An Application of Graph Diffusion for Gesture Classification

by

Perry Samuel Voisin

Department of Statistical Science
Duke University

Date: _____

Approved:

_____

Fan Li, Advisor

_____

Sayan Mukherjee

_____

David Banks

_____

John Harer

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science
in the Department of Statistical Science
in the Graduate School of
Duke University

2020

ABSTRACT

An Application of Graph Diffusion for Gesture Classification

by

Perry Samuel Voisin

Department of Statistical Science
Duke University

Date: _____

Approved:

_____

Fan Li, Advisor

_____

Sayan Mukherjee

_____

David Banks

_____

John Harer

An abstract of a thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science
in the Department of Statistical Science
in the Graduate School of
Duke University

2020

Abstract

Reliable and widely available robotic prostheses have long been a dream of science fiction writers and researchers alike. The problem of sufficiently generalizable gesture recognition algorithms and technology remains a barrier to these ambitions despite numerous advances in computer science, engineering, and machine learning. Often the failure of a particular algorithm to generalize to the population at large is due to superficial characteristics of subjects in the training data set. These superficial characteristics are captured and integrated into the signal intended to capture the gesture being performed. This work applies methods developed in computer vision and graph theory to a set of time series modalities with the goal of identifying pertinent features and diffusing superficial noise.

Acknowledgements

Many thanks to my committee members, Dr. Sayan Mukherjee, Dr. David Banks, and Dr. John Harer, for their guidance on this project. I would like to further express my gratitude for Dr. John Harer for inspiring this project and introducing me to the field of topological data analysis. The many office hours spent discussing math and science were well spent. A special thank you to Dr. Paul Bendich for opening my mind to the blessings and curses of high-dimensional geometry.

I must express my deepest appreciation to my wife, Taylor. Few of my accomplishments at Duke would have been possible without her constant support.

Contents

Abstract **iv**

Acknowledgements **v**

List of Figures **viii**

List of Tables **ix**

1 Introduction **1**

1.1 Background 1

1.2 Research Question 2

2 Related Work **3**

2.1 Graph Methods for Metric Fusion 3

2.2 Cross-Modal Sensor Fusion for Time Series Data 4

3 Data **5**

3.1 Data Decription 5

3.2 Data Preparation and Pre-Processing 6

4 Methodology **8**

4.1 Definitions 8

4.2 Self-Similarity 9

4.3 Sensor Fusion 11

4.4 Persistence Diagrams and Images 13

4.5 Image Comparison via CNN 14

5 Results **16**

6 Conclusions **18**

Bibliography **19**

List of Figures

3.1 sEMG recordings as captured by the Myo armband. 7

3.2 sEMG modalities after RMS smoothing. 7

4.1 SSMs for each gesture class performed by one subject. 10

4.2 Multiple performances "Left" gesture performed by one subject. . . . 10

4.3 Left" gesture performed by different subjects. 11

4.4 Examples fused similarity matrices for each gesture class. 12

4.5 Example Persistence Diagram and Persistence Image for SSM. 13

4.6 Example Persistence Diagram and Persistence Image for a fused similarity template. 14

List of Tables

5.1 Classification accuracy scores for persistence image vectors. 16

5.2 Maximum accuracy scores for AlexNet. 17

5.3 Classification accuracy scores for unsupervised distance-based (1-NN) classification. 17

Chapter 1

Introduction

Gesture recognition is a supervised classification problem in which a gesture's membership in one of potentially many categories is inferred from a data set of sensor readings. For the purposes of this paper, methods for gesture recognition can be divided into two subsets based on the type of sensor device employed: (1) methods utilizing cameras or radar fixed to a set reference point relative to the subject performing the gesture and (2) methods utilizing sensors embedded in a device worn by the subject. This paper focuses on methods applicable to the latter. Specifically, the research presented here focuses on motion captured through an array of sensors embedded in an armband fixed to the subject's forearm.

1.1 Background

The nervous system activates skeletal muscles to contract by transmitting electrical impulses. These impulses, known as *myoelectricity*, can be detected and analyzed using incumbent medical technology known as an *electromyograph* (EMG). Traditionally, EMGs have been used by medical personnel to diagnose conditions and disorders relating to the neuro-muscular system.

There are two primary categories of EMG - the intramuscular electromyograph and the surface electromyograph (sEMG). The intramuscular EMG uses subcutaneous sensors implanted in the skeletal muscle through thin needles containing electrodes. Alternatively, the surface electromyograph only requires two or more electrodes to be adhered to the surface of the subject's skin. This approach is less invasive than the intramuscular EMG but comes at the cost of decreased precision. A patient may exhibit superficial characteristics (e.g. perspiration, body-fat content, fatigue, etc.) which are detrimental to quality of sEMG sensor readings [LKK+18].

