# Study on human motion recognition using dimensionality reduction


**Mehdy Bohlool**

Advisor: Dr. Eraldo Ribeiro

Computer Vision and Bio-Inspired Computing Laboratory

Department of Computer Sciences

# Contents

- **Motivation and Background**
- Dimensionality reduction methods
- Dataset
- Implementation
- Experiments
- Conclusion

# Motivation and Background

- Beneficence for Human Computer Interaction
  - Entertainment, surveillance and security
- High dimensionality
- Two approaches:
  - Each frame a pattern
    - Recognizing sequence of patterns
    - Ribeiro and Blackburn work[1]
  - Whole motion as a pattern
    - Which feature?
    - Papanikolopoulos and Masoud work[5]

# Contents

- Motivation and Background
- **Dimensionality reduction methods**
- Dataset
- Implementation
- Experiments
- Conclusion

# Dimensionality reduction

- Multidimensional scaling (MDS)
- Isomap
- Locally Linear Embedding (LLE)

# Multidimensional scaling (MDS)

- Approach in [10] is Implemented
- Minimizing lease square cost function

$$\sigma^2(W) = \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_{ij}\left(q_{ij}(W) - d_{ij}(X)\right)^2$$

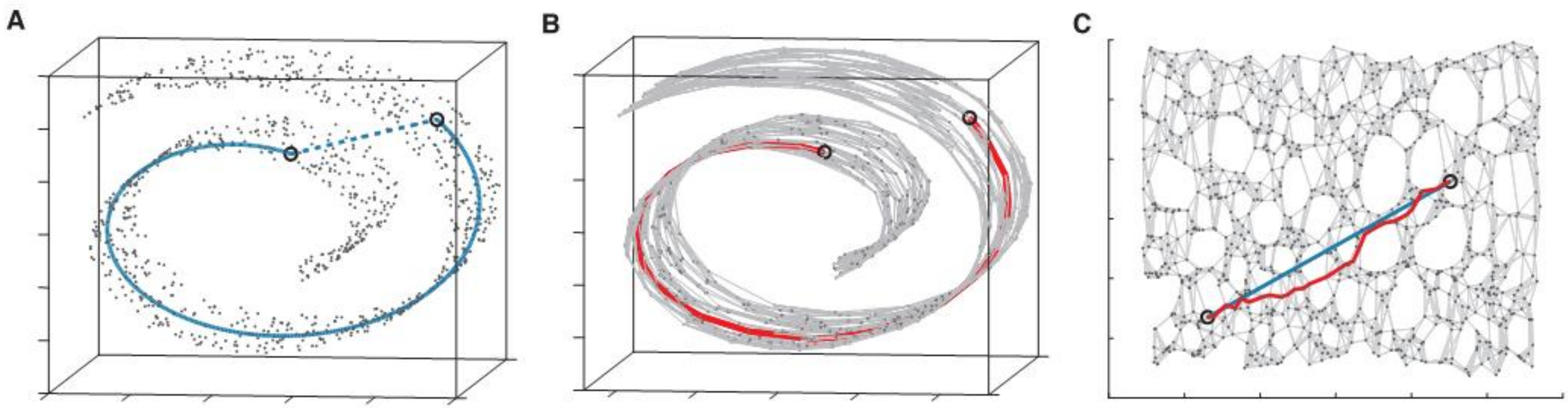
- The answer is a W matrix that satisfy this equation:

$$AW^{t+1} = D(W^t)W^t$$

$$A = \sum_i \sum_j \alpha_{ij}(x_i - x_j)(x_i - x_j)^T$$

$$D(W) = \sum_i \sum_{j, q_{ij}\neq 0} \alpha_{ij} \frac{d_{ij}(X)}{q_{ij}(W)}(x_i - x_j)(x_i - x_j)^T$$

