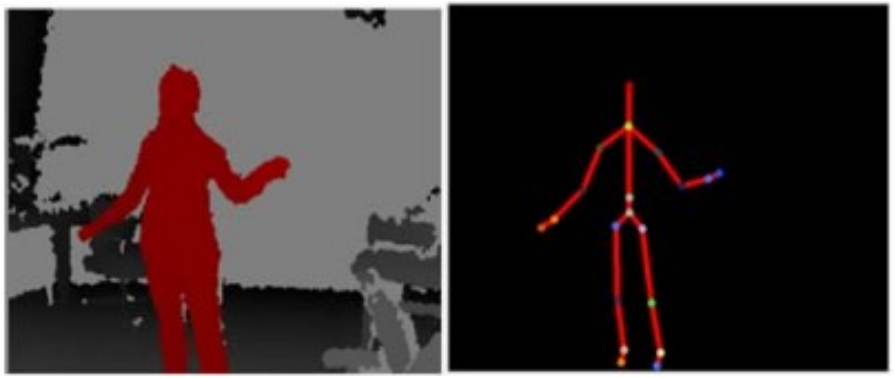
Robot-Centric Activity Recognition 'in the Wild'

Ilaria Gori, Jivko Sinapov, Priyanka Khante,
Peter Stone and J. K. Aggrawal

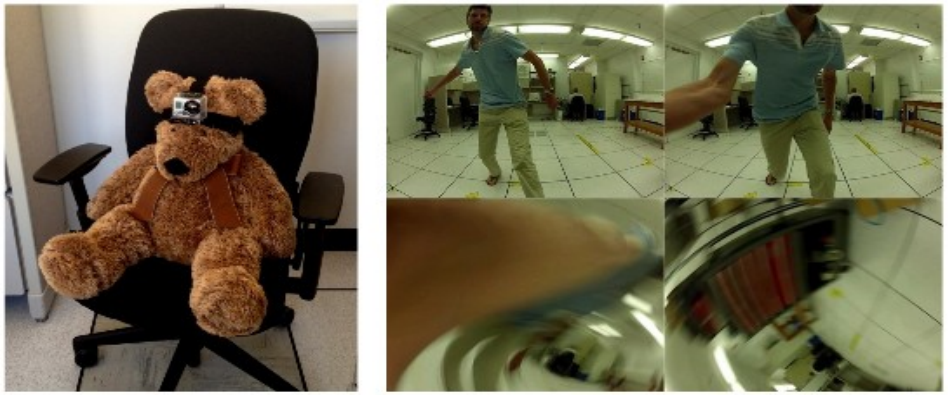
University of Texas at Austin, Austin TX 78712, USA
{ilaria.gori,aggarwaljk}@utexas.edu,
{jsinapov,pkhante,pstone}@cs.utexas.edu

Motivation

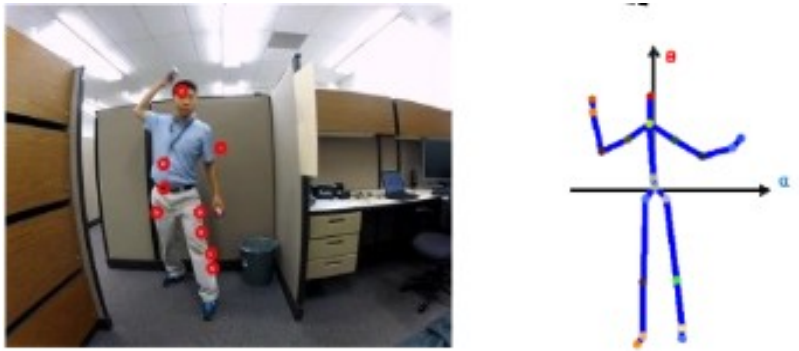
Related Work



(Xia *et al*. 2011)



(Ryoo and Matthies 2013)

