Local Probabilistic Regression for Activity-Independent Human Pose Inference

Raquel Urtasun and Trevor Darrell

Massachusetts Institute of Technology, Cambridge, MA 02139 USA

Learning a mapping from visual observation to articulated body configuration is the foundation of discriminative approaches to pose estimation; such methods (Agarwal & Triggs, 2006; Sminchisescu et al., 2005) have recently become popular due to their ability to estimate pose from a single image without initialization. We are interested in the discriminative inference of arbitrary poses without restriction to a relatively limited set of predefined activities, e.g., running or walking, and wish to have a method which can perform inference efficiently enough to provide pose estimates at interactive rates (i.e. near-real time). Learning such a transformation is extremely challenging, due to the multimodality of the mapping, the high dimensionality of the input and output spaces, and the fact that activity-independent pose mappings have considerable variability and therefore require very large training sets to be accurately defined.

In this paper we develop a method to learn a complex appearance-to-pose mapping for arbitrary motions using probabilistic regression. We take advantage of Gaussian Process (GP) models, which offer a general framework for probabilistic regression and have been shown to generalize well when the training data are few in number (Rasmussen & Williams, 2006; Urtasun et al., 2005). However, current GP models are limited in their ability to handle large training sets, allowing at most a few thousand training examples (Lawrence et al., 2003; Quiñonero-Candela & Rasmussen, 2005; Snelson & Ghahramani,). Also, in their standard form, GP models do not directly handle multimodality, and assume a single set of hyperparameters is sufficient to model the distribution of the data. Adapting the models locally is critical for human pose estimation since the training data density, noise levels and/or smoothness may vary considerably across the pose space.

We propose a new sparsification technique for Gaussian Processes, where local regressors are defined online for each test point. Local neighborhoods are very small, so training and inference are efficient. The use of a GP framework offers accurate probabilistic pose estimates from small neighborhoods, and naturally defines a redundancy criteria for pruning. Our method's computational complexity and memory requirements are dramatically reduced when compared to classic GP inference: inference is very fast with large databases of hundreds of thousands of examples. By using an online strategy our technique adapts to local regions of the space and does not suffer from the boundary problems that can affect static sparsification techniques or offline mixture models. Our method handles multimodality by forming online mixture components which are local both in terms of appearance and pose.

Our method can determine when training examples are redundant given the rest of the database, and use this criteria for pruning. Finally, we show accurate pose estimation results on both synthetic (Poser) and real images (Humaneva) of hands and whole body poses, using a variety of input feature types, with databases ranging from 10^2 to 10^5 examples.

Bibliography

Agarwal, A., & Triggs, B. (2006). Recovering 3D human pose from monocular images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*, 44–58.

Lawrence, N. D., Seeger, M., & Herbrich, R. (2003). Fast sparse Gaussian process methods: The informative vector machine. In *Neural information processing systems*, 609–616. Cambridge, MA: MIT Press.

Quiñonero-Candela, J., & Rasmussen, C. E. (2005). A unifying view of sparse approximate gaussian process regression. *Journal of Machine Learning Research*, *6*, 1939–1959.

Rasmussen, C. E., & Williams, C. K. (2006). *Gaussian process for machine learning*. MIT Press.

Sminchisescu, C., Kanaujia, A., Li, Z., & Metaxas, D. (2005). Discriminative Density Propagation for 3D Human Motion Estimation. *Conference on Computer Vision and Pattern Recognition* (pp. 390–397). San Diego, CA.

Snelson, E., & Ghahramani, Z. Sparse gaussian processes using pseudo-inputs. *Neural Information Processing Systems*.

Urtasun, R., Fleet, D. J., Hertzman, A., & Fua, P. (2005). Priors for people tracking from small training sets. *International Conference on Computer Vision* (pp. 403–410). Beijing, China.