The problem of gesture recognition using electromyography has captured researchers' attention since the late 1960s. A considerable amount of research has been performed since that time with the goal of developing and improving robotic prosthetic limbs. In fact, the *Non-Invasive Adaptive Hand Prosthetics* ("NinaPro") project now acquires, stores, and distributes access to data sets of sEMG recordings for this reason [AGC+14] [PS18]. A number of private firms have also conducted research in the area of gestures recognition with hopes of bringing advanced human-computer interfaces (HCIs) to market as wearable technology or "wearables" become increasingly popular. Researchers in the domain have routinely achieved classification accuracy scores in excess of 85% using primarily linear discriminant analysis and artificial neural-network classifiers. However, myoelectric prostheses and other HCIs have not become popular despite decades of positive results. There are many reasons for this, but one of the most significant issues is poor generalizability of the products to the population at large [PBW+11].

1.2 Research Question

Recently developed methods for metric fusion (section 2.1) and time-series data analysis (section 2.2) provide new tools with which to analyze output from a variety of sensors. The research presented here seeks to determine whether these methods might benefit the field of gestures recognition by identifying key features of sEMG recordings. It may also be possible to improve the classification accuracy of supervised gesture recognition algorithms by pre-processing sEMG recordings to amplify their identifying features prior to model fitting.

Chapter 2

Related Work

2.1 Graph Methods for Metric Fusion

In their 2012 paper *Unsupervised Metric Fusion by Cross Diffusion* [WJW+12], Wang *et al.* develop an algorithm for combining an arbitrary number of features or "views" of an observation sampled from a manifold. In this work, the authors generate two graphs, G and $\mathcal{G}$. The vertices of both graphs are the same. However, G is a complete graph whose edges are defined by a similarity function between vertices, and therefore G is a representation of global similarity. The vertices in $\mathcal{G}$ are connected to only the nearest k vertices meaning $\mathcal{G}$ is a representation of local similarity. Both G and $\mathcal{G}$ may be created for each view of an observation. For example, in a data set consisting of images, the different views may the raw image (1) and an edge-detected version of the image (2). Then we would have a set of four graphs, G_1, $\mathcal{G}_1$, G_2, and $\mathcal{G}_2$. A cross-diffusion process is performed between these graphs for a pre-specified number of iterations, and the adjacency matrices at the final iteration are averaged together. The result of the process is a single adjacency matrix representing a graph whose edges represent similarity between points through a similarity measure that takes into account the global and local information of the input views.

Readers seeking a more thorough explanation of the algorithm are encouraged to read the original paper [WJW+12] as well as the author's foundational paper on graph diffusion, *Affinity Learning via Self-diffusion for Image Segmentation and Clustering* [WT12].

2.2 Cross-Modal Sensor Fusion for Time Series Data

There are two motivating works for the research presented here. The first is *Geometric Cross-Modal Comparison of Heterogeneous Sensor Data*, by Tralie, *et al.* [TSB+18].The authors detail the construction of *Self-Similarity Matrices* (SSMs) derived from time series data in a d-dimensional Euclidean space, termed a *time-ordered point cloud* or "TOPC" by the authors, generated by a number of sensors focused on some moving object. The construction of SSMs as it relates to gesture recognition is detailed in Chapter 4 (Methodology). Traile *et al.* also describe two methods of comparing SSMs. The first is *Isometry-blind Dynamic Time Warping* (IBDTW) [Tra17] while the second method compares topological features of SSMs by way of their persistence diagrams.

The second motivating paper is *Multi-scale Geometric Summaries for Similarity-based Sensor Fusion*, by Tralie, Bendich, and Harer [TBH19]. In their paper, Tralie, Bendich, and Harer use the metric fusion algorithm described in section 2.1 to fuse SSMs generated from differing modality types (e.g. audio and video recordings). The method is used to successfully identify patterns in a spoken series of digits. This illustrates that the SNF approach is appropriate for application to time series data similar to the gestures data presented here.

Chapter 3

Data

3.1 Data Decription

The data set used for this research is the *EMG Data for Gestures Data Set* hosted by the University of California, Irvine Machine Learning Repository [Rep19]. This data set was collected by Lobov, *et al.* for their 2018 paper [LKK+18]. In this paper subjects were asked to perform a series of six to seven distinct gestures. Each gesture was performed four times by each subject. The myoelectric readings for each gesture are recorded by a device known as the *Myo gesture control armband* which is situated on the subject's forearm. The armband is equipped with eight evenly spaced sensors. The armband was positioned so that four sensors were situated directly atop four main muscles in the forearm - *flexor carpi radialis*, *flexor carpi ulanis*, *extensor carpi radialis longus*, *extensor digitorum*, and *extensor carpi ulanis*.