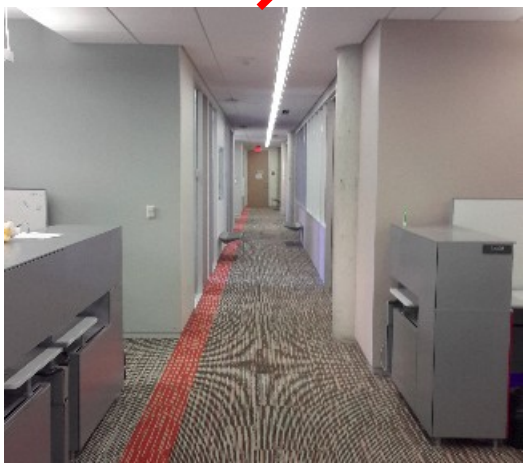


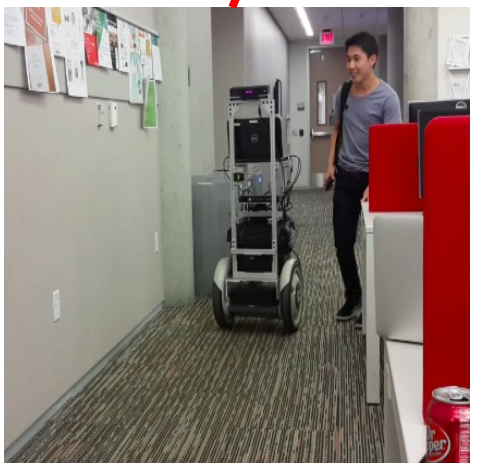
(Ryoo *et al.* 2015)

Limitations of Existing Work

- The activities were specified by the researchers ahead of the experiment
- The activities were performed by a small number (5 to 8) of 'actors'
- The robot is either stationary or teleoperated

Dataset Collection



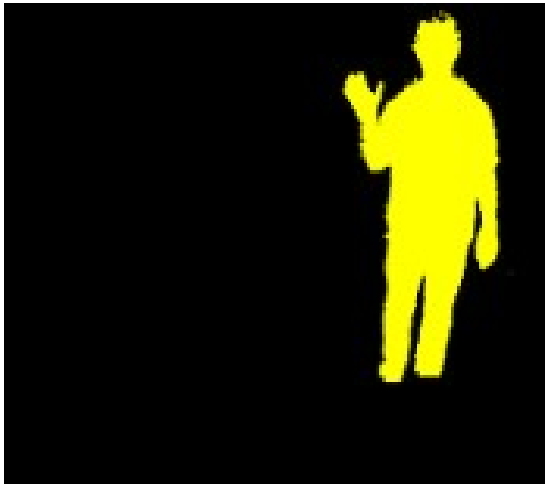


Video

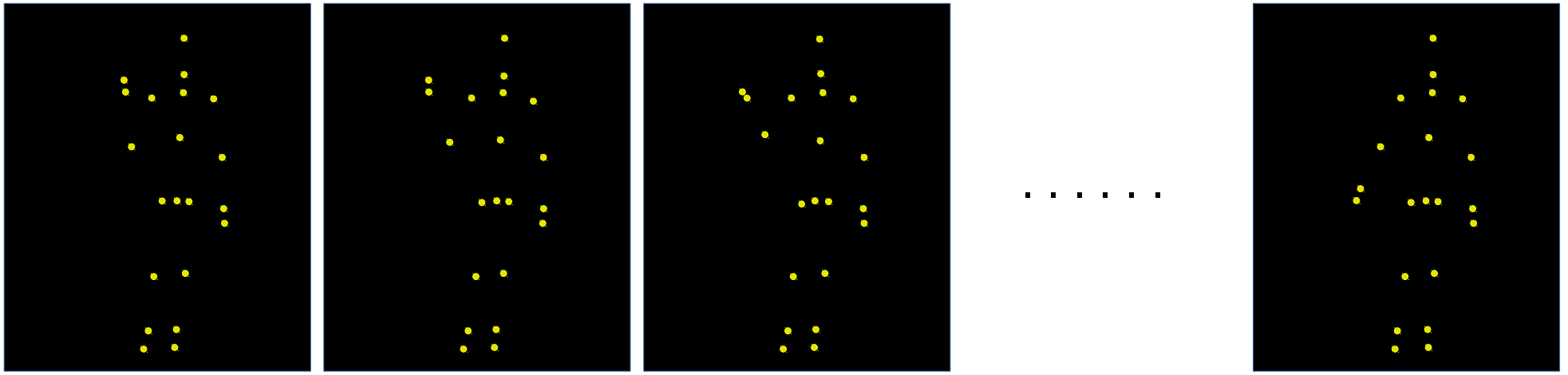
Dataset Collection

- Robot: Segbot
- Environment: 3rd Floor of GDC, spanning a public undergraduate lab and a graduate lab
- The robot autonomously traversed the environment for 1-2 hours a day over the course of 6 days covering ~14 km total
- Whenever the robot's Kinect 2.0 detected a person, the robot recorded a range of visual and non-visual data which was later used for classification

