FICC 2018

5-6 April, Singapore

A Generic Multi-Modal Dynamic Gesture Recognition System using Machine Learning

Gautham Krishna Gudur, Solarillion Foundation, Chennai

Ankith A Prabhu, SRM University, Chennai

Solarillion Foundation

INTRODUCTION

Gesture

A movement of part of the body, especially a hand or the head, to express an idea or meaning.

Gesture Recognition

The perception of non-verbal communication through an interface that identifies gestures using mathematical, probabilistic and statistical methods.

GESTURE RECOGNITION APPLICATIONS

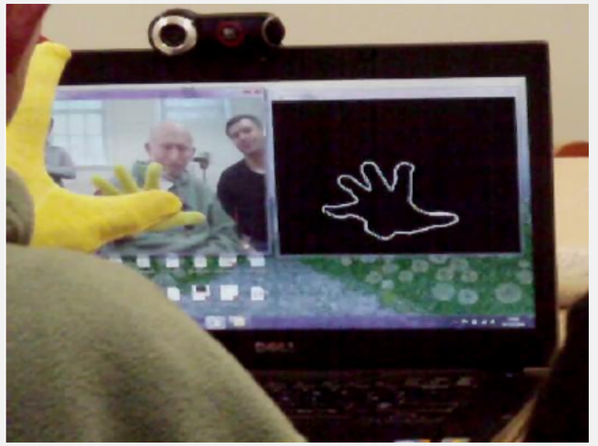
- Developing aids for the hearing impaired using Sign Language interpretation
- Virtual Gaming
- Smart Home Environments
- Socially-Assistive Robotics
- Affective Computing

Solarillion Foundation

GESTURE RECOGNITION APPROACHES

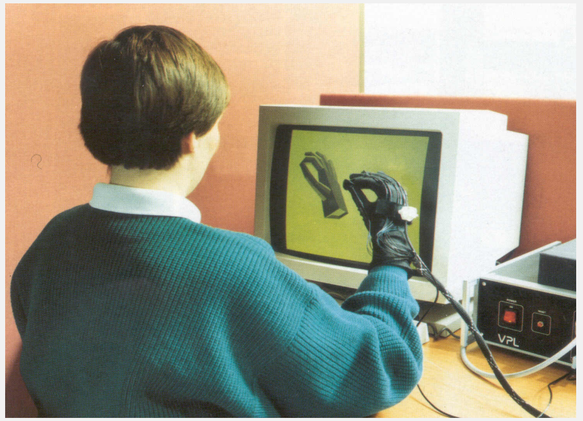
Vision Based Approach

- Camera for tracking movements
- Higher Computational Overheads



Sensor (Haptic) Based Approach

- Cost-effective and Computationally Efficient
- Dependent on Environmental Stimuli
- Examples: Accelerometer, Gyroscope



Solarillion Foundation

GESTURE RECOGNITION TYPES

Static Gesture Recognition

- Uniquely characterized by identifying Start and End points
- Analogous to an Image

Dynamic Gesture Recognition

- Requires the entire sequence of Gesture sample
- Analogous to Video

CHALLENGES IN CONVENTIONAL GESTURE RECOGNITION SYSTEMS

- Requires instrumented and multifarious sensors
- Failure to generalize across multiple users (modally inflexible)
- Gesture specific feature extraction
- Failure to handle disparate speeds of same gestures signed by various users
- Application specificity and deployment in Low-Cost Platforms

