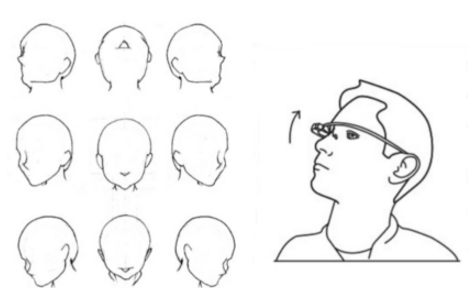
GlassGesture: Exploring Head Gesture Interface of Smart Glasses

WILLIAM & MARY
CHARTERED 1693

Shanhe Yi, Zhengrui Qin, Ed Novak, Yafeng Yin, and Qun Li
College of William and Mary



Introduction

Google Glass users use touchpad gestures and voice commands to interactive with the device to fulfil tasks like taking picture, selecting menus, navigating the content on screen display.

However, this user interface on Google Glass is less than ideal:

- Gestures on the touchpad is error-prone
- Voice cannot be applied in every scenario.
- Authentication on the touchpad is hard to use.

In GlassGesture, we uses simple gestures (Fig. 3) to control the device. The user can enter alphabets and digits by “drawing” it using the Glass. We also use those gestures as “password” to authenticate users by extracting unique features from those movements.

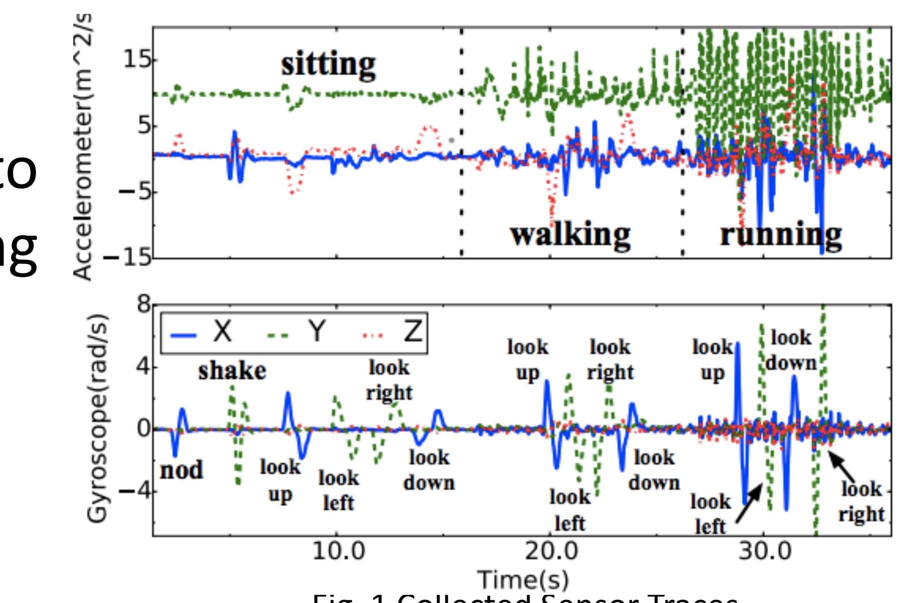


Fig. 1 Collected Sensor Traces

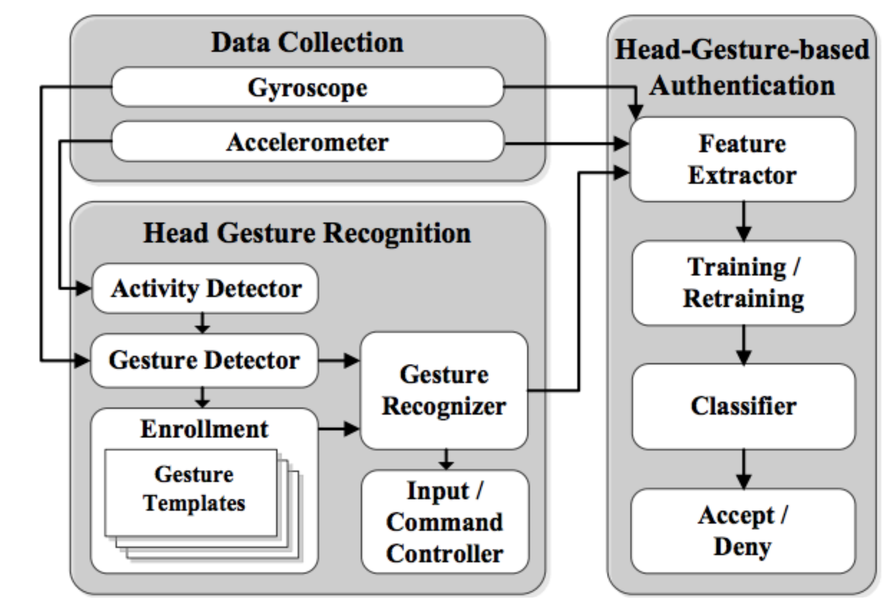


Fig. 2 System Architecture

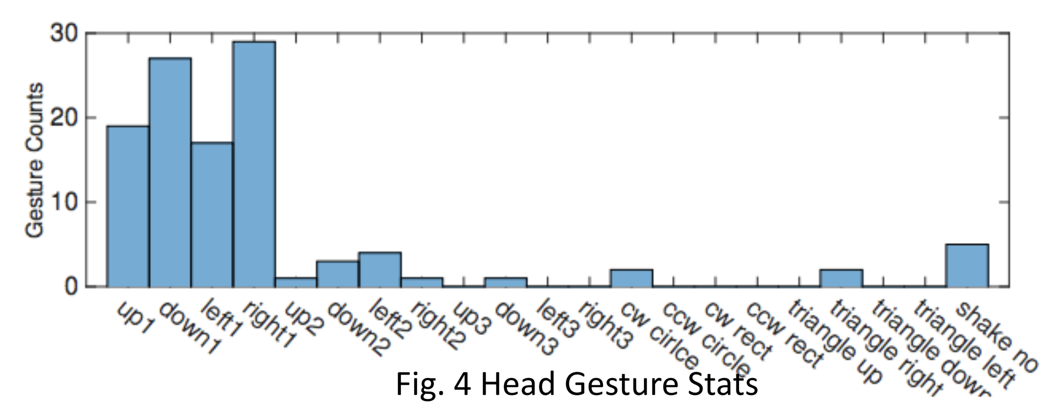
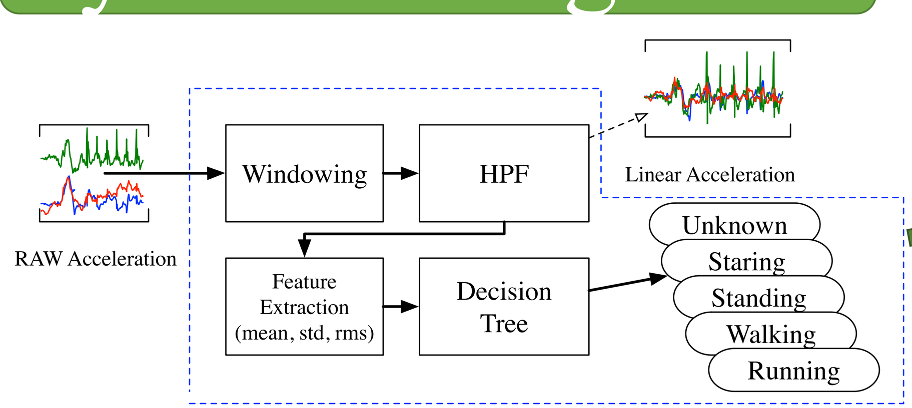


