Bayesian Body Localization Using Mixture of Nonlinear Shape Models

Jiayong Zhang[1], Robert Collins[2], Yanxi Liu[1]

[1]Carnegie Mellon University [2]Pennsylvania State University

Abstract

We present a 2D model-based approach to localizing a human body in images viewed from arbitrary and unknown angles. The central component is a statistical shape representation of the nonrigid and articulated body contours, where a nonlinear deformation is decomposed based on the concept of parts. Several image cues are combined to relate the body configuration to the observed image, with self-occlusion explicitly treated. To accommodate large viewpoint changes, a mixture of view-dependent models is employed. Inference is done by direct sampling of the posterior mixture, using Sequential Monte Carlo (SMC) simulation enhanced with annealing and kernel move.

Features

- ✓ Handle subjects with large shape differences
- ✓ Handle arbitrary and unknown view angle
- ✓ Label individual limb contours (rather than a closed curve of the body)
- ✓ Encode both shape and pose
- ✓ Does not require preselection of a "correct" viewpoint
- ✓ Computationally independent of the number of view-dependent models

Assumptions

✓Torso approximately parallel to the imaging plane
✓ Weak perspective ✓ Body within field of view
✓No external occlusion ✓ Background model given

Motivation

What does "body localization" mean?

✓Detect all human figures ✓ Find limb shapes and positions

Why does it matter?

- ✓ Fundamental in computer vision (Marr et al, 1978)(O'Rourke et al, 1980)(Hogg, 1983). . .
- ✓ Applications in surveillance, HCI, entertainment…

Approach

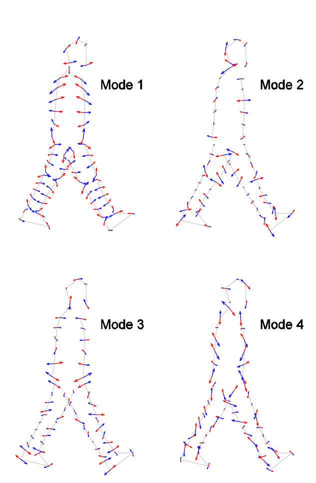
Model-based, 2D, Top-down, Static

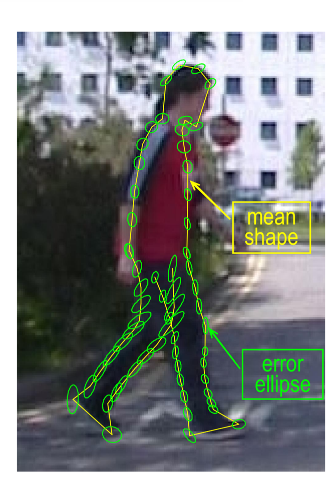
Bayesian Formulation

- ✓ Impose strong constraints from prior knowledge of possible shapes/poses
- ✓ Combine multiple image cues in a robust fashion

$$p(\Omega) \quad + \quad p(\mathcal{I}|\Omega) \quad \Rightarrow \quad p(\Omega|\mathcal{I})$$

Shape Prior **Image Likelihood** **Posterior Distrib.**





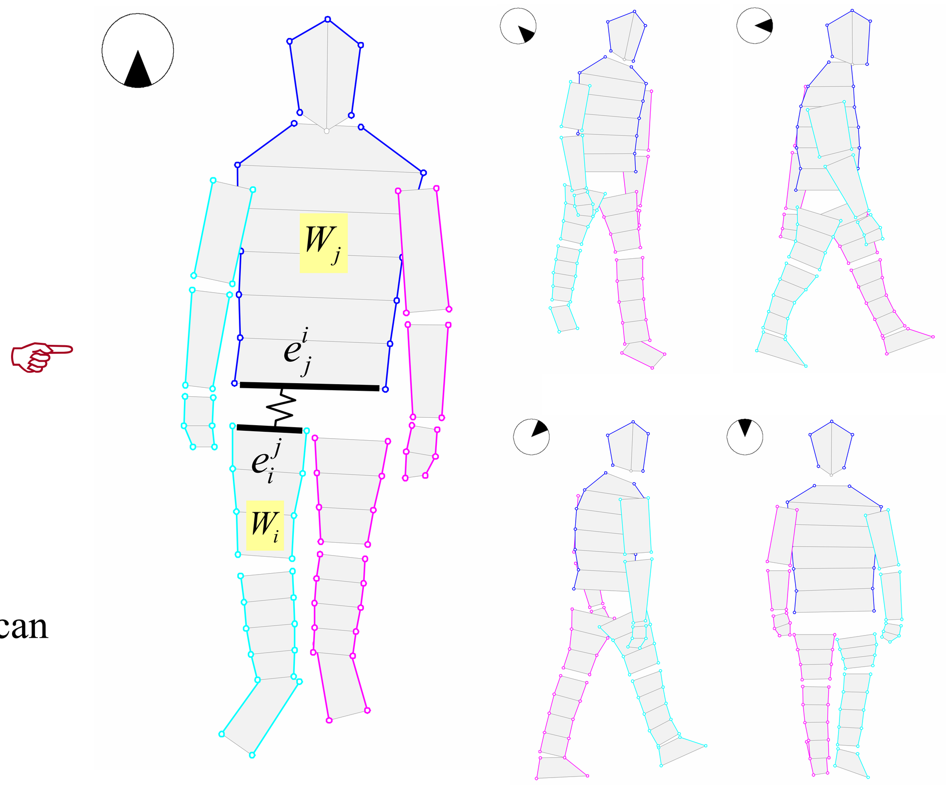
Mixture Shape Model

$$p(\Omega|\mathcal{I}) \sim \sum_{\chi} p(\mathcal{I}|\Omega,\chi)\, p(\Omega|\chi)\, p(\chi)$$