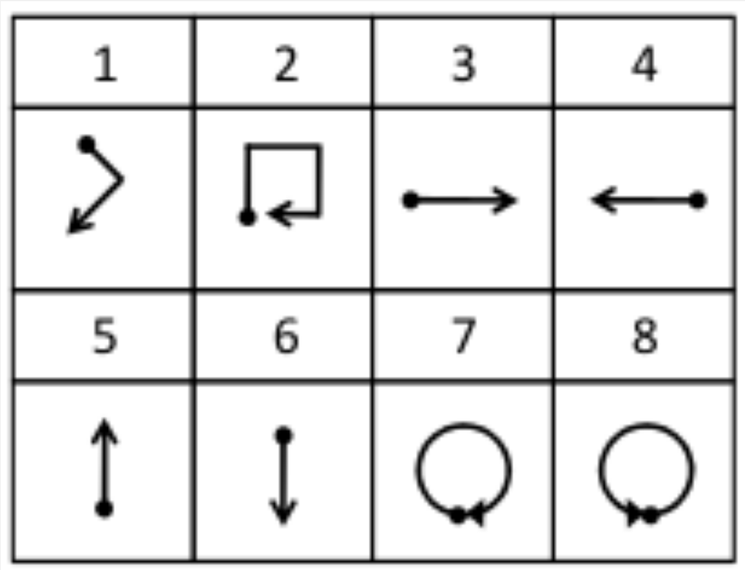
GOALS OF OUR PROPOSED SYSTEM

- To utilize accelerometer-based datasets for Sensor Minimization (containing g-values)
- To handle disparate speeds across gestures by leveraging unique features
- To incorporate Multi-Modality across users
- To provide the user the flexibility to emphasize between classification accuracy and classification time
- To deploy on a low-cost platform

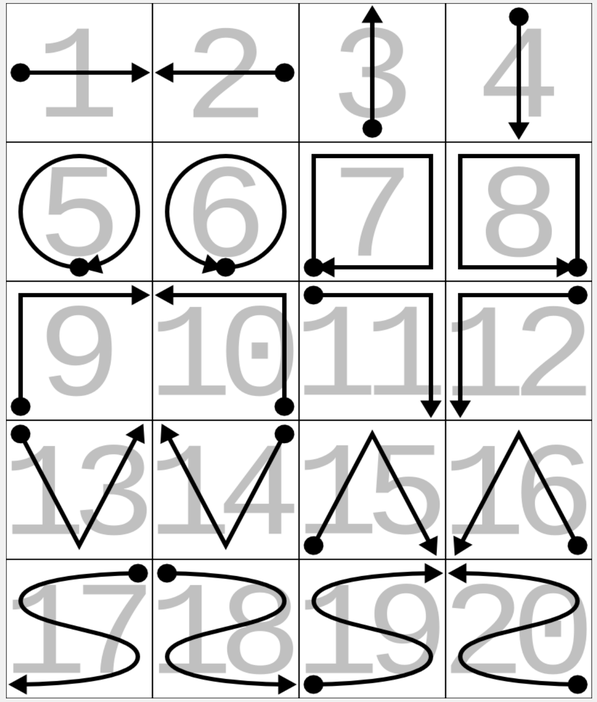
GESTURE DATASETS

Dataset	Users (U)	Gestures (N_G)	Samples per Gesture (S_G)	Days (N_D)	N_{GS}
uWave (D_u)	8	8	10	7	4480
Sony (D_S)	8	20	20	-	3200

uWave Gestures



Sony Gestures



Solarillion Foundation

FEATURE EXTRACTION

- To generalize the datasets, we eliminate the dependency on time from the datasets. No other pre-processing was done to both the datasets as the g-values would be altered, making the datasets lossy

- The features are characterized across two domains: Time and Frequency.

- Two transforms: Fast Fourier Transform (*FFT*) and Hilbert Transform (*HT*), are utilized to convert from time domain to frequency domain.

- Most of the features in frequency domain are *HT* specific. The novelty in *HT* is that it reduces the incoming data by considering only the imaginary values, thereby decreasing the computation time of the feature vector, while increasing the accuracy.