# Isomap

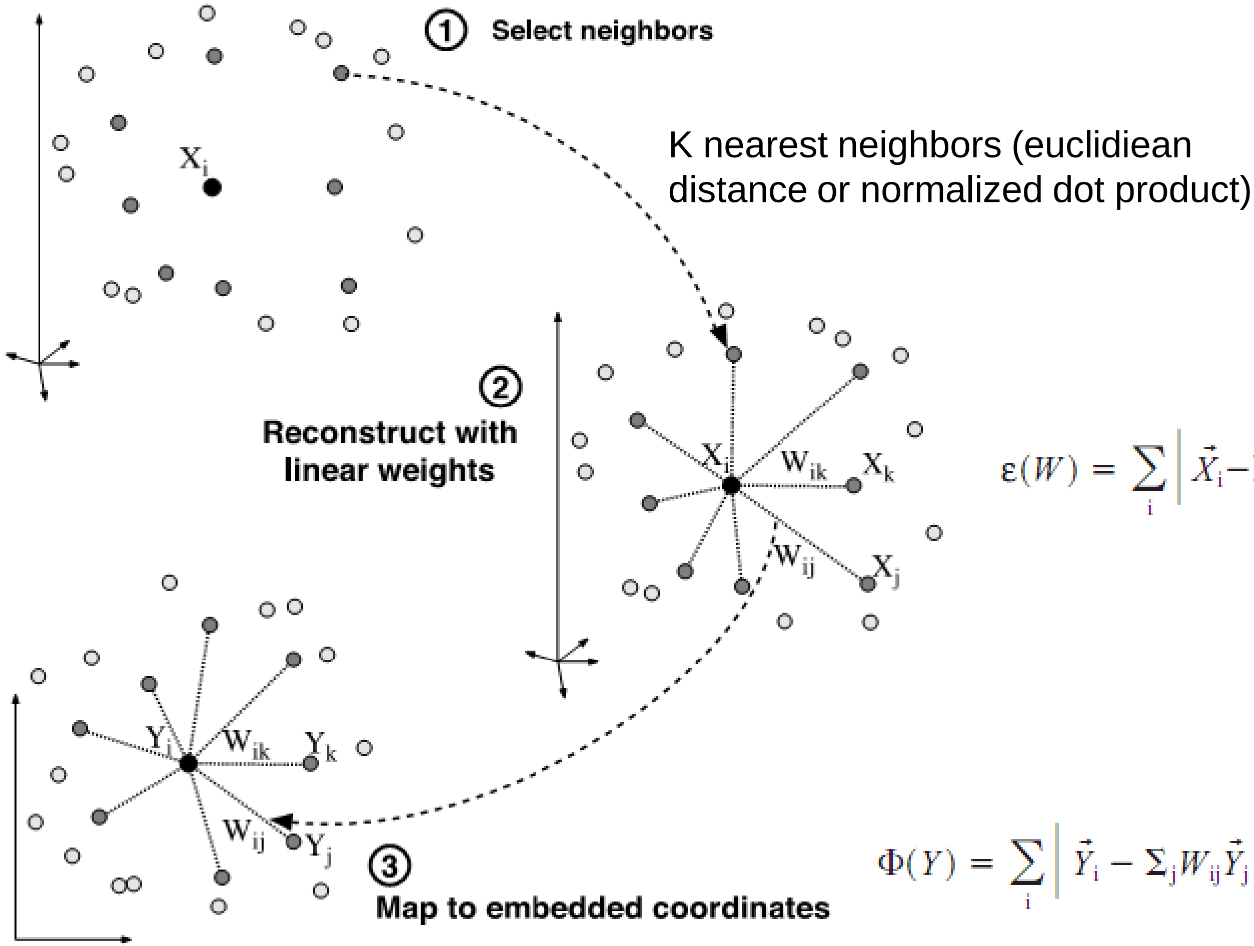



Swiss roll dataset

neighborhood graph K=7

2D embedding
recovered by Isomap

# Locally Linear Embedding (LLE)



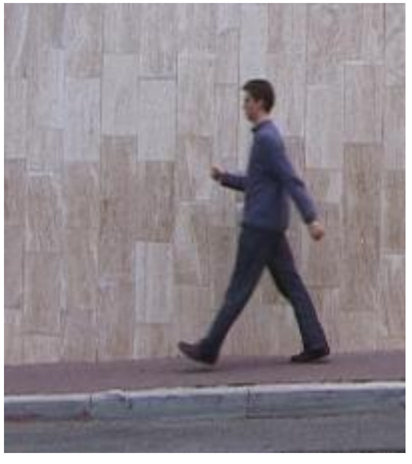

$$\varepsilon(W) = \sum_i \left| \vec{X}_i - \Sigma_j W_{ij} \vec{X}_j \right|^2$$

$$\Phi(Y) = \sum_i \left| \vec{Y}_i - \Sigma_j W_{ij} \vec{Y}_j \right|^2$$

# Contents

# Dataset

- Dataset created by Irani etal. 2005
- Consist of 9 actions with 9 or 10 sample for each action (83 samples in total)
- Backgrounds partially provided.



Walk

Jack

Run

# Contents

- Motivation and Background
- Dimensionality reduction methods
- Dataset
- **Implementation**
- Experiments
- Conclusion

# Implementation

1. **Finding Background**
2. Feature extraction
3. Dimensionality reduction
4. Classification/Recognition
5. Comparing classifiers

# Finding Background

- Backrounds were partially provided in the dataset.
- This method used for finding background points:
- A pixel belong to background if it is common between all frames in the same location.

$$B_{xy} = P_{xy}^{f} \text{ where } f = \operatorname*{argmax}_{k} (S_{xy}^{k})$$

- B is background, P is pattern, S is sum of similarities between pixel at position (x,y) with all pixels in the same locations on other frames.

# Finding Background

# Implementation

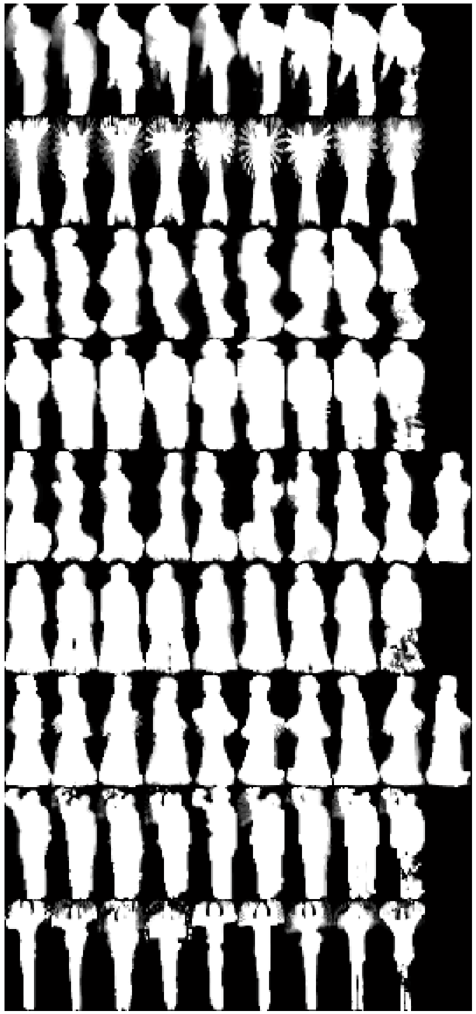
1. Finding Background
2. **Feature extraction**
3. Dimensionality reduction
4. Classification/Recognition
5. Comparing classifiers

