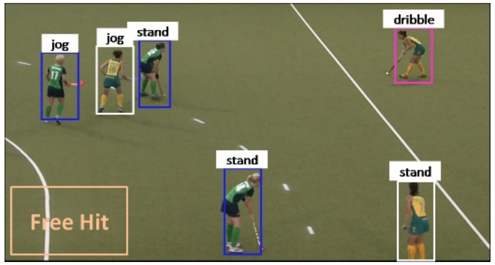
Analysis and Evaluation of Kinect-based Action Recognition Algorithms

School of Computer Science and Software Engineering
Lei Wang
Email: 21676963@student.uwa.edu.au
Supervisor: A/Prof Du Huynh

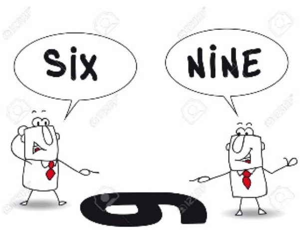
October 2017

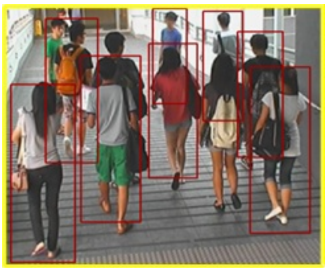
Applications and Issues

Applications of human action recognition:



Challenging issues:

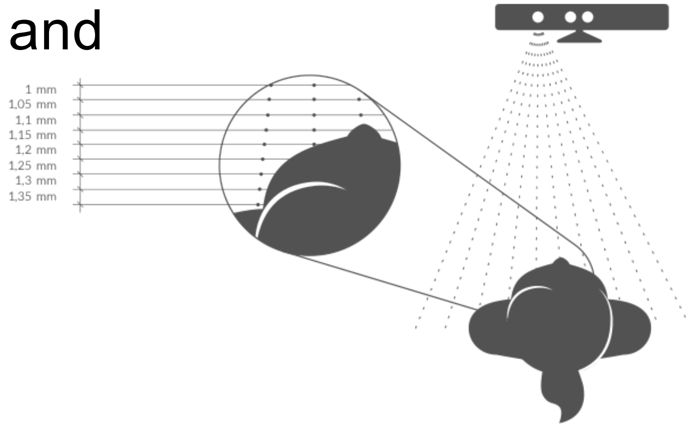




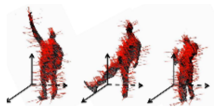
Kinect sensor and Techniques

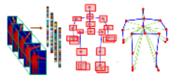
THE UNIVERSITY OF WESTERN AUSTRALIA

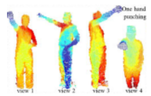
- Kinect sensor
  - Records real time depth sequences
  - Captures 3D information. Advantages:
    - Extra body shape information
    - Insensitive to illumination conditions and the colour of human clothes

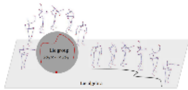


- State-of-the-art techniques
  - HON4D (Oreifej et al., 2013)
  - HDG (Rahmani et al., 2014)
  - HOPC (Rahmani et al., 2016)
  - RBD (Vemulapalli et al., 2016)



