THE UNIVERSITY *of* LIVERPOOL

ADVANCED MATHEMATICAL MORPHOLOGY AND ITS APPLICATION TO SIGNAL PROCESSING

Thesis submitted in accordance with the
requirements of the University of Liverpool
for the degree of Doctor of Philosophy

in

Electrical Engineering and Electronics

by

Tianyao JI, B.Sc.(Eng.), B.A., M.Sc.(Eng.)

November 2009

ADVANCED MATHEMATICAL MORPHOLOGY AND ITS APPLICATION TO SIGNAL PROCESSING

by

Tianyao JI

To my family

Acknowledgements

First and foremost, I would like to express my deepest appreciation and acknowledgement to my supervisor, Professor Henry Wu, not only for his dedicated supervision and inspiring discussions at every stage of this research, but also for his intellectual advice and guidance in both academic and personal level.

I would also like to thank The Duncan Norman Charitable Trust and Overseas Research Students Awards Scheme for the financial support they provided for my research work at The University of Liverpool and living.

My gratitude goes to all of my colleagues, especially Dr. Zhen Lu, Mr. Mengshi Li, Mr. Dayu Shi and Mr. Jonathan Buse for their kind help throughout my studies.

Finally, I am greatly indebted to my parents and husband, for their love, encouragement, patience and understanding. I dedicate this thesis to them for their unwavering love and trust.

Abstract

Mathematical morphology (MM) is a non-linear approach to signal and image processing, based on set theory, integral geometry, topology, lattice algebra, random function, *etc.* Ever since Matheron and Serra laid the foundation of MM in 1964, the research has flourished in the following decades, especially after the 1980s. MM has been a very active area of research all over the world. However, MM research undergoes a bottleneck recently. A major issue that limits the fast development of MM is that most work undertaken in this area is mainly concerned with image processing and its potential has not been fully explored for applications in signal processing and for 'shape' processing in a high dimensional space.

This thesis is concerned with the development of advanced morphological operators and their application to signal processing. Effective may it be, classic MM does not meet the requirement of feature extraction of complex signals. In this thesis, traditional morphological operators are extended towards four directions, including morphological decomposition schemes, multiscale morphological filters, biologically inspired optimal morphological filters, and embedding-based techniques. The advanced morphological operators developed in the these are applied to the broad areas of electrocardiogram (ECG) signal analysis, power system protection and image processing.

Inspired by the success of the wavelet transform (WT) in signal/image processing, the framework of multi-resolution decomposition scheme has been introduced to MM. Based on the general framework that encompasses most existing linear and nonlinear multi-resolution decompositions, a multi-resolution morphological filter is proposed. With the help of the lifting scheme, mor-

phological lifting filter is presented as an improvement of the multi-resolution morphological filter, due to its feature recovery ability. Applying the proposed filters to ECG signal analysis, which requires the removal of impulsive noise and the detection of characteristic waveforms, the results are very satisfactory. The WT is a linear tool in its original form, however, MM has made the nonlinear extensions possible. The morphological gradient wavelet (MGW) is such a nonlinear WT. Thanks to its advantage in feature extraction, the MGW has been successfully applied to detect power quality disturbances, which is a crucial issue in power system protection.

The Fourier transform (FT) obtains a frequency spectrum of the signal and the frequency is a measure of the spectral content. In the same manner, the shape and size content of a signal can be measured by geometrical structure and a pattern spectrum of the signal can be obtained through a set of size-related structuring elements (SEs). Such a study is called multiscale shape description. Two advanced multiscale morphological operators have been proposed in this thesis, both equipped with targeted filtering ability for noise removal of ECG signals. The adaptive multiscale morphological filter aims to reject high frequency noise without affecting the feature waveforms, and an approach hybridising empirical mode decomposition (EMD) and MM is designed to remove the baseline wander.

In order to reinforce the filtering ability of basic morphological filters, a new morphological filter is designed based on soft MM. Moreover, linear filtering techniques are also combined with the morphological ones to fully explore their potential. To guarantee best performance, the parameters of such filters are optimised by evolutionary algorithms (EAs). For the former filter, an improved particle swarm optimiser with passive congregation (PSOPC) subject to the least mean square error criterion is used to optimise the parameters. For the latter, a bacterial swarming algorithm (BSA) is presented to optimally combine linear and morphological filtering techniques. The optimally assignment of morphological operators, the SEs and other parameters makes the filters useful in areas of artificial intelligence.

It has been proved that MM can be extended to n dimension. The thesis devotes a chapter on processing a signal in the corresponding phase space, which has a higher dimension than the time or shape domain. Such a scheme is proposed on the believe that the features of a signal can be more clearly revealed if it is embedded to the phase space through a proper embedding strategy. The embedding-based signal processing technique is applied to three applications: feature waveform detection of ECG signals, phasor measurement of power system signals and disturbance detection for power quality monitoring and analysis.

Declaration

The author hereby declares that this thesis is a record of work carried out in the Department of Electrical Engineering and Electronics at the University of Liverpool during the period from October 2006 to September 2009. The thesis is original in content except where otherwise indicated.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation and Objective

Mathematical morphology (MM) has been recognised as a powerful tool for image/signal processing. Having experienced flourishing development for decades, MM research now faces new challenges. A majority of work in this field concentrates on image processing, whereas the potential of MM in complex signal processing has been underestimated. As the development of modern technology, target signals met in practice become more and more complex, and the processing requirement in terms of accuracy, adaptability, speed, *etc*, are more and more demanding. Traditional approaches or technologies, especially the linear ones and including the classic MM, not only suffer from incapacity towards these issues but also limit the development of signal processing. Therefore, there is an urgent need to develop generic advanced morphological operators that are able to meet the requirement of complex signal processing.

Research on investigating MM's potential in frequency analysis has always been insufficient. For many signals, frequency is a crucial feature. Hence, it is attractive to investigate the characteristics of morphological operators in the frequency domain and the relationship between the MM-based transformation and the frequency-based transformation. Unfortunately, this area has not been explored. The mathematical background of MM implies the difficulty

of bridging MM to frequency analysis, however, it is still worthwhile to start the research by developing morphological operators that can process frequency information and comparing its performance with traditional signal processing techniques that work well under given conditions in the frequency domain.

Extending MM to a higher dimensional space is another topic that has long been ignored, although its mathematical background bolsters its function in a space of any dimension. It is believed that embedding a signal to the phase space through a proper embedding strategy, the features of the signal can be more clearly revealed. Since the accurate extraction of transient features is desirable for a wide range of applications in military defense, vehicle navigation, safety monitoring, traffic and road plan, *etc*, it is necessary to conduct research on feature extraction in the phase space.

The advanced morphological operators are applied to the following three applications in this thesis: electrocardiogram (ECG) signal analysis, power system protection and image denoising. ECG signals record the electrical activity of the heart of a patient and the waveforms provide important information for clinic diagnosis. They show evidence of disease or reveal rhythm problems when heart beat is abnormal. Therefore, it requires accurate identification and measurement of feature waveforms.

Power system protection is closely related to the electric power sector. There are hundred and thousand millions of protection relays operating in power systems and plants in the world. The methodologies of existing protective relaying has not been changed for over 50 years. It is based on the Fourier transform that is time consuming, computationally complicated and insufficient to extract the features of transient fault signals. In addition, existing relays operate based on a long sampling window, with a low response, and are large in size and prone to suffer from electromagnetic disturbances, because they are constructed from a large number of electronics components and boards. Hence, it is an urgent issue to develop a new generation of protective relaying algorithms that can avoid these problems and have the potential to be mapped to a system on a chip (SOC) to enhance the reliability of morphological

operations and reduce the size and cost of the protection relays.

Noise removal is an underlying issue for image processing, especially in industrial applications as noise widely appears in raw images. Various filters have been proposed to reject noise, however, most of them are designed specifically for a certain type of noise. Therefore, a generic strategy on designing an appropriate filter aiming at removing compound noise meanwhile preserving the details of the image is greatly in need. These areas are closely related to our daily life and the public will benefit greatly from the outcome of the research.

1.2 Mathematical Morphology

1.2.1 History

MM was introduced in 1964 by Matheron [1] and Serra [2, 3], both researchers at the Paris School of Mines in Fontainebleau. Their objective was to characterise physical or mechanical properties of certain materials, such as the permeability of porous media, by examining the geometrical structure [4]. They introduced a set formalism to analyse binary images, which allows the image objects to be processed by simple operations such as unions, intersections, complementation and translations, and published their first reports concerning some of the methods that are part of MM now. Matheron rediscovered Minkowski algebra, which defined set addition (dilation) in 1903. In that period, Matheron had undertaken the study of permeability for porous media in relation with their geometry on a more theoretical level [5] and published a seminal book entitled *Random Sets and Integral Geometry* [1], which is considered to have laid down the foundations of MM.

Since its launching in 1964, MM grew rapidly in the 1970s on the theoretical level. The major contribution of Matheron's work during this period includes topological foundations, random sets, increasing mappings, convexity and several models of random sets. Based on iterative processing, binary thinning, skeletons, ultimate erosion, conditional bisectors and their geodesic framework were introduced. With the significant development of automated

visual inspection, substantial developments in morphology were stimulated in the 1980s. The most important development of MM in that period was the setting of the method in the mathematical framework of complete lattices. The theory of the complete lattice provides a compact theoretical foundation for grey-scale morphological operations. The theory of morphological filtering was also presented in the 1980s. Some real-time applications of morphological filtering were developed, such as the Delft image processor (DIP). In the 1990s, the growth of MM focused on its applications, which included robot vision, medical imaging, visual inspection, texture and scene analysis, *etc.* The theoretical works on the concept of connection, in combination with the connected filters, have made MM a remarkable powerful tool for segmentation [6, 7]. The slope transform was developed by endowing morphological operators with eigenfunctions and their related transfer functions. Furthermore, a relatively new approach of MM, soft morphology, was introduced by Koskinen [8]. The recent trends in soft morphology including the algorithms and implementations can be found in [9], [10] and [11]. Other extensions of MM include fuzzy morphologies, which attempt to apply fuzzy set theory to MM [12, 13, 14]. MM has also been hybridised with other techniques to explore its potential, such as artificial neural networks (ANNs) [15], evolutionary algorithms (EAs) [16, 17, 18, 19], *etc.* With over three decades of development, MM has become a powerful tool for geometrical shape analysis.

A major research topic of MM is digital image analysis. The development in both methods and applications has been equally boosted and there has always been a close interconnection between them. Mature and integrated methods have been formed, such as image filtering, image segmentation and classification, image measurements, pattern recognition, and texture analysis and synthesis, *etc.* As for the application of MM, it includes visual inspection and quality control, optical character recognition and document processing, materials science, geosciences, and life science. An overview of the development of MM can be found in two milestone monographs by Serra [2] and Soille [20], respectively.

MM has been widely used in the areas of image processing, but only a few investigations have been attempted for signal processing. In this field, MM prefers to depict the profile of signal waveforms in the time domain directly, in contrast to many traditional methodologies that place emphasis on the response in the frequency domain, such as the Fourier transform (FT) and the wavelet transform (WT).

The underlying basis of morphological filtering technique is to process signals by a function, known as the structuring element (SE) or the structuring function generally. An SE slides through the signal as a moving window, inspects its interaction with the signal, and detects specific features in the neighbourhood of every point in the signal. A morphological filter is composed of various combinations of fundamental morphological operators, which can preserve or suppress the feature represented by SE and obtain a signal with only components of interest. When operating upon a signal of complicated shapes, morphological filters are capable of decomposing a signal into certain parts. These parts are separated from the background for identification or other purposes, and the main shape characteristics of the signal are preserved. Therefore, for a distorted signal, its underlying shapes can be identified, reconstructed and enhanced. Furthermore, the mathematical calculation involved in MM includes only addition, subtraction, maximum and minimum operations, without any multiplication or division. Hence, morphological operators calculate faster than traditional integral transforms when processing the same signal.

In contrast to linear frequency domain techniques, MM is concerned with the shape of a signal waveform in the time domain. When used for the extraction of waveform components, MM has the following merits in comparison with the integral transform-based methods:

1. The morphological operators have fast and simple calculations without using multiplication and division operations.

2. It is applicable to non-periodic transient signals and not restricted to periodic signals.

3. MM uses a much smaller size of sampling window in real-time signal processing, as it does not require the information of the full signal components.

4. It is able to accurately and reliably extract the signal components without causing any distortion, as it is a time-domain signal processing method without performing any signal integral transforms.

1.2.2 Mathematical Background

Having been developed for several decades, MM has become a powerful tool for signal and image processing, especially for geometrical shape analysis. Considering its mathematical background, MM is defined on a complete lattice. Before giving the definition of a complete lattice, we need to understand the concept of a partially ordered set, which is also called a poset. A partially ordered set is a set in which a binary relation '$\leq$' is defined for certain pairs of elements. The binary relation '$\leq$' over a set $\mathcal{P}$ satisfies the following conditions for all elements $x, y, z \in \mathcal{P}$:

1. *Reflexive:* $\forall\ x,\ x \leq x$.

2. *Antisymmetry:* If $x \leq y$ and $y \leq x$, then $x = y$.

3. *Transitivity:* If $x \leq y$ and $y \leq z$, then $x \leq z$.

Given two partially ordered sets A and B and arbitrary elements a and x, the following definitions can be developed:

1. *Translation:* The translation of A by x, denoted by $(A)_x$, is defined as $(A)_x = \{a + x | a \in A\}$.

2. *Reflection:* The reflection of A, denoted by $\check{A}$, is defined as $\check{A} = \{-a | a \in A\}$. Reflection is also called transposition.

3. *Complement:* The complement of A, denoted by A^c, is defined as $A^c = \{x | x \notin A\}$.

4. *Difference:* The difference between two sets A and B, denoted by $A-B$, is defined as $A-B=\{x|x\in A, x\notin B\}=A\bigcap B^c$. Based on this operation, the complement of set A can also be defined as $A^c=\{x|x\in I-A\}$, where I is the universal set.

The partially ordered set formalises the intuitive concept of an ordering relation, which plays a key role in MM [20]. Aside from partial ordering, there also exists a total ordering relation. A totally ordered set has a strengthened relation of '<': for any two elements x and y, exactly one of $x<y$, $x=y$, $x>y$ is true. The property of transitivity on a totally ordered set becomes $x<y$ and $y<z$ implies $x<z$.

A poset, $(\mathcal{P},\leq)$, is a lattice if any two elements of it, x and y, have a greatest lower bound (*i.e.* infimum), $x\wedge y$, and a least upper bound (*i.e.* supremum), $x\vee y$. A lattice, $\mathcal{J}$, is a complete lattice if each of its subsets has an infimum and a supremum in $\mathcal{J}$. A complete lattice satisfies the following properties: for subsets X, Y and Z,

1. *Commutativity:* $X\vee Y=Y\vee X$, $X\wedge Y=Y\wedge X$.

2. *Associativity:* $(X\vee Y)\vee Z=X\vee(Y\vee Z)$, $(X\wedge Y)\wedge Z=X\wedge(Y\wedge Z)$.

1.3 Three Major Application Areas

The advanced morphological operators are mainly applied to three areas: ECG signal analysis, power system protection and image processing. All the advanced morphological operators developed in this thesis have been applied to tackle an issue from these areas and the narration style of algorithm followed by application is found throughout the thesis. This section introduces in detail the characteristics of the signals met in the three areas, addresses the objectives of processing them, and summarises relevant work reported in recent literature.

1.3.1 ECG Signals

ECG records heart electrical signals activated first by the atria then by the ventricles. The analysis of the ECG signals is a well-known non-invasive technique to detect the electrical heart activity. A normal heart cycle is reflected in the ECG by three feature waveforms: the contraction of the atria results in a P wave, followed by a QRS complex, which is produced by the contraction of the ventricles, and a T wave, which results from the subsequent return of the ventricular mass to a rest state [21]. To be more intuitive, a fragment of an ECG signal is shown in Fig. 1.1.

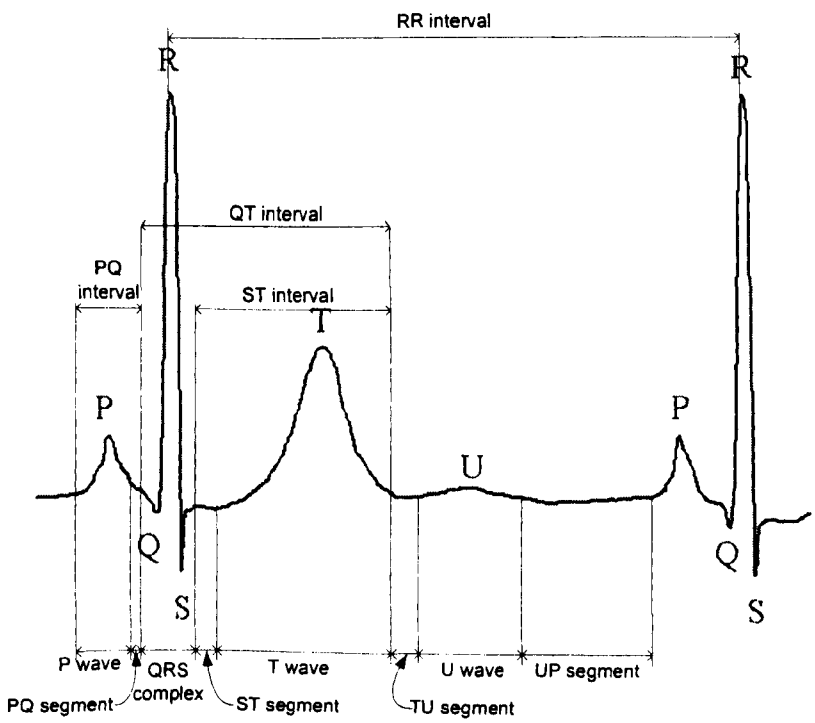


Figure 1.1: Feature waveforms of an ECG signal.

The ECG signal has a time periodicity allowing to define an elementary beat composed by specific waveforms, appearing periodically in time [22]. Figure 1.1 shows a heartbeat and its respective waveform labels. The QRS complexes represent the ventricular activity of the heart [23]. The R waves have the

highest amplitude and in practice, an RR interval can be used to define a cycle of the ECG signal.

An ECG signal may be disturbed by impulsive noise, due to electrical activity of muscles [24], high frequency noise (usually modelled by Gaussian white noise), due to electromagnetism interference and grounding of biomedical instruments and measuring circuits [25], and low frequency baseline wander, due to the respiratory cycle. Therefore, noise removal is usually the first step of ECG analysis. In some cases, the noise has the same magnitude and frequency as the feature waveforms. Therefore, it is important to minimise the distortion in the feature waveforms so as to keep the features of most interest in terms of clinical diagnosis, while removing the noise [26]. Paper [27] proposed a progressive umbra-filling (PUF) procedure to adaptively assign the SEs of openings or closings, so as to provide different filtering scales with respect to different spacial positions. The PUF procedure first extracts the feature parts (*e.g.* jumping edges) from a signal; then, assigns different filtering scales to the feature parts and the other parts, respectively. In this manner, the feature waveforms of the signal can be preserved while the noise is removed.

Accurate extraction the feature waveforms is a much concerned topic. Some algorithms detect QRS complexes only. Paper [23] studied and compared three non-linear QRS detection algorithms. Two inherent characteristics of the QRS complex are that first, it has high frequency content compared to the rest of the ECG signal; second, it has large amplitude. Hence, the difference or derivative of the QRS complexes yields a larger value than the remainder of the signal. This feature can be exploited by squaring the differences or by multiplying successive differences together. Other algorithms tried to obtain all the characteristic points of the ECG, *i.e.* the onset, peak and end of the P waves, QRS complexes and T waves [28].

Classification of the feature waveforms has also aroused considerate amount of interest. A number of papers are devoted to this field. In [29], several ANN architectures were designed to classify the ECG signals integrating the most common features: arrhythmia, myocardial ischemia, chronic alterations. In

[22], a hidden Markov model-based approach was proposed for automatic beat segmentation and classification. The hidden Markov models were trained on a large set of waveform examples and gave a good representation of the morphologies that can be found in an ECG signal. Some literature focuses on the extraction of P waves and their morphology classification and several clinical studies connected certain P-wave properties, such as its width and morphology, with anomalies in the electrical atrial conduction and atrial pathology [28]. Paper [30] presented comparative study of the classification of QRS complexes in five heartbeat types.

In the simulation studies of this thesis, test ECG signals come from MIT-BIH Arrhythmia Database [31]. They were digitised at 360 samples per second with 11-bit resolution over ±5 mV range. Sample values thus range from 0 to 2047 inclusive, with a value of 1024 corresponding to 0 mV.

This thesis proposes a multi-resolution morphological filter and a morphological lifting filter for the removal of impulsive noise, an adaptive morphological filter for the removal of Gaussian white noise, and a filter engaging empirical mode decomposition (EMD) and MM for the removal of baseline wander.

The proposed multi-resolution morphological filter can also be applied to extract the feature waveforms. It uses the fact that R waves have a highest magnitude and the magnitude and width of the P waves are smaller than those of the T waves to separate the waveforms level by level. As an alternative, an embedding-based scheme is designed to identify the feature waveforms and moreover, the waveforms can be classified by their geometric properties in the phase space.

1.3.2 Power System Signals

Transient disturbances of power systems

Power quality has been one of the main concerns over the operation of utilities and manufacturing industries. In general sense, anything that causes the power system voltage or current to deviate from its ideal sinusoidal waveform

could be considered as a power quality disturbance [32]. According to the IEEE 1159 standard [33], which is a well-established power quality standard, a given disturbance is generally considered to belong to one of the two broader categories: steady-state disturbances and temporary disturbances. Steady-state disturbances are of long duration and change little with time, including waveform distortion caused by harmonics, noise and power frequency variations. Temporary disturbances, on the other hand, usually appear in a steady-state and as such a disturbance disappears, the power system will return to normal or to a new steady-state. Temporary disturbances are further divided into several different types including long duration voltage variation, lasting longer than one minute, short duration voltage variation, lasting longer than 0.5 cycles of the power frequency but less than or equal to 1 minute, and transient, lasting shorter than 50 ms, *etc.* The following disturbances are considered and dealt with in this thesis.

- **Transient** A transient is an undesirable momentary deviation of the supply voltage or load current. Transients can be broadly classified into two categories - impulsive and oscillatory.
  - *Impulsive transient*: An impulsive transient is a sudden, non-power frequency change in the steady-state condition, and it is unidirectional in polarity (primarily either positive or negative). The most common cause of impulsive transients is lightning.
  - *Oscillatory transient*: An oscillatory transient consists of a voltage or current whose instantaneous value changes polarity rapidly, *i.e.* it includes both positive or negative polarity value.
- **Short duration variation** Short duration voltage variations are almost always caused by fault conditions, the energisation of large loads that require high starting currents, or intermittent loose connections in power wiring. Depending on the fault location and the system conditions, the fault can cause either temporary voltage rises (swells) or voltage drops (sags), or a complete loss of voltage (interruptions).

- *Voltage swell*: A voltage swell is a short term increase of system voltage. Such an event is often caused by an abrupt reduction in load or appears on the un-faulted phases of a three-phase circuit, where a single-phase short circuit has occurred. Swell may stress any delicate equipment components causing premature failure.
- *Voltage sag*: A voltage sag is a significant voltage reduction for a relatively short duration. It may be caused by faults, increased load demand and transitional events such as starting large electrical motors.
- *Interruption*: A momentary loss of voltage on a power system can be called a momentary interruption. Such a disturbance describes a drop of 90-100 percent of the rated system voltage. They are often detected down-line from switch-gear such as circuit breakers, re-closers or fuses.

- **Waveform distortion** Waveform distortion is a steady-state deviation from an ideal sine wave of power frequency principally characterised by the spectral content of the deviation. There are five primary types of waveform distortion: DC offset, harmonics, inter-harmonics notching and noise.
  - *Harmonics*: Harmonics are sinusoidal voltages or currents having frequencies that are integer multiples of the fundamental frequency (the frequency at which the supply system is designed to operate). Harmonic distortion exists due to the nonlinear characteristics of devices and loads on the power system.
  - *Notching*: Notching is a periodic disturbance of opposite polarity from the waveform lasting shorter than a half of a cycle. It may be caused by lightning, static discharges, utility switching operations, starting or stopping major equipment or machinery, *etc.*
- **Power frequency variation** Small changes in the fundamental frequency occur when the balance between the load and the capacity of

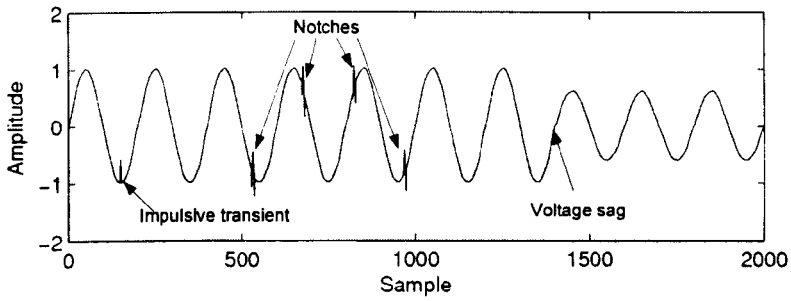


Figure 1.2: A simulated power system disturbance signal.

the available generation changes. The size of the frequency shift and its duration depend on the load characteristics and the response of the generation system to load changes.

