2012 American Control Conference
Fairmont Queen Elizabeth, Montréal, Canada
June 27-June 29, 2012

Optimally Robust Extrema Filters for Time Series Data

Pramod K.Vemulapalli, *Student Member, IEEE,*
Vishal Monga, *Member, IEEE,* and Sean N.Brennan, *Member, IEEE*

***Abstract*— Feature vectors encoded by using extrema are known to be immune to different types of distortions of the original time series [1]. This property enables them to be effective in a wide range of pattern matching applications for time series data [2] [3]. The process of extracting extrema is usually preceded by a filtering step to reduce noise and to bring out prominent features in a time series. The core contribution of this paper is a methodology based on eigenanalysis to optimize the filter that would lead to robust extrema being extracted from the filtered signal. In this context, robustness is understood as the ability of the extrema from a signal to remain intact in spite of distortions to the signal. The paper then demonstrates that the 'optimally robust' extrema outperform extrema obtained from using traditional filters in a time series pattern matching (subsequence matching) task on real and simulated datasets in the presence of bias, scale factor, and outlier distortions in the query signal.**

I. INTRODUCTION

The task of extracting feature vectors from time series data is of fundamental importance in accomplishing a wide range of pattern recognition tasks. A wide array of solutions have been proposed to solve this problem, and one could classify these approaches into three categories, namely: dimensionality reduction methods, distance metric methods, and interest point methods. A brief description of each approach is provided below.

1) **Dimensionality reduction methods:** In these methods, given a particular time series, a window of a certain length is chosen and the window is slid across the time series to extract all possible subsequences [4]. This initial step is often referred to as the sliding window method. A dimension reduction technique is then applied to each subsequence to obtain a feature vector to describe it. Different types of dimension reduction techniques have been proposed in literature. These have included extracting coefficients from Discrete Fourier Transform [4] , Discrete Wavelet Transform [5], Discrete Cosine Transform [4], and Singular Value Decomposition [6] of the subsequence. Methods have also been developed to represent a subsequence using piecewise constant values [7], piecewise linear functions [8], chebyshev polynomials [9], and via symbols [10].
2) **Distance metric methods:** The distance metric based methods completely circumvent the need for feature-based encoding of time series by taking the windowed raw signal as the feature vector and by introducing different methods to compute the distance between the raw signal and the query signal. It is important to note that these methods are commonly used within the sliding window framework. While euclidean distance is the most straightforward method of comparison, this metric has been found to be brittle [11]. Dynamic time warping methods (DTW) have been proposed to mitigate the brittleness associated with the euclidean distance measure by allowing for temporal distortion [11]. Euclidean and DTW based methods are not robust to outliers, and this lead to "edit distance" methods. The concept of edit distance was borrowed from matching strings, and these methods enable matching by ignoring the dissimilar parts of the given time series [12] [13] [14].
3) **Interest point methods:** The fundamental difference between the interest point method of generating features when compared to the above methods is that it is not necessary to use the sliding window method in this case. Previous work has shown that interest point based features are immune to many types of distortions [1]. Thus, they have been utilized in commercial music identification applications [15]. The computational and memory benefits offered by interest point based encoding have been clearly demonstrated [3] in a vehicle localization application [16] amongst others. Relevant to this work are extrema points, a particular category of interest points which are easy to identify. The use of features developed from the extrema points of a signal has been presented in literature [1]. The terms 'Extrema points' or 'Extrema' refer to the maxima and minima that occur in a signal and are used synonymously throughout this paper. These prominent artifacts are utilized in a number of interest point based feature vector generation schemes [1] [3] [15].

Of the above methods, this work focuses on extrema based methods as they possess certain inherent capabilities that make them desirable in a number of different pattern recognition / feature encoding tasks. Some of these properties are listed below.

1) As the number of extrema in a signal loosely cor-

Pramod.K.Vemulapalli is a graduate student in the Department of Mechanical Engineering, The Pennsylvania State University, State College, PA 16802 USA (phone: 814-321-2800; e-mail: pkv106@psu.edu).

Vishal Monga is an assistant professor in the Department of Electrical Engineering, Pennsylvania State University, State College, PA 16802 USA. (e-mail: vmonga@engr.psu.edu).

Sean.N.Brennan is an associate professor in the Department of Mechanical Engineering, The Pennsylvania State University, State College, PA 16802 USA (phone: 814-863-2430; e-mail: sbrennan@psu.edu).

responds to the amount of variation present in the signal, this method of encoding will automatically lead to more features at locations corresponding to high variation and vice versa.

