Recognizing Gestures with Ambient Light

Raghav H. Venkatnarayan
Muhammad Shahzad

North Carolina State University
Raleigh NC USA

Motivation

- Gesture Recognition enables various interactive applications.

Gaming

Health Care

Smart Homes

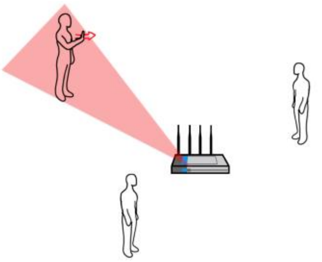
AR

- Multiple Modalities

Wearables

Sound

Vision/IR

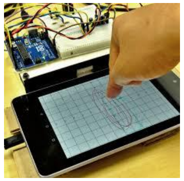


RF

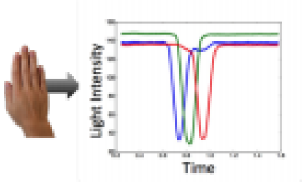
Motivation

- Gesture Recognition using Ambient Light Signals
  - Ubiquitous:
    Light Sources are available everywhere
  - Non-invasive:
    Movements can be sensed from shadows
  - Preserve Privacy:
    Signals do not leak through walls

Existing Approaches

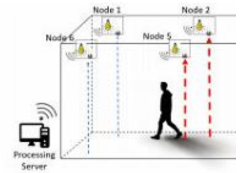


Okuli
(MobiCom ‘15)

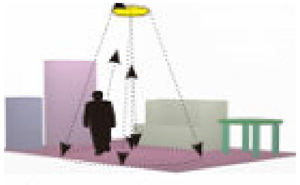


GestureLite
(DTR ‘16)

Limited Range
(< 30cm)



VLAS
(VLCS ‘16)



CeilingSee
(PerCom ‘16)

Limited Resolution
(Room-Level Semantics)

LiSense
(MobiCom ‘15)

StarLight
(MobiSys ‘16)

Active Sensing :

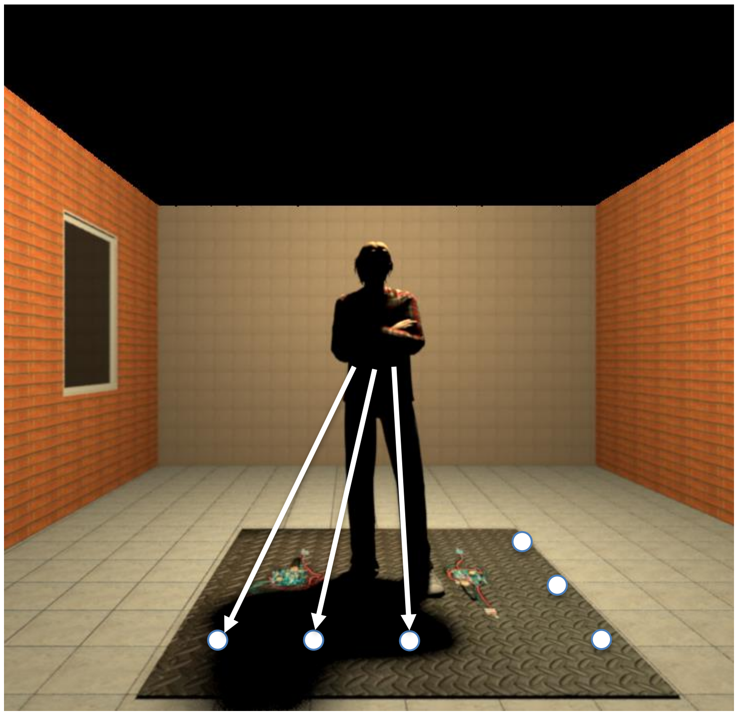
Controlled lighting infrastructure
(Modulated LED lights)

Problem Statement

Design a passive, *ambient* light based gesture recognition system

- Unmodulated Light Sources
- Agnostic to changing lighting conditions
- Agnostic to changing user position and orientation
- Recognize gestures of any given user

Approach

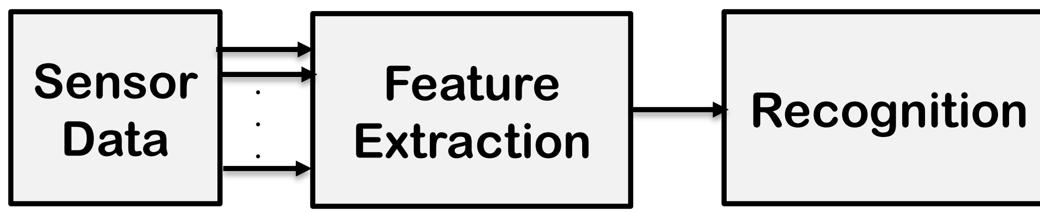


<u>Observations</u>:

- Shadows follow movements
- Different gestures create distinct shadow patterns on the floor

<u>Idea:</u>

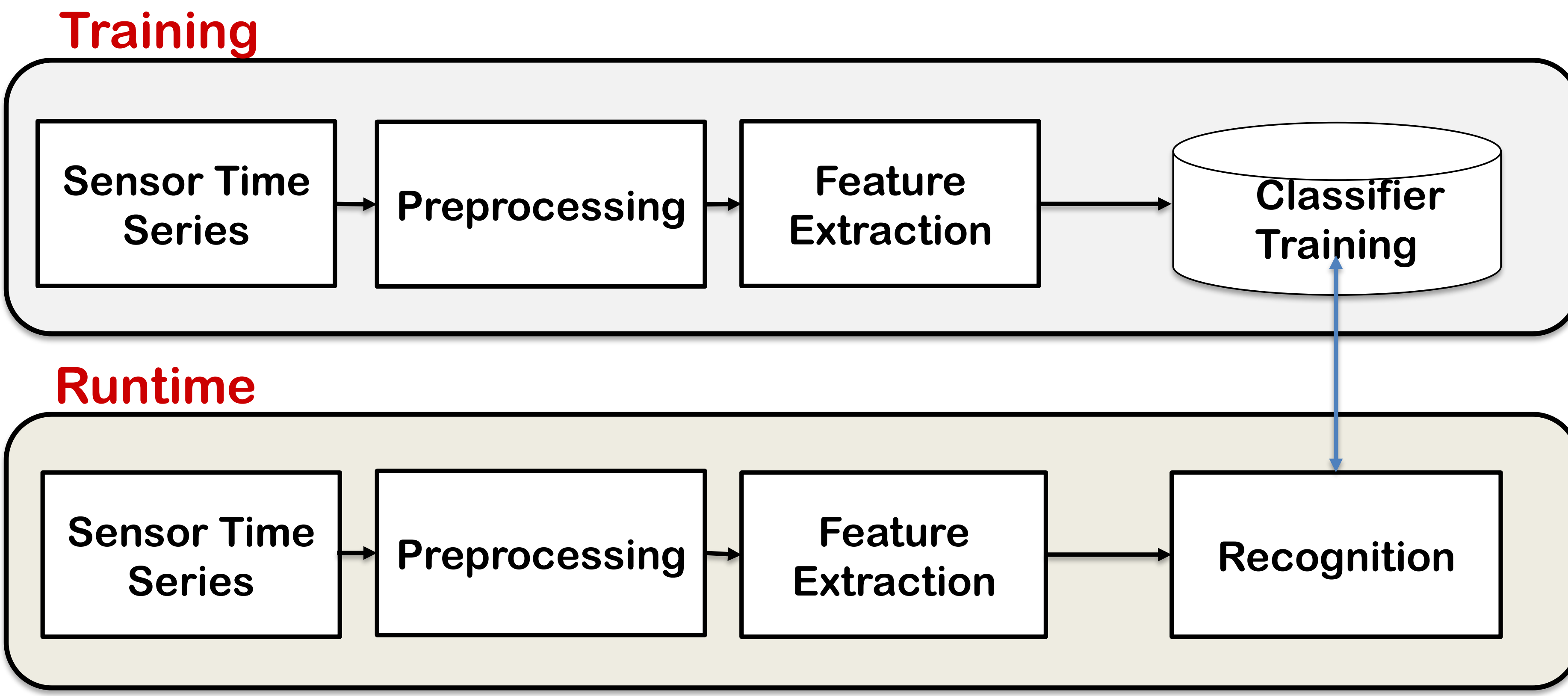
- Instrument floor to learn shadow patterns using ML models and infer gestures.