# Feature Extraction

- Two type of features tested
- Motion History:
  - Collect all binarize foregrounds in one image using this equation: $F^{t+1} = \alpha F^{t} + (1-\alpha) I_t$
- Sum all frames together and normilize it

# Feature Extraction

Motion History



Bend

Jack

Jump

P-Jump

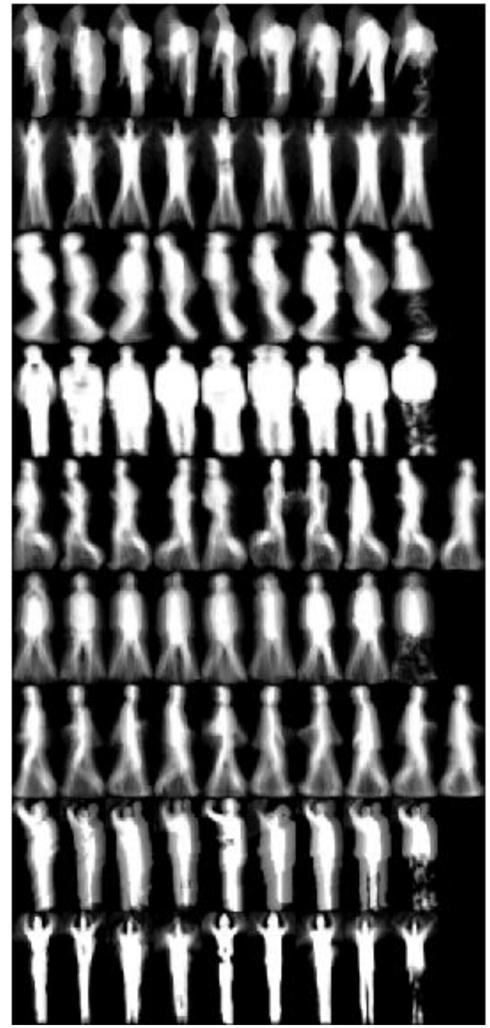
Run

Side

Walk

Wave

Wave2

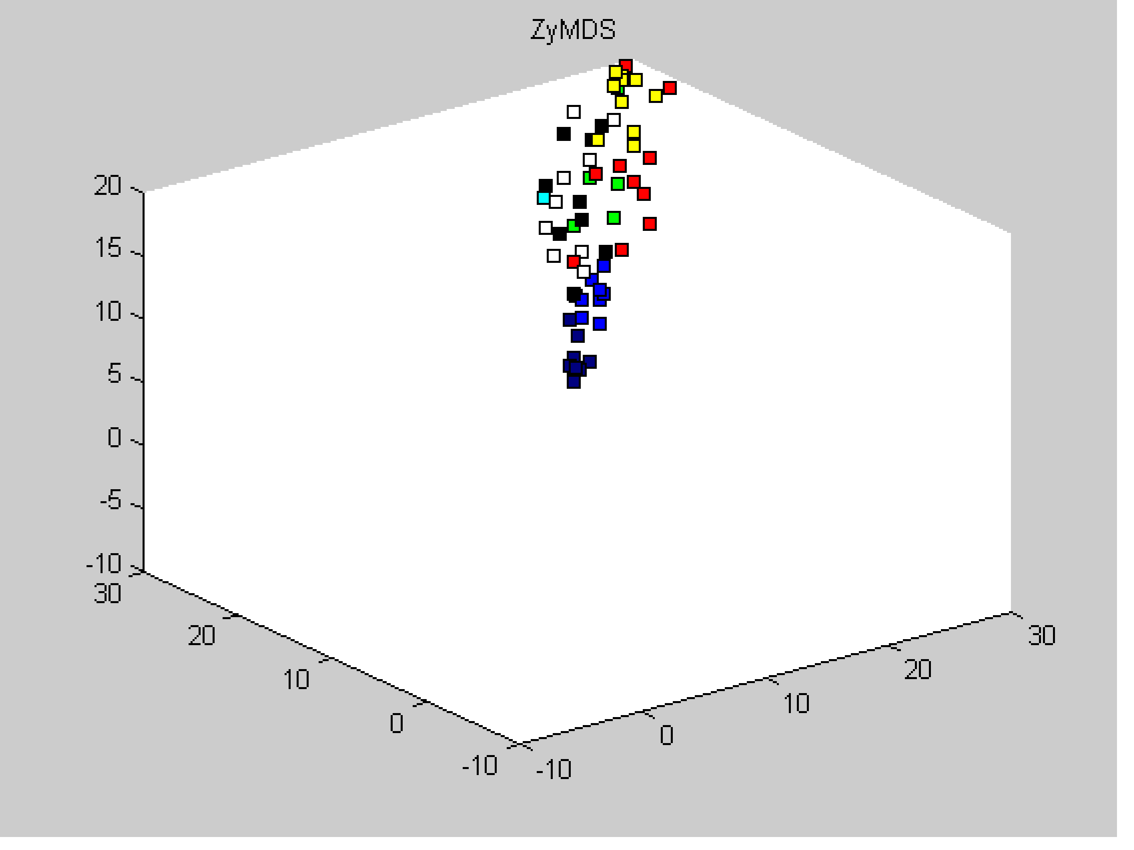


Normalized Sum

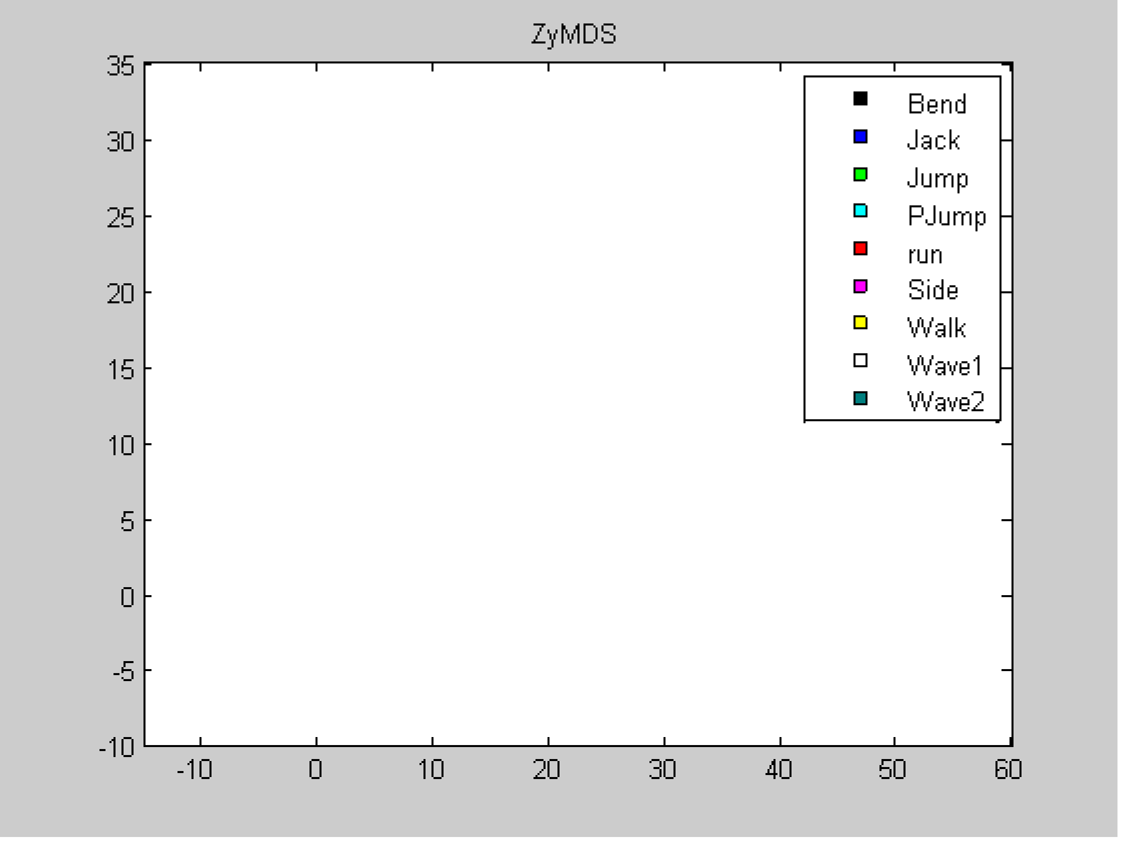
# Implementation

1. Finding Background
2. Feature extraction
3. **Dimensionality reduction**
4. Classification/Recognition
5. Comparing classifiers