2) The number of extrema in a signal or data usually constitute a small portion of the amount of information in the signal. Thus, feature encoding with the aid of extrema, has an inherent lossy data compression associated with it.
3) The extrema themselves tend to be robust to a large number of distortions, where robustness is the ability to survive intact in spite of distortions being introduced into the signal. Perng *et al* in [1] have shown that extrema can survive certain severe distortions that are expected to drastically effect the feature vectors obtained from other methods.

Given the inherent advantages that extrema methods provide and the broad scope of their application, this paper proposes an optimization technique to extract robust extrema from time series. The capabilities of features generated from these extrema are then demonstrated through a subsequence matching problem. The remainder of this paper is organized in the following manner. Section 2 details a generic outline that is used to create feature vectors from raw data while using extrema. Section 3 details an optimization based on eigenanalysis that maximizes the robustness of the extrema by identifying the filter that would lead to optimally robust extrema. In the next section, an application for the proposed framework is demonstrated where pattern recognition using robust extrema is compared with extrema obtained from using traditional filters on both real and simulated data. Conclusions and directions for future work are discussed in section 6.

II. BACKGROUND

Extrema based techniques create feature vectors by encoding the amplitude, relative locations or other properties of the extrema. A block diagram illustrating the steps involved in a generic process that seeks to encode features from extrema is shown in Fig. 1. The first step in the process is filtering the signal to reduce noise and/or to enhance significant aspects of the time series data. The second step involves extracting the extrema from the filtered signal. The process of extracting the extrema itself could be based on using simple thresholds or by considering additional properties of an extrema that try to ascertain whether each extrema is "significant" or not. The properties of these extrema are encoded into feature vectors in the last step.

Given a particular pattern recognition method involving extrema, one could optimize each one of the above steps (filtering, extrema detection, and encoding) to enhance the result of the method. The optimization itself can be based on the data or the type of distortions involved in the pattern recognition problem. The next section presents an optimization method for the filtering step.

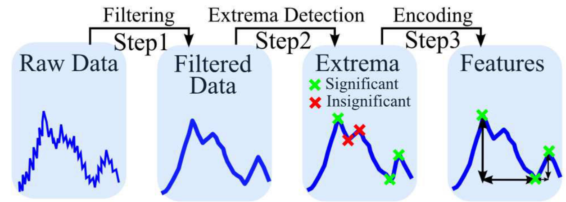

Fig. 1. The filters used in the experimental tests.

III. OPTIMALLY ROBUST EXTREMA

A. Derivation for the Optimally Robust Extrema Filter (OREF)

The most desirable extrema are those that remain identifiable and unaffected when distortions are introduced into a signal and can be reffered to as robust extrema. A filter that results in the most robust extrema from being extracted from the filtered signal can be defined as the "Optimally Robust Extrema filter (OREF)" in this context. In order to find the OREF that maximizes the robustness of the extrema, it is necessary to geometrically visualize the process that occurs when selecting an extrema. The following derivation demonstrates that the extrema selection process is equivalent to a geometric problem of selecting data points in a hyperspace. It also shows that the filtering operation can be interpreted as bounding the selected extrema by utilizing two hyperplanes. The derivation is as follows:

For a given discrete signal $x[n]$ and an acausal finite impulse response filter $h[n]$ with $2N+1$ coefficients, the corresponding filtered signal is denoted by

$$y[n] = \sum_{i=-N}^{N} b_i x[n-i] \tag{1}$$

where b_i are the filter taps for the filter $h[n]$,

$$As\ h[n]\ has\ 2N+1\ taps,\ b_i = 0,\ i \notin [-N, N] \tag{2}$$

It is desirable to have an odd number of taps in the filter so that the filter has an identical number of taps on either side of a particular point. This selection is advisable for most signals unless there are any specific reasons to choose a filter with different number of taps on either side of a point.
If $y[n_0]$ is a maxima of the filtered signal then by definition it must satisfy the following properties:

$$y[n_0] > y[n_0+1]\ and\ y[n_0] > y[n_0-1] \tag{3}$$

Considering the first condition

$$y[n_0] > y[n_0+1] \tag{4}$$

Substituting (1) into the above condition,

$$\sum_{i=-N}^{N} b_i x[n_0-i] > \sum_{i=-N}^{N} b_i x[n_0+1-i] \tag{5}$$

$$\iff \sum_{i=-N}^{N} b_i x[n_0-i] > \sum_{i=-N-1}^{N-1} b_{i+1} x[n_0-i] \tag{6}$$

$$\iff \sum_{i=-N-1}^{N+1} b_i x[n_0 - i] > \sum_{i=-N-1}^{N-1} b_{i+1} x[n_0 - i] \quad (7)$$
$$as\ b_{-N-1} = b_{N+1} = b_{N+2} = 0 (from\ (2))$$