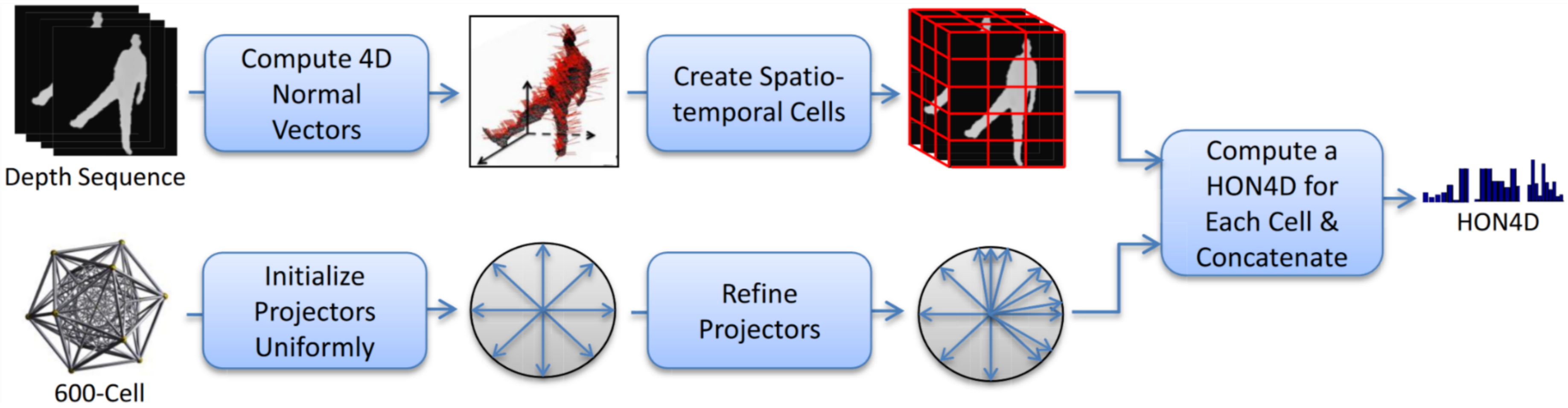






Algorithms to be Analyzed and Evaluated

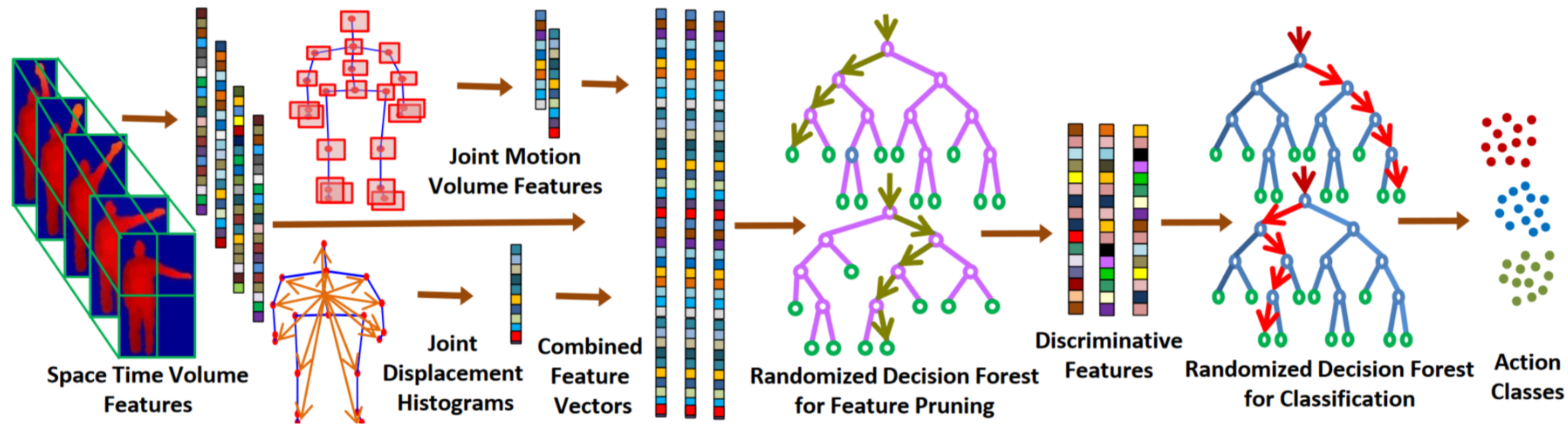
HON4D --- Histogram of Oriented 4D Normals (Oreifej et al., 2013)



- Geometry and motion of human action were captured
- A 4D space was quantised using a 600-cell polychoron
- 120 vertices were used as projectors
- More vertices were induced randomly to increase the difference between two similar action classes

Algorithms to be Analyzed and Evaluated

HDG --- Histograms of Depth Gradients (Rahmani et al., 2014)



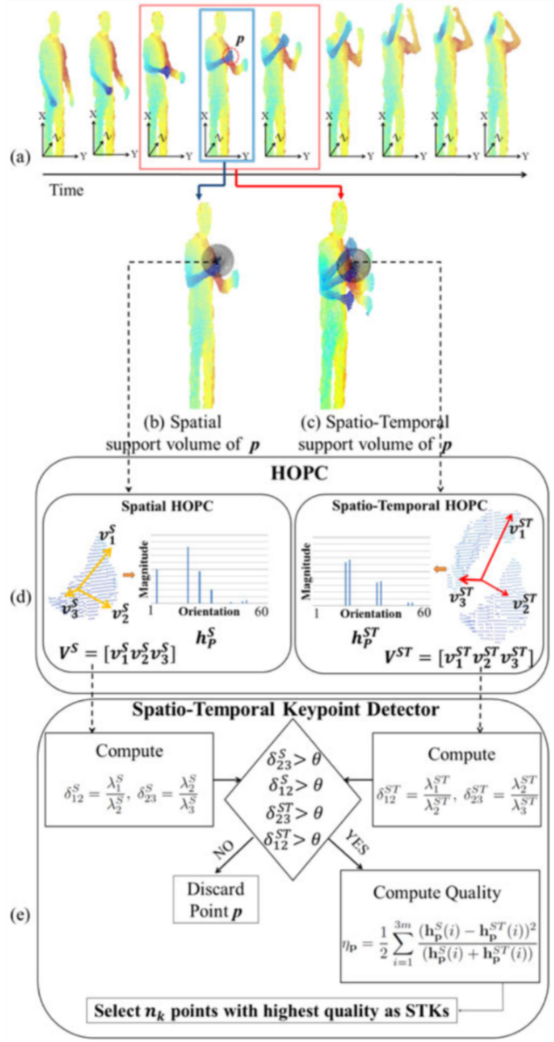
- A concatenation of 4 descriptors
  - Histograms of depth (hod)
  - Histograms of depth derivatives (hodg)
  - Histograms of joint position differences (jpd)
  - Histograms of joint movement volume (jmv)
- Two random decision forests were trained

THE UNIVERSITY OF WESTERN AUSTRALIA

Algorithms to be Analyzed and Evaluated

HOPC --- Histogram of Oriented Principal Components (Rahmani et al., 2016)

- For a sequence of 3D pointclouds
  - HOPC is extracted at each point
  - Two types of support volume were defined
    - Spatial support volume
    - Spatio-temporal support volume
  - Principal component analysis was applied
  - Spatio-temporal keypoints (STKs) detection
  - A quality factor for detecting significant motion variations

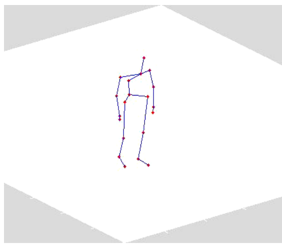


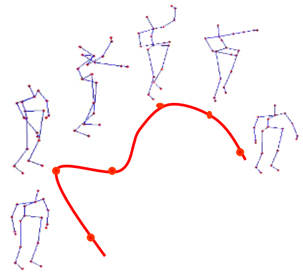
Algorithms to be Analyzed and Evaluated

THE UNIVERSITY OF WESTERN AUSTRALIA

RBD --- Rotation-based Descriptor (Vemulapalli et al., 2016)