Figure 1.2 demonstrates a disturbance signal containing three types of disturbances – an impulsive transient, four notches and a voltage sag. In order to examine the power quality, voltage or current waveforms may be recorded continuously using power monitoring instruments which generate digitised time series of sampled data [34]. The detection is based on the time series. A common way to assess and monitor the recorded data is through frequency-based approaches, such as discrete Fourier transform (DFT) and WT [35, 36, 37]. DFT has a successful application to periodic signals, but fails to track a transient signal due to the limitation that it must perform in a window of a fixed length [38]. Unlike DFT, WT constructs a string of time-frequency representations of a signal and the representations are of different resolutions. Hence, it is more capable of identifying the details of localised transients. However, it requires a sampling window of a certain length to perform integral calculations, which increases the computational burden and impairs its attraction. Besides, WT does not reflect the shape information in the time domain, which is essential for the analysis of some particular distorted portion of power system signals. Other approaches involve the analysis of the root-mean-square (RMS) value of voltage. In [39], the RMS-based measurement is used for detection and classification of short and long duration variations, and digital

filtering and mathematical morphology are used for transients and waveform distortions.

Having ascertained that a disturbance has occurred in the power system, the next task is to determine the type of the disturbance. Many artificial intelligence-based algorithms have been proposed for disturbances classification of power system signals. Some of the algorithms make use of neural networks [40] and/or expert systems [35, 41]. Merged with different detection approaches, they can provide satisfactory classification results. However, all of these techniques rely on massive data for training, which greatly limits their practical applications.

In this thesis, test disturbance signals are simulated to contain up to 40 harmonics and Gaussian noise with a signal-to-noise ratio (SNR) of 30 $\sim$ 60 dB, and they may also have a $\pm 5\%$ fundamental frequency shift. As for each disturbance, its location, duration and magnitude are randomly selected within a range [33] so that they are at different levels. Impulsive transients are simulated at millisecond scale, which means they last over 1 ms but shorter than 0.5 cycles. Such a short undervoltage cannot be characterised effectively as a change in the RMS value of the fundamental frequency value [33]. Oscillatory transients are simulated to contain low frequency (300 $\sim$ 900 Hz) component only, with a duration of 0.3 $\sim$ 50 ms and a magnitude of 0 $\sim$ 4 p.u. The voltage of swells and sags in the range of 1.1 $\sim$ 1.8 p.u. and 0.1 $\sim$ 0.9 p.u., respectively, and both disturbances last within the instantaneous range, which is a half of a cycle at least and 30 cycles at most. For interruptions, the reduced voltage is in the range of 0 $\sim$ 0.1 p.u. and last between 0.5 cycles to 3 s. The magnitude of notches is limited within 0.2 $\sim$ 0.6 p.u. and their frequency is 10 $\sim$ 20 times of the signal's fundamental frequency. A test signal may contain up to 6 notches and their interval is 1/4 $\sim$ 1 cycle. Main parameter settings are listed in Table 1.1, which cover major situations that may occur in power systems.

Table 1.1: Main parameter settings of each type of disturbance

Power disturbance	Duration	Magnitude (p.u.)
Millisecond impulsive transients	> 1 ms	$0 \sim 1$
Low frequency oscillatory transients	$0.3 \sim 50$ ms	$0 \sim 4$
Instantaneous voltage sag	$0.5 \sim 30$ cycles	$0.1 \sim 0.9$
Instantaneous voltage swell	$0.5 \sim 30$ cycles	$1.1 \sim 1.8$
Momentary interruption	$0.5 \sim 150$ cycles	$0 \sim 0.1$
Notching	$1 \sim 4$ ms	$0.2 \sim 0.6$

Fault currents in power systems

When a fault occurs in a power system, the current or voltage signal in the transmission line will contain harmonics and exponentially decaying DC offset components. The source current or voltage signal has the standard sinusoidal waveform:

$$I_0(t) = A_0 \cos(\omega t + \phi) \tag{1.3.1}$$

where $\omega = 2\pi/T$ with T the period. On the other hand, a fault current or voltage signal is considered to comprise a fundamental sinusoidal component, a series of harmonics and an exponential decaying DC offset, which can be expressed as:

$$I_{\rm f}(t) = I_1(t) + I_{\rm h}(t) + I_{\rm dc}(t) \tag{1.3.2}$$

$$I_1(t) = A_1 \cos(\omega t + \varphi) \tag{1.3.3}$$

$$I_{\rm h}(t) = \sum_{j=2} I_j(t) = \sum_{j=2} A_j \cos(j\omega t + \varphi) \tag{1.3.4}$$

$$I_{\rm dc}(t) = Be^{-\beta t} \tag{1.3.5}$$

where I_1 is the fundamental component and it has the same frequency as the source signal but has a phase shift of an angle of $\varphi - \phi$, $I_{\rm h}$ is the combination of the harmonics, and $I_{\rm dc}$ denotes the DC offset. Therefore, the signal including a source part and a fault part can be expressed by:

$$I(t) = \begin{cases} I_0(t), & 0 < t < t_s \\ I_{\rm f}(t - t_s), & t \geq t_s \end{cases} \tag{1.3.6}$$

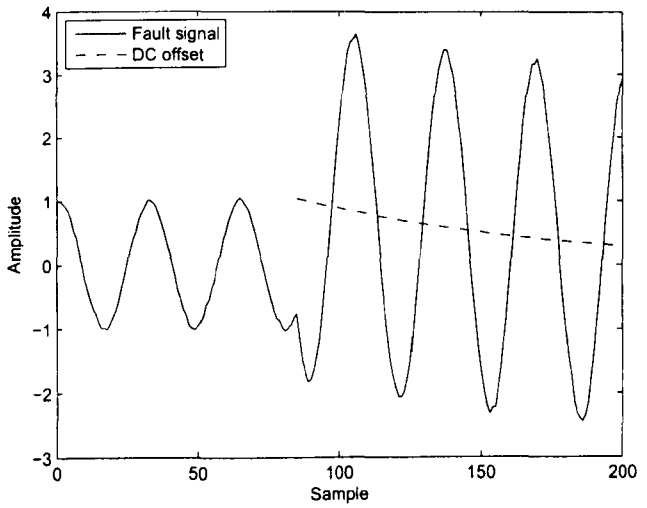


Figure 1.3: A typical fault current in power systems.

where t_s denotes the time instant that the fault occurs. A typical example of a fault signal is given in Fig. 1.3, which contains an exponential decaying DC offset and 35 dB Gaussian noise. A traditional way of analysing fault currents is the full-cycle and half-cycle FTs, as described in Appendix A.

1.3.3 Compound Noise in Images

Raw images often contain Gaussian white noise, periodic noise and salt & pepper noise. Gaussian white noise widely appears in all practical situations. It contains equal power within a fixed bandwidth at any centre frequency and is uncorrelated in time. Periodic noise is mainly caused by electrical interference from data collecting devices, such as image scanners, capturing sensors, and video cameras. Apparently, it has a well-defined frequency. Salt & pepper noise is a type of non-Gaussian noise and typically seen on images. It represents itself as randomly occurring white and black pixels.

This thesis sets up a generic framework of designing an optimal filter to remove the compound noise. The filter contains morphological operation and may also engage linear filtering technique if desired, and has a set of adjustable

parameters. EAs are used to search for the most suitable parameters, which is a contribution in introducing new blood to MM.

1.4 Major Contributions of Thesis

The major contributions of this research include the fundamental study of morphological operators, development of advanced morphological operators based on the basic ones and their application to signal processing, in particular for ECG signals and power system signals, and image processing. The contributions are summarised as follows.

- The thesis digs up in the mathematical background of MM and gives proof for certain properties of basic morphological operators.
- After a through study of a set of morphological operators that are suitable for signal processing in one dimension, advanced morphological operators have been developed, some of which are hybridised with other signal processing techniques, to fulfill specific tasks. Nonetheless, the operators can be adopted in other applications.
- A research of employing EAs in the design of optimal morphological filters is conducted. Two EAs are employed in the optimisation of morphological filters for the removal of periodic noise and Gaussian and non-Gaussian noise, respectively. In such a manner, it is ensured that the filters designed can filter out the noise and have the best performance according to pre-set criteria.
- The embedding theorem, which was originally proposed for dynamical systems, is for the first time applied to signal processing to view a signal from more than one angle and thus, hidden features can be extracted. This pioneer work presents a generic scheme that can be adopted by a wide range of applications.

- Initial investigation of MM in power quality analysis is carried out and three schemes for disturbance detection and location are proposed. Furthermore, the classification of power quality events is conducted based on the detection results.

- An embedding-based method is proposed for phasor measurement to replace the FT, which is commonly used but requires complex integral calculation and thus increases the computation burden. The new method measures the amplitude, the phase angle and the phase difference of current and voltage signals sample by sample and the calculation involves two or four samples only. Moreover, the method can also be used to estimate the shift of the fundamental frequency.

- Several MM-based algorithms are developed for ECG signal processing, including the removal of impulsive noise, high frequency noise and low frequency baseline wander, and the identification of the feature waveforms. As the frequency spectrum of the noise and feature waveform overlaps with one another, the MM-based algorithms that focus on the shape of the signals are more applicable.

1.5 Contents of Thesis

The thesis is structured as follows.

Chapter 2 This chapter introduces the mathematical background of MM. It starts with the basic morphological operators, dilation and erosion, in binary case and extends them to grey scale version. Afterwards, the basic operators evolve to a set of sequential operators, such as opening and closing. Other types of MM-based schemes are also introduced in this chapter, including soft MM, multi-resolution MM and multiscale MM, from which the advanced morphological operators proposed in the following chapters are developed.

Chapter 3 In this chapter, three morphological decomposition algorithms are introduced. For the purpose of noise removal, a multi-resolution morphological filter and a morphological lifting filter are designed. The latter is an improvement over the former in that it can partly recover the detailed information lost during the filtering procedure. The multi-resolution morphological filter is applied to ECG signal analysis. Using it at different levels, the multi-resolution morphological filter can remove the impulsive noise or separate the feature waveforms. For the purpose of feature extraction, a morphological gradient wavelet (MGW) is developed. It involves morphological operation in the decomposition process and presents the gradient information in the output. The MGW is applied to power system disturbance detection to detect and locate the disturbances occurring in a power system signal. Its performance is compared with a classic WT – the Daubechies DB4 WT (DB4), and the result shows the advantage of the MGW in terms of accuracy and computational complexity.

Chapter 4 Morphological operators can be designed into a multiscale version to extract target information scale by scale. This chapter presents two advanced multiscale morphological filters. The first one is an adaptive multiscale morphological filter that can reject high frequency noise without affecting the feature waveforms. It is used to process an ECG signal from a larger scale to a smaller scale in a recursive manner to distinguish high frequency noise from feature waveforms according to a set of preset criteria. The output is a much smoother curve that still follows the trends of the feature waveforms. The second filter is a fusion of EMD and MM. It removes low frequency noise whose spectrum has a wide span and partly overlaps with that of the feature waveforms. The hybrid filter is applied to remove the baseline wander of ECG signals, which traditional filter techniques are unable to deal with, and provides a stable signal for subsequent automatic processing and reliable visual interpretation.

Chapter 5 Morphological filters may contain a range of parameters and the performance of the filters can be enhanced by adjusting the parameters according to the requirement of the task. The scheme of designing optimal morphological filters is described in this chapter. Two types of morphological filters are designed in this chapter and their adjustable parameters are optimised by two EAs, namely particle swarm optimiser with passive congregation (PSOPC) and bacterial swarming algorithm (BSA), respectively. Simulation studies demonstrate that the optimal filters are more powerful than traditional ones in removing compound noise, which is often the case met in practice.

Chapter 6 This chapter is concerned with a novel signal processing technique developed based on the embedding theorem and on the belief that the features of a signal can be more clearly revealed in a high-dimensional space. After the introduction of the embedding theorem, the embedding-based technique is applied to deal with three types of problems. The first one is the detection and classification of ECG feature waveforms. The feature waveforms and the baseline form different shapes in the phase space, and the embedded signal is therefore used to separate them from one another. Consequently, clinically important parameters, such as the onset, length, peak and end of a P or T wave, can be measured for further analysis and diagnosis. Moreover, the geometric information presented by the embedded signal is used to classify the feature waveforms, which is also often required in clinic. The second application area is phasor measurement of power system signals, which used to be dominated by the FT-based algorithm. This chapter proposes a fast yet reliable embedding-based method for the measurement of amplitude, phase, fundamental frequency, frequency shift, *etc.* In the area of power quality protection, the disturbances are extracted using the embedding-based technique with the help of morphological operation, data clustering technique and algebra-based algorithms. In the phase space, the embedded disturbances and the embedded normal signal form distinguish shapes,

which ensures a more accurate detection and location of the disturbances. An on-line classification strategy is also presented based on the disturbances extracted.

Chapter 7 Finally, the thesis is concluded in this chapter and suggestions for future work are given.

1.6 Auto-bibliography

Book

1. Wu, Q.H., Lu, Z., **Ji, T.Y.**, *Protective Relaying of Power Systems Using Mathematical Morphology*, Dordrecht, Heidelberg, London, New York: Springer-Verlag, 2009.

Journal Papers

1. **Ji, T.Y.**, Lu, Z., Wu, Q.H., *Optimal soft morphological filter for periodic noise removal using a particle swarm optimiser with passive congregation*, in Signal Processing, 87(11), November 2007, pp. 2799-2809.

2. **Ji, T.Y.**, Lu, Z., Wu, Q.H., Ji, Z., *Baseline normalisation of ECG signals using empirical mode decomposition and mathematical morphology*, in Electronics Letters, 44(2), January 2008, pp. 82-83.

3. **Ji, T.Y.**, Lu, Z., Wu, Q.H., *Detection of power disturbances using morphological gradient wavelet*, in Signal Processing, 88(2), February 2008, pp. 255-267.

4. **Ji, T.Y.**, Wu, Q.H., *Disturbance detection using an improved hit-or-miss transform*, International Journal of Modelling, Identification and Control, 8(1) the special issue of Stochastic Systems and Applications, 2009.

5. **Ji, T.Y.**, Li, M.S., Lu, Z., Wu, Q.H., *Optimal estimation of harmonics in dynamic environment using an adaptive bacterial swarming algorithm*, submitted to IET Power Electronics Journal.

6. **Ji, T.Y.**, Lu, Z., Wu, Q.H., *Disturbance Detection, Location and Classification in Phase Space*, submitted to IEEE Transactions on Circuits and Systems I.

7. Lu, Z., **Ji, T.Y.**, Tang, W.H, Wu, Q.H., *Optimal harmonic estimation using a particle swarm optimiser*, in IEEE Transactions on Power Delivery, 23(2), April 2008, pp. 1166-1174.

8. Lu, Z., **Ji, T.Y.**, Wu, Q.H., *A morphological transform for exponentially decaying DC-offset removal*, in Electronics Letters, 44(9), April 24 2008, pp. 595-U104.

9. Lu, Z., Tang, W.H., **Ji, T.Y.**, Wu, Q.H., *A morphological scheme for inrush identification in transformer protection*, in IEEE Transactions on Power Delivery, 24(2), April 2009, pp. 560-568.

10. Lu, Z., **Ji, T.Y.**, Wu, Q.H., *EHV transmission line protection using a morphological lifting scheme*, accepted by Electric Power Systems Research, 2009.

11. Wu, Q.H., **Ji, T.Y.**, Li, M.S., Lu, Z., *Distributed optimal power flow using bacterial swarming algorithm*, in International Journal of Modelling, Identification and Control, the special issue of Modelling and Control Electrical Systems, 2009.

12. Li, M.S., **Ji, T.Y.**, Tang, W.J., Wu, Q.H., Saunders, J., *Bacterial foraging algorithm with varying population*, submitted to Biosystems.

13. Wu, Q.H., **Ji, T.Y.**, Li, M.S., Lu, Z., *Group search optimiser with multiple producers for reactive power dispatch*, submitted to IEEE Transactions on Power Systems.

14. Li, M.S., **Ji, T.Y.**, Lu, Z., Wu, Q.H., *Distributed optimal power flow using bacterial swarming algorithm*, submitted to IEEE Transactions on Power Systems.

Conference Papers

15. **Ji, T.Y.**, Lu, Z., Wu, Q.H., A particle swarm optimiser applied to soft morphological filters for periodic noise reduction, in Lecture Notes on Computer Science (LNCS 4448), The Ninth European Workshop on Evolutionary Computation in Image Analysis and Signal Processing, EvoIASP2007, Valencia, Spain, 11-13 April 2007, pp. 367-374, ISBN 978-3-540-71804-8.

16. **Ji, T.Y.**, Lu, Z., Wu, Q.H., Multi-resolution morphological operators for electrocardiogram signal analysis, in The 26th Chinese Control Conference, Zhangjiajie, China, 26-31 July 2007, Vol 3, pp. 425-429.

17. **Ji, T.Y.**, Lu, Z., Tang, W.H., Wu, Q.H., Disturbance detection in the phase space through GK clustering, in The Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies, DRPT2008, Nanjing, China, 6-9 April 2008.

18. **Ji, T.Y.**, Li, M.S., Lu, Z., Wu, Q.H., Optimal filter design using a bacterial swarming algorithm, in 2008 IEEE Congress on Evolutionary Computation, CEC 2008, Hong Kong, 1-6 June 2008, pp. 452-458.

19. Lu, Z., Tang, W.H., **Ji, T.Y.**, Wu, Q.H., A phase selector based on mathematical morphology for double circuit transmission lines, in The Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies, DRPT2008, Nanjing, China, 6-9 April 2008.

20. Wu, Q.H., Lu, Z., Li, M.S., **Ji, T.Y.**, Optimal placement of FACTS devices by a group search optimiser with multiple producer, in 2008 IEEE Congress on Evolutionary Computation, CEC 2008, Hong Kong, 1-6 June 2008, pp. 1033-1039.

21. Li, M.S., **Ji, T.Y.**, Lu, Z., Wu, Q.H., Optimal harmonic estimation using dynamic bacterial swarming algorithm, in 2008 IEEE Congress on

Evolutionary Computation, CEC 2008, WCCI2008, Hong Kong, 1-6 June 2008, pp. 1302-1308.

22. Lu, Z., **Ji, T.Y.**, Wu, Q.H., An adaptive distance relaying algorithm with a morphological fault detector embedded, in IEEE Power & Energy Society General Meeting, Calgary, Alberta, Canada, 26-30 July 2009.

Chapter 2

Mathematical Morphology

2.1 Basic Morphological Operators

2.1.1 Definitions for Binary Operations

Minkowski Addition and Subtraction

The main function of morphological operators is to extract relevant structures of a set. The extraction is usually done by the interaction between the set and another set, which is called structuring element (SE). The shape of the SE is pre-defined according to some *a priori* knowledge about the shape of the signal. There are two basic morphological operators, dilation and erosion, which form a pair of dual transforms. They are derived from Minkowski set theory. Therefore, we start this section from the introduction to Minkowski addition and subtraction.

Minkowski addition is a binary operation of two sets A and B in Euclidean space, named after Hermann Minkowski. It is denoted by $\overset{M}{\oplus}$ and is defined as the result of adding every element of A to every element of B:

$$A \overset{M}{\oplus} B = \{a + b | a \in A, b \in B\} = \bigcup_{b \in B} (A)_b. \tag{2.1.1}$$

It is assumed that $A \overset{M}{\oplus} \{0\} = A$ and $A \overset{M}{\ominus} \emptyset = \emptyset$. The dual operation is called

Minkowski subtraction, denoted by $\overset{M}{\ominus}$, and defined as:

$$A \overset{M}{\ominus} B = \bigcap_{b \in B} (A)_b. \tag{2.1.2}$$

The following relation will be true for Minkowski subtraction:

$$A \overset{M}{\ominus} B = \{x | (\check{B})_x \subseteq A\} = (A^c \overset{M}{\oplus} B)^c. \tag{2.1.3}$$

Proof:

$$\begin{aligned}
\because \quad & z \in \left(\bigcap_{b \in B} (A)_b \right) \\
\Rightarrow \quad & z \in (A)_b \\
\Rightarrow \quad & z \in \{a + b | a \in A, b \in B\} \\
\Rightarrow \quad & z \in \{x | x - b \in A. b \in B\}
\end{aligned}$$

$$\begin{aligned}
\text{and} \quad \because \quad & \{x | (\check{B})_x \subseteq A\} = \{x | x - b \in A, b \in B\} \\
\therefore \quad & A \overset{M}{\ominus} B = \{x | (\check{B})_x \subseteq A\}.
\end{aligned}$$

$$\begin{aligned}
\because \; A^c \overset{M}{\oplus} B &= \bigcup_{b \in B} (I - A)_b = I - \bigcap_{b \in B} (A)_b \\
&= \left(\bigcap_{b \in B} (A)_b \right)^c = \left(A \overset{M}{\ominus} B \right)^c \\
\therefore \; A \overset{M}{\ominus} B &= (A^c \overset{M}{\oplus} B)^c.
\end{aligned}$$

Next, we will prove that Minkowski addition satisfies the properties of commutativity and associativity. Similar methods can be used to prove that Minkowski subtraction also satisfies these properties.

1. *Commutativity:* $A \overset{M}{\oplus} B = B \overset{M}{\oplus} A$.

   Proof:

$$\begin{aligned}
A \overset{M}{\oplus} B &= \{x | x = a + b, a \in A, b \in B\} \\
&= \{x | x = b + a, b \in B, a \in A\} = B \overset{M}{\oplus} A.
\end{aligned}$$

2. *Associativity:* $(A \overset{M}{\oplus} B) \overset{M}{\oplus} C = A \overset{M}{\oplus} (B \overset{M}{\oplus} C)$.

Proof:

$$\begin{aligned} A \overset{M}{\oplus} (B \overset{M}{\oplus} C) &= \{a + (b + c) | a \in A, b + c \in B \overset{M}{\oplus} C\} \\ &= \{a + b + c | a \in A, b \in B, c \in C\} \\ &= \{(a + b) + c | a + b \in A \overset{M}{\oplus} B, c \in C\} \\ &= (A \overset{M}{\oplus} B) \overset{M}{\oplus} C. \end{aligned}$$

Binary Dilation and Erosion

The definition of dilation is similar to Minkowski addition. Replacing the operator $\overset{M}{\oplus}$ by $\ominus$ in (2.1.1) and use a reflected SE, $\check{B}$, we have

$$A \oplus B = \bigcup_{b \in \check{B}} (A)_b = \bigcup_{b \in B} (A)_{-b} = \bigcup_{b \in B} \{x | x = a - b, a \in A\}. \tag{2.1.4}$$

Erosion derives from Minkowski subtraction, but this time b comes from the reflection set of B. The definition of erosion is

$$A \ominus B = \bigcap_{b \in \check{B}} (A)_b = \bigcap_{b \in B} (A)_{-b} = \bigcap_{b \in B} \{x | x = a - b, a \in A\}. \tag{2.1.5}$$

Since a binary image is a digital image that has only two possible values for each pixel, it is very convenient to describe a binary image using the notion of a set. A binary image is often considered as a set I, while an object in it is considered as a subset $X \subseteq I$. Letting sets A and B represent two binary images in the above definitions, we have the operations of dilation and erosion for binary image processing.

The two images A and B function differently in image processing. Generally speaking, A is the image being processed, while B serves as an SE that slides as a probe across image A and interacts with each pixel of A. Obviously, the size of B should be much smaller than that of A. To have a clear view of this process, we give an example in Fig. 2.1 to show how dilation and erosion function between a binary image and an SE. Here, the origin of B is set at $(0, 0)$.