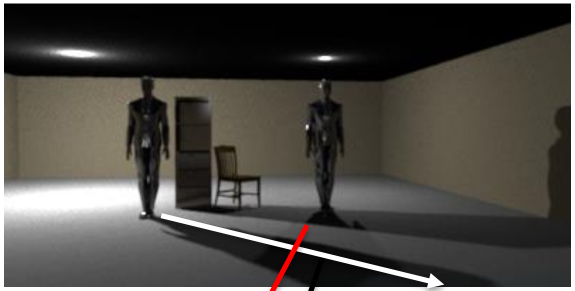
<u>Contributions</u>

- Capturing features agnostic to different lighting conditions, user positions and orientations

Overview


Training
Sensor Time Series
Preprocessing
Feature Extraction
Classifier Training
Runtime
Sensor Time Series
Preprocessing
Feature Extraction
Recognition

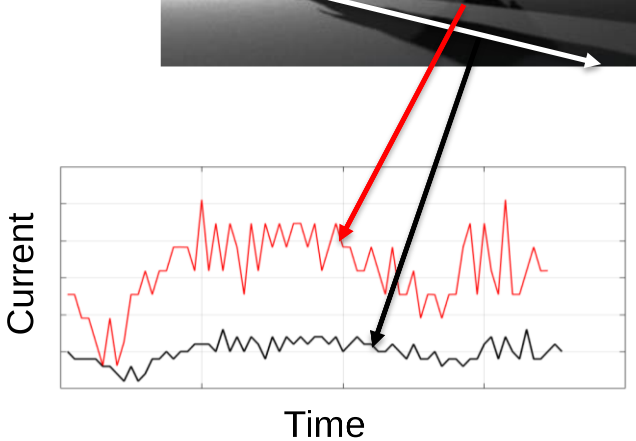
I. Preprocessing



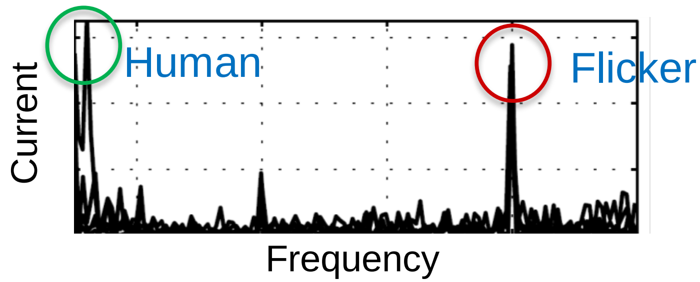
<u>1. Denoising:</u>

Separating signal from the noise

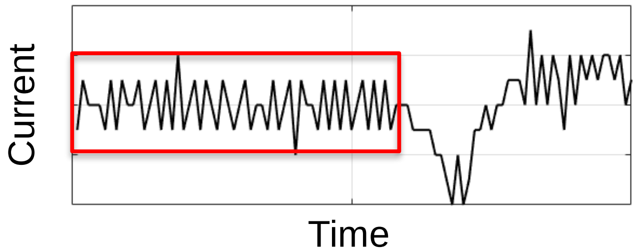
(i) Stray Shadows and Reflectors :
Varying photocurrent levels with environmental changes.



(ii) Light Source Flicker (AC Powered):
Fluctuations of comparable magnitude
-Well localized in Frequency Domain



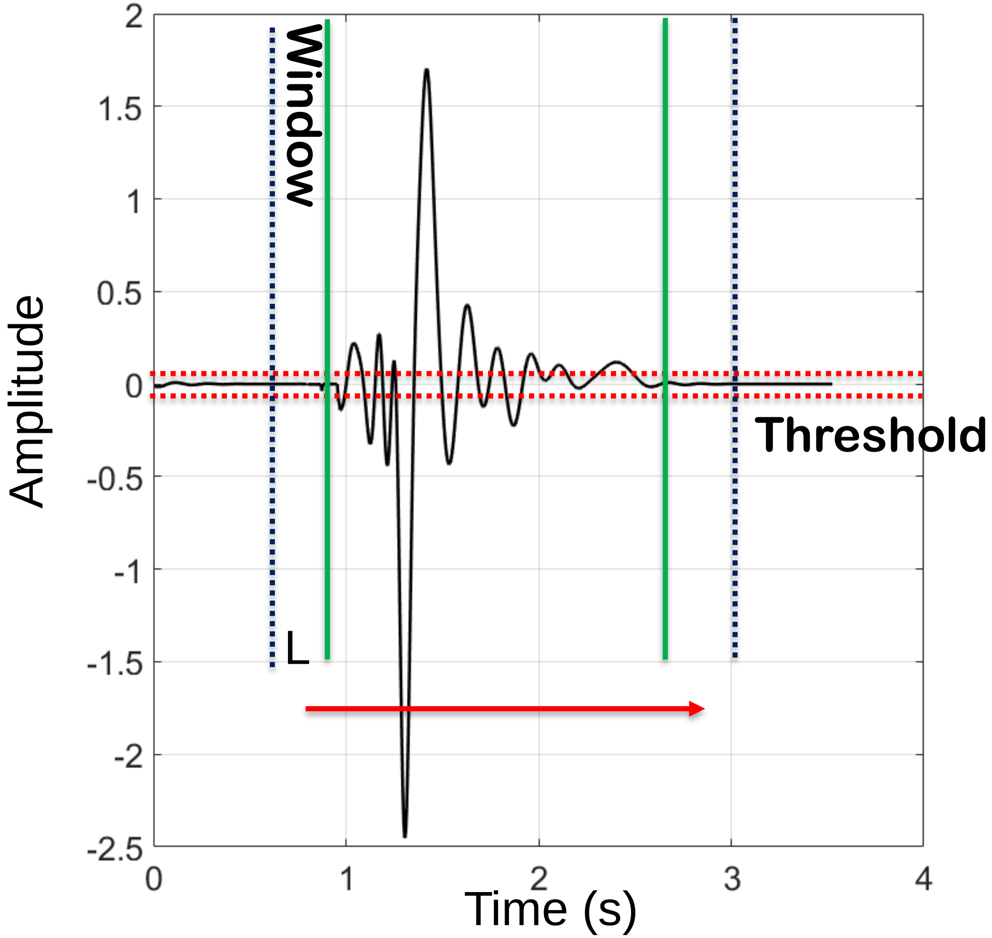
(iii) Shot noise
Spurious burst noises
-Well localized in Time Domain



