Predicting Dynamical Evolution of Human Activities from a Single Image

Suhas Lohit†, Ankan Bansal*, Nitesh Shroff#, Jaishanker Pillai$, Pavan Turaga† and Rama Chellappa*

ARIZONA STATE UNIVERSITY

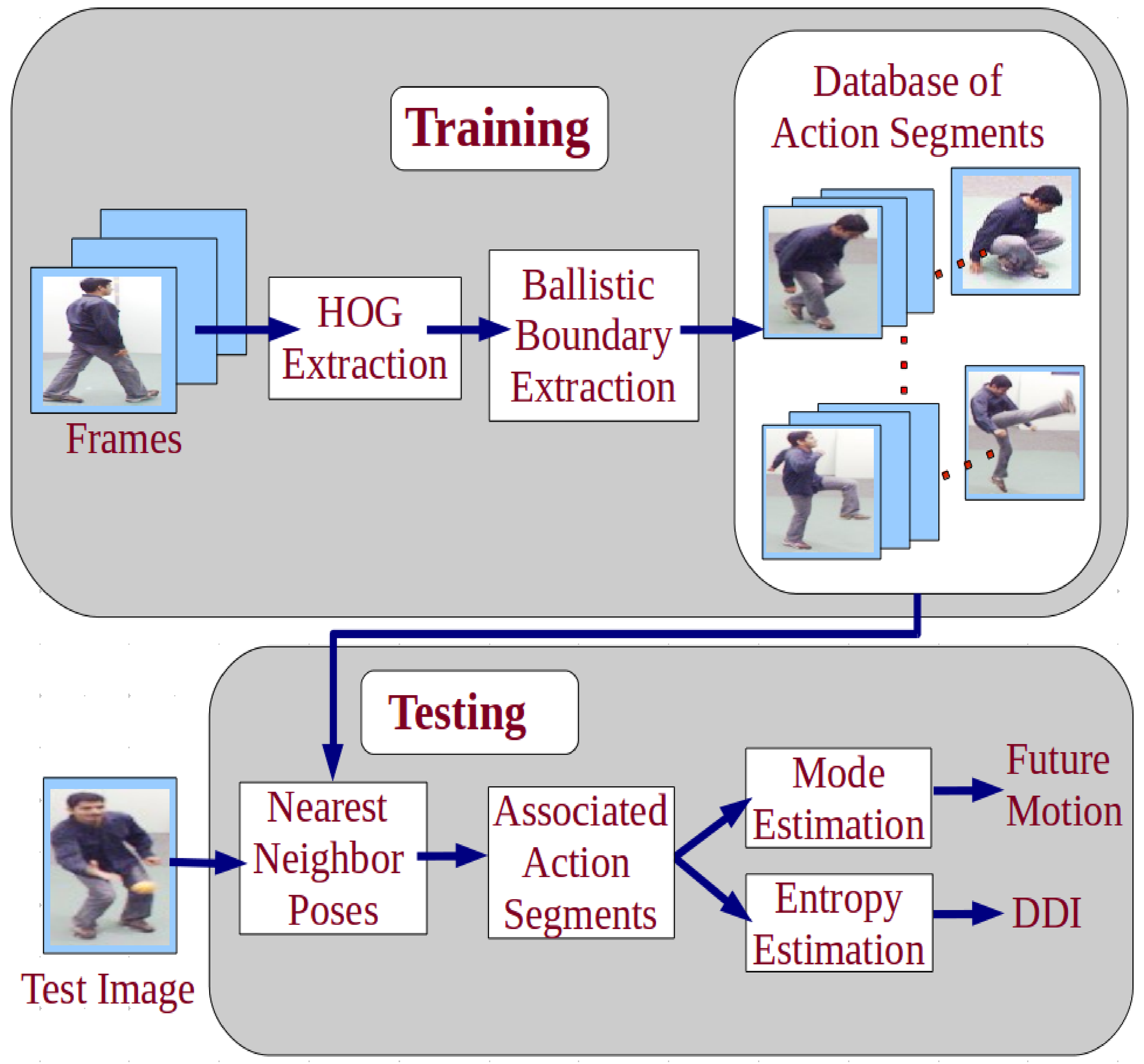
†Geometric Media Lab, Arizona State University
*Center for Automation Research, University of Maryland, College Park
#Zoox Inc., $Google Research