The data set consists of seven distinct gesture classes: a hand at wrest, a hand clenched in a fist, an extended palm, and four directional gestures in which the subject angles their hand at the wrist. The four directional gestures were performed using the right hand, fingers extended, and palm facing in toward the torso. The specifics of each gesture are as follows:

- Gesture 1 ("left"): The subject angles the hand inward to the left toward the torso (wrist flexion).
- Gesture 2 ("right"): The subject angles the hand outward to the right away from the torso (wrist extension).
- Gesture 3 ("up"): The subject angles the hand upward while maintaining the orien-

tation of the palm (radial deviation).

- Gesture 4 ("down"): The subject angles the hand downward while maintaining the orientation of the palm (ulnar deviation).

The methods presented here were applied to the four directional gestures only. This is because each gesture in that subset represents a mirror image of a counterpart gesture also in that subset. The "left" gesture is the counterpart of the "right" gesture; the "up" gesture is the counterpart of the "down" gesture. The muscles involved in performing a given gesture and its counterpart alternate synergist and antagonist roles, and the four sensors placed atop the four main muscles capture the motion accordingly. The extended palm and clenched fist also represent counterpart gestures but were ultimately excluded from the analysis because few subjects performed the extended palm gesture.

3.2 Data Preparation and Pre-Processing

Raw electromyograph recordings captured by the Myo gesture control armband are characterized by high volatility in amplitude. The recordings also tend to truncate peaks and troughs of the waveforms. This is due to insensitivity in the sEMG sensors in the sEMG sensors [LKK+18]. Figure 3.1 depicts an example of raw electromyograph recordings.

A common method for smoothing modalities such as those depicted in figure 3.1 is by smoothing the modalities as a pre-processing step using a root mean square sliding window. The root mean square for modality m of a window beginning at time-index w with window size n is calculated as

$$RMS(m, w) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} m[w \cdot i]^2} \quad \text{where } [w \cdot i] \text{ indexes modality } m \tag{3.1}$$

Another parameter that must be chosen for RMS smoothing is the step size between windows. Lobov, *et al.* chose a window size of $200ms$ with a $100ms$ step size [LKK+18]. A $100ms$ window size and a $50ms$ step size were chosen as the RMS parameters for the analysis presented below on the basis of some simple sensitivity analysis.

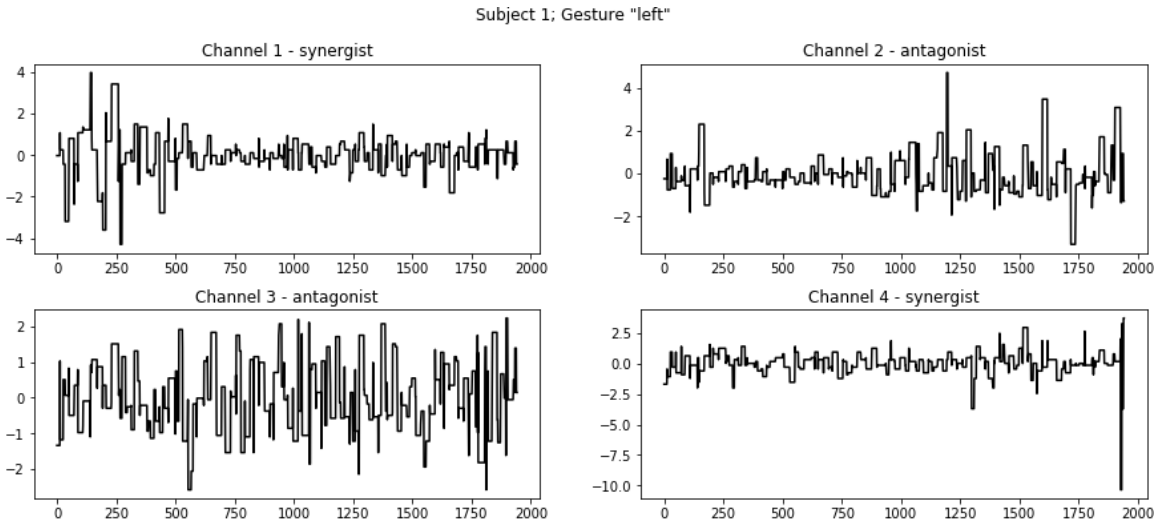

Figure 3.1: sEMG recordings as captured by the Myo armband.

