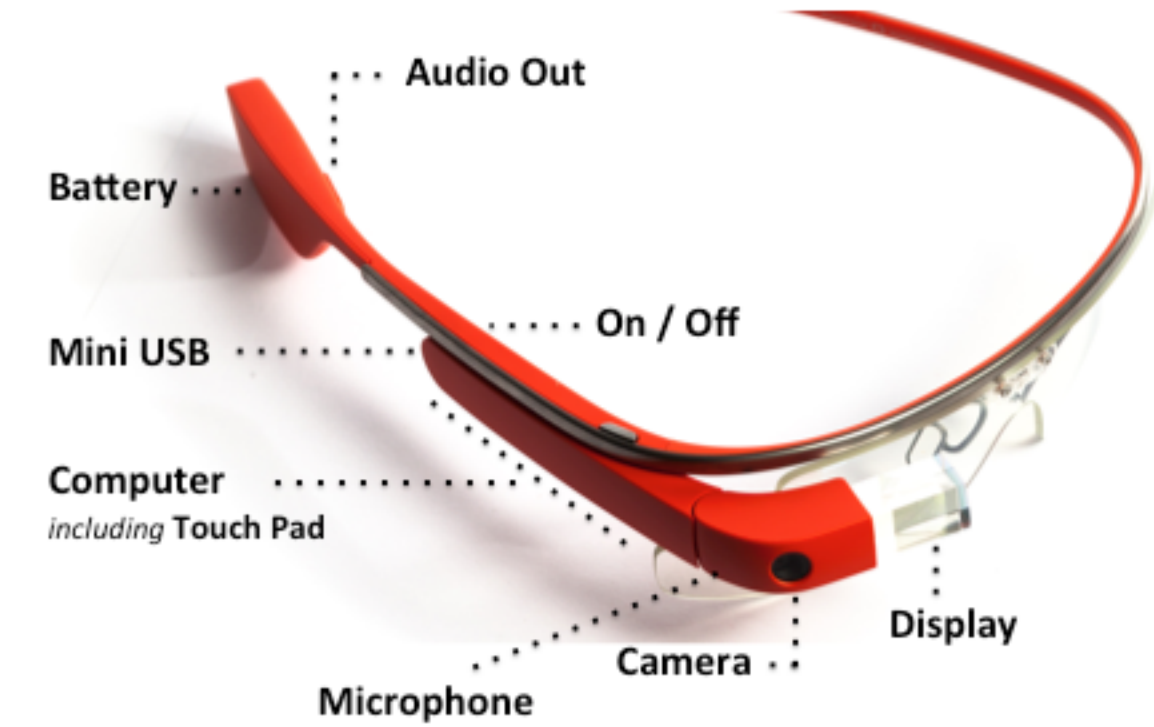
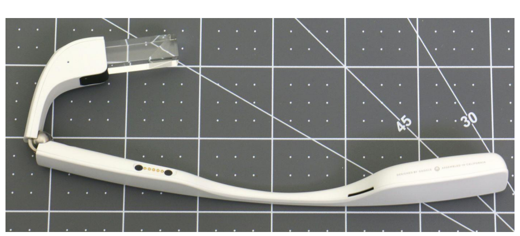
GlassGesture: Exploring Head Gesture Interface of Smart Glasses

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

WILLIAM & MARY
CHARTERED 1693

Smart Glasses - Some setbacks, but still a promising market





WILLIAM & MARY

Limitations of Current Input Methods of Google Glass

- Voice recognition
  - unpredictable result
  - embarrassing using it outdoor/one-on-one meeting/conference
  - pick up louder voices from others than the wearer

Microphone

Limitations of Current Input Methods of Google Glass

- Touchpad gesture
  - not hands-free, need to hold your hand to the side of forehead, hard to use in walking or driving
  - non-intuitive tapping and swiping, error-prone
  - narrow and slim, some gestures can be confused

Touchpad

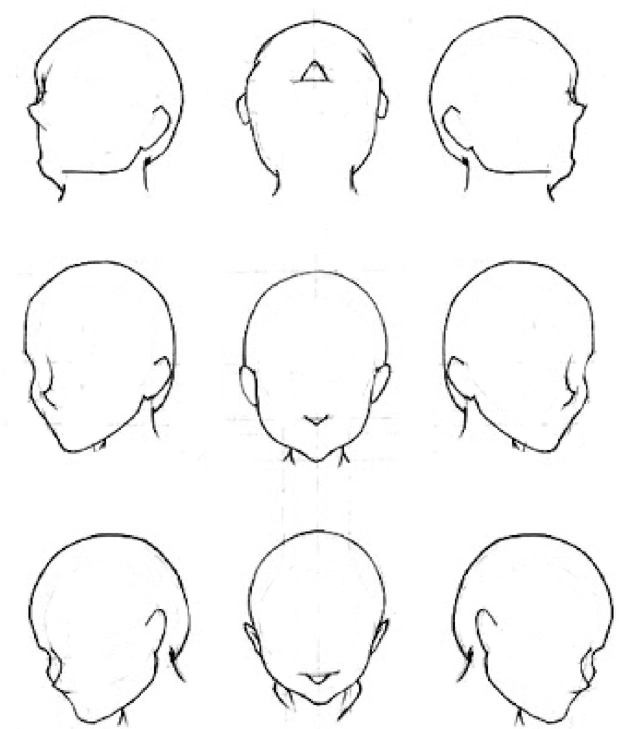
Limitations of Current Input Methods of Google Glass

- Authentication based on Touchpad gesture Input
    - limited combinations
    - some gestures are difficult to perform correctly on that touchpad
    - this “password” is hard to remember
    - susceptible to shoulder-surfing attack

Available gestures

- Swipe forward
- Two-finger swipe forward
- Swipe back
- Two-finger swipe back
- Hook swipe - Swipe forward then back in the same motion. Back and then forward works too.
- Two-finger swipe.
- Tap
- Two-finger tap

Head Gesture Interface



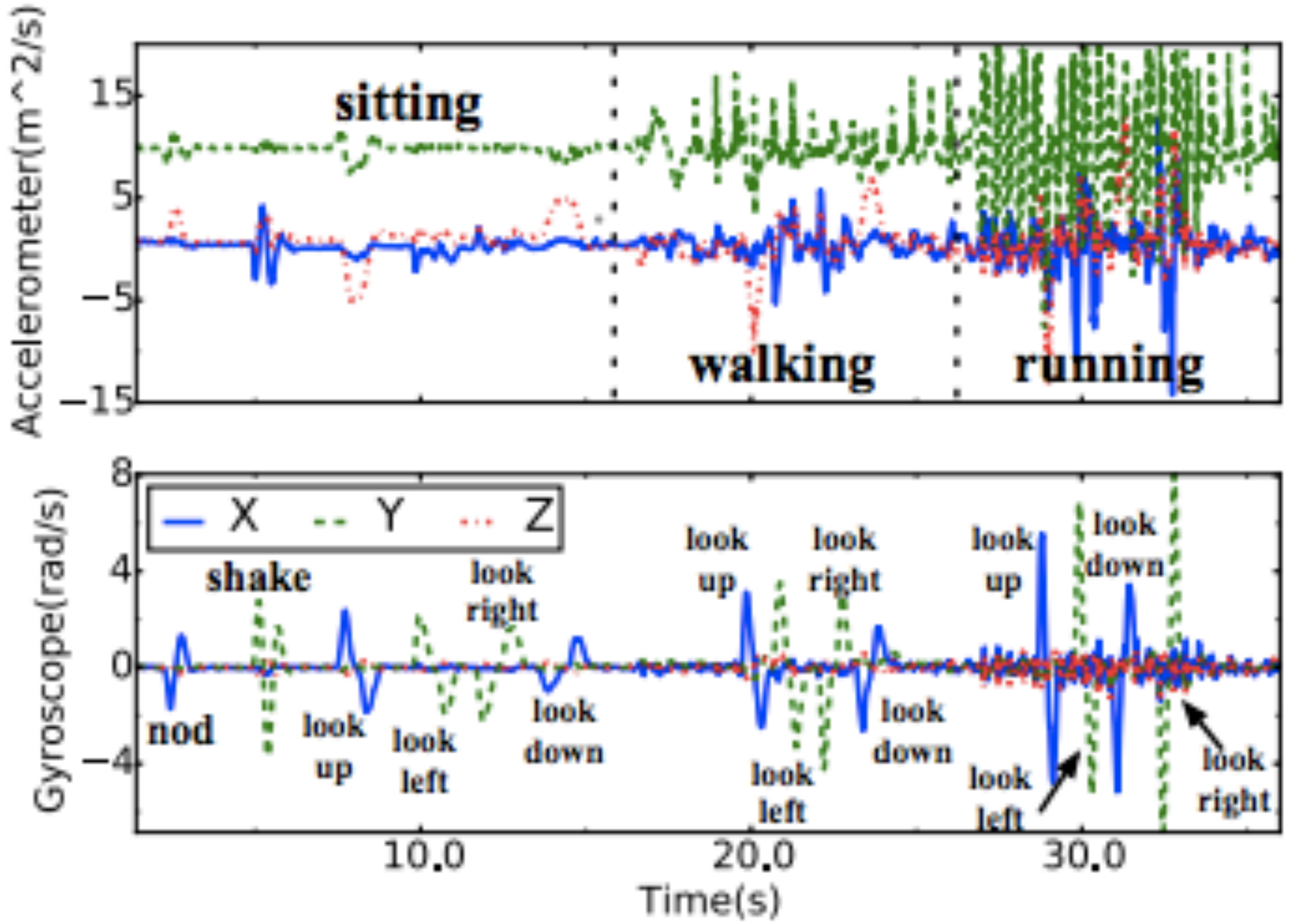
- intuitive
- minimal user effort (HCI design principle)
    - hands-free
    - easy-to-use
- can be used as password

A head gesture is short burst of several discrete or consecutive movements of user's head.

