Computer-Assisted Labeling of Motor Stereotypies in Video

Joshua Fasching[1], Nicholas Walczak[1], William D. Toczyski[1], Kathryn Cullen M.D.[2], Guillermo Sapiro[3], Vassilios Morellas[1], Nikolaos Papanikolopoulos[1]

[1]Department of Computer Science and Engineering, [2]Department of Psychiatry, University of Minnesota
[3]Department of Electrical and Computer Engineering, Duke University

Introduction: *What* and *Why*

Stereotypies: Repetitive and unvarying behavioral sequences lacking any obvious function or eliciting stimuli (Goldman, *et al.,* 2009).

Classifying Stereotypies: Correlated to numerous possible developmental disorders, including Autism and Rett Syndrome.

This research: New tools towards the automated identification of motor stereotypic behaviors.

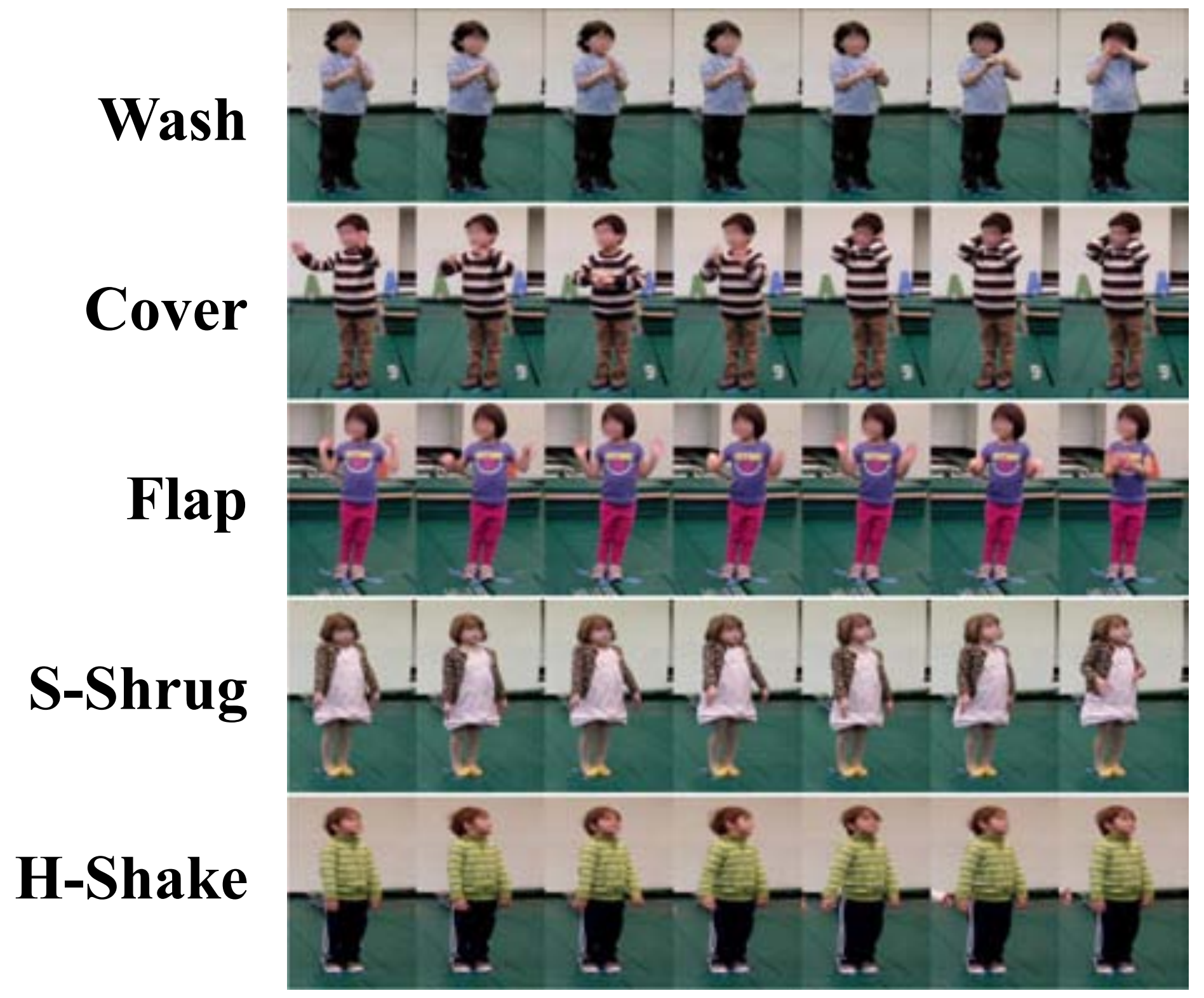
Long-Term Technical Objective: The collection and efficient processing of extremely large and very diverse video and 3D-based datasets (on scales until now impractical).

