From Pictorial Structures to Deformable Structures

Silvia Zuffi[1] **Oren Freifeld**[2] **Michael J. Black**[3,1]
[1]Department of Computer Science and [2]Division of Applied Mathematics, Brown University, Providence, RI, USA
[3]Max Plank Institute for Intelligent Systems, Tübingen, Germany
zuffi@cs.brown.edu, freifeld@dam.brown.edu, black@is.mpg.de

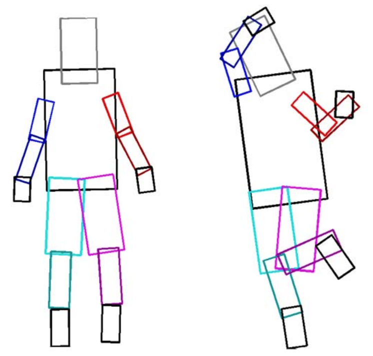
Results

We test the DS model for pose estimation on the "Buffy the Vampire Slayer" data set.

Method	Torso	Head	U. Arms	L. Arms	Total
Baseline (PS)	97.0	92.3	86.3	52.1	77.7
Our (NL)	99.2	97.9	90.9	10.4	66.6
Our (NS)	99.2	97.5	94.0	50.4	80.9
Our (DS)	99.6	99.2	94.7	62.8	85.6
Eichner *et al.*	98.7	97.9	82.8	59.8	80.1
CPS	100	96.2	95.3	63.0	85.5
Yang *et al.*	100	99.6	96.6	70.9	89.1

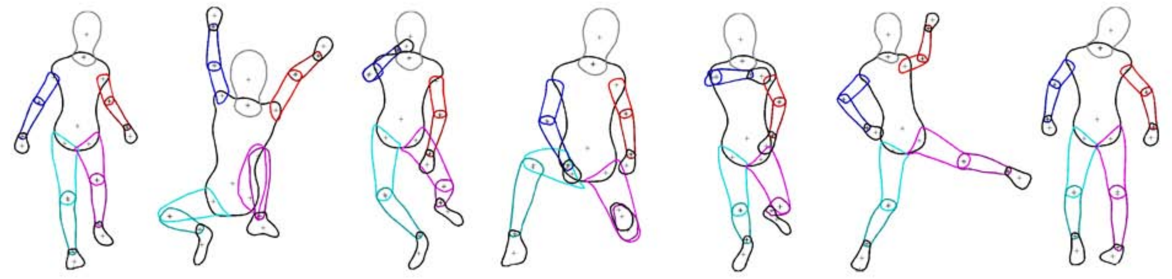
Table 1. PCP scores (see text) for our model without likelihood (NL), our model with a fixed shape (NS), and our full model (DS), with shape variation. PS is the implementation of [3]. We also compare with the current state of the art: CPS [30] and Yang *et al.* [1].

Introduction



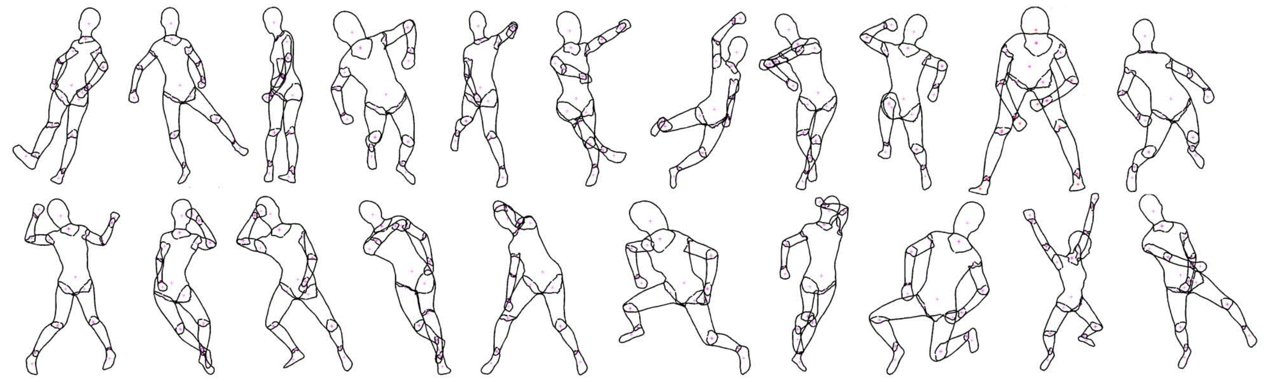
Pictorial Structures (PS) models do not represent shape deformations induced by pose.

Contribution: Deformable Structures (DS) are a generative model of 2D human shape that can represent pose-dependent shape deformations.



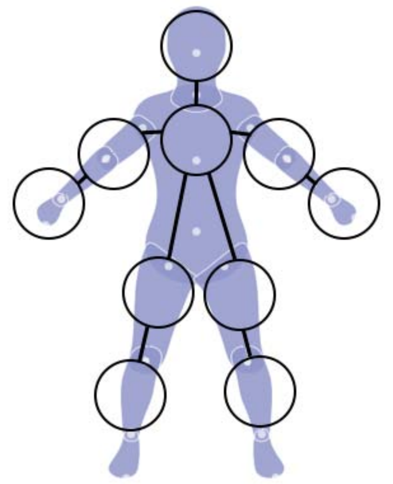
Training data

Training contours are derived by SCAPE, a realistic, parametric 3D model of articulated human shape, projecting random poses with random cameras.



The model is gender and person specific.