Contributions

- Increased the input space of Google Glass by enabling small, easy-to-perform head gestures
- Investigated gesture library exclusively for head movements
- speed up gesture template matching with a novel scheme
- design a two-factor authentication scheme using head gestures
- collected 6000 samples from 18 users for extensive evaluations.

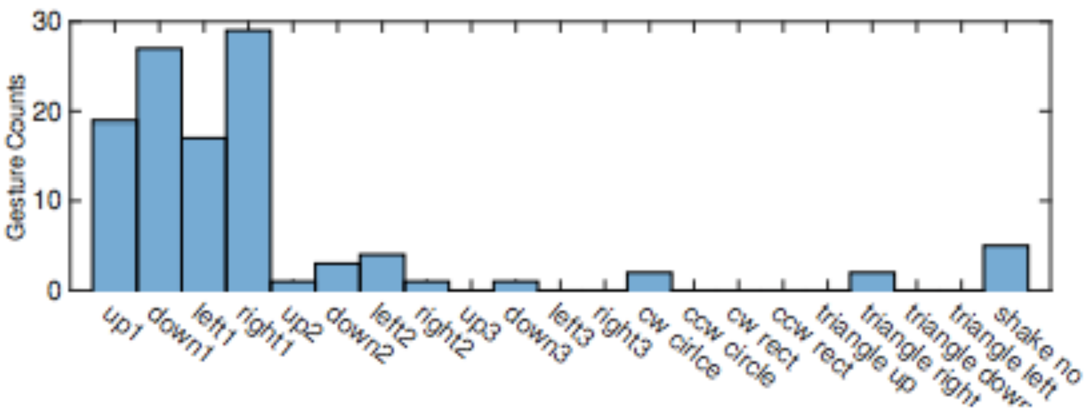
Observations and Challenges



- Different activities add different amount of noise.
- Different gestures show different distributions on three axis.
- Head gestures consist more rotations than accelerations.
- Head gestures are used rather frequently, need an efficient recognition scheme.

WILLIAM & MARY
CHARTERED 1693

Head Gesture Library



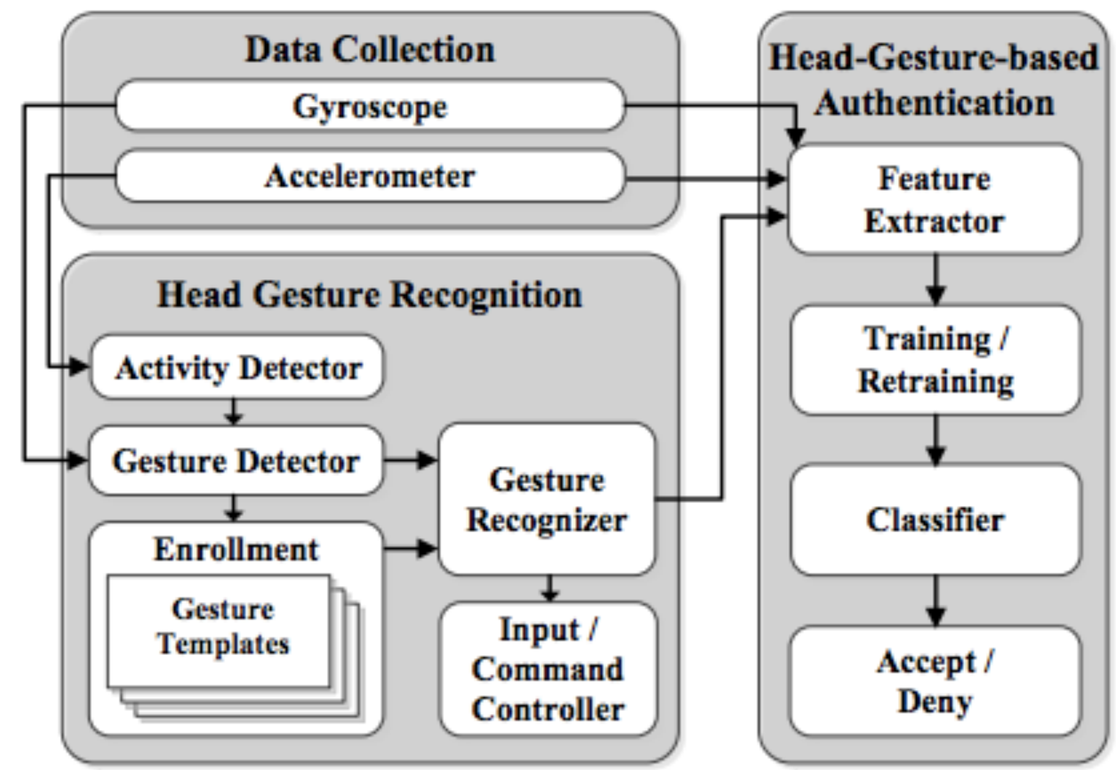
Gesture	Styles	Number of strokes	Easy to perform	Frequency in Fig. 4	Easy to repeat	Decision
1	up and down	3+	5.2	low	no	keep
2	up/down/left/right	1	4.9	high	yes (81%)	keep,repeat
3	left and right	3+	4.4	low	no	keep
4	cw/ccw	1	3.0	very low	neutral	keep
5	cw/ccw, directions	3	2.2	very low	no	drop
6	cw/ccw,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

Daily movements are mostly simple gestures, which we should avoid.

Through the survey, we identify several qualifiable gestures. And we repeat those daily head gestures for 3 times to make them useable. We also support alphabet and number input.