Example Human Detection


Example Human Detection



Recorded Data

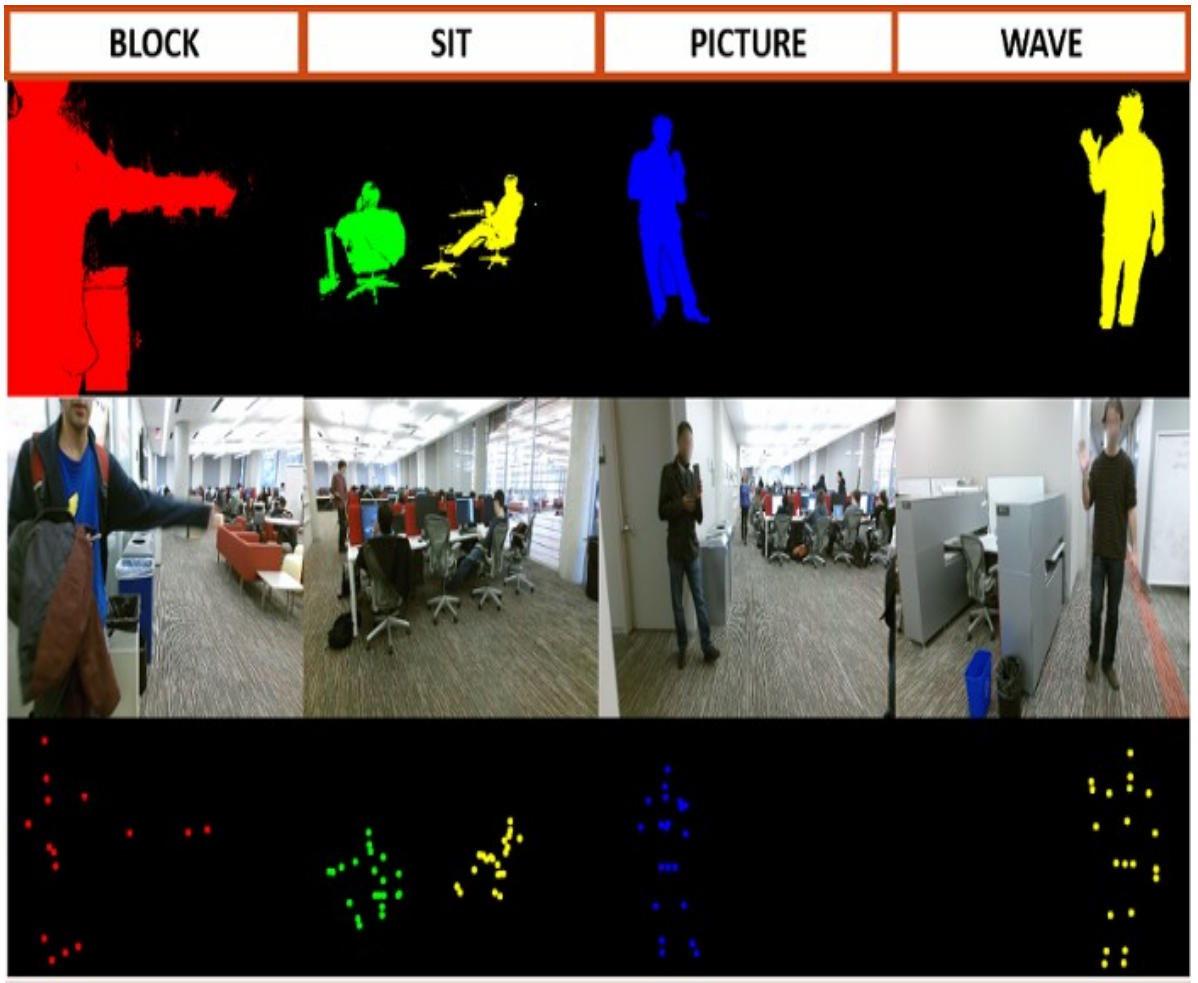
Features	Dimension	Range	Sampling Rate (Hz)
RGB images	512×424×3	{0,255}	50
Depth images	512×424	{0,65535}	50
3D Joints	21x3	$\mathbb{R}$	50
2D Joints	21x2	{0,512}×{0,424}	50
Robot's pose on the map	7	$\mathbb{R}$	1.5
Robot's odometry	7	$\mathbb{R}$	100
Activity label	1	{1,10}	–