FEATURE EXTRACTION

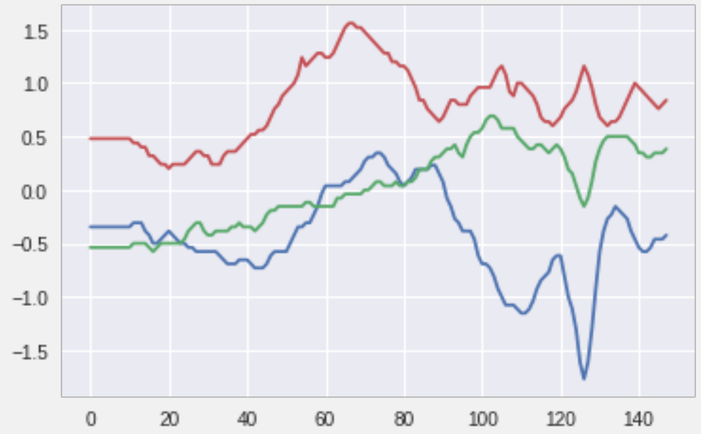
Features\Domain	Time	Frequency	
		FFT	HT
Mean	✓ (T^1)	✗	✓ (H^1)
Skew	✓ (T^2)	✗	✓ (H^2)
Kurtosis	✓ (T^3)	✗	✗
PM correlation coefficients	✓ (T^4)	✗	✗
Cross correlation	✓ (T^5)	✗	✗
Energy	✗	✓ (F^1)	✓ (H^3)
Minimum	✗	✗	✓ (H^4)
Maximum	✗	✗	✓ (H^5)

- These are the set of unique features that accurately represent a gesture signed by the user.

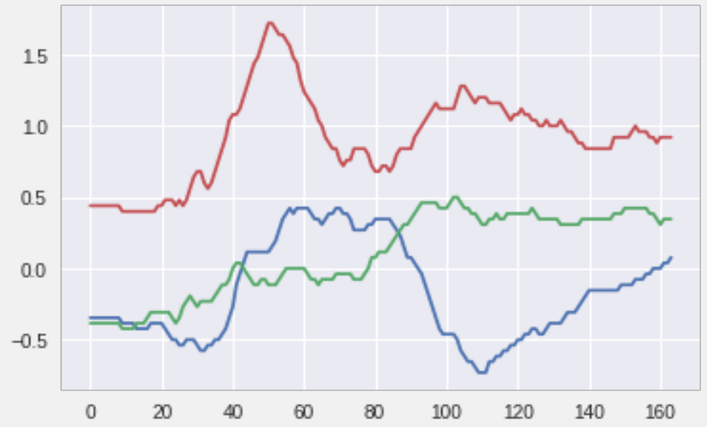
- Each gesture is characterized by a feature vector, and we arrive at this final set of features by recursive feature elimination across all domains and transforms.

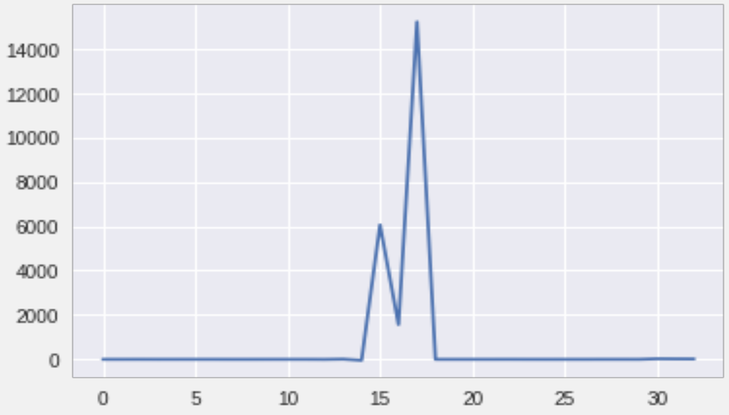
Solarillion Foundation

FEATURE EXTRACTION

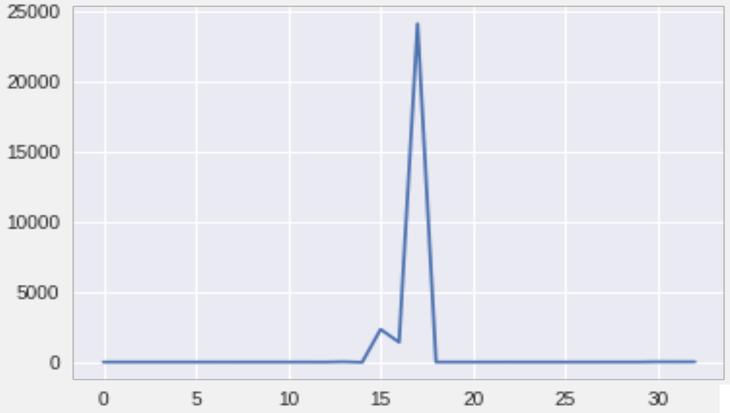


INPUT





FEATURE VECTOR



Solarillion Foundation

END-USER MODELING

The end-user is provided the choice of any one of three proposed modes of operation:

- ***User Dependent (UD)*** mode is an estimator of how well the system performs when the train-test split is between the gestures of a single user.
- ***Mixed User (UM)*** is representative of the complete set of gestures across all participants.
- ***User Independent (UI)*** mode employs a stratified k-fold cross validation technique which corresponds to training on a number of users and testing on the rest.