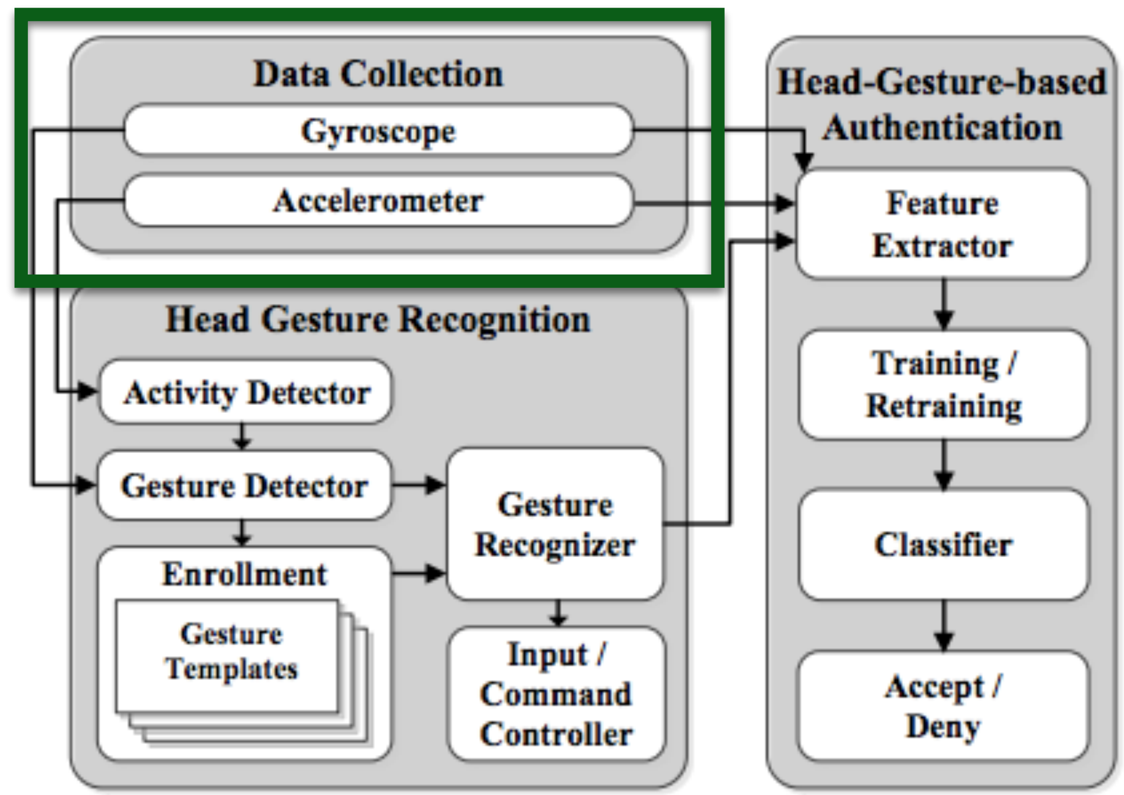
System Architecture

- Data Collection
- Head Gesture Recognition
- Head Gesture-based Authentication



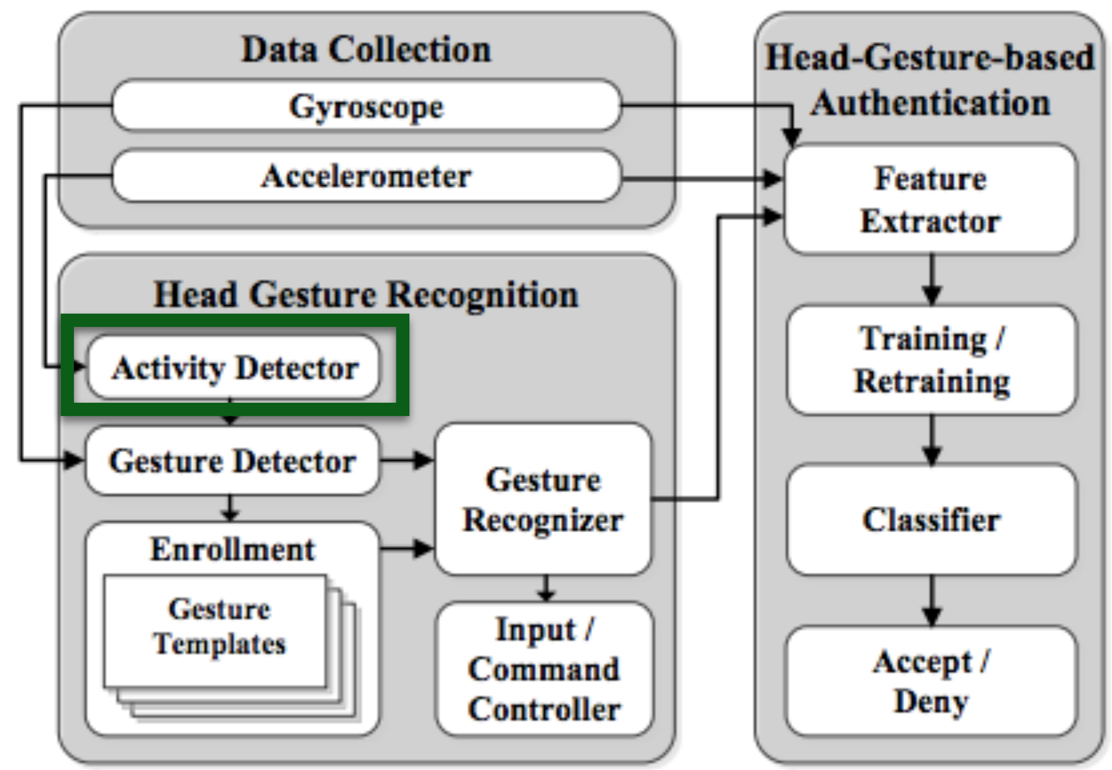
System Architecture

- Data Collection
- Head Gesture Recognition
- Head Gesture-based Authentication



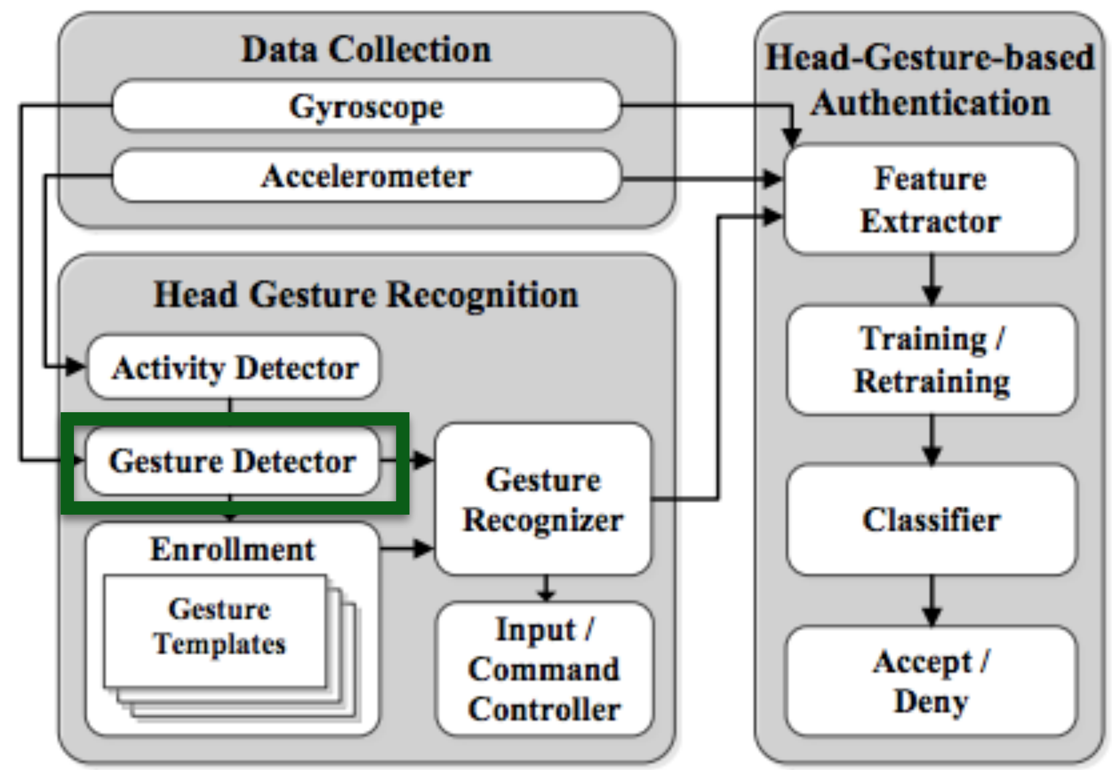
System Architecture

- Data Collection
- Head Gesture Recognition
- Head Gesture-based Authentication



System Architecture

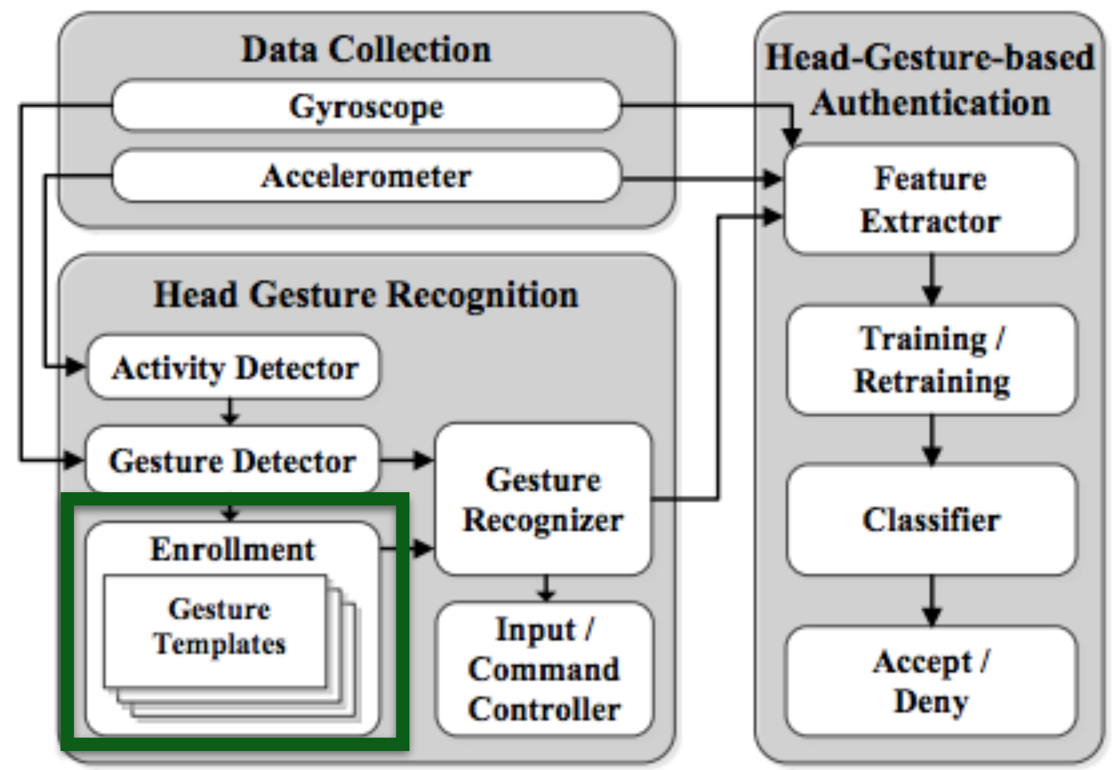
- Data Collection
- Head Gesture Recognition
- Head Gesture-based Authentication



