Articulated Human Detection and Pose Estimation

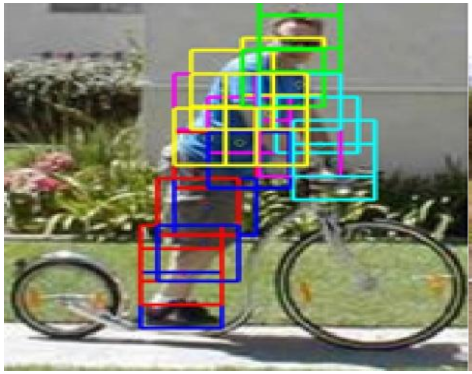
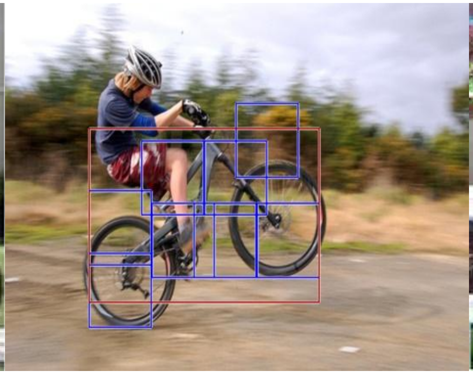
Anant Raj & Triya Bhattacharya

Mentor – Amitabha Mukerjee

- Introduction
- Problem Statement
- Previous related work
- Our Approach
- A brief about algorithm
- References

Introduction

- <u>Major Challenges</u>
  - Articulated Object detection
  - Human Detection and Tracking
  - Pose Estimation

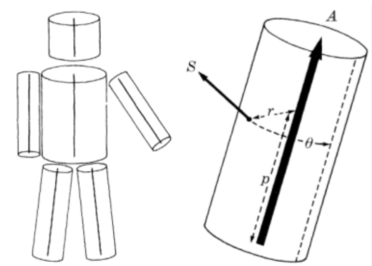
GOAL:

- To recover the pose of an articulated object which consist of joints and rigid parts(Human).

Previous Work

Part Representation

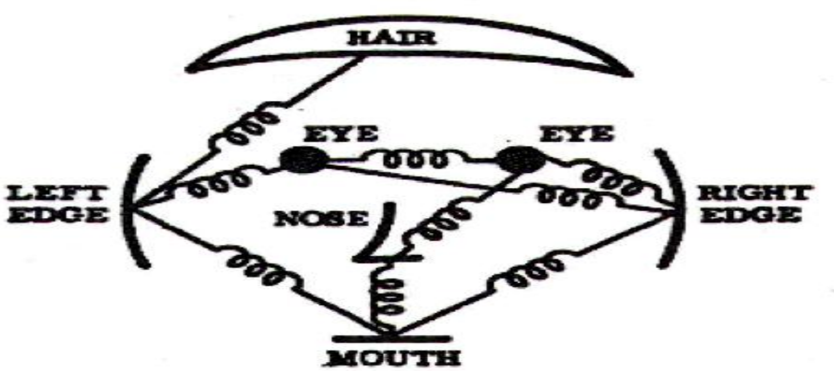
- Head, Torso, Arm, Leg
- Location, rotation, Scaling.

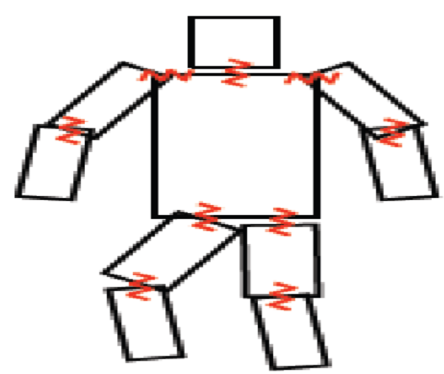


Marr & Nishihara 1978

Pictorial Structure

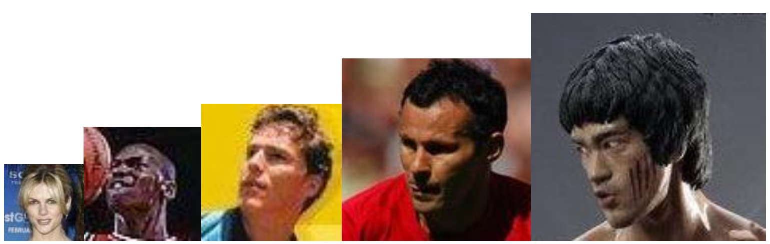
- Pairwise Spring.