$$\iff \sum_{i=-N-1}^{N+1} (b_i - b_{i+1}) x[n_0 - i] > 0 \quad (8)$$

Performing a similar computation for

$$y[n_0] > y[n_0 - 1] \quad (9)$$

We obtain

$$\sum_{i=-N-1}^{N+1} (b_i - b_{i-1}) x[n_0 - i] > 0 \quad (10)$$

Let

$$\alpha_i = b_i - b_{i-1} \quad (11)$$

Then substituting $i = i + 1$

$$\alpha_{i+1} = b_{i+1} - b_i = -(b_i - b_{i+1}) \quad (12)$$

$$\therefore b_i - b_{i+1} = -\alpha_{i+1} \quad (13)$$

Substituting (11) into (10) we obtain

$$\sum_{i=-N-1}^{N+1} \alpha_i x[n_0 - i] > 0 \quad (14)$$

Substituting (13) into (8) we obtain

$$\sum_{i=-N-1}^{N+1} -\alpha_{i+1} x[n_0 - i] > 0 \quad (15)$$

Given the two equations (14) and (15) that need to be satisfied for $y[n_0]$ to be a maxima, these conditions can be interpreted in a geometric manner. Consider the '$2N + 3$' long sequence of $x[n]$ where $n \in [n_0 - N - 1, n_0 + N + 1]$ as a point $\overline{x}$ in a $2N+3$ dimensional space. Let the sequence α_i where $i \in [-N-1, N+1]$ be denoted by a vector $\overline{\alpha_1}$ and the sequence $-\alpha_{i+1}$ where $i \in [-N-1, N+1]$ be denoted by vector $\overline{\alpha_2}$ in the same $2N + 3$ dimensional hyperspace. Then the conditions given by (14) and (15) are required for $Y[n_0]$ to be a maxima can be interpreted in the following manner:

1) Imagine a hyper plane passing through the origin and perpendicular to $\overline{\alpha_1}$. Any point $\overline{x}$ ($2N+3$ dimensional) that lies to one side of this hyperplane will satisfy

$$\sum_{i=-N-1}^{N+1} \alpha_i x[n_0 - i] > 0 \quad (16)$$

and all the points that lie on the other side will satisfy

$$\sum_{i=-N-1}^{N+1} \alpha_i x[n_0 - i] < 0 \quad (17)$$

2) One can similarly imagine a hyperplane corresponding to $\overline{\alpha_2}$ that divides the entire hyperspace into two regions corresponding to each of the below conditions

$$\sum_{i=-N-1}^{N+1} -\alpha_{i+1} x[n_0 - i] > 0 \text{ and} \qquad \sum_{i=-N-1}^{N+1} -\alpha_{i+1} x[n_0 - i] < 0 \quad (18)$$

3) From the above information, the region where

$$\sum_{i=-N-1}^{N+1} \alpha_i x[n_0 - i] > 0 \text{ and} \qquad \sum_{i=-N-1}^{N+1} -\alpha_{i+1} x[n_0 - i] > 0 \quad (19)$$

are satisfied is given by the intersection of two of the regions created by the hyperplanes perpendicular to $\overline{\alpha_1}$ and $\overline{\alpha_2}$.

The above explanation is illustrated in Fig. 2 which shows a two dimensional projection of the $2N+3$ dimensional space.

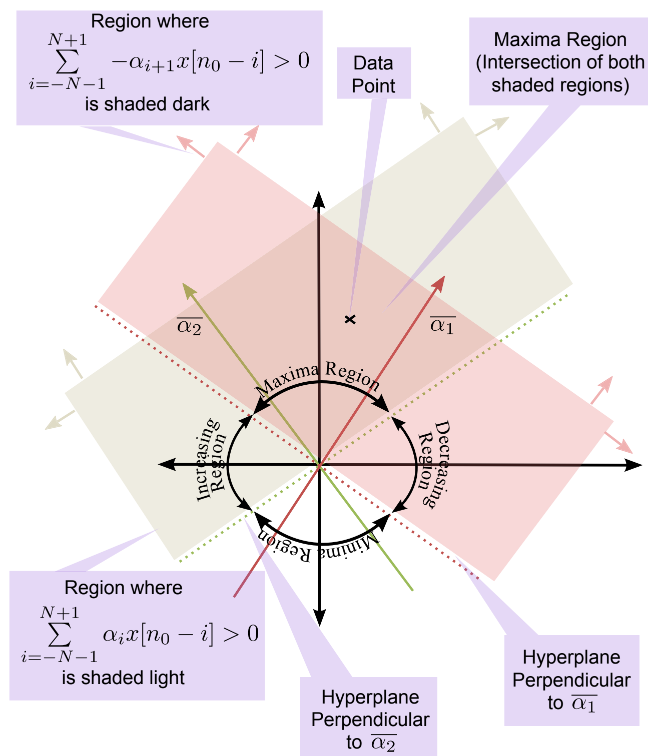

Fig. 2. A two dimensional projection of the filter hyperplanes and the regions associated with conditions given in equation (19).