Figure 2.1 illustrates an important property of dilation and erosion – duality, which means that applying dilation to A is equivalent to applying erosion

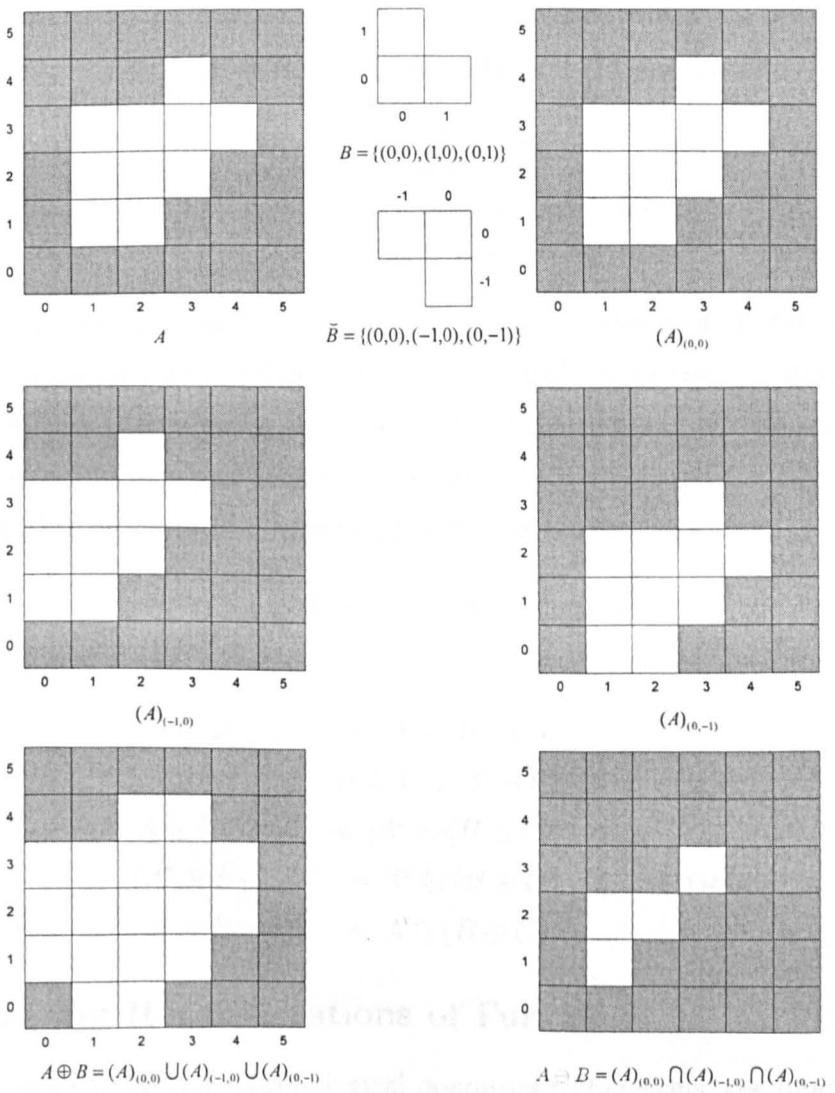


Figure 2.1: Binary dilation and erosion of a binary image

to its complement A^c. This property can be expressed as:

$$A \oplus B = (A^c \ominus B)^c \tag{2.1.6}$$

$$A \ominus B = (A^c \oplus B)^c. \tag{2.1.7}$$

Proof:

$$\begin{aligned} A^c \oplus B &= \bigcup_{b \in \check{B}} (I - A)_b = \bigcup_{b \in B} (I - A)_{-b} \\ &= I - \bigcap_{b \in B} (A)_{-b} \\ &= I - A \ominus B = (A \ominus B)^c \\ \Rightarrow \quad A \ominus B &= (A^c \oplus B)^c. \end{aligned}$$

The property of duality illustrates that the processing of dilation and erosion is not reversible and there is no inverse transform for the operators. As we can see from the following sections, applying dilation and erosion alternately actually produces a pair of new operations.

Another property of dilation and erosion is the distributivity:

$$A \ominus B \ominus C = A \ominus (B \oplus C). \tag{2.1.8}$$

Proof:

$$\begin{aligned} A \ominus B \ominus C &= (A^c \oplus B)^c \ominus C \\ &= [(A^c \oplus B) \oplus C]^c \\ A \ominus (B \oplus C) &= [A^c \oplus (B \oplus C)]^c \\ \because \ (A^c \oplus B) \oplus C &= A^c \oplus (B \oplus C) \quad (\textit{Associativity}) \\ \therefore \quad A \ominus B \ominus C &= A \ominus (B \oplus C). \end{aligned}$$

2.1.2 Set Representations of Functions

In order to extend morphological operators to functions, the functions are represented by their umbra [42], which is defined as:

$$U(f) = \{(x, a) | a \leq f(x)\}. \tag{2.1.9}$$

Hence, a d-dimensional function $f(x)$ is represented by a $(d+1)$-dimensional set. Obviously, the umbra is the set of points below the surface represented by

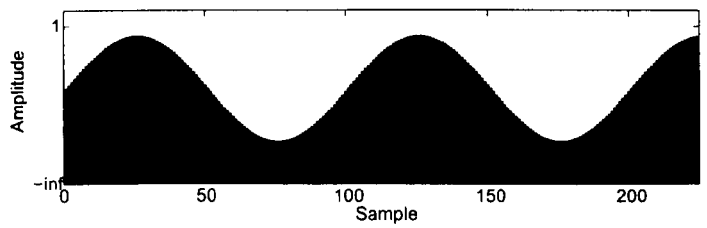


Figure 2.2: Umbrae $U(f)$ of a sinusoidal function f

$f(x)$. After getting the umbra, binary morphological operators can be applied to the signal. In general, the umbra set extends to $a = -\infty$, and the function f can be reconstructed from its umbra since:

$$f(x) = \max\{a|(x, a) \in U(f)\}, \ \forall x. \tag{2.1.10}$$

Figure 2.2 shows, as an example, the umbra of a sinusoidal function, where the umbra of $f(x)$ is the shaded region. We can easily show that $f \leq g \Leftrightarrow U(f) \subseteq U(g)$.

Some definitions for grey-scale operations based on sets are defined as follows [43]:

1. *Grey-scale union:* The union of two functions f and g, denoted by $f \vee g$, is defined as:

$$(f \vee g)(x) = f(x) \vee g(x). \tag{2.1.11}$$

There is a one-to-one correspondence between the union of functions and the set union:

$$U(f \vee g) = U(f) \cup U(g). \tag{2.1.12}$$

2. *Grey-scale intersection:* The intersection of two functions f and g, denoted by $f \wedge g$, is defined as:

$$(f \wedge g)(x) = f(x) \wedge g(x). \tag{2.1.13}$$

A similar one-to-one correspondence exists for the function and the set intersection:

$$U(f \wedge g) = U(f) \cap U(g). \tag{2.1.14}$$

3. *Grey-scale transpose:* The transpose $\check{f}$ of a function f is defined as:

$$\check{f}(x) = f(-x). \tag{2.1.15}$$

4. *Grey-scale complement:* The complement f^c of a function f is defined as:

$$f^c(x) = -f(x). \tag{2.1.16}$$

For an eight-bit grey-scale digital image, each pixel of which can have $2^8 = 256$ possible values to carry the full and only information about its intensity, the image can be represented as a set whose components are in $\mathbb{Z}^3$. In this case, two components of each element of the set refer to the coordinates of a pixel, and the third corresponds to its discrete intensity value. For a signal, the set is defined in $\mathbb{Z}^2$ with each element corresponding to a sample of the digitised signal. Similarly, the first component of each element represents the coordinate and the second represents its value. Sets in a higher dimensional space can contain other attributes, such as the colour information of an image.

2.1.3 Grey-Scale Dilation and Erosion

In order to use MM in signal processing where most signals are not binary, morphological operators should be extended to a grey-scale level. Instead of performing dilation and erosion by union and intersection as in the binary case, they are performed by algebraic addition and subtraction in the grey-scale case. Let f denote a signal and g denote an SE, and the length of g be considerably shorter than that of f [44]. Dilation and erosion are defined as follows:

$$f \oplus g(x) = \max_s\{f(x+s) + g(s) | (x+s) \in \mathcal{D}_f, s \in \mathcal{D}_g\} \tag{2.1.17}$$

$$f \ominus g(x) = \min_s\{f(x+s) - g(s) | (x+s) \in \mathcal{D}_f, s \in \mathcal{D}_g\} \tag{2.1.18}$$

where $\mathcal{D}_f$, $\mathcal{D}_g$ are the definition domains of f and g, respectively. For example, suppose the SE, g, has a length of five samples with its origin in the middle. In this case, the domain of g is given by $\mathcal{D}_g = \{-2, -1, 0, 1, 2\}$. The dilation

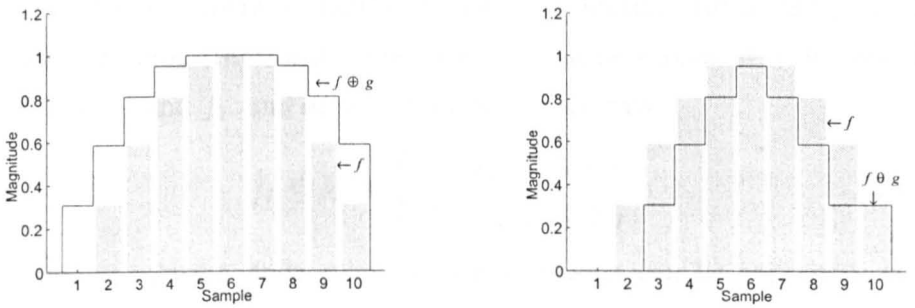


Figure 2.3: Grey-scale dilation and erosion of a one-dimensional signal. (a) Dilation. (b) Erosion.

and erosion of f by g are therefore calculated from

$$\begin{aligned} f \oplus g(x) &= \max\{f(x-2)+g(-2), f(x-1)+g(-1), f(x)+g(0), \\ &\qquad f(x+1)+g(1), f(x+2)+g(2)\} \end{aligned}$$

and

$$\begin{aligned} f \ominus g(x) &= \min\{f(x-2)-g(-2), f(x-1)-g(-1), f(x)-g(0), \\ &\qquad f(x+1)-g(1), f(x+2)-g(2)\} \end{aligned}$$

respectively. Intuitively, dilation can be imagined as swelling or expanding, while erosion can be thought of as a shrinking procedure.

As explained previously, the SE is a small set used to probe the signal under study. A simple case is that the SE has the form of $g(s) \equiv 0$, $s \in \mathcal{D}_g$, which is referred to as a 'flat SE'. Hence, definitions of dilation and erosion degrade to:

$$f \oplus g(x) = \max_s\{f(x+s) | (x+s) \in \mathcal{D}_f, s \in \mathcal{D}_g\} \tag{2.1.19}$$

$$f \ominus g(x) = \min_s\{f(x+s) | (x+s) \in \mathcal{D}_f, s \in \mathcal{D}_g\}. \tag{2.1.20}$$

The function of g is to indicate which samples are involved when processing the current sample. For a binary signal, the SE, g, must be flat. The dilation and erosion of a one-dimensional signal are illustrated in Figs. 2.3a and b, respectively. Both operations use a flat SE of length 3: $g(-1) = g(0) = g(1) = 0$.

Figure 2.3 demonstrates another property of dilation and erosion, *i.e.* that they are increasing transforms. The property can be expressed as follows. For two signals, f_1 and f_2, and an arbitrary SE, g, we have:

$$f_1 \leq f_2 \Rightarrow \begin{cases} f_1 \oplus g \leq f_2 \oplus g \\ f_1 \ominus g \leq f_2 \ominus g \end{cases} . \tag{2.1.21}$$

The ordering relation between dilation and erosion can be expressed as the erosion of a signal by an SE being less than or equal to its dilation by the same SE: $f \ominus g \leq f \oplus g$. If the SE contains its origin, which means processing a sample of the signal within a window that contains the sample, the following ordering exists:

$$f \ominus g \leq f \leq f \oplus g. \tag{2.1.22}$$

2.2 Morphological Filters

2.2.1 Definitions of Morphological Filters

Morphological filters are non-linear signal transforms that locally modify the geometrical features of signals or image objects. The idempotence and increasing properties are necessary and sufficient conditions for a transform, ψ, to be a morphological filter:

ψ is a morphological filter $\Leftrightarrow$ ψ is increasing and idempotent.

The property of idempotence implies that applying a morphological filter twice to a signal is equivalent to applying it only once:

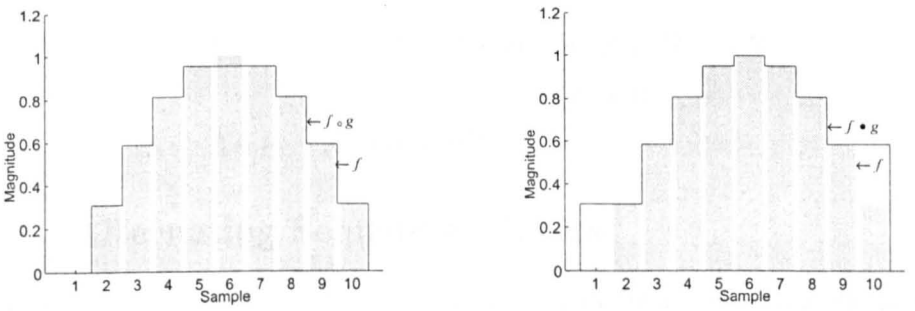
ψ is idempotent $\Leftrightarrow \psi\psi = \psi$.

The increasing property ensures that the ordering relation on signals is preserved after being filtered if the same SE is employed.

2.2.2 Opening and Closing

Opening is an operator that performs dilation on a signal eroded by the same SE. The definition is given as follows:

$$f \circ g = (f \ominus g) \oplus g \tag{2.2.1}$$



(a) (b)

Figure 2.4: Grey-scale opening and closing of a one-dimensional signal. (a) Opening. (b) Closing

where f is the signal, g is the SE, and $\circ$ denotes the opening operator. Opening can recover most structures lost by erosion, except for those completely erased by erosion. Closing, on the other hand, can be defined by its duality as:

$$f \bullet g = (f \oplus g) \ominus g. \tag{2.2.2}$$

Usually, opening and closing are also denoted by operators γ and ϕ, respectively. The results of performing opening and closing on the signal used in the previous section by the same SE are illustrated in Fig. 2.4.

Morphological opening and closing are both increasing transforms:

$$f_1 \leq f_2 \Rightarrow \begin{cases} \gamma(f_1) \leq \gamma(f_2) \\ \phi(f_1) \leq \phi(f_2) \end{cases}. \tag{2.2.3}$$

Moreover, successive applications of openings or closings do not further modify the signal, which means that they are both idempotent transforms:

$$\begin{aligned} \gamma\gamma &= \gamma \\ \phi\phi &= \phi. \end{aligned} \tag{2.2.4}$$

Apparently, opening and closing fulfill the conditions of morphological filters.

Opening and closing are also a pair of dual transforms:

$$A \circ B = (A^c \bullet B)^c \tag{2.2.5}$$

$$A \bullet B = (A^c \circ B)^c. \tag{2.2.6}$$

Proof:

$$\begin{aligned} (A \circ B)^c &= (A \ominus B \oplus B)^c = (A \ominus B)^c \ominus B \\ &= A^c \oplus B \ominus B = A^c \bullet B \\ \Rightarrow \quad A \circ B &= (A^c \bullet B)^c. \end{aligned}$$

2.2.3 Alternating Sequential Filters

Opening and closing are the basic morphological filters and new filters can be designed from their sequential combinations, such as an opening followed by a closing or vice versa. In fact, all the following combinations are filters: $\gamma\phi$, $\phi\gamma$, $\gamma\phi\gamma$ and $\phi\gamma\phi$. Moreover, for these new filters, the ordering relations of

$$\gamma \leq \gamma\phi\gamma \leq \begin{matrix} \gamma\phi \\ \phi\gamma \end{matrix} \leq \phi\gamma\phi \leq \phi \tag{2.2.7}$$

are always satisfied [20]. The pair of dual filters $\gamma\phi$ and $\phi\gamma$ are called opening-closing and closing-opening filters, and they have almost the same filtering effects. Therefore, in practice, usually only one of them is employed.

In some applications, such as when the objects under processing are over a wide range of sizes, there is a need to alternatively use openings and closings with an SE of an increasing size. This sequential application of the basic operators is called an alternating sequential filter. Since the four types of sequential combinations of opening and closing are all morphological filters, four alternating sequential filters can be developed hereafter. Let γ_i and ϕ_i be a pair of dual operators with an SE of size i. Suppose the size of SE increases from i to j. Therefore, the four types of alternating sequential filters are given as:

$$f_{\text{aoc}} = (\gamma_j\phi_j)\ldots(\gamma_i\phi_i) \tag{2.2.8}$$

$$f_{\text{aco}} = (\phi_j\gamma_j)\ldots(\phi_i\gamma_i) \tag{2.2.9}$$

$$f_{\text{aoco}} = (\gamma_j\phi_j\gamma_j)\ldots(\gamma_i\phi_i\gamma_i) \tag{2.2.10}$$

$$f_{\text{acoc}} = (\phi_j\gamma_j\phi_j)\ldots(\phi_i\gamma_i\phi_i). \tag{2.2.11}$$

2.3 Soft Mathematical Morphology

Soft MM (SMM) is an extension of standard MM. Instead of applying local maximum and minimum operations, SMM uses a more general weighted order statistics [45]. Besides, the SE used in SMM is divided into two subsets: the hard centre and the soft boundary. In comparison with standard MM, SMM is more robust in noisy conditions and less sensitive to small variations in the shapes of the objects [46].

Given sets $A, B \subseteq \mathbb{Z}^2$ and $A \subseteq B$, B is divided into two subsets: the hard centre A and the soft boundary $B - A$, where $B - A$ means the difference of the two sets. Soft dilation and soft erosion of an image f by an SE $[B, A, k]$ are defined as:

$$\begin{aligned} f \oplus [B, A, k] &= \max^{(k)}\{k \diamond (f(x-\alpha) + A(\alpha)) | \alpha \in \mathcal{D}_A\} \\ &\quad \bigcup \{f(x-\beta) + B(\beta) | \beta \in \mathcal{D}_{B-A}\} \qquad (2.3.1) \\ f \ominus [B, A, k] &= \min^{(k)}\{k \diamond (f(x+\alpha) - A(\alpha)) | \alpha \in \mathcal{D}_A\} \\ &\quad \bigcup \{f(x+\beta) - B(\beta) | \beta \in \mathcal{D}_{B-A}\} \qquad (2.3.2) \end{aligned}$$

respectively, where $\max^{(k)}$ and $\min^{(k)}$ denote the k^{th} largest and smallest value in the set respectively; $\diamond$ is the repetition operator and $k \diamond f(a) = \{f(a), f(a) \cdot \cdot, f(a)\}$ (k times); $\mathcal{D}_A$, $\mathcal{D}_{B-A}$ represent the field of definition of A and $B - A$ respectively.

Consequently, soft opening and soft closing are defined as:

$$f \circ [B, A, k] = (f \ominus [B, A, k]) \oplus [B, A, k] \qquad (2.3.3)$$

$$f \bullet [B, A, k] = (f \oplus [B, A, k]) \ominus [B, A, k] \qquad (2.3.4)$$

respectively. Accordingly, soft open-closing and soft close-opening are realised through cascade connection of the soft opening and closing in different orders. The two operators are defined as:

$$OC(f, [B, A, k]) = (f \circ [B, A, k]) \bullet [B, A, k] \qquad (2.3.5)$$

$$CO(f, [B, A, k]) = (f \bullet [B, A, k]) \circ [B, A, k]. \qquad (2.3.6)$$

2.4 Multi-resolution Decomposition Schemes

The multi-resolution decomposition scheme provides convenient and effective approaches for signal processing [47]. The core idea is to remove high frequencies in a signal, thus to obtain a reduced-scale version of the original signal. Therefore, by repeating this procedure, a collection of coarse signals at different levels are produced. Meanwhile, a collection of detail signals can be constructed by subtracting the coarse signal at level $i+1$ from the coarse signal at level i.

Let $J \subseteq \mathbb{N}$ be an index set [1] indicating the levels in a multi-resolution decomposition scheme; J can be finite or infinite. Therefore, for a domain V_j ($j \in J$), signals of level j belong to V_j. The analysis operator $\psi^{\uparrow}$ decomposes a signal in the direction of increasing, i.e. $\psi_j^{\uparrow} : V_j \rightarrow V_{j+1}$. On the other hand, the synthesis operator $\psi^{\downarrow}$ proceeds in the direction of decreasing, i.e. $\psi_j^{\downarrow} : V_{j+1} \rightarrow V_j$. In other words, the analysis operator $\psi^{\uparrow}$ maps a signal to a higher level and reduces information, while $\psi^{\downarrow}$ maps it to a lower level.

Composing analysis operators successively, a signal at any level i can be transferred to a higher level j. The composed analysis operator is denoted as $\psi_{i,j}^{\uparrow}$ and is defined to be:

$$\psi_{i,j}^{\uparrow} = \psi_{j-1}^{\uparrow}\psi_{j-2}^{\uparrow}\cdots\psi_{i}^{\uparrow}, \quad j > i. \tag{2.4.1}$$

It maps an element in V_i to an element in V_j. It should be noted that here the operations are carried from right to left. Likewise, the composed synthesis operator:

$$\psi_{j,i}^{\downarrow} = \psi_{i}^{\downarrow}\psi_{i+1}^{\downarrow}\cdots\psi_{j-1}^{\downarrow}, \quad j > i \tag{2.4.2}$$

takes the signal back from level j to level i. Finally, the composition operator that takes a signal from level i to level j and then back to level i is defined as:

$$\hat{\psi}_{i,j} = \psi_{j,i}^{\downarrow}\psi_{i,j}^{\uparrow}, \quad j > i \tag{2.4.3}$$

[1] A set I serves as an *index set* for the family $\mathcal{F} = \{A_\alpha\}$ of sets if, for every $\alpha \in I$, there exists a set of A_α in the family $\mathcal{F}$. The index set I can be any set, finite or infinite. Very often, we use the set of nonnegative integers $\{1, 2, 3, \ldots\}$ as an index set [48].

where $\hat{\psi}_{i,j}$ is viewed as an approximation operator because analysis operator $\psi_j^{\uparrow}$ reduces signal information and, in general, the synthesis operator $\psi_j^{\downarrow}$ cannot compensate the information lost in the analysing procedure.

Analysis and synthesis operators play an important role in the construction of a decomposition scheme. If there is no information lost during the analysis and synthesis procedure, then this decomposition scheme would be perfect. The Pyramid transform [49] and morphological wavelets [50] discussed in the following subsections are designed to not modify the decomposition when the analysis and synthesis steps repeat.

2.4.1 Pyramid Transform

Pyramid transform is a multi-resolution decomposition scheme that does not influence the decomposition when analysis and synthesis steps are repeated. A premise of pyramid transform is the so-called pyramid condition [49]. The analysis and synthesis operators $\psi_j^{\uparrow}$ and $\psi_j^{\downarrow}$ are said to satisfy the pyramid condition if $\psi_j^{\uparrow}\psi_j^{\downarrow} = \mathrm{id}$ on V_{j+1}, where id represents the identity operator.

Although analysis operator $\psi_j^{\uparrow}$ is the left inverse of synthesis operator $\psi_j^{\downarrow}$, it is in general not the right inverse of the latter, which means that $\psi_j^{\downarrow}\psi_j^{\uparrow}(x)$ is only an approximation of $x \in V_j$. Denoting the approximation of x by $\hat{x}$, *i.e.* $\hat{x} = \hat{\psi}_{j,j+1}(x) = \psi_j^{\downarrow}\psi_j^{\uparrow}(x) \in \hat{V}_j$, it is assumed that there exists a subtraction operator $(x, \hat{x}) \mapsto x \dot{-} \hat{x}$ mapping $V_j \times \hat{V}_j$ into a set Y_j, as well as an addition operator $(\hat{x}, y) \mapsto \hat{x} \dot{+} y$ mapping $\hat{V}_j \times Y_j$ into V_j, where $y = x \dot{-} \hat{x}$ is the detail signal that contains information about x that is not present in $\hat{x}$.

If a signal x can be reconstructed from its approximation $\hat{x}$ and the detail signal y, *i.e.* $\hat{x} \dot{+} (x \dot{-} \hat{x}) = x$, the signal can be decomposed recursively as:

$$\begin{aligned} x \rightarrow \{y_0, x_1\} &\rightarrow \{y_0, y_1, x_2\} \rightarrow \cdots \\ &\rightarrow \{y_0, y_1, \cdots, y_j, x_{j+1}\} \rightarrow \cdots \end{aligned} \tag{2.4.4}$$

where

$$\begin{cases} x_0 = x \in V_0 \\ x_{j+1} = \psi_j^{\uparrow}(x_j) \in V_{j+1}, \quad j \geq 0. \\ y_j = x_j \dot{-} \psi_j^{\downarrow}(x_{j+1}) \in Y_j \end{cases} \tag{2.4.5}$$

This scheme is called the pyramid transform and the original signal x_0 can be exactly reconstructed from x_{j+1} and $y_j, y_{j-1}, \ldots, y_0$ by means of the backward recursion $x_j = \psi_j^{\downarrow}(x_{j+1}) \dot{+} y_j$.

