

Captured Motion Recognition with Principal Component Analysis

Robin Chartrand
rchartra@uw.edu

February 5, 2024

Abstract

A high dimensional dataset of spatial and temporal data of a robot performing three different movements was analyzed and reduced using singular value decomposition and principal component analysis. It was found that the dimensions of the data could be greatly reduced to 6% of the original dimensions while still capturing 95% of the dataset's information. Each movement was found to have a distinct structure in PCA space, and using the centroid of each movement an algorithm was developed to predict the movement being performed by new samples with a stable accuracy of 95%.

1 Introduction

The three dimensional coordinates of each of 38 joints on the humanoid OptimuS-VD robot were recorded over a 1.4 second interval as it performed 3 different movements: walking, running, and jumping. Such a high dimensional dataset is difficult to understand and visualize so a lower dimensional representation of the data was sought after to gain an understanding of the major structures of each movement. Principal component analysis using singular value decomposition is one of the primary methods for this task and was the method used in this study. The goal was then to use the major structures to differentiate between the three different movements in order to be able to predict which movement was being performed when given a new sample.

2 Theoretical Background

Singular value decomposition (SVD) is a method of decomposing a matrix A into the product of three matrices:

$$A = U\Sigma V^T$$

If we treat A as an operator acting on a vector x then the three matrices U, Σ, V^T represent in turn the change of basis into the new space, the rescaling of x , and the rotation of x . When the data is appropriately centered around zero the U matrix is of particular importance as its columns (modes) make up a basis of principal axes pointing in the direction of increasing variance. SVD is therefore approximating the covariance matrix of a Gaussian distribution centered at 0 with the modes making up the eigenvectors of the covariance

matrix. Σ is a diagonal matrix where the values along the diagonal are known as singular values σ_i , which are the eigenvalues of the covariance matrix. These singular values can be used to quantify the Frobenius norm or “energy” E added by each mode to the representation of the data. By looking at the cumulative energy (1) for each mode we can determine how good of an approximation of the original dataset we get from a k mode representation.

$$CE_k = \sum_{i=1}^k \frac{\sigma_i^2}{E_{tot}} \quad (1)$$

Once energies are analyzed in many cases it is clear that most of the energy can be captured by a lower dimensional representation of the data. Unwanted modes can be removed and the original data can be either reconstructed through SVD or projected onto the lower dimensional PCA space. This latter method can be helpful for visualization. By keeping just the first two or three pca modes the data is now in a plottable form that allows you to analyze the dominant structures in the data.

3 Algorithm Implementation and Development

The data was first compiled into a 114 by 1500 array. The columns were composed of the three movements each having 5 samples with 100 timeframes, and the rows were the x dimensions, y dimensions, and z dimensions of each of the 38 joints of the robot. The mean of the data was then centered at zero and SVD was performed using the Numpy python package [1]. The resulting U matrix had 114 modes (Figure 1), which were used to visualize lower dimensional representations of the data using the Plotly [3] and Matplotlib [2] python libraries. In addition, the singular values were used to calculate the cumulative energy captured by each additional mode used in the representation.

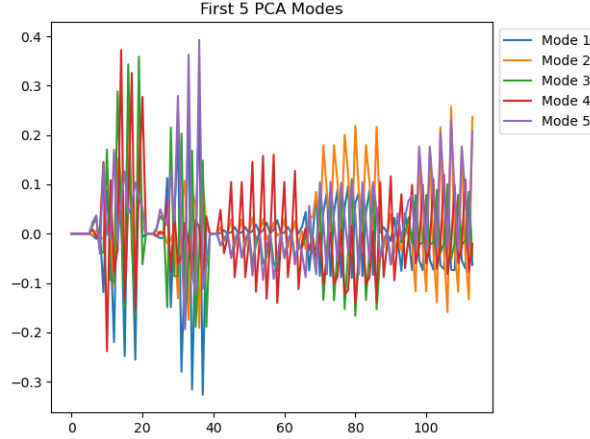


Figure 1: The first 5 PCA modes (columns of the U matrix) of the dataset

Next, for each k value, the training data set was projected onto the k mode PCA space. The projection was then divided by movement and the centroid of the samples for each movement was calculated. Our method for classifying each sample was then to calculate the centroid of each movement in PCA space, then determine which centroid the sample was

closest to when projected onto the same PCA space. The sample was then labeled according to the centroid with the shortest Euclidean distance from the sample. Each training sample was labeled according to this method and then the labels were compared to the known ground truth labels using the python library Scikit-Learn’s [4] `accuracy_score()` method to determine the accuracy of the training. The method was then used on 3 test samples withheld from the training process to get a more realistic sense of the method’s accuracy.

