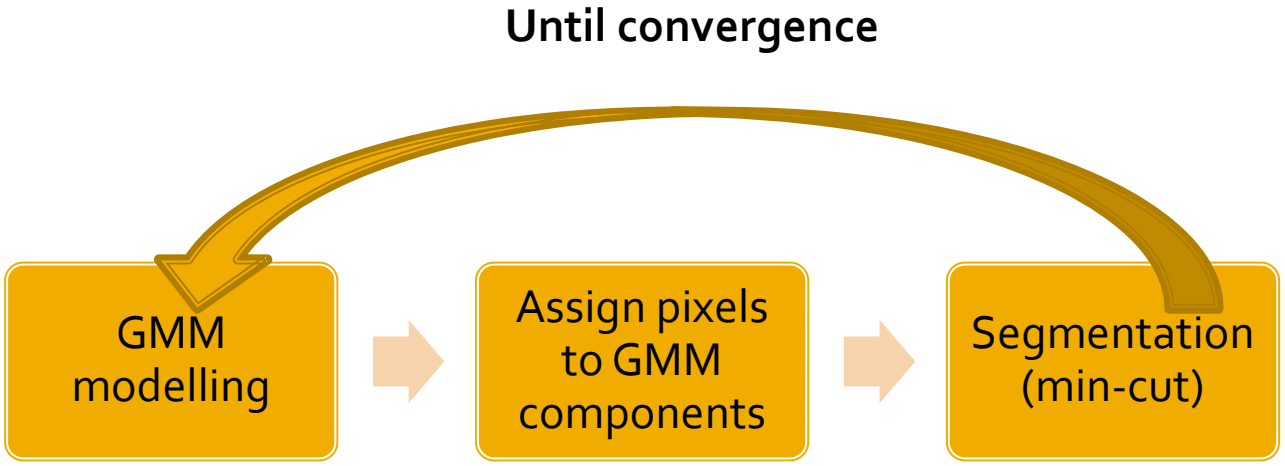
Minimum cut == minimum energy

$$|C| = \sum_{i \in C} \omega_i$$

[3] C Rother, V Kolmogorov, A Blake. "**Grabcut: Interactive foreground extraction using iterated graph cuts**" , *ACM Transactions on Graphics, 2004*.

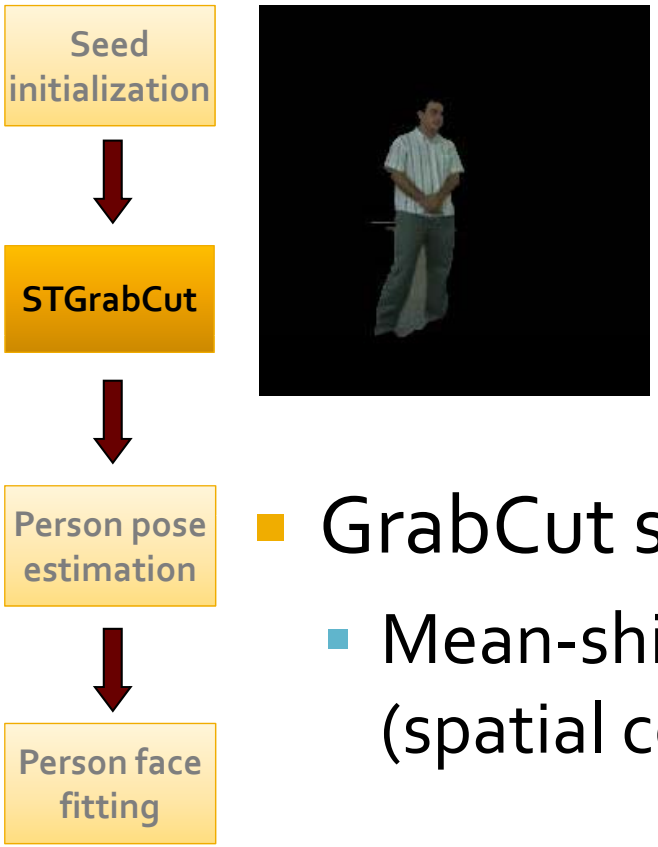
GrabCut [3]

- Iterative procedure

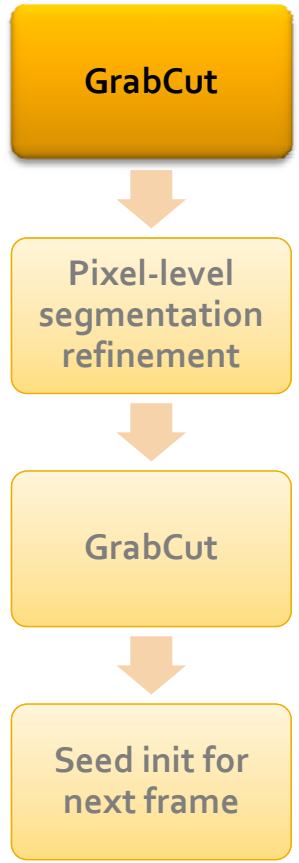


[3] C Rother, V Kolmogorov, A Blake. "**Grabcut: Interactive foreground extraction using iterated graph cuts**" , *ACM Transactions on Graphics, 2004*.