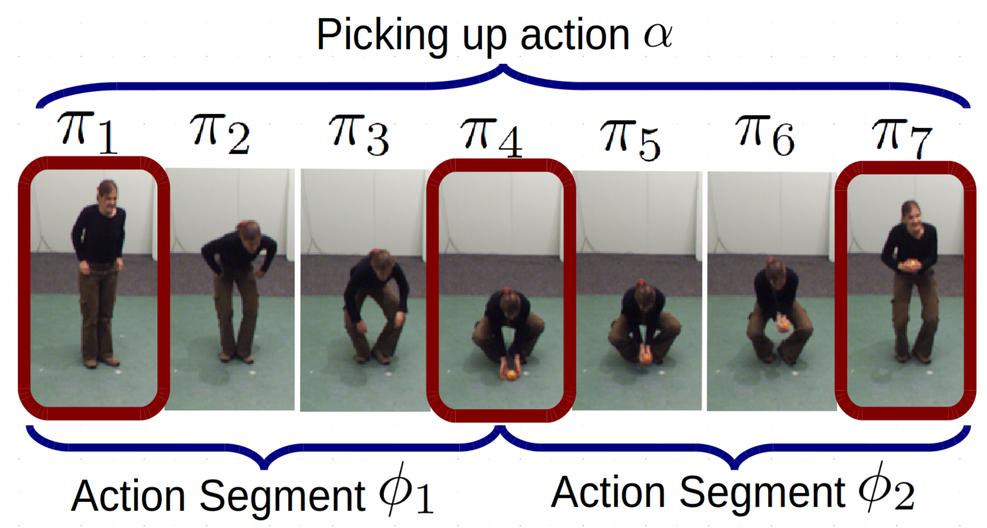
UNIVERSITY OF MARYLAND

Motion Inference from Human Pose

- Posture of a human body is an important indicator of ensuing motion e.g. the experiments of Hirai and Hiraki [1]
- Given a single pose, we use statistical inference on the Grassmannian to predict the future motion of a human computationally
- We quantify the extent to which future motion is constrained by a given pose
- We apply this approach to a variety of vision problems like human motion prediction and activity recognition

Computational Approach for Predicting Human Activity from a Single Image





- We define degree of dynamic information (DDI) present in a pose:

$$\mathcal{H}(\phi|\pi) = -\int_{\phi\in\Phi} \mathcal{P}(\phi|\pi)\log(\mathcal{P}(\phi|\pi))d\phi$$

$$\mathrm{DDI}(\pi) = \exp[-\mathcal{H}(\phi|\pi)]$$

Estimating the conditional distribution

- Modeling action segments using LDS

$$z_\phi(t+1) = A_\phi z_\phi(t) + v_\phi(t), v_\phi(t) \sim N(0,\Xi)$$
$$y_\phi(t) = C_\phi z_\phi(t) + w_\phi(t), w_\phi(t) \sim N(0,\Theta)$$

$$\Omega_\phi^\top = \left[C_\phi^\top, (C_\phi A_\phi)^\top, \ldots, (C_\phi A_\phi^{m-1})^\top, \ldots\right]$$

$$\zeta^2(\Omega_i,\Omega_j) = p - tr(\Omega_j^T\Omega_i\Omega_i^T\Omega_j)$$

- Density estimation on the Grassmannian [3]

$$\hat{\mathcal{P}}(\phi|\pi_s) = c_1 \sum_{\phi_i\in\mathcal{N}_{\phi(\pi_s)}} \Psi(M^{-\frac{1}{2}}(I_d - \Omega_i^\top\Omega\Omega^\top\Omega_i)M^{-\frac{1}{2}})$$