Morphological pyramid is a special pyramid transform satisfying the following conditions [49]: (a) all domains V_j are complete lattices; and (b) the pair $(\psi_j^{\uparrow}, \psi_j^{\downarrow})$ is an adjunction between V_j and V_{j+1}. To satisfy the pyramid condition, $\psi_j^{\uparrow}$ is an erosion and $\psi_j^{\downarrow}$ is a dilation.

2.4.2 Morphological Wavelet

The morphological wavelet is an improvement over Pyramid transform in that it engages two analysis operators and one synthesis operator. A formal definition of the morphological wavelet is presented as follows. Assume that there exist sets V_j and W_j. V_j is referred to as the signal space at level j and W_j the detail space at level j. The two analysis operators together decompose a signal in the direction of increasing j. The signal analysis operator, $\psi_j^{\uparrow}$, maps a signal from V_j to V_{j+1}, i.e. $\psi_j^{\uparrow} : V_j \rightarrow V_{j+1}$, while the detail analysis operator, $w_j^{\uparrow}$, maps it from V_j to W_{j+1}, i.e. $w_j^{\uparrow} : V_j \rightarrow W_{j+1}$. On the other hand, a synthesis operator proceeds in the direction of decreasing j, denoted as $\Psi_j^{\downarrow} : V_{j+1} \times W_{j+1} \rightarrow V_j$.

In order to yield a complete signal representation, the mappings $(\psi_j^{\uparrow}, \omega_j^{\uparrow}) : V_j \rightarrow V_{j+1} \times W_{j+1}$ and $\Psi_j^{\downarrow} : V_{j+1} \times W_{j+1} \rightarrow V_j$ should be inverses of each other, which means that the following condition:

$$\Psi_j^{\downarrow}(\psi_j^{\uparrow}(x), \omega_j^{\uparrow}(x)) = x, \; x \in V_j \tag{2.4.6}$$

should be fulfilled. This is called the perfect reconstruction condition, and

$$\begin{cases} \psi_j^{\uparrow}(\Psi_j^{\downarrow}(x, y)) = x, & x \in V_{j+1}, y \in W_{j+1} \\ \omega_j^{\uparrow}(\Psi_j^{\downarrow}(x, y)) = y, & x \in V_{j+1}, y \in W_{j+1} \end{cases} \tag{2.4.7}$$

where x is called the approximation signal and y is the detail signal. Therefore, decomposing an input signal $x_0 \in V_0$ with the following recursive analysis scheme:

$$x_0 \rightarrow \{x_1, y_1\} \rightarrow \{x_2, y_2, y_1\} \rightarrow \cdots \rightarrow \{x_j, y_j, y_{j-1}, \cdots, y_1\} \rightarrow \cdots \quad (2.4.8)$$

where

$$x_{j+1} = \psi_j^{\uparrow}(x_j) \in V_{j+1} \quad (2.4.9)$$

$$y_{j+1} = w_j^{\uparrow}(x_j) \in W_{j+1}. \quad (2.4.10)$$

x_0 can be exactly reconstructed from x_j and y_j, y_{j-1}, $\ldots$, y_1 by means of the following recursive synthesis scheme:

$$x_{j-1} = \Psi_{j-1}^{\downarrow}(x_j, y_j). \quad (2.4.11)$$

This signal representation scheme is referred to as the morphological wavelet decomposition scheme and is illustrated in Fig. 2.5.

2.4.3 The Lifting Scheme

The lifting scheme or simple lifting was originally developed as an alternative way to construct the wavelets used in (the first generation) WT, which leads to the so-called second generation WT [51, 52]. The second generation WT not only preserves the good features of the the first generation WT, namely time-frequency localisation and fast algorithms, but can also extend beyond simple geometries. As an alternate implementation of classical WTs, the lifting scheme provides an entirely spatial-domain interpretation of the transform, as opposed to the traditional frequency domain-based constructions. The feature of local spatial interpretation enables us to adapt the transform to the data, thereby introducing non-linearities in the process of multi-scale transforms. Compared with FTs using the same filter all the time and wavelets being translation and dilation of one given function, lifting adapts local data irregularities in the transform process.

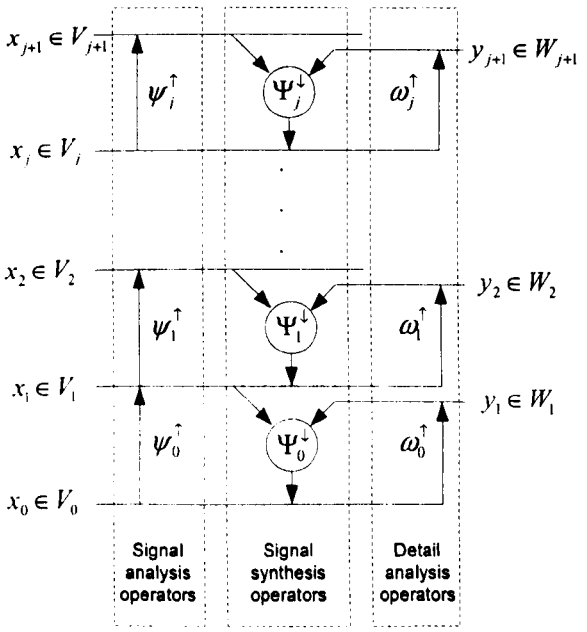


Figure 2.5: Sample stages of the morphological wavelet decomposition scheme

The basic idea of the lifting scheme is very simple: the redundancy is removed by using the correlation in the data. To reach this goal, a signal x_k is first split into two sets - the even set $\{x_k | k \text{ is even}\}$ and the odd set $\{x_k | k \text{ is odd}\}$. Then, the odd set is predicted from the even set. Finally, the prediction error, which is the difference between the odd-indexed sample and its prediction, is used to update the even-indexed data. The three steps are noted as split, predict and update, respectively, as shown in Fig. 2.6.

1. *Split stage:* Let $x(n)$ represent a discrete signal. Firstly, $x(n)$ can be split into even and odd components, $x_e(n)$ and $x_o(n)$, respectively, where

$$\begin{aligned} x_e(n) &= x(2n) \\ x_o(n) &= x(2n+1). \end{aligned} \tag{2.4.12}$$

If the signal $x(n)$ corresponds to the sample of a smooth and slowly varying function, then the components of $x_e(n)$ and $x_o(n)$ are highly

correlated. Therefore, it should be possible to accurately predict each odd polyphase coefficient from the nearby even polyphase coefficients.

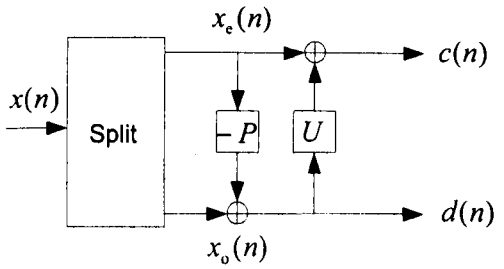


Figure 2.6: The typical lifting steps (analysis): split, predict (P) and update (U). $c(n)$ and $d(n)$ are the approximation and detail signals, respectively

2. *Predict stage:* In the interpolating formulation of lifting, we can predict the odd polyphase coefficients $x_o(n)$ from the neighbouring even coefficients $x_e(n)$. Normally, the predictor for each $x_o(n)$ is a linear combination of neighbouring even coefficients:

$$P(x_e(n)) = \sum_l p_l x_e(n+l) \tag{2.4.13}$$

where the value of l lies on how many even coefficients are used for the prediction. If N ($N = 2D$) coefficients are used in symmetry, then:

$$-D+1 \leq l \leq D. \tag{2.4.14}$$

The prediction calculation is conducted within a data window of a fixed length. For each $x_o(n)$, its neighbouring even coefficients at both the left and right sides are involved in the calculation. As the window slides through the signal, the prediction for all odd coefficients is performed. If $n+l$ exceeds the range of the signal, the signal will be extended with zero samples, and this process is called zero padding.

A new representation of $x(n)$ may be obtained by replacing $x_o(n)$ with the prediction residual, $d(n)$:

$$d(n) = x_o(n) - P(x_e(n)). \tag{2.4.15}$$

The prediction residual will be small if the underlying signal is locally smooth. Furthermore, the same information of the original signal $x(n)$ is preserved in the residuals $d(n)$ since the odd polyphase coefficients $x_o(n)$ can be recovered by noting that

$$x_o(n) = d(n) + P(x_e(n)). \tag{2.4.16}$$

Actually, the prediction procedure is equivalent to applying a high pass filter to the original signal $x(n)$.

3. *Update stage:* The last lifting step transforms the even polyphase coefficients, $x_e(n)$, into a low pass filtered and sub-sampled version of $x(n)$ by updating $x_e(n)$ with a linear combination of the prediction residuals, $d(n)$, where $x_e(n)$ is replaced with

$$c(n) = x_e(n) + U(d(n)) \tag{2.4.17}$$

and normally $U(d)$ is a linear combination of the values of d:

$$U(d(n)) = \sum_l u_l d(n+l). \tag{2.4.18}$$

No information is lost in the update step. Assuming the same U is chosen for synthesis, given $d(n)$ and $c(n)$, we have

$$x_e(n) = c(n) - U(d(n)). \tag{2.4.19}$$

The inverse lifting stage is shown in Fig. 2.7. It corresponds to a critically sampled perfect reconstruction filter bank with c and d at half a rate.

2.5 Multiscale Morphological Operators

Scale space is an important concept used in image processing and pattern recognition [53]. It aims at viewing and analysing an image, or more generally a signal, at different scales by a smoothing operation. The advantages of describing a signal in such a way has been expatiated in depth in [54, 55, 56].

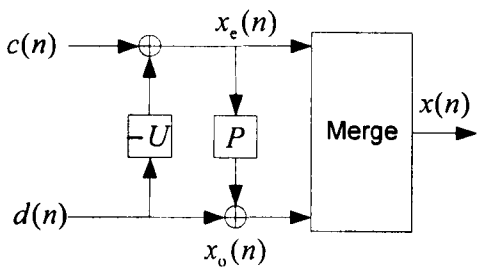


Figure 2.7: The typical inverse lifting steps (synthesis): update (U), predict (P) and merge

It makes the choice at what scale visual observations are to be made explicit; otherwise, the choice for an observation scale was often hidden somewhere in the definition of the operators. By embedding the stack of descriptions in a mathematical space called the scale space, the information obtained at one scale can be related to that at another scale. Therefore, the features of the signal can be tracked from scale to scale.

Multiscale morphological operators involve a finite set of SEs of different sizes. Let g be an SE as defined above. The function

$$ng = \underbrace{g \oplus g \oplus \ldots \oplus g}_{(n-1) \text{ times}} \tag{2.5.1}$$

defines a function pattern of size $n = 0, 1, \ldots$ [57]. Hence, taking the distributivity into consideration, multiscale erosion and multiscale dilation are defined as:

$$f \ominus ng = f \underbrace{\ominus g \ominus \ldots \ominus g}_{n \text{ times}} \tag{2.5.2}$$

$$f \oplus ng = f \underbrace{\oplus g \oplus \ldots \oplus g}_{n \text{ times}} \tag{2.5.3}$$

respectively. When $n = 0$, $f \ominus ng = f$ and $f \oplus ng = f$. Consequently, multiscale opening and multiscale closing are defined as:

$$f \circ ng = (f \ominus ng) \oplus ng \tag{2.5.4}$$

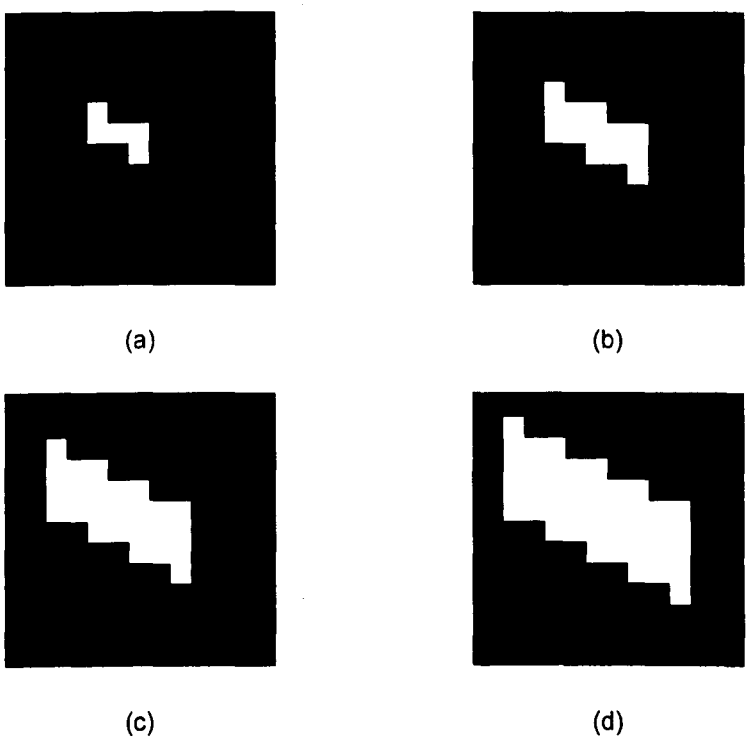


Figure 2.8: An example of a set of binary multiscale SEs. (a) B. (b) $B \oplus B$. (c) $3B$. (d) $4B$

$$
\begin{aligned}
& = (f \underbrace{\bigcirc g \bigcirc \ldots \bigcirc g}_{n \text{ times}}) \underbrace{\bigoplus g \bigoplus \ldots \bigoplus g}_{n \text{ times}} \\
f \bullet ng & = (f \ominus ng) \ominus ng \qquad (2.5.5) \\
& = (f \underbrace{\oplus g \oplus \ldots \oplus g}_{n \text{ times}}) \underbrace{\ominus g \ominus \ldots \ominus g}_{n \text{ times}}
\end{aligned}
$$

respectively. An example of a set of binary multiscale SEs is given in Fig. 2.8.

2.6 Conclusion

This chapter studies the mathematical background of MM and reviews the development of the basic morphological operators. It also gives brief proofs for some important properties of the operators. Basic morphological operators

include dilation and erosion, which can be defined in both binary and grey scale cases. This thesis deals with the grey scale version only. A set of alternating sequential filters can be developed from the cascade connection of dilation and erosion. The simplest ones are opening and closing, which have desirable properties for filtering.

A set of more complex morphological operators evolving from the basic ones are summarised afterwards, including soft MM, morphological wavelet and multiscale morphological operators. Soft MM replaces maximisation and minimisation by an arbitrary order, which makes MM more flexible. Morphological wavelet is a special case of multi-resolution decomposition scheme. It can be viewed as the non-linear version of the WT by engaging morphological operation in the analysis and synthesis procedure. As its name implies, multiscale morphological operators process a signal with a range of SEs of different sizes and the information extracted at each scale are different from one another. In such a manner, the signal can be represented according to its shape and size content and geometrical structure. The advanced morphological operators proposed in the following chapters are developed based on the operators introduced in this chapter.

Chapter 3

Morphological Decomposition Algorithms

Multi-resolution decomposition scheme is considered as a powerful tool for signal processing. This chapter proposes a multi-resolution morphological filter based on morphological wavelet. In addition, as an improvement of the multi-resolution morphological filter, a morphological lifting filter has been presented using the lifting scheme. Applying the proposed filters to ECG signal analysis, which requires the removal of impulsive noise and the detection of characteristic waveforms, the results are very satisfactory.

This chapter also presents another morphological decomposition algorithm - morphological gradient wavelet (MGW). It is a nonlinear WT that involves morphological operation in the decomposition process and presents the gradient information in the output. MGW is applied to power systems to detect disturbances.

3.1 Multi-resolution Morphological Filter

A multi-resolution morphological filter can be constructed through the combination of a morphological filter and a multi-resolution decomposition scheme, where the filter also serves as the decomposition operator.

3.1.1 Algorithm Description

Open-closing and close-opening are two commonly used filters derived from opening and closing, defined as:

$$\mathrm{OC}(f) = f \circ g \bullet g \tag{3.1.1}$$

$$\mathrm{CO}(f) = f \bullet g \circ g \tag{3.1.2}$$

respectively. Through the cascade connection of opening and closing, both the positive and negative noise can be eliminated using the same SE. Moreover, to weaken the statistical deviation caused by the in open-closing and close-opening filters, the two filters are combined together to form a new one:

$$\vartheta(f) = [\mathrm{OC}(f) + \mathrm{CO}(f)]/2. \tag{3.1.3}$$

In this filter, the SE functions as a moving window through the signal to extract the high-frequency components. If the length of the SE is longer than the width of the noise in a segment of the signal, the noise can be eliminated. An appropriate SE is determined by trial and error so that the noise can be eliminated.

According to [58], the filter should satisfy the following two conditions: (a) They must be smoothing functions, that is, they should generate a lower resolution approximation of the input. (b) They must not add a DC level to the input. However, most morphological operators do not fulfill these conditions simultaneously. For example, opening and closing have property (a), but do not have property (b), because $f \circ g \leq f$ and $f \bullet g \geq f$ [2]. The same applies to their compositions $\mathrm{OC}(f)$ and $\mathrm{CO}(f)$. Fortunately, the average of those operations tends to generate an output for which the DC level is very close to that of the input. Thus, the ϑ with a flat SE is employed here for it suffices condition (a) and approximately suffices (b).

Hybridising the above ϑ filter and the morphological wavelet scheme, a multi-resolution morphological filter is constructed with the analysis operators $\psi_j^{\uparrow}$ and $w_j^{\uparrow}$:

$$\psi_j^{\uparrow}(x_j) = x_{j+1} = \vartheta(x_j) \tag{3.1.4}$$

$$w_j^{\uparrow}(x_j) = y_{j+1} = \mathrm{id} - \vartheta(x_j) \tag{3.1.5}$$

$$\vartheta(x_j) = (\mathrm{OC}(x_j) + \mathrm{CO}(x_j))/2 \tag{3.1.6}$$

where $x_j \in V_j$, $x_{j+1} \in V_{j+1}$, $y_{j+1} \in W_{j+1}$, and id denotes the identity transform [20]:

$$\forall f, \mathrm{id}(f) = f.$$

The size of the SE increases as the level goes up. It should be noted that in the application of filtering, the synthesis operator $\Psi_j^{\downarrow}$ is unnecessary.

3.1.2 Simulation Studies on ECG Signal Analysis

A sinusoidal signal polluted by white noise is selected here to test the performance of the proposed filter. The signal is filtered 4 times and each time by a flat SE with the length of $j+1$ ($j = 1, 2, 3, 4$). The results are shown in Fig. 3.1, which includes the approximation signals x_j and the detail signals y_j at each level j. In order to compare the improvements, the signal-to-noise ratio (SNR) is used to evaluate the amount of noise remaining in each level of the signal. The value of SNR is defined as:

$$\mathrm{SNR} = 10 \times \log_{10} \frac{S_\sigma}{N_\sigma} \tag{3.1.7}$$

where S is a noise-free signal and N is the noise. For a given signal X, X_σ is defined as:

$$X_\sigma = \sum_{l=0}^{L-1} (X(l) - \mu_X)^2 \tag{3.1.8}$$

where μ_X and L are the mean and length of X, respectively.

In order to evaluate the performance of the multi-resolution morphological filter, the values of SNR of the original signal and the decomposed signals are calculated and shown in Table 3.1. And so is the computation time required in each decomposition phase. As it can be seen from the table, SNR increases as the level ascends, but the increment is descending. At level 4, the improvement is not evident, which implies that it is enough to perform the decomposition three times. Most computation time consumed by the decomposition is due

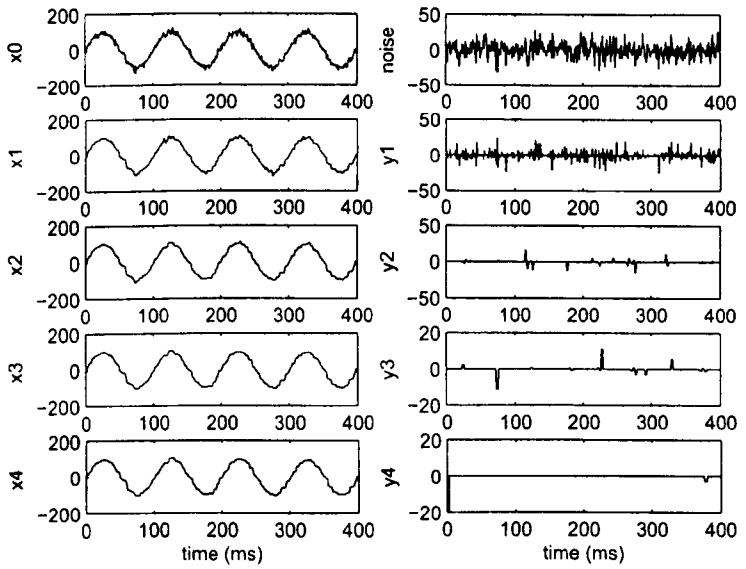


Figure 3.1: Performance of multi-resolution morphological filter on a sinusoidal signal

to the ranking operation involved in (3.1.6). As the number of ranking each decomposition takes depends highly on the signal it processes, it is impossible to estimate the computation expense due to the complex operation applied to the specific signal. Hence, we record the average computation time of 30 runs taken by (3.1.6) at each level, respectively. All the simulation studies are carried out on a DELL PC with a 2.80 GHz Intel Pentium D CPU and 1.99 GB of RAM using MATLAB 7.2.0.232. As the table indicates, as the level increases, it takes more time to get the signal filtered. This reflects the fact that a longer SE is involved as the level increases. The computation time is also used as an index to compare the performance of multi-resolution morphological filter with that of the morphological lifting filter, which is introduced in the next section.

As a practical application, the multi-resolution morphological filter is used to remove the impulsive noise of ECG signals. Noise suppression is typically the first step in ECG signal processing. It mainly deals with the elimination of the impulsive noise caused by muscle activities and power line interference

Table 3.1: Performance of multi-resolution morphological filter: SNR of the original signal and the decomposed signals and computation time at each level

Signal	SNR (dB)	Time (s)
x_0	33.6074	/
x_1	41.0862	0.0952
x_2	42.9527	0.0981
x_3	43.0812	0.1008
x_4	43.1734	0.1040

[59], meanwhile preserves the shape of the original signal. Previous studies in [60] used the filter of (3.1.3), however, in that case the length of the SE must be determined in advance.

An ECG signal contaminated by impulsive noise is shown in Fig. 3.2(a). Since this time, the noise is not added artificially, it is impossible to calculate the SNR of the filtered signal at each level. However, intuitively, the output signal is smoother and the shape of the waveform is well preserved, although part of the R wave has also been treated as noise and been eliminated because it is too narrow.

Increasing the length of the SE used at each level, the multi-resolution morphological filter can be used to extract the feature waveforms according to their width. Noticing that the QRS complexes are short in width and the R waves have a high amplitude, the first step is to determine the location of the peak of the R waves, which is also the location of the QRS complexes. Then, the multi-resolution morphological filter is used to decompose the signal in a window that covers the QRS complexes. The length of the SE increases from 3 at a step of 2. The decomposition procedure ends automatically until there is no evident decrease in the energy carried by the approximation signal. Therefore, the QRS complexes are extracted by subtracting the filtered signal from the original signal. In this manner, the only *a priori* knowledge required is the length of the window that covers the QRS complexes, which can be obtained from observation or experience.

Similar processes are applied to detect the P waves and T waves. The ap-

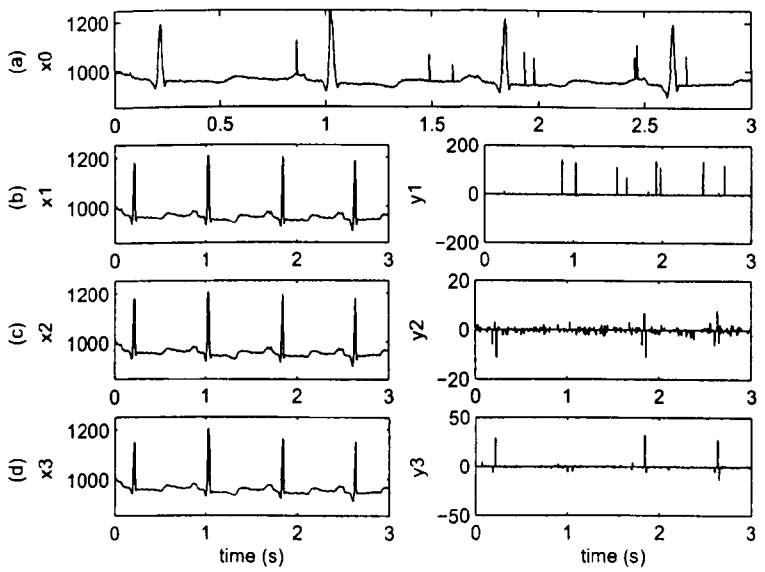


Figure 3.2: ECG noise removal using multi-resolution morphological filter

proach takes the advantage that the P wave comes before the QRS complex and the T wave comes after it. Since the duration of the former is usually shorter than that of the latter, P wave is the second to be extracted. The decomposition procedure is carried out based on the signal with QRS complexes eliminated. Likewise, when extracting the T wave, the decomposition procedure is carried out on the signal with both QRS complexes and P waves eliminated. As for the baseline wander, a lowpass filter is applied to the residual signal and the output is the estimated baseline.