viewpoint index

Topologies of five basic component models

- ✓ Landmarks are grouped into parts with depth order.
- ✓ Part W_i is virtually attached to part W_j through two edges (e_i^j, e_j^i), which constitute a flexible joint.
- ✓ A fixed landmark ordering is specified so that the shape can be traversed by growing one quadrangle at a time.
- ✓ Eight component models uniformly distributed in [0,2π].



Decompose Prior by Parts

$$p(\mathbf{v}_{1:K}) = p(W_1)\prod_{(i,j)\in\mathcal{J}} \underbrace{p(\mathbf{e}_i^j|\mathbf{e}_j^i)}_{p(x_i,y_i)p(\rho_i)p(\theta_i)}\underbrace{p(W_i|\mathbf{e}_i^j)}_{p(\mathbf{r}_{i,:})}$$

Procrustes residuals

Dependencies among parts

- ✓ Continuity at joints
- ✓ Relative lengths of limbs

Correlation between $\boldsymbol{r}_{i,:}$ and $||e_i^j||$

Introduce conditioning variable $\gamma_i^j = \|\mathbf{e}_i^j\|/l_1$

$$p(\mathbf{v}_{1:K}) \propto \prod_{(i,j)\in\mathcal{J}} p(x_i,y_i|\gamma_j^i)p(\rho_i|\gamma_j^i)p(\theta_i) \prod_k p(\mathbf{r}_{i,k}|\mathbf{r}_{i,1:k-1},\gamma_i^j).$$

Decompose Likelihood by Clusters

$$p(\mathcal{I}|\Omega) \propto \prod_t \frac{p(\mathcal{I}_{R_t}|FG)}{p(\mathcal{I}_{R_t}|BG)} = \prod_t \phi(\mathbf{v}_{Q_t})$$

- ✓ Inhomogeneous model: $\phi(\mathbf{v}_{Q_t}) \Longrightarrow \phi(\mathbf{v}_{Q_t}|\ell_{Q_t})$
- ✓ Correlation: $\phi(\mathbf{v}_{Q_t}) \Longrightarrow \prod_{C\in\mathfrak{C}_t}\phi(\mathbf{v}_C)$
- ✓ Self-occlusion: $\psi(\mathbf{v}_C, \mathbf{v}_{\mathfrak{C}_{1:t-1}}) = \dfrac{\phi(\mathbf{v}_C,\mathbf{v}_{C'})}{\phi(\mathbf{v}_C)\phi(\mathbf{v}_{C'})}$
- ✓ Non-generative model: $\phi(\mathbf{v}_C) = \dfrac{p(\mathcal{F}_{R_C}|FG)}{p(\mathcal{F}_{R_C}|BG)}$
- ✓ Multiple cues: $z \in$ {edge gradient, silhouette, skin color, region similarity}

$$p(\mathcal{I}|\Omega) \propto \prod_t \prod_z \prod_{C\in\mathfrak{C}_t^z} \phi^z(\mathbf{v}_C|\ell_C)\,\psi^z(\mathbf{v}_C,\mathbf{v}_{\mathfrak{C}^z_{1:t-1}})$$

Inference

Sequential structure on Ω

⇨ Expand Ω and collect I incrementally

⇨ Search via Sequential Monte Carlo (SMC)

- ✓ a.k.a. particle filters, condensation
- ✓ probabilistic version of beam search

$$\Gamma_t = \begin{cases} p(x_i,y_i|\gamma_j^i,\chi)p(\rho_i|\gamma_j^i,\chi)p(\theta_i|\chi) & Q_t \text{ is joint} \\ p(\mathbf{r}_{i,k-1:k}|\mathbf{r}_{i,1:k-2},\gamma_i^j,\chi) & \text{otherwise} \end{cases}$$

Proposal

$$p(\chi,\Omega|\mathcal{I}) \propto p(\chi)\prod_t \Gamma_t \cdot \Phi_t$$

Importance Weight

$$\Phi_t = \prod_z \prod_{C\in\mathfrak{C}_t^z} \phi^z(\mathbf{v}_C|\ell_C,\chi)\,\psi^z(\mathbf{v}_C,\mathbf{v}_{\mathfrak{C}^z_{1:t-1}}|\chi)$$

Multiple Models

Common strategies

✓ Fit before select ✓ Select before fit

Our approach

- ✓ Directly sample posterior mixture ✓ Output sample representation of the posterior mixture $\left\{\chi^{(i)}, \mathbf{v}_{0:K}^{(i)}\right\}_{i=1}^N$
- ✓ Parallel search with dynamic resource allocation