# MDS 3D

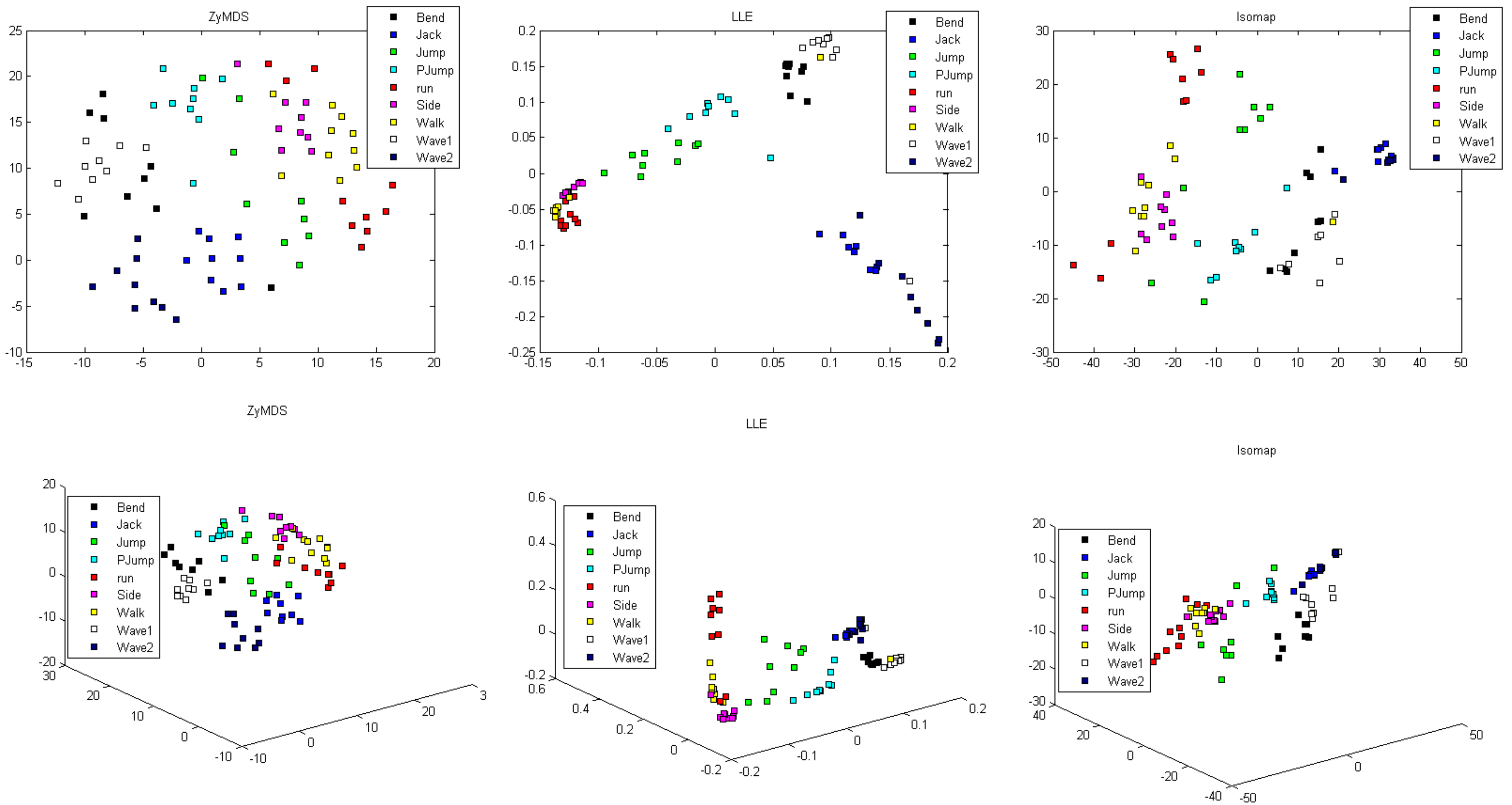


# MDS 2D

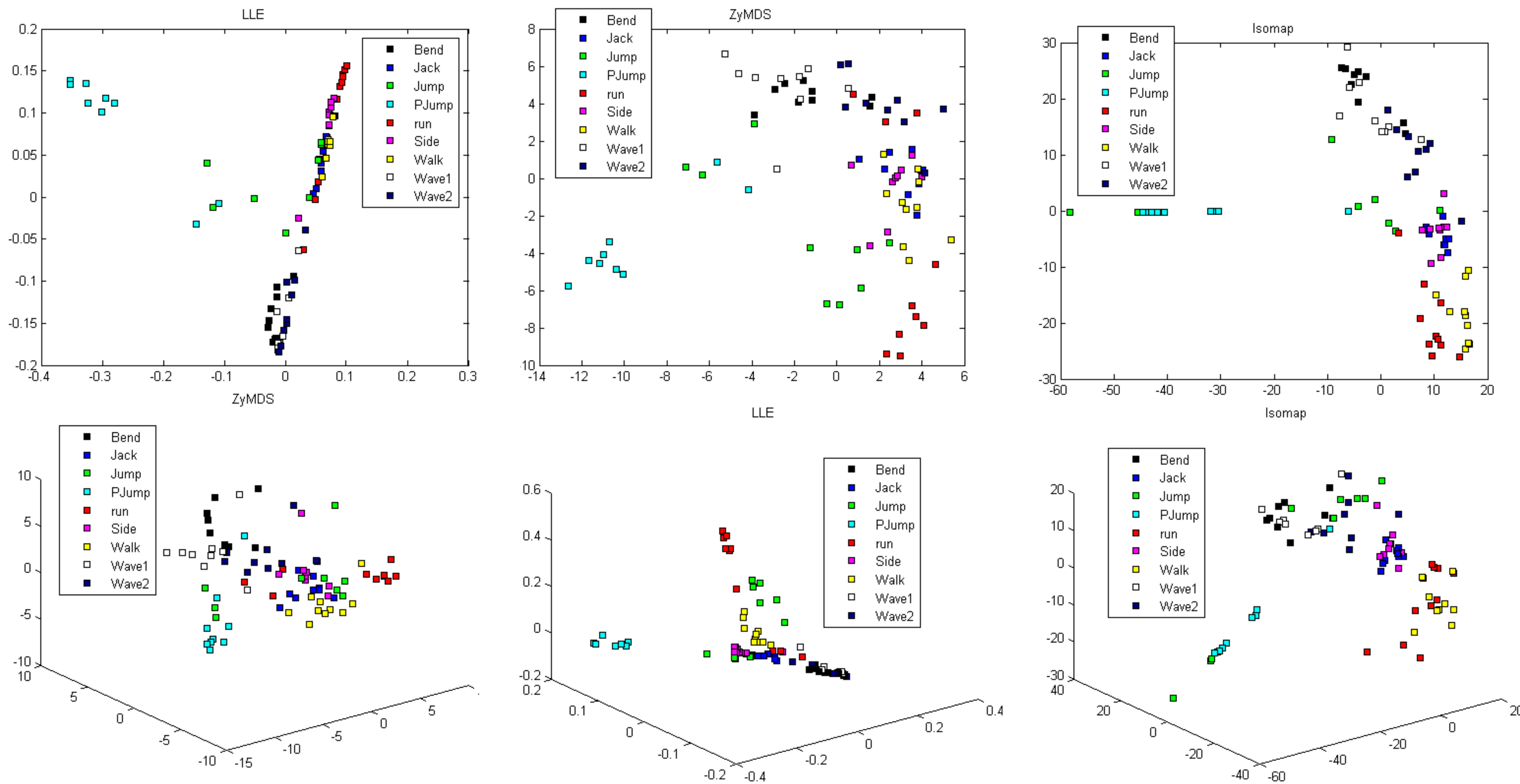



Interesting:
-Look at the Aqua (blue-green) squares.
-3 red squares at the top (3 runs in the other direction)