I. Preprocessing

2. Gesture Detection:

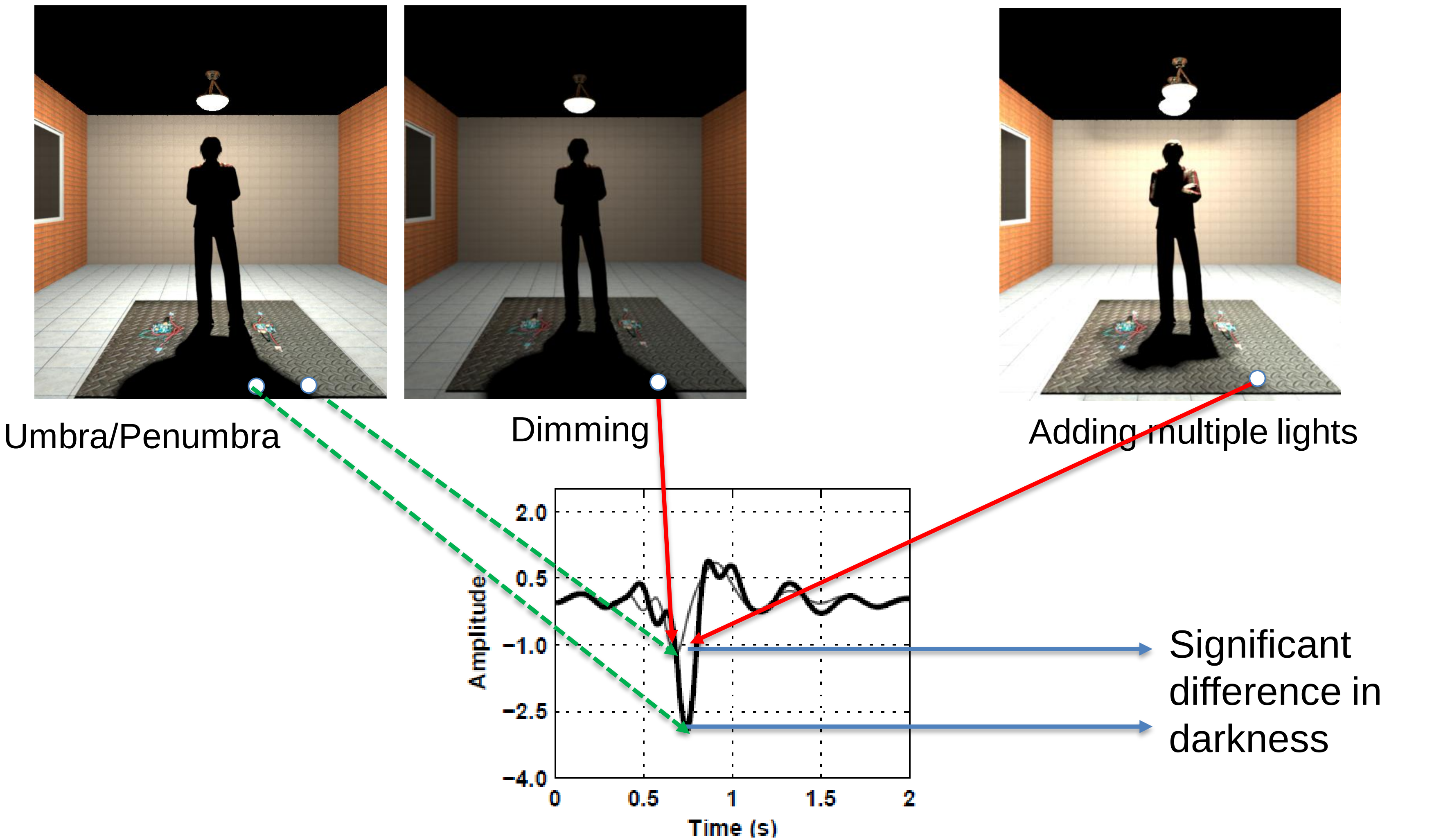
Thresholding with a sliding window



I. Preprocessing

3. Standardization

Handling changes in intensity across sensors

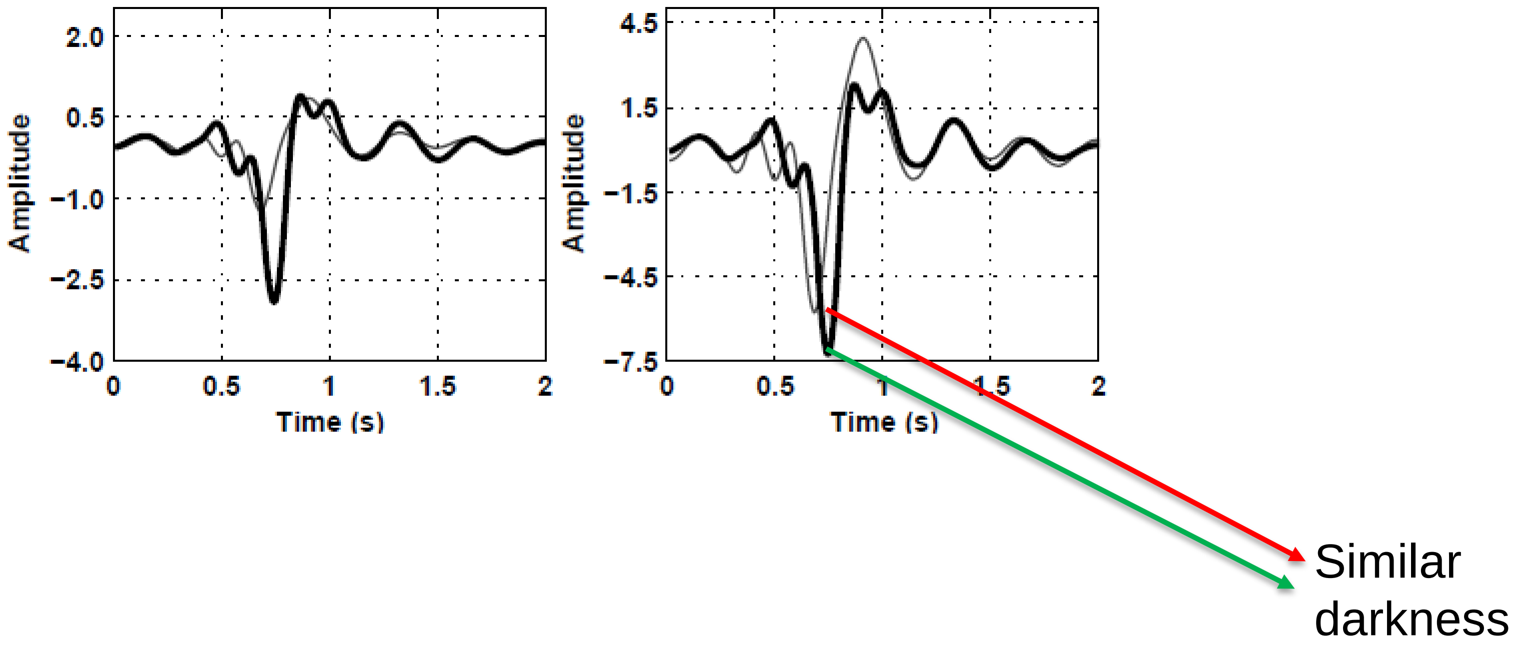


I. Preprocessing

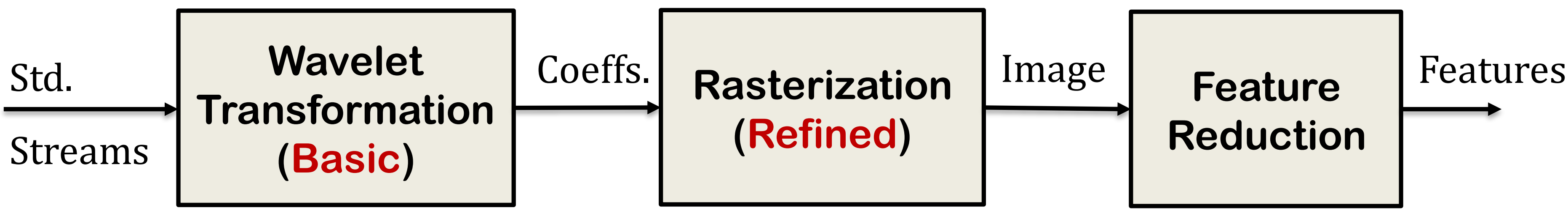
3. Standardization

Handling changes in intensity across sensors

Solution : Scale each sensor time series by deviation



II. Feature Extraction

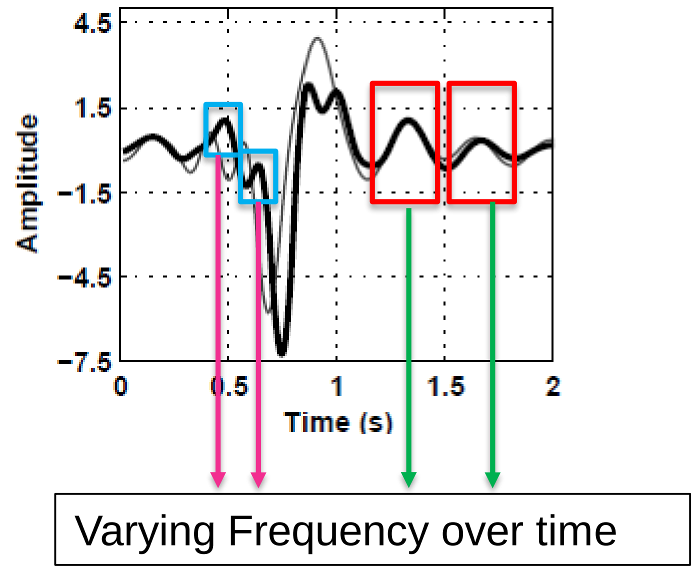


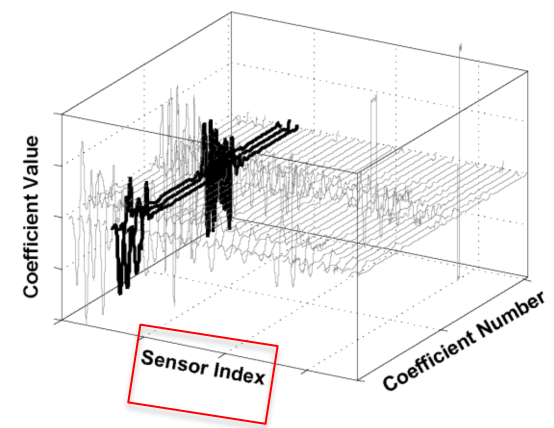
II. Feature Extraction

<u>1. Wavelet Transformation:</u>

Objective:
Characterize shape of the signal.