Recorded Data

Features	Dimension	Range	Sampling Rate (Hz)
RGB images	512×424×3	{0,255}	50
Depth images	512×424	{0,65535}	50
3D Joints	21x3	$\mathbb{R}$	50
2D Joints	21x2	{0,512}×{0,424}	50
Robot's pose on the map	7	$\mathbb{R}$	1.5
Robot's odometry	7	$\mathbb{R}$	100
Activity label	1	{1,10}	–

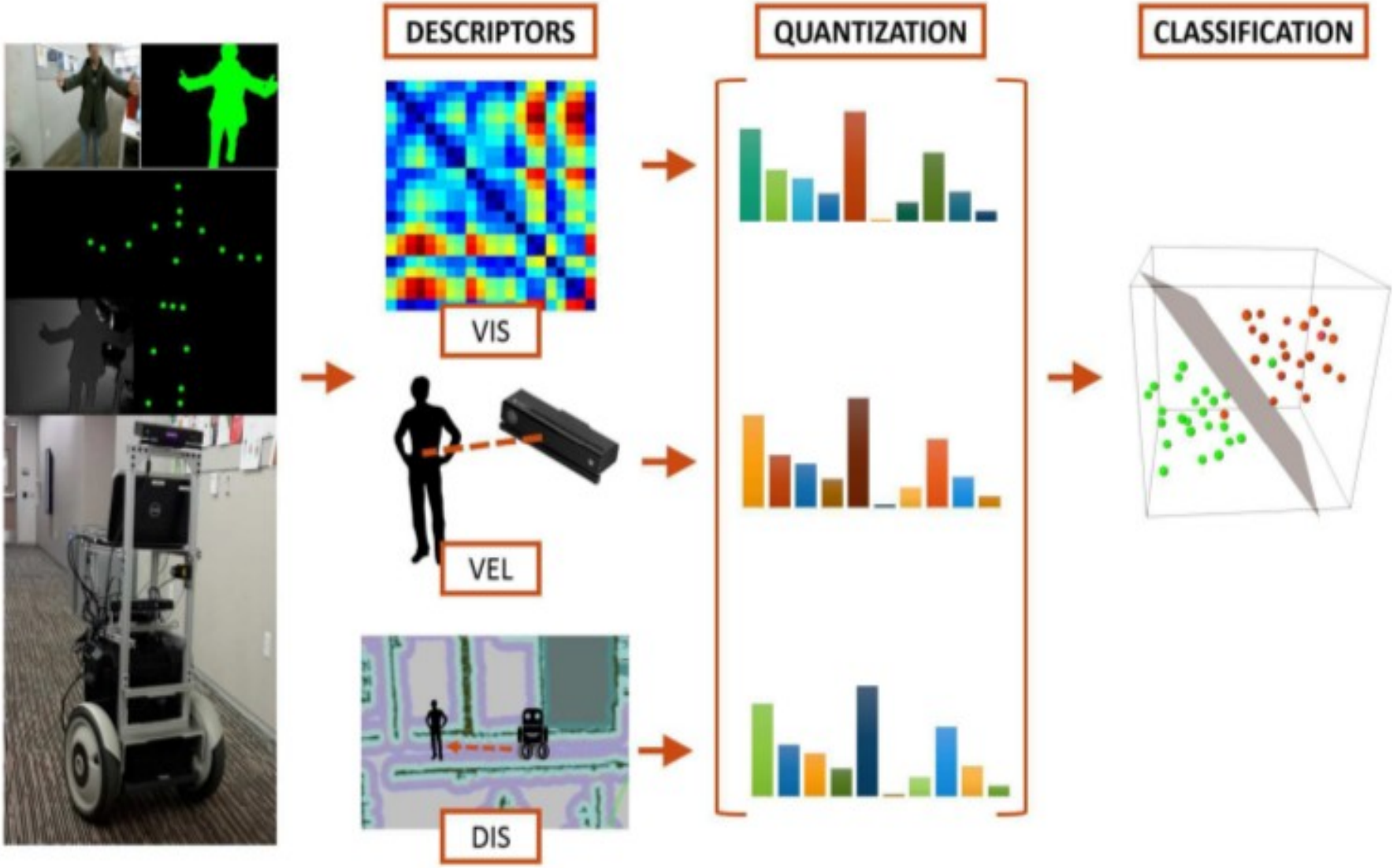
Dataset size: ~ 140 GB
Available upon request