Methodology



- GrabCut segmentation
  - Mean-shift GMM initialization (spatial coherence)



Methodology

Seed initialization

STGrabCut

Person pose estimation

Person face fitting

- Biased to convex regions!
- Background seeds for pixels:

$$\left\{ z_i \Big| \sum_{k=t-j}^{t} \frac{p(z_i|\alpha_i = 0, k_i, \boldsymbol{\theta}^k)}{j} > \sum_{k=t-j}^{t} \frac{p(z_i|\alpha_i = 1, k_i, \boldsymbol{\theta}^k)}{j} \right\}$$

GrabCut

Pixel-level segmentation refinement

GrabCut

Seed init for next frame

Methodology

Seed initialization

STGrabCut

Person pose estimation

Person face fitting

- GrabCut re-iteration
  - Infer segmentation with new background seeds

GrabCut

Pixel-level segmentation refinement

GrabCut

Seed init for next frame

Methodology

Seed initialization

STGrabCut

Person pose estimation

Person face fitting

- Initialization for next frame

$$
\begin{aligned}
T_F &= \{z_i \in A \ominus ST_d\} \\
T_U &= \{z_i \in A \oplus ST_e\} \setminus T_F \\
T_B &= \{z_i, i = 1..N\} \setminus (T_F \cup T_U)
\end{aligned}
$$

GrabCut

Pixel-level segmentation refinement

GrabCut

Seed init for next frame

Methodology

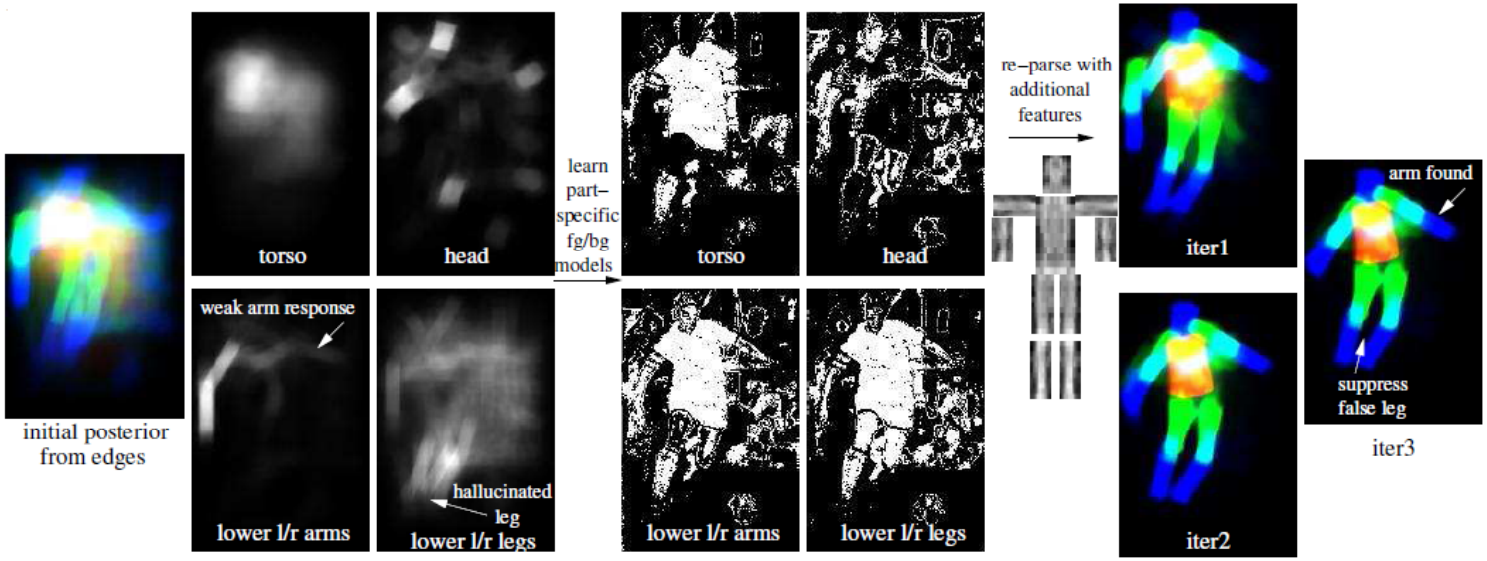
Seed initialization

STGrabCut

Person pose estimation

Person face fitting

- CRF-based [1] pose recovery



$$P(L|f_t) \propto \exp\left(\sum_{(i,j)\in E} \Psi(l_i, l_j) + \sum_i \Phi(l_i|f_t)\right)$$

[1] Ramanan, D. "**Learning to Parse Images of Articulated Bodies** ", *In NIPS, 2006*.

Methodology

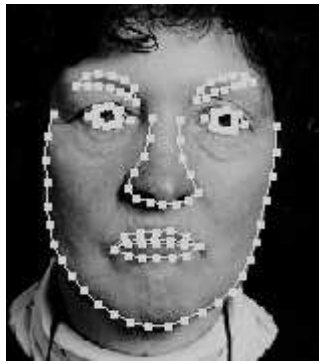
Seed initialization

STGrabCut

Person pose estimation

Person face fitting

- AAM-based [2] face recovery



- 3 meshes: 1 frontal, 2 lateral

$$\Im^{t+1} = \min_{\Im^{t+1}} \{E_{\Im_F}, E_{\Im_R}, E_{\Im_L}\}, \Im^{t+1} \in \nu(\Im^t)$$

[2] T. Cootes, J.Edwards and C. Taylor, "**Active Appearance models.**", *IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998*.