From the above derivation, it can be seen that the hyperspace is divided into four regions. While one of the four regions is the maxima region the conditions that are satisfied in the other regions will lead to a corresponding minima region, a decreasing region and an increasing region for a particular filter. Fig. 2 illustrates a two dimensional projection of the $2N + 3$ dimensional space that identifies these regions. For a given signal x[n], one can extract all possible $2N+3$ long subsequences of points and populate them in a feature space and build hyper planes corresponding to a particular filter. Then all the subsequences that lie within the maxima region of the feature space will correspond to maxima in the filtered signal. This concept is illustrated in Fig. 2.

B. Robustness

Given that a noise signal could be added to the original signal, robustness is defined as the ability of the maxima and minima of the filtered signal to remain intact in spite of the addition of noise. In terms of the above geometric interpretation, that would mean that, after the addition of noise, the subsequences present in the maxima region remain in the maxima region and similarly other subsequences from other regions (minima, decreasing, increasing) should remain in their respective regions. This would ensure that all the maxima and minima are intact and no new extrema are formed.

Given the above explanation, the robustness of a subsequence can now be defined as the sum of the squared distances of a subsequence to both the hyper planes of a filter. Given this definition, it should be possible to find an OREF such that the sum of the squared distances to the hyperplanes of all the subsequences corresponding to that filter is maximized. It is important to note that the particular definition of robustness has implications on the nature of the distortions that the signal is likely to experience in pattern matching. For example, if one assumes that the noise to the signal is i.i.d, then the nature of distribution of a particular subsequence in the $2N+3$ dimensional hyperspace would be spherical. The outer radius of such a sphere would indicate the severity of the distortion and as long as this outer radius is less than the perpendicular distance of the subsequence to the filter hyperplanes, the point would not change its state from being an extrema or non-extrema. Thus, the current definition of robustness is apt in case of an i.i.d noise whose distribution's range is less than twice the perpendicular distance of a point to the closest boundary hyperplane.

C. Derivation for the Optimally Robust Extrema Filter (OREF)

Given the vector

$$\overline{\alpha_1} = [\alpha_{-N-1}\ \alpha_{-N}\ \alpha_{-N+1}\ \ldots\ \ldots\ \alpha_0\ \ldots\ \ldots\ \alpha_N\ \alpha_{N+1}\ ] \tag{20}$$

As the filter is of length $2N+1$ so $\alpha_{-N-1}=0$. Let

$$\overline{\alpha_2} = -[\alpha_{-N}\ \alpha_{-N+1}\ \alpha_{-N+2}\ \ldots\ \ldots\ \alpha_1\ \ldots\ \ldots\ \alpha_{N+1}\ \alpha_{N+2}\ ] \tag{21}$$

Similarly, as the filter is of length $2N+1$ so $\alpha_{N+2}=0$. Let the above two vectors represent perpendicular unit vectors to the hyperplanes for a particular filter. Therefore, the requirement that they be unit vectors imposes a constraint for α_1 and α_2 which is given by (22).

$$\sum_{i=-N}^{N+1} \alpha_i^2 = 1 \tag{22}$$

From (2) and (11), one obtains the additional constraint

$$\sum_{i=-N}^{N+1} \alpha_i = 0 \tag{23}$$

The above condition can be rewritten as

$$\alpha_{N+1} = -(\alpha_{-N} + \alpha_{-N+1} + \ldots.. + \alpha_N) \tag{24}$$

The sum of squared perpendicular distances of a particular point represented by a sequence

$$[x_{-N-1}\ x_{-N}\ x_{-N+1}\ \ldots\ \ldots\ x_0\ \ldots\ \ldots\ x_{N-1}\ x_N\ x_{N+1}\ ] \tag{25}$$

is given by the square of the dot product of the subsequence to $\overline{\alpha_1}$ and $\overline{\alpha_2}$

$$[\sum_{i=-N}^{N+1} \alpha_i x_i]^2 + [\sum_{i=-N}^{N+1} -\alpha_i x_{i-1}]^2 \tag{26}$$