Step 1: Skeletal Representation

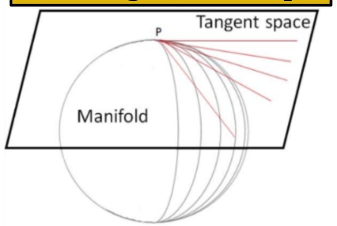




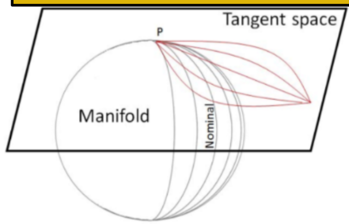
Step 2: Nominal curve computation using DTW

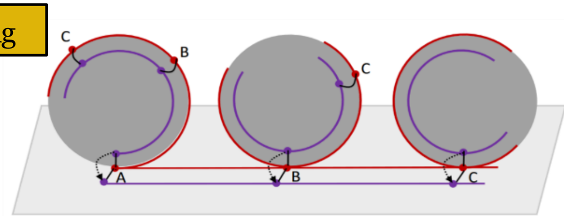
Step 3: Unwrapping while rolling

RBD-logarithm map



RBD-unwrapping while rolling





Step 4: SVM classification

RBD-FTP representation

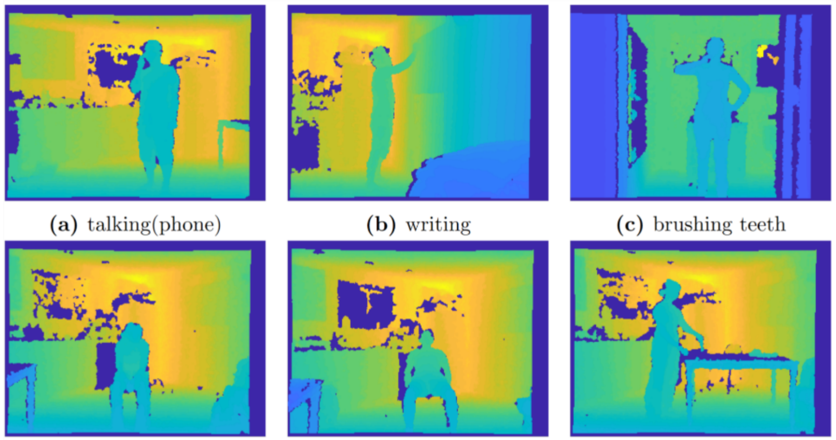
- 3D rotations are members of the special orthogonal group SO_3
- Human actions were represented as curves after skeleton representation
- Dynamic Time Warping (DTW) handles the rate variations
- Rolling maps were used for flattening SO_3
- Fourier Temporal Pyramid (FTP) representation for each unwrapped curve

THE UNIVERSITY OF WESTERN AUSTRALIA

Experimental Datasets

5 benchmark datasets:

Datasets	Classes	Subjects	Views	Sensor	Modalities	Year
MSRAction3D	20	10	1	Kinect v1	Depth + 3DJoints	2010
3D Action Pairs	12	10	1	Kinect v1	RGB + Depth + 3DJoints	2013
Cornel Activity Dataset (CAD-60)	14	4	-	Kinect v1	RGB + Depth + 3DJoints	2011
UWA3D Single View	30	10	1	Kinect v1	RGB + Depth + 3DJoints	2014
UWA3D Multiview	30	9	4	Kinect v1	RGB + Depth + 3DJoints	2015



(a) talking(phone) (b) writing (c) brushing teeth

(d) talking(couch) (e) relaxing(couch) (f) cooking(stirring)

