Estimating Body Pose from A Single Image

Pedro F. Felzenszwalb

Artificial Intelligence Laboratory
Massachusetts Institue Of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: The problem is to estimate the body pose of a person from a single picture taken from an uncallibrated camera.

Motivation: The specific problem is motivated by numerous applications that fall in the general categories of human computer interaction, video surveillance, virtual reality, computer animation, and others. Another motivation is the development of general computer vision techniques for matching and recognition of articulated and flexible models.

Previous Work: There are a number of systems that tackle the specific problem defined here, the latest one being [2]. They use fairly generic machine learning techniques to find a mapping from image silhouettes to body pose. Our work takes a more geometric approach by modeling the body as an articulated model. There are a number of techniques for matching articulated models to images, see [1] for a brief overview.

Approach: We started with the techniques developed in [1]. Those techniques allow us to model an object as a collection of parts connected by joints. Our body model consists of rigid body parts (torso, head, limbs) and flexible revolute joints connecting the parts in a way to approximate the human skeleton. We use generic apperance models for the individual body parts. The characteristics of the joints connecting different body parts encode a prior model of "typical" body poses. This prior model is learned from examples (motion capture data available on the web). Figure 1 shows samples from the learned probability distribution over body poses. The algorithms in [1] allows us to efficiently find the best body pose for an image, in the sense that the pose is "typical" and each individual body part matches the image according to the apperance models.

Difficulty: One of the main challenges is to develop generic apperance models for the individual body parts. So far we have assumed the camera and the background are static. This allows us to use simple background subtraction to characterize image pixels that are likely to belong to the body. Our assumptions would fail if there were more than one person in the scene, or other moving objects around.

Future Work: One direction we want to pursue is to develop better apperance models for the individual body parts. Another goal is to make the geometric model richer by characterizing interactions between parts that are not directly connected by joints.

Research Support: This work is supported by a grant from NTT for the Variable Viewpoint Reality.

References:

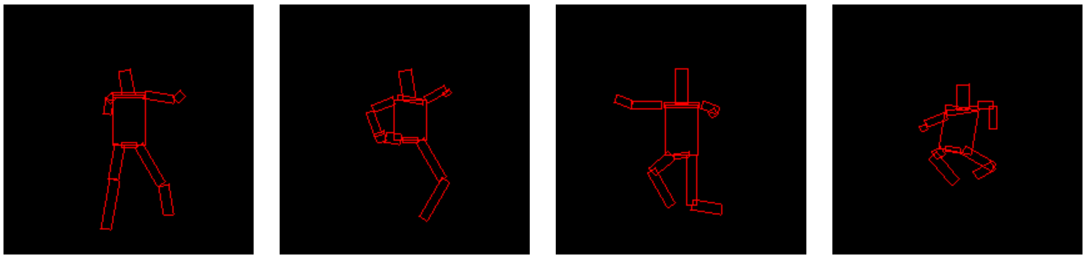

Figure 1: Sampling poses from the prior model. These are "typical body poses" to our system.

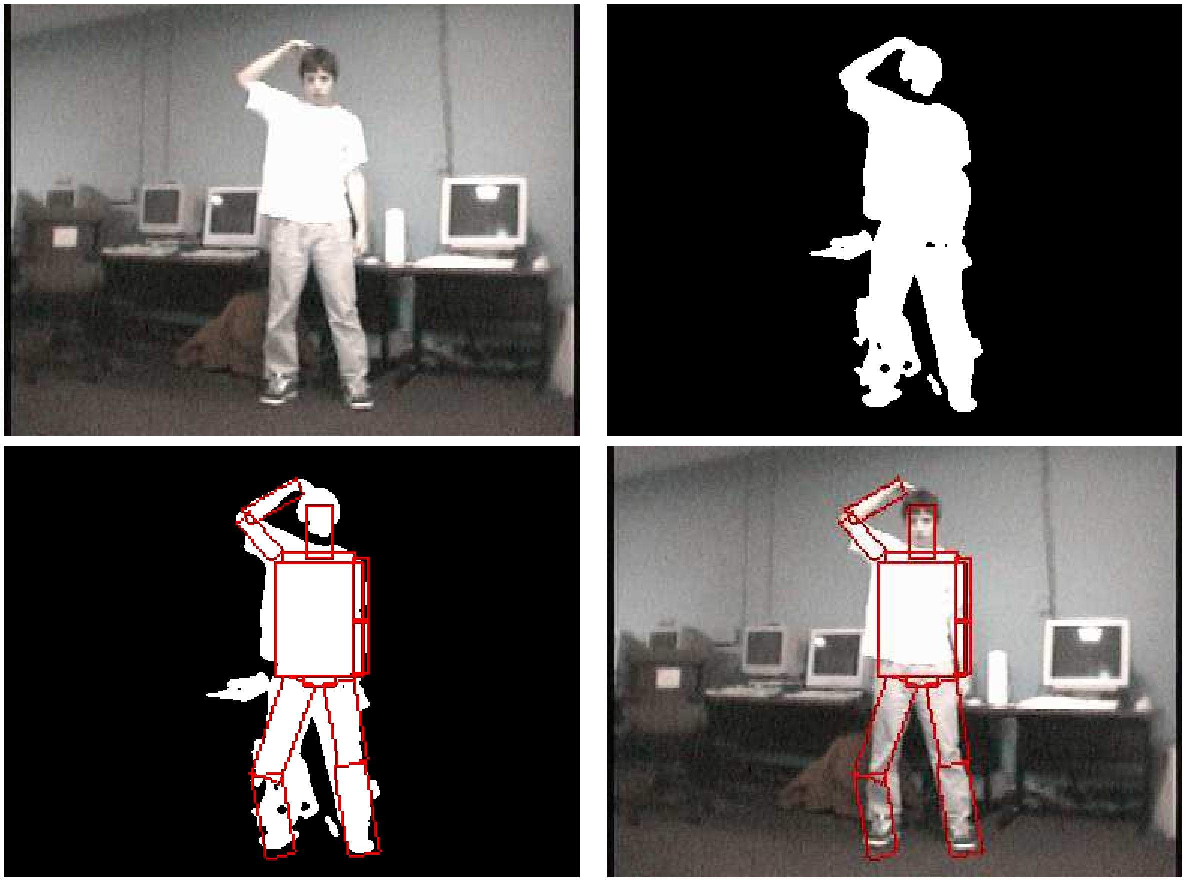

Figure 2: Matching results. The top left image is the original input. The top right image is the silhouette extracted by background subtraction. The remaining images show the estimated body pose.

[1] P. Felzenszwalb and D. Huttenlocher. Efficient Matching of Pictorial Structures. *CVPR*, 2000.

[2] R. Rosales and S. Sclaroff. Inferring Body Pose without Tracking Body Parts. *CVPR*, 2000.