# Dimensionality Reduction : Feature 1



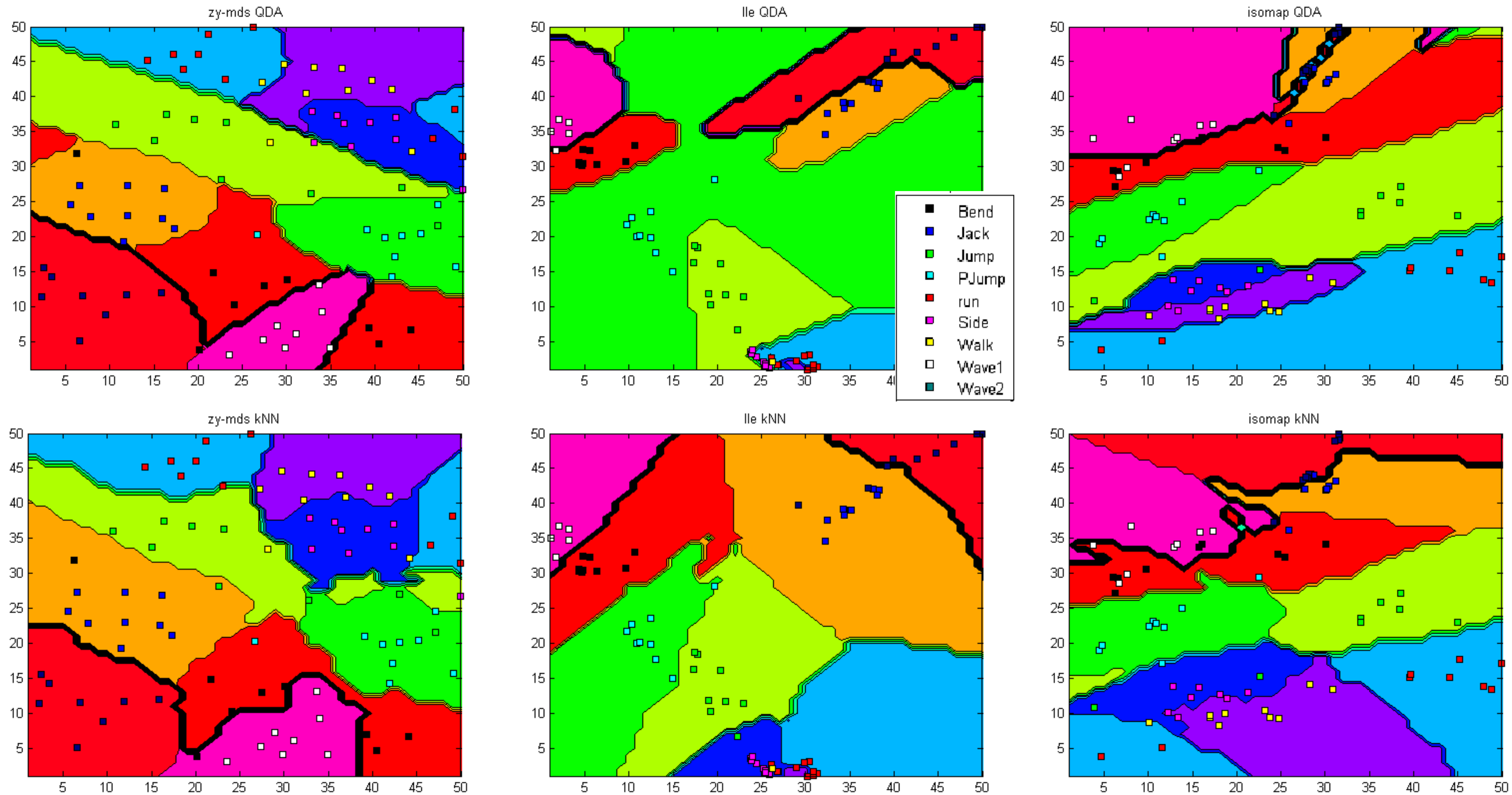
# Dimensionality Reduction : Feature 2



# Implementation

1. Finding Background
2. Feature extraction
3. Dimensionality reduction
4. **Classification/Recognition**
5. Comparing classifiers

# Classification



# Implementation

1. Finding Background
2. Feature extraction
3. Dimensionality reduction
4. Classification/Recognition
5. **Comparing classifiers**

