Methods for Dimensionality Reduction, Compression, and Feature Extraction for Higher-Dimensional Data

Josh Martin
University of Nebraska-Lincoln
Lincoln, USA
joshmartin0212@gmail.com

Reginald Bolman
University of Nebraska-Lincoln
Lincoln, USA
rbolman2@unl.edu

***Abstract*—As numerous domains such as biology, astronomy, geography, etc. produce increasing volumes of data there is a need for sophisticated automated tools to manage this data. While there has been much work done in the fields of clustering, statistical analysis, machine learning etc. real-world data sets are often hindered by the so-called curse of dimensionality. Most common clustering algorithms fail to produce meaningful results on data sets with high dimensionality.**

While various techniques have been studied to address this problem such as "manifold learning", principal component analysis, or many univariate/multivariate feature selection methods such as ANOVA, there exist clear shortcomings with each of these methods. First, not all data can be assumed to be locally smooth, not all data can be assumed to be linear, and not all models can be assumed univariate especially with respect to highly non-linear data sets with multiple exogenous and endogenous variables such as hyperspectral images. Additionally the inherent sparsity of objects within these data sets often leads to results of little to no significance when applying the aforementioned methods.

Due to issues such as these, much research has gone into utilizing wavelet transform image processing techniques for dimensionality reduction, feature detection/extraction problems, and the related problem of data compression. This paper will overview the different advances in utilizing wavelets to address the problems of feature extraction, dimensionality reduction, and compression from the context of a variety of fields within science and mathematics.

I. Introduction

Over the past 40 years, many advances have been made across multiple fields creating a myriad of real world applications for dimensionality reduction, data compression and feature extraction.

As it turns out, the problem of data compression and data clustering is mathematically very similar, to the problems of feature detection and dimensionality reduction and often employ recurring techniques in order to arrive at a solution. For example, the problem of data compression is analogous to the problem of reducing the dimension of a dataspace for a stochastic analysis (i.e. removing unwanted variables reduces the total amount of information present within any given signal) hence there exist certain methods that can be utilized when solving both problems. Within this paper, we overview the methods utilized for dimensionality reduction, compression and feature extraction. We then cover the background of each methodology and introduce any pertinent mathematical concepts. Finally, we overview the current implementations of wavelets to solving the problems of dimensionality reduction, compression and feature extraction.

As humans, our perception of the world is fundamentally biased. Being that our sole experience with the environment around us is from the context of three or less dimensions, it can be extremely difficult, if not impossible, to conceptualize data that exists in higher-dimensions. For this reason, especially when visualizations of higher-dimensional data are needed, a tool for dimensionality reduction is not only useful, but necessary.

Many modern problems in science and engineering deal with situations where data exists in three, four, or even higher dimensions. Weber notes that a sequential scan can outperform various sophisticated techniques used for data retrieval when dealing with multivariate data which implies that dimensionality reduction is not only useful but vital when processing data of ten or more dimensions (Weber). Additionally, it can be shown that many common clustering and feature extraction algorithms fall short of producing meaningful results due to the inherent sparsity of higher dimensional data. Thus, before the analysis of many of these higher dimensional data sets can be meaningful, one must necessarily reduce the dimensionality of the system.

Perhaps the most common approach to the problem of dimensionality reduction is Principal Component Analysis (PCA), a technique commonly used in multivariate statistics. This method creates new uncorrelated variables that attempt to successively maximize variance (Kailing). Finding these new variables, also called the principal components, reduces to solving an eigenvalue/eigenvector problem, and the new variables are defined by the data set at hand (Kailing). However, one major issue with PCA is the assumption of stochastic linearity within the data set. Since not all data can be assumed to be locally smooth and not all data can be assumed to be stochastically linear, this creates a major shortcoming for the PCA technique, potentially limiting its practicality for analyzing multivariate data of ten or more dimensions.

Another method for dimensionality reduction that overcomes the assumption of linearity within the data is the ”manifold learning” technique. Manifold learning is an approach to non-linear dimensionality reduction predicated on the idea that data embedded within higher dimensions can be restrained to a lower dimensional representation through a fitting function (Lawrence Clayton algorithms for manifold learning). Manifold Learning therefore attempts to generalize linear frameworks such as PCA to be sensitive to non-linear structures within data sets (Lunga). Structural analysis and forecasting of macroeconomic variables, portfolio selection and volatility matrix estimations, and studying co-expression of genes within biological systems often require a large number of temporally observed variables with moderate sample sizes, the problem of regularization i.e. the problem of errors created when using a fitting function that results in overfitting one's data set becomes the primary difficulty for the proper implementation of manifold learning techniques.