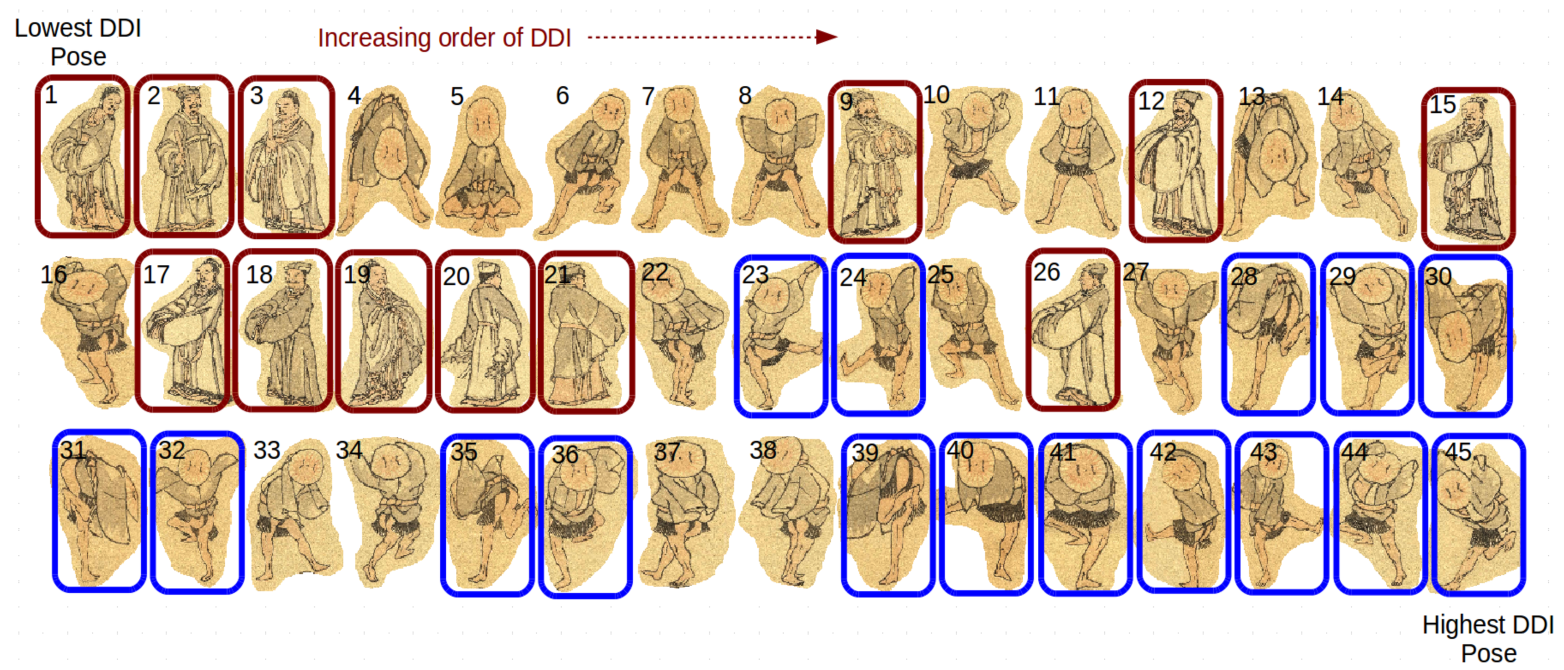
- Statistical inference on the estimated density

$$\hat{\phi}(\pi_s) = \arg\max_{\phi_i\in\mathcal{N}_{\phi(\pi_s)}} \hat{\mathcal{P}}(\phi_i|\pi_s)$$