Sample depth images from CAD-60

THE UNIVERSITY OF WESTERN AUSTRALIA

Experimental Settings

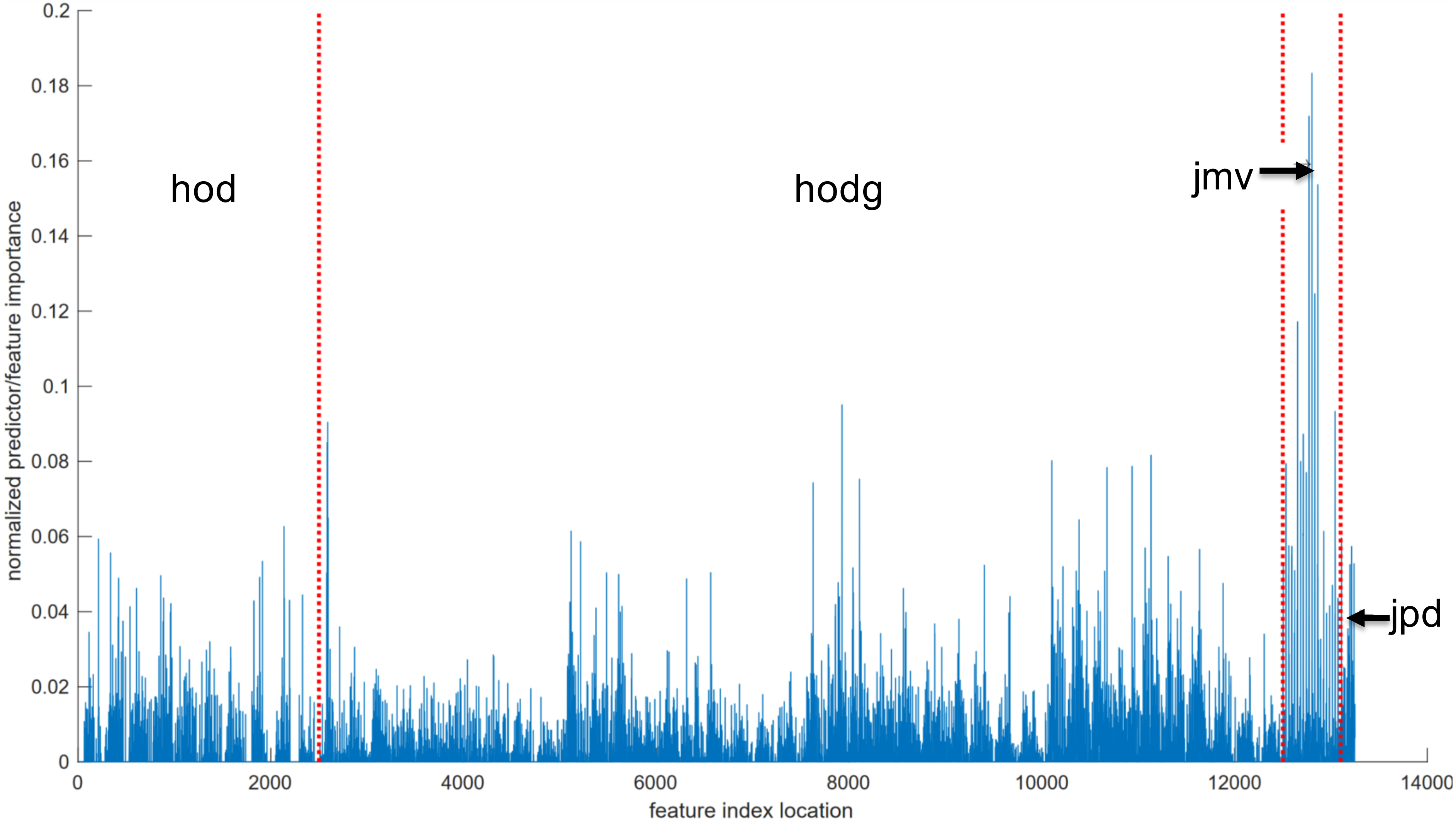
- HDG was implemented in Matlab.
- HON4D, HOPC and RBD were modified from the original authors' codes.

- For the UWA3D Multiview Dataset, a **cross-view action recognition** strategy is used; for the other 4 datasets, half of the subjects' data are used for training and the others for testing.

- **Confusion matrices** are used to illustrate the recognition accuracy of these algorithms.