4 Computational Results

The PCA modes were found and then the cumulative energy was calculated for each PCA mode from the singular values (Figure 2).

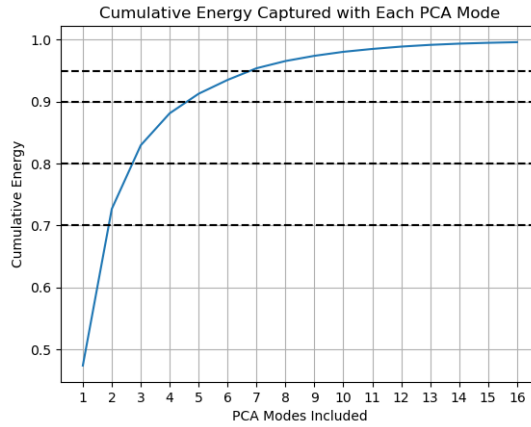


Figure 2: The cumulative energy captured with each new mode added to the representation for the first 16 modes.

We can see that 2 PCA modes captured approximately 70% of the energy. Using just 2 PCA modes the dataset can be visualized as a 2D plot. By projecting the original data onto each PCA mode we can determine the dominant structures of each movement (Figure 3).

There are clear differences between each of the three movements. Walking and running look similar but at different levels of the second mode. This makes intuitive sense as running and walking are very similar movements; the movement is just faster when one is running. In contrast, jumping has little variation along the first modal axis, while having a similar amount of variation as the other two movements on the second modal axis. This makes it seem that the first PCA mode captures outward movement of the body. When you’re jumping you mostly move up and down without extending your limbs outward very much, in contrast to running and walking where your limbs extend much more in both directions. The second modal axis can be interpreted as capturing the energy or variation of the movement, with running involving more changes in body position than walking, and walking involving more changes in body position than jumping.

Three PCA modes capture approximately 80% of the energy of the data (Figure 2). With a three dimensional truncated PCA representation the dataset can be visualized as a 3D plot (Figure 4).

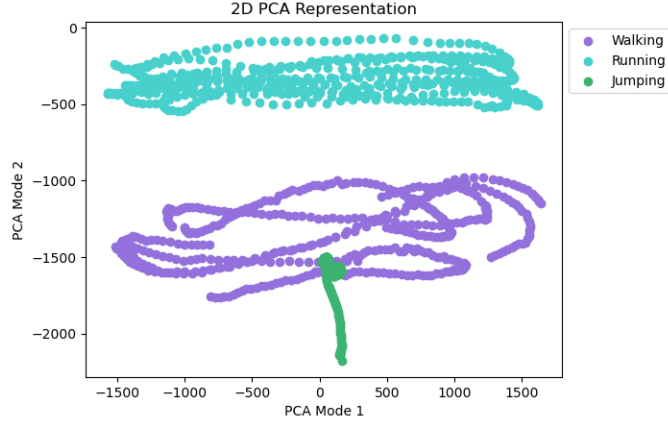


Figure 3: The two PCA mode projection of the data labeled by movement performed.

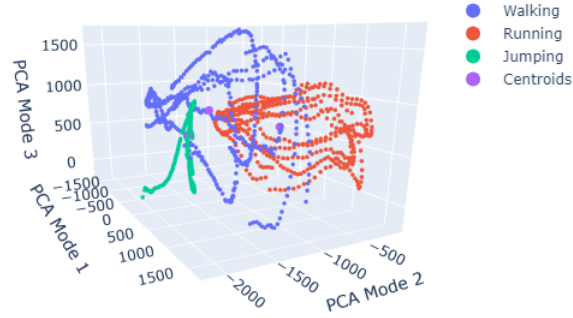


Figure 4: The three PCA mode projection of the data labeled by movement performed. The 3D mean (centroid) of each movement is also shown.

There are very similar structures to the 2D representation when viewed along the 1st and 2nd mode axis and each movement has distinct locations in the 3D space, though walking and jumping are closer together and overlap a bit. Each movement spans a similar range along the third dimension, with walking having a slightly greater range than running and jumping which are relatively similar. In order to capture 90% of the energy one would need 5 modes and to capture 95% one would need 7 modes. These dimensions can no longer be visualized, but in terms of dimension reduction this is an incredible result. 95% of the information can be captured by just 6% of the dimensions! The fact that each movement has a distinct location in PCA space serves as a method to predict which movement is being performed in an unlabeled sample. The centroid for each movement was found (Figure 4) and for each sample the distance to each centroid was calculated. The sample was then labeled as performing the movement whose centroid was closest to the sample in PCA space. Figure 5 shows the accuracy of the method When performed on the training samples for each dimension k of the PCA space.

