



Human Layout Detection using Discrete Optimization Methods

Arpit Mittal¹, Andrew Zisserman¹ and Phil Torr¹¹
University of Oxford¹, Oxford Brookes University¹¹

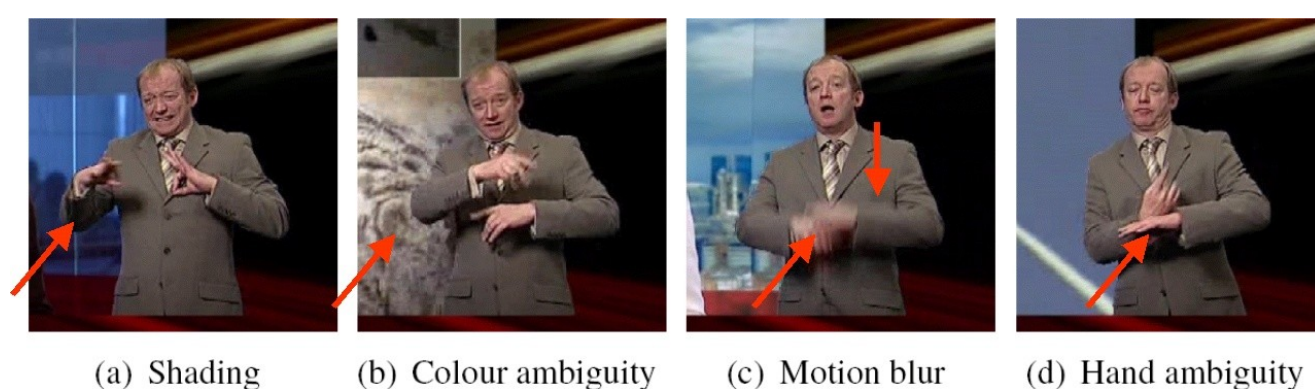
1. Objective

- Pose detection of human beings in images and video frames.
- Estimate the position of the head, torso, arms and hands.

2. Overview

Why is it difficult?

- Self-occlusions
- Motion Blur
- Varying Background
- Illumination Effects
- Vast number of possible configurations



3. Proposed Method

- Employ a generative model to perform layout detection.
- Time-efficient model fitting using Pictorial Structures to sample plausible layout configurations.
- Discrete Techniques to optimize upon possible configuration space

4. Mathematical Model

- Labelling for each part is associated with the following energy:

$$Q(f) = \sum_{a \in \mathcal{V}} \Theta_{a,f(a)} + \sum_{(a,b) \in \mathcal{E}} \Theta_{ab,f(a)f(b)} + b$$

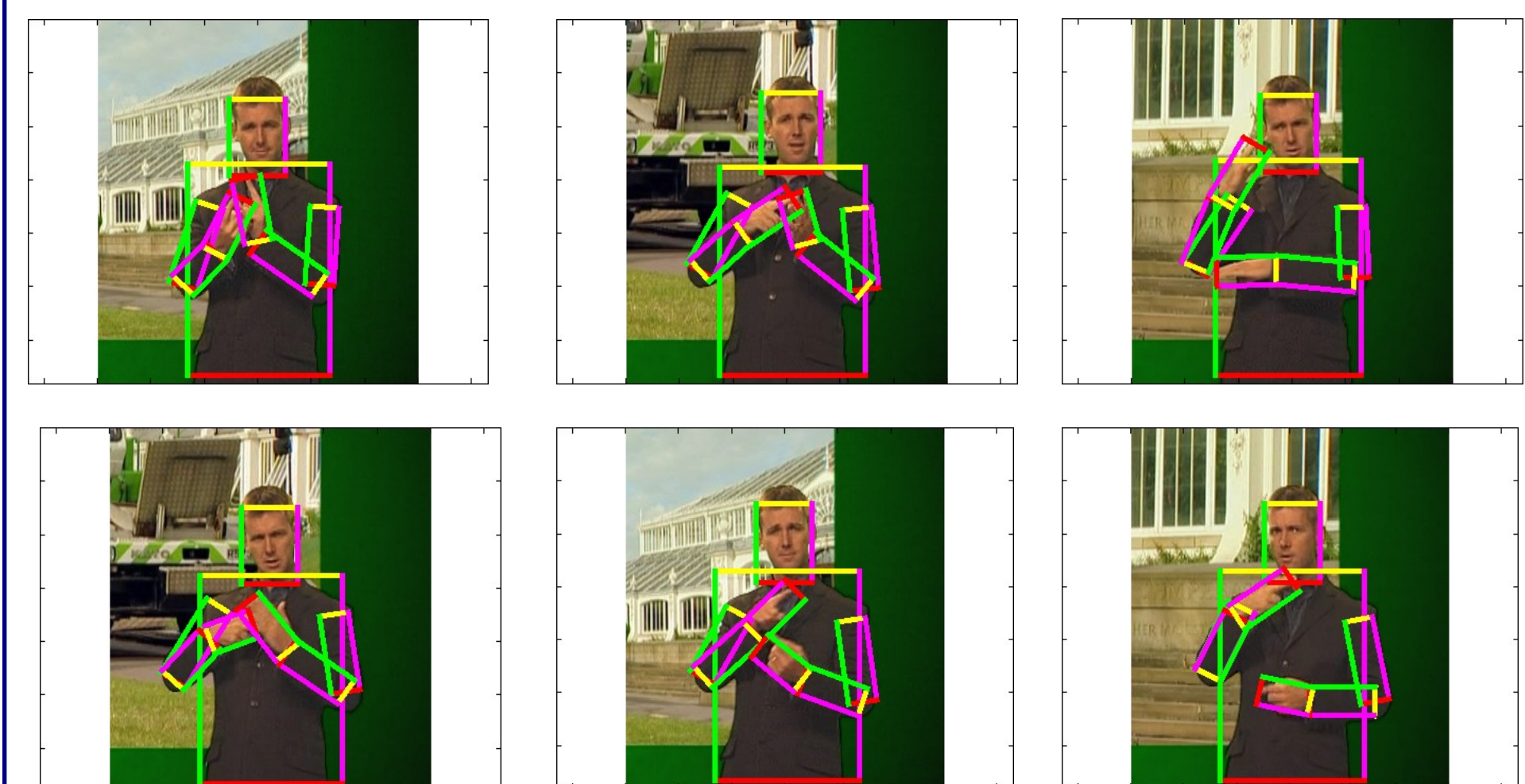
– Here $\Theta_{a,f(a)}$ is the unary term, $\Theta_{ab,f(a)f(b)}$ is the pairwise term and b is the offset