Long-Term Clinical Objective: Very large-scale multimodal data analysis of long-term observations for clinical evaluation and data mining.

Methods

Dataset

- Children were recorded while playing a mimicry game led by their teacher at the University of Minnesota's Shirley G. Moore Laboratory School (with IRB approval).
- Children performed actions which simulated motor stereotypies
- Training data was created by manually annotating information in the video (see examples below).
- This dataset of stereotypic-like behaviors was used to train a computational model for distinguishing stereotypic behaviors.



Total Number of Subjects	16
Number of Videos Per Action	
Hand Washing	22
Ear Covering	44
Hand Flapping	20
Shoulder Shrugging	61
Head Shaking	25
Total	172

Task 1: Automatic Classification

- Motion and appearance descriptors suitable for automatic classification of motor stereotypies in video were extracted following the work of Wang et al. (2011).

***How it works*:**

- First, initialize thousands of *virtual point trackers* equally spaced in the image space of the video (see below).

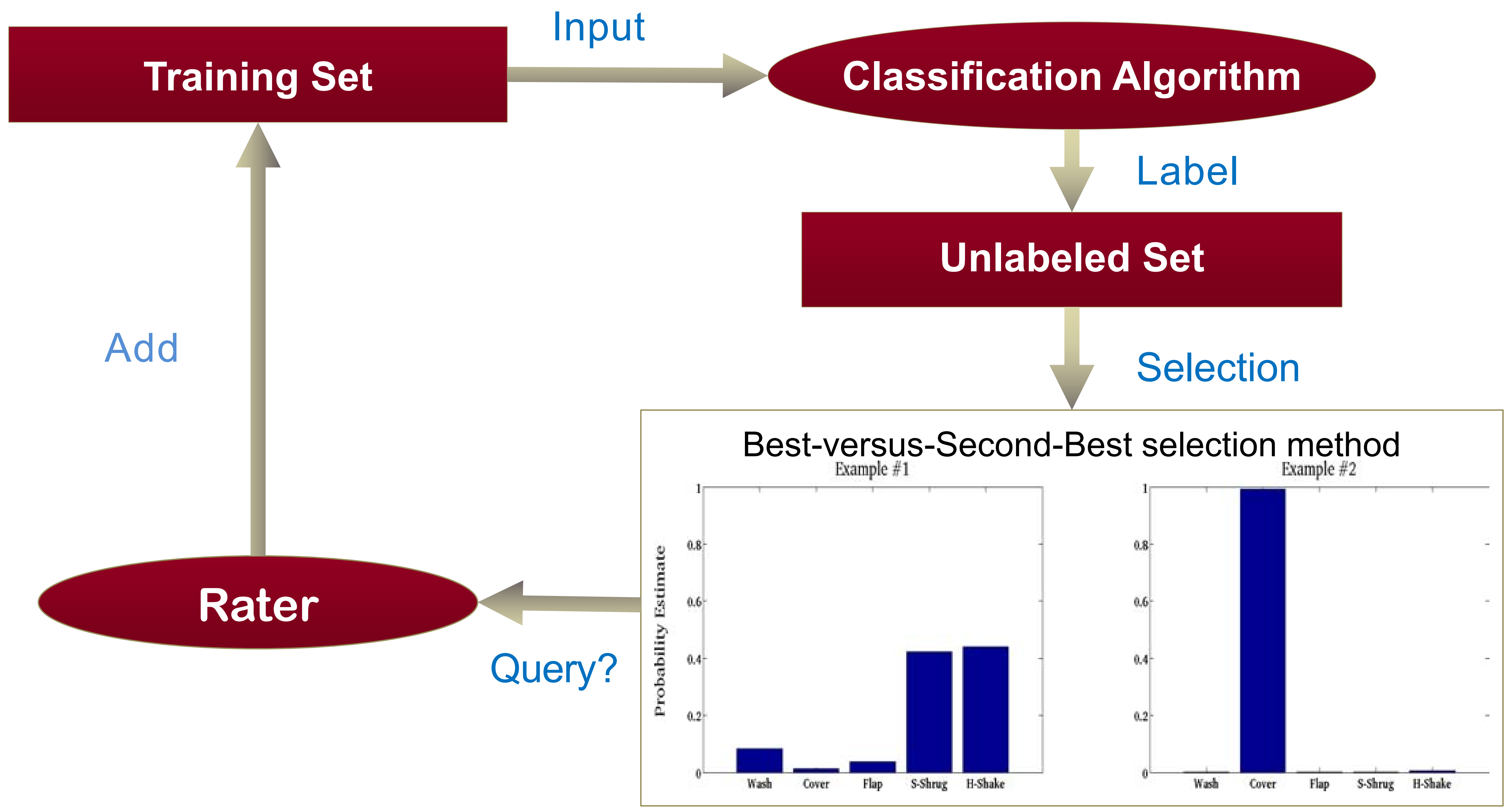
Note: under 10% of virtual point trackers shown (for clarity)

- Retain trackers that contain long-term fluid motion and use them as a spotlight for extraction of descriptors useful for low-level image analysis.
- These descriptors, using *vector quantization*, allow a video to be transformed into a single data vector that is amenable to statistical and machine learning techniques.
- *Learn* a set of support vector machines (SVM) with the X^2-kernel $k(x,y) = \exp\left(-\gamma \sum_i \frac{(x_i-y_i)^2}{x_i-y_i}\right)$ to cluster and classify the different stereotypic events, thus automatically labeling them.