Activity Labels



Activity	Num Samples
Picture	6
Wave	12
False	608
Block	23
Pass by	153
Walk away	68
Approach	33
Sit	150
Stand	106
Side pass	45

System Overview



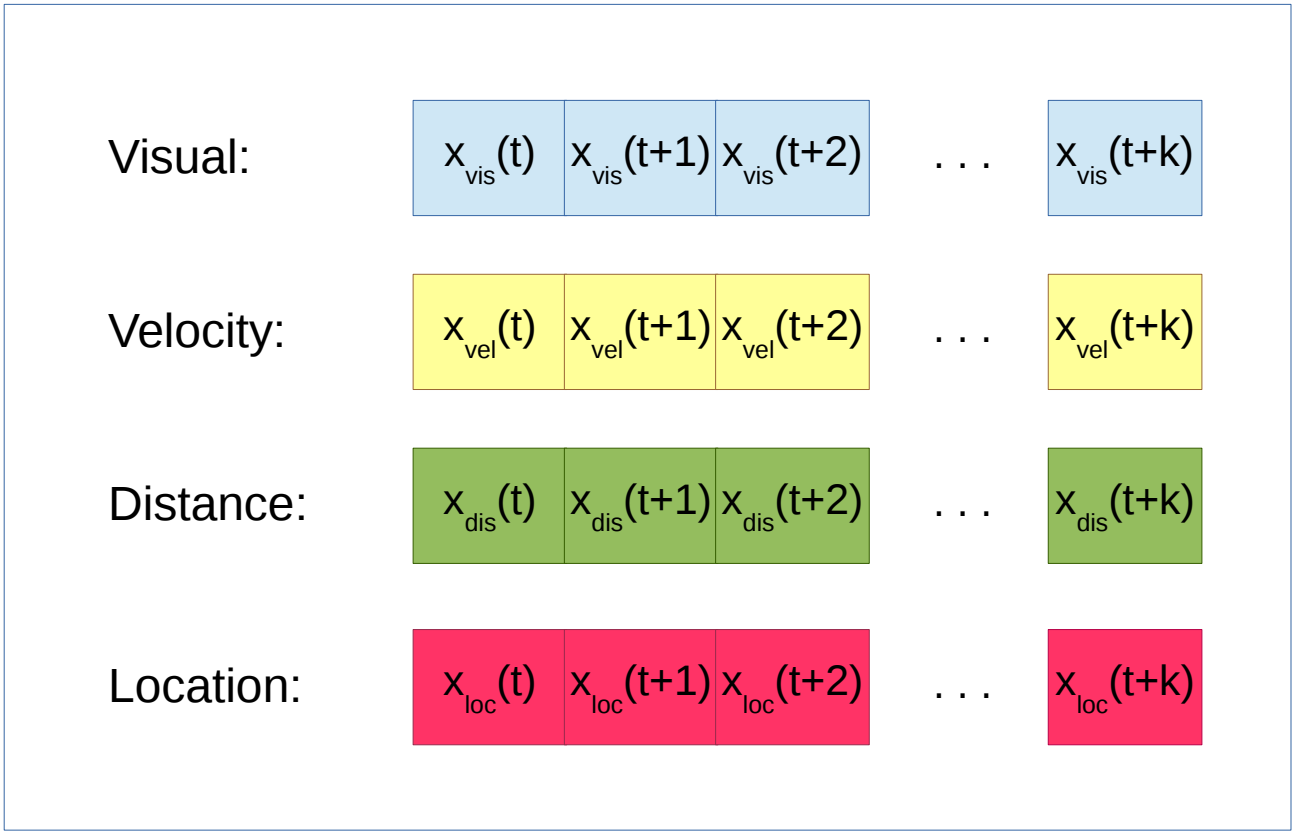
Visual Features

- Histogram of 3D Joints (HOJ3D)
- Covariance of Joint Positions over Time (COV)
- Histogram of Direction Vectors (HODV)
- Histogram of Oriented 4D Normals (HON4D)
- Pairwise Relational Matrix (PRM)