Outline

- Introduction
- Methodology
  - Spatio-Temporal GrabCut → Human segmentaton
  - Active Appearance models → Face fitting
  - Conditional Random Field → Limb recovery
- **Validation**
  - **Existent public data sets**
  - **New Human Body Limb data set**
- Conclusions and future work

Validation

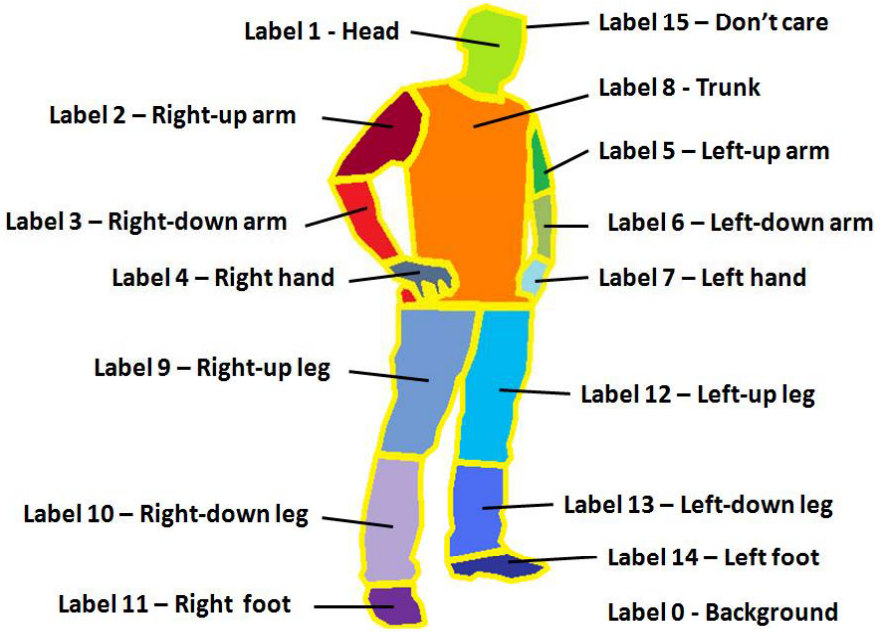
- CVSG data set [6]
  - Video sequence: 307 frames
  - Ground truth

- Undergraduate thesis defense
  - 4 video sequences: 720 frames each

[6] **F. Tiburzi, M. Escudero, J. Bescos, and J. Martinez**."A ground-truth for motion-based video-object segmentation" I*EEE International Conference on Image Processing (Workshop on Multimedia Information Retrieval, 2008*

Validation

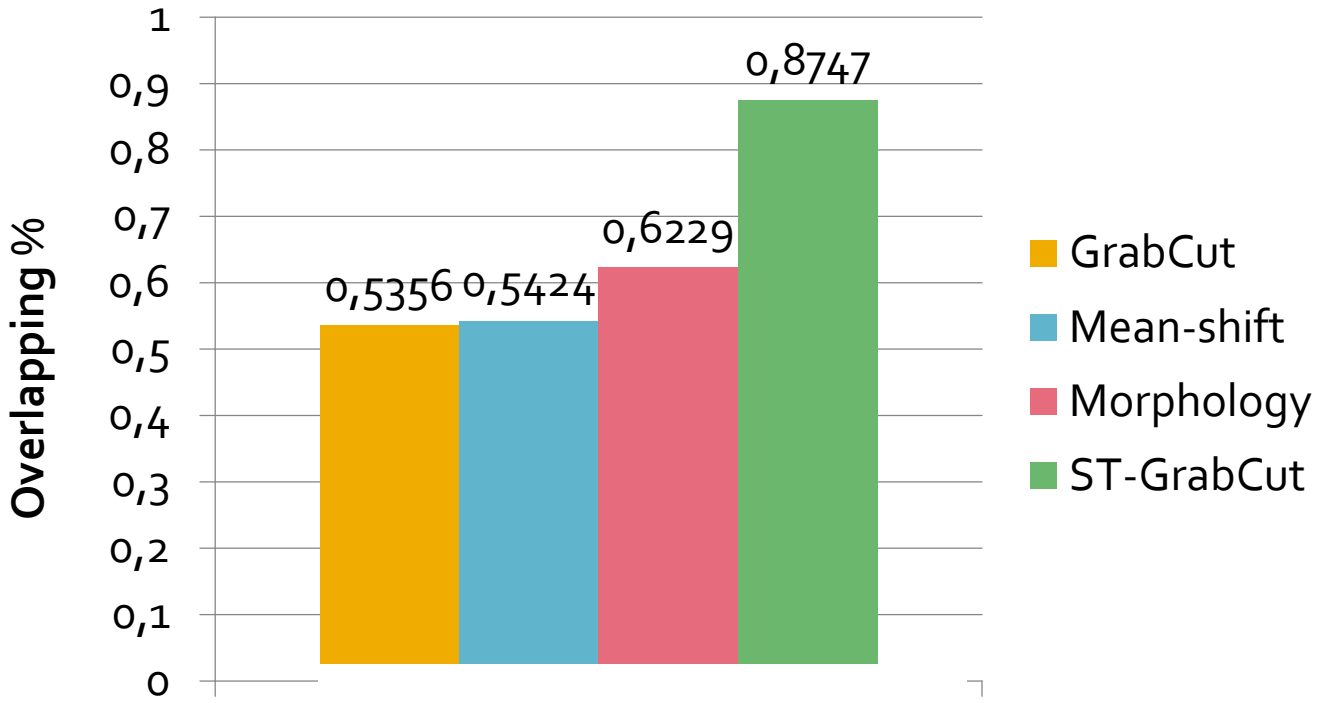
- New human body limb data set
  - 227 images
  - 25 different people
  - Ground truth