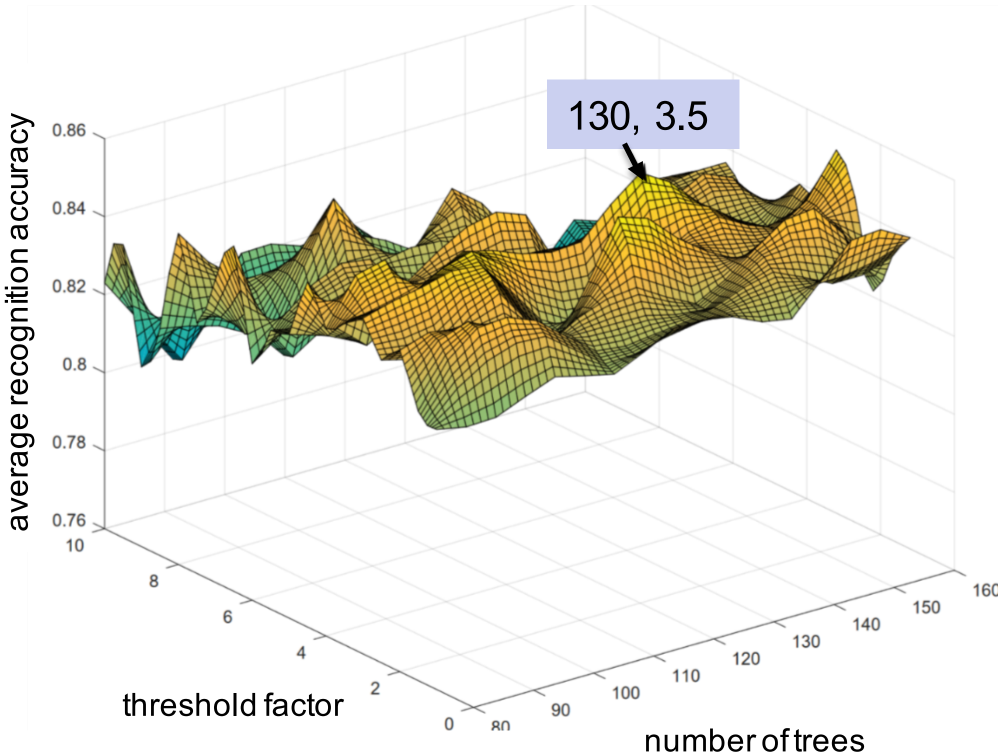
THE UNIVERSITY OF WESTERN AUSTRALIA

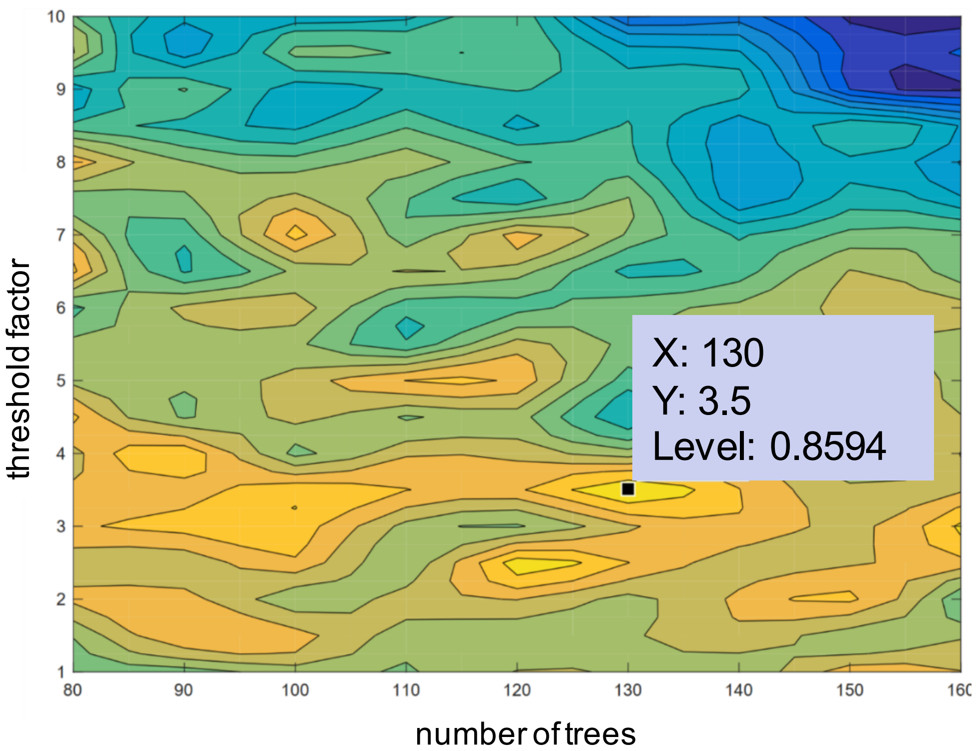
Feature importance normalization for HDG