- For this work, we considered the form of potentials to be the following:

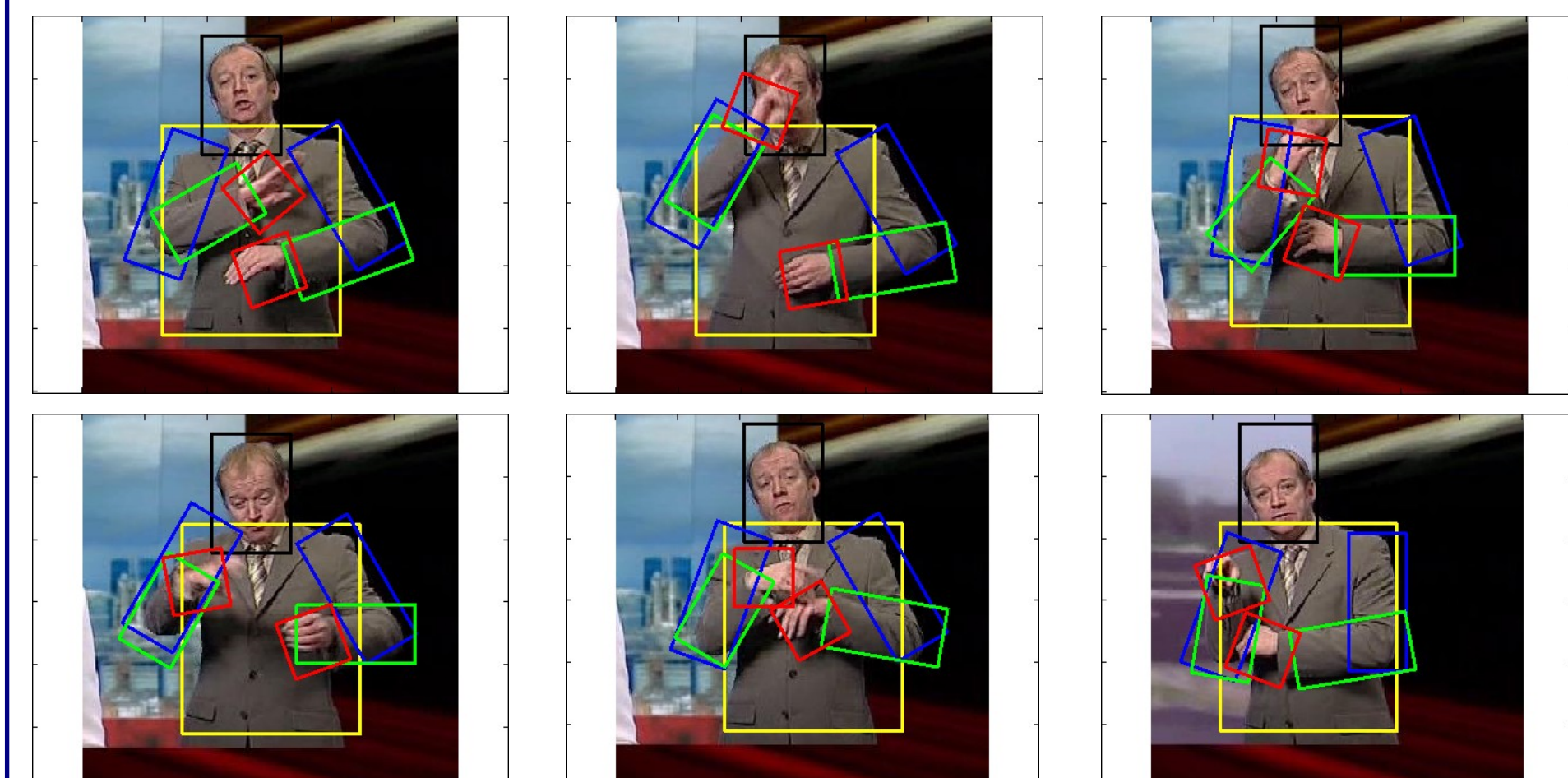
$$\Theta_{a,f(a)} = w_a^T \Phi_{a,f(a)}, \quad \Theta_{ab,f(a)f(b)} = w_{ab}^T \Phi_{ab,f(a)f(b)}$$

– Here $\Phi_{a,f(a)}$ is the feature vector for part a , computed using the image data (HOG and Color features) and $\Phi_{ab,f(a)f(b)}$ is the pairwise feature defined for valid configuration of parts a and b . Pairwise features obey pott's model.

5. Training Data



6. Sample Results



* On the "Signer" dataset (Ref: http://www.robots.ox.ac.uk/~vgg/data/sign_language/)