Figure 3.3 shows an example of feature waveform extraction. The original ECG signal is shown in Fig. 3.3(a). Figures 3.3(b)~(d) show the results of the extracted QRS complexes, P waves and T waves. Applying a lowpass filter to the residual signal, the estimated baseline is shown in Fig. 3.3(e). Finally, subtracting the estimated baseline from the original signal, the normalised signal is given in Fig. 3.3(f). As it can be seen from these figures, there is no miss-extraction and the detected waveforms are identical to the original.

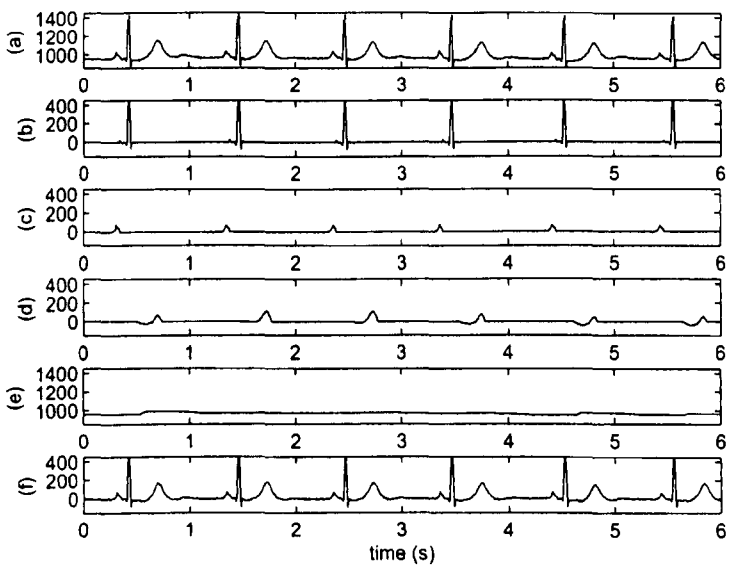


Figure 3.3: Feature waveforms extraction using multi-resolution morphological operators. (a) The original signal. (b) The extracted QRS complexes. (c) The extracted P waves. (d) The extracted T waves. (e) The extracted baseline. (f) Normalised output signal (the baseline wander is eliminated)

3.2 Morphological Lifting Filter

As discussed in the previous section, by using a flat SE, a multi-resolution morphological filter simply replaces the curves with a series of flat lines, thus causes the loss of some features in the approximation waveforms. Besides, as the length of the SE increases, the filtered signal becomes more distorted. Therefore, a compensation process is needed for the recovery of the waveform.

3.2.1 Algorithm Introduction

Based on the lifting scheme introduced in section 2.4.3, the signal analysis operator $\psi^{\uparrow}$ and the detail analysis operator $w^{\uparrow}$ can be modified respectively to improve the quality of the filter mentioned in the previous section. At level j, the information in the detail signal y_j is used to modify the decomposed signal x_j to x'_j. After that, y_j is updated according to the modified decomposed signal x'_j. The procedure is described as:

$$x'_j = x_j - \lambda(y_j) \tag{3.2.1}$$

$$y'_j = y_j - \pi(x'_j) \tag{3.2.2}$$

where λ and π are operators mapping from W_j to V_j and from V_j to W_j, respectively. In the update step, the samples of the detail signal y_j are divided into two groups, zero and non-zero, according to their values. The non-zero samples correspond to the points needing update. These samples can be replaced by a linear, cubic or polynomial interpolation. For example, at level j, if a sample $y_{j,l}$ of the detail signal has a non-zero value, it can be updated by a certain interpolation using

$$y_{j,l-2}, y_{j,l+2} \Rightarrow y_{j,l} \tag{3.2.3}$$

or

$$y_{j,l-4}, y_{j,l-2}, y_{j,l+2}, y_{j,l+4} \Rightarrow y_{j,l}. \tag{3.2.4}$$

In general, if n samples in a segment of a signal are updated, a polynomial p of degree $n-1$ is constructed. The value at point l can be calculated from

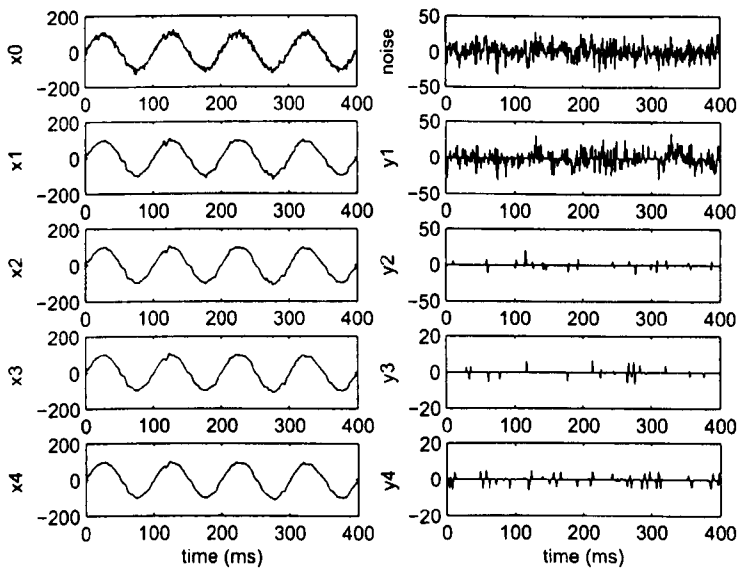


Figure 3.4: Performance of morphological lifting filter on a sinusoidal signal

$y_{j,l} = p(x_{j,l+k})$, where l is a non-zero point in the detail signal y_j and $l + k$ are points at which y_j has zero values.

3.2.2 Tests on ECG Signals

To evaluate the effectiveness of the morphological lifting filter, the same voltage waveform is used as a reference. The signal is also filtered 4 times and each time the SE is the same as that used in the previous study. The interpolation method is linear interpolation. The results are shown in Fig. 3.4, illustrating that concomitant with noise removal, the sinusoidal features of the voltage waveform are well kept and the output signals are smoother than those in Fig. 3.1. The performance is also demonstrated quantitatively by SNR and computation time, as shown in Table 3.2. In comparison with the results shown in Table 3.1, the values of SNR are increased, indicating that the performance of the filter has been improved as the updating process applied. Due to the same reason, the decomposition procedures are more time-consuming.

Applying the morphological lifting filter to impulsive noise filtering of ECG

Table 3.2: Performance of morphological lifting filter: SNR of the original signal and the decomposed signals and computation time at each level

Signal	SNR (dB)	Time (s)
x_0	33.6074	/
x_1	41.1492	0.2015
x_2	43.8999	0.1118
x_3	44.8389	0.1111
x_4	48.0149	0.1232

signals, one of the simulation results is given in Fig. 3.5. The input signal is the same as the one in Fig. 3.2(a). Comparing the two figures, it can be seen that detail signals obtained by the morphological lifting filter are cleaner, including impulsive noise and only a little amount of details. It shows that the morphological lifting filter preserves the details of the signal while rejecting the impulsive noise.

3.3 Morphological Gradient Wavelet

As stated in section 2.4.2, morphological wavelets engage morphological operation in decomposition and reconstruction process. To make the framework of morphological wavelet more comprehensible, the morphological Haar wavelet (MHW), proposed in [50], is introduced here. MHW is different from the classical linear Haar wavelet in that the linear signal analysis filter of classical Haar wavelet is replaced by a morphological operator, *e.g.* taking the minimum or maximum over two samples. The analysis and synthesis operators are defined as:

$$\psi^{\uparrow}(x)(n) = x(2n) \wedge x(2n+1),\ x \in V_j \tag{3.3.1}$$

$$\omega^{\uparrow}(x)(n) = x(2n) - x(2n+1),\ x \in V_j \tag{3.3.2}$$

$$\begin{aligned} \psi^{\downarrow}(x)(2n) &= x(n),\ x \in V_{j+1} \\ \psi^{\downarrow}(x)(2n+1) &= x(n),\ x \in V_{j+1} \end{aligned} \tag{3.3.3}$$

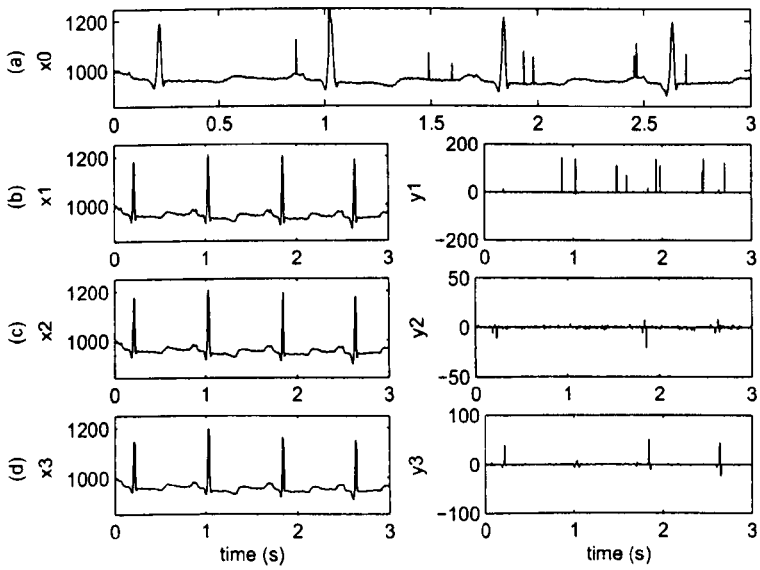


Figure 3.5: ECG noise removal using morphological lifting filter

$$\begin{aligned} \omega^{\downarrow}(y)(2n) &= y(n) \vee 0, \ y \in W_{j+1} \\ \omega^{\downarrow}(y)(2n+1) &= -(y(n) \wedge 0), \ y \in W_{j+1} \end{aligned} \tag{3.3.4}$$

$$\Psi^{\downarrow}(x,y)(n) = \psi^{\downarrow}(x)(n) + \omega^{\downarrow}(y)(n), \ x \in V_{j+1}, y \in W_{j+1} \tag{3.3.5}$$

where n is the sample index, '$\wedge$' denotes the minimisation operation and '$\vee$' denotes the maximisation operation.

3.3.1 Algorithm Design

In order to take into consideration the gradient of each sample of the signal, this thesis proposes MGW. Derived from [45], a morphological median operator termed by $\overset{m}{\oplus}$ is defined as:

$$(f \overset{m}{\oplus} g)(x) = \text{median}\{f(x+s) + g(s) | (x+s) \in \mathcal{D}_f, s \in \mathcal{D}_g\} \tag{3.3.6}$$

where f is the signal and g is the structuring element (SE), $\mathcal{D}_f$ and $\mathcal{D}_g$ represent the field of definition of f and g respectively. Apparently, instead of

applying local maximum and minimum operations as dilation and erosion do, morphological median operator returns the median value.

If the morphological median operator employs a flat SE whose length is 3, (3.3.6) becomes:

$$(f \overset{\text{m}}{\oplus} g)(x) = \text{median}\{f(x-1), f(x), f(x+1)\}. \tag{3.3.7}$$

In MGW, the above equation is selected as the signal analysis operator, which is rewritten in (3.3.8). Consequently, the detail analysis operator of MGW is defined as depicted in (3.3.9).

$$\begin{aligned} \psi^{\uparrow}(x_j)(n) &= \text{median}\{x_j(2n-1), x_j(2n), x_j(2n+1)\} \\ &= x_{j+1}(n) \qquad (3.3.8) \\ \omega^{\uparrow}(x_j)(n) &= x_j(2n+1) - 2x_{j+1}(n) + x_j(2n-1) \\ &= y_{j+1}(n). \qquad (3.3.9) \end{aligned}$$

In order to construct the synthesis operator, two intermediate variables are needed as follows:

$$t_{\text{R},j+1}(n) = x_j(2n+1) - x_j(2n) \tag{3.3.10}$$

$$t_{\text{L},j+1}(n) = x_j(2n) - x_j(2n-1) \tag{3.3.11}$$

where $t_{\text{R},j+1}(n)$ and $t_{\text{L},j+1}(n)$ denote the right and left gradient of $x_j(2n)$ respectively. The synthesis operator is shown in Table 3.3.

Figure 3.6 shows the performance of MHW and MGW on the same signal as used in [50]. To avoid misunderstanding, approximation analysis operator and detail analysis operator of MHW are denoted by $\psi_{\text{H}}^{\uparrow}$ and $\omega_{\text{H}}^{\uparrow}$ respectively, while those of MGW are denoted by $\psi_{\text{G}}^{\uparrow}$ and $\omega_{\text{G}}^{\uparrow}$ respectively. As it can be seen, the detail signal obtained by $\omega_{\text{H}}^{\uparrow}$ is zero at points where the input signal is constant, while that obtained by $\omega_{\text{G}}^{\uparrow}$ is zero at points where the input signal is linear. Moreover, MGW not only detects every point where a change of gradient occurs, but also indicates the change quantitatively. These features make MGW really suitable for the detection of sudden changes in gradient, as explicitly revealed in Fig. 3.6(e). On the contrary, MHW is not as applicable as MGW.

Table 3.3: The synthesis operator $\Psi_j^{\downarrow}$

if $t_{\mathrm{R},j+1} \cdot t_{\mathrm{L},j+1} \geq 0$
  $\Psi_j^{\downarrow}(2n) = x_{j+1}(n)$
  $\Psi_j^{\downarrow}(2n+1) = x_{j+1}(n) + \frac{1}{2}(y_{j+1}(n) + t_{\mathrm{R,j+1}}(n) + t_{\mathrm{L,j+1}}(n))$
else
  if $|t_{\mathrm{R,j+1}}(n)| \geq |t_{\mathrm{L,j+1}}(n)|$
    $\Psi_j^{\downarrow}(2n) = x_{j+1}(n) + t_{\mathrm{L},j+1}(n)$
    $\Psi_j^{\downarrow}(2n+1) = x_{j+1}(n) + y_{j+1}(n)$
  else
    $\Psi_j^{\downarrow}(2n) = x_{j+1}(n) + y_{j+1}(n) + t_{\mathrm{L},j+1}(n)$
    $\Psi_j^{\downarrow}(2n+1) = x_{j+1}(n)$
  end if
end if

3.3.2 Application to Power System Disturbance Detection

In this section, MGW is applied to disturbance detection of power system signals. The aim is to extract the features of the disturbances and to detect their location and duration. Five types of power disturbances are investigated, including impulsive transients, oscillatory transients, momentary interruption, voltage swell and periodical notching. For the purpose of comparison, a widely used wavelet scheme, Daubechies DB4 WT (DB4), is employed to verify the superiority of MGW. The Daubechies wavelets are the most used for the detection, localisation and classification of disturbances, and among them, DB4 is the shortest and therefore, is more localised in time than others Daubechies wavelets except DB1, which is the same as the Haar wavelet. It is reported in [61] that DB4 allows accurate time location of transient components, has much simpler implementation than other Daubechies wavelets, and has excellent performance in disturbance detection.

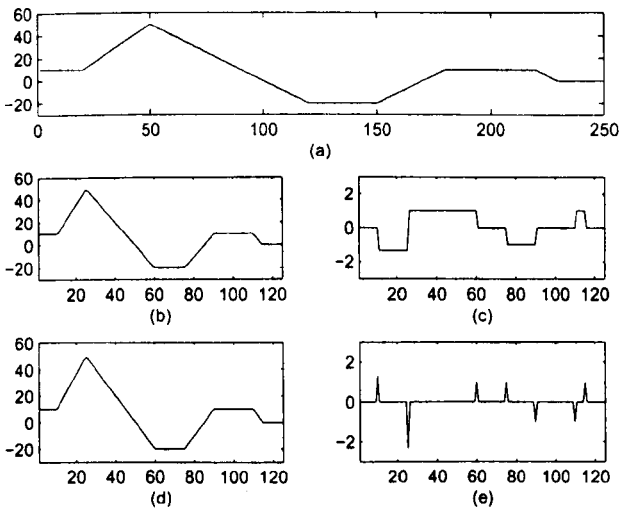


Figure 3.6: Performance of MHW and MGW with analysis operators $\psi_{\mathrm{H}}^{\uparrow}, \omega_{\mathrm{H}}^{\uparrow}$ and $\psi_{\mathrm{G}}^{\uparrow}, \omega_{\mathrm{G}}^{\uparrow}$. (a) Original signal. (b) Approximation signal obtained by $\psi_{\mathrm{H}}^{\uparrow}$. (c) Detail signal obtained by $\omega_{\mathrm{H}}^{\uparrow}$. (d) Approximation signal obtained by $\psi_{\mathrm{G}}^{\uparrow}$. (e) Detail signal obtained by $\omega_{\mathrm{G}}^{\uparrow}$.

Criterion

Before carrying out the simulation studies, the criterion used to evaluate the performance should be introduced. Instead of observing the changes in detail signals visually, a criterion that can quantify the output is needed. Meanwhile, since gradient operators are sensitive to noise, the effect of noise ought to be weakened. Based on these two matters, the criterion of entropy [62] is utilised here. The entropy E of a signal is introduced as:

$$E(s) = \sum_i E(s_i) \tag{3.3.12}$$

where s is the signal and s_i is the coefficients of s in the i^{th} orthonormal basis. There are different entropy criteria. A most basic one is Shannon entropy, which is defined as:

$$E(s) = -\sum_i s_i^2 \log(s_i^2). \tag{3.3.13}$$

In the following simulation studies, the entropy is calculated within the window of the SE. Therefore, the disturbances in a signal, which correspond to sudden

changes in energy, are magnified in the entropy signal. At the same time, since the small oscillations of the detail signal caused by noise share similar energy, their influence is greatly suppressed in the entropy signal.

In order to evaluate the accuracy of detection results, an index of *degree of match* is calculated to describe the similarity between the detected disturbance and the actual disturbance. The degree of match focuses on two properties of a sample interval, location and duration. Location refers to the centre of the sample interval and duration refers to the length. The definition of degree of match is presented hereafter. Let a and b be the beginning and the end of the actual disturbance respectively, and the beginning and the end of the detected disturbance are termed a' and b' respectively. Degree of match, σ, is hereby given as follows.

$$\sigma_1 = \min\left\{\frac{l}{l'}, \frac{l'}{l}\right\} \tag{3.3.14}$$

$$\sigma_2 = \frac{(l+l')/2 - |c-c'|}{(l+l')/2} \tag{3.3.15}$$

$$\sigma = \sigma_1 \cdot \sigma_2 \times 100\% \tag{3.3.16}$$

where $c = (a+b)/2$ and $c' = (a'+b')/2$ are the centre of actual disturbance and detected disturbance respectively; $l = b - a + 1$ and $l' = b' - a' + 1$ are the length of them respectively; and if $\sigma_2 < 0$, σ_2 is reset to 0.

Impulsive transients

Figures 3.7(a) and 3.8(a) illustrate a typical impulsive transient. Note also that a Gaussian white noise is added to the signal in the simulation studies. The detail signals obtained by MGW and DB4 are shown in Figs. 3.7(b) and 3.8(b), respectively. The detail signal obtained by MGW is at level 1 of the decomposition process, and the length of the approximation signal and detail signal is half of the length of the original signal, which are 400 and 800 samples respectively in this case. As for DB4, an interpolation process is provided by MATLAB so that the approximation and detail signals are of the same length as the original signal whatever level they are at. As it can be seen

from Figs. 3.7(b) and 3.8(b), the transient results in two high impulses in the detail signal, which are visually obvious. On the other hand, it is difficult to quantitatively determine the location and duration of the disturbance based on the detail signals, because the noise also causes fairly high impulses. However, when referring to the entropy in Figs. 3.7(c) and 3.8(c), it can be seen that impulses caused by the transient are magnified and those caused by the noise are compressed.

Comparing these two schemes, the former is more capable of detecting the disturbance, which can be concluded from both the detail signals and the entropy signals. The waveform of entropy in Fig. 3.8(c) contains too many high-frequency components, thus it is not suitable for indicating the location and duration of the transient. In contrast, the beginning and end of the transient can be easily obtained from the position of the local minimum of Fig. 3.7(c). The detection results are shown in Table 3.4, from which it can be seen that MGW precisely detects the location of the transient, with the degree of match of 100%; whereas DB4 gives an incorrect result, with the degree of match of only 41.85%.

As stated previously, the detail signal obtained by MGW is at level 1, while that obtained by DB4 is at level 2. Hence, in this case, MGW is more efficient in extracting the features of the disturbance. A comparison of computation time is given in Table 3.5, which demonstrates the superiority of MGW in computation expense. The computation time refers to the time taken by the MGW or the DB4 to process the signal at level 1.

Oscillatory transients

A simulation signal with an oscillatory transient occurring during the period of a sinusoidal wave is demonstrated in Figs. 3.9(a) and 3.10(a). The detail signals are shown in Figs. 3.9(b) and 3.10(b), respectively, from which it can be found that the location of the peaks corresponds to the time scale when the oscillatory transient occurs. The entropy signals are shown in Figs. 3.9(c) and 3.10(c), respectively. In this case, the location and duration of the

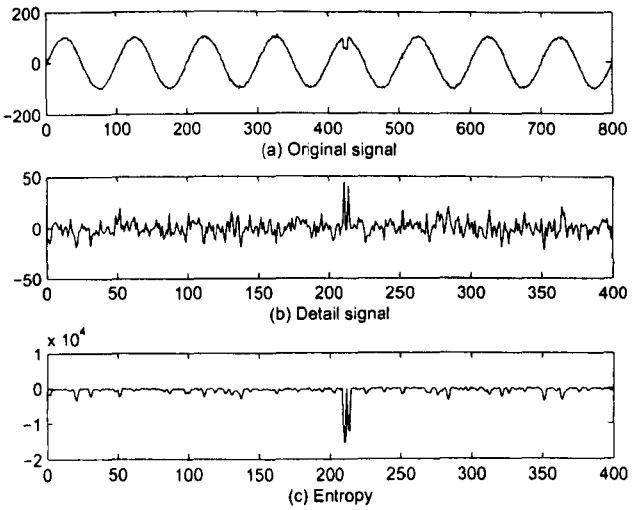


Figure 3.7: Detection of an impulsive transient using MGW.

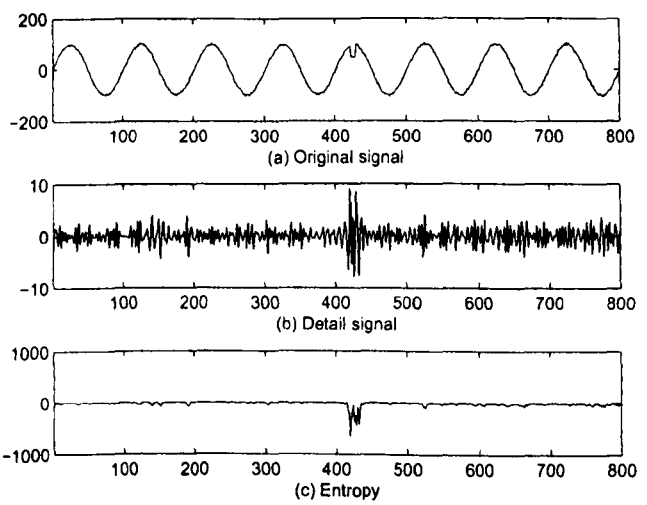


Figure 3.8: Detection of an impulsive transient using DB4.