Model



$$p(L|I,\Theta) = \frac{1}{Z} \prod_{i=1..M} \phi_i(\mathbf{l}_i) \prod_{(i,j)\in E} \psi_{ij}(\mathbf{l}_i,\mathbf{l}_j|\Theta_{ij})$$

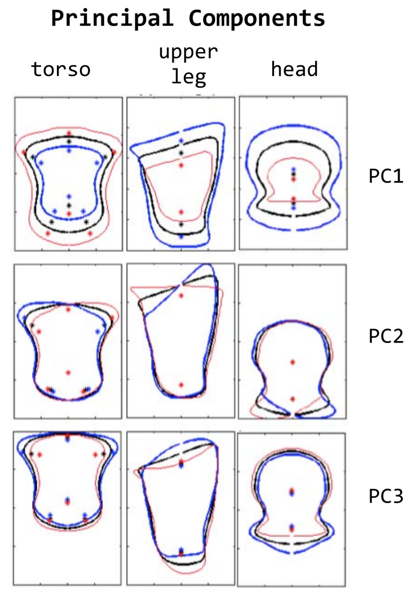
$$\mathbf{l}_i = (\mathbf{c}_i, \theta_i, \mathbf{z}_i)$$

$\mathbf{c}_i$ = location, θ_i = orientation, $\mathbf{z}_i$ = shape.

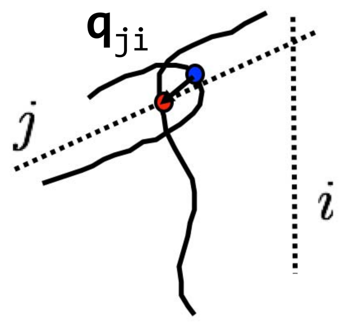
Shape representation

$$\begin{bmatrix} \mathbf{s}_i \\ \mathbf{p}_i \end{bmatrix} = \mathbf{B}_i \mathbf{z}_i + \mathbf{m}_i$$

$\mathbf{s}_i$ = contour points, $\mathbf{p}_i$ = joint points, $\mathbf{z}_i$ = PCA coefficients, $\mathbf{m}_i$ = mean shape, $\mathbf{B}_i$ = basis components.



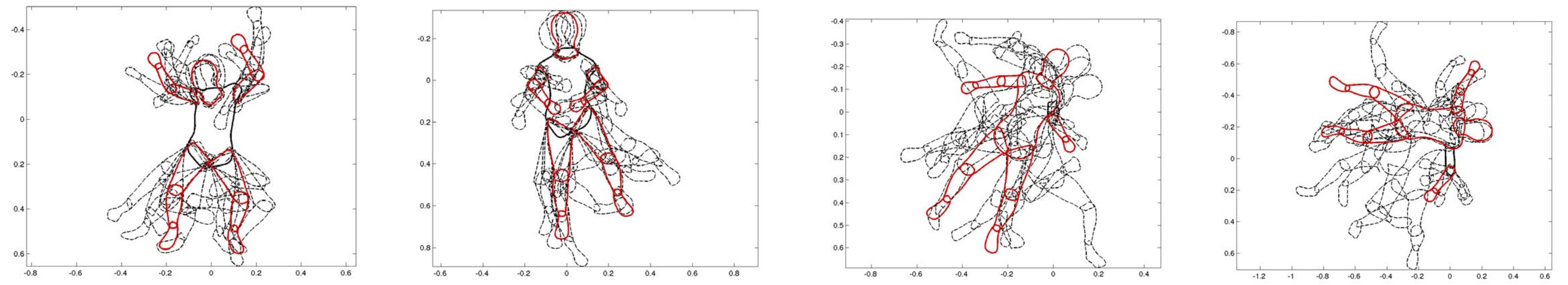
Probabilistic model



$$\psi_{ij}(\mathbf{l}_i,\mathbf{l}_j|\Theta_{ij}) = \mathcal{N}(\mathbf{z}_j, \sin(\theta_{ji}), \cos(\theta_{ji}), \mathbf{q}_{ji}, t_j, \mathbf{z}_i, t_i|\mu_{ij}, \Sigma_{ij})$$

θ_{ij} = relative angle, t_i and t_j = part lengths, $\mathbf{q}_{ij}$ = vector between joint points.

Sampling



Likelihood

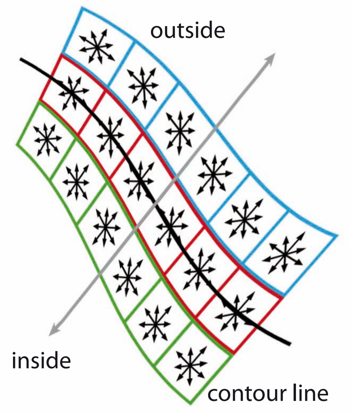
$$\phi_i(\mathbf{l}_i) = \phi_i^{\text{contour}}(\mathbf{l}_i)\phi_i^{\text{color}}(\mathbf{l}_i)$$

Contour-based likelihood

Images annotated with a DS-based annotation tool.

$$\phi_i^{\text{contour}}(\mathbf{l}_i) = \frac{1}{1+\exp(a_i f_i(h_i(\mathbf{l}_i)) + b_i))}$$

f_i = output of a linear SVM classifier, a_i and b_i = calibration parameters, h_i = set of HOG descriptors computed at contour locations and *steered* along the contour direction.



Color-based likelihood

$$\phi_i^{\text{color}}(\mathbf{l}_i) = \prod_{r \in M(\mathbf{l}_i)} hist(r)$$

$M(\mathbf{l}_i)$ = set of pixels of part i in state $\mathbf{l}_i$, hist = histogram of skin colors or upper body colors.

Inference

Due to the high dimensional variables and continuous state space, inference is performed with a particle-based version of Max-Product BP.