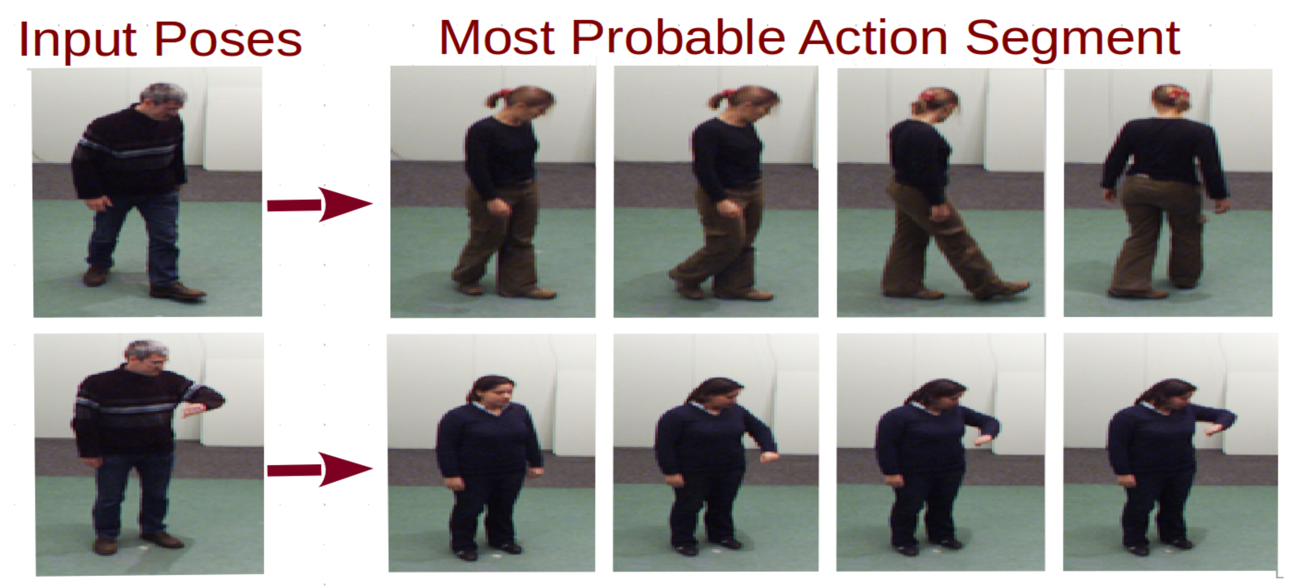
Applications and Experimental Results

1. Perceptual Evaluation on Manga Images

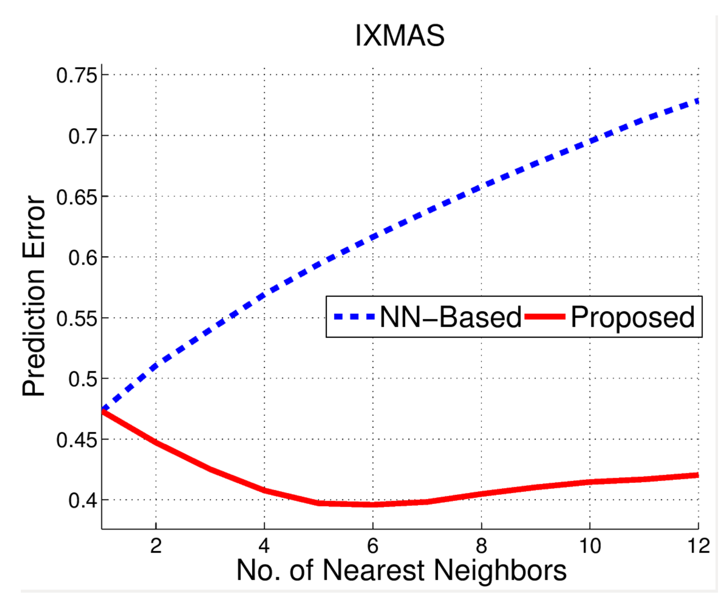
- We estimate the amount of motion information in the Hokusai Manga image database
- Osaka et al. [4] showed that the dancer images activated the motion sensitive regions of the visual cortex, while the priest images did not
- Our DDI metric shows very similar trends