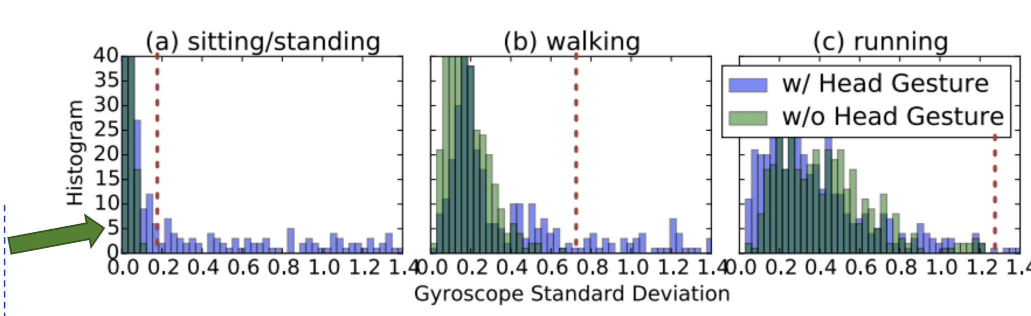
Fig. 4 Head Gesture Stats

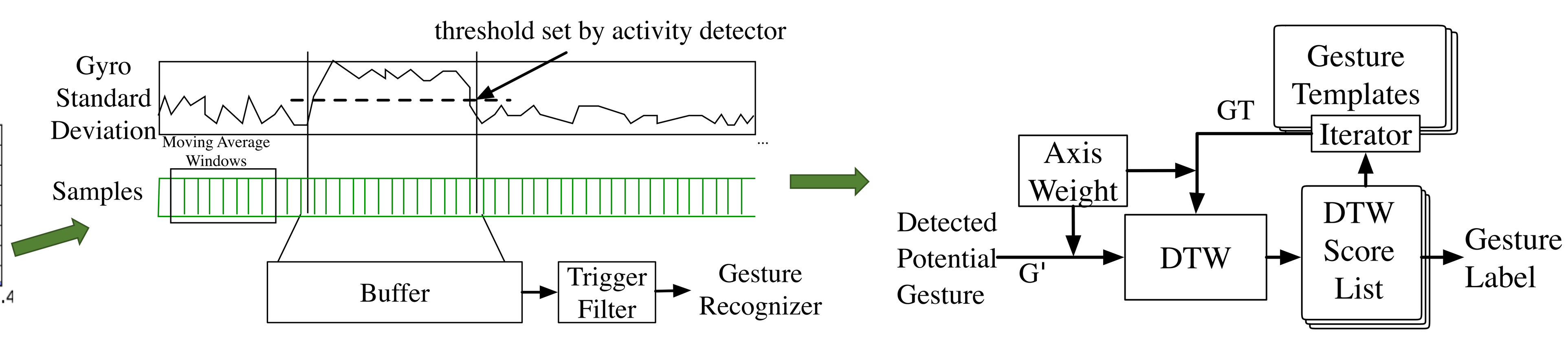
Gesture	Styles	Number of strokes	Easy to perform	Frequency in Fig. 4	Easy to repeat	Decision
1	up and down	3+	5.2	low	no	keep
2	up/down/left/right	1	4.9	high	yes (81%)	keep,repeat
3	left and right	3+	4.4	low	no	keep
4	cw/ccw	1	3.0	very low	neutral	keep
5	cw/ccw, directions	3	2.2	very low	no	drop
6	cw/ccw, start points	4	1.4	very low	no	drop

Fig. 3 Head Gesture candidates

System Design







Efficient Similarity Search: We use DTW to match templates to recognize gestures. DTW is pair-wise which is not efficient when there is too many templates. To reduce the complexity, we build a k-d tree for kNN search.

Algorithm 1 Build KD-Tree

1: **procedure** BUILD KDTREES($\mathbf{T}$, n_{ED})
2: **for** each template t in $\mathbf{T}$ **do**
3: downsampling to length-n_{ED}
4: stored in $\mathbf{T}_{down}$.
5: **end for**
6: Build a KD Tree from $\mathbf{T}_{down}$ using Euclidean distance, as $\mathbf{Tr}$
7: **end procedure**

Algorithm 2 kNN search.

1: **procedure** KNN SEARCH($\mathbf{Tr}$, t, k)
2: put k nearest neighbors of target t in tree $\mathbf{Tr}$ into $\mathbf{C}$.
3: **for** each candidate in $\mathbf{C}$ **do**
4: run DTW on target and candidate.
5: **end for**
6: return index of minimal DTW distances
7: **end procedure**

Head Gesture based Authentication:

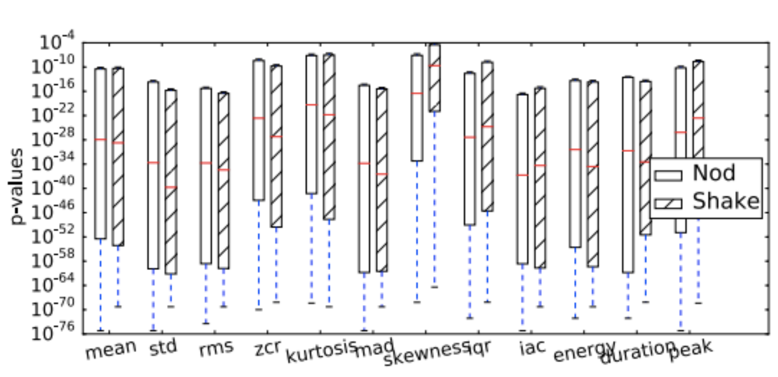
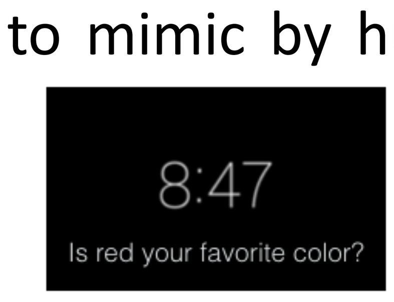


Fig. 3 K-S test of selected features

A “yes-or-no” question on the near-eye display will be answered using head gestures. Selected features will be extracted and put into the trained classifier.