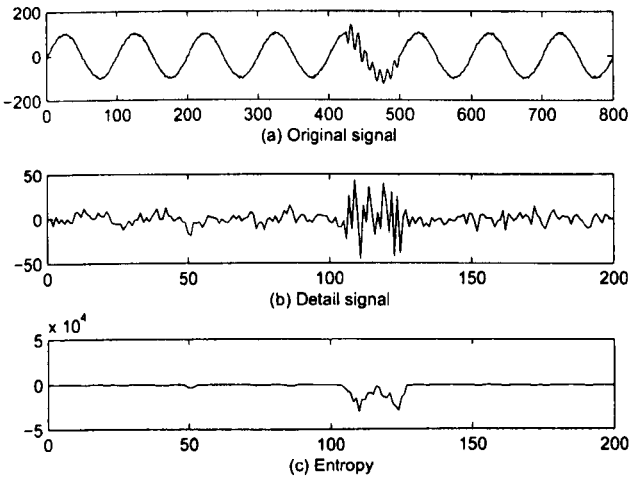


Figure 3.9: Detection of oscillatory transients using MGW.

disturbance can be estimated by recording the locations of a series of points with distinct large magnitude values, which is the same technique as used in periodic notch detection. For the Fourier transform-based methods, such a disturbance may not be easy to detect. Nonetheless, the problem is solved by MGW and DB4 thanks to their outstanding extraction capability. As far as the detection results are concerned, MGW gains a more veracious result, although neither scheme reaches the 100% accuracy. In Table 3.5, the comparison of the computation cost indicates that MGW is more efficient than DB4. More oscillatory transients have been simulated, by varying the oscillatory frequency, amplitude and damping exponent respectively, and similar results are obtained.

Interruption

Figures 3.11(a) and 3.12(a) describe a momentary interruption beginning from the third cycle of the signal and lasting 4 cycles. The detail signals are depicted in Figs. 3.11(b) and 3.12(b), respectively, and the entropy signals in 3.11(c) and 3.12(c), respectively. The two peaks contained in Fig. 3.11(c) clearly indicate the start and end of the occurrence of the disturbance. On the

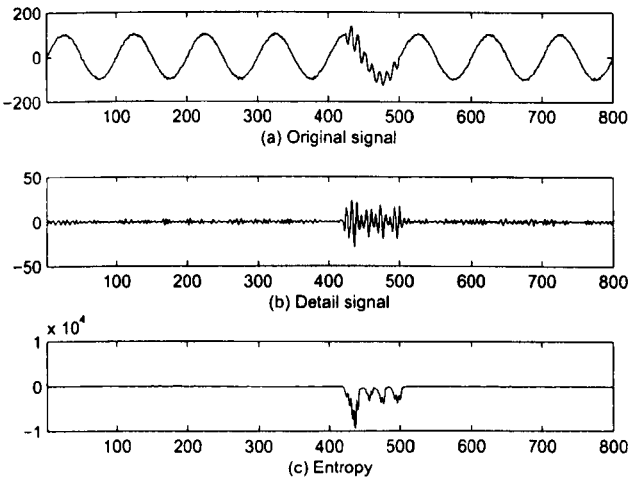


Figure 3.10: Detection of oscillatory transients using DB4.

contrary, the entropy in Fig. 3.12(c) suffers from the drawback that it contains too many high-frequency components, which makes it unable to determine the exact point where the interruption begins or ends. Comparisons of veracity and computation expense between the two methods are shown in Table 3.4 and Table 3.5 respectively, from which it can be concluded that MGW is more accurate and faster than DB4.

Voltage swell

In Figs. 3.13(a) and 3.14(a), a voltage swell lasting from the third to the sixth cycle in a fundamental sinusoidal waveform is presented. Similarly, the detail signals and the entropy signals are shown in Figs. 3.13(b)(c) and 3.14(b)(c), respectively. As it can be seen, several impulses with a similar amplitude are presented in Fig. 3.14(c), and it is impossible to determine which one is resulted from the swell, which means that DB4 is not suitable for detecting the beginning and the end of the voltage swell. In contrast, two entropy impulses are clearly seen in Fig. 3.13(c), and they occur exactly at the position where the voltage swell starts and ends. The detection results of the

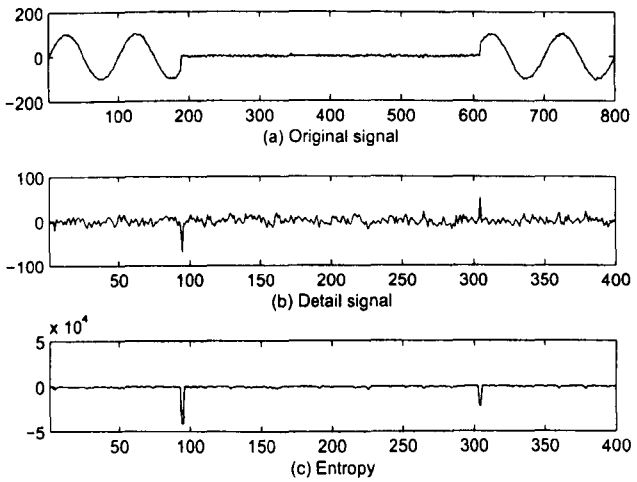


Figure 3.11: Detection of momentary interruption using MGW.

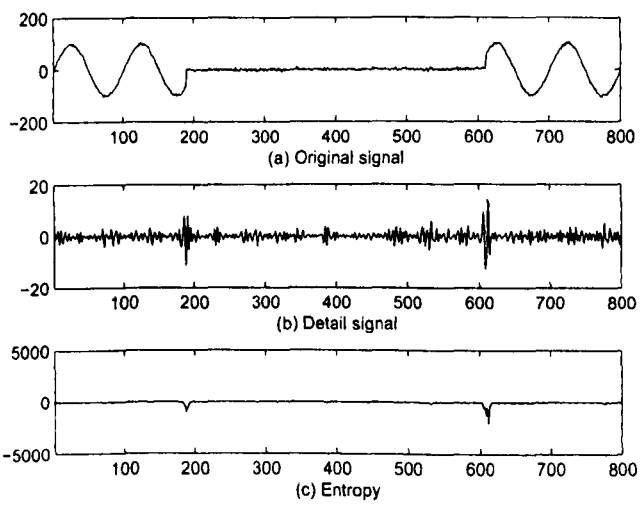


Figure 3.12: Detection of momentary interruption using DB4.

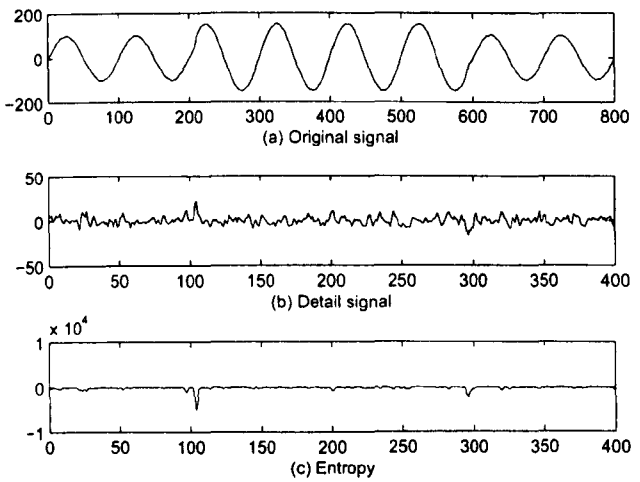


Figure 3.13: Detection of voltage swell using MGW.

two algorithms are demonstrated in Table 3.4.

Notching

A fundamental waveform affected by a periodical sinusoidal notching from the third to the sixth cycle is shown in Figs. 3.15(a) and 3.16(a), and the detail signals are given in Figs. 3.15(b) and 3.16(b), respectively. As it can be seen, the position and duration of the notches are clearly revealed in the detail signals. This time for MGW, the detail signal is gained at level 2, however, the computation efficiency is not greatly affected. The computation time of MGW is still an order of magnitude shorter than that of DB4, as shown in Table 3.5. Figure 3.15(c) illustrates that the disturbance duration can be estimated based on a series of points that have distinct large magnitude values. But the signal shown in Fig. 3.16(c) is not smooth enough to find a proper threshold to distinguish the energy upsurge. Thus, large errors may occur in the detection result. For MGW, because the detail signal is obtained at level 2, the detection result at level 2 should be multiplied by 4 to compare with the actual position. Hence, the detection error, if any, will be magnified. The quantitative results

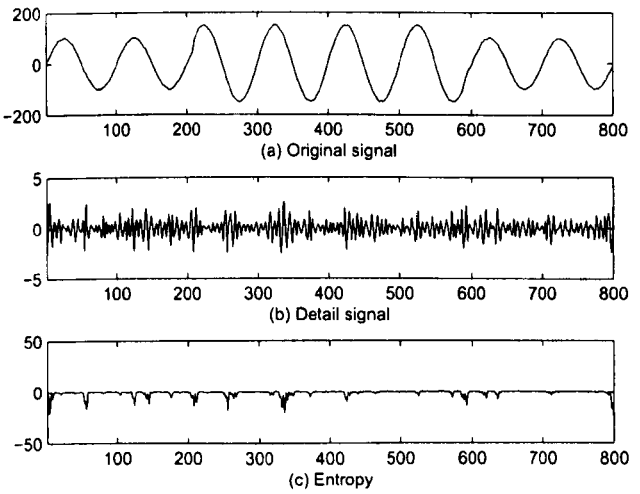


Figure 3.14: Detection of voltage swell using DB4.

are shown in Table 3.4. As it can be seen, a 1-sample error in the detail signal results in a 4-sample error in the final result of MGW. Different tests have been undertaken by changing notching frequency, amplitude and shape, respectively, and similar results have been achieved.

3.4 Conclusion

Morphological pyramid transform and morphological wavelet are two typical morphological decomposition algorithms. Based on this knowledge, this chapter proposes a multi-resolution morphological filter, which uses a morphological filter as the signal analysis operator in the decomposition procedure of morphological wavelet. Experimental results have shown that in each repetition, the output signal is less noise-corrupted and the computation is very effective.

Noticing that the multi-resolution morphological filter results in the loss of detailed information, a morphological lifting filter is proposed to recover the lost features, based on the lifting scheme. Simulation studies have shown that

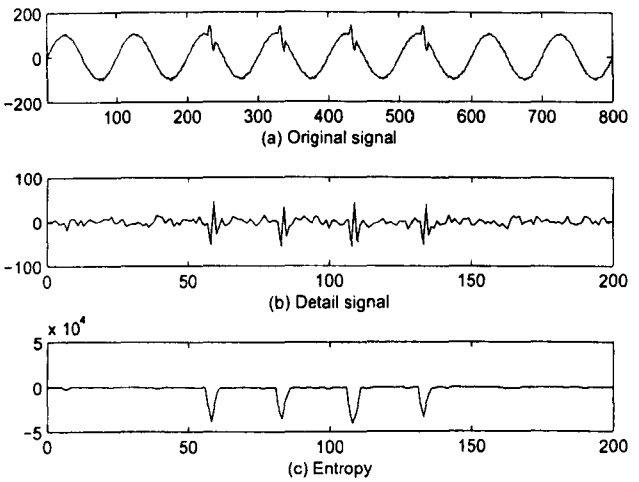


Figure 3.15: Detection of notching using MGW.

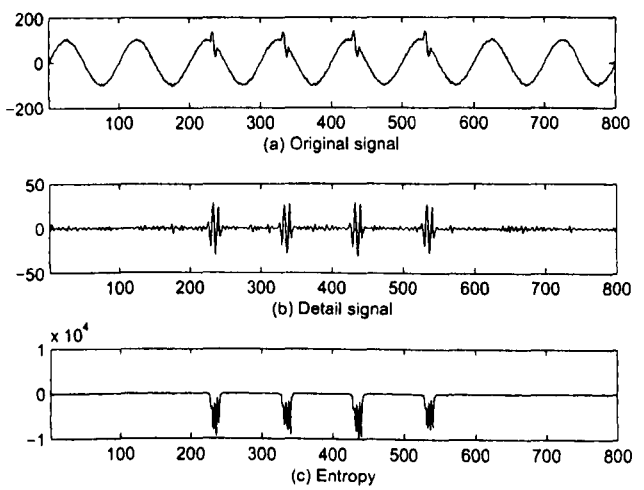


Figure 3.16: Detection of notching using DB4.

Table 3.4: Detection results of MGW and DB4

Power disturbances	Location and duration of the disturbances				
	Actual position	MGW		DB4	
		detected position	degree of match	detected position	degree of match
Impulsive transient	422-428	422-428	100%	417-432	41.85%
Oscillatory transient	425-500	424-500	98.06%	421-500	92.56%
Interruption	190-610	190-610	100%	188-612	99.06%
Voltage swell	210-590	210-590	100%	4-800	47.64%
Notching	228-240	224-240	66.27%	error	0
	328-340	324-340	66.27%	error	0
	428-440	424-440	66.27%	error	0
	528-540	524-540	66.27%	error	0

morphological lifting filter achieves better performance in noise removal than multi-resolution morphological filter, and the increase in computation time is tolerable.

The proposed filters have also been applied to ECG signal analysis, including noise removal and waveform detection. In noise removal, both the filters achieve a good performance in reducing impulsive noise meanwhile preserving the shape of the signal. In waveform detection, multi-resolution morphological filter is applied to detect characteristic waveforms, the P wave, the QRS complex and the T wave of an ECG signal, and the simulation results are satisfactory.

The MGW is also proposed based on the framework of morphological wavelet. In comparison with MHW, which is derived from Haar wavelet, MGW is more capable of presenting gradient information in the detail signal. MGW is applied to detect the location and duration of power disturbances. The simulation studies have demonstrated that the proposed method is able to detect different types of power disturbances even in a noisy environment. The adoption of entropy magnifies the effect of power disturbances while alleviates that

Table 3.5: Computation time of MGW and DB4

Power disturbances	Computation time (s)	
	MGW	DB4
Impulsive transient	0.0235	0.3473
Oscillatory transient	0.0286	0.3454
Interruption	0.0236	0.3434
Voltage swell	0.0296	0.3694
Notching	0.0285	0.3456

of the noise, so that the features of the disturbances can be more easily extracted from the entropy signal than from the detail signal. In comparison with DB4, the magnitude of entropy is much larger and the waveform of entropy contains much less high-frequency components, which may cause misjudgement when determining the start and end of a disturbance. Furthermore, all the simulation studies show that the computation time of MGW is remarkably reduced, compared with DB4.

Chapter 4

Multiscale Morphological Filters

This chapter focuses on advanced morphological filters developed based on multiscale morphological operators. Target signals are ECG signals, the spectrum of which covers a wide range of the frequency domain and thus, it is difficult to separate noise using frequency-based filters. As mentioned in section 1.3.1, ECG signals may be contaminated by white Gaussian noise and baseline wander.

An adaptive multiscale morphological filter is developed to reject high frequency noise without affecting the feature waveforms. The filter processes an ECG signal from a larger scale to a smaller scale in a recursive manner and distinguishes noise from feature waveforms according to a set of pre-set criteria.

To remove the baseline wander, section 4.2 describes an approach hybridising empirical mode decomposition (EMD) and MM. The removal of baseline wander is an important step in ECG signal processing, not only to produce a stable signal for subsequent automatic processing, but also for reliable visual interpretation [7]. The normalised signal should contain as little distortion of characteristic waveforms as possible.

4.1 Adaptive Multiscale Morphological Filter

Noise removal is the first step of almost all the signal processing tasks. As the spectrum of pure ECG signals overlaps with that of the noise, it is difficult

to separate them in the frequency domain. It is also difficult to select a proper SE of a basic morphological filter, such as opening and closing, to remove the noise meanwhile preserve the features. If the SE is too short, the noise cannot be effectively removed; if the SE is too large, the feature waveforms may be filtered as well. Moreover, such a filter will smooth the boundary of the feature waveforms. An example is given in Fig. 4.1 to illustrate the phenomena. In order to make it clearer, a fragment of PR segment and a fragment of the QRS complex are enlarged and shown in the figure, respectively. The ECG signal is filtered by two openings with a flat SE of size 3 and size 11, respectively. When the SE is short, the resulting signal still vibrates at a high frequency, which gives the evidence that there exists a large amount of noise. On the other hand, when the SE is too long, although the UP segment is smoothed, the R wave is impaired at the same time. Therefore, it is necessary to design an adaptive filter that can distinguish feature waves and use a set of multiscale SEs, *i.e.* SEs of different sizes, accordingly. The adaptive multiscale morphological filter used in this thesis is a simplification and improvement of the progressive umbra-filling (PUF) procedure proposed in [27]. In the following subsections, adaptive multiscale opening and closing are introduced, respectively, and the output of the adaptive multiscale morphological filter is the average of the opening and closing.

4.1.1 Adaptive Multiscale Opening

In order to fit the waveforms of ECG signals, the SE takes a circular function rather than a flat one. The SE is defined as:

$$g(s) = \sqrt{r^2 - s^2}, \ -r \leq s \leq r \tag{4.1.1}$$

where r is the radius of the circular function and the length of g is $2r + 1$. The advantage of using a circular SE over a flat SE is demonstrated in Fig. 4.2. Here, the length of the SEs are 21 and the origins of the SE are their geometry centres. As it can be seen, the circular SE generates a curve that fits the waveform of the input signal better than the flat SE does. According to

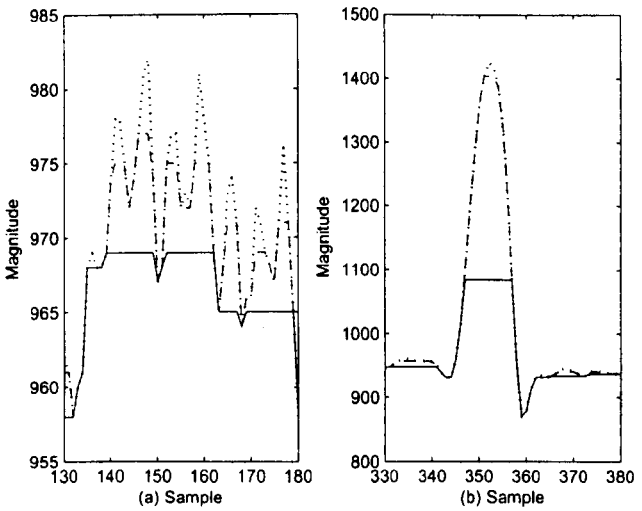


Figure 4.1: Filtering results by openings with SEs of different sizes. Dotted line: input signal. Dashed line: the size of SE is 3. Solid line: the size of SE is 11.

[57], the multiscale SE at scale n can be defined by

$$ng(s) = \sqrt{(nr)^2 - s^2}, \ -nr \leq s \leq nr. \tag{4.1.2}$$

Adaptive multiscale opening performs in a recursive way. In the first iteration, the SE of $n_0 g$, which is long enough to cover the width of P waves, is used: and in the next iteration, the SE of $(n_0 - \Delta n)g$ is used, where Δn is a pre-set step. During the process, the length of the SE decreases at a fixed rate iteration by iteration. The protocol of adaptive multiscale opening is given as follows.

- Step 1: Denote the input signal by f_0. Denote the SE used to process f_0 by g_0. Denote the scale of g_0 by n_0, *i.e.* $g_0 = n_0 g$. Denote the radius of g_0 by r_0. Set $i := 1$.

- Step 2: Calculate the opening of f_{i-1} by (2.2.1): $f_i = f_{i-1} \circ g_{i-1}$.

- Step 3: Calculate the opening residue: $R_i = f_{i-1} - f_i$. If $|R_i(t)| < T_R$, where T is a pre-set threshold with a small value, the residue of $R_i(t)$ is

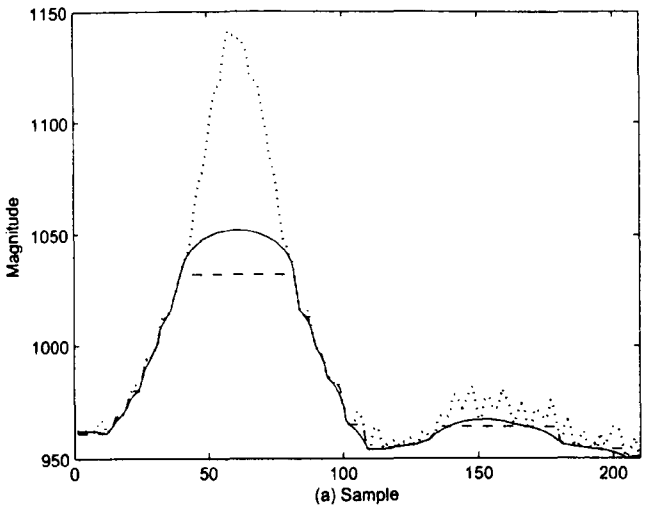


Figure 4.2: Filtering results by openings with SEs of different shapes. Dotted line: input signal. Dashed line: using a flat SE. Solid line: using a circular SE.

considered as noise. Otherwise, a decision-making process is performed to determine if it is noise or feature.

- Step 3.1: For each segment of $R_i(t) \geq T$, $t_1 \leq t \leq t_2$, calculate its width, $w = t_2 - t_1$, and its maximum magnitude, $v = \max_t\{R_i(t)\}$.
- Step 3.2: If either of the two indicators exceeds a pre-defined threshold, that is $w > T_w$ or $v > T_v$, the segment is considered to be a feature that should not be removed. Denote these $R_i(t)$ by a set $\mathcal{R}$. In order to make the width indicator adaptive, the threshold T_w changes with r.
- Step 3.3: Otherwise, the segment is considered to be noise.

- Step 4: The segments that are considered to be features are added back to f_i according to their position:

$$\tilde{f}_i(t) = \begin{cases} f_i(t) + R_i(t), & \text{if } R_i(t) \in \mathcal{R}, \\ f_i(t), & \text{otherwise.} \end{cases} \tag{4.1.3}$$

- Step 5: Set $f_{i+1} := \tilde{f}_i$. Generate g_{i+1}, which is at scale $n_{i+1} = n_i - \Delta n$. Set $i := i + 1$. Go to Step 2.

- Step 6: The process ends when $r_i = 1$ or when $|R_i(t)| < T_R$ for all t. The final output is denoted by f_o.

4.1.2 High Frequency Noise Removal of ECG Signals

The adaptive multiscale closing can be performed in the same manner. However, in Step 3, $R_i = f_i - f_{i-1}$ should be used to calculate the closing residue, due to the fact that $f_i = f_{i-1} \bullet g_{i-1} > f_{i-1}$. The final output of adaptive multiscale closing is denoted by f_c. Therefore, the filtering result of the adaptive multiscale morphological filter is the average of f_o and f_c. An example is given in Fig. 4.3 to show the performance of the filter. The parameters are set to be $n_0 = 25$, $\Delta n = 4$, $T_R = 1$, $T_{wi} = r_i/5$ and $T_v = 20$. In order to make it clearer, the figure is enlarged at the first UP segment and the second T wave. As it can be seen, the noise is smoothed out, the feature waveforms are retained, and the detail information, such as the tops of P waves, R waves and T waves and the valleys of Q waves and S waves, is well kept.

4.2 Baseline Normalisation Using Empirical Mode Decomposition and Mathematical Morphology

4.2.1 Empirical Mode Decomposition

The aim of EMD is to decompose the signal into a series of intrinsic mode functions (IMFs). An IMF is a function that satisfies two conditions: (a) in the whole data set, the number of extrema and the number of zero crossings must either equal or differ at most by one; and (b) at any point, the mean value of the envelopes defined by local maxima and minima respectively is zero [63].

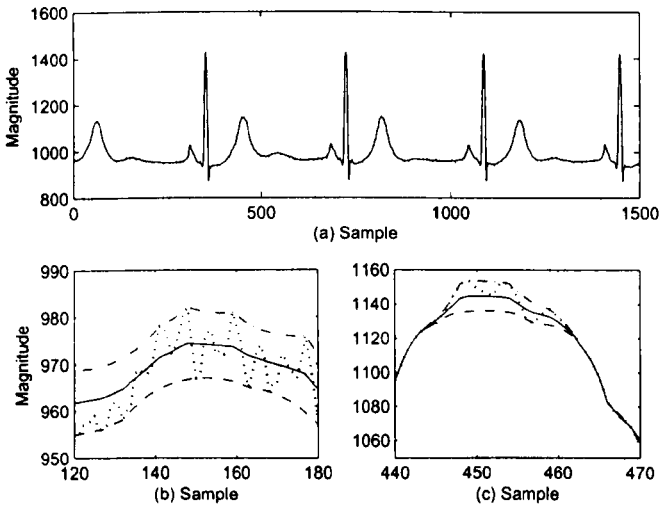


Figure 4.3: (a) Filtering result by the adaptive multiscale morphological filter. (b) The enlargement of the first UP segment. (c) The enlargement of the top of the second T wave. Dotted line: input signal. Dashed line: result of the adaptive multiscale opening. Dash-dot line: result of the adaptive multiscale closing. Solid line: average of adaptive multiscale opening and closing.

The process of finding the IMF, which is called the sifting process, is described as follows.