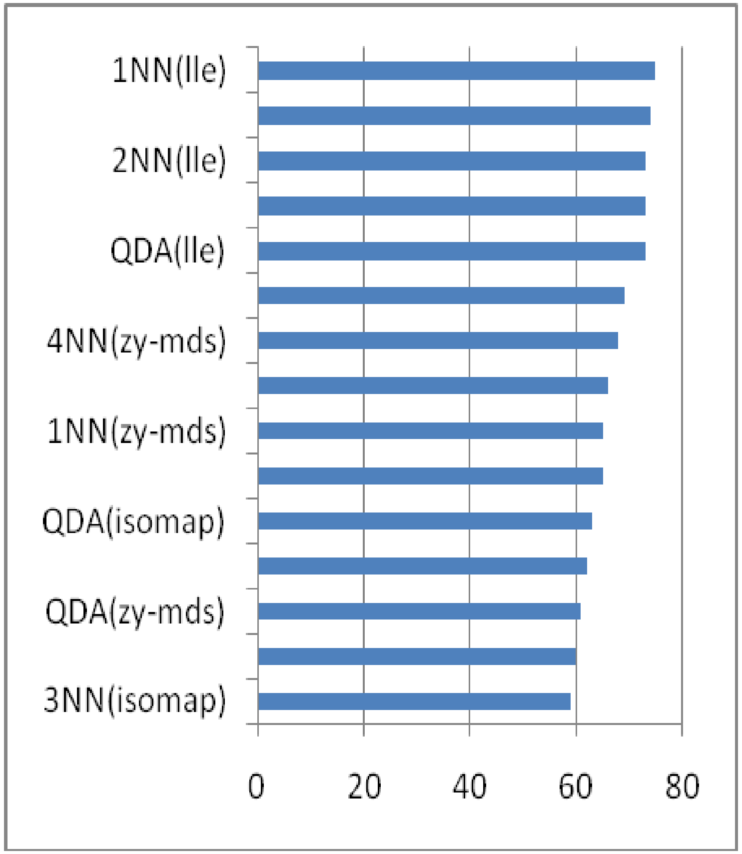
# Comparing Classifiers

- Leave one out cross validation (LOOCV)
- Fisher's Sign Test
  - $H_o$ is classifier one is not better than classifier two
  - If $\pi$ value is less than $\alpha$=0.10 then we can reject $H_o$

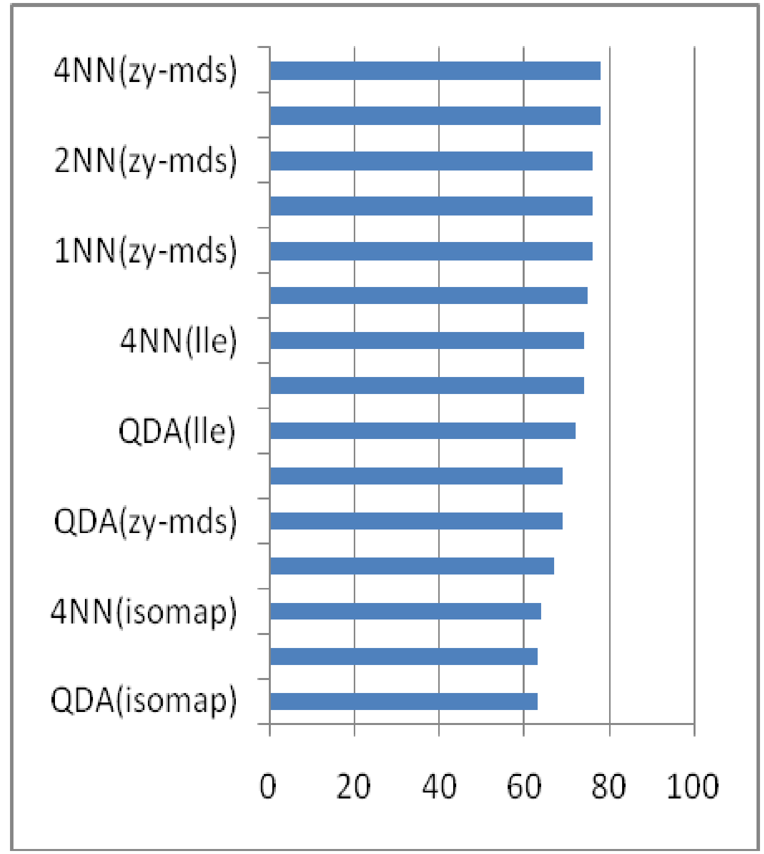
$$\pi(\hat{s}_1, \hat{s}_2, N_1, N_2) = \sum_{\hat{s}=\hat{s}_1}^{\hat{s}_m} \frac{\binom{N_1}{s}\binom{N_2}{\hat{s}_1 + \hat{s}_2 - s}}{\binom{N_1 + N_2}{\hat{s}_1 + \hat{s}_2}}$$

# Contents

# Comparing dimensions




2D




3D

# Comparing classifiers

Ho : Classifier on the rows are not more accurate than classifiers on the columns