- **Feature dimension reduction** using random decision forest

Optimization of Hyperparameters for HDG

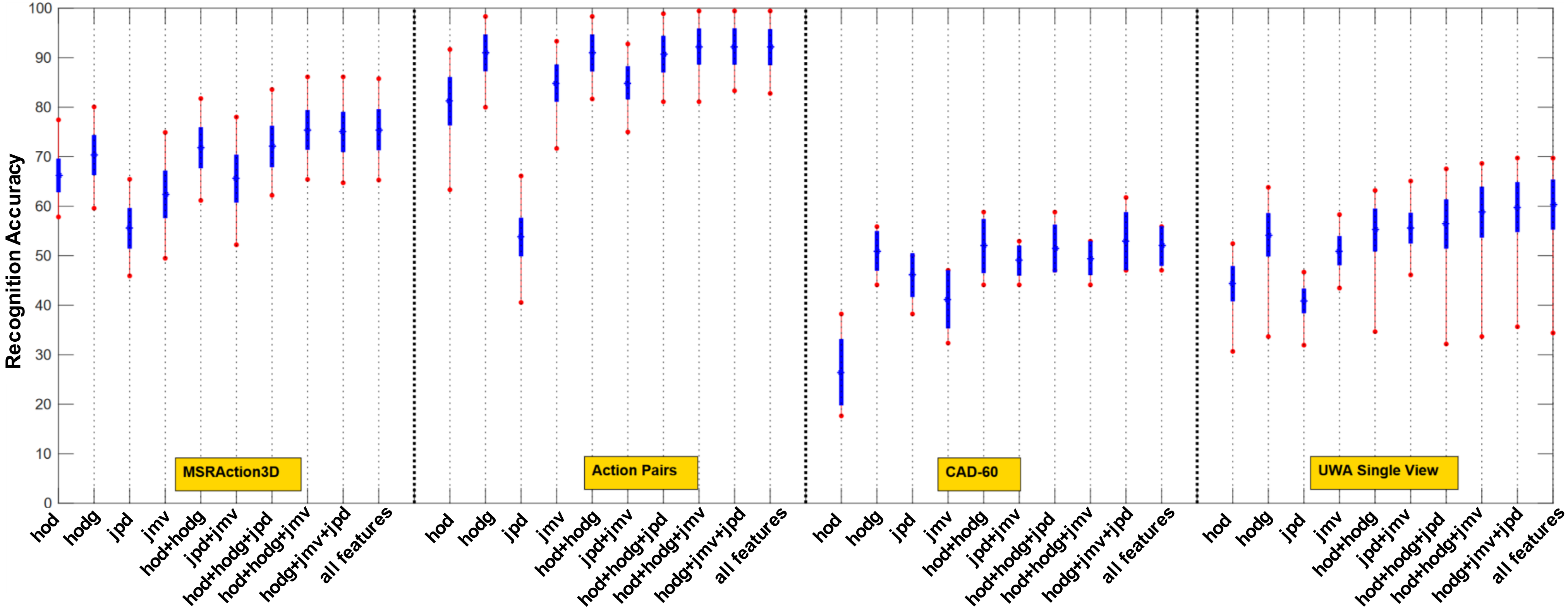




- Involving 2 hyperparameters: *number of trees* and *threshold factor*

THE UNIVERSITY OF WESTERN AUSTRALIA

Results and Discussions for the first 4 datasets



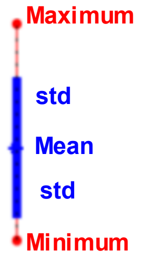
hod = histograms of depth

hodg = histograms of depth derivatives

jpd = joint position differences

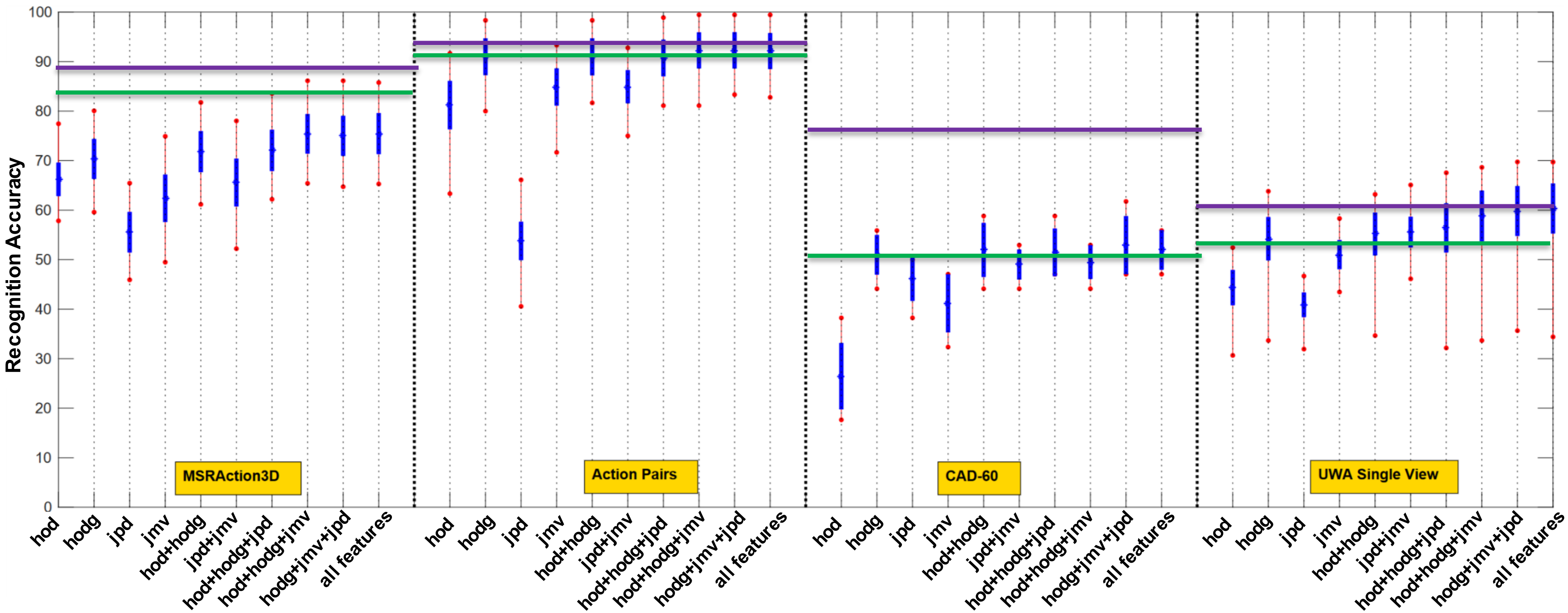
jmv = joint movement volume features

HDG:



THE UNIVERSITY OF WESTERN AUSTRALIA

Results and Discussions for the first 4 datasets



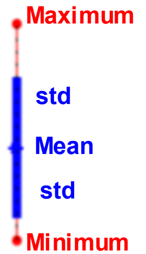
hod = histograms of depth

hodg = histograms of depth derivatives

jpd = joint position differences

jmv = joint movement volume features

HDG:



RBD-FTP representation:

HOPC:

Results and Discussions for the UWA3D Multiview Dataset

THE UNIVERSITY OF WESTERN AUSTRALIA

Training view	V_1 & V_2		V_1 & V_3		V_1 & V_4		V_2 & V_3		V_2 & V_4		V_3 & V_4		Mean
Testing view	V_3	V_4	V_2	V_4	V_2	V_3	V_1	V_4	V_1	V_3	V_1	V_2	
HON4D	31.1	23.0	21.9	10.0	36.6	32.6	47.0	22.7	36.6	16.5	41.4	26.8	28.9
HOPC	25.7	20.6	16.2	12.0	21.1	29.5	38.3	13.9	29.7	7.8	41.3	18.4	22.9
Holistic HOPC*	32.3	25.2	27.4	17.0	38.6	38.8	42.9	25.9	36.1	27.0	42.2	28.5	31.8
Local HOPC+STK-D*	52.7	51.8	**59.0**	**57.5**	42.8	44.2	58.1	38.4	63.2	43.8	66.3	48.0	52.2
RBD-logarithm map	48.2	47.4	45.5	44.9	46.3	52.7	62.2	46.3	57.7	45.8	61.3	40.3	49.9
RBD-unwrapping while rolling	50.4	45.7	44.0	44.5	40.8	49.6	57.4	44.4	57.6	47.4	59.2	40.8	48.5
RBD-FTP representation	54.9	55.9	50.0	54.9	48.1	56.0	66.5	**57.2**	62.5	54.0	68.9	43.6	56.0
HDG-hod	22.5	17.4	12.5	10.0	19.6	20.4	26.7	13.0	18.7	10.0	27.9	17.2	18.0
HDG-hodg	26.9	34.2	20.3	18.6	34.7	26.7	41.0	29.2	29.4	11.8	40.7	28.8	28.5
HDG-jpd	36.3	32.4	31.8	35.5	34.4	38.4	44.2	30.0	44.5	33.7	44.4	34.0	36.6
HDG-jmv	57.2	59.3	**59.3**	54.3	56.8	50.6	63.4	52.4	65.7	53.7	67.7	56.9	58.1
HDG-hod+hodg	26.6	33.6	17.9	19.3	34.4	26.2	40.5	27.6	28.6	11.6	38.4	29.0	27.8
HDG-jpd+jmv	**61.0**	61.8	**59.3**	**56.0**	60.0	57.4	68.8	54.2	**71.1**	**57.2**	69.7	59.0	**61.3**
HDG-hod+hodg+jpd	31.0	43.5	25.7	21.4	45.9	31.1	53.2	35.7	38.0	11.6	49.7	38.3	35.4
HDG-hod+hodg+jmv	59.0	**62.2**	58.1	52.0	62.5	57.1	66.0	54.2	67.7	52.7	**70.3**	**61.1**	60.2
HDG-hodg+jpd+jmv	58.2	61.8	54.8	47.6	**63.5**	**58.7**	**69.0**	52.3	64.9	47.1	67.2	59.4	58.7
HDG-all features	**60.9**	**64.3**	57.9	54.6	**62.6**	**59.2**	**68.9**	**55.8**	**69.8**	**55.2**	**71.8**	**62.6**	**61.9**