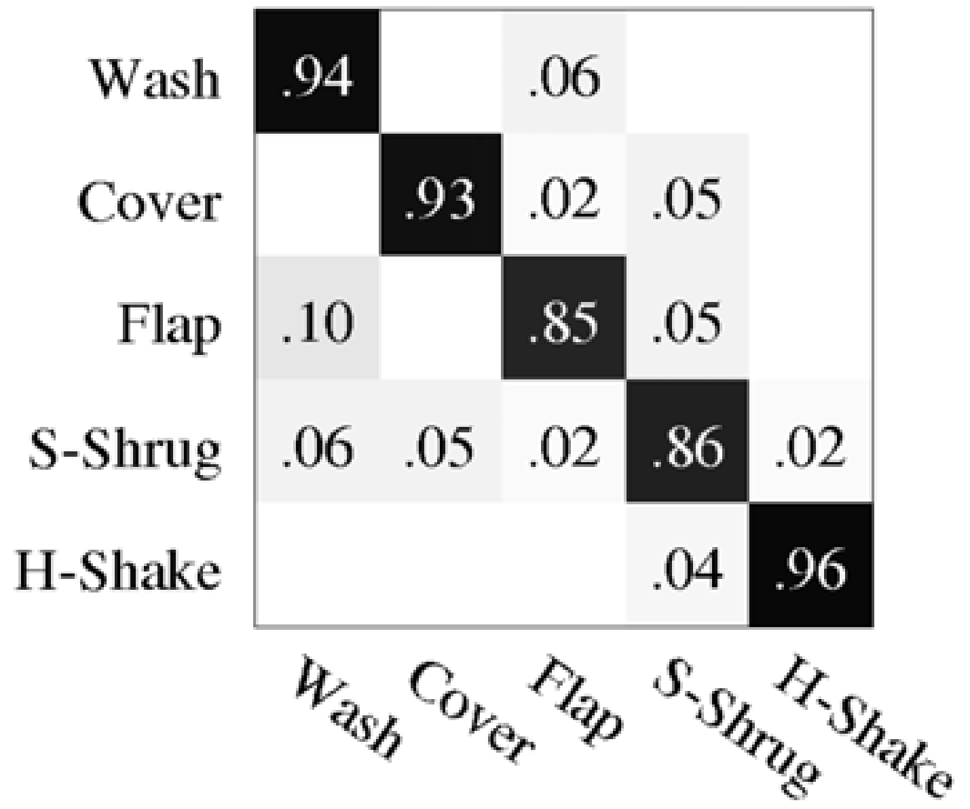
Task 2: Computer-Assisted Labeling

- For a large set of videos it may become prohibitive to label enough videos to achieve the best classification performance.
- How do we select the best examples for human labeling to achieve the best performance?
- Our approach uses the Best-versus-Second-Best selection method (Joshi et al. 2009).