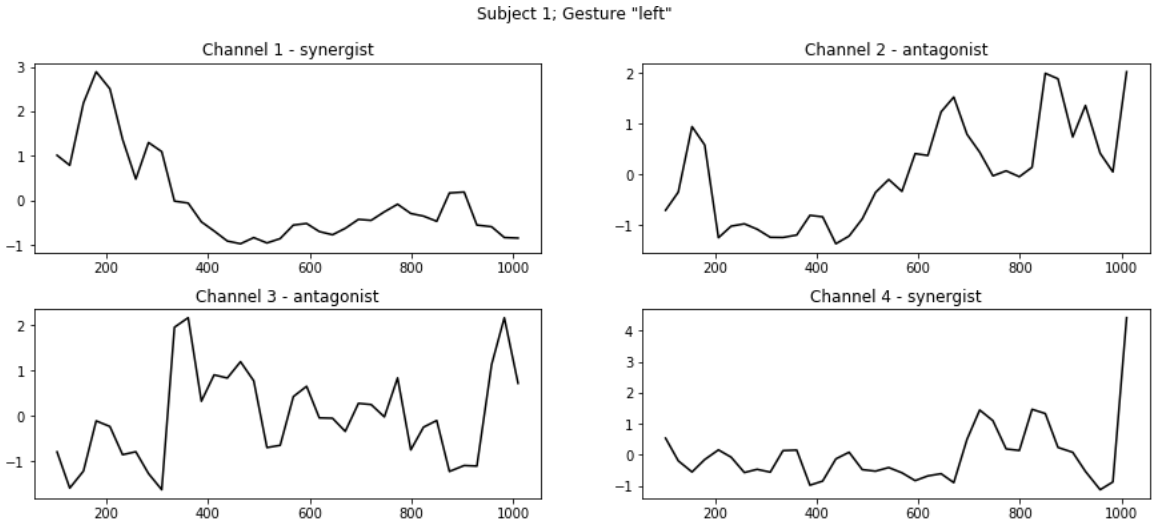

Figure 3.2: sEMG modalities after RMS smoothing.

Chapter 4

Methodology

4.1 Definitions

As described in the preceding sections, our primary goal is the fusion of a number of views or features of a gesture we are seeking to classify. To this end we must first rigorously define the problem at hand and the objects of interest in our data set.

Let $[\tau_0, \tau_n]$ be a continuous interval, and define $s_\ell : [\tau_0, \tau_n] \to \mathbb{R}$ to be a continuous map for indices $\ell = 1, 2, \ldots, m$. Further, let $S : [\tau_0, \tau_N] \to \mathbb{R}^m$ be a continuous map with a vector-valued output defined as

$$S := \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \end{bmatrix}$$

Finally, we may define a continuous interval of tuples in the form $(\tau_i, S(\tau_i))$ and ordered by τ:

$$\mathcal{A} = \left\{ [\ (\tau_0, S_\gamma(\tau_0)), (\tau_n, S_\gamma(\tau_n))\ ] : \gamma \in \mathcal{G} \right\}$$

where $\mathcal{G}$ is the set of all possible maps S.

From a practical standpoint, we may view the set of sEMG sensors as the maps $s_1, \ldots s_\ell$ as each sensor outputs a value $s_k(t_i) \in \mathbb{R}$ for a given point t_i sampled discretely from the continuous interval $[\tau_0, \tau_N]$. We may now characterize the observed gestures data set as $A \subset \mathcal{A}$ where

$$A = \left\{ (t_0, S_g(t_0)), (t_1, S_g(t_1)), \ldots, (t_n, S_g(t_n)) : g \in G \right\}$$

where G is the set of *observed* maps S sampled from $\mathcal{G}$.

4.2 Self-Similarity

We may now define R to be a self-similarity matrix (SSM) with elements $R_{ij} = f(\mathbf{x}_i, \mathbf{x}_j)$ where f is a distance metric or similarity kernel of some kind, the most common of which being the Gaussian kernel. Given a suitable choice of metric, the SSM provides a means of visually comparing modalities as they develop over time. Traile, *et al.* provide numerous examples of this under a variety of conditions in [TSB+18] and [TBH19]; Traile also depicts examples of *cross-similarity timewarping matrices* in [Tra17]. Examples of SSMs generated from the gestures data set will follow shortly. However, an explicit definition of the metric ρ is necessary to calculate these SSMs.

A common algorithm for measuring similarity between time series is *Dynamic Time Warping* ("DTW") or multi-dimensional DTW ("MD-DTW"). In *Multi-Dimensional Dynamic Time Warping for Gesture Recognition*, Holt, Reinders, and Hendriks develop MD-DTW for comparing sign language gestures [ten07]. This approach does not suit our purposes for a key reason: The dynamic time warping "distance" is not a true metric as it does not satisfy the triangle-inequality. This can lead to complications for geometry-based classification algorithms as the notion of a reference point becomes ambiguous when distances are considered between more than two data points. Of course, the standard metrics used in machine learning applications (L_1, L_2, L_∞, etc.) are available, but these metrics do not take into account previous steps in the time series. To retain this information we utilize the *cumulative Euclidean distance*:

$$\rho(\mathbf{x}_k, \mathbf{x}_\ell) = \Big|\Big| \sum_{i=0}^{k} \mathbf{x}_i - \sum_{j=0}^{\ell} \mathbf{x}_j \Big|\Big|_2$$

The cumulative Euclidean distance (CED) is a true metric and may be used in the computation of SSMs.