Approach:
Extract a joint signature in time and frequency domains using Discrete Wavelet Transform

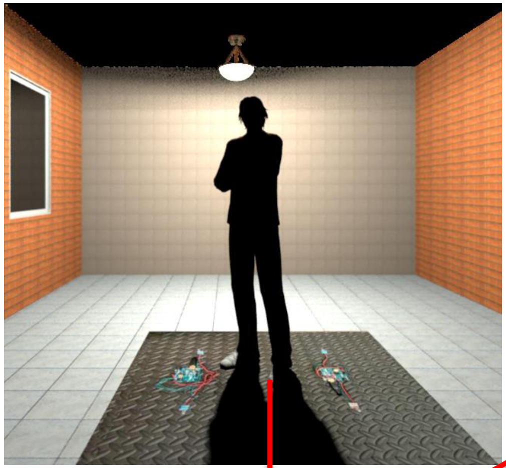




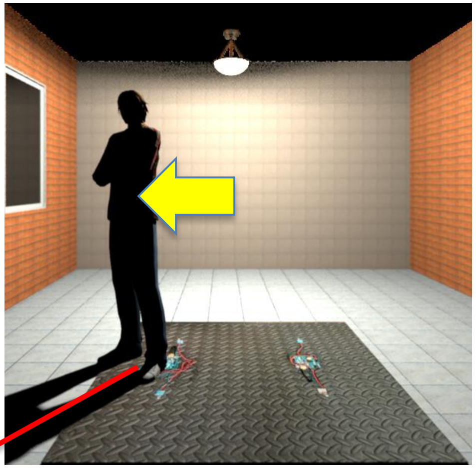
II. Feature Extraction

2. Rasterization:

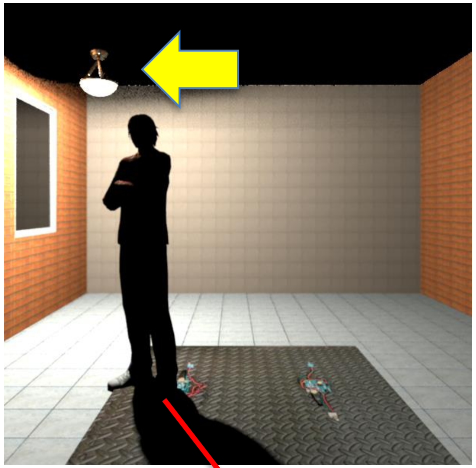
Handling changes in features caused by shifts in position of light sources or position of users



Example Position



Shift in User Position



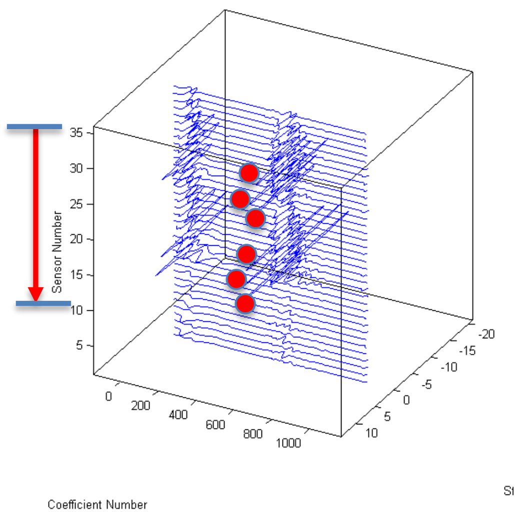
Shift in Light Source

Effect : Changing Direction and length of shadows across samples

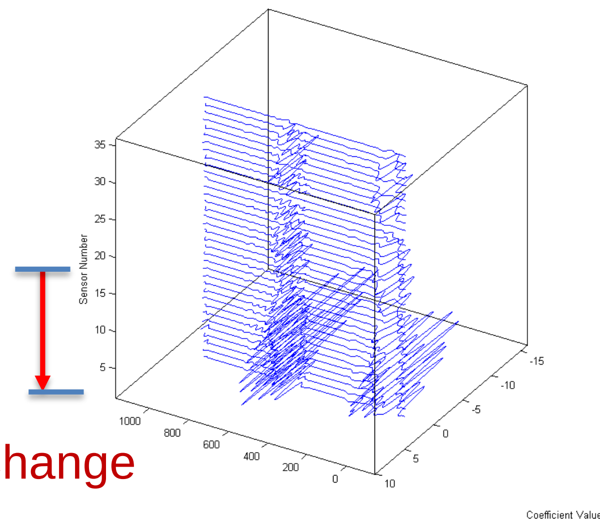
II. Feature Extraction

2. Rasterization:

Example:



Sensors Change



6 sensors see variations

10 sensors see variations

Need a way to negate the effects of change in shadow length/direction

II. Feature Extraction

2. Rasterization:

Existing Approaches:

- Identify blockage of individual light sources using Frequency Modulation to localize shadows
- Shadows can then be scaled, translated or rotated

- **Cannot be applied to unmodulated / unknown light sources**

II. Feature Extraction

2. Rasterization:

How to handle variations in position of light sources or position of users with unmodulated light sources?

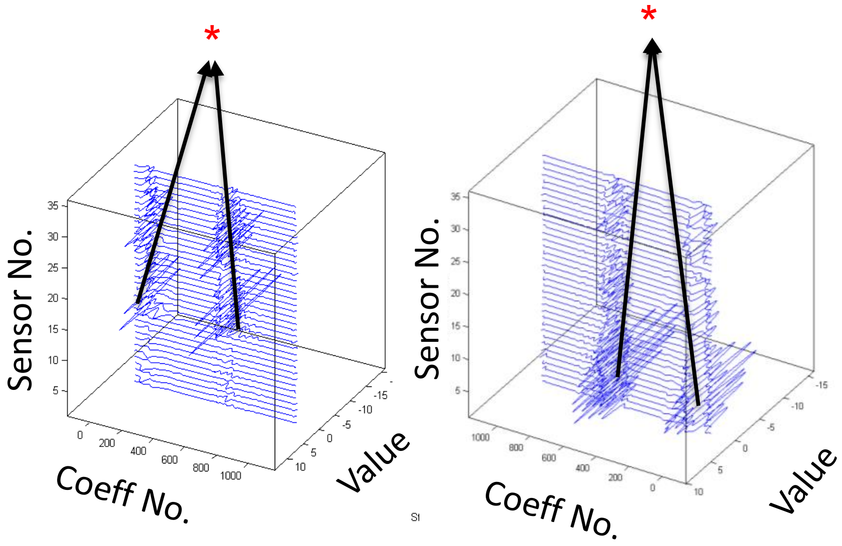
Observations:

1) Sensor values still have similar patterns * due to same blocking source

2) More light sources =>
Multiple redundant shadows

3) Change in shadow length =>
Change in No. of sensors

4) Change in shadow direction =>
Change in index of sensors



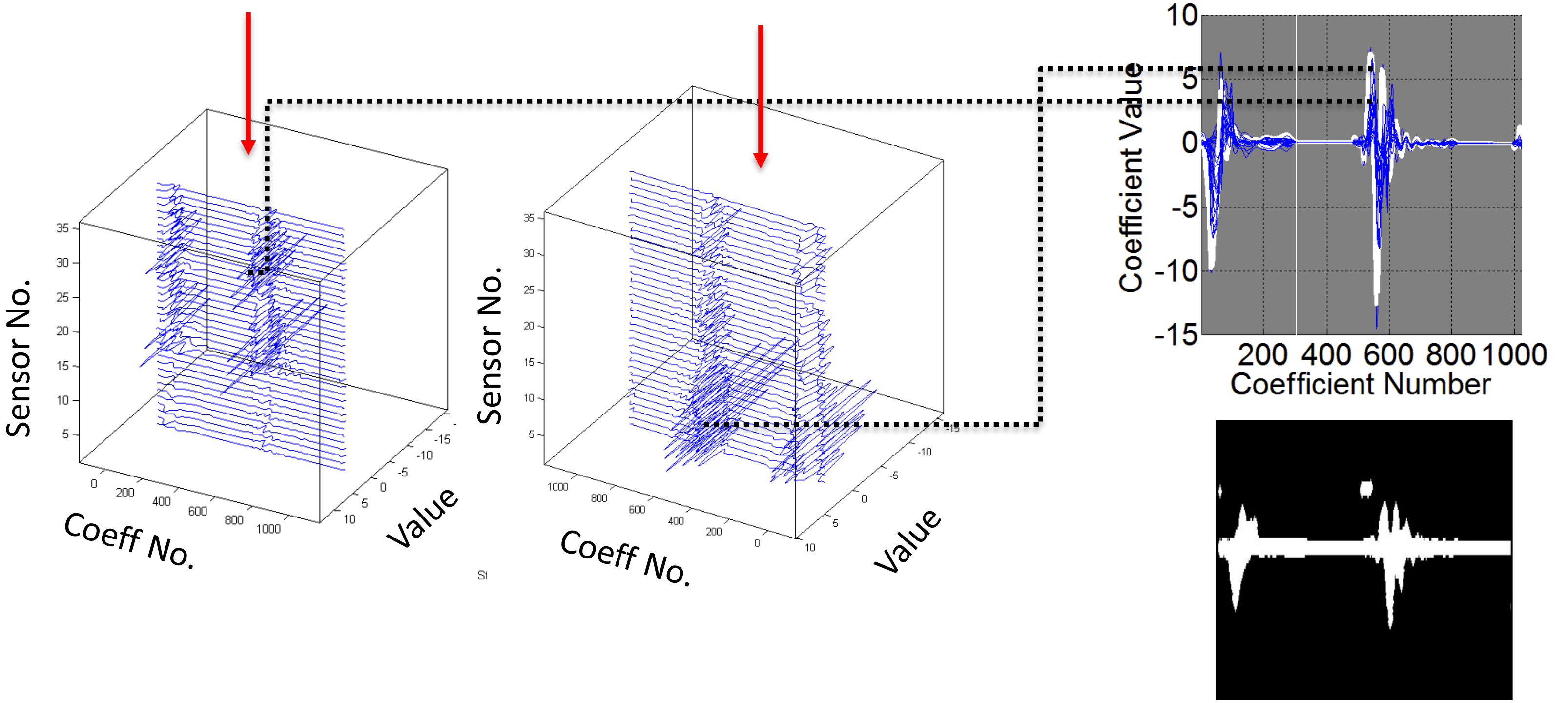
II. Feature Extraction

2. Rasterization:

How to handle variations in position of light sources or position of users with unmodulated light sources?