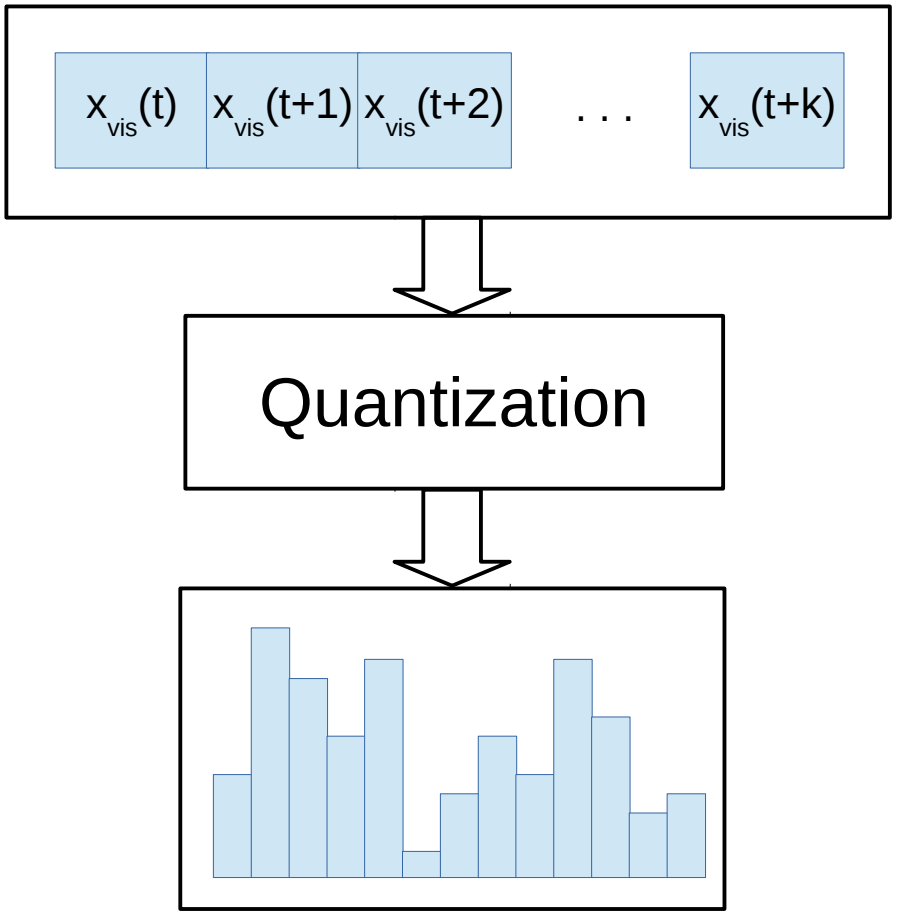
Additional Features

- **Human-Robot Velocity Features:** The direction in which the human moves with respect to the robot
- **Distance Features:** The distance between the human and robot over time
- **Localization Features:** The robot's pose (position and orientation) in the map

Example Feature Sequence



Feature Quantization



Feature Quantizations

- The computed features for each descriptor were quantized using k-means
- Bag-of-Words representation was obtained by counting the occurrence of each “word” over the course of each video
- The BoW representations of all descriptors were concatenated to obtain a final feature vector

Evaluation

- Evaluation was performed using 5-fold cross validation
- Because the dataset was unbalanced, the *kappa statistic* was used to measure performance

Probability of correct classification by classifier

Probability of correct classification by chance

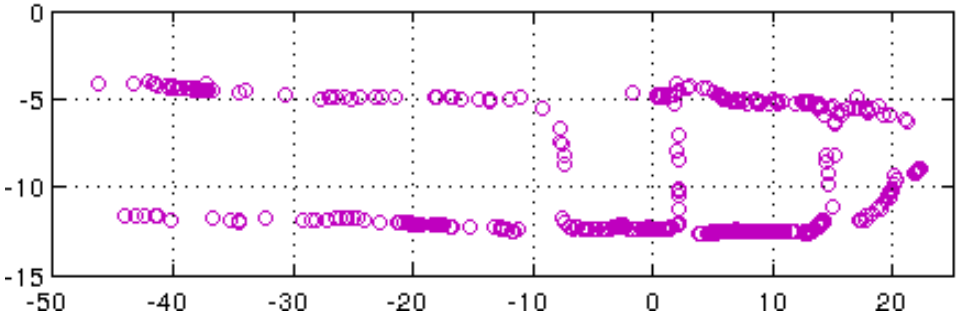
$$K = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

