# Learning to Recover 3D Human Pose from Silhouettes

Ankur Agarwal, Bill Triggs

# Learning to Reconstruct 3D Human Pose and Motion from Silhouettes

**Ankur Agarwal** and **Bill Triggs**
GRAVIR-CNRS-INRIA, 655 avenue de l'Europe, 38330 Montbonnot, France.
*Ankur.Agarwal@inrialpes.fr, Bill.Triggs@inrialpes.fr*

We will describe our ongoing work on learning-based methods for recovering 3D human body pose and motion from single images and from monocular image sequences. The methods work directly with raw image observations and require neither an explicit 3D body model nor a prior labelling of body parts in the image. Instead, they recover the body pose or motion by direct nonlinear regression against shape descriptors extracted automatically from image silhouettes or contours. For improved resistance to segmentation errors and occlusions, we use a robust shape representation: histograms of locally-supported shape-contexts descriptors. The image description is thus related to (generalized / robustified versions of) Brand's 'Shadow Puppetry' [1], Mori & Malik's shape context method [2] and Shakhnarovich *et al*'s contour based method [3].

We regress the current pose (body joint angles) against both silhouette shape and (in the tracking-based schemes) the previous 1–2 poses. The regression is nontrivial owing to high dimensionality, sparse training data, and the fact that recovering pose from monocular image observations is inherently multi-valued owing to pervasive kinematic ambiguities. For tracking, we evaluated several different regression dependency structures designed to reduce these reconstruction ambiguities while capturing the dynamics, observations and the correcting effect of observation updates (all nonlinear and unknown a priori).

We tested a number of different regression methods on these problems, including regularized least squares, Support Vector Regression and Relevance Vector Machine (RVM) regression [4], over both linear and kernel bases. In general the kernelized methods do best and the RVM framework provides much sparser regressors without compromising performance. But linear least squares (over our very nonlinear shape description) also performs surprisingly well. If time permits, we will also sketch our novel scalable continuation-based RVM training algorithm.

The methods are trained using real human motion capture data, to ensure that they capture both the global structure and the fine details of human motion. However to improve model coverage and make the most of the limited amount of training data available, we currently re-synthesize the corresponding training images from a range of different viewpoints.

We have tested our models both quantitatively on independently captured test sequences and qualitatively on videos of typical human motions. On the test sequences, we are currently getting mean angular errors of about 6–7 degrees — a factor of about 3 better than the current state of the art for the much simpler upper-body-only problem.

**Keywords:** visual learning, human motion, nonlinear regression, kernel methods.

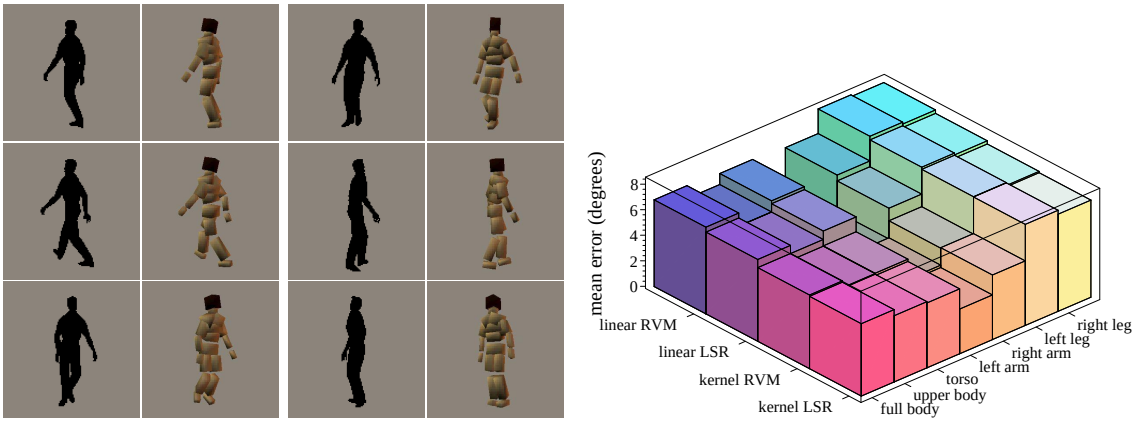


Figure 1: Some sample pose reconstructions for a test sequence in which the subject walks in a decreasing spiral, and a summary of the accuracy of several of our static pose regressors on this sequence, for different combinations of body parts.

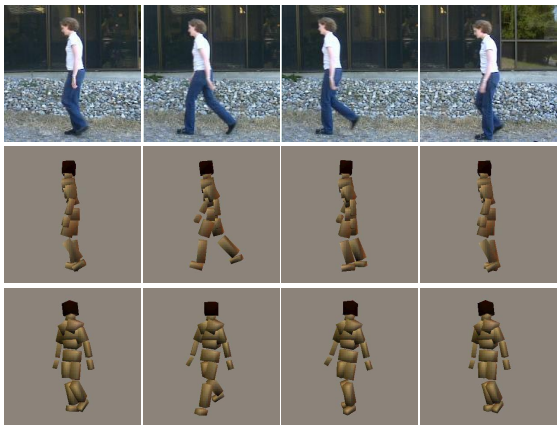

Figure 2: 3D poses reconstructed from a real test sequence (from www.nada.kth.se/∼hedvig/data.html), viewed from the original viewpoint and a new one. The last two columns illustrate limitations of our current system. In the third column, a noisy silhouette causes slight mis-estimation of the lower right leg, while the final column the reconstruction is incorrect owing to the left-leg right-leg ambiguity in the silhouette.

[1] M. Brand. Shadow puppetry. In *Int. Conf. Computer Vision*, pages 1237–1244, 1999.

[2] G. Mori and J. Malik. Estimating human body configurations using shape context matching. In *European Conf. Computer Vision*, volume 3, pages 666–680, 2002.

[3] G. Shakhnarovich, P. Viola, and T. Darrell. Fast pose estimation with parameter sensitive hashing. In *Int. Conf. Computer Vision*, pages 750–757, 2003.

[4] M. Tipping. Sparse bayesian learning and the relevance vector machine. *J. Machine Learning Research*, 1:211–244, 2001.