Detangling People: Individuating Multiple Close People and Their Body Parts via Region Assembly

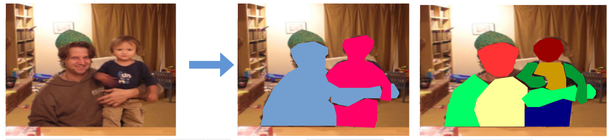
Hao Jiang
Boston College

Kristen Grauman
University of Texas at Austin

Contribution

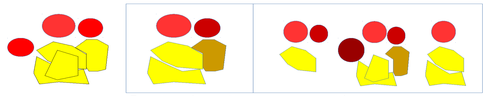
Problem: Parsing "entangled" people in crowded scenes
Our idea: Human segmentation as region assembly
Main steps:

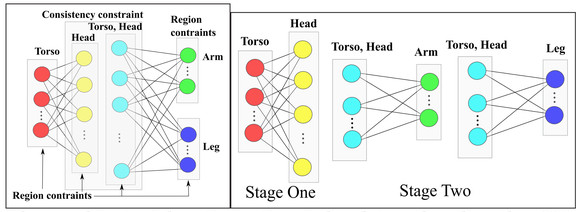
1. Detect regions and body part candidates
2. Optimize their assembly and individuate people



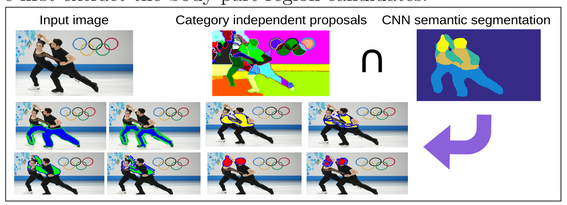
Model

Human segmentation can be formulated as a grouping problem.



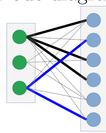


The combinatorial optimization is hard to solve directly. We decompose it into three easier ones.
We first extract the body part region candidates:

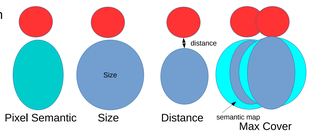


Optimizing Region Assembly

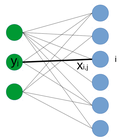
Node diagram



Person blob schematic

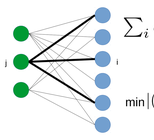


(1) Part-instance association cost



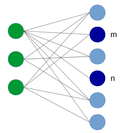
$$U(\mathcal{X}) = \sum_{i,j} c_{i,j} x_{i,j} + \xi \sum_j p_j y_j$$
$$y_j \geq x_{i,j}, \forall i$$
$$\sum_j x_{i,j} \leq 1$$

(2) Size constraint



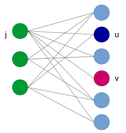
$$\sum_i a_i x_{i,j} \leq s_j^2 b$$
$$\min |(\sum_i a_i x_{i,j}/s_j^2) - l|$$

(3) Region overlap exclusion



$$z_i = \sum_j x_{i,j}$$
$$z_m + z_n \leq 1 \quad \text{if } q_{m,n} > \tau$$

(4) Color exclusion



$$x_{u,j} + x_{v,j} \leq 1 \quad \text{if } h_{u,v} > \varepsilon$$

(5) Max covering constraint

$$R = \sum_i r_i z_i$$
max(R)

The optimization has the following special structure:

$$\min_{x,y,e} \{g^T x + w^T y + \phi 1^T e\} \quad (1)$$
$$\text{s.t. } Ax \leq 1, Bx + Ce + Dy \leq f, \ e \geq 0, \ x, y \text{ are binary.}$$

Its Lagrangian relaxation is:

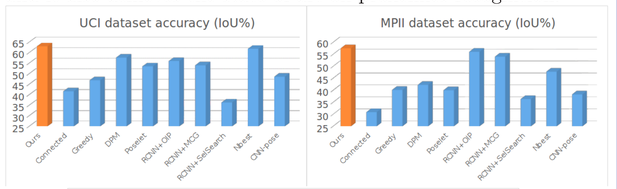
$$\max_{\nu} \min_{x,y,e} \{g^T x + w^T y + \phi 1^T e + \nu^T (Bx + Ce + Dy - f)\}$$
$$\text{s.t. } Ax \leq 1, 0 \leq e \leq M, \quad x, y \text{ are binary}, \nu \geq 0, \quad (2)$$

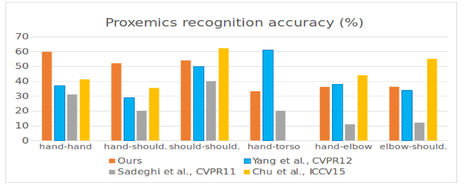
For each ν, it is efficiently solved by decomposing into:

$$[\text{P1}]: \min_x (g^T + \nu^T B)x, \text{ s.t. } Ax \leq 1, \ x \text{ is binary.} \quad (3)$$
$$[\text{P2}]: \min_y (w^T + \nu^T D)y, \text{ s.t. } y \text{ is binary.} \quad (4)$$
$$[\text{P3}]: \min_e (\phi 1^T + \nu^T C)e, \text{ s.t. } 0 \leq e \leq M. \quad (5)$$

Experimental Results

Our branch and bound method gives *global optimal body part assembly*. It gives results superior to different competing approaches, and it also achieves the state of art on proxemics recognition.





Comparison with people detector based approaches:

Comparison with approaches using pose estimation:

Experimental Results (Cont'd)

Sample results of our method:

Sample results for proxemics recognition:

Conclusion

Our novel method segments human instances and labels their body part regions. It is robust to complex human interactions, occlusions, and difficult poses, and it is rotation and scale invariant. Our results compare favorably to a wide array of alternative methods, and we improve the state of art on proxemics recognition.