Therefore, the total sum of the squared perpendicular distances (U_R) of all the $'M-N'$ points in the hyperspace is given by

$$U_R = \sum_{\substack{j\,=\,N\,+\,1\\ (Sum\ over\\ all\ points)}}^{M} [[\sum_{i=-N}^{N+1} \alpha_i x_{i+j}]^2 + [\sum_{i=-N}^{N+1} -\alpha_i x_{i+j-1}]^2] \tag{27}$$

where $x_0, x_1, x_2, x_3, x_4, \ldots\ldots\ldots, x_{M+N+1}$ represents the time series for which an OREF is being sought. Substituting (24) into the above equation we obtain

$$\begin{aligned} U_R = \sum_{j=N+1}^{M} [[\sum_{i=-N}^{N} \alpha_i (x_{i+j} - x_{j+N+1})]^2 + \\ [\sum_{i=-N}^{N+1} \alpha_i (x_{i+j-1} - x_{j+N})]^2] \end{aligned} \tag{28}$$

Let y_j denote a $2N+1$ vector whose i^{th} element is denoted by $(x_{-N-1+i+j} - x_{N+1+j})$. Next, let z_j denote a $2N+1$ vector whose i^{th} element is denoted by $(x_{-N-1+i+j-1} - x_{N+j})$. Finally, let α denote a $2N+1$ vector whose i^{th} element is denoted by α_{-N-1+i}. Then the above equation can be written in the matrix form as

$$U_R = \sum_{j=N+1}^{M} [\alpha^T y_j y_j^T \alpha + \alpha^T z_j z_j^T \alpha] \tag{29}$$

$$\iff U_R = \alpha^T [\sum_{j=N+1}^{M} [y_j y_j^T + z_j z_j^T]] \alpha \tag{30}$$

$$Let\ X_{Data} = \left[\sum_{j=N+1}^{M} [y_j y_j^T + z_j z_j^T] \right] \tag{31}$$

Clearly X_{Data} is positive semi definite from its very definition (28), but for most datasets for which $N << M$, the matrix X_{Data} will be positive definite. There are a few times series which are deterministic, in which case the time series satisfies a certain recurrence relation, and which satisfy the condition $N << M$ but would result in $det(X_{Data}) = 0$. As analysis of such 'special' time series is not the objective of this paper, further derivation is performed under the assumption $N << M$ and that X_{Data} is positive definite. Then, (30) can be expressed as

$$U_R = \alpha^T X_{Data} \alpha \tag{32}$$

In order to maximize U_R under the condition (22) one can utilize the Lagrangian multiplier method [17].Thus L given in (33) needs to be minimized

$$L = -\alpha^T X_{Data} \alpha + \nu[\sum_{i=-N}^{N+1} (\alpha_i)^2 - 1] \tag{33}$$

Using (24)

$$L = -\alpha^T X_{Data} \alpha + \nu[\sum_{i=-N}^{N} (\alpha_i)^2 + (\sum_{i=-N}^{N} (\alpha_i))^2 - 1] \tag{34}$$

Let $j_{1,2N+1}$ denote a $2N+1$ unit vector, then the above equation can be simplified to

$$L = -\alpha^T X_{Data} \alpha + \nu[\alpha^T I \alpha + \alpha^T j_{1,2N+1}^T j_{1,2N+1} \alpha - 1] \tag{35}$$

Let J_{2N+1} denote a $2N+1$ by $2N+1$ unit matrix,

$$L = -\alpha^T X_{Data} \alpha + \nu[\alpha^T [I + J_{2N+1}] \alpha - 1] \tag{36}$$

Given that there is one quadratic program with one quadratic inequality constraint, this problem is often referred to as the *trust region sub-problem* in mathematical literature [18][19]. Following the Lagrangian multiplier method, the stationary points are given by

$$\nabla_\alpha L = 0 \; and \; \nabla_\nu L = 0 \tag{37}$$

The condition $\nabla_\nu L = 0$ results in the Karush-Kuhn-Tucker condition

$$\alpha^T [I + J_{2N+1}] \alpha - 1 = 0 \tag{38}$$

Once the α vector satisfying the condition $\nabla_\alpha L = 0$ is obtained, it can be multiplied by a constant to satisfy (38) provided that the modified vector can still satisfy $\nabla_\alpha L = 0$. Solving for $\nabla_\alpha L = 0$

$$\nabla_\alpha L = 2(-X_{Data} \alpha + \nu[I + J_{2N+1}] \alpha) = 0 \tag{39}$$

$$\therefore X_{Data} \alpha = \nu[I + J_{2N+1}] \alpha \tag{40}$$

As ν is a scalar,

$$[I + J_{2N+1}]^{-1} X_{Data} \alpha = \nu \alpha \tag{41}$$

The above equation is the generalized eigenvalue problem. Given that X_{Data} is positive definite, one can rewrite the above equation in the form

$$[I + J_{2N+1}]^{-1} (X_{Data})^{1/2} (X_{Data})^{1/2} \alpha = \nu \alpha \tag{42}$$

where $(X_{Data})^{1/2}$ can be obtained from eigenvalue decomposition. Multiplying both sides of (42) with $(X_{Data})^{1/2}$

$$\begin{aligned} (X_{Data})^{1/2} [I + J_{2N+1}]^{-1} (X_{Data})^{1/2} (X_{Data})^{1/2} \alpha \\ = \nu (X_{Data})^{1/2} \alpha \end{aligned} \tag{43}$$