*This result is obtained from original authors' paper for comparison.

hod = histograms of depth

hodg = histograms of depth derivatives

jpd = joint position differences

jmv = joint movement volume features

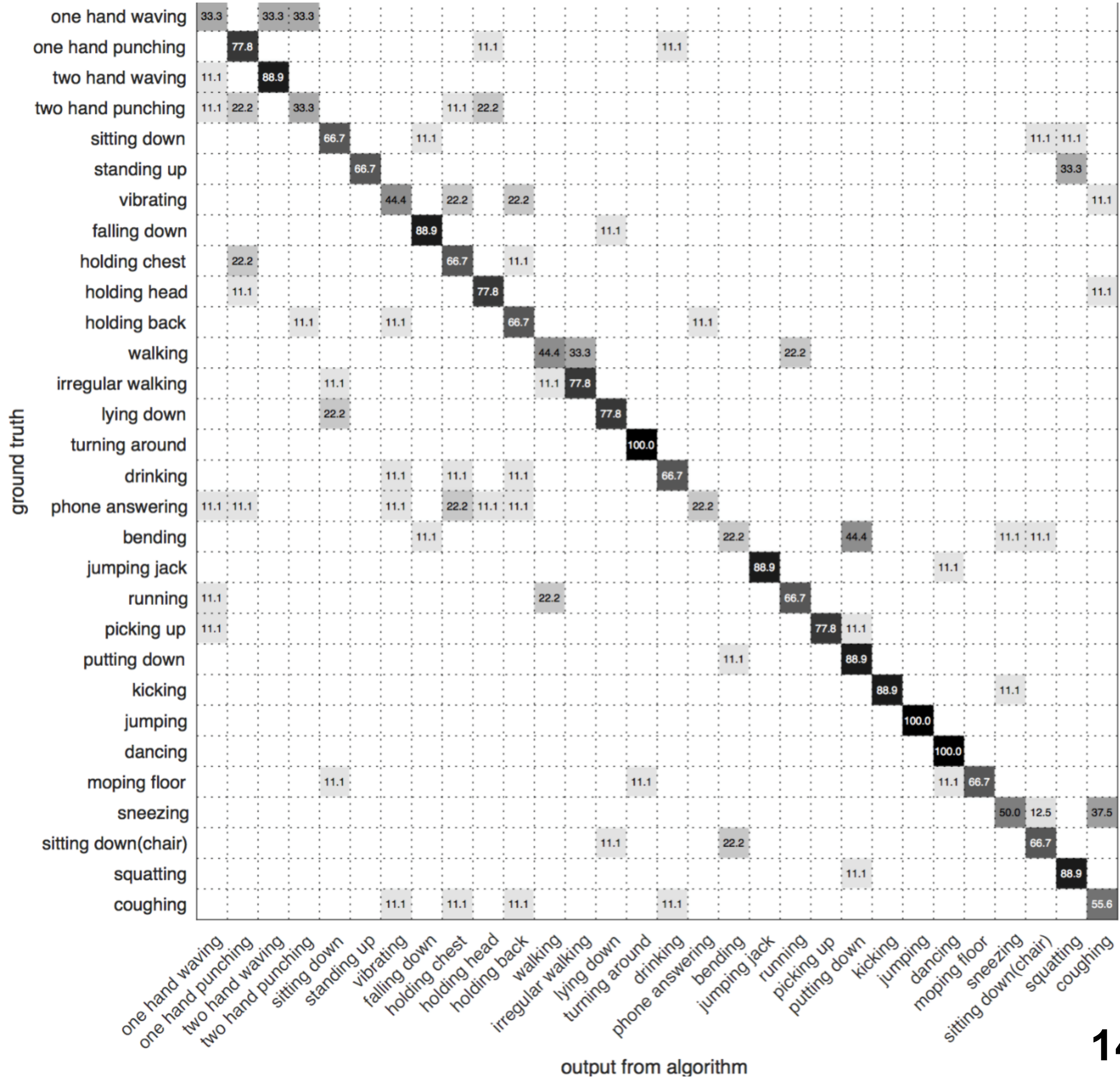
View 1 (**V_1**):	Front view
View 2 (**V_2**):	Left view
View 3 (**V_3**):	Right view
View 4 (**V_4**):	Top view