Just as one might be interested in dimensionality reduction techniques in order to reduce the inherent errors created by sparse higher dimensional data sets by either creating a function to constrain data within a lower dimensional embedding, or by reducing one's data set to data which contains properties which are pertinent, one might instead be interested in the problem of reducing the data that exists within a sample space while maintaining the ability of a user to represent and interpret the data meaningfully utilizes i.e. the problem of “compression”. Much of the study of compression, particularly as it relates to lossy compression is fixated on 2D data. Seismic data is an example of data that typically occupies three or four dimensions. Additionally, seismic data often contains a high degree of anistropy and contains large volumes of noise (Vilasenor). Characteristics such as these tend to cause two-dimensional coding approaches based on wavelets to at best achieve only modest compression ratios.

Vilasenor et al. present a wavelet based approach for compression of geophysical data achieving compression ratios greater than 100:1 with no observable loss. Wavelet based approaches are particularly useful in seismic data, which often contain large regions of data with features that are fairly repetitive and occur across a common spatial scale. Multidimensional data is transformed by repeated application of a 1D wavelet filter bank across all dimensions of the data. Seismic data is then filtered recursively on the low frequency information. This technique while helpful for work on seismic data may not be applicable to other forms of multidimensional data.

Feature extraction is a problem that has wide applicability within many of the scientific and engineering fields. Beyond the standard application for image recognition, much of the work on feature extraction with respect to data of higher-dimensions comes in automatic real-time recognition of segments or abnormalities within signals. Particularly in the medical field, the ability to detect abnormalities in a signal can be the difference between life and death for a patient.

One such problem in medicine where the automated discrimination of a segment within a signal is crucial is electroencephalography. Electroencephalography (EEG) is the measurement of electrical activity in different regions of the brain. Although this method was invented in 1924, it remains to be considered an important clinical tool in neurosurgery, psychiatry, and pediatrics. Primarily EEG is used for seizure detection and the diagnosis of epilepsy. The presence of transients (sharp spikes) in the waveform of the EEG is one of the early signs of a seizure. As the seizure progresses, the transient slowly transforms into a more regular high amplitude quasi-periodic oscillations (Unser). While the shape and size of these waveforms can vary dramatically from patient to patient, the general pattern of oscillation tends to remains fairly consistent. Because of this non-linearity and non-regularity, EEG is another type of data where WT seem to be an appropriate and promising detection tool (Unser). In this context, a real time processing WT algorithm is desirable to be able to detect abnormal transients whilst they occur. For this reason, the Mallat WT is an appropriate choice for its ability to handle real time analysis, and relatively low computational complexity.

Through the use of evoked response potentials (ERP's), the sensitivity of EEG can be dramatically improved. The method of ERP utilizes an external stimuli which could be auditory, visual, or somatosensory, and then calculates the latency from the time of stimulation being applied (Unser). Thakor et al. characterized the shape of partially averaged somatosensory ERP's using global WT and showed that these measures provided a reliable tracking of the time course of neurological injury.

Wavelets have also been applied in an attempt to discriminate between beat-to-beat fluctuations of the heart rate while applying varying physiological stimuli (Unser). Additionally, Senhadji et al. have proven the ability of wavelet-based feature extraction methods to discriminate between normal and abnormal cardiac pattern. WT has similarly been applied for use in detection of late ventricular potentials or VLP's (Unser) - small signals with a frequency of 40 Hz that have been found to be associated with coronary heart disease, myocardial infections, and ventricular arrhythmia. VLP's are most frequently found in the terminal part of the QRS complex and in the beginning of the ST segment, but have been shown to rarely occur in any other portion of the QRS complex (Li Zeng).

Apart from medical data, when dealing with hyperspectral data, the dominant image processing tasks are compression and feature recognition. Compression and feature recognition are fundamentally related. Hyperspectral data contains large volumes of information that needs to be processed depending on the application. The discrete wavelet transform is considered the ideal tool for this type of data structure (Scholl). There are many applications that require processing on these data sets to be done extremely efficiently. Additionally, the higher number of dimensions occupied by the data increases the number of different ways to do these transforms. Scholl et al. present a 3D wavelet transform that allows the resolution in both the spatial domain and spectral domain to be adjusted separately. The algorithm presented by Scholl et al. utilizes the Mallat wavelet to analyze the composition of rock and soil, and attempts to detect for precious metals or oil (Scholl). Additionally, the MODWT wavelet was used on 3D hyperspectral data obtained from NASA for feature recognition. While the techniques tested by Scholl et al. showed much promise in terms of level of compression with minimal loss of observable data, it was concluded that there are applications where the complexity of these compression algorithms may not be necessary (Scholl).

The Daubechies wavelet transform is another proposed method for feature extraction when dealing with multispectral data. The wavelet's inherent multiresolutional properties can be used for multispectral and hyperspectral remote sensing. Various wavelet-based features have been applied to the problem of automatic classification for specific ground vegetation captured in hyperspectral signatures (Bruce). The experimental results of Bruce et al. once again demonstrate the promising discriminant capability of the wavelet-based approaches.