1. Identify the maxima and minima of the data $X(t)$, and generate the envelope by connecting maxima and minima points with cubic splines respectively;

2. Determine the local mean, $m_{10}(t)$, by averaging the envelope;

3. Since IMF should have zero local mean, subtract the mean from the data $h_{10}(t) = X(t) - m_{10}(t)$;

4. $h_{10}(t)$ is probably not an IMF. Thus, treat $h_{10}(t)$ as the input and repeat the above procedure to get $h_{11}(t) = h_{10}(t) - m_{11}(t)$; repeat $h_{1k}(t) = h_{1(k-1)}(t) - m_{1k}(t)$ as necessary until $h_{1k}(t)$ is an IMF. Here, $m_{1k}(t)$ is the local mean of $h_{1(k-1)}(t)$.

To guarantee that the IMF components retain enough physical sense of both amplitude and frequency modulations, a criterion is needed for the sifting process to stop. This is accomplished by limiting the size of the standard deviation, SD, computed from the two consecutive sifting results. The definition of SD is shown as follows.

$$\mathrm{SD} = \sum_{t=0}^{T} \frac{|h_{1(k-1)}(t) - h_{1k}(t)|^2}{h_{1(k-1)}^2(t)}. \tag{4.2.1}$$

Usually, the threshold of SD is set between 0.2 and 0.3.

Once the first sifting process returns an IMF, define $c_1 = h_{1k}$. This is the first component that contains the finest temporal scale in the signal. To obtain a series of IMFs, the procedure of the EMD method is described as follows.

1. Generate the residue, r_1, of the data by subtracting c_1 from $X(t)$: $X(t) - c_1(t) = r_1(t)$;

2. Since $r_1(t)$ still contains the information of longer period components, it is treated as the new data and subjected to the same sifting process as described above. This procedure is repeated on all the residues $r_j(t)$:

$$r_1(t) - c_2(t) = r_2(t), \ \cdots, \ r_{n-1}(t) - c_n(t) = r_n(t); \tag{4.2.2}$$

3. The repetition terminates when the residue, $r_n(t)$, is either a constant, a monotonic slope, or a function with only one extremum.

Consequently, the original signal can be reconstructed from the superposition of the components:

$$X(t) = \sum_{i=1}^{n} c_i(t) + r_n(t). \tag{4.2.3}$$

The last residue $r_n(t)$ is usually considered as the last IMF $c_{n+1}(t)$ and (4.2.3) is rewritten as:

$$X(t) = \sum_{i=1}^{n+1} c_i(t). \tag{4.2.4}$$

4.2.2 Baseline Normalisation Approach Using Empirical Mode Decomposition and Mathematical Morphology

Baseline wander corresponds to low frequency information of the ECG signal, however, it cannot be simply considered as the summation of certain highest order IMFs. As a matter of fact, a higher order IMF contains more baseline wander components and less signal components, while a lower order IMF contains more signal components and less baseline wander components [64]. Therefore, a series of low pass filters should be developed to extract the baseline wander component from each IMF. The filter applied here is the average of opening and closing:

$$\varpi_i = (c_i \circ g_i + c_i \bullet g_i)/2 \tag{4.2.5}$$

where c_i is the i^{th} IMF, g_i is a flat SE ($g_i(s) \equiv 0$) with the length of i^2, and ϖ_i is the extracted baseline from c_i.

The extraction procedure is carried out from the highest order IMF, $c_{n+1}(t)$, and stops when the variance of $\varpi_i(t)$ is small enough [64], which means that the extracted baseline is so flat that the corresponding IMF can be considered to contain no baseline wander component at all. Variance is defined as:

$$\text{var}(\varpi_i(t)) = \frac{1}{L}\sum_{t=1}^{L}[\varpi_i(t) - \mu_{\varpi_i}]^2, \tag{4.2.6}$$

where μ_{ϖ_i} is the mean value of $\varpi_i(t)$. The threshold, which determines if $\text{var}(\varpi_i(t))$ is sufficiently small, is selected by trial since it is difficult to be set automatically.

Assume that the baseline extraction procedure terminates at the m^{th} IMF, therefore, the baseline wander can be calculated from the summation:

$$\text{BW}(t) = \sum_{i=m+1}^{n+1} \varpi_i(t), \tag{4.2.7}$$

where $\text{BW}(t)$ is the extracted baseline. Consequently, the signal containing a normalised baseline, which is the so-called normalised signal, can be obtained by subtracting $\text{BW}(t)$ from the original signal $X(t)$.

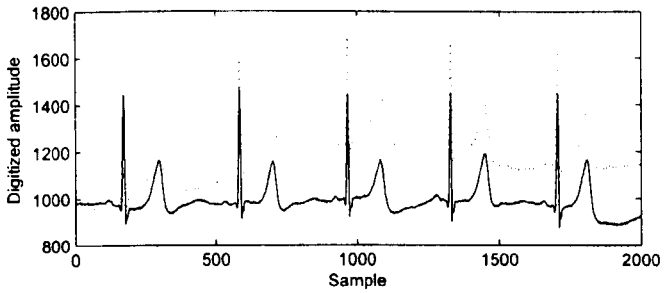


Figure 4.4: Baseline normalisation. Dotted line: the simulated ECG signal; Solid line: the normalised ECG signal.

In order to testify the applicability of the proposed method, an artificial baseline wander is added to the test signal. The resulted signal is hereby referred to as the simulated signal. Denote the test signal and the simulated signal by S and S_s, respectively. The artificial baseline wander is generated from a random signal filtered by a low pass filter. The samples of the random signal are uniformly distributed in an interval of $[0, c]$, where c is called the amplitude of artificial baseline wander. In Fig. 4.4, the dotted line shows the simulated signal of $c = 500$, while the normalised signal is illustrated in solid line. It can be seen that the proposed method works well under the condition of severe baseline distortion. As the test signals are not absolutely baseline wander free, the normalisation process removes both artificial baseline wander and the baseline wander existing in the test signal. Thus, the normalised signal is not identical to the test signal.

To analyse the performance of the proposed method qualitatively, a correlation criterion is used, which is defined as $\mathrm{COR} = \dfrac{X \cdot Y}{\|X\| \cdot \|Y\|}$, where X and Y are the two signals under investigation. The value of COR reflects the similitude between the two signals. If they are identical, COR = 1; if they are inverse, COR = -1. Another criterion is the signal-to-noise ratio (SNR) [7], which is defined as $\mathrm{SNR} = 10 \times \log_{10} \dfrac{S_\sigma}{N_\sigma}$, where $N = S_s - \mathrm{BW} - S$, and the subscript σ is the same as defined in (3.1.8).

In the simulation studies, for each signal, five artificial baseline wanders with the amplitude of 100, 300, 500, 700 and 900 are added respectively. Ta-

Table 4.1: Performance of the proposed method and Sun's method subject to the correlation criterion

ECG signal No.	Amplitude of artificial baseline wander	COR		
		Simulated signal	Proposed method	Sun's method
100	100	0.539	0.615	0.502
	300	-0.386	0.572	0.505
	500	-0.562	0.518	0.505
	700	-0.623	0.301	0.498
	900	-0.643	0.588	0.507
113	100	0.894	0.920	0.688
	300	0.397	0.886	0.685
	500	0.128	0.838	0.685
	700	0.001	0.802	0.684
	900	-0.08	0.861	0.681
222	100	0.657	0.962	0.947
	300	0.382	0.938	0.903
	500	0.306	0.926	0.868
	700	0.271	0.909	0.706
	900	0.250	0.884	0.809

bles 4.1 and 4.2 list the normalisation results of three ECG signals subject to the correlation and SNR criteria, respectively. In all the cases, the proposed method enhances COR and SNR greatly, except that for ECG113, when the amplitude of the artificial baseline wander is 100, the SNR of the output signal decreases. This is mainly because the baseline wander contained in the test signal is larger than the artificial baseline wander. Therefore, the normalisation approach removes both baseline wanders and results in an output signal that is quite different from the input one. Using other ECG signals as the input, similar results are gained.

The performance of the proposed approach is also compared with the MM-

Table 4.2: Performance of the proposed method and Sun's method subject to the SNR criterion

ECG signal No.	Amplitude of artificial baseline wander	SNR		
		Simulated signal	Proposed method	Sun's method
100	100	6.429	7.731	6.111
	300	-3.150	12.350	5.635
	500	-7.555	10.216	4.958
	700	-10.726	7.193	4.471
	900	-12.904	5.113	3.817
113	100	14.276	8.341	3.290
	300	5.148	6.081	3.266
	500	0.521	9.674	3.274
	700	-2.249	5.302	3.257
	900	-4.776	4.786	3.200
222	100	4.533	17.652	10.853
	300	-4.898	10.154	9.256
	500	-9.611	9.679	8.685
	700	-12.818	8.115	7.898
	900	-14.870	7.107	6.788

based method proposed in [7]. The COR and SNR values of the normalised signals obtained by Sun's method are also listed in Tables 4.1 and 4.2, respectively. In most cases, the proposed method gains better results than Sun's method, under both correlation and SNR criteria. Another advantage of the proposed approach is that it requires no *a priori* knowledge of the ECG signal, while Sun's method demands the length of the P wave. However, when it comes to the computation time, the proposed method is much more time-consuming than the other, due to the complex EMD decomposition procedure.

4.3 Conclusion

Two multiscale morphological filters are proposed in this chapter to remove the noise whose spectrum overlaps with that of the signal. The adaptive multiscale morphological filter, developed from the PUF procedure, removes the high frequency noise scale by scale while keeps the waveforms of interest unaffected. The approach hybridising EMD and MM aims at the removal of low frequency noise. The noisy signal is firstly decomposed into a series of IMFs, and each IMF is filtered by a morphological filter, whose SE is at the scale determined by the order of the IMF. In this manner, the signal information is preserved while the noise is removed.

In application, the two filters are tested on ECG signals to remove high frequency noise and low frequency baseline wander, respectively. The simulation studies show that both filters can preserve the feature waveforms of the ECG signals while removing the noise. The output of the adaptive multiscale morphological filter is much smoother than the noisy signal meanwhile follows the trend of the feature waveforms. The EMD-MM-based filter normalises the baseline wander so that the ECG signal has an even base, without affecting the shape of the feature waveforms.

Chapter 5

Optimal Morphological Filters

Noise removal is an underlying issue of image processing. Morphological filtering is a nonlinear filtering technique that becomes very popular, due to its ability to preserve the crucial features of image objects, such as the size and the shape. However, in some cases, basic morphological filters such as opening and closing may seem inadequate. In order to serve special purposes, morphological filters may collaborate with other filtering techniques and the parameters can be optimised to be most suitable.

One of the special case is periodic noise removal. Periodic noise widely appears in raw images and is often accompanied with Gaussian white noise. The removal of such compound noise is a challenging problem in image processing. To avoid using the time-consuming methods such as Fourier transform (FT), a simple and efficient spatial filter called optimal soft morphological filter (OSMF) is proposed in this chapter. The filter is a combination of basic soft morphological operators and the combination parameters are optimised by an improved particle swarm optimiser with passive congregation (PSOPC) [65] subject to the least mean square error criterion. The mean square error is a common measure of the difference between the filtered image and the original image when the noise is Gaussian noise. As the assessment of the periodic noise applied here is long enough that it fulfills the Gaussian distribution, the least mean square error criterion is used here as the cost function of the optimisation.

This chapter also proposes a generic approach to designing an optimal filter

which combines linear and morphological filtering techniques, so that both Gaussian and non-Gaussian noise can be rejected. The optimisation process is performed by a bacterial swarming algorithm (BSA), which is derived from the bacterial foraging algorithm (BFA) and involves underlying mechanisms of bacterial chemotaxis and quorum sensing. As linear filters are ideal for eliminating additive Gaussian white noise and morphological filters can be exclusively designed to remove non-Gaussian noise, the combination filter may take advantage from both.

The optimisation algorithms employed in this chapter are evolutionary algorithms (EAs). They are biologically inspired computational methodologies that have been intensively studied and widely applied to solve various scientific and engineering problems. The algorithms of PSOPC and BSA are introduced in detail in sections 5.2.1 and 5.3.1, respectively.

5.1 Introduction

Noise removal is a fundamental problem in image processing. In the past years, a variety of filters have been proposed aiming at the reduction of different noises. Periodic noise is a kind of noise that almost all raw images suffer from, due to electrical interference from data collecting devices, such as image scanners, capturing sensors, and video cameras.

Since periodic noise has a well-defined frequency, a usual approach is to eliminate the noise in the frequency domain. In recent research, the following three frequency methods have been proposed: (a) Wiener filter [66][67], which needs a proper and precise noise model to be built and is very complicated in computation; (b) spectrum amplitude thresholding, which is suitable only for truly periodic noise with high energy in peaks [67][68]; (c) spectral median filter [67], an effective filter with a semi-automatic detector of the peaks in the spectrum amplitude. Nonetheless, despite all their advantages, frequency filters are always computationally time-consuming and are not acceptable for time constraint applications. The inefficiency is mainly caused by the conver-

sion between the space and frequency domains and the noise peak detecting procedure.

Section 5.2 aims at the removal of the periodic noise using OSMF, a nonlinear space domain filter based on soft mathematical morphology (SMM), which requires less computational time than other traditional filters. The design of OSMF and its application to the removal of a typical type of periodic noise - sinusoidal noise is intensively studied and discussed. An optimisation technique using PSOPC is described in section 5.2.1 to evolve the OSMF, because PSOPC has shown a faster convergence rate than many other evolutionary algorithms such as the genetic algorithm (GA) and has fewer parameters to adjust. Moreover, by introducing passive congregation, the information sharing mechanism is improved and the optimisation result is more accurate. The filtering results of OSMF are compared with those obtained by spectral median filter.

Although morphological filtering technique has great ability to preserve the crucial features of image objects, linear filters should not be excluded from image processing because they are ideal for eliminating additive Gaussian white noise, which widely exists in raw images. Therefore, it is useful to hybridise both linear and morphological filtering strategies so that the combined filter can deal with both Gaussian and non-Gaussian noise.

To devise a proper combined filter also requires the optimisation of the parameters, which has always been a popular issue. However, most of the published works focus on a single type of filter and use GA to perform the optimisation. It seems that GA has dominated the optimisation problems in image processing, such as the optimisation of partition-based weighted sum filters in [69], image rejection waveguide filters in [70], morphological filters in [71], to name a few. The MRL filter, which hybridises linear and nonlinear filtering techniques, has been proposed in [72]. However, it was also optimised by GA. As for the design of the optimal combined filter (OCF) in section 5.3.3, it is performed by BSA, an improvement of the algorithm recently proposed in [73] and [74].

5.2 Optimal Soft Morphological Filter Design Using Particle Swarm Optimiser with Passive Congregation

5.2.1 Particle Swarm Optimiser with Passive Congregation

Particle swarm optimiser (PSO) is a population based algorithm developed in 1995 [75][76], inspired by social behaviour of animals such as birds flocking and fish schooling. It shares many similarities with other iteration based evolutionary computation techniques: initialise the system with a group of randomly generated population, evaluate fitness values to update the population, and search for the optimal solutions by updating generations, the strategy of which is based on the previous generations.

In PSO, the population is called swarm and each individual in the population is called a particle. As stated before, after being initialised with a group of random particles, the swarm need to be updated according to certain rules. The updating algorithm of PSO is: in every iteration, each particle is updated by two best values: (a) the best solution it has achieved so far, called personal best (*pbest*); (b) the best solution achieved by any particle in the population in this iteration. If the best solution is among all the particles, it is called global best (*gbest*); if it is taken from some smaller number of adjacent particles, it is called local best (*lbest*). After finding the two best values, the particle updates its velocity and position using the following equations [75][76]:

$$V_i^{k+1} = wV_i^k + c_1 r_1 (P_i^k - X_i^k) + c_2 r_2 (P_g^k - X_i^k) \tag{5.2.1}$$

$$X_i^{k+1} = X_i^k + V_i^{k+1} \tag{5.2.2}$$

where V_i^k is the velocity of the i^{th} particle in the k^{th} iteration; X_i is the position of the i^{th} particle; P_i is the *pbest* position of the i^{th} particle; P_g is the *gbest* position; r_1, r_2 are two random numbers between $(0, 1)$; w, c_1, c_2 are learning factors, usually being set to $w = 0.75$, $c_1 = c_2 = 2.05$ [77]. It should also

be noted that particles' velocities on each dimension are restricted within a predefined range $[0, V_{\max}]$. If the velocity tends to exceed this range, it is limited to $V_{\max}$.

Passive congregation is a special biological mechanism used in PSO, based on the fact that group members can get necessary information from not only the environment but also their neighbours [78]. Therefore, individuals in the swarm have more options to obtain information, which helps to minimise the chance of missed detection and incorrect interpretations. To keep the model simple and uniform with standard PSO, the PSOPC is defined as [65]:

$$V_i^{k+1} = \omega V_i^k + c_1 r_1 (P_i^k - X_i^k) + c_2 r_2 (P_{\mathrm{g}}^k - X_i^k) + c_3 r_3 (R_i^k - X_i^k) \tag{5.2.3}$$

$$X_i^{k+1} = X_i^k + V_i^{k+1} \tag{5.2.4}$$

where R_i is a particle randomly selected from the swarm; c_3 is the passive congregation coefficient and is set to $c_3 = 0.6$; r_3 is a uniform random sequence in the range $(0, 1)$: $r_3 \sim U(0, 1)$. The pseudo code for PSOPC is listed in Table 5.1.

5.2.2 Criteria

Peak signal-to-noise ratio

To evaluate the quality of the filtered image quantitatively, the criterion of peak signal-to-noise ratio (PSNR) is employed, which is defined as:

$$\mathrm{PSNR} = 20 \log_{10}(255/\sqrt{\mathrm{MSE}}) \tag{5.2.5}$$

$$\mathrm{MSE} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} [I_{\mathrm{o}}(i, j) - I_{\mathrm{f}}(i, j)]^2 \tag{5.2.6}$$

where MSE stands for mean square error, $I_{\mathrm{o}}(i, j)$ and $I_{\mathrm{f}}(i, j)$ are the original image and the filtered (output) image respectively, M and N denote the dimensions of the image.

Table 5.1: Pseudocode for the PSOPC algorithm

Set $k := 0$;

Randomly initialise positions and velocities of all particles;

WHILE (the termination conditions are not met)

FOR (each particle i in the swarm)

Calculate fitness:	Calculate the fitness value of current particle: $f(X_i)$;
Update *pbest*:	Compare the fitness value of *pbest* with $f(X_i)$. If $f(X_i)$ is better than the fitness value of *pbest*, then set *pbest* to the current position X_i;
Update *gbest*:	Find the global best position of the swarm. If $f(X_i)$ is better than the fitness value of *gbest*, then *gbest* is set to the position of the current particle X_i;
Update R_i:	Randomly select a particle from the swarm as R_i;
Update velocities:	Calculate velocities V_i using (5.2.3). If $V_i > V_{\max}$ then $V_i = V_{\max}$. If $V_i < V_{\min}$ then $V_i = V_{\min}$;
Update positions:	Calculate positions X_i using (5.2.4);

END FOR

Set $k := k + 1$;

END WHILE

Shape error

In image processing, information is also carried on shapes. Therefore, to indicate the filters' ability of shape preservation, the criterion of shape error is introduced [79]. The difference between the original and filtered image is defined as:

$$e_s^\gamma = \frac{1}{MN} \sum_{(i,j)=(1,1)}^{(M,N)} \sum_{(i',j')\in w} | I_o(i,j) - I_o(i',j') - (I_f(i,j) - I_f(i',j')) |^\gamma \quad (5.2.7)$$

where M, N, $I_o(i,j)$, $I_f(i,j)$ have the same meaning as in (5.2.6), γ is set to 2, (i',j') specifies the floating windows around (i,j) with the masking element w, which is a 2×2 window.

The criterion of PSNR is used to direct the optimisation process and to evaluate the performance of the filter. The criterion of shape error, on the other hand, is only applied to assess the filtering results.

5.2.3 Optimal Soft Morphological Filter for Periodic Noise Removal

Optimisation process

Noting that the soft dilation suppresses the negative parts of the noise while the soft erosion suppresses the positive parts, a filter can be designed as the combination of the two basic operators as shown below:

$$f_{smf} = \alpha \cdot (f \oplus [B,A,k]) + \beta \cdot (f \ominus [B,A,k]) \quad (5.2.8)$$

where $[B,A,k]$ is the structuring element (SE) as defined in section 2.3; $0 < \alpha < 1$ and $0 < \beta < 1$ suffice $\alpha + \beta = 1$. Obviously, in this general formula, the values of α, β, k and B, A influence the filter's performance directly. Therefore, the optimisation of these parameters is needed to achieve the best filtering results, and the filter with the optimised parameters is the OSMF we aim to find.

In order to design the OSMF, the following parameters should be considered:

- The size of the SE;
- The shape of the hard centre, also the shape of the soft boundary;
- The repetition operator k;
- The weight coefficients α and β;
- The soft operators employed by the filter.

The size of the SE relies on the period length of the noise. In order to determine the range from which the size of the SE is chosen, the period length of the noise is estimated before carrying out the optimisation. The size of the SE is chosen from 3×3 to $(2 \cdot \llcorner L/2 \lrcorner + 1) \times (2 \cdot \llcorner L/2 \lrcorner + 1)$, where L is the estimated length of the noise period, $\llcorner L/2 \lrcorner$ denotes the largest integer that is less than or equals $L/2$, and $(2 \cdot \llcorner L/2 \lrcorner + 1)$ is ensured to be odd.

Considering symmetry, the shape of the hard centre has only three different types, *i.e.* crossing, square and diamond. An example is shown in (5.2.9)~(5.2.11) when the size of the SE is 5×5.

$$\text{crossing:} \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \tag{5.2.9}$$

$$\text{square:} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{5.2.10}$$

$$\text{diamond:} \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \tag{5.2.11}$$

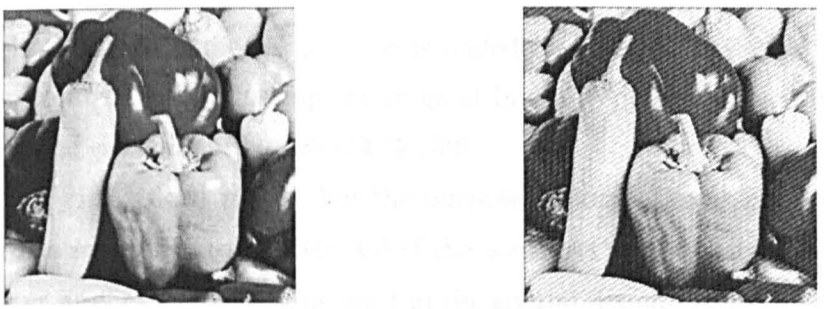
Accordingly, the repetition parameter is an integer between 1 and the length of the SE. The weight coefficients may vary from 0 to 1, as long as their summation equals 1. The fitness function is (5.2.6), which is more suitable for PSOPC than (5.2.5) as the target value can be set to 0.

The optimisation procedure is carried out under the following conditions. Image Pepper as shown in Fig. 5.1(a) is selected as the original image. When optimising an integer parameter, the corresponding value of the particle is equiprobably mapped to a valid integer. When evaluating the fitness function, the MSE is only calculated within a small part of the image in order to reduce the computational complexity. When initialising the system, the size of the swarm is chosen to be 30 and the maximum iteration is set to 300. It is testified by experiments that the fitness value remains at 419.8367 after iterating 300 times or after the size of population reaches 30. Each time the period length is assigned with a new value, the optimisation process is carried out over again. Altering the period length of the sinusoidal noise $1/\omega$ from 2 to 12, the optimisation results are listed below:

- The size of the SE is subject to the length of the noise period, as shown in Table 5.2;
- The hard centre of the SE is of the shape of crossing;
- The repetition operator k also has different values in different noise conditions, as shown in Table 5.3;
- The weight coefficients α and β are not the same when different experiments are carried out in the same conditions, but vary slightly around 0.5. Therefore, considering the adaptivity of the optimisation algorithm, the weighted coefficients are set to be $\alpha = \beta = 0.5$.

Table 5.2: Relationship between the size of the SE and the period length of the sinusoidal noise

Period Length	Size of SE
$2 \sim 3$	3×3
$4 \sim 7$	5×5
$8 \sim 12$	7×7



(a) Original image

(b) Original image contaminated by sinusoidal noise $\aleph(1/4, 30, 30)$

Figure 5.1: Add periodic noise to the original image.

An OSMF can be achieved based on the original image and the noisy image, by the above optimisation process, without requiring the knowledge of the frequency of periodic noise. However, if the frequency of the noise changes, the optimisation process could be carried out again to update the OSMF parameters. In many practical applications, the frequency of the noise is determined by electrical attribute of the data acquisition devices, which is helpful for selecting the range of OSMF parameters.