Solution: Map all sensor coefficients into a 2D image

Redundant / Similar patterns merge

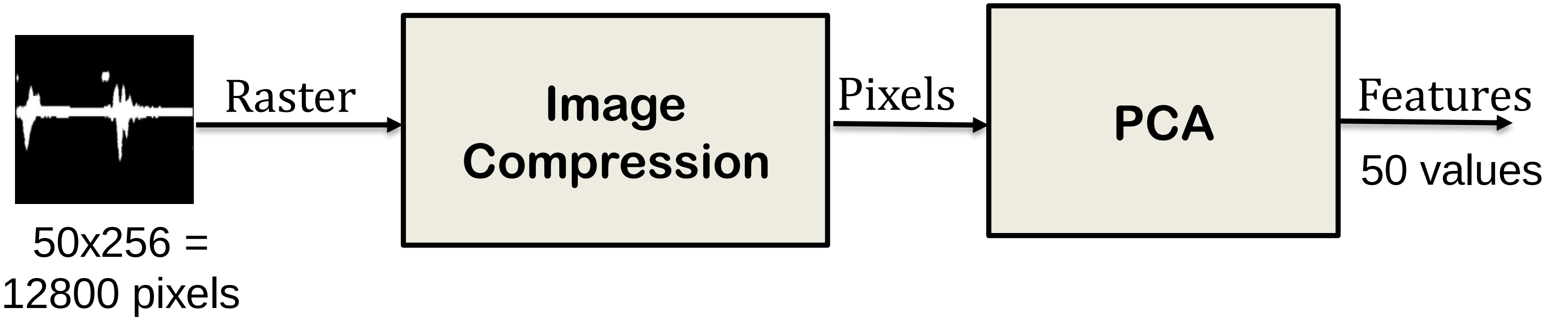


II. Feature Extraction

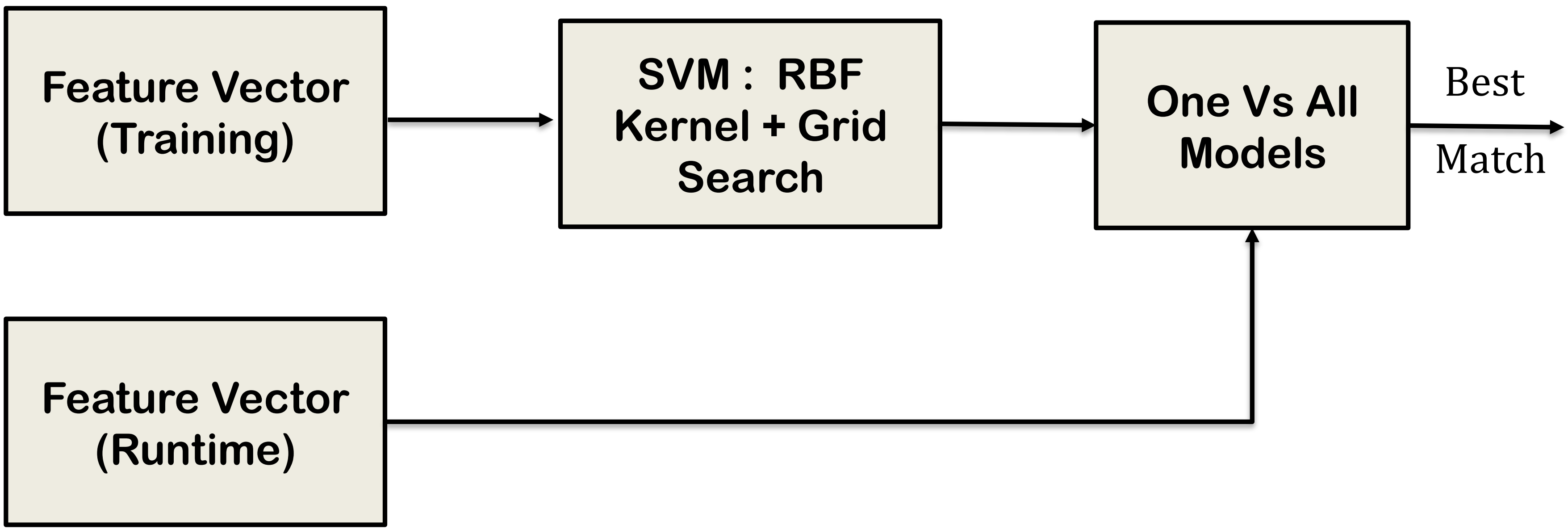
3. Feature Reduction:

Objective:
Extract only features of high classification potential

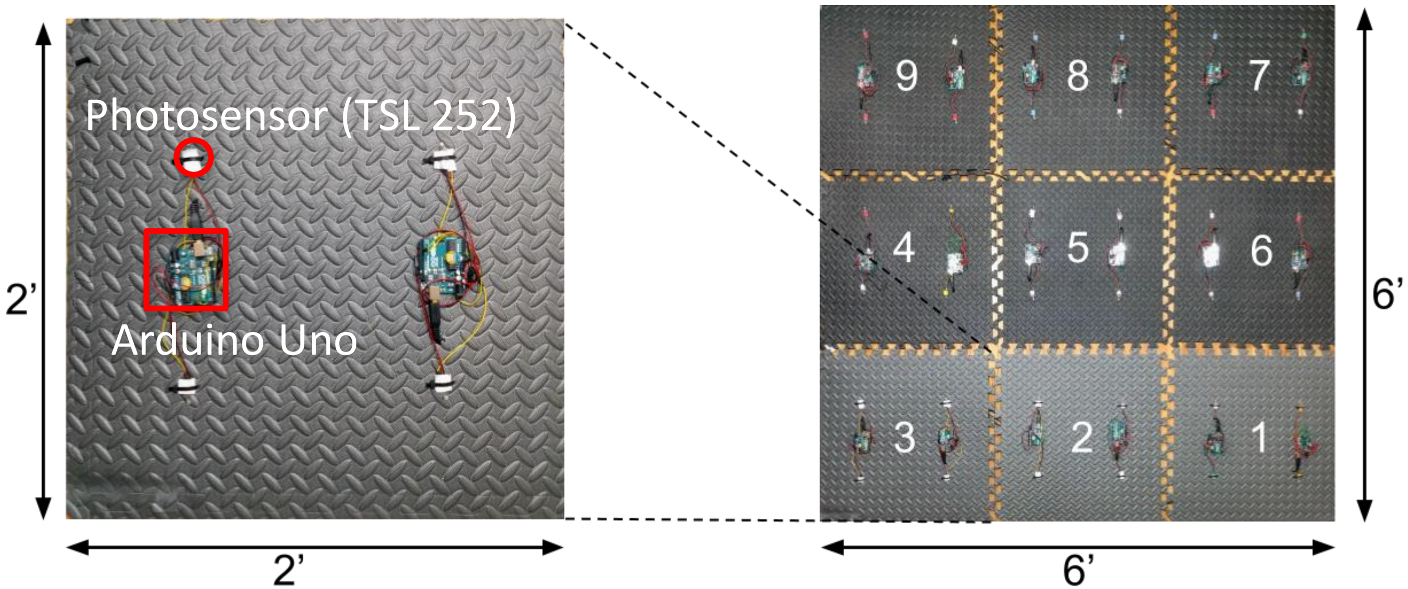
Approach:
Dimensionality Reduction



III. Classifier Training / Recognition



Implementation

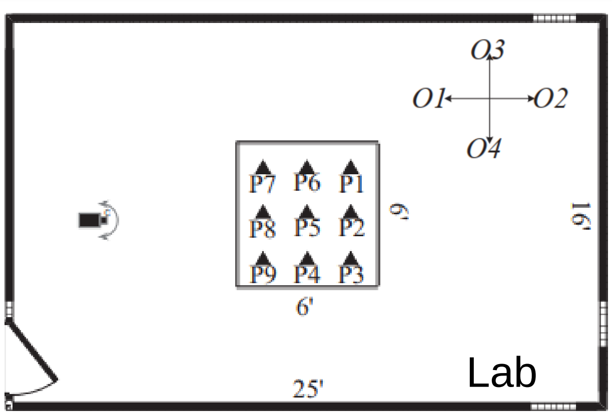


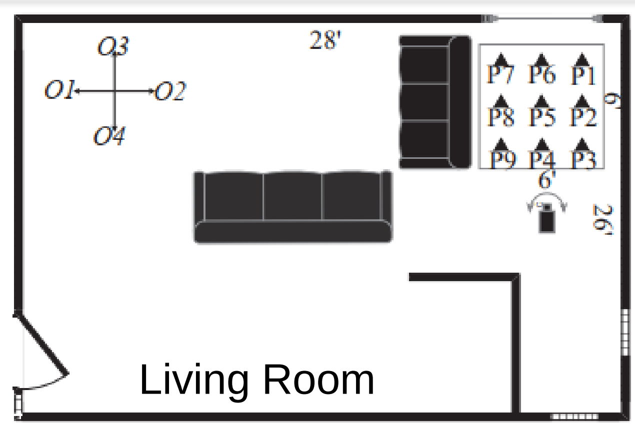
Sensor Density	1/sq.ft
Production System Cost	\$ 0.2 /sq.ft
Avg Cost of Carpeting	\$ 3 /sq.ft

Evaluation

- **Volunteers** : 20
  Height : 150cm – 180cm
- **Gestures** : 5
  Clap, Hug, Jump, Punch, Step

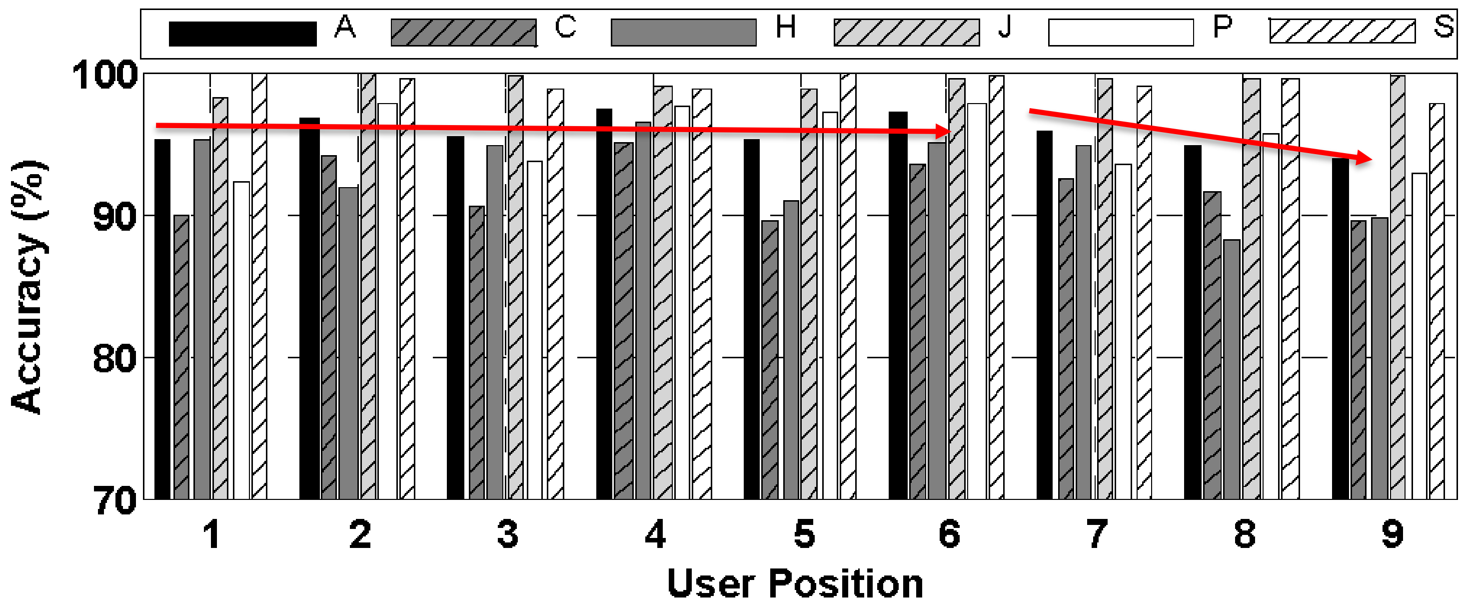
- **Positions** : 9
- **Orientations** : 4
- **Lighting Conditions** : 11
- **Environments** : 2 (15175 Samples)





Evaluation

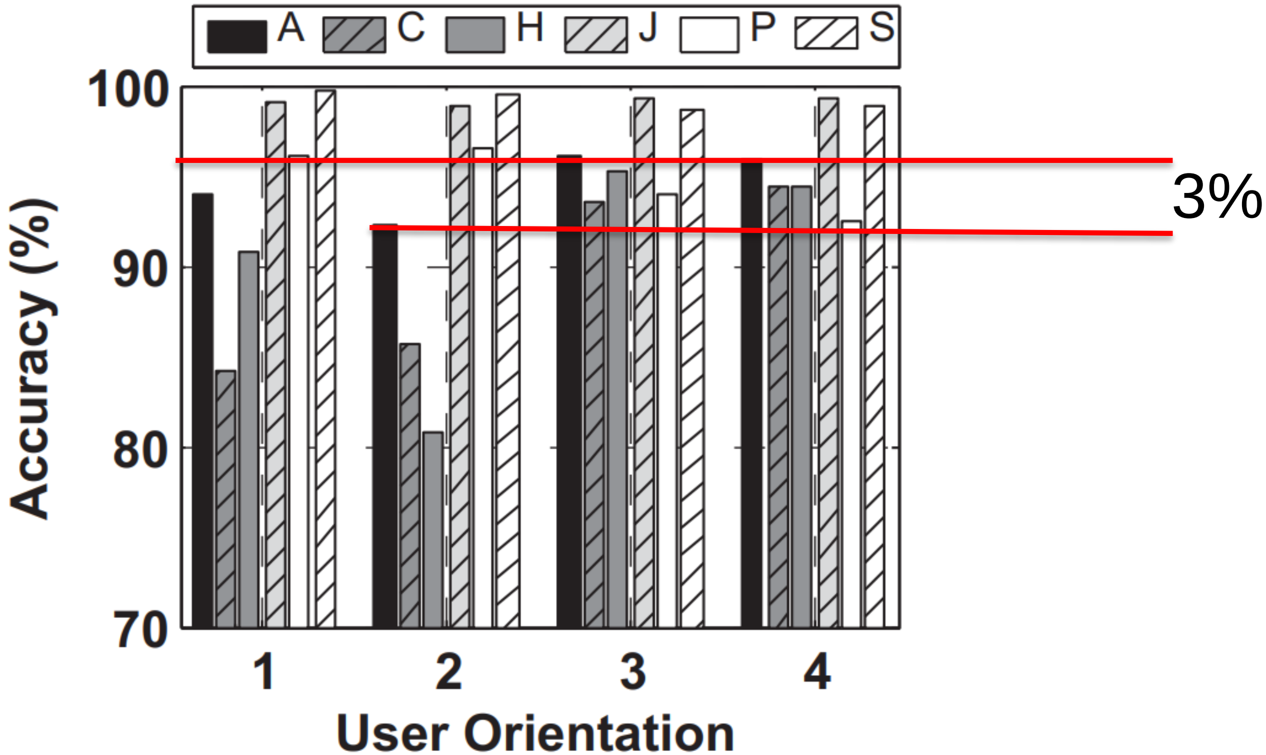
1. Recognition Accuracy : Unseen User Positions