II. Background

As a result of the shortcomings of methods such as manifold learning and principal component analysis with respect to the problems of dimensionality reduction and feature extraction much research has gone into utilizing wavelet transform techniques as a solution. There are essentially two types of wavelet decomposition: the redundant (continuous wavelet transform (CWT) or wavelet frames), and the non-redundant (orthogonal, semi-orthogonal, or biorthogonal wavelet bases) (Unser). It has been shown in the literature that redundant wavelet transforms are usually preferable for signal analyses, feature extraction, and detection tasks because they provide a description that is shift-invariant; Meaning that a shift in the independent variable of the input signal causes a corresponding shift in the output signal (Unser). In contrast, non-redundant wavelet transforms are more suited for when it is desirable to perform some kind of data reduction, or when the orthogonality of the representation is an important factor (Unser).

A. *Principal Component Analysis*

As previously stated, principal component analysis (PCA) is a linear technique for reduction of dimensionality within a dataset. PCA works by creating new uncorrelated variables that attempt to successively maximize variance (Kailing). Once these new variables, also known as the principal components, are found the problem reduces to solving an eigenvalue/eigenvector problem (Kailing).

When looking to utilize PCA we consider a dataset with n entities or individuals (Kailing). These data values define pn-dimensional vectors $x1, \ldots, x_p$ or equivalently, an $n \times p$ matrix X, whose jth column is the vector x_j of observations on the jth variable (Kailing). Next, the goal of PCA is to seek a linear combination of the columns of maxtrix X with the maximum variance (Lunga). These linear combinations can be found using $\sum_{j=1}^{p} a_j x_j = Xa$, where a is a vector of the constants $a_1, a_2, \ldots, a_p$ (Kailing). The variance of any of the derived linear combinations can be calculated by $var(Xa) = a'Sa$ (Kailing). It must be noted as well that for this problem to have a well-defined solution, an additional restriction must be imposed, the most common of which involving working with unit-norm vectors (Lunga). Once the variance for each linear combination has been found, it is a simple problem of finding the maximum value and selecting said vector.

B. *Manifold Learning*

Manifold learning is an non-linear approach to the problem of dimensionality reduction. The inherent difficulty with data sets of higher-dimensions is that they can be very difficult to visualize. While there are various motivations for dimensionality reduction, it is necessary for the ability to visualize a high-dimensionality dataset in a way that humans can intuit.

There are several implementations of manifold learning. While each strategy utilizes different fundamental approaches and algorithms, each is used for the same goal, a non-linear dimensionality reduction. The various implementations are optimal for use on different types of data and have differing time complexities. Below are a few implementations of manifold learning with their respective algorithms and time complexities.

*a) **Isomap**:* The Isomap technique is one the earliest known approaches to manifold learning, and is short for Isometric Mapping (Li). The Isomap can be seen as an extension of Multi-dimensional Scaling (MDS) also known as Kernel PCA (Lunga). Isomap looks for a lower-dimensional embedding which maintains geodesic distances between all points.

The algorithm for Isomap is comprised of three steps:

1) **Nearest Neighbor Search:** While any NN search can be used, the cost is approximately $O[D \log(k) N \log(N)]$, for k nearest neighbors of N points in D dimensions (Li).
2) **Shortest-Path Graph Search:** The most efficient known algorithms for shortest path search are Dijkstra's Algorithm, which is approximately $O[N^2(k + \log(N))]$, or the Floyd-Warshall algorithm, which is $O[N^3]$ (Lunga).
3) **Partial Eigenvalue Decomposition:** The embedding is encoded in the eigenvectors corresponding to the d largest eigenvalues of the $N \times N$ isomap kernel. For a dense solver, the cost is approximately $O[dN^2]$ (Li).

Therefore the overall complexity for the Isomap implementation is: $O[D \log(k) N \log(N)] + O[N^2(k + \log(N))] + O[dN^2]$ Where:

- **N**: Number of training data points.
- **D**: Input Dimension.
- **k**: Number of Nearest Neighbors.
- **d**: Output Dimension.

*b) **Locally Linear Embedding**:* Locally linear embedding (LLE) looks for a lower-dimensional projection of the data that preserves distance within local neighborhoods. This can be though of as a series of local Principal Component Analyses which are globally compared to find the best non-linear embedding (Roweis).

The standard implementation for LLE has three steps:

1) **Nearest Neighbors Search:** As discussed above in the Isomap section, the complexity of NN search is $O[D \log(k) N \log(N)]$, for k nearest neighbors of N points in D dimensions (Li)
2) **Weight Matrix Construction:** The construction of the LLE weight matrix involves the solution of a $k \times k$ linear equation for each of the local neighborhoods. This results in a complexity of $O[DNk^3]$ (Roweis).
3) **Partial Eigenvalue Decomposition:** As discussed above in the Isomap section, the complexity of this decomposition is: $O[dN^2]$ (Li).