http://www.maia.ub.es/~sergio/Code.html

Validation: Segmentation

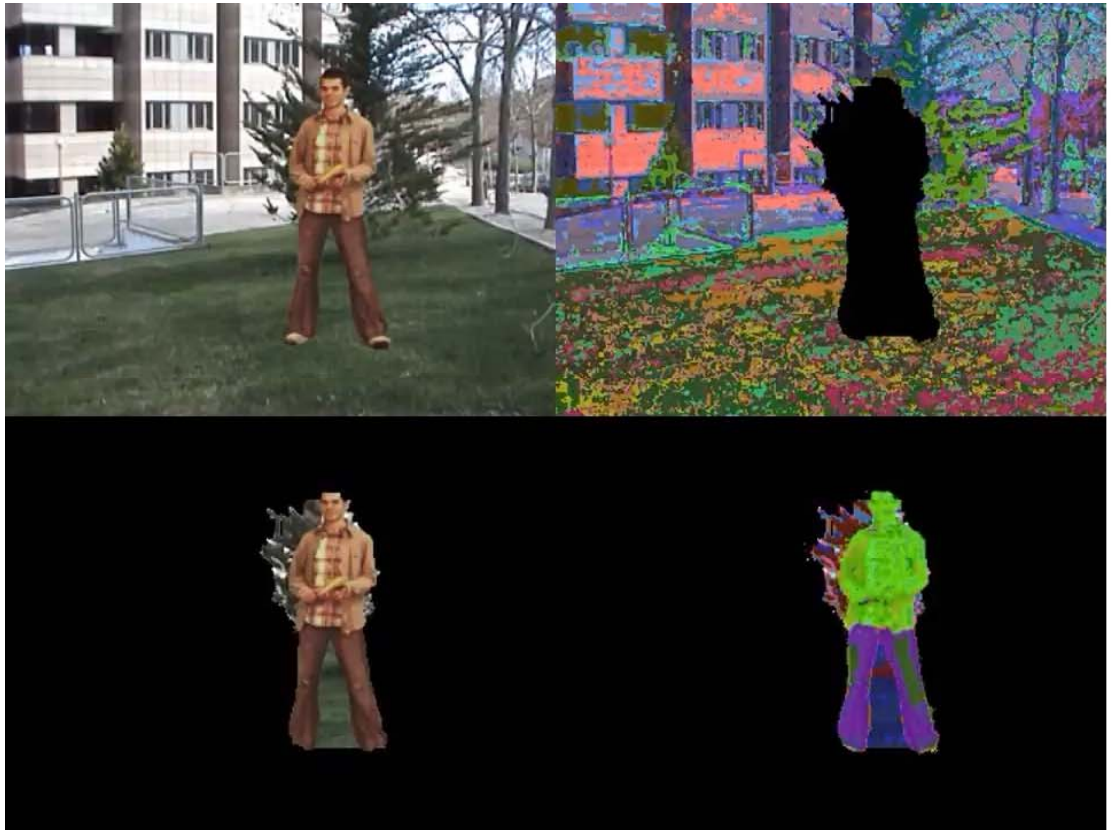
- Quantitative results (CVSG data set)