This scheme has several advantages:

- no need to remember anything.
- taking minimal effort from user.
- more secure, hard to mimic by human eye observing.



Classification:

- Use an ensemble one-class SVM
- Grid search with constraint on the false positive rate to find a set of parameters
- Majority voting on the classification decisions from multiple SVM instances.

Evaluation

We have evaluated the system in terms of gesture recognition in different activities, gesture recognition performance and authentication performance. We have collected 6000 samples from 18 users for head gesture authentication evaluation.

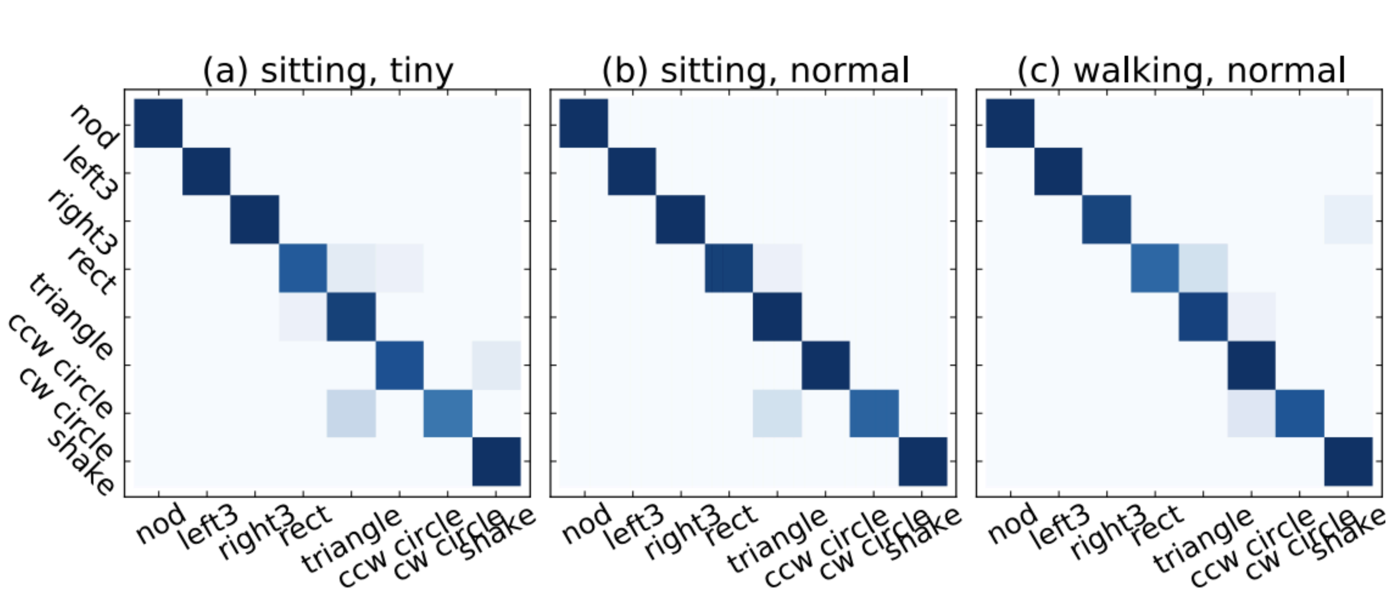


Fig. 5 (a) Confusion matrix of command gestures (sitting, tiny). TPR: 92.87%, FPR: 5.7%. (b) Confusion matrix of command gestures (sitting, normal). TPR: 96.99%, FPR: 2.4%. (c) Confusion matrix of command gestures (walking, normal). TPR: 94.88%, FPR: 4.6%