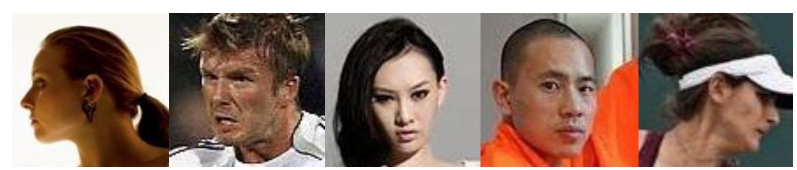




Problems:

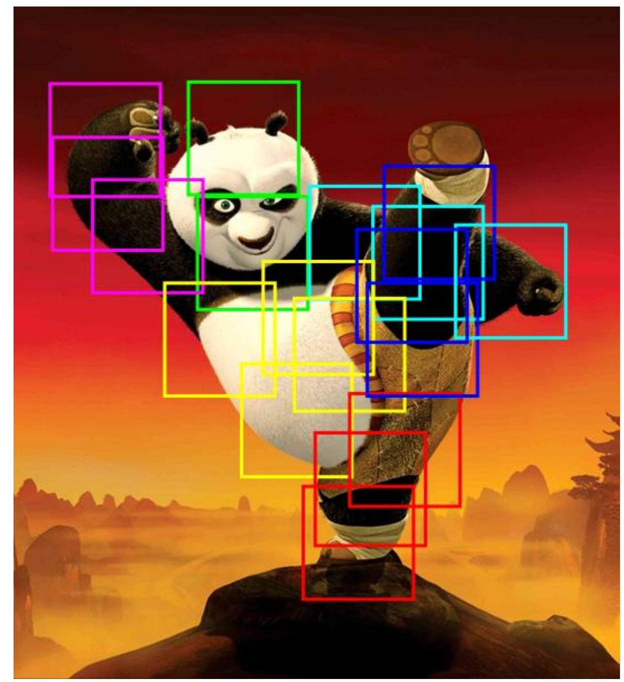
- Many degrees of freedom to be estimated.
- Limbs vary greatly in appearance.
- Need more no. of training images and also complications in inference.





That’s Why

- Mini Parts Model.
- It can approximate deformations.



Co-occurrence Model

- Compatibility function for part type

$$S(t) = \sum_{i \in V} b_i^{t_i} + \sum_{ij \in E} b_{ij}^{t_i, t_j}$$

  where parameter $b_i^{t_i}$ favors particular type assignment for part *i* *and* $\mathrm{b}_{\mathrm{ij}}^{\mathrm{t_i t_j}}$ favors particular co-occurrence of part type i and j.

- G(V,E) is a graph whose edges specify pair of parts having consistent relations.

Flexible Mixture of Parts.

- Total score associated with a configuration of part types and positions –

$$S(I, L, M) = \sum_{i \in V} \alpha_i^{m_i} \cdot \phi(I, l_i) + \sum_{ij \in E} \beta_{ij}^{m_i m_j} \cdot \psi(l_i, l_j) + S(M)$$

m_i: Mixture of part i

$\alpha_i^{m_i}$: Unary template for part i with mixture m_i

$\beta_{ij}^{m_i m_j}$: Pairwise springs between part i with mixture m_i and part j with mixture m_j

Inference and Learning

- Test Phase
  - Maximize the total score over L and M.
  - Dynamic programming starting from the leaf of graph G(V,E).
  - Given: - Image (I)
  - Need to compute - Part locations , part Mixture
  - Algorithm

    - $(L^*,M^*) = \arg\max (S(I,L,M))$

- Train Phase
  - Supervised Learning.
  - Given – Image and known location of the parts.
  - Need to learn –
    - Unary Templates α_i
    - Spatial Features β_{ij}
    - Co-occurrence $S(M)$

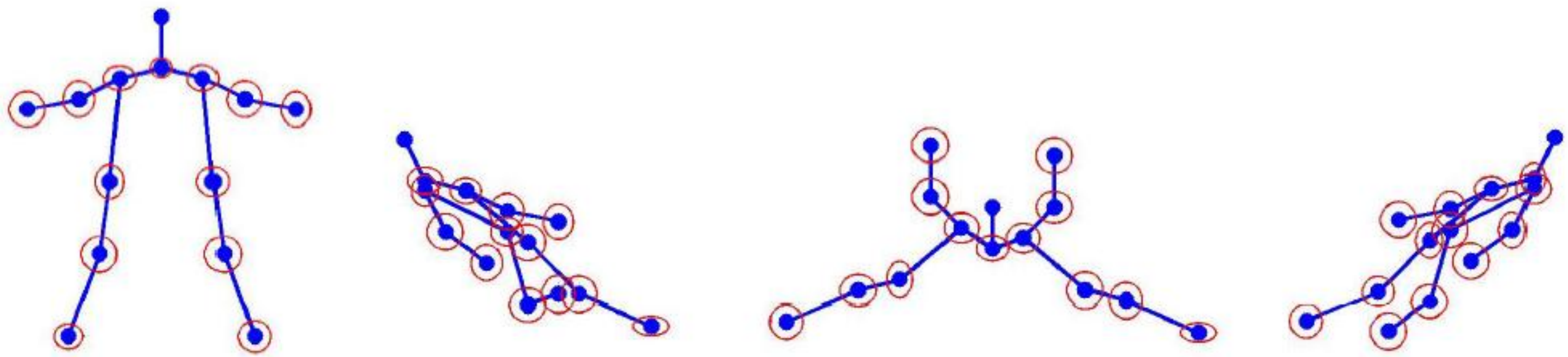
References :