We also define another useful metric for time-series distance-based comparisons: the fading time-series distance (FTSD):

$$\rho(\mathbf{x}_k, \mathbf{x}_\ell) = |k - \ell| \cdot ||\mathbf{x}_k - \mathbf{x}_\ell||_2$$

The FTSD metric is also non-negative, symmetric, and sub-additive. In the event that $\mathbf{x}_k = \mathbf{x}_\ell$ while $k \neq \ell$, then $\rho(\mathbf{x}_k, \mathbf{x}_\ell) = 0$ despite the penalty. We assuming a continuous CDF over the vector space, we have that $Pr(\mathbf{x}_k = \mathbf{x}_\ell) = 0$.

Some example SSMs generated using both metrics are depicted in figures 4.1, 4.2, and 4.3 below.

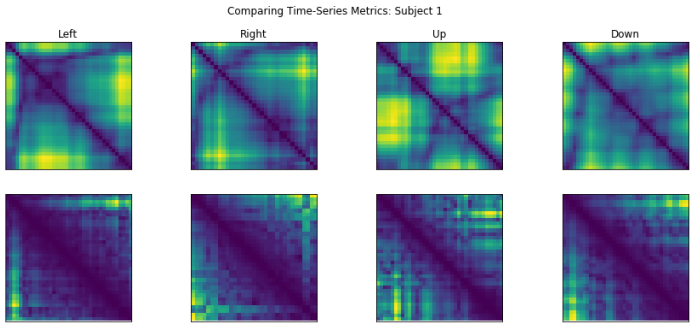

Figure 4.1: SSMs for each gesture class performed by one subject.

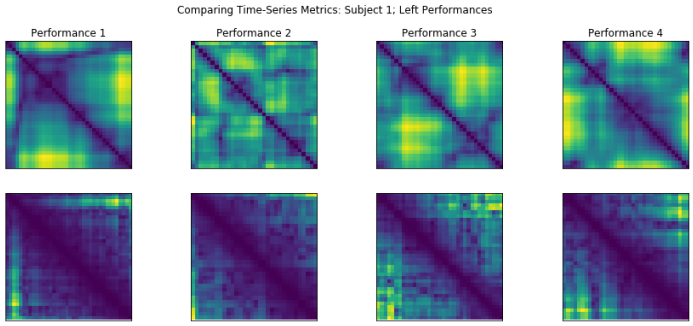

Figure 4.2: Multiple performances "Left" gesture performed by one subject.

We calculate an SSM similar to those in figures 4.1, 4.2, and 4.3 for each gesture in the data set as a baseline against which we can compare the fused SSMs to be described in section 4.3.

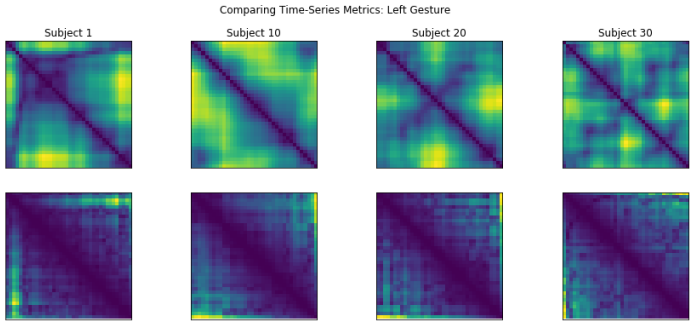

Figure 4.3: Left" gesture performed by different subjects.

4.3 Sensor Fusion

Having defined the self-similarity matrix, we may now describe the similarity network fusion process as it relates to gesture recognition. In the previous section we calculated a single SSM for each gesture which combined all four modalities using our metric of choice. Under the the metric fusion approach, we delay combining the metrics and instead calculate one SSM for each modality. Elements of this matrix are calculated using a similarity kernel and may be considered the adjacency matrix of a complete graph, $G^{(m)}$ where m represents the modality. We then calculate a second adjacency matrix $\mathcal{G}^{(m)}$ to represent a graph in which connections between vertices are made on a k-nearest neighbors basis.

We next normalize the adjacency matrices for all $2m$ graphs $G^{(1)}, \ldots, G^{(m)}$ and $\mathcal{G}^{(1)}, \ldots, \mathcal{G}^{(m)}$. The resulting local adjacency matrices, denoted $\mathcal{P}^{(1)}, \ldots, \mathcal{P}^{(m)}$ in [WJW+12]. The normalized adjacency matrices for $G^{(1)}, \ldots, G^{(m)}$ are denoted as $P^{(1)}, \ldots, P^{(m)}$ and can now be thought of as random walk over the fully connected graphs with which they are associated. The m views of a given gesture may be iteratively fused together over a set number of iterations. The cross-diffusion process as defined by Wang, *et al.* is described below for the t^{th} iteration. Note that Wang, *et al.* use the regularization parameter η to prevent

over-dispersion of the graph.