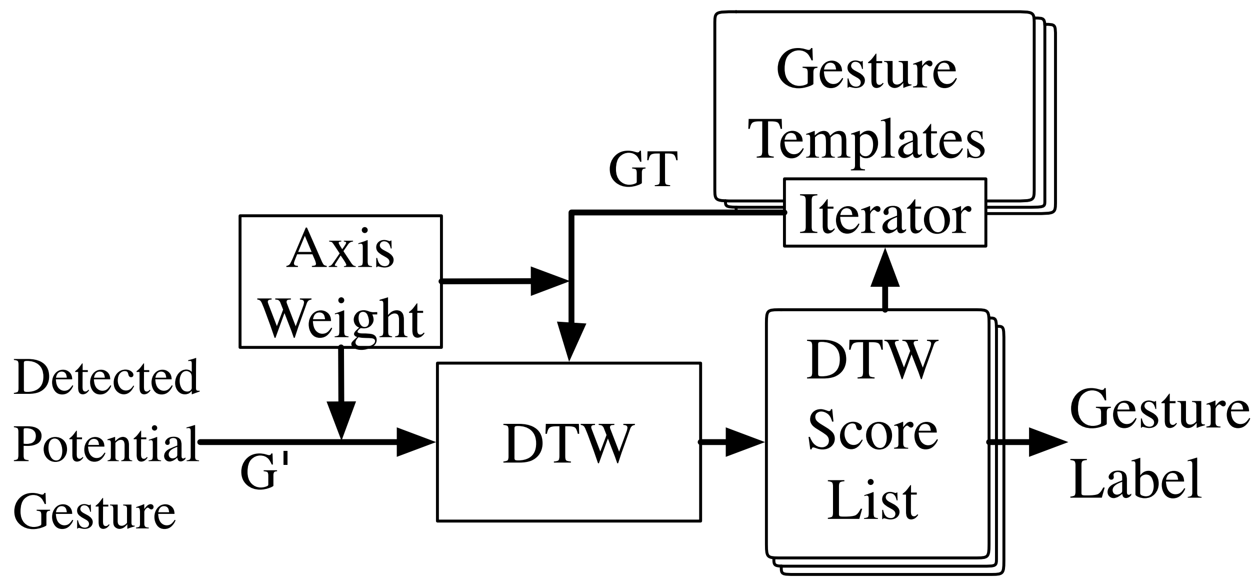
WILLIAM & MARY
CHARTERED 1693

System Architecture

- Data Collection
- Head Gesture Recognition
- Head Gesture-based Authentication



Gesture Recognizer



- To match gesture and templates, we use Dynamic Time Warping [Liu et. al. uWave].

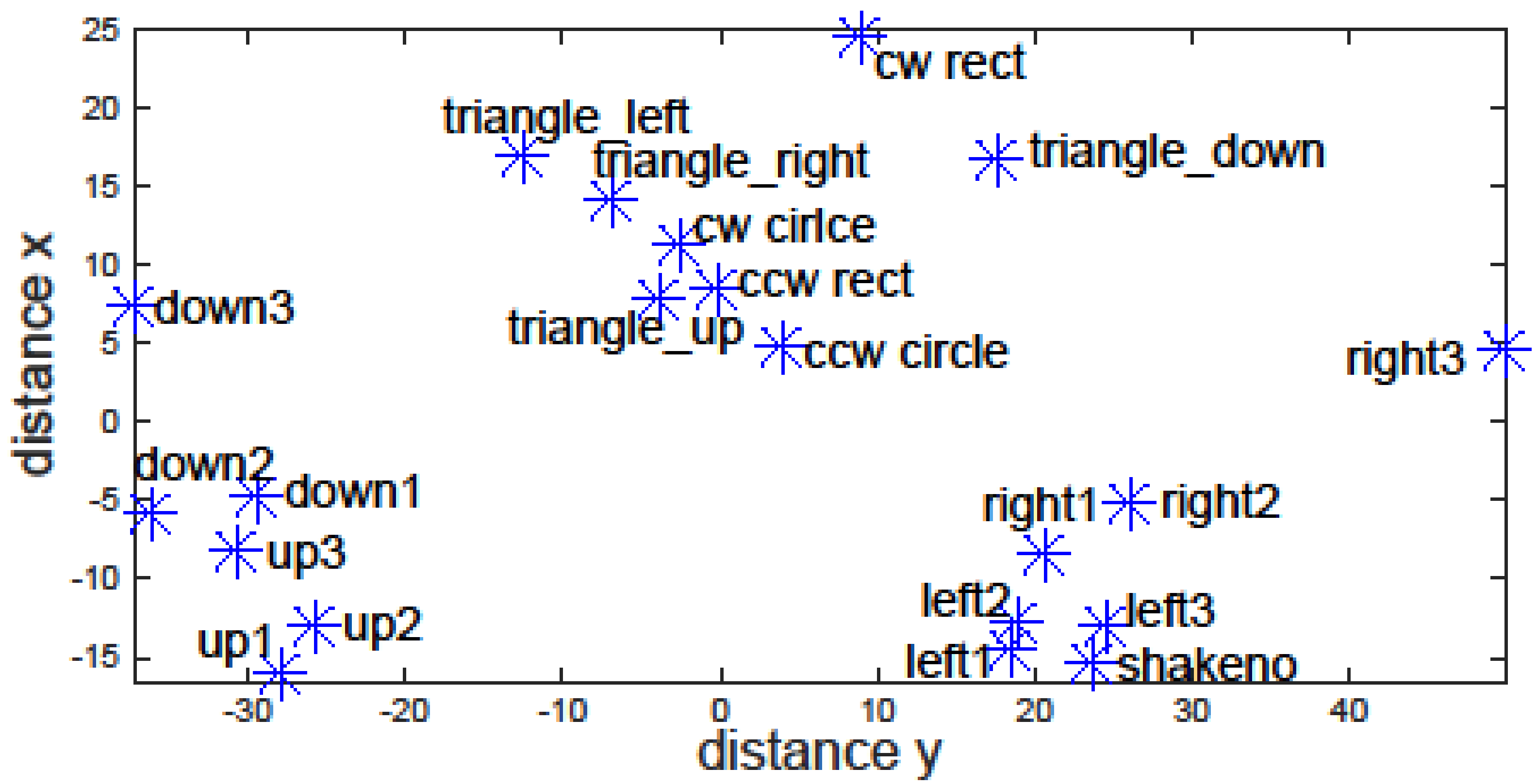
$$\mathrm{dtw}(G, Gt) = \sqrt{w_x D^2_{l,l_t}(x) + w_y D^2_{l,l_t}(y) + w_z D^2_{l,l_t}(z)}$$

- add weights to axis to better the accuracy

$$D_{i,j} = d(G(i), Gt(j)) + \min\{D_{i-1,j-1}, D_{i,j-1}, D_{i-1,j}\}$$

$$w_x = \frac{\mathrm{std}(Gt_x)}{\mathrm{std}(Gt_x) + \mathrm{std}(Gt_y) + \mathrm{std}(Gt_z)}$$

Improving Gesture Recognition Efficiency

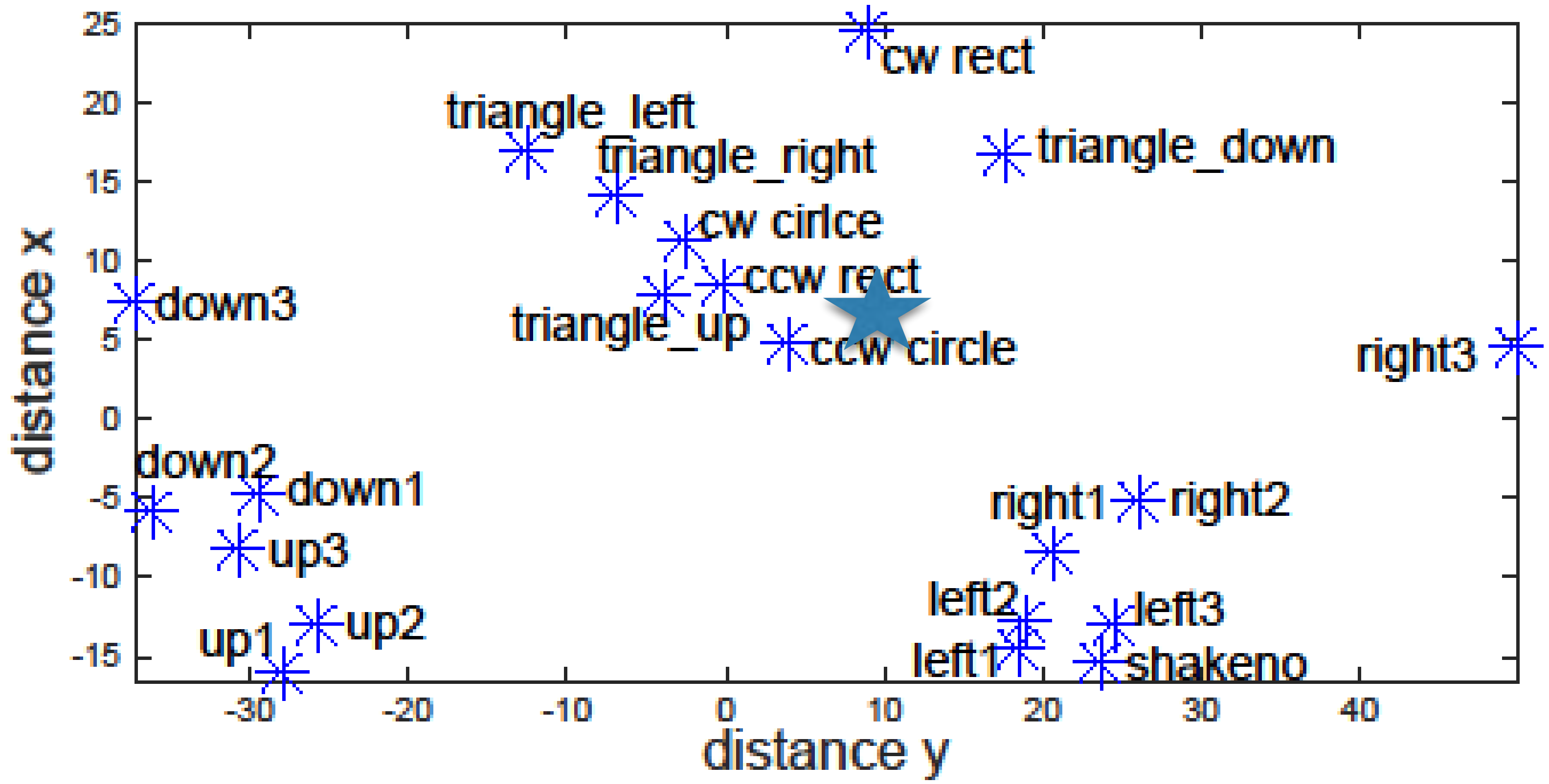


- Similarities in gesture templates: build a KD tree and perform kNN query on it.