Annealing: Gradually increase peakiness of likelihood term to avoid being trapped in local maxima during early stage of search, e.g.

$$\ln\phi(t) = \xi_t \ln\phi = \frac{t}{T}\ln\phi$$

MCMC: Move each particle once after every resampling, using Metropolis update to avoid sample attrition

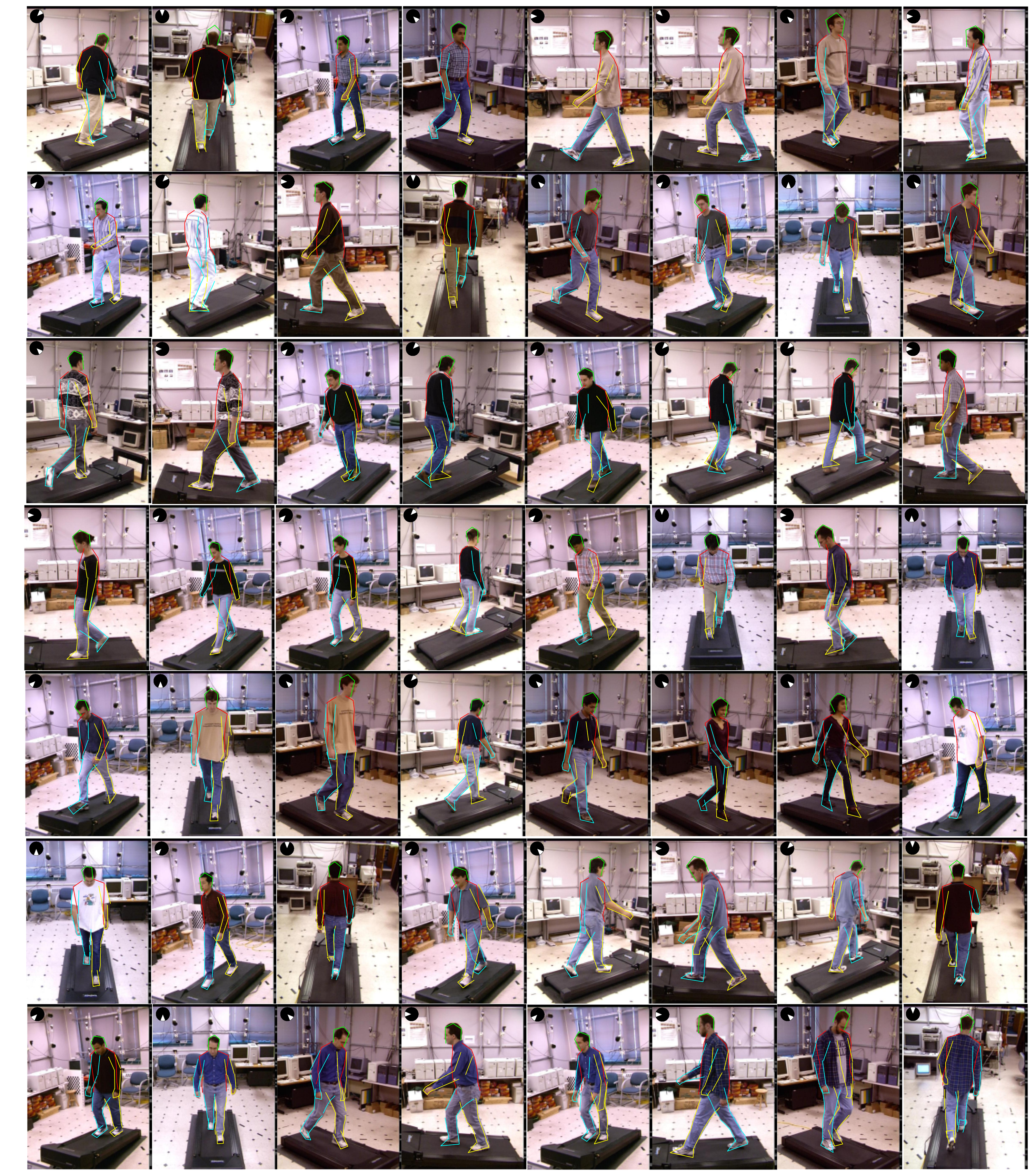
- ✓ Propose $\tilde{\mathbf{v}}_{0:2t}^{(i)} \sim N(\mathbf{v}_{0:2t}^{(i)}, \eta_t\Sigma_t)$
- ✓ Accept with prob $\min(1, p(\chi^{(i)},\tilde{\mathbf{v}}_{0:2t}^{(i)}|\mathcal{I})/p(\chi^{(i)},\mathbf{v}_{0:2t}^{(i)}|\mathcal{I}))$

Preliminary Results

Collect training data by interactive tracking

✓ CMU MoBo Database ✓ 25 subjects × 6 views × 50 frames ✓ slow-walk

Test I ✓ 300 random frames ✓ incline-walk & fast-walk



Test II ✓ 175 frames ✓ walk in a circle