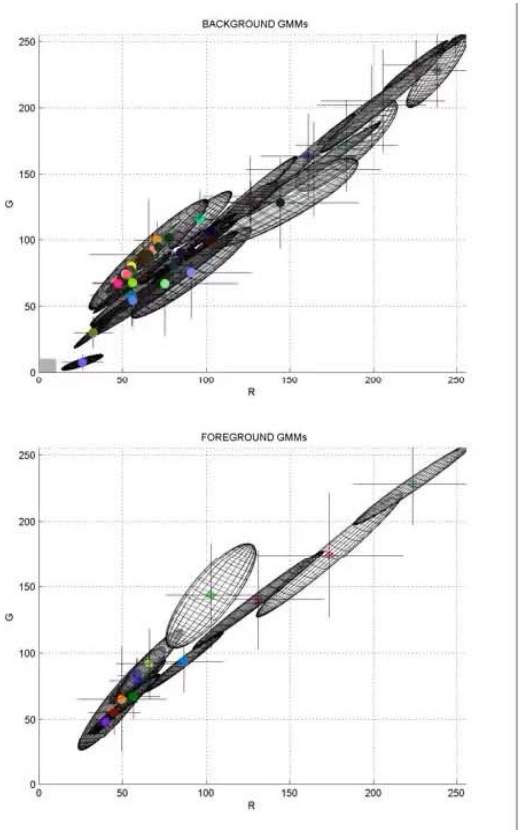


$$O = \frac{\sum M_{GC} \cap M_{GT}}{\sum M_{GC} \cup M_{GT}}$$

Validation: Segmentation

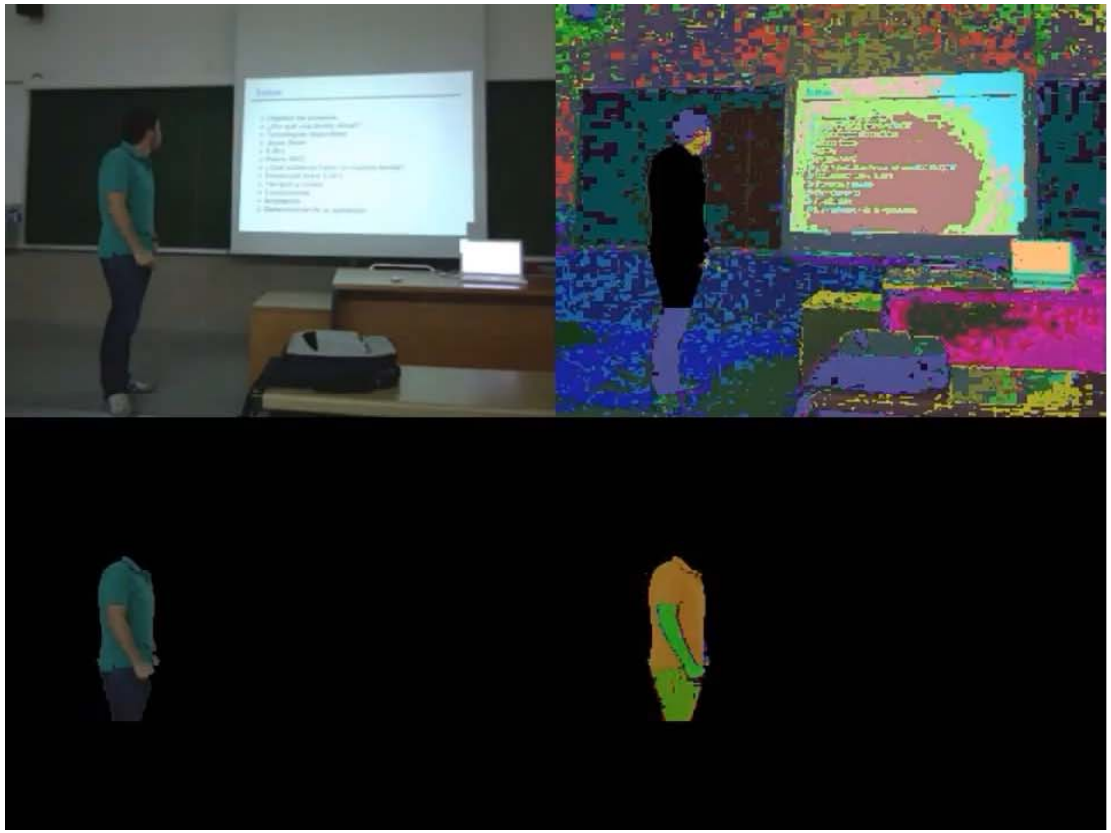
- Qualitative results (CVSG data set)

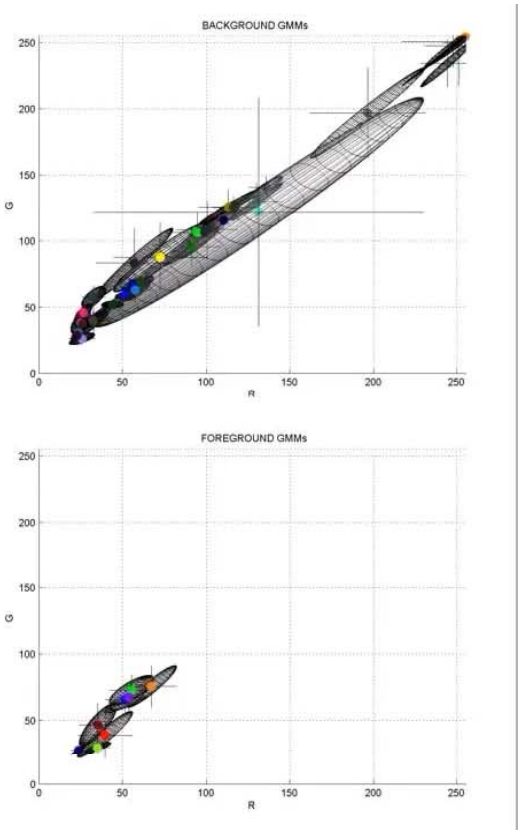




Validation: Segmentation

- Qualitative results (UB data set)