Substituting $w = (X_{Data})^{1/2} \alpha$ into (43) results in the regular eigenvalue problem

$$(X_{Data})^{1/2} [I + J_{2N+1}]^{-1} (X_{Data})^{1/2} w = \nu w \tag{44}$$

Thus the eigenvalues (ν_k) and eigenvectors (w_k) corresponding to the symmetric positive semi definite matrix $(X_{Data})^{1/2} [I + J_{2N+1}]^{-1} (X_{Data})^{1/2}$, will lead to the solution $(X_{Data})^{-1/2} w_k$ for the $'\alpha'$ vector. Given that the following problem is a trust region sub-problem, it has been shown that strong duality is satisfied [18] and so the Lagrangian relaxation for this non-convex problem is exact [19]. We define the Lagrange dual function as $g : R \rightarrow R$ as the minimum value of the lagrangian over a.

$$g(\nu) = inf_{a \in Domain} - a^T X_{Data} a + \nu[a^T [I + J_{2N+1}] a - 1] \tag{45}$$

As the stationary points satisfy (40), (45) can be simplified to

$$\begin{aligned} g(\nu) = inf_{a \in Domain} - a^T \nu [I + J_{2N+1}] a + \\ \nu[a^T [I + J_{2N+1}] a - 1] \end{aligned} \tag{46}$$

Therefore, the Lagrange dual function is

$$g(\nu) = inf_{a \in Domain} - \nu \tag{47}$$

Hence the optimal value for the function in (32) is given by the largest eigenvalue. Thus, from the different eigenvector solutions that are obtained, the most optimal solution is given by the eigenvector corresponding to the maximum eigenvalue. The above optimization process contained no constraints on the number of extrema that result from the filtered signal. Therefore, this procedure could theoretically result in a filter for which no extrema are created or a situation in which each and every point in the signal is an extrema. In case one obtains no extrema points from the OREF (corresponding to the maximal eigenvalue), then one can utilize subsequent eigenvalues and eigenvectors that follow the maximal eigenvalue to obtain OREFs that result in extrema. In case all the values or a large percentage of the signal values are chosen as extrema then it is advisable to smooth the signal and then use the optimization process to extract the filter.

IV. Experimental Results

The purpose of this section is to demonstrate the advantages obtained from using the extrema from the OREF as compared to using other traditional filters in a given pattern matching task. The datasets used in the experiments are described in subsection IV-A and the experimental procedure and final results are presented in subsection IV-B. The traditional filters which are used in the comparison process and the OREFs which are obtained from utilizing the method described in section III on the different datasets described in subsection IV-A are shown in Fig. 3.

A. *Data for experimental tests to illustrate robustness of the derived filter*

The following data was used as a part of the experiments:

1) Pitch Data (Real): The dataset consists of road pitch data collected from an in-vehicle data acquisition system [2]. The data was collected as a part of the NCHRP 22-21 median design project.
2) EEG Data (Real): EEG time series data obtained from different trials on human subjects is concatenated to obtain the time series used in this analysis. The data was downloaded from the following source [20].
3) Gaussian Random Walk (GRW) (Simulated): The data was obtained from a MATLAB simulation. The MATLAB code to obtain the simulated data is given by $Data = cumsum(randn(2^{16}, 1));$

B. *Test for subsequence matching ability*

The procedure used to compare the extrema obtained from using different filters, shown in Fig. 3, in a pattern/subsequence matching task is described in this subsection. The problem of matching a given query time signal to an existing time series dataset is called *time series subsequence matching*. The problem in encountered in various applications [2] dealing with time series data. A wide variety of solutions utilizing sliding window methods for dimension reduction or using different types of distance measures have been proposed [4],[11],[12]. However, extrema based methods offer certain advantages while solving the subsequence matching problem because of their ability to handle complex distortions [1]. For example, scale factor noise, bias noise and outliers are commonly encountered distortions in time series data. However, common index-based subsequence matching methods based on euclidean distance [4] and DTW [11] cannot handle outliers while certain edit distance based methods like Longest Common Sub-Sequence (LCSS) [12] are not designed to handle scale and outlier noise simultaneously.