THE UNIVERSITY OF WESTERN AUSTRALIA

Confusion Matrix for the UWA3D Multiview Dataset

View 1 ($\mathbf{V}_1$):	Front view
View 2 ($\mathbf{V}_2$):	Left view
View 3 ($\mathbf{V}_3$):	Right view
View 4 ($\mathbf{V}_4$):	Top view

HDG-all features on the UWA3D Multiview Dataset when V3 and V4 are used for training and V1 is used for testing



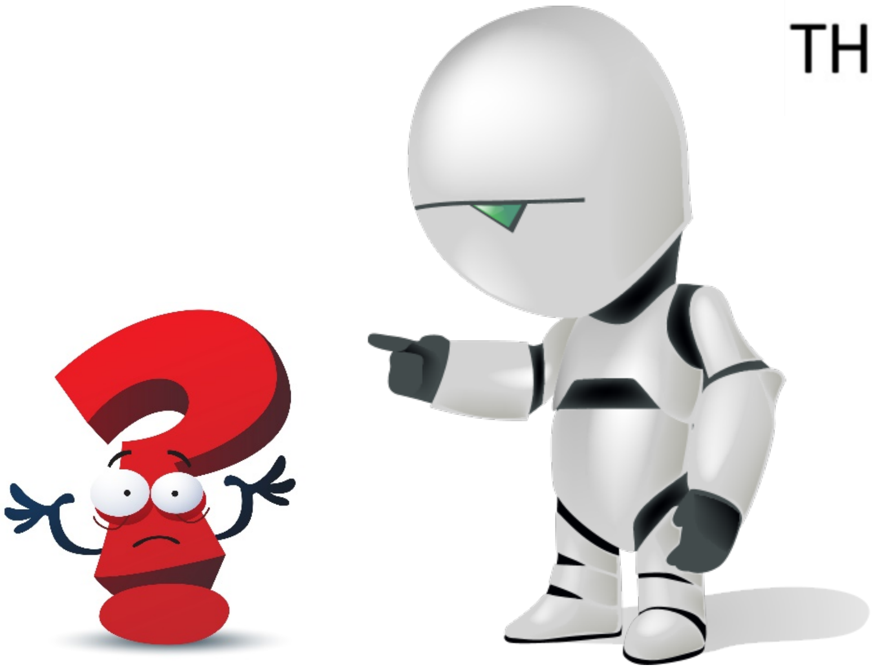
Conclusion and Future Work

- Skeleton features are more robust for cross-view action recognition.
- HDG-all features performs better than other state-of-the-art approaches for cross-view action recognition.
- HOPC and RBD is more robust to noise, human body size and action speed variations

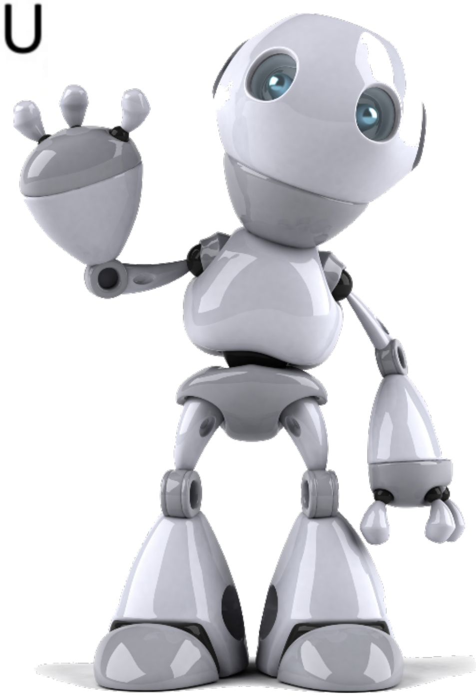
- Future work: build a convolutional neural network (CNN) architecture to make it easier, faster and more robust than existing approaches in dealing with challenging issues.

Acknowledgment

- I am grateful to A/Prof Du Huynh for her valuable suggestions and discussions.
- I am also thank the authors of HON4D, HOPC and RBD for making their codes publicly available.



THANK YOU



References

- O. Oreifej and Z. Liu. HON4D: Histogram of Oriented 4D Normals for Activity Recognition from Depth Sequences. In *CVPR*, pages 716–723, 2013.
- H. Rahmani, A. Mahmood, D. Q. Huynh, and A. Mian. Real Time Action Recognition Using Histograms of Depth Gradients and Random Decision Forests. In *WACV*, pages 626–633, 2014.
- H. Rahmani, A. Mahmood, D. Q. Huynh, and A. Mian. HOPC: Histogram of Oriented Principal Components of 3D Pointclouds for Action RecognitionPrincipal Components of 3D Pointclouds for Action Recognition. In *ECCV*, pages 742–757, 2014.
- H. Rahmani, A. Mahmood, D. Huynh, and A. Mian. Histogram of Oriented Principal Components for Cross-View Action Recognition. *TPAMI*, pages 2430–2443, December 2016.
- R. Vemulapalli and R. Chellappa. Rolling Rotations for Recognizing Human Actions from 3D Skeletal Data. *CVPR*, pages 4471–4479, 2016.
- R. Vemulapalli, F. Arrate, and R. Chellappa. Human Action Recognition by Representing 3D Skeletons as Points in a Lie Group. *CVPR*, pages 588–595, 2014.