| | 4NN(zy-mds-3) | 1NN(zy-mds-3) | 1NN(lle-2) | 3NN(lle-2) | QDA(lle-2) | QDA(lle-3) | QDA(zy-mds-3) | 1NN(isomap-3) | 3NN(zy-mds-2) | 4NN(zy-mds-2) | 2NN(isomap-3) | 1NN(isomap-2) | 1NN(zy-mds-2) | 2NN(isomap-2) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4NN(zy-mds-3) | | 0.274 | 0.267 | 0.185 | 0.084 | 0.080 | 0.019 | 0.012 | 0.019 | 0.011 | 0.004 | 0.004 | 0.002 | 0.001 |
| 1NN(zy-mds-3) | | | 0.600 | 0.490 | 0.305 | 0.296 | 0.111 | 0.080 | 0.111 | 0.076 | 0.035 | 0.033 | 0.021 | 0.014 |
| 1NN(lle-2) | | | | 0.500 | 0.314 | 0.305 | 0.116 | 0.084 | 0.116 | 0.080 | 0.038 | 0.035 | 0.023 | 0.015 |
| 3NN(lle-2) | | | | | 0.412 | 0.402 | 0.175 | 0.132 | 0.175 | 0.126 | 0.064 | 0.060 | 0.040 | 0.029 |
| QDA(lle-2) | | | | | | 0.583 | 0.316 | 0.254 | 0.316 | 0.244 | 0.143 | 0.135 | 0.097 | 0.073 |
| QDA(lle-3) | | | | | | | 0.327 | 0.265 | 0.327 | 0.254 | 0.150 | 0.142 | 0.103 | 0.078 |
| QDA(zy-mds-3) | | | | | | | | 0.514 | 0.584 | 0.500 | 0.355 | 0.341 | 0.273 | 0.224 |
| 1NN(isomap-3) | | | | | | | | | 0.650 | 0.568 | 0.420 | 0.406 | 0.332 | 0.278 |
| 3NN(zy-mds-2) | | | | | | | | | | 0.500 | 0.355 | 0.341 | 0.273 | 0.224 |
| 4NN(zy-mds-2) | | | | | | | | | | | 0.435 | 0.420 | 0.345 | 0.291 |
| 2NN(isomap-3) | | | | | | | | | | | | 0.563 | 0.484 | 0.424 |
| 1NN(isomap-2) | | | | | | | | | | | | | 0.500 | 0.439 |
| 1NN(zy-mds-2) | | | | | | | | | | | | | | 0.516 |
| 2NN(isomap-2) | | | | | | | | | | | | | | |
| correct classification | 78 | 76 | 75 | 74 | 73 | 72 | 69 | 69 | 69 | 68 | 67 | 66 | 65 | 65 |

# Comparing actions: Confusion matrix

Confusion matrix for best (approximately) classifier: QDA, LLE, 2D

Actual

Predicted

| | Bend | Jack | Jump | PJump | run | Side | Walk | Wave1 | Wave2 |
|---|---|---|---|---|---|---|---|---|---|
| Bend | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Jack | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Jump | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| PJump | 0.000 | 0.000 | 0.222 | 0.778 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| run | 0.000 | 0.000 | 0.000 | 0.000 | 0.800 | 0.000 | 0.200 | 0.000 | 0.000 |
| Side | 0.000 | 0.000 | 0.111 | 0.000 | 0.000 | 0.875 | 0.000 | 0.000 | 0.000 |
| Walk | 0.000 | 0.000 | 0.000 | 0.000 | 0.200 | 0.000 | 0.800 | 0.000 | 0.000 |
| Wave1 | 0.111 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.889 | 0.000 |
| Wave2 | 0.000 | 0.111 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.889 |

# Contents

# Conclusion

- Different classifier and dimensionality reduction methods applied
- Lack of enough patterns make it hard to decide on best classifier
- With more classifier even methods such as LLP can be applied.
- For future works testing more classifiers such as Parzen Window Classifier maybe feasible.

# References


1. Blackburn J., Ribeiro E., "Human Motion Recognition Using Isomap and Dynamic Time Warping", ICCV Workshop on Human Motion. Rio de Janeiro, Brazil, 2007.
2. Sminchisescu C., Kanaujia A., Li Z., Metaxas D., "Conditional Models for Contextual Human Motion Recognition", Computer Vision and Image Understanding, Special Issue on Modeling People: Shape, Appearance, Movement and Behavior, Volume 104, Issues 2-3, November-December 2006, Pages 210-220.
3. Tenenbaum J. B., Silva V., Langford J. C., "A Global Geometric Framework for Nonlinear Dimensionality reduction". Science 290 (5500): 2319-2323, 22 December 2000.
4. Wikipedia, "Principal component analysis", http://en.wikipedia.org/wiki/Principal_component_analysis, visited October 11, 2009.
5. Masoud O., Papanikolopoulos N., "A method for human action recognition". Image Vision Computing, 21(8):729–743, 2003.
6. Wang L., Suter D., "Analyzing human movements from silhouettes using manifold learning". IEEE International Conference on Video and Signal Based Surveillance, page 7, Washington, DC, USA, 2006.
7. Roweis S. T., Saul L. K., "Nonlinear dimensionality reduction by locally linear embedding". Science, 290(5500):2323–2326, December 2000.
8. He X., Niyogi P., "Locality preserving projection", Advances in neural information processing systems, 2003
9. Wikipedia, "Isomap", http://en.wikipedia.org/wiki/Isomap, visited October 11, 2009.
10. Andrew R. Webb, Multidimensional scaling by iterative majorization using radial basis functions, Pattern Recognition, Volume 28, Issue 5, May 1995, Pages 753-759, ISSN 0031-3203, DOI: 10.1016/0031-3203(94)00135-9.
11. L.J.P. van der Maaten and G.E. Hinton. "Visualizing High-Dimensional Data Using t-SNE". Journal of Machine Learning Research 9(Nov):2579-2605, 2008.
12. L. Wang, D. Suter, "Analyzing human movements from silhouettes using manifold learning.", IEEE International Conference on Video and Signal Based Surveillance, page 7, Washington DC, USA, 2006.
13. Fisher, R.A. "Statistical Methods for Research Workers". Oliver and Boyd, 1954.
14. Parzen, E., "On estimation of a probability density function and mode", The annals of mathematical statistics, vol. 33(3) pp. 1065-1076, 1962.

# Thank you!