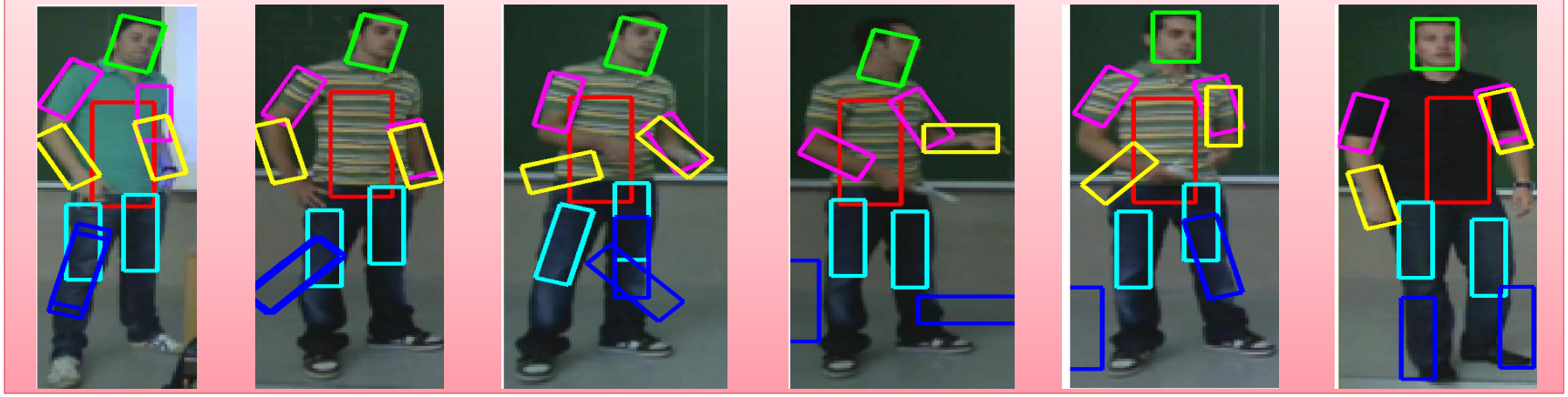




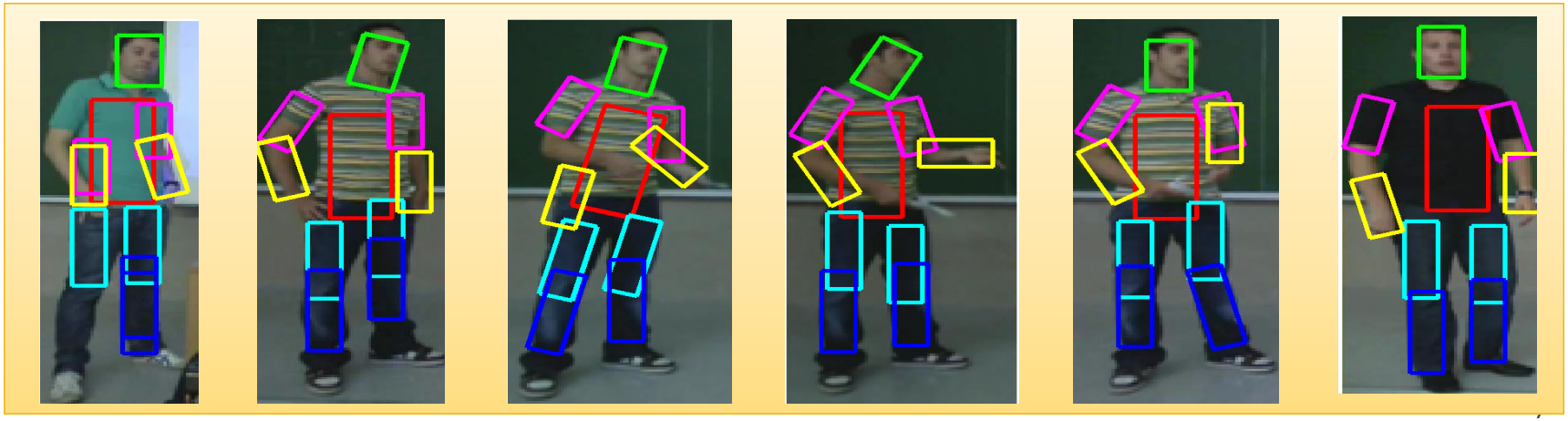
Validation: Pose recovery

- Body pose recovery

No segmentation

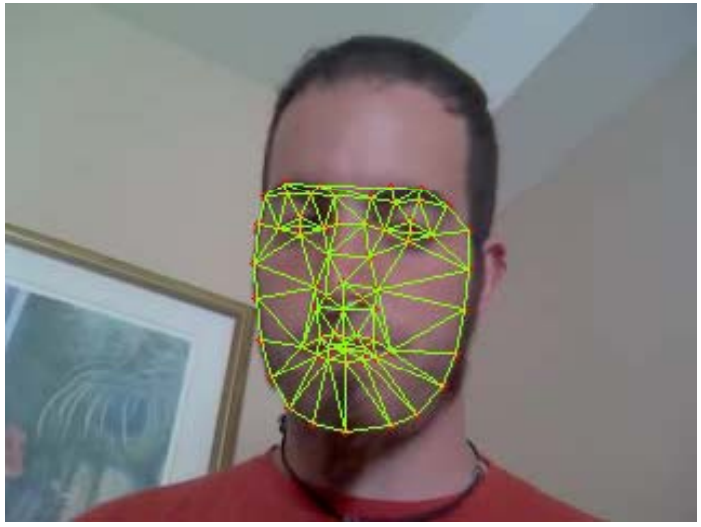


ST-GrabCut



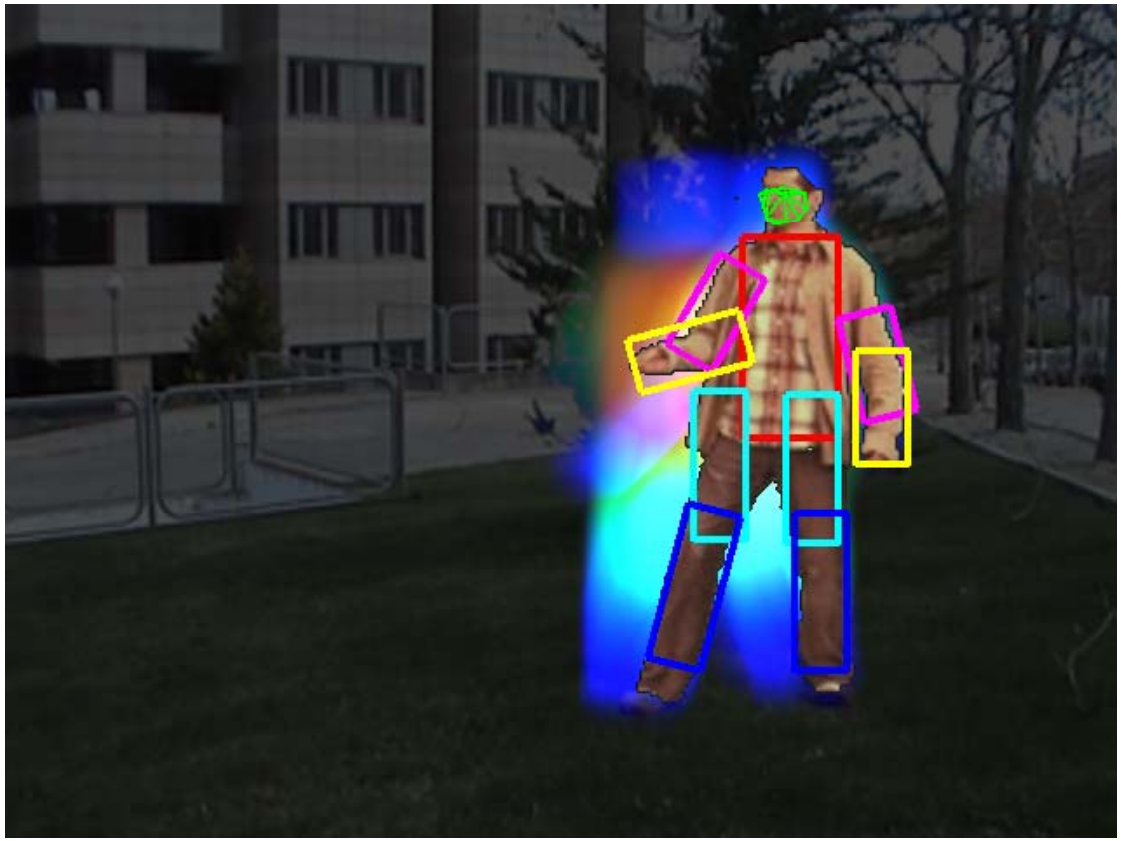
Validation: Face recovery

- Face recovery



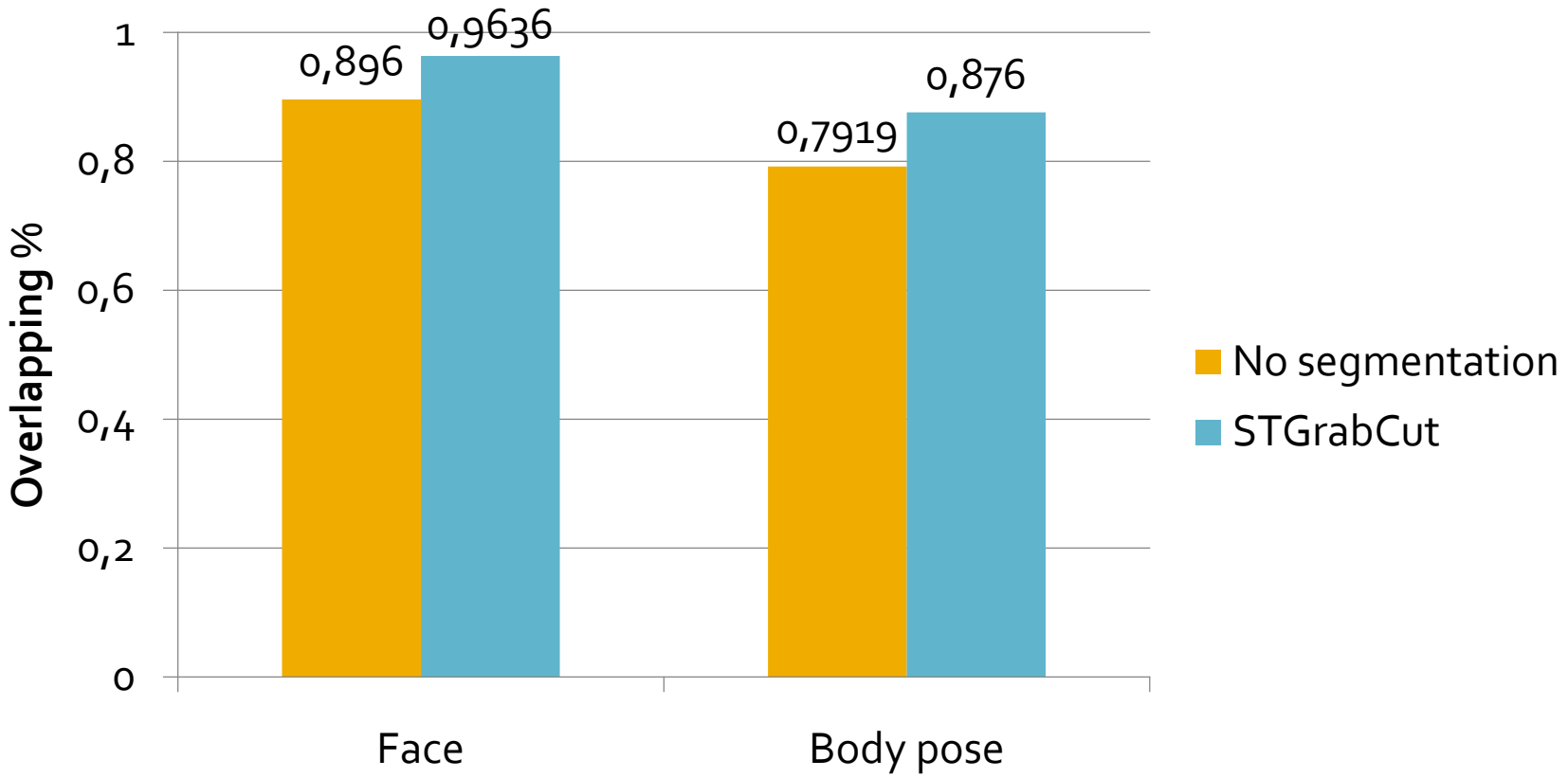
Validation

- Temporal joint body and face recovery



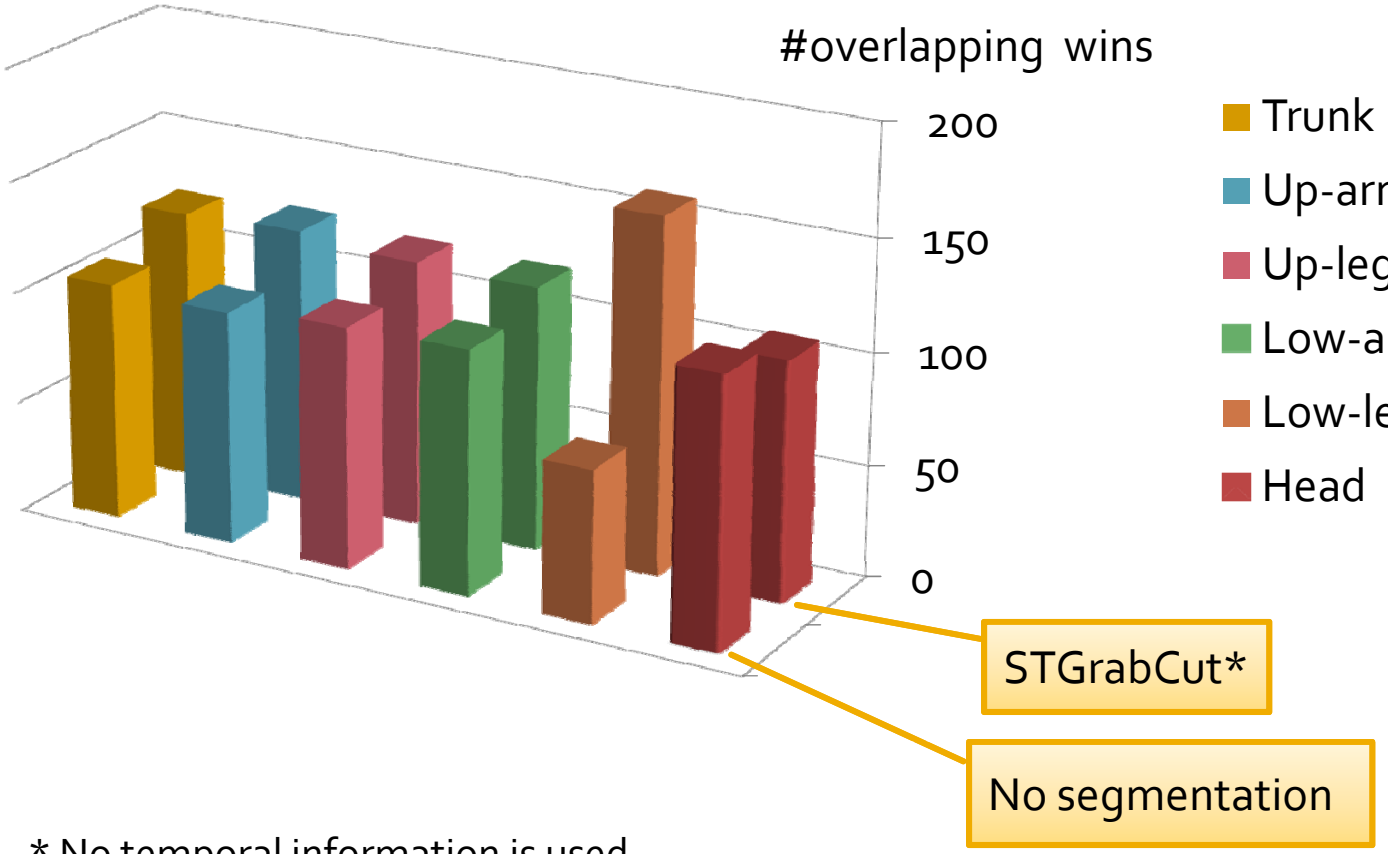
Validation

- Face and body pose recovery (CVSG)



Validation