In this particular set of experiments, 31 query signals are extracted from a time series of 65536 points obtained from the datasets in subsection IV-A. Each query signal contains 256 data points and is corrupted with bias noise, scale factor noise, and outliers before being tested for subsequence matching. The bias noise for a query signal was obtained by randomly selecting a value from a uniform distribution in the following interval [-std dev(query signal),+std dev(query signal)]. Similarly, the scale factor noise was obtained by randomly selecting a value from a uniform distribution in the following interval [0.75,1.25]. The outliers are simulated by multiplying a given number of randomly selected points in a query signal by a large scale factor. The large scale factor is chosen such that it results in a 15% matching accuracy when the above subsequence matching task is performed using the "standard" sliding window based euclidean distance matching method [4] in the presence of three outliers in the query signal. The query signal is also corrupted with a small amount of Gaussian noise (approximately 14dB SNR) that is often encountered in time series data in order to make the matching process more realistic.

The subsequence matching was performed by using the feature encoding / KD-tree methodology described in [2] for all the different filters shown in Fig. 3. In [2], the feature vectors are encoded at different scales by using the wavelet transform. However, in order to simplify the comparison process, features are extracted from a 'single scale' (with a filter containing 12 taps) in this paper. The entire subsequence matching process with all the 31 query signals, for each dataset, is repeated 30 times so as to obtain a statistical average of the performance accuracy.

A brief description of the subsequence matching process from [2] is presented here for the sake of completeness. The first step (Indexing phase) in the extrema based method is to build a KD-tree using the features generated from time series as described in section 2 of [2]. In the online phase, the features generated from the query signals are matched with the features present in the KD-tree database. The matches in the database yield different estimates of the location from which the query signal was extracted. These locations are put into a histogram to obtain the most agreed upon location estimate for the time series subsequence as described in [2]. If the location estimate from the matching process was within a certain threshold distance (within 25 data points 'or' approximately 10% of the length of the query signal) from the true location of extraction, then the result is deemed accurate. The main difference between the methodology in section 2 of [2] and the methodology followed in this paper is that, instead of using the wavelet transform for filtering, this paper used the filters shown in Fig. 3.

It is important to note that the same distortions, noise levels and accuracy thresholds were used while testing all the different filters. In the feature generation process, the threshold in the extrema detection step described in Fig. 1 is adjusted such that the number of extrema from each filter are approximately the same. This process ensures that the computational effort required for the different filtering techniques is also identical so that no filter has any undue advantages.

The results from these filters are compared in Fig. 4 to Fig. 6 for the different datasets and for different number of outliers introduced into the query signal. The results clearly show that the extrema from OREF outperform the traditional filters in the given subsequence matching task. The optimization method presented in section III was mainly aimed towards the robustness of the extrema created after filtering. It can be seen that this property is an important determinant of the performance in the above subsequence matching task. However, it is important to understand that because the above filter wasn't optimized for returning maximum accuracy in subsequence matching, it may be possible to find another linear filter that may exceed the performance of the filter in this particular subsequence matching application.

V. CONCLUSIONS AND FUTURE WORK

The core contribution of this paper is to demonstrate the possibility of optimizing filters in order to enable extraction

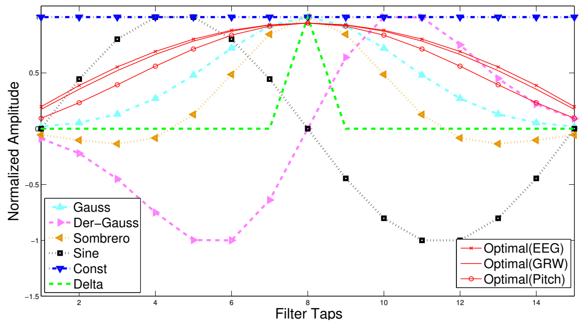

Fig. 3. The filters used in the experimental tests.

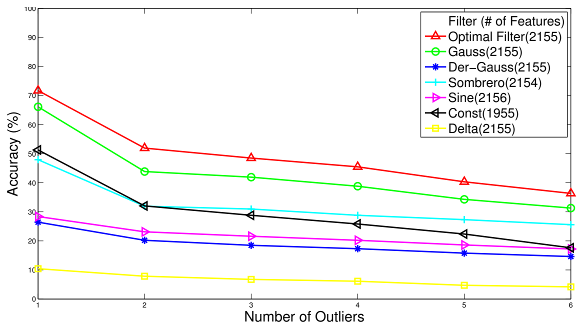

Fig. 4. The subsequence matching results for Pitch data using extrema features obtained from using different filters.

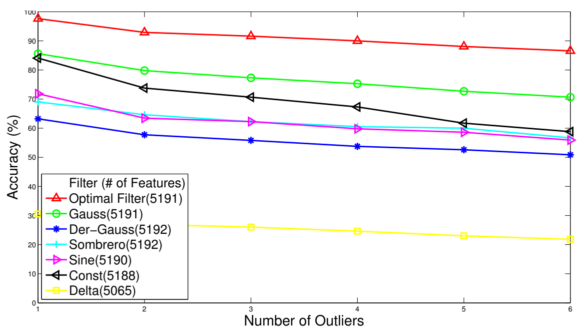

Fig. 5. The subsequence matching results for EEG data using extrema features obtained from using different filters.