Therefore the total complexity for LLE is: $O[D \log(k) N \log(N)] + O[DNk^3] + O[dN^2]$
Where:

- **N**: Number of training data points.
- **D**: Input Dimension.
- **k**: Number of Nearest Neighbors.
- **d**: Output Dimension.

*c) **Spectral Embedding**:* Spectral Embedding is a manifold learning implementation that is predicated on the concept of non-linear embedding. Scikit-learn implements Laplacian Eigenmaps, which finds a low dimensional representation of the data using a spectral decomposition of the graph Laplacian (Belkin). The resulting graph generated as a result of this spectral composition can be considered to be a discrete approximation of the low dimensional manifold in the high dimensional space (Belkin). By minimizing the cost function based on the derived graph, it is ensured that points that are located close to each other on the manifold are mapped close to each other in the low dimensional space (Lunga). This preserves the local distances within the data.

The Spectral Embedding algorithm is comprised of three steps:

1) **Weighted Graph Construction:** The first step is to transform the raw input data into a graph representation. This is done using an adjacency matrix representation. As stated in the LLE section, the complexity of weighted graph construction is: $O[DNk^3]$ (Roweis).
2) **Graph Laplacian Construction:** The unnormalized Laplacian graph is constructed as $L = D - A$ and the normalized graph is constructed as $L = D^{-\frac{1}{2}}(D - A)D^{-\frac{1}{2}}$
3) **Partial Eigenvalue Decomposition:**The eigenvalue decomposition is done on the generated weighted graph from the first step. As in the LLE, the complexity of this step is: $O[dN^2]$

Therefore the total complexity for Spectral Embedding is: $O[D \log(k) N \log(N)] + O[DNk^3] + O[dN^2]$
Where:

- **N**: Number of training data points.
- **D**: Input Dimension.
- **k**: Number of Nearest Neighbors.
- **d**: Output Dimension.

*C. **4-D Zerotree Extension (The Hybrid Tree)***

The 4-D Zerotree extension also known as the Hybrid Tree is a file structure proposed by Chakrabarti et al. to index high-dimensionality data (Chakrabarti). Chakrabarti's algorithm allows one to perform dimensionality reduction dynamically wherein a sampling of the dimensions are initially reduced by utilizing a hybrid search tree that can contract and extend feature vectors, the algorithm then adds new dimensions only when the additional discriminatory power is absolutely necessary. The dimensionality of the data set is considered the feature space, and each variable is represented as a vector within the feature space. The goal is to use as few features as necessary to discriminate among the objects. This technique aligns with the intuitive way that humans classify objects. For example, in zoology, the species are grouped in a few broad classes first, using a few features (e.g., vertebrates versus invertebrates). As the classification is further refined, more and more features are gradually used (e.g., warm-blooded versus cold-blooded, or lungs versus gills) (Chakrabarti).

The Hybrid Tree organizes data into a hierarchical structure similar to any other tree. Next, feature vectors are clustered into leaf nodes of the tree, and the description of their Minimum Bounding Region (MBR) is stored in the parent node (Chakrabarti). Parent nodes are recursively grouped as well until the root is formed. When comparing to a tree that uses a fixed number of features, the Hybrid Tree provides a higher fanout at the top levels, by using only the feature vectors necessary and eliminating the irrelevant ones (Berchtold). As more objects are inserted into the tree, more features might be needed to discriminate among the objects. At that time, new features are introduced. The addition of features on a "when needed" basis allows the Hybrid Tree to beat the "curse of dimensionality".

*D. **Daubechies Wavelets***

The Daubechies wavelets are a family of orthogonal wavelets defining a discrete wavelet transform and characterized by a maximal number of vanishing moments for some given support. For each wavelet type in Daubechies family of wavelets, there exists a father wavelet scaling function which can be used to generate an orthogonal multiresolutional analysis (Hazewinkel).

The Daubechies wavelets are chosen in an attempt to maximize the number A of vanishing moments for the given support width 2A - 1. Two naming schemes are currently used when referring to Daubechies wavelets. D*N* when using the length or number of taps, and db*A* when referring to the number of vanishing moments (Hazewinkel). Therefore, D4 and db2 would be the same wavelet transformation simply referenced by different parameter.

Within the 2^{A-1} potential solutions of the algebraic equations for the moment and orthogonality conditions, the one whose scaling filter has extremal phase is chosen. Additionally, Daubechies wavelet transforms are easy to put into practice using the fast wavelet transform, a mathematical algorithm designed to turn a waveform or signal in the time domain into a sequence of coefficients based on an orthogonal basis of small finite waves, or wavelets (Hazewinkel). Daubechies wavelets are utilized in solving numerous different problems, some of the most common among them being: self-similarity search, feature recognition, and detection of signal discontinuities.