$$P_{t+1}^{(1)} = \frac{1}{m-1} \cdot \mathcal{P}^{(1)} \times \sum_{k \neq 1} P_t^{(k)} \times \mathcal{P}^{(1)\prime} + \eta I$$
$$P_{t+1}^{(2)} = \frac{1}{m-1} \cdot \mathcal{P}^{(2)} \times \sum_{k \neq 2} P_t^{(k)} \times \mathcal{P}^{(2)\prime} + \eta I$$
$$\vdots$$
$$P_{t+1}^{(m)} = \frac{1}{m-1} \cdot \mathcal{P}^{(m)} \times \sum_{k \neq m} P_t^{(k)} \times \mathcal{P}^{(m)\prime} + \eta I$$

The resulting m matrices are then averaged to create a single *fused similarity matrix*.

Tralie, Bendich, and Harer offer a slight re-parameterization of the normalized adjacency matrices P and $\mathcal{P}$ in [TBH19]. Their parameterization does not explicitly use the regularization parameter. Through some experimentation, including a small regularization parameter, $\eta \leq 1$, resulted in fused similarity matrices with the most obvious off-diagonal similarities. Some examples of fused similarity matrices are presented in figure 4.2.

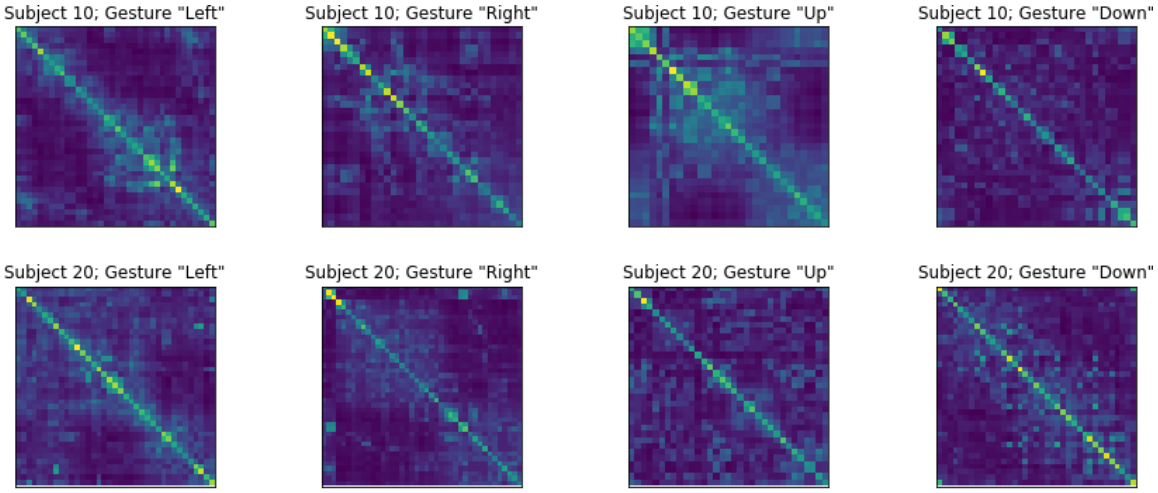

Figure 4.4: Examples fused similarity matrices for each gesture class.

4.4 Persistence Diagrams and Images

Having calculated the distance SSMs and the fused similarity templates, we may now begin the task of comparing the matrices. There is a practical problem preventing direct comparison of the observations or training a classifier on the matrices - the matrices are not of uniform size. This is due to subjects performing gestures over different time intervals.

It is possible to interpolate the images to a standard size, which we do as a point of comparison (see Ch. 5). We may also consider persistent diagrams and persistence images.

Persistence diagrams are a tool used in topological data analysis to visualize a cloud of data points by way of its topological features. In short, the persistence diagram depicts birth and death times of cycles in each homology class. Readers interested in the topic may find a more complete description of the properties of persistence diagrams in [EH10] and [CSEH05]. The *persistence image* represents an attempt to vectorize the persistence diagram in a stable manner. This improves the quality of models trained using persistence images [AEK+17].

For the purposes of this study, we benefit from the fact that we pre-define the size of the persistence image by its pixelation. Another benefit of this approach is that it takes into account the topological information of the SSM or SNF template in a manner similar to that of a person performing a visual inspection. Points that are high or low in the SSM will appear similar and have similar birth and death times.