Just two modes had an accuracy of 88%, and at around 11 modes the accuracy was

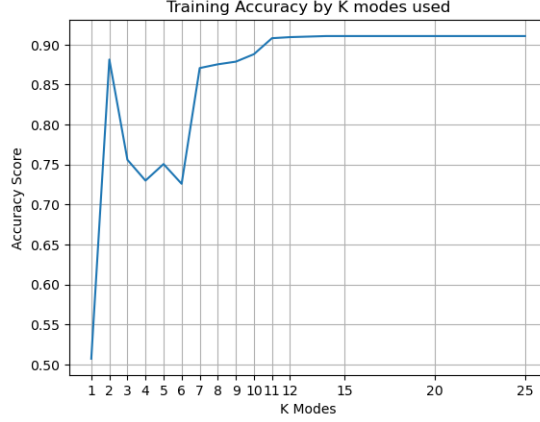


Figure 5: Percentage of correctly labeled training samples for each k modes forming the PCA space.

approximately 91%. Adding further modes beyond this point had a negligible effect on the accuracy of the method, so $k = 11$ is the optimal k value for this classification method. It is quite interesting however that while 2 modes had a high accuracy, the accuracy decreases to around 75% when using 3 to 6 modes before spiking back up to 87% at 7 modes. Upon inspection of the labels the most common misclassification is between walking and jumping. This makes sense when we look at the visualizations of the movements in PCA space (Figure 3) as these two movement structures were close together and overlapped a bit while running had a structure more distinct from the rest.

The method was then tested on 300 time samples that were not used to determine the centroids to see if the method was successful for predicting the movement of new samples as well. The accuracy score for each dimension k of the PCA space is shown in Figure 6.

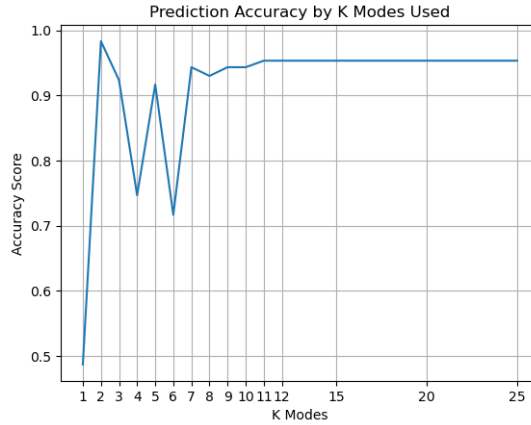


Figure 6: Percentage of correctly labeled test samples for each k modes forming the PCA space.

Once again just two modes had an outstanding accuracy, this time correctly classifying

98% of the samples. The accuracy is once again unstable as k increases until around 11 modes where the accuracy levels out at around 95%. The testing accuracy ended up being higher overall than the training accuracy, which is quite an interesting and unusual result.

5 Summary and Conclusions

Singular value decomposition and principal component analysis were an effective method for dimension reduction in our dataset allowing the dimensions to be greatly reduced. This was helpful for understanding the dominant structures of the robot’s movement when performing each movement by allowing us to visualize the data. The dimension reduction further shows that not every variable is needed to capture most of the variance in the data. This means it is possible to still have valuable data in the situation where some of the joint’s sensors aren’t functional or reporting accurate data. Using this truncated PCA space we can predict the movement being performed by a new sample with high accuracy by seeing which movement structure the sample is closest to in truncated PCA space. While two modes tends to have a high accuracy for this prediction algorithm, for practical implementation it is safest to include 11 modes as it has the highest stable accuracy and it includes much more of the cumulative energy in its prediction. However in cases where the dimensions need to be severely reduced 2 modes will work quite well.

6 Acknowledgements

I would like to thank Professor Shlizerman for his direction on using singular value decomposition for dimension reduction. I am also grateful to Saba Heravi for her direction on centroids in k mode pca space. Finally I would like to thank Grace Zhou for her discussion on projecting samples onto truncated PCA space.

References

- [1] C. R. Harris, K. J. Millman, and et. al. Array programming with numpy. *Nature*, 585(7825):357–362, 2020.
- [2] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.
- [3] P. T. Inc. Collaborative data science, 2015.
- [4] F. Pedregosa, G. Varoquaux, and et. al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.