The scaling function for Daubechies wavelet is defined as follows:

$\phi(x) = \sqrt{2}\sum_{k=0}^{N-1} c_k \phi(2x-k)$

With the property that:

$\int \phi(x)dx = 1$

Additionally, the mother wavelet can be defined as:

$\psi(x) = \sqrt{2}\sum_{k=0}^{N-1} (-1)^k c_{N-1-K} \phi(2x-k)$

E. HAAR Wavelet

The Haar wavelet is a sequence of rescaled square-shaped functions which together form a wavelet family or basis. This transform is similar in many ways to a Fourier analysis and allows a target function to be represented in terms of an orthonormal basis over a given interval. While there are various applications for the Haar wavelet, the primary use in practice is reduction of dimensionality. The Haar wavelet is widely considered as the first know wavelet basis and was proposed by Alfréd Haar in 1909 (Haar). It has become one of the most commonly and extensively used wavelets for educational purposes. The Haar wavelet is actually a special case of Daubechies wavelet and is sometimes known as Db1 (Haar).

There are however some notable disadvantages to the Haar wavelet. One such technical disadvantage of the Haar wavelet is that it is not continuous, and therefore not differentiable (Haar). However, this property can be seen as an advantage when dealing with signals that have sudden transitions, such as monitoring for tool failure in machines.

The Haar wavelet's mother wavelet function $\psi(t)$ can be described as:

$$\psi(t) = \begin{cases} 1 & 0 \leq t < \frac{1}{2}, \\ -1 & \frac{1}{2} \leq t < 1, \\ 0 & otherwise \end{cases}$$

The Haar wavelet's scaling function $\varphi(t)$ can be described as:

$$\varphi(t) = \begin{cases} 1 & 0 \leq t < 1, \\ 0 & otherwise \end{cases}$$

There are several properties to note for the Haar wavelet.

- Any continuous real function can be approximated uniformly by linear combinations of $\varphi(t), \varphi(2t), \varphi(4t), \ldots, \varphi(2^n t), \ldots$ and their shifted functions. This also can be extended to any function spaces where the function can be approximated by continuous functions.
- Any continuous real function on [0, 1] can be approximated uniformly on [0, 1] by linear combinations of the constant function $1, \psi(t), \psi(2t), \psi(4t), \ldots, \psi(2^n t), \ldots$ and their shifted functions.
- Orthogonality presents in the form:

$$\int_{-\infty}^{\infty} 2^{(n+n_1)/2}\psi(2^n t-k)\psi(2^{n_1}t-k_1)dt = \delta_{n,n_1}\delta_{k,k_1}$$

Where $\delta_{ij} = \begin{cases} 0 & if i \neq j \\ 1 & if i = j \end{cases}$

F. MODWT

The maximal overlap discrete wavelet transform, or MODWT, is a transform very similar to the generic discrete wavelet transform also known as DWT. The transform is denoted using the MODWT coefficients $d_{j,k}$ (Lark). When examining a sequence of MODWT coefficients the locations k progress in unit steps rather than steps of $2j$ as in the DWT transform (Lark). Unlike the DWT however, MODWT is not orthonormal. This is because the MODWT coefficients are obtained without the subsampling step present in DWT the new coefficients at any given scale are no longer orthogonal to each other (Dghais). The wavelet can be used to perform a decomposition of an input vector, and the inverse can be performed to reconstruct the original vector. Perhaps the most crucial difference between DWT and MODWT is that MODWT is defined naturally for all samples sizes (Dghais). Therefore MODWT can handle any input size n whereas the standard DWT cannot. Furthermore, MODWT is transform invariant, meaning a shift in the signal will not change the pattern of the wavelet transform coefficients (Dghais).

The generic DWT transform, as well as MODWT can be defined as follows:

$\psi(t) = \sum_n h_1(n)\sqrt{2}\varphi(2t-n), n \in Z$

The mother wavelet function psi for DWT can be defined as:

$\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}}\psi(\frac{t-k2^j}{2^j})$

The wavelet coefficients c_j and d_j for the generic DWT can be calculated using:

$c_j = \langle g(t), \varphi_{j,k}(t)\rangle = \int g(t)\varphi_{j,k}(t)dt$

$d_j = \langle g(t), \psi_{j,k}(t)\rangle = \int g(t)\psi_{j,k}(t)dt$

In order to determine the coefficients for MODWT, one must simply take the coefficients found using the formula for the generic DWT and divide them by $2^{\frac{j}{2}}$

G. Mallat Wavelet

The Mallat wavelet transform, also known as the "Fast Wavelet Transform" was first introduced in 1989 by Stéphane Mallat (Mallat). The algorithm was designed to turn a waveform or signal in the time domain into a sequence of coefficients based on an orthogonal basis of small finite waves, or wavelets (Mallat). Mallat's wavelet transform can easily be extended to higher-dimensional signals, such as hyperspectral images, by replacing the time domain with the space domain (Hubbard). This fact makes Mallats wavelet particularly useful for tasks involving processing of data or higher-dimensions.