Solarillion Foundation

EXPERIMENT

- Initially, seven Machine Learning (*ML*) paradigms were implemented to classify gesture samples across the various modes.

- *ML* classifiers were chosen over traditional algorithms like Dynamic Time Warping (DTW), as generalizing a look-up table (template) for each gesture is computationally inefficient in the user-independent paradigm.

- The *ML* classification algorithms used here are,

  - Extremely Randomized Trees (Extra Trees)
  - Random Forest
  - Gradient Boosting
  - Bagging
  - Decision Trees
  - Naive Bayes
  - Ridge Classifier

Solarillion Foundation

EXPERIMENT

- The seven classifiers were simulated on a single board computing platform – Raspberry Pi Zero.
- Raspberry Pi Zero priced at 5$, makes it a low cost alternative to conventional computing modules.

Solarillion Foundation

RESULTS

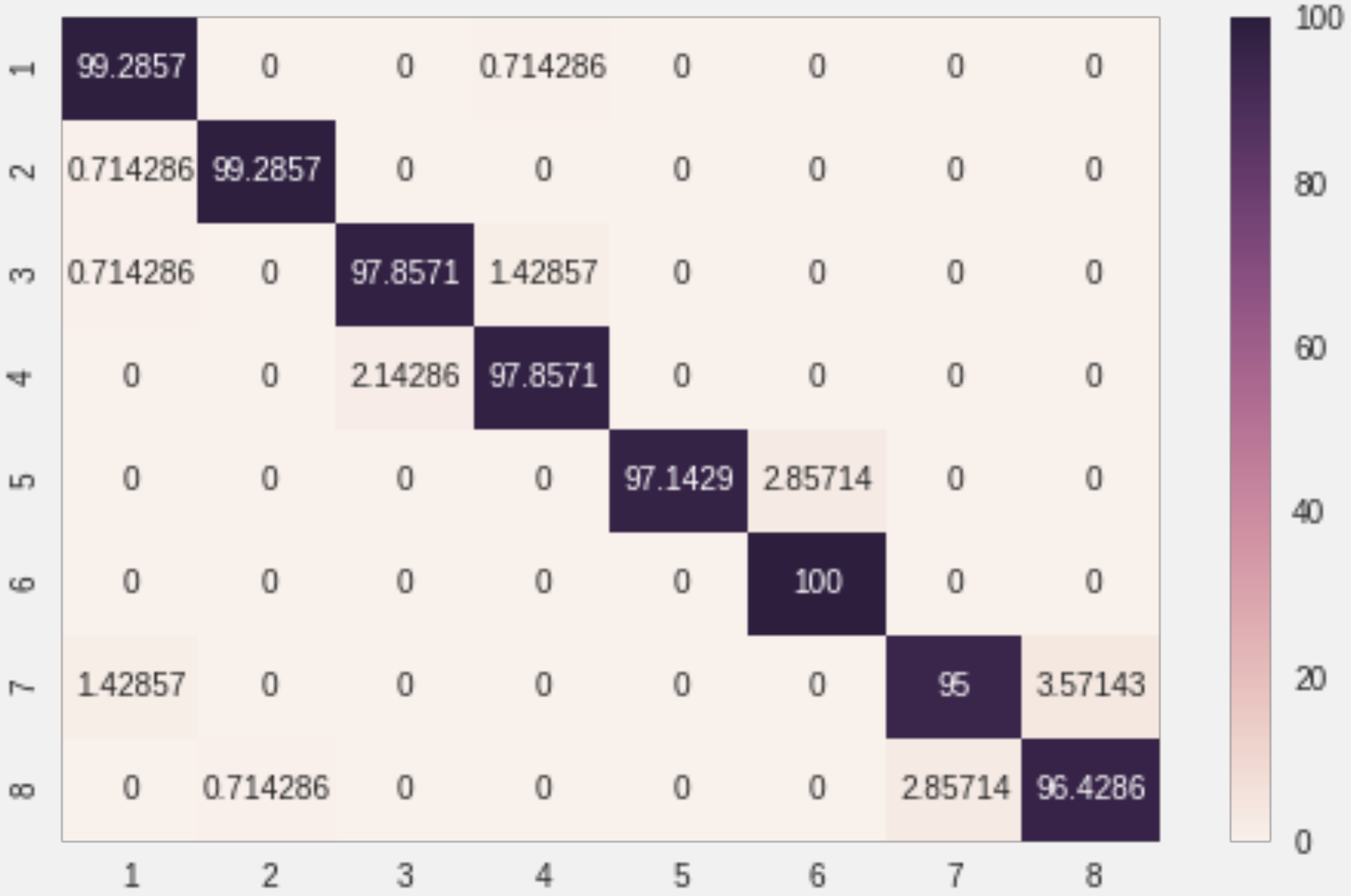
	uWave (D_u)						Sony (D_s)					
Classifier\Mode	User Dependent (U_D)		Mixed User (U_M)		User Independent (U_D)		User Dependent (U_D)		Mixed User (U_M)		User Independent (U_D)	
	Acc	Time	Acc	Time	Acc	Time	Acc	Time	Acc	Time	Acc	Time
Extra Trees	97.76	0.6287	97.85	0.6873	82.49	0.6853	95.88	0.6038	98.63	0.6538	75.1	0.669
Random Forest	97.41	0.6991	95.45	0.73	77.91	0.6995	95.25	0.6887	97.13	0.7041	70.13	0.686
Gradient Boosting	93.75	0.0072	94.38	0.0076	75.64	0.0078	90.5	0.0055	95.5	0.0057	66.41	0.0057
Bagging	92.74	0.1526	94.19	0.1527	76.64	0.1528	93.5	0.1673	93.37	0.1527	62.44	0.1529
Decision Trees	89.55	0.0015	84.11	0.0053	66.73	0.0015	84.13	0.0014	86.25	0.0051	50.9	0.0015
Naive Bayes	91.16	0.0273	71.96	0.0292	64.66	0.0273	91.38	0.0118	65.5	0.0116	54.35	0.0117
Ridge Classifier	97.5	0.0013	83.84	0.0013	74.64	0.0013	94.13	0.0013	77.75	0.0014	61.59	0.0013