- DTW distance is non-metric. There are errors if directly used in building KD tree: Use euclidean distance to build the KD Tree, use DTW distance for template matching in the result of kNN query.

- Down-sampling the samples used for KD tree build and DTW matching.

Improving Gesture Recognition Efficiency

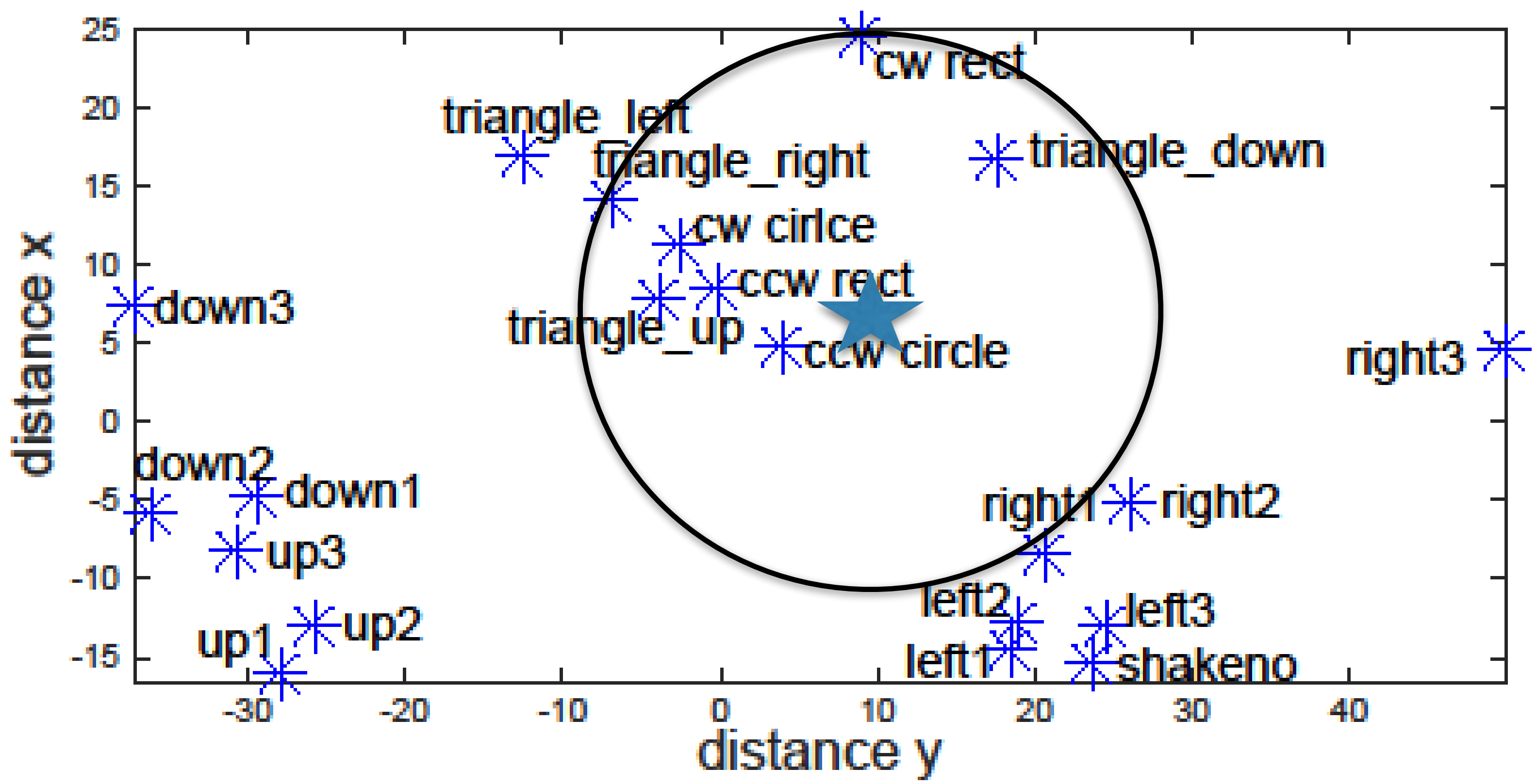


- Similarities in gesture templates: build a KD tree and perform kNN query on it.

- DTW distance is non-metric. There are errors if directly used in building KD tree: Use euclidean distance to build the KD Tree, use DTW distance for template matching in the result of kNN query.

- Down-sampling the samples used for KD tree build and DTW matching.

Improving Gesture Recognition Efficiency



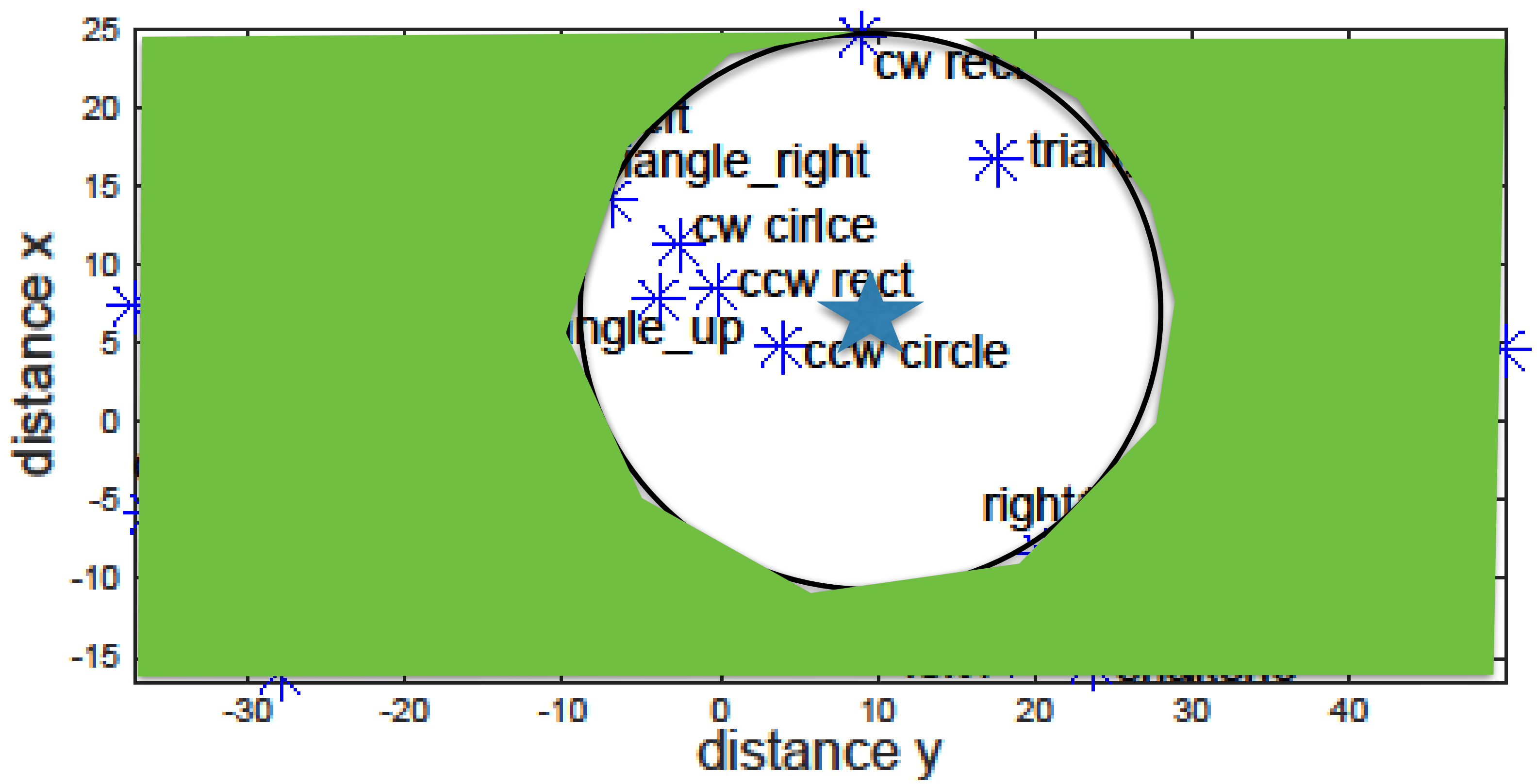
- Similarities in gesture templates: build a KD tree and perform kNN query on it.