Mallat's transform is predicated on the idea of a finitely generated orthogonal multiresolution analysis, or MRA (Mallat). Using the terms dictated by MRA, one selects a sampling scale J with a sampling rate of 2^J per unit interval, and then proceeds to project the signal f onto the space V_J by computing the scalar products: $s_n^J := 2^J \langle f(t), \phi(2^J t - n) \rangle$, where ϕ is the scaling function for the selected wavelet transform (Hubbard). The MRA is categorized by the scaling sequence $a = (a_{-N}, \ldots, a_0, \ldots, a_N)$, and the wavelet sequence $b = (b_{-N}, \ldots, b_0, \ldots, b_N)$ (Mallat). These sequences allow one to compute the wavelet coefficients d_n^k, without needing to calculate the integrals in the corresponding scalar products (Hubbard).

III. Overview

A. *Dimensionality Reduction*

As previously discussed, one such prominent tool for the reduction of dimensions, is Chakrabarti's hybrid tree algorithm. Chakrabarti et al. performed extensive experimentation with two primary goals in mind. First, to evaluate the various design decisions made in the creation of the hybrid tree; second, to compare the hybrid tree with other competitive techniques (Chakrabarti). The experiments conducted using Chakrabarti's hybrid tree was performed over the following two real world datasets:

1) **The FOURIER dataset:** This dataset contains approximately 1.2 million 16D vectors produced by fourier transformations of polygons
2) **The COLHIST dataset:** This dataset contains color histograms derived from approximately 70,000 colored imaged obtained from the Corel Database. Chakrabarti et al. generated 16, 32, and 64 dimensional vectors by extracting $4 \times 4, 8 \times 4, and 8 \times 8$ color histograms from the images in the database.

Queries were then randomly distributed within the data space with chosen ranges aimed at achieving close to constant selectivity (Chakrabarti). The selectivity in Chakrabarti et al. experiments was maintained at a constant 0.07% for the FOURIER dataset and 0.2% for the COLHIST dataset (Chakrabarti).

The specific features of the algorithm that were of interest during testing were (1) the impact of the proprietary node splitting algorithms and (2) the effect of live space optimization in the hybrid tree (Chakrabarti). The performance of the algorithm was measured in two ways. First, by measuring the average number of disk accesses required to execute any given query; second by calculating the average CPU time required to execute the given query. Chakrabarti et al. found that the hybrid tree consistently outperformed more standard tree structures when dealing with the increase in dimensionality (Chakrabarti). As the dimensionality increased, Chakrabarti's algorithm performed linearly better in both data accesses and CPU time (Chakrabarti). While a linear improvement may seem trivial, when dealing with datasets with hundreds of thousands, if not millions of points in 64 or more dimensions, this improvement could save large amounts of CPU time and significantly decrease the number of data accesses needed to execute a query.

B. *Compression*

Various industries, the petroleum industry among them, use seismic reflections as one of the primary tools to analyze geological data. Many of the modern seismic data stores contain terabytes upon terabytes of data (Vilasenor). Compression becomes a crucial technique in order to store these datasets. Given seismic data exists in higher-dimensions, standardized compression algorithms such as JPEG and MPEG are not particularly effective. The inherent block-based nature of the approach of JPEG and MPEG leads to block artifacts in the reconstruction (Vilasenor). Additionally JPEG is a two-dimensional standard and does not provide extensions for three or four-dimensional data. For these reasons, wavelet based approaches are significantly better suited for use with seismic data.

Vilasenor et al. applied the MODWT wavelet transform in a field trial on geophysical data captured in the North Sea (Vilasenor). Data was collected at a rate of approximately 10 Mbits per second, and the data was subsampled by a factor of 4 using the conventional industry standard techniques at the time of the trial. The data was transmitted via satellite on a 128 Kbit/sec link to a processing center (Vilasenor). Over the course of the trial nearly half a terabyte of seismic data was collected.

Using the MODWT wavelet the collected seismic data was able to be compressed at ratios exceeding 100:1 (Vilasenor). Vilasenor et al. state that their method of compression for many datasets could very well exceed this level by a considerable margin; however, 100:1 represents a threshold at which data can be predictably be compressed to regardless of the sparsity and structure of the data (Vilasenor). Additionally, with the compression algorithm used, 3D seismic data was able to be transmitted in real-time from the field to the processing center. Currently operational workstations no not have capability for real-time transmission without requiring highly specialized hardware components (Vilasenor). Vilasenor et al. also found that their algorithm lead to other significant benefits such as the ability to increase the online volume of data, the reduction or archival storage volume, and an increase of the speed of network transmission (Vilasenor).

C. *Feature Extraction*

Many experiments have performed in order to assess the consistency and accuracy of the various algorithms used for feature recognition. In practice, many instances of feature recognition tasks require extremely high accuracy whether that be due to the safe-critical nature of the task, or the monetary risk associated.