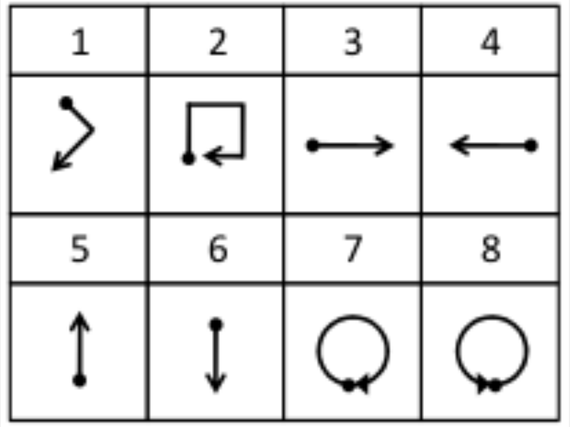
Solarillion Foundation

RESULTS

- From this initial set of seven classifiers, Extra Trees, Gradient Boosting and Ridge Classifier were chosen based on their computational characteristics, as is evident from the Table.
- Trade-off between accuracies and classification time per gesture sample is done based on real-world user requirements.
- Also, depending on the application the system is used for, the end-user has the liberty of choosing the mode of operation.

CONFUSION MATRIX





Solarillion Foundation

THANK YOU!

Contact

Gautham Krishna Gudur

Research Assistant

Solarillion Foundation

gauthamkrishna.gudur@gmail.com

Ankith A Prabhu

Research Assistant

Solarillion Foundation

ankithprabhu@ymail.com

Solarillion Foundation