- DTW distance is non-metric. There are errors if directly used in building KD tree: Use euclidean distance to build the KD Tree, use DTW distance for template matching in the result of kNN query.

- Down-sampling the samples used for KD tree build and DTW matching.

William & Mary
Chartered 1693

Improving Gesture Recognition Efficiency



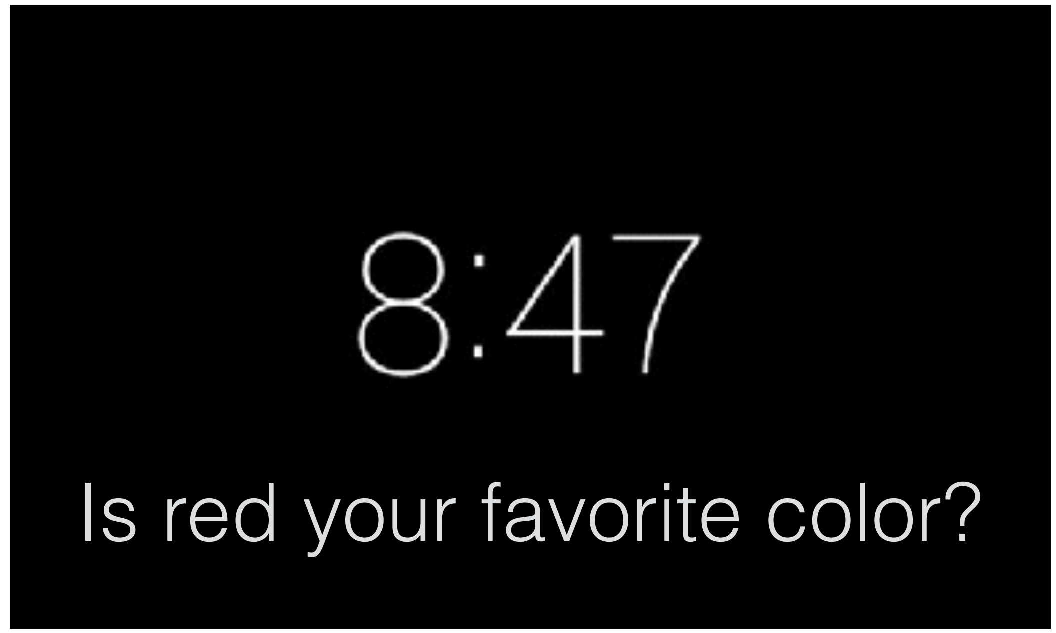
- Similarities in gesture templates: build a KD tree and perform kNN query on it.

- DTW distance is non-metric. There are errors if directly used in building KD tree: Use euclidean distance to build the KD Tree, use DTW distance for template matching in the result of kNN query.

- Down-sampling the samples used for KD tree build and DTW matching.

WILLIAM & MARY
CHARTERED 1693

Head Gesture based Authentication



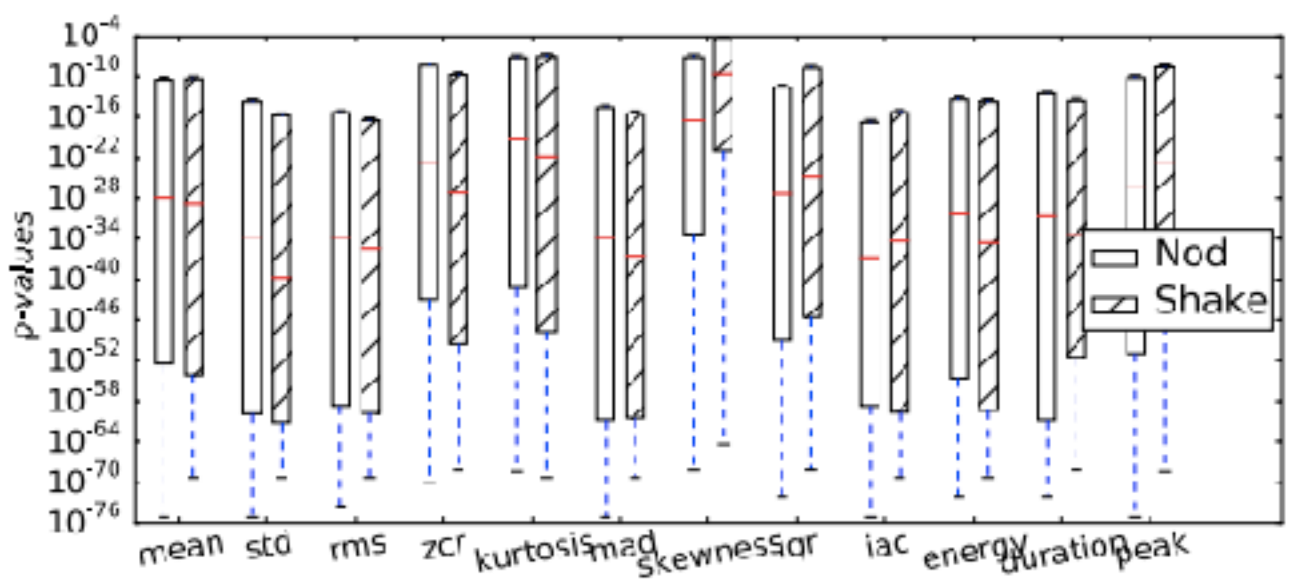
- private near-eye display showing yes-or-no questions
- user answers with head gestures (shaking/nodding)
- answer will be verified, feature will be extracted

prevent shoulder surfing attack, no need to remember anything, simple gestures take almost no effort, features are hard to observe by human eye

WILLIAM & MARY
CHARTERED 1693

Data Set and Feature Set

- Data set: 18 users (gender: m/f: 14/4; age: 20-30: 11, 30-40: 5, 40+: 2)
- Feature set
  - mean, std, rms, median absolute deviation(mad), zero-crossing rate(zcr), and inter-quartile range(iqr) , energy, duration, inter-axis correlation(iac) etc.
  - peak features: average peak-to-peak duration, average peak amplitude, number of peaks.



One-class classifier

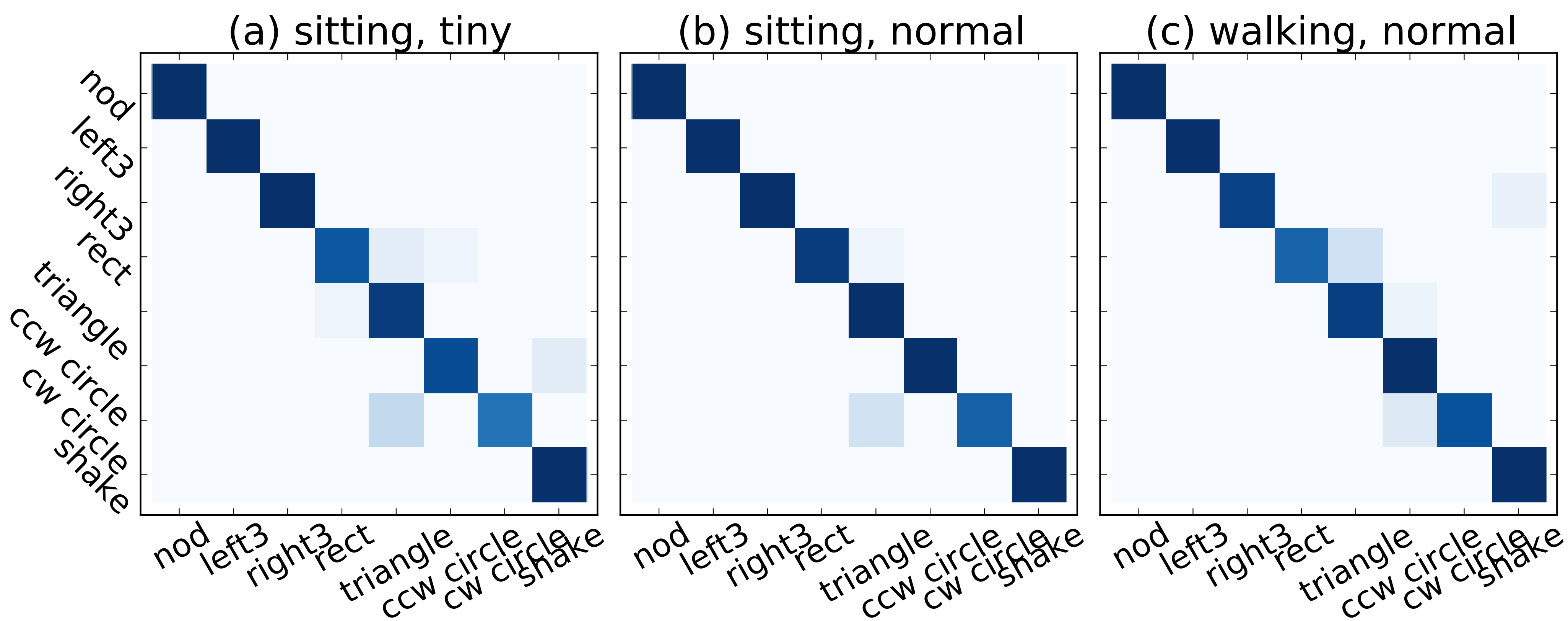
- one class SVM classifier (OCSVM):
  - RBF kernel
  - grid search on parameter space to find the best one
- one class ensemble SVM classifier (OCESVM):
  - RBF kernel
  - grid search on parameter space to find multiple instances with constraint FPR < 1%
  - majority votes from results of those top-r SVM instances

Feature Selection

- SVM Recursive feature elimination (RFE) [Guyon et. al., 2002]
  - recursively construct SVM models
  - remove low weight feature every step
  - needs to be multi-class
- Turn the one-class training set into a multi-class one
  - divide the training set into several groups evenly and manually
  - assign each group a different virtual class label
- Apply RFE, get feature ranking
  - eliminate features ranked in the top (e=3).