- Body pose recovery



* No temporal information is used

Outline

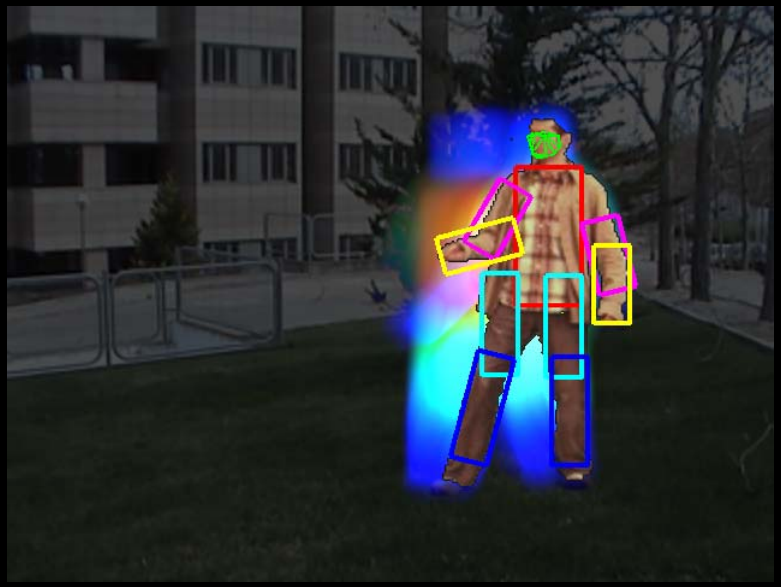
- Introduction
- Methodology
  - Spatio-Temporal GrabCut → Human segmentaton
  - Active Appearance models → Face fitting
  - Conditional Random Field → Limb recovery
- Validation
  - Public data sets
  - Human Body Limb data set
- **Conclusions and future work**

Conclusions

- Extension of GrabCut for human segmentation
  - Fully automatic method
  - Temporal coherence
  - Segmentation convexity problem
- Face recovery with temporal coherence
- New human body limb database
- Human segmentation helps to retrieve face and body pose

Future work

- Include temporal coherence inside graph cuts framework
  - Extended graph for image volumes
  - New temporal potential
- Improve segmentation using face and pose recovery feedback



Thank you!

Questions?

Antonio Hernández
Miguel Reyes
Sergio Escalera
Petia Radeva

CVC Centre de Visió per Computador

UNIVERSITAT DE BARCELONA