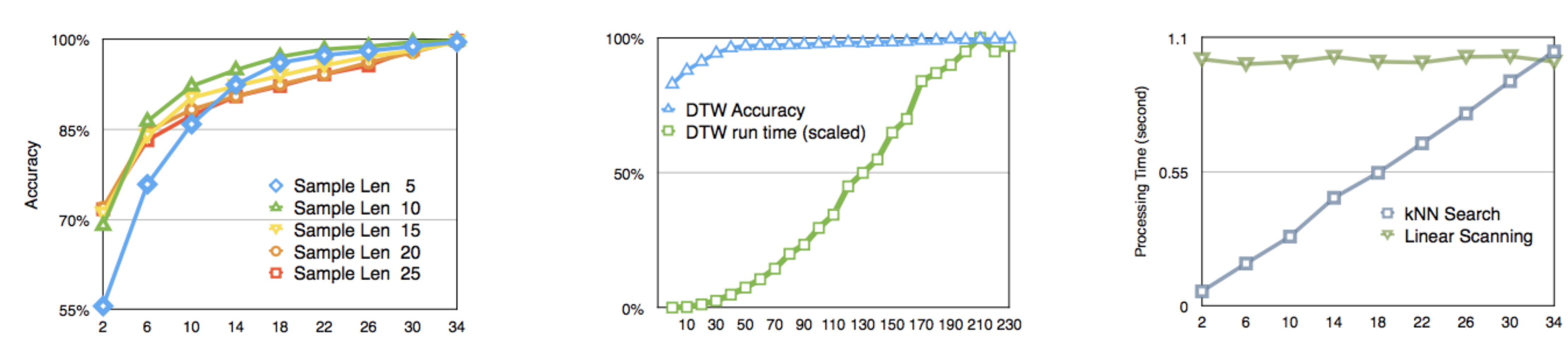


Fig. 6 (left) Accuracy changes with sampling lengths and numbers of nearest neighbors. Fig. 4 (mid) Accuracy and running time using DTW change with sampling length. Fig. 4 (right) running time comparison between our scheme and linear scanning. The running time will be reduced by 55% when n_{ED}=10, n_{dtw}=50 and k=14.

Single	TPR	FPR
GlassGesture Nod	92.43% (+/-3.09)	0.09% (+/-0.22)
GlassGesture Shake	92.33% (+/-3.32)	0.17% (+/-0.33)
GlassGesture Left3	89.08% (+/-6.36)	0.48% (+/-0.79)
GlassGesture Right3	89.61% (+/-5.99)	0.52% (+/-0.87)
Multiple and Comparison	**TPR**	**FPR**
GlassGesture (2 gestures)	99.16%	0.61%
Touchpad+ Voice (5 events) [1]	97.14%	1.27%
Touchscreen GEAT (3 gestures) [2]	98.2%	1.1%

Table. 1 The TPR and FPR of authentication of different gestures and comparison of other works in combining multiple instances.

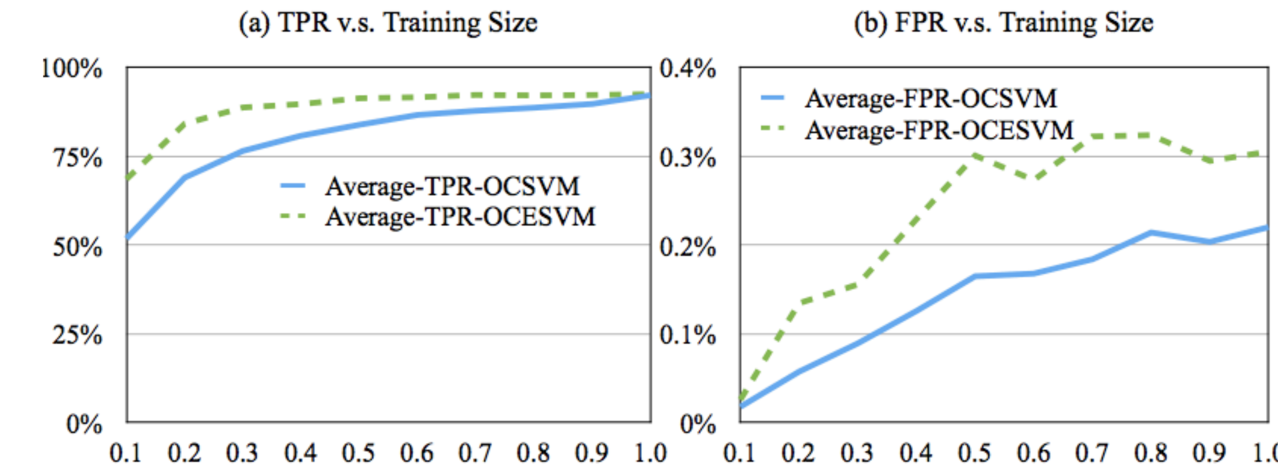


Fig. 7 The average TPR and FPR change with different ratios of training samples. One-class ensemble SVM (OCESM) requires less training samples with higher TPR and the cost of a small FPR loss.

GlassGesture: Exploring Head Gesture Interface of Smart Glasses , Shanhe Yi, Zhengrui Qin, Ed Novak, Yafeng Yin, and Qun Li, *The 35th Annual IEEE International Conference on Computer Communications (INFOCOM)* , San Francisco, CA, 2016.