Wavelet Transforms (WT) have been proven to be an extremely useful tool for the time-frequency analysis and characterization of the primary heart sounds captured. It has even been shown that WT can pick up sound components that could not be detected by the other methods such as ultrasound (Unser). The information can be extracted from the signal by decomposing the signal on a wavelet orthonormal basis of $L^2(R)$(Akay). $L^2(R)$ denotes the vector space of measurable, square-integrable one-dimensional functions f(x). Then the projection of the signal f(x) onto the orthonormal bases of $L^2(R)$ can be estimated using a combination of Daubechies Wavelet Transform and the Mallat Wavelet (Akay). Then by applying a low-pass filter to the result of this transform in order to remove the high frequency content we can produce an orthogonal projection of a signal f(x) on the vector space which will represent the discrete detailed signal, effectively classifying the signal as standard heart sounds or murmur.

When dealing with ECG signals, perhaps the most critical step in the analysis is accurate detection of the QRS complex. While there exist many algorithms to achieve this task, wavelet based approaches provide excellent performance with studies finding a detection rate of 99.8% (Unser). Furthermore, Li Zeng et al. have had great levels of success implementing their 4-D extension of the well known zerotree algorithm on ECG data. They found that not only were they able to achieve detection rates of over 99.6%, but also managed to compress the 4-D ECG data at rates of 2000:1.

Beyond medical applications, feature extraction is commonly used for the discrimination of objects within hyperspectral data. Bruce et al. used the DWT and MODWT wavelet transforms for hyperspectral feature extraction. The automated classification used by Bruce et al. consistently provided over 95% and 80% classification accuracy for end-member and mixed-signature applications, respectively (Bruce). When compared to conventional feature extraction methods, the wavelet transform approach is shown to significantly increase the overall classification accuracy.

IV. Conclusions

There has been an explosion in not only the sheer volume of data that is represented digitally, but also the number of dimensions that this data occupies for many modern engineering and science problems. Therefore, it will become increasingly important that we develop efficient algorithms for the problems of feature extraction, compression, and dimensionality reduction in order to manage this data and produce reliable predictive models. The breadth of the applicability for techniques for managing data of higher dimensions is enormous. From medical applications that can save lives through detection of abnormalities within signals, to geophysical applications that can explore vast regions of land and detect precious natural resources, the possibilities seem near infinite.

While much work has gone into techniques to manage data in a general sense, such as the discussed methods of manifold learning and principal component analysis, real world datasets which are seen in modern engineering and science are rarely locally smooth, inherently sparse, and often violate the statistical assumptions of normality, linearity, etc rendering techniques such as PCA or manifold learning questionable for performing an analysis of data of higher-dimensions. Further, even if accurate results are generated with non-parametric classification and feature extraction algorithms such as DAFE or DBFE, the underlying models fitness to the data cannot be tested.

Herein lies the opportunity for the applicability of wavelet based techniques. The literature shows wavelet based methods show great promise in the related problems of feature extraction, dimensionality reduction, and compression (Akay). With respect to dimensionality reduction, when compared to PCA, wavelet-based techniques for dimensionality reduction showed a significant increase in not only the time taken to complete the reduction, but also in the amount of data lost (Bruce). Similarly for feature extraction problems, wavelet transformations showed not only unparalleled accuracy, but also allowed for real-time processing in situations where it was never before possible (Bruce). With respect to the problem of compression, wavelet based approaches utilizing wavelets such as MODWT or HAAR have achieved results of at worst 100:1 and in some cases even exceeding compression ratios of 2000:1. These are results that could have never been attainable without the use of wavelet signal processing techniques.

Upon aggregating and analyzing the results of studies from a variety of fields focused on the processing of multi-dimensional data, one clear commonality emerges; wavelet based methods consistently out perform many of the industry standard practices for the problems of dimensionality reduction, feature recognition, and compression. While the study of wavelets is only about a hundred years old, their promise and capabilities cannot be understated. Therefore, we conclude that the variety of applications for wavelets and algorithms derived from wavelets warrants additional research and further experimentation.

V. References

- Akay, Metin, et al. "Time-Frequency Analysis of the Electrocortical Activity during Maturation Using Wavelet Transform." Biological Cybernetics, vol. 71, no. 2, 1994, pp. 169–176., doi:10.1007/s004220050079.
- Belkin, Mikhail, and Partha Niyogi. "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation." Neural Computation, vol. 15, no. 6, 2003, pp. 1373–1396., doi:10.1162/089976603321780317.
- Berchtold, S., et al. "Independent Quantization: an Index Compression Technique for High-Dimensional Data Spaces." Proceedings of 16th International Conference on Data Engineering (Cat. No.00CB37073), doi:10.1109/icde.2000.839456.
- Bruce, L.m., et al. "Dimensionality Reduction of Hyperspectral Data Using Discrete Wavelet Transform Feature Extraction." IEEE Transactions on Geoscience and Remote Sensing, vol. 40, no. 10, 2002, pp. 2331–2338., doi:10.1109/tgrs.2002.804721.
- Chakrabarti, K., and S. Mehrotra. "The Hybrid Tree: an Index Structure for High Dimensional Feature Spaces." Proceedings 15th International Conference on Data Engineering (Cat. No.99CB36337), 1999, doi:10.1109/icde.1999.754960.
- Dghais, Amel Abdoullah Ahmed, and Mohd Tahir Ismail . "A Comparative Study between Discrete Wavelet Transform and Maximal Overlap Discrete Wavelet Trans-

form for Testing Stationarity." International Journal of Mathematical Science and Engineering , vol. 7, no. 12, 2013.