Classification Results

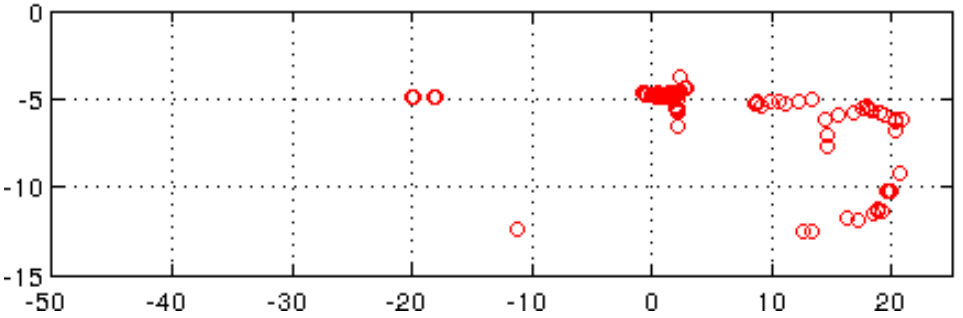
	Vision Only	Vision + Distance + Velocity
COV [6]	0.329	0.440
HOJ3D [16]	0.515	0.633
HODV [3]	0.624	0.649
PRM	0.547	0.660
HON4D [11]	0.756	0.762

Can the robot exploit the spatial structure of activities?

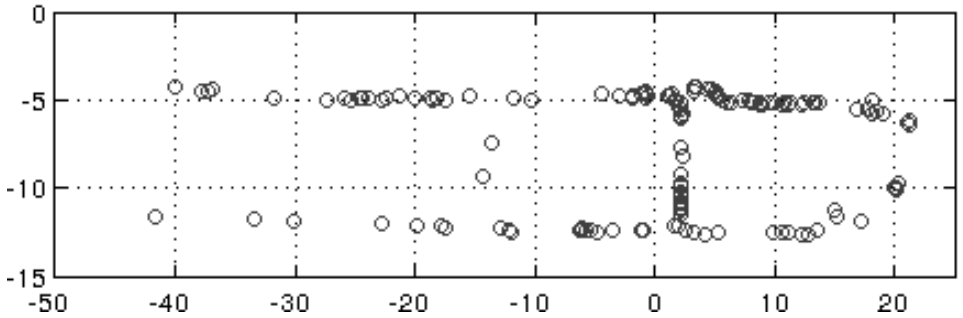
Can the robot exploit the spatial structure of activities?



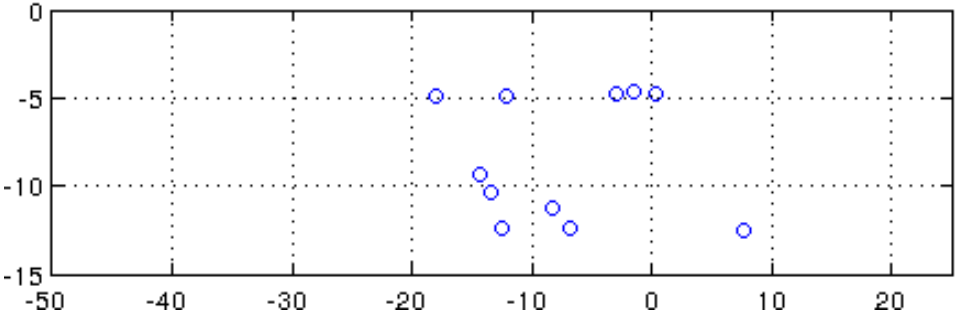
“false detection”



“sit”



“walk away”



“wave”