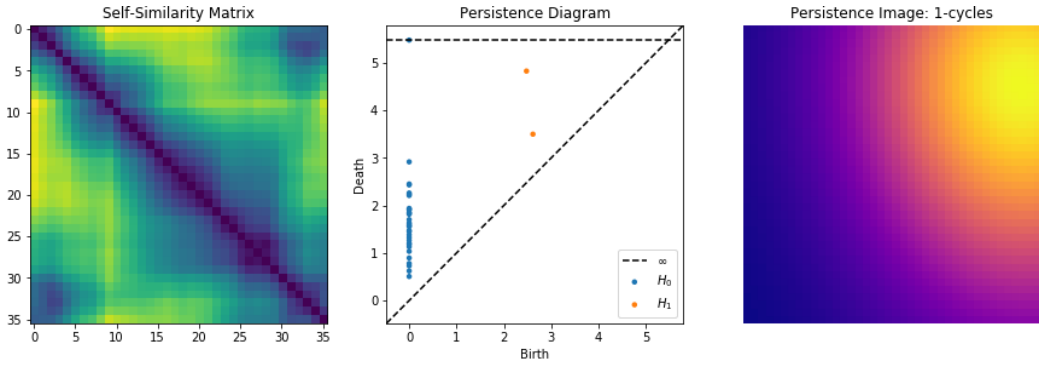

Figure 4.5: Example Persistence Diagram and Persistence Image for SSM.

The H_1 cycle in the center plot of figure 4.3 with the greatest persistence represents 1-cycles in the top right and bottom left of the SSM image (i.e. the dark blue corners) finding a union with the "trough" that is the diagonal on which all values are zero. Figure 4.4 depicts the similarity template, persistence diagram, and H_1 persistence image for the same gesture as figure 4.3.

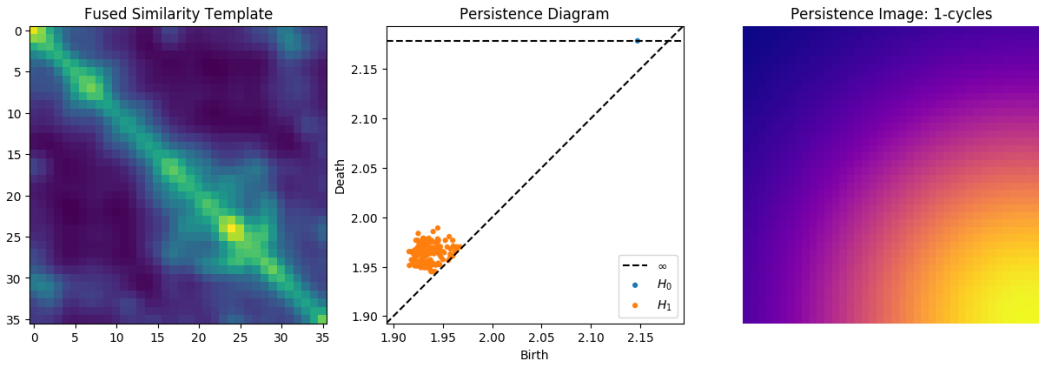

Figure 4.6: Example Persistence Diagram and Persistence Image for a fused similarity template.

There is far more local heterogeneity among values of the fused similarity template. We can see this upon inspection of the left most plot. However, the persistence diagram and persistence image show this more clearly with a large clustering of low-persistence H_1 cycles. In the next section we explore how this difference in topological features affects classification accuracy.

4.5 Image Comparison via CNN

As described in the previous section, it is possible to compare the SSMs and fused similarity templates using a convolutional neural network as they are images. By interpolating the images to be of a uniform size we may predefine a CNN architecture to delineate between gesture classes.

In order to compare the pre-processing techniques applied in this paper, we turn to the AlexNet CNN architecture [KSH12]. AlexNet is known to be an effective architecture

for image classification. It consists of eight layers, five of which are convolutional layers with the remainder being dense layers. A seperate AlexNet CNN was trained on images generated with each pre-processing method described. The classification accuracies of these models are described in Chapter 5.

Chapter 5

Results

Table 5.1 below depicts the classification accuracy scores of a variety of models on three sets of persistence images generated from the original gestures data set: (1) persistence images generated from the SSMs calculated using the cumulative Euclidean distance metric, (2) persistence images generated by applying the ISOMAP algorithm to reduce 4-dimensional vectors to 1-dimension time series vectors then calculating SSMs with the cumulated time series metric on the 1-dimensional vectors - this set was included to provide another point of comparison [TSB+18] and (3) persistence images for the fused similarity graphs for each gesture.

Model	SSM	ISOMAP SSM	Fused Similarity Matrix
KNN	26.74%	21.37%	27.15%
Multiple Logistic Regression	31.27%	26.24%	24.82%
Support Vector Classifier	26.4%	26.07%	24.82%
Random Forest	25.36%	21.54%	27.76%
Multi-layer Perceptron	24.82%	25.0%	25.0%

Table 5.1: Classification accuracy scores for persistence image vectors.

Table 5.2 depicts classification accuracies of AlexNet trained on a stratified 80/20 train-test sample of gestures treated with the different pre-processing techniques discussed here including the ISOMAP SSM. The pre-processed gestures were then used to train and test identical versions of the AlexNet CNN architecture.