- Haar, Alfréd (1910), "Zur Theorie der orthogonalen Funktionensysteme", Mathematische Annalen, 69 (3): 331–371, doi:10.1007/BF01456326, hdl:2027/uc1.b2619563
- Han, Te Sun, and Shun-ichi Amari. "Statistical Inference Under Multiterminal Data Compression." Information Theory, 2009, doi:10.1109/9780470544907.ch11.
- Hazewinkel, Michiel, ed. (2001) [1994], "Daubechies wavelets", Encyclopedia of Mathematics, Springer Science+Business Media B.V. / Kluwer Academic Publishers, ISBN 978-1-55608-010-4
- Hubbard The World According to Wavelets: The Story of a Mathematical Technique in the Making (1998 Peters) p. 184
- Kaewpijit, Sinthop, et al. "Spectral Data Reduction via Wavelet Decomposition." Wavelet and Independent Component Analysis Applications IX, 2002, doi:10.1117/12.458728.
- Kailing, Karin, et al. "Density-Connected Subspace Clustering for High-Dimensional Data." Proceedings of the 2004 SIAM International Conference on Data Mining, 2004, doi:10.1137/1.9781611972740.23.
tem Lark, R. M., and R. Webster. "Changes in Variance and Correlation of Soil Properties with Scale and Location: Analysis Using an Adapted Maximal Overlap Discrete Wavelet Transform." European Journal of Soil Science, vol. 52, no. 4, 2001, pp. 547–562., doi:10.1046/j.1365-2389.2001.00420.x.
- Li, M. M. Crawford, and J. Tian, "Local manifold learningbased k-nearest-neighbor for hyperspectral image classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 48, no. 11, pp. 4099–4109, 2010.
- Lunga, Dalton, et al. "Manifold-Learning-Based Feature Extraction for Classification of Hyperspectral Data: A Review of Advances in Manifold Learning." IEEE Signal Processing Magazine, vol. 31, no. 1, 2014, pp. 55–66., doi:10.1109/msp.2013.2279894.
- Mallat "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation" IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 2, no. 7. July 1989.
- Roweis, S. T. "Nonlinear Dimensionality Reduction by Locally Linear Embedding." Science, vol. 290, no. 5500, 2000, pp. 2323–2326., doi:10.1126/science.290.5500.2323.
- Sanchez, Jorge, and Florent Perronnin. "High-Dimensional Signature Compression for Large-Scale Image Classification." Cvpr 2011, 2011, doi:10.1109/cvpr.2011.5995504.
- Senhadji, G. Carrault, J. 5. Bellanger, and G. Passariello, "Comparing wavelet transforms for recognizing cardiac patterns," IEEE Eng. in Med. and Biol. Mag., vol. 14, no. 2, pp. 167-173, 1995.
- Scholl, James F., and Eustace L. Dereniak. "Higher-Dimensional Wavelet Transforms for Hyperspectral Data Compression and Feature Recognition." Mathematics of Data/Image Coding, Compression, and Encryption VI, with Applications, 2004, doi:10.1117/12.508086.
- Thakor, X.-R. Guo, Y.-C. Sun, and D. F. Hanky, "Multiresolution wavelet analysis of evoked potentials," IEEE Trans. Biomed. Eng., vol. 40, pp. 1085-1094, Nov. 1993.
- Unser, M., and A. Aldroubi. "A Review of Wavelets in Biomedical Applications." Proceedings of the IEEE, vol. 84, no. 4, 1996, pp. 626–638., doi:10.1109/5.488704.
- Villasenor, J.d., et al. "Seismic Data Compression Using High-Dimensional Wavelet Transforms." Proceedings of Data Compression Conference - DCC '96, 1996, doi:10.1109/dcc.1996.488345.
- Weber, Roger, et al. "A Quantitative Analysis and Performance Study for Similarity-Search Methods in High-Dimensional Spaces." Institute of Information Systems ETH Zentrum, 1998.
- Zeng, Li, et al. "Four-Dimensional Wavelet Compression of Arbitrarily Sized Echocardiographic Data." IEEE Transactions on Medical Imaging, vol. 21, no. 9, 2002, pp. 1179–1187., doi:10.1109/tmi.2002.804424.