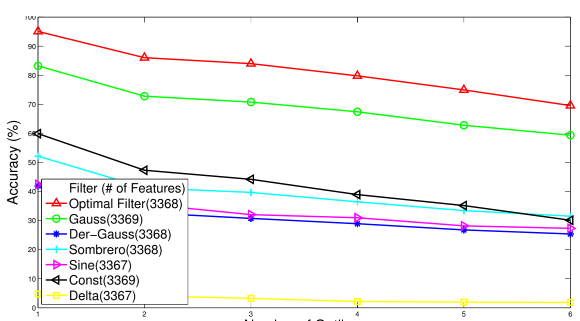

Fig. 6. The subsequence matching results for GRW data using extrema features obtained from using different filters.

of extrema with desirable properties. The methodology described here can be extended to a multi-scale extrema encoding methodology like [2] by performing the optimization at each scale separately. Future work could be directed towards building methods for controlling other important properties such as uniqueness or cardinality of the extrema extracted from the data. Also, Fig. 3 shows that the OREFs for different datasets have similar shapes and future work could also be directed towards deriving an analytical equation to represent this filter which could be useful for a wide range of datasets.

REFERENCES

[1] C.-S. Perng, H. Wang, S. R. Zhang, and D. S. Parker, "Landmarks: a new model for similarity-based pattern querying in time series databases," in *Proc. ICDE*, 2000.

[2] P. Vemulapalli, A. Dean, and S. Brennan, "Pitch based vehicle localization using time series subsequence matching with multi-scale extrema features," in *Proc. ACC*, June 2011.

[3] S. Kadetotad, P. Vemulapalli, S. Brennan, and C. Lagoa, "Terrain-aided localization using feature-based particle filtering," in *Proc. DSCC*, Nov 2011.

[4] C. Faloutsos, M. Ranganathan, and Y. Manolopoulos, "Fast subsequence matching in time-series databases," in *Proc. ACM SIGMOD*, 1994.

[5] K. Chan and A.-C. Fu, "Efficient time series matching by wavelets," in *Proc. ICDE*, 1999.

[6] F. Korn, H. V. Jagadish, and C. Faloutsos, "Efficiently supporting ad hoc queries in large datasets of time sequences," in *Proc. SIGMOD*, 1997.

[7] E. J. Keogh, K. Chakrabarti, M. J. Pazzani, and S. Mehrotra, "Dimensionality reduction for fast similarity search in large time series databases," in *Knowl. Inf. Syst*, 2001.

[8] Q. Chen, L. Chen, X. Lian, and Y. Liu, "Indexable PLA for efficient similarity search," in *Proc. VLDB*, 2007.

[9] Y. Cai, "Indexing spatio-temporal trajectories with chebyshev polynomials," in *Proc. SIGMOD*, 2004.

[10] J. Lin, L. Wei, and et al., "Experiencing SAX: A novel symbolic representation of time series," in *Data Mining and Knowledge Discovery*, 2007.

[11] E. Keogh, "Exact indexing of dynamic time warping," in *Proc. VLDB*, 2002.

[12] M. Vlachos, D. Gunopoulos, and G. Kollios, "Discovering similar multidimensional trajectories," in *Proc. ICDE*, 2002.

[13] P.-F. Marteau, "Time warp edit distance with stiffness adjustment for time series matching," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 2, pp. 306 –318, feb. 2009.

[14] L. Chen and R. Ng, "On the marriage of lp-norms and edit distance," in *Proc. VLDB*, 2004.

[15] A. Wang, "An industrial-strength audio search algorithm," in *Proc. ISMIR*, 2003.

[16] A. Dean, P. Vemulapalli, and S. Brennan, "Highway evaluation of terrain-aided localization using particle filters," in *Proc. DSCC*, Nov 2008.

[17] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, Mar. 2004.

[18] R. J. Stern and H. Wolkowicz, "Indefinite trust region subproblems and nonsymmetric eigenvalue perturbations," in *SIAM Journal of Optimization*, vol. 5, no. 2, 1995, pp. 286–313.

[19] K. Anstreicher, X. Chen, H. Wolkowicz, and Y.-X. Yuan, "Strong duality for a trust-region type relaxation of the quadratic assignment problem," *Linear Algebra and its Applications*, vol. 301, no. 1-3, pp. 121 – 136, 1999.

[20] R. Andrzejak, K. Lehnertz, C. Rieke, F. Mormann, P. David, and C. lger, "Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Phys. Rev. E*, 2001.