Resulting classification accuracy scores are disappointingly low. In fact, greater accuracy can be achieved by simply selecting one of the four modalities, calculating its 1-

Pre-Processing Method	In-Sample Acc.	Validation Acc.
CED SSM	29.35%	22.41%
FTSD SSM	28.48%	36.21%
ISOMAP SSM	95.87%	37.93%
SNF	28.70%	25.86%

Table 5.2: Maximum accuracy scores for AlexNet.

nearest-neighbor and selecting that as its class. Table 5.2 depicts these results under the cumulative Euclidean distance metric described in section 4.2.

Distance	Modality 1	Modality 2	Modality 3	Modality 4
Cumulated Euclidean Distance	36.81%	38.28%	39.99%	41.15%

Table 5.3: Classification accuracy scores for unsupervised distance-based (1-NN) classification.

It is believed that the low classification accuracy scores are the result of a loss of information in the transformations to SSM, fused similarity template, or persistence image. Continued parameter tweaking may alleviate some of these issues. Another possible avenue for improvement may be to reduce SSMs back to time series modalities - this time a 1-D fusion of the originally four dimensional time series - and apply a recurrent neural network.

Chapter 6

Conclusions

This paper introduced the concept of similarity network fusion to the problem of gesture recognition. Intermediate results do seem to provide some heuristic insight into the connectedness and temporal similarities of sEMG recordings for a given gesture. However, process of fusing the modalities and the persistence images generated from fused SSMs seem to mask some temporal information. While fusing sEMG modalities in this way does not seem to simplify the problem of gesture recognition, it may yet be possible to fuse views of a gesture through modifications to the SNF algorithm. Future work may consider fusing many different gestures performed by a single subject and then modeling gestures hierarchically. Additional research in these methods may also consider fusing modalities based on synergist-antagonist pairs, synergist-only groupings, or antagonist-only groupings. Finally, future research may consider directed graphs in the fusion process to better capture the temporal nature of time-ordered data.

Bibliography

[AEK+17] H. Adams, Tegan Emerson, M. Kirby, Rachel Neville, C. Peterson, P. Shipman, Sofya Chepushtanova, Eric Hanson, Francis Motta, and L. Ziegelmeier. Persistence images: A stable vector representation of persistent homology. *ArXiv*, 18:1–35, 02 2017.

[AGC+14] Manfredo Atzori, Arjan Gijsberts, Claudio Castellini, Barbara Caputo, Anne-Gabrielle Mittaz Hager, Simone Elsig, Giorgio Giatsidis, Franco Bassetto, and Henning Müller. Electromyography data for non-invasive naturally-controlled robotic hand prostheses. *Nature*, 1, 12 2014.

[CSEH05] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. In *Discrete and Computational Geometry - DCG*, volume 37, pages 263–271, 01 2005.

[EH10] Herbert Edelsbrunner and John Harer. *Computational Topology.* American Mathematical Society, 2010.

[KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton. Imagenet classification with deep convolutional neural networks. *Neural Information Processing Systems*, 25, 01 2012.

[LKK+18] Sergey Lobov, Nadezhda Krilova, Innokentiy Kastalskiy, Victor Kazantsev, and Valeri Makarov. Latent factors limiting the performance of semg-interfaces. *Sensors*, 18:1122, 04 2018.

[PBW+11] B. Peerdeman, Daphne Boere, Heidi Witteveen, Rianne Veld, Hermie Hermens, Stefano Stramigioli, Johan Rietman, Peter Veltink, and Sarthak Misra. Myoelectric forearm prostheses: State of the art from a user-centered perspective. *Journal of rehabilitation research and development*, 48:719–37, 07 2011.

[PS18] Angkoon Phinyomark and Erik Scheme. Emg pattern recognition in the era of big data and deep learning. *Big Data and Cognitive Computing*, 2:21, 08 2018.

[Rep19] UCI Machine Learning Repository. Emg data for gestures data set. online, 2019.

[TBH19] Christopher Tralie, Paul Bendich, and John Harer. Multi-scale geometric summaries for similarity-based sensor fusion. Conference Paper, 03 2019.

[ten07] *Multi-Dimensional Dynamic Time Warping for Gesture Recognition*, 01 2007.

[Tra17] Christopher Tralie. Self-similarity based time warping. *ArXiv*, 11 2017.

[TSB+18] Christopher Tralie, Abraham Smith, Nathan Borggren, Jay Hineman, Paul Bendich, Peter Zulch, and John Harer. Geometric cross-modal comparison of heterogeneous sensor data. In *IEEE Aerospace Conference*, pages 1–10, 03 2018.

[WJW+12] Bo Wang, Jiayan Jiang, Wei Wang, Zhi-Hua Zhou, and Z. Tu. Unsupervised metric fusion by cross diffusion. In *Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2997–3004, 06 2012.

[WT12] Bo Wang and Z. Tu. Affinity learning via self-diffusion for image segmentation and clustering. In *Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2312–2319, 06 2012.