WILLIAM & MARY
CHARTERED 1693

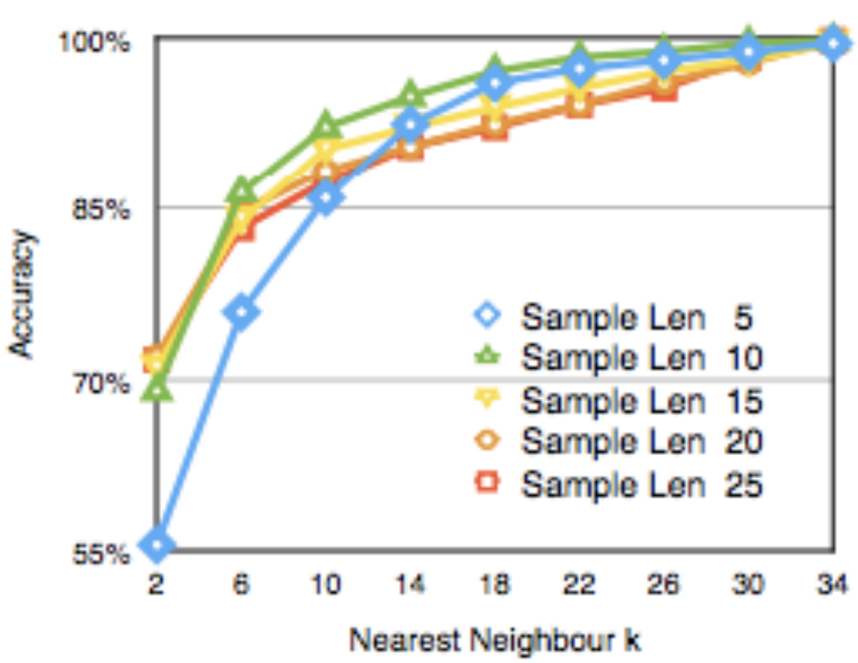
Evaluation - Gesture Recognition Accuracy

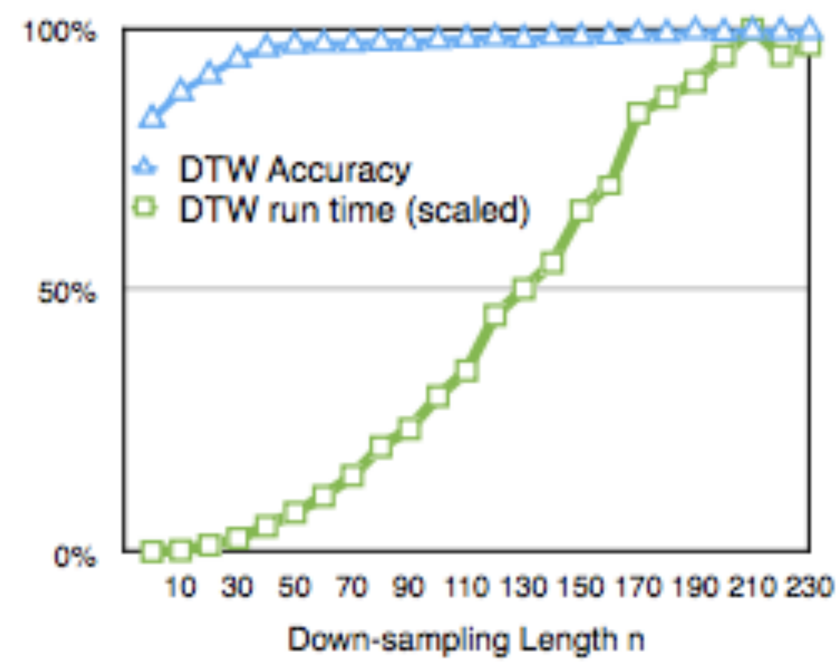


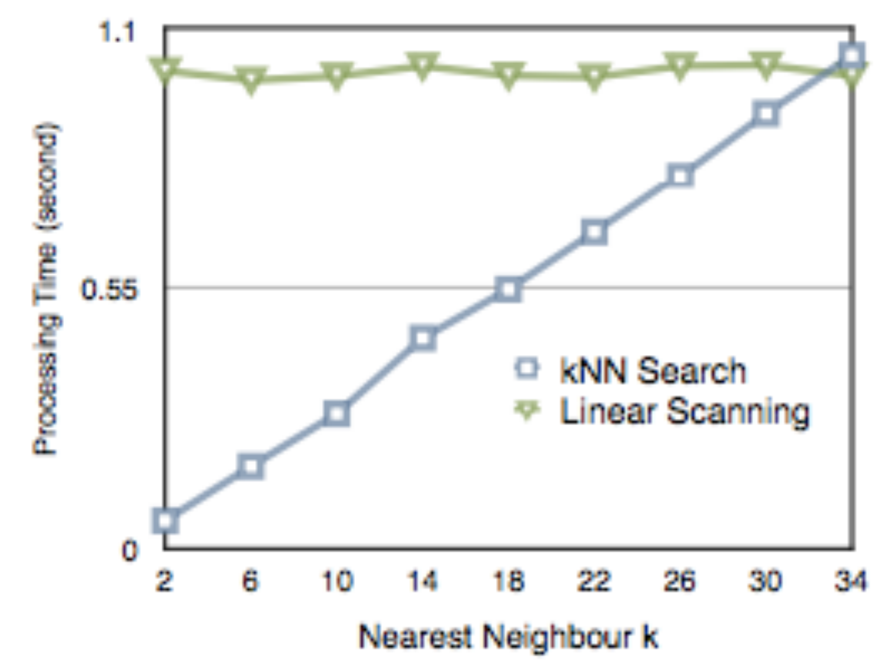
sitting, tiny: TPR 92.87% FPR 5.7%, sitting,normal: TPR 96.99% FPR 2.4%, and walking,normal: TPR 94.88% FPR 4.6%

number and alphabets input: 35/36 are 100%, only one instance of 9 is mis-recognized as 7.