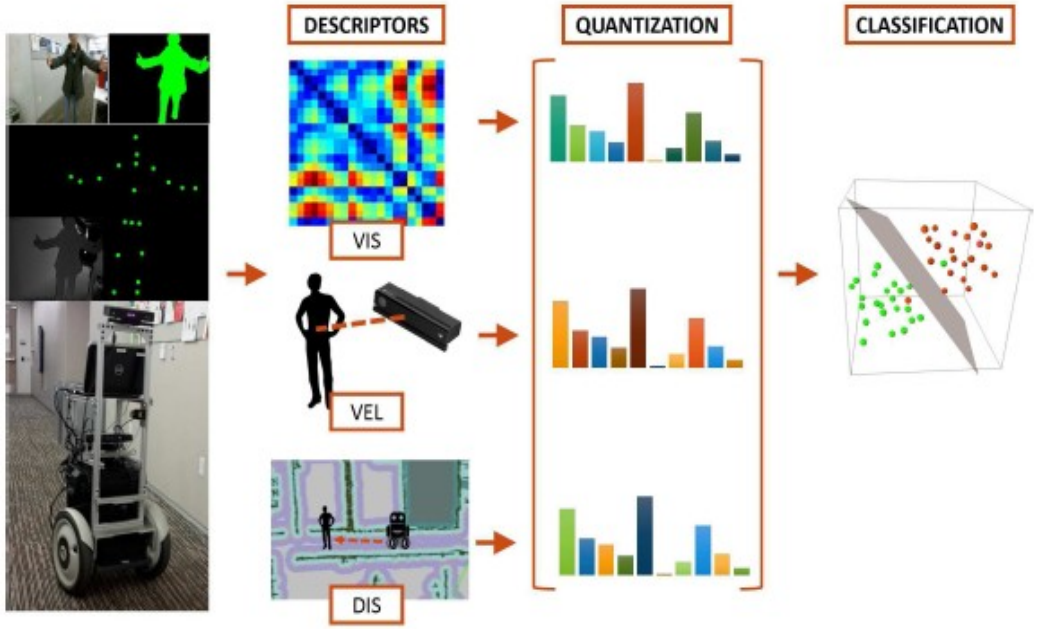
Classification Results

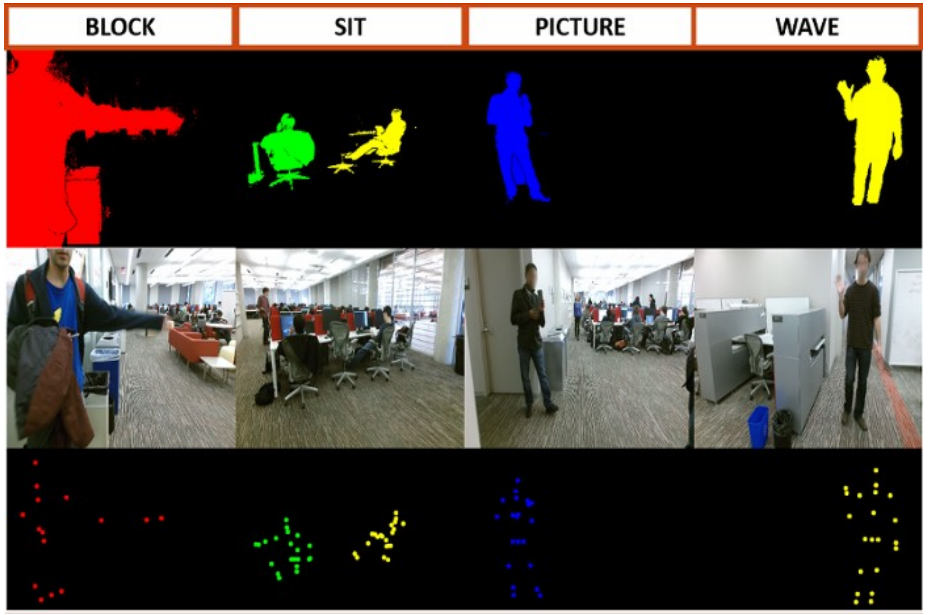
	Vision Only	Vision + Distance + Velocity	Vision + Distance + Velocity + Localization
COV [6]	0.329	0.440	0.462
HOJ3D [16]	0.515	0.633	0.651
HODV [3]	0.624	0.649	0.660
PRM	0.547	0.660	0.671
HON4D [11]	0.756	0.762	0.764

Summary and Conclusion

- Conducted largest experiment in robot-centric activity recognition to-date
- Dataset is available upon request
- Evaluated 5 different visual features
- Demonstrated that non-visual features can improve classification results

Thank you!



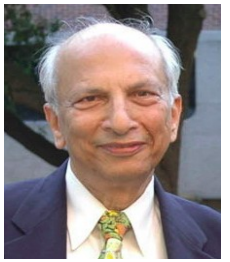



Ilaria Gori


Jivko Sinapov


Priyanka Khante


Peter Stone


J.K. Aggarwal

http://www.cs.utexas.edu/~larg/bwi_web/