2. Human Motion Prediction

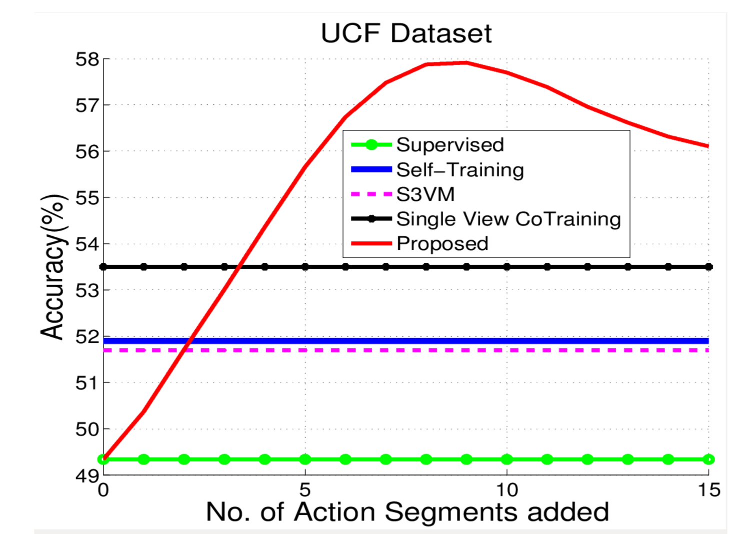


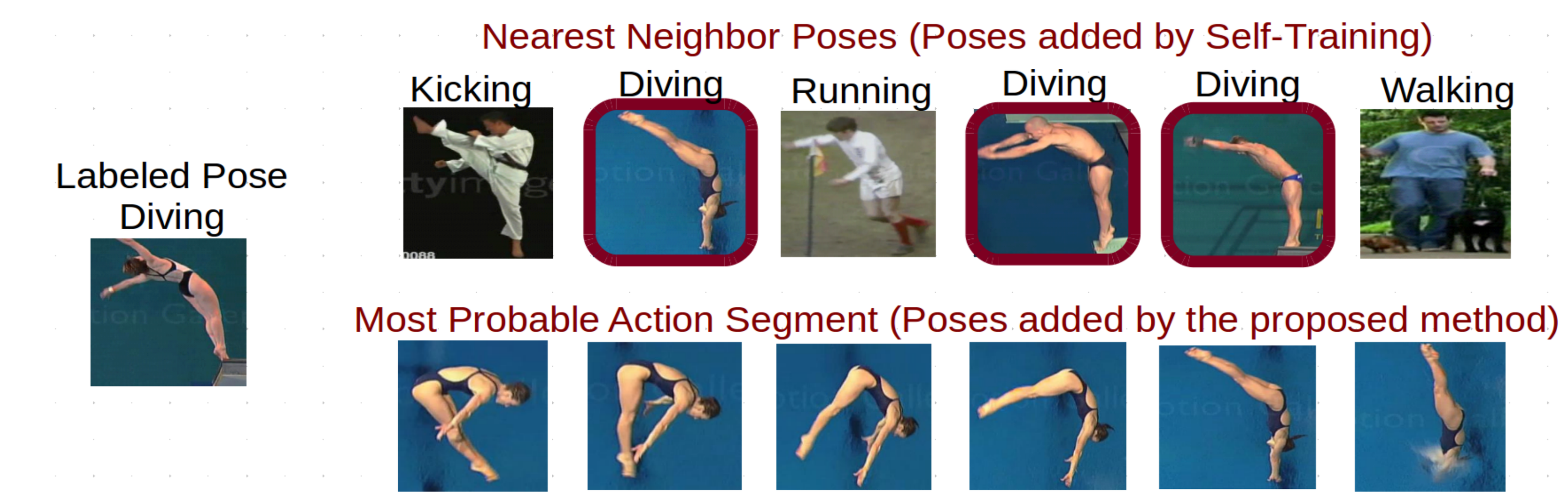
- We use IXMAS dataset to predict motion given a single pose
- We used first nine subjects in the first view as training data and predicted future motion for each pose of last subject.



3. Semi-supervised Single-Image Action Recognition

- We evaluate the label propagation technique for semi-supervised activity recognition
- We considered nine actions of UCF Sports Activity dataset, using 2 subjects for training, 2 for testing and 8 as unlabeled data





Poses added by the proposed method differ from the labeled ones, thus improving training data diversity.

[1] M. Hirai and K. Hiraki. The relative importance of spatial versus temporal structure in the perception of biological motion: An event-related potential study. Cognition, 99(1), 2006
[2] S. N. P. Vitaladevuni, V. Kellokumpu, and L. S. Davis. Action recognition using ballistic dynamics. In International Conference on Computer Vision and Pattern Recognition, 2008
[3] Y. Chikuse. Statistics on Special Manifolds. Springer-Verlag, 2003.
[4] N. Osaka, D. Matsuyoshi, T. Ikeda, and O. M. Implied motion because of instability in Hokusai Manga activates the human motion-sensitive extrastriate visual cortex: an FMRI study of the impact of visual art. Neuroreport, 21(4), 2010.