Evaluation: Gesture Recognition Performance



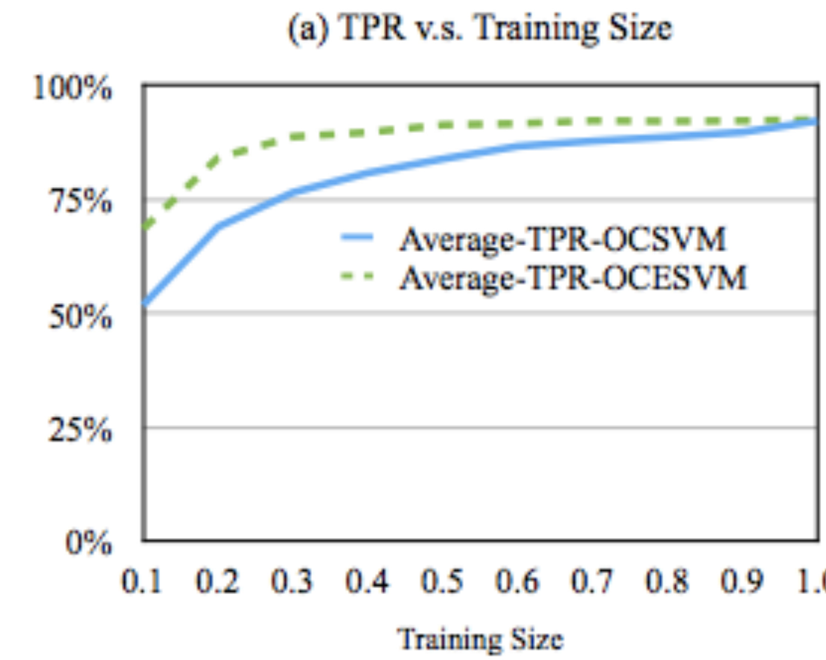


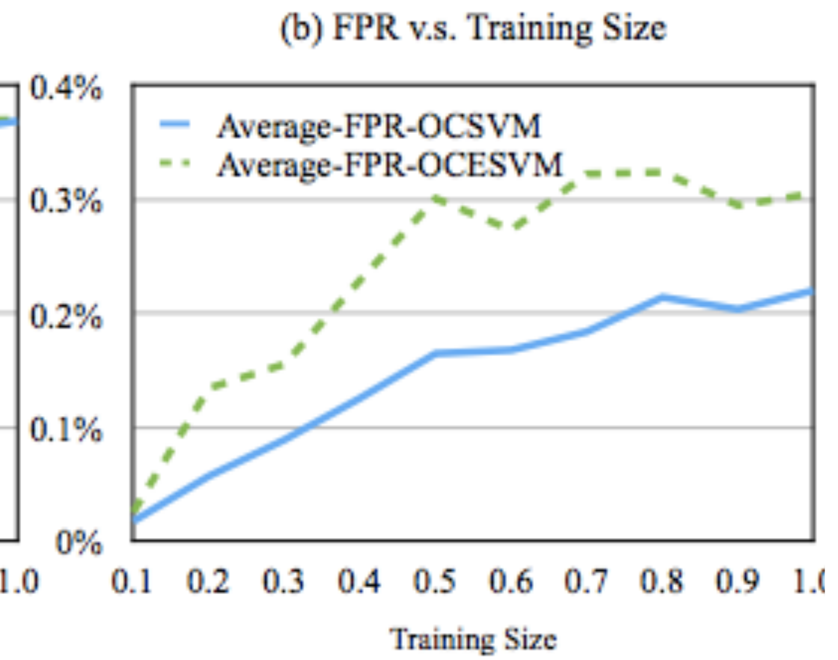


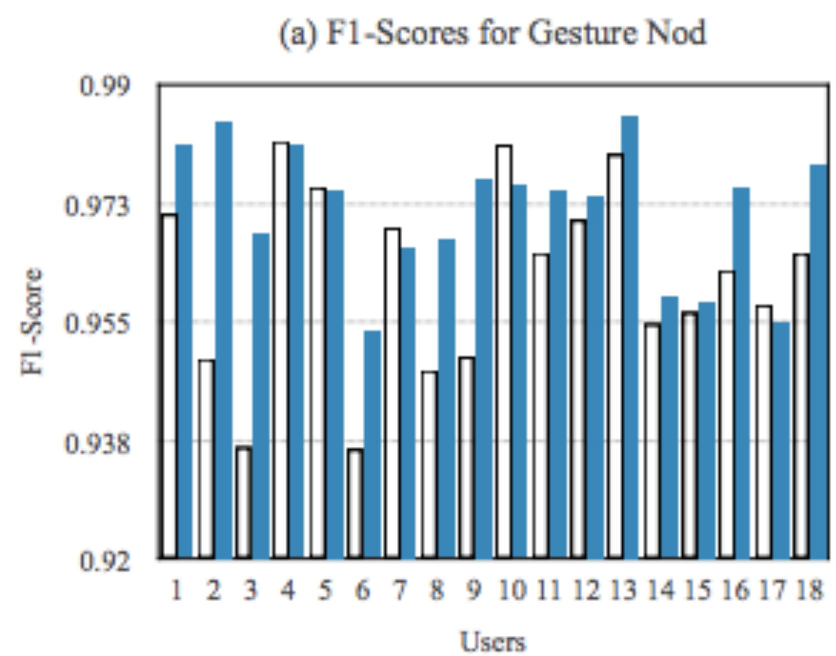
- **sampling every n_ed=10 samples gives best accuracy**
- **downsampling the input for DTW (n_dtw), reduce the run time significantly while slightly reducing the accuracy**
- **k=14 (n_ed=10, n_dtw=50), reduces the time cost by 55%**

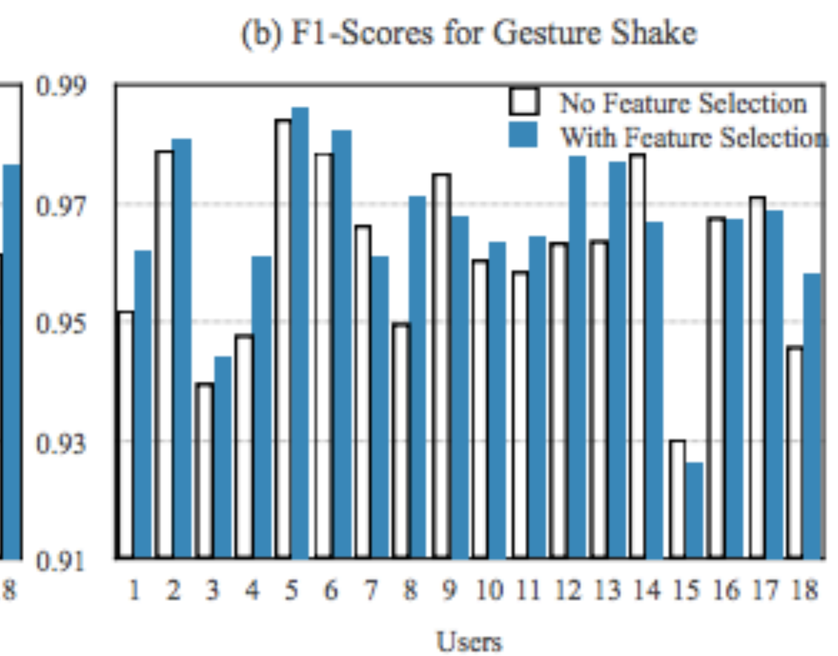
Evaluation: Gesture Authentication, impact of training size and feature selection

- 30 samples is sufficient to achieve 70% TPR and low FPR 0.3%
- OCESVM shows gain of TPR and slight deterioration of FPT when training size is small
- for majority of users (13/18 and 12/18 respectively), feature selection improves the classification









WILLIAM & MARY
CHARTERED 1693

Evaluation: Gesture Authentication Accuracy

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) [19]	97.14%	1.27%
Touchscreen GEAT (3 gestures) [13]	98.2%	1.1%

- **Our scheme requires fewer gestures, and better result when multiple gestures are combined**

Conclusion

- We proposed GlassGesture, which provides a new gesture-based user interface with gesture recognition and authentication, which enable users to use head gesture as input and protect Glass from unauthorized attackers.
- GlassGesture achieves a gesture recognition accuracy near 96%.
- For authentication, GlassGesture can accept authorized users in near 92% of trials, and reject attackers in near 99% of trials. **We also show that in 100 trials imitators cannot successfully masquerade as the authorized user even once.**

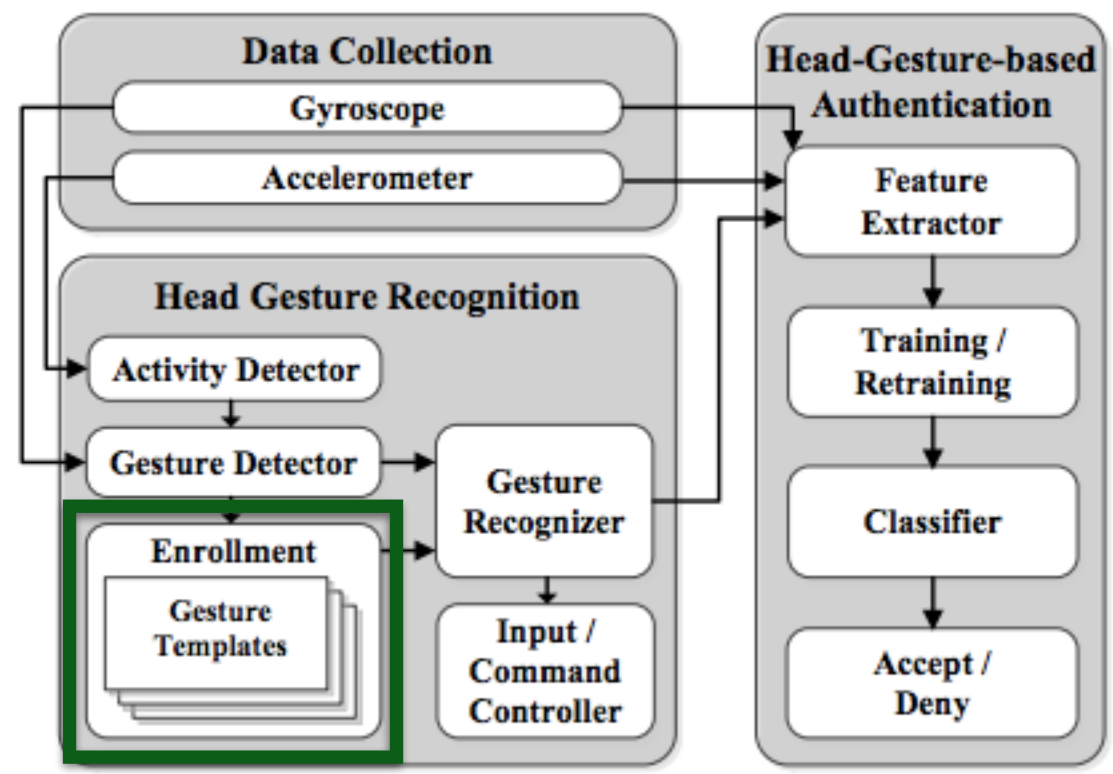
WILLIAM & MARY
CHARTERED 1693

End. Thank you.

Q&A

Gesture Templates Enrollment

- offline enrollment
- user supplies gesture samples for at least three instances
- using a clustering algorithm and choosing the cluster center as template



Activity and Gesture Detectors

- Activity Detector
  - Use a decision tree to get activity context
  - features: mean, std, rms
  - 98% accuracy in preliminary experiment
- Gesture Detector
  - find potential gesture in various activities
  - use a histogram based method to determine thresholds