Average : 95.2%

Evaluation

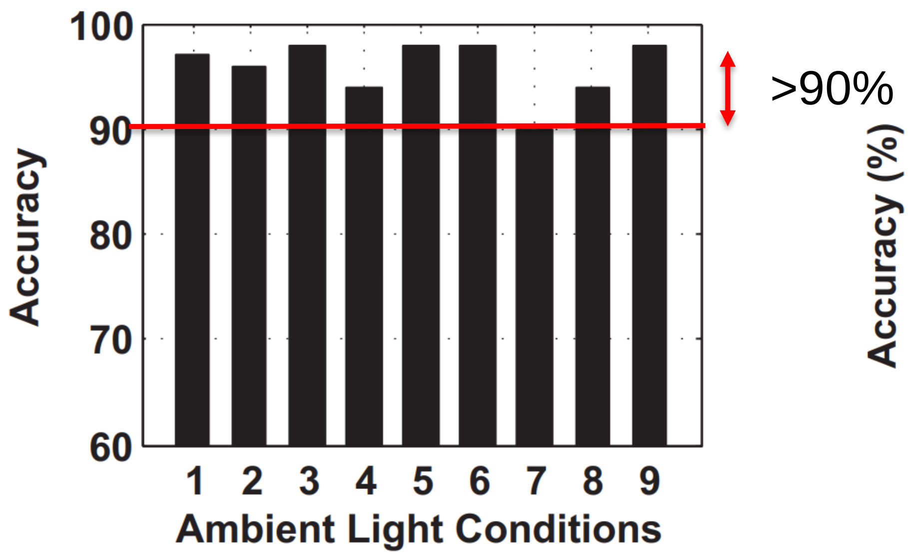
2. Recognition Accuracy : Unseen User Orientations

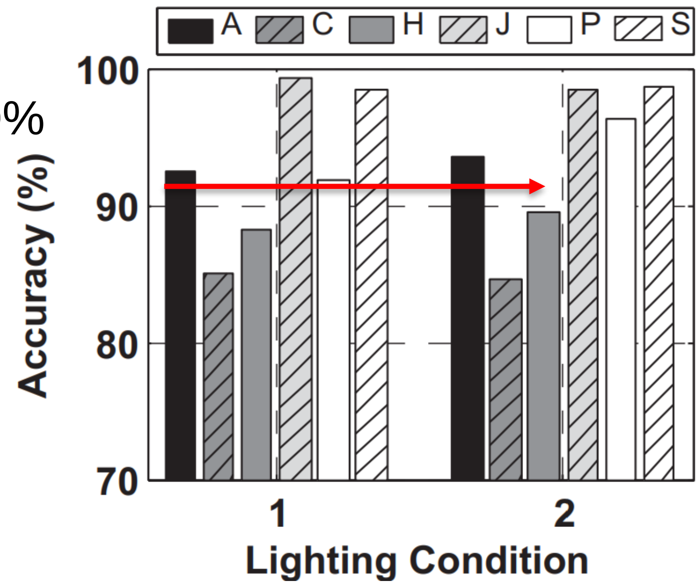


Average : 94.5%

Evaluation

3. Recognition Accuracy : Unseen lighting conditions





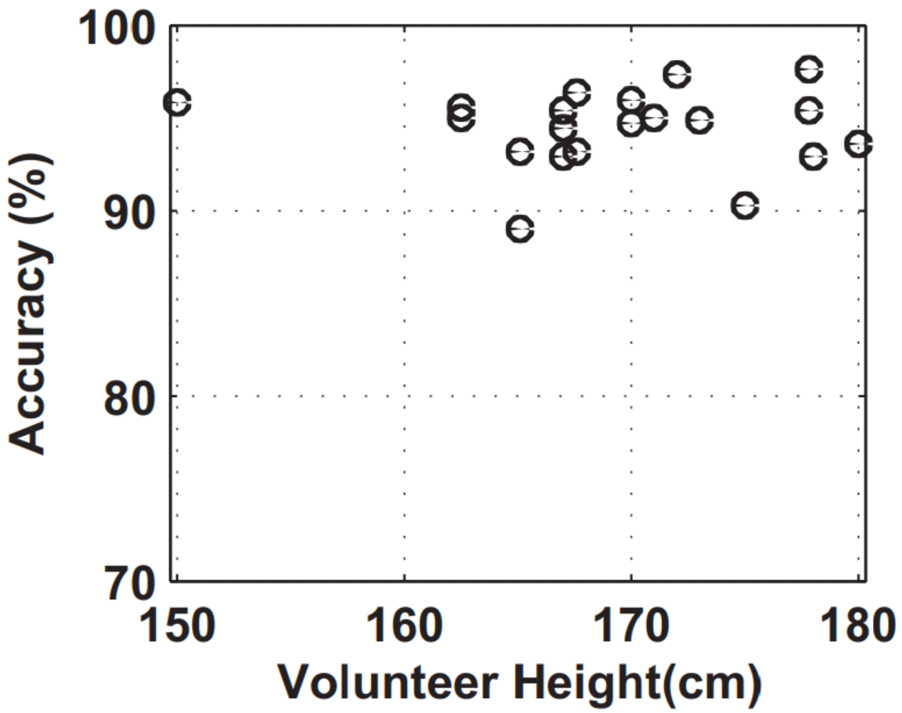
Living Room (Sunlight)
[200,1700 lux]

Lab (Fluorescent)
[280,320 lux]

Average : 96.1% , 93%

Evaluation

4. Recognition Accuracy : Unseen Users



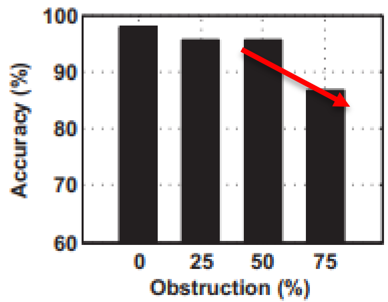
Average : 94.64%

Key Takeaways

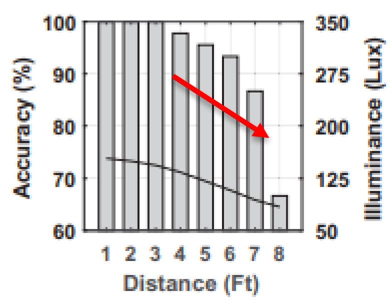
1. Demonstrated a gesture recognition system using only ambient light.
2. Developed feature extraction methods agnostic to changing lighting conditions, user positions, user orientations, users.
3. Extensively evaluated a prototype using low-cost commercially available sensors.
4. Demonstrated average accuracy (96%) comparable to existing RF-based gesture recognition systems

Limitations

1. Obstructions



2. Sensitivity to low-illumination levels (<300 Lux)



3. Users cannot walk while performing gesture