***How it works*:**



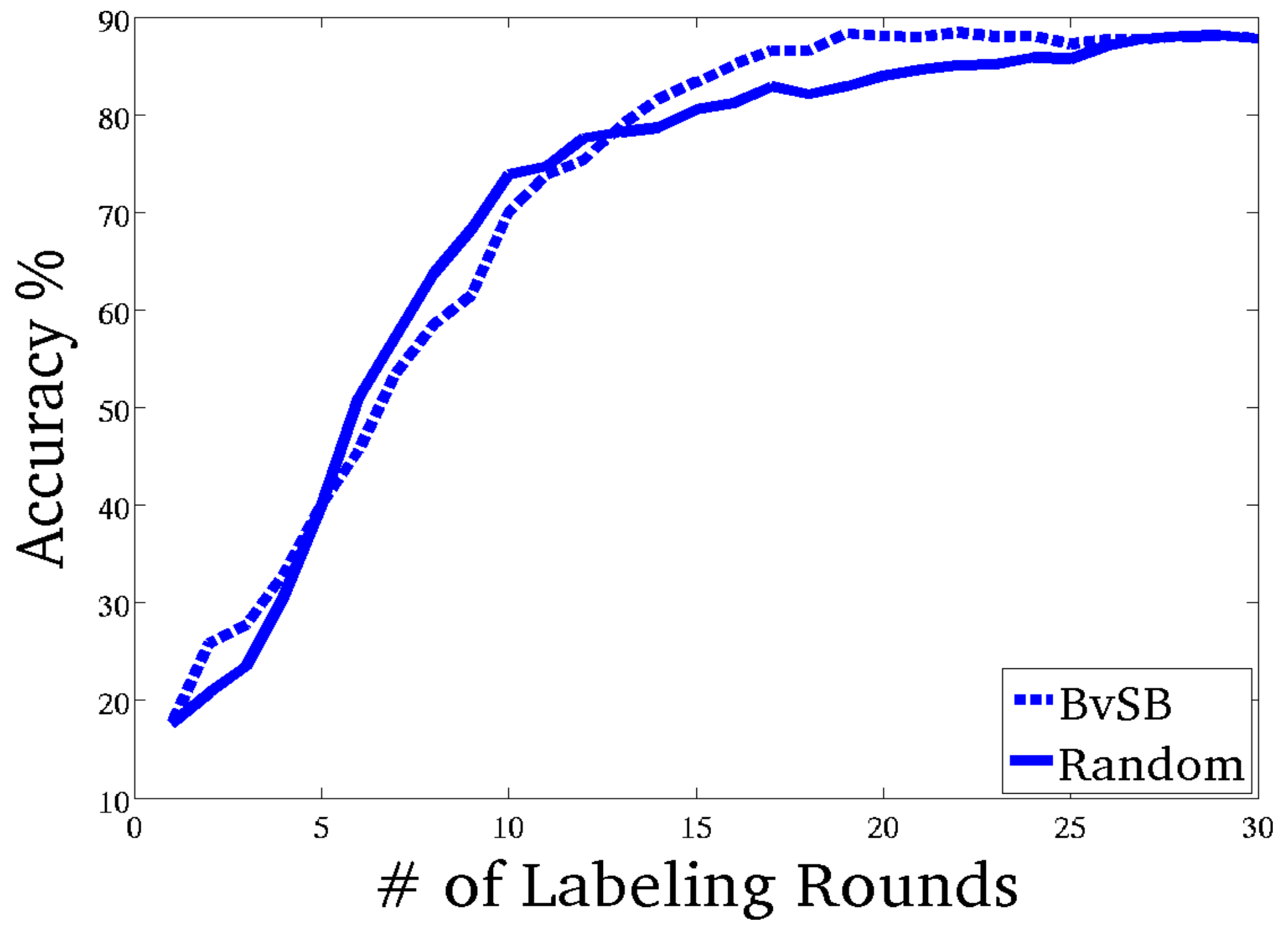
Results



	Wash	Cover	Flap	S-Shrug	H-Shake
Wash	.94		.06		
Cover		.93	.02	.05	
Flap	.10		.85	.05	
S-Shrug	.06	.05	.02	.86	.02
H-Shake				.04	.96

Left: Confusion Matrix for automatic classification of motor stereotypic behaviors in video using Motion Boundary Histogram video features computed from densely sampled trajectories. The "head-shaking" and "shoulder shrugging" actions are abbreviated as "H-Shake" and "S-Shrug" respectively.

Right: Classification accuracy comparing the performance of Best-versus-Second-Best active learning labeling strategy versus random selection for each round. In each round, five examples are selected. The Best-versus-Second-Best active learning strategy achieves the upper limit of classification performance earlier than random selection.



Conclusions

Overall, these methods represent progress in overcoming the impractical amounts of manual video annotation currently demanded of any large observational study of human subjects with high volumes of raw data recordings. Future progress for our work will include

- Future tests of our learned classifiers using actual (not-mimicked) examples of motor stereotypic behavior to verify our reliance on using the more easily acquired mimicked examples for system training,
- Extending these computer-assisted labeling methods to other pre-defined behaviors of known relevance to early mental illness diagnostics, and
- Incorporating a subsystem for automatic 'stereotypic action alerts' into our larger pre-school classroom monitoring system.

References:

A. Joshi, F. Porikli, and N. Papanikolopoulos. "Multi-class active learning for image classification.", IEEE Conference on Computer Vision and Pattern Recognition, pages 2372-2379. IEEE, Jun 2009.

S. Goldman, C. Wang, M. W. Salgado, P. E. Greene, M. Kim, and I. Rapin."Motor stereotypies in children with autism and other developmental disorders". Developmental Medicine & Child Neurology, 51(1):30–38, 2009.

H. Wang, A. Klaser, C. Schmid, and C.-L. Liu. "Action recognition by dense trajectories".
IEEE Conference on Computer Vision and Pattern Recognition, pages 3169–3176. IEEE, Jun 2011.

Acknowledgments

This material is based upon work supported by the National Science Foundation through grants #IIP-0443945, #CNS-0821474, #IIP-0934327, #CNS-1039741, and #SMA-1028076. Additional support for GS from NGA, ONR, ARO, and NSSEFF (AFOSR) is also gratefully acknowledged.