1. http://phoenix.ics.uci.edu/software/pose/
2. Yang, Yi, and Deva Ramanan. "Articulated pose estimation with flexible mixtures-of-parts." *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011.
3. Felzenszwalb, Pedro F., and Daniel P. Huttenlocher. "Pictorial structures for object recognition." *International Journal of Computer Vision* 61.1 (2005): 55-79.

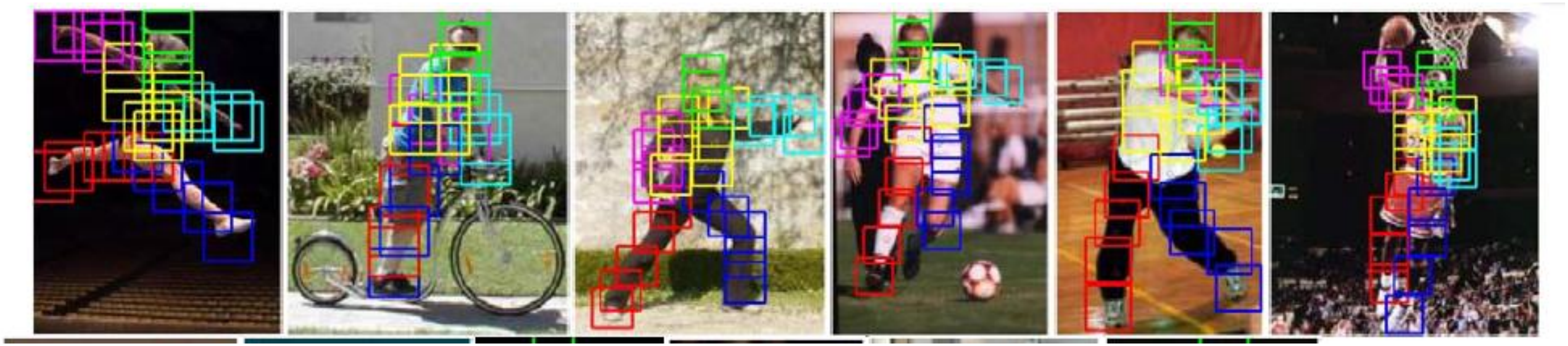
Questions !!

Thanks.

Experimental Results (as described in the paper)



- Visualization for 14 parts and 4 local mixtures, trained on the Parse dataset



- Examples where algorithm was successful

A Deeper Look:

- Assign $p_i = (x, y)$ for the pixel location of part i and t_i for the mixture component of part I
- Compatibility function that computes local and pairwise score
- Inference corresponds to maximizing score of part i over p and t
- Local score of part i computed by collecting messages from children of i using Dynamic programming
- Once messages sent from root (i=1) best score is computed
- Using supervised learning, a predictive model is generated, using scores >1 as positive examples, and <-1 as negative

- Score can be written in the form-

$$S(I, z) = \sum_{i \in V} \phi_i(I, z_i) + \sum_{ij \in E} \psi_{ij}(z_i, z_j),$$

$$\text{where} \quad \phi_i(I, z_i) = w_i^{t_i} \cdot \phi(I, l_i) + b_i^{t_i}$$

$$\psi_{ij}(z_i, z_j) = w_{ij}^{t_i, t_j} \cdot \psi(l_i - l_j) + b_{ij}^{t_i, t_j}$$

- Compute the message part i passes to its parent j by

$$\text{score}_i(z_i) = \phi_i(I, z_i) + \sum_{k \in \text{kids}(i)} m_k(z_i)$$

$$m_i(z_j) = \max_{z_i} \left[\text{score}_i(z_i) + \psi_{ij}(z_i, z_j) \right]$$

Where upper equ. Computes the local score and the lower computes possible orientation of part i and j.

- *Score S can be written in linear form*

  $S(I, z) = \beta \cdot \Phi(I, z)$.

- Learn a model like

$$
\begin{aligned}
& \arg \min_{w,\xi_n \geq 0} && \frac{1}{2}\beta \cdot \beta + C\sum_n \xi_n \\
& \text{s.t.} \quad \forall n \in \text{pos} && \beta \cdot \Phi(I_n, z_n) \geq 1 - \xi_n \\
& \quad \forall n \in \text{neg}, \forall z && \beta \cdot \Phi(I_n, z) \leq -1 + \xi_n
\end{aligned}
$$

- The above constraint states that positive examples should score better than 1 (the margin), while negative examples, for all configurations of part positions and types, should score less than -1.