Articulated Human Detection and Pose Estimation

Anant Raj (anantraj@iitk.ac.in), Triya Bhattacharya (triya@iitk.ac.in)

Supervisor: Dr. Amitabha Mukerjee

Dept. of Computer Science and Engineering

Problem Statement : [1]

The aim of the project is to estimete the pose of Articulated Human in static 2D images with flexible mixture of parts **[1].** The main idea behind this work is that “”mini part” model can approximate deformation” as explained in the paper **[1]**.

Motivtion and Related Work :

Articulated pose estimation is one of the core problems in object detection and Computer Vision. Any improvement in this area will lead to solve some challenging problem and will immediatelly impact in the area of Human-Machine interface and intelligent systems. That's why people have devoted a large amount of time in doing research to correctly estimate the articulated pose.

A very early and classic work in the area of object detection is to represent object in pictorial structural framework **[2].** After that there became a long tradition of using pictorial structure for human **[3],[4]**. In this approach the appearance of object is decomposed in local part templete and a geometric constraints is imposed on the joints. But still full body pose estimation can't be done effectively using this methos also because of the many degrees of freedom present in the huma body. Along with this the appearance of limbs is also continuously changing with clothing and angle of view. These difficulties complicate inference as one must typically search images with a large number of warped (rotated and foreshortened) templates **[1]**.

Method Chosen and Datasets [1]

/* The content of this section is fully taken from the paper [1]*/

The approach described in **[1]** is based on approximation of deformation as mini parts. Let us assume that l_i is the pixel location of part i in Image I and t_i is for the mixture component of part i. Compatibility function for part types is defined as a sum of local and pairwise score. For example, if part types correspond to orientations and part i and j are on the same rigid limb, then the pairwise parameters reflects the co-occurance of part I and j (Highly positive for consistent orientation t_i and t_j and negative for inconsistence one).

Full score associated with a configuration of part types and positions can be written as a sum of two parameteres which represents appearance and deformation respectively. Inference corresponds to maximizing the score over given parameters. The whole structure can be represented as a graphical model with K-node relational graph G(V,E) whose edge specify which pairs of parts are constrained to have consistent relations**[1]**. This model can force rigdity on a collection of parts. So in this tree structured graph the overall score can be maximized with dynamic programming starting from leaves and moving upstream.

Supervised learning algorithm is used to learn the parameters. Constraints are defined in such way that positive examples result in score greater that 1 and negative examples results in score less than -1. This form of learning is known as structural SVM **[1]**.

Image Parse dataset **[5]** and the Buffy Stickmen dataset [6], [7] is used for evaluating results. Along with this we will also be generating our own image sets with the help of camera for evaluation. They have also provide the source code on their web.

Extension of Work :

After achieving the basic task of pose estimation , if time permits then we will try to detect multiple people(2 people) in a complex image and the interaction between them.

References :

[1] Yang, Yi and Ramanan, Deva . Articulated pose estimation with flexible mixtures-of-parts

[2]M. Fischler and R. Elschlager, “The representation and matching of pictorial structures,” IEEE Transactions on Computers, vol. 100, no. 1, pp. 67–92, 1973.

[3]P. Felzenszwalb and D. Huttenlocher, “Pictorial structures for object recognition,” International Journal of Computer Vision, vol. 61, no. 1, pp. 55–79, 2005.

[4]Ronfard, Rémi, Cordelia Schmid, and Bill Triggs. "Learning to parse pictures of people." *Computer Vision—ECCV 2002* (2006): 700-714.

[5]D. Ramanan, “Learning to parse images of articulated bodies,” in Advances in Neural Information Processing System, 2007.

[6]V. Ferrari, M. Marin-Jimenez, and A. Zisserman, “Progressive search space reduction for human pose estimation,” in IEEE Conference on Computer Vision and Pattern Recognition, 2008.

[7]M. Eichner, M. Marin-Jimenez, A. Zisserman, and V. Ferrari, “2d articulated human pose estimation and retrieval in (almost) unconstrained still images,” International Journal of Computer Vision, vol. 99, no. 2, pp. 190–214, 2012.