Simulation results and analysis

Sinusoidal noise. Sinusoidal noise is a typical periodic noise, denoted as $\aleph(\omega, \theta, \alpha)$, where ω, θ and α represent the frequency, angle and amplitude of the sinusoidal noise respectively; ω means that the noise is with a period length of $1/\omega$ pixels; θ varies from 0 to 180; α reflects the changes in gray value caused

Table 5.3: Relationship between the value of k and the period length of the sinusoidal noise

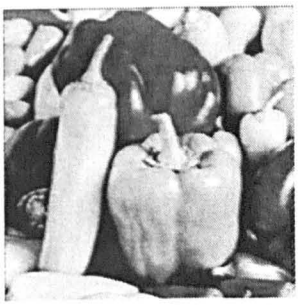
Period length	Repetition operator k
$2 \sim 4$	2
$5 \sim 7$	3
$8 \sim 10$	5
$11 \sim 12$	1

by the noise and $\alpha = 0$ means no noise is added to the image. Figure 5.1(b) shows the result of contaminating the original Image Pepper as shown in Fig. 5.1(a) with the sinusoidal noise $\aleph(1/4, 30, 30)$.

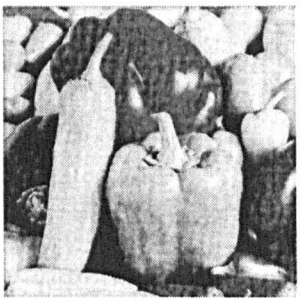
Removal of sinusoidal noise. For the purpose of comparison, spectral median filter is also applied to the removal of the noise. As stated in [67], spectral median filter applies the technique used in the spatial domain median filter to spectrum amplitude image, and checks whether the amplitude of pixel (i, j) (denoted as X_{ij}) is a peak. If X_{ij} suffices $\frac{X_{ij}}{\text{MED}_{m\times n}(X_{ij})} \geq \Theta$, then it is a peak and should be replaced by $\text{MED}_{m\times n}(X_{ij})$, which represents the median value in the $m \times n$ neighbourhood of pixel (i, j). The advantage of spectral median filter is that it does not require an estimation of the frequency of the noise, but parameters m, n, Θ should be selected properly. Since it is not discussed in [67] how to determine the parameters, they are chosen by trial and are set to be $m = 9$, $n = 9$ and $\Theta = 2.2$.

Figure 5.2 shows the results of removing sinusoidal noise $\aleph(1/4, 30, 30)$ using (a) OSMF and (b) spectral median filter. Intuitively, OSMF suppresses the noise but blurs the edge details, while spectral median filter blurs the whole image.

To quantitatively analyse their performance, the PSNR of the output images is calculated. Meanwhile, to illustrate their ability in reducing the low-frequency noise, the periodic length $1/\omega$ of the sinusoidal noise $\aleph(\omega, 30, 30)$ is varied from 2 to 12. The comparison of the results of the two filters is shown in Fig. 5.3(a). As can be seen from the figure, before filtering, the PSNRs are close to 21.9 dB. As the width of noise increases, the PSNR of OSMF decreases



(a) OSMF



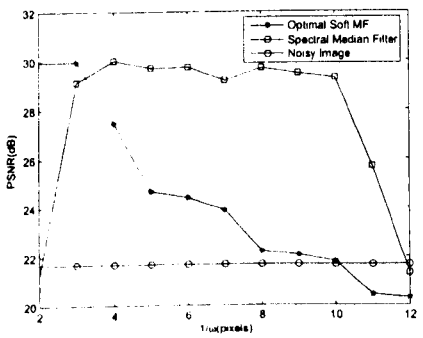
(b) spectral median filter

Figure 5.2: Comparison between OSMF and spectral median filter in the removal of sinusoidal noise $\aleph(1/4, 30, 30)$.

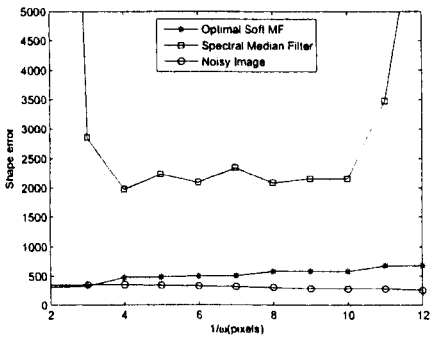
severely. Even though the size of the SE is enlarged according to the period length of the noise, the sinusoidal noise can not be removed in the space domain. On the other hand, the PSNR of spectral median filter maintains at a high level thanks to a noise peak detecting procedure, which is time-consuming yet effective.

Shape preservation. The criterion of shape error is calculated according to (5.2.7) to indicate the difference of the shapes between the original image and the image after filtering. The simulation studies are carried out on Image Pepper with the sinusoidal noise of $\aleph(\omega, 30, 30)$ with $1/\omega = 2, 3, \ldots, 12$, respectively. The performance is compared in Fig. 5.3(b), from which it can be seen that spectral median filter introduces a larger shape error than does OSMF.

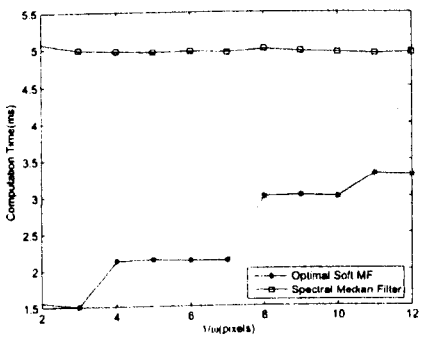
To be intuitive, corrupting the original image with the noise $\aleph(1/4, 30, 30)$, the difference images of the two filtered images and the original one are shown in Fig. 5.4, respectively, where the black pixels represent the lost details (the pixels are darkened for the purpose of observation). As demonstrated in the figure, for OSMF, the lost details concentrate on the edges while spectral median filter causes information reduction over the whole image. Like other spatial filters, morphological filters use the 'substitution' strategy that replaces the gray value of the central pixel by a value calculated from the pixels within a local window according to a certain rule. Hence, it is difficult to distinguish between



(a) PSNR

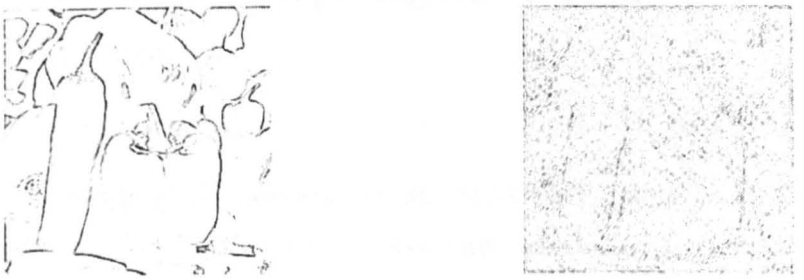


(b) Shape error



(c) Computation time

Figure 5.3: Comparison of the performance between OSMF and spectral median filter in sinusoidal noise removal.



(a) OSMF (b) spectral median filter

Figure 5.4: Comparison of the performance between OSMF and spectral median filter in shape preservation.

the edges where gray values change suddenly and the noise where gray values also change remarkably. This is the main reason why the loss of information caused by OSMF concentrates on the edges.

Computation efficiency. Although the overall performance of OSMF is not as good as spectral median filter in noise reduction, the computation time of the former is much less. For spectral median filter, the forward and inverse FTs require a total of $2N^2 \log_2 N$ complex multiplications, *i.e.* $8N^2 \log_2 N$ real multiplications for an $N \times N$ image [67]. On the other hand, the computation complexity of OSMF is only $O(N^2M^2)$. Figure 5.3(c) shows the computation time required by the two filters in the same conditions as stated above. All simulation studies are undertaken in MATLAB 7.0 running on a DELL PC with a 2.80 GHz Intel Pentium D CPU and 1.99 GB of RAM. As can be seen, the computation time required by OSMF increases as the size of the SE enlarges, while spectral median filter is always time-consuming.

5.2.4 Optimal Soft Morphological Filter in Low Frequency Condition

Optimisation process

As can be seen from the previous section, the OSMF gained so far does not behave well enough in low frequency conditions, which is mainly because the dilation and erosion displace the boundaries. Therefore, other soft operators should be employed to the filter and the general formula of (5.2.8) is extended to:

$$f_{\text{smf}} = \alpha \cdot (f \otimes [B, A, k]) + \beta \cdot (f \oslash [B, A, k]) \tag{5.2.12}$$

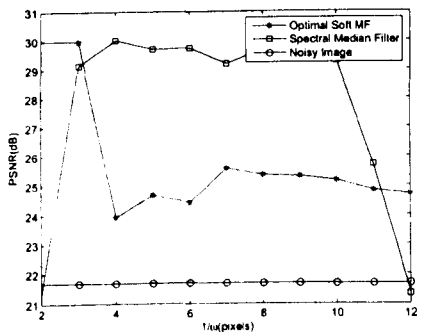
where $\otimes$ represents soft dilation, soft opening or soft open-closing. Accordingly, $\oslash$ represents soft erosion, soft closing or soft close-opening. According to the previous knowledge, the size of the SE is determined by Table 5.2 and the shape of the SE is 'crossing'. Nevertheless, there are still three parameters to be optimised, *i.e.* the repetition operator k, the weight coefficients α and β, and the choice of the soft operators. The optimisation condition is the same as depicted in section 5.2.3, and the optimisation results are shown in Table 5.4.

Table 5.4: Relationship between the period length of the sinusoidal noise and the design of OSMF

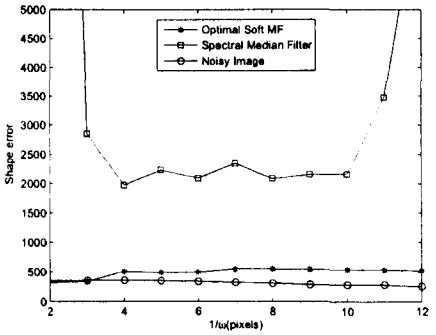
Period length	Size of SE	Design of OSMF
$2 \sim 3$	3	$0.5(f \oplus [B, A, 2]) + 0.5(f \ominus [B, A, 2])$
4	5	$0.5(f \oplus [B, A, 2]) + 0.5(f \ominus [B, A, 2])$
$5 \sim 6$	5	$0.5(f \oplus [B, A, 3]) + 0.5(f \ominus [B, A, 3])$
$7 \sim 12$	7	$0.5(f \circ [B, A, 2]) + 0.5(f \bullet [B, A, 2])$
$9 \sim 10$	7	$0.5(f \circ [B, A, 4]) + 0.5(f \bullet [B, A, 4])$
$11 \sim 12$	7	$0.5(f \circ [B, A, 4]) + 0.5(f \bullet [B, A, 4])$

Simulation results and analysis

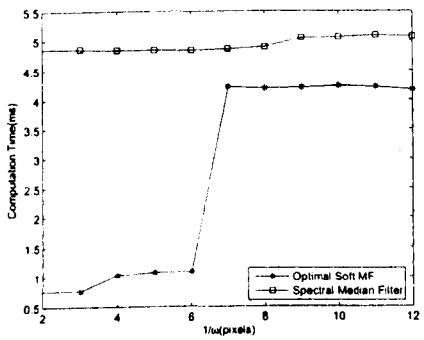
Removal of sinusoidal noise. The comparison of filtering results are also carried out between OSMF and spectral median filter in the simulation stud-



(a) PSNR



(b) Shape error



(c) Computation time

Figure 5.5: Comparison of the performance between OSMF and SMF in low frequency condition.

ies, as shown in Fig. 5.5(a). Compared with Fig. 5.3(a), it can be seen that the performance of OSMF has been greatly improved by applying more complicated soft operators. Furthermore, when the period length increases to 12 pixels, the PSNR of OSMF is even higher than that of spectral median filter, for the frequency of the noise is too low to be distinguished from that of the original image by the frequency filter.

Shape preservation. The shape errors of the filtered images are shown in Fig. 5.5(b), which is quite similar to Fig. 5.3(b). In low frequency conditions, the shape error only declines slightly, which illustrates that the adoption of soft opening and soft closing has no significant influence in shape preservation.

Computation efficiency. Unlike the former two indicators, the computation

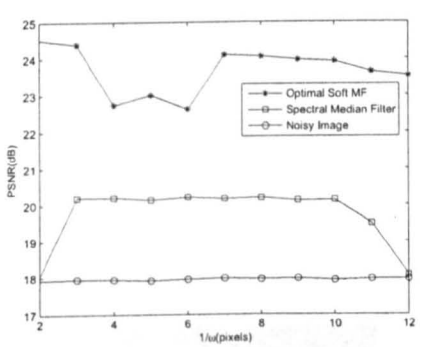
time of OSMF increases greatly when soft opening and soft closing are applied, as shown in Fig. 5.5(c). However, even in this condition, OSMF is still less time-consuming than SMF.

5.2.5 Optimal Soft Morphological Filter for Compound Noise Removal

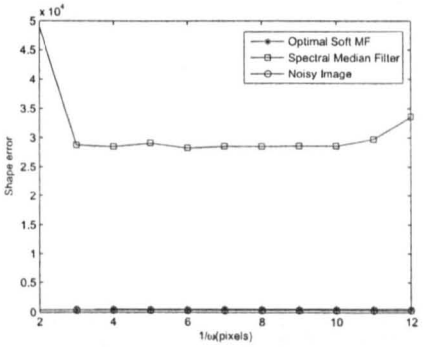
In practice, periodic noise is usually accompanied with random noise, such as Gaussian white noise, which is a good approximation of many real-world situations. Therefore, it is necessary to evaluate OSMF's ability in reducing such compound noise.

Corrupting Image Pepper by both sinusoidal noise $\aleph(1/\omega, 30, 30)$ and Gaussian white noise $\mathcal{N}(0, 0.01)$, then filtering it by OSMF and spectral median filter, respectively, the simulation results are shown in Fig. 5.6. It should be noted that both filters are the same as used in section 5.2.4. As can be seen from the figures, when Gaussian white noise is added, OSMF outperforms spectral median filter greatly as far as PSNR is concerned. The PSNR of spectral median filter descends significantly. On the contrary, the PSNR of OSMF only descends remarkably when the frequency is high. Spectral median filter cannot remove Gaussian white noise because the spectrum of the noise is buried in that of the image and is hard to be eliminated by the low-pass filter. On the other hand, in the space domain, the pixel denoting Gaussian white noise is easy to be replaced by other pixels representing the image signal in its neighbourhood using the sorting strategy of SMM.

The shape error of spectral median filter also increases dramatically, while that of OSMF remains at a low level. The computation time is not depicted because it is irrelative to the noise added to the image. Applying both OSMF and spectral median filter to two other images, the results are shown in Figs. 5.7~5.8, which demonstrate that the proposed filter achieves a better performance. Therefore, it can be concluded that OSMF is capable of reducing the periodic noise and is robust.



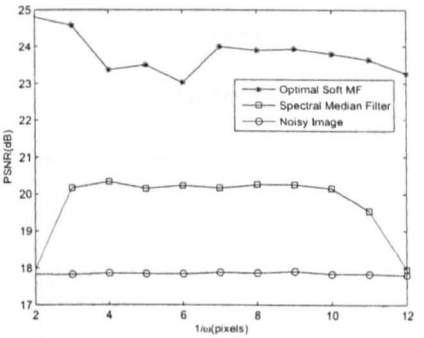
(a) PSNR



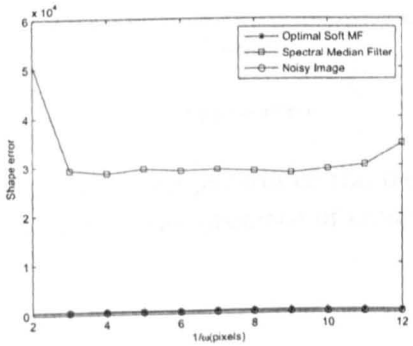
(b) Shape error

Figure 5.6: Comparison of the performance between OSMF and spectral median filter in the presence of compound noise.

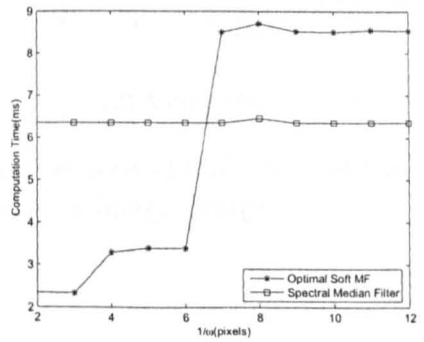
(a) Image Lena



(b) PSNR



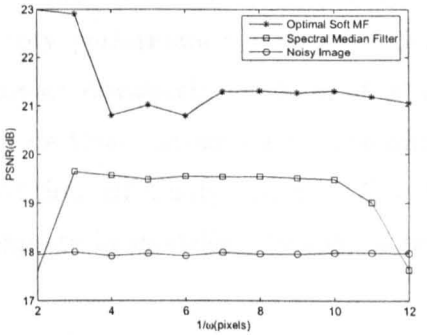
(c) Shape error



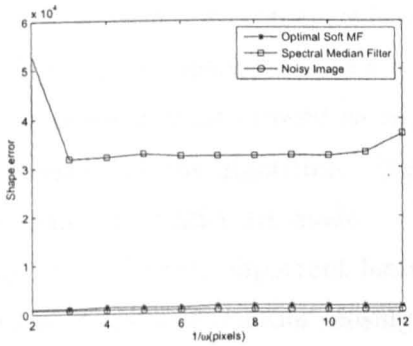
(d) Computation time

Figure 5.7: Comparison of the performance between OSMF and spectral median filter in the presence of compound noise on Image Lena.

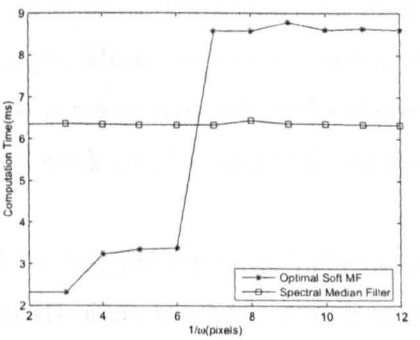
(a) Image Bridge



(b) PSNR



(c) Shape error



(d) Computation time

Figure 5.8: Comparison of the performance between OSMF and spectral median filter in the presence of compound noise on Image Bridge.

5.3 Optimal Combined Filter Design Using Bacterial Swarming Algorithm

5.3.1 Bacterial Swarming Algorithm

BSA is developed from BFA [80], which is one of the EAs that has received great attention recently. It is based on the study of the *E. coli* chemotaxis behaviour and is claimed to have a satisfactory performance in optimisation problems. However, BFA suffers from a major drawback which most EAs can not avoid: the optimisation process may be time consuming in searching along the randomly selected directions. In addition, BFA only describes *E. coli* chemotaxis phenomenon, which seems inadequate in modelling biological behaviours.

To overcome these problems, BSA is proposed to improve the performance of BFA. In the chemotaxis behaviour of BSA, which fulfills the searching process, the bacterial rotation angle is calculated by Polar-to-Cartesian coordinate transform and is restricted within a certain range. Moreover, the behaviour of quorum sensing is introduced to accelerate the convergence rate and enhance the diversity of the algorithm. Based on the work of [73] and [74], further improvements of BSA are made.

In BSA, the two important features used to describe bacterial behaviours are chemotaxis and quorum sensing. Chemotaxis offers the basic search principle of BSA, and quorum sensing enables BSA to escape from local optima. In order to describe these features, two mathematical models are constructed correspondingly. During the optimisation process, they are performed orderly in each iteration.

Chemotaxis

E.coli bacteria sense simple chemicals in the environment, and are able to decide whether the nutrients at a certain location are getting better or worse [81]. Bacteria swim by rotating thin, helical filaments known as flagella

driven by a reversible motor embedded in the cell wall. *E.coli* have $8 \sim 10$ flagella placed randomly on the cell body [82]. In the chemotaxis behaviour, the motor runs either clockwise or counterclockwise with the different direction of protons flowing through the cytoderm. When the motors turn clockwise, the flagellar filaments work independently, which leads to an erratic displacement. This behaviour is called tumble. When the motors turn counterclockwise, the filaments rotate in the same direction, thus push the bacterium steadily forward. This behaviour is called run. The alternation of tumble and run is presented as a biased random walk.

The chemotaxis behaviour is modelled by a tumble-run process that consists of a tumble step and several run steps. The tumble-run process follows gradient searching principles, which means the bacteria's position is updated in the run steps by the gradient information provided by the tumble step. Determining the rotation angle taken by a tumble action in an n-dimensional search space can be described as follows. Suppose the p^{th} bacterium, at the k^{th} tumble-run process (*i.e.* the k^{th} iteration), has a current position $X_p^k \in \mathbb{R}^n$, a rotation angle $\varphi_p^k = (\varphi_{p1}^k, \varphi_{p2}^k, ..., \varphi_{p(n-1)}^k) \in \mathbb{R}^{n-1}$ and a tumble length $D_p^k(\varphi_p^k) = (d_{p1}^k, d_{p2}^k, ..., d_{pn}^k) \in \mathbb{R}^n$, which can be calculated from φ_p^k via a Polar-to-Cartesian coordinate transform:

$$\begin{aligned} d_{p1}^k &= \prod_{i=1}^{n-1} \cos\left(\varphi_{pi}^k\right) \\ d_{pj}^k &= \sin\left(\varphi_{p(j-1)}^k\right) \prod_{i=p}^{n-1} \cos\left(\varphi_{pi}^k\right) \quad j = 2, 3, ... n-1 \\ d_{pn}^k &= \sin\left(\varphi_{p(n-1)}^k\right). \end{aligned} \tag{5.3.1}$$

The maximal rotation angle $\theta_{\max}$ is related to the number of the dimensions of the objective function, which can be formulated as:

$$\theta_{\max} = \frac{\pi}{\text{round}(\sqrt{n+1})} \tag{5.3.2}$$

where n is the number of dimensions of the objective function.

At the k^{th} tumble-run process, the p^{th} bacterium generates a random rotation angle, which falls in the range of $[0, \theta_{\max}]$. Then during the run steps, the

bacterium keeps moving along this rotation angle direction. A step of tumble and run can be expressed as:

$$\hat{\varphi}_p^k = \varphi_p^k + r_2\theta_{\max}/2 \tag{5.3.3}$$

$$\hat{X}_p^k(1) = X_p^k + r_1 l_{\max} D_p^k(\hat{\varphi}_p^k) \tag{5.3.4}$$

$$\varphi_p^{k+1} = \hat{\varphi}_p^k \tag{5.3.5}$$

where φ_p^k and X_p^k indicate the rotation angle and the position of the p^{th} bacterium at the beginning of the k_{th} iteration, respectively; $\hat{\varphi}_p^k$ is its rotation angle after the tumble step; $\hat{X}_p^k(1)$ is its position immediately after the first run step; $r_1 \in \mathbb{R}$ is a normally distributed random number generated from $\mathcal{N}(0,1)$ with $\mathcal{N}$ standing for normal distribution; $r_2 \in \mathbb{R}^{n-1}$ is a random sequence with a range of [0,1]; $l_{\max}$ is the maximal step length of a run; finally, φ_p^{k+1} is the rotation angle at the beginning of the next iteration, *i.e.* the $(k+1)^{\text{th}}$ iteration.

Once the angle is decided by the tumble step, the bacterium will run for a maximum of n_c steps, or until reaching a position with a worse evaluation value. The position of the p_{th} bacterium is updated at the h_{th} $(h > 1)$ run step as follows:

$$\hat{X}_p^k(h) = \hat{X}_p^k(h-1) + r_1 l_{\max} D_p^k(\hat{\varphi}_p^k). \tag{5.3.6}$$

After n_c steps of run process, the bacterium stops at position $\hat{X}_p^k(n_c)$.

Quorum sensing

A bacterium uses a batch of receptors to sense the signals coming from external substances. The bacterium also has an inducer, which is a molecule inside the bacterium, to start the gene expression [80]. When the inducer binds the receptor, it activates the transcription of certain genes, including those for inducer synthesis. This process, called quorum sensing, was discovered by Miller explaining the cell-to-cell communication in [83].

Quorum sensing can occur within a single bacterial species as well as between disparate species. In BSA, most nutrients locate around optima, which correspond to better fitness values. Based on this assumption, the density of the inducer is increased if the fitness value is better. Therefore, in the single

bacterial species case, each bacterium is attracted by the signal randomly, and a bacterium's position is updated by:

$$X_p^{k+1} = \delta r_3 (X_{\text{best}} - \hat{X}_p^k) \tag{5.3.7}$$

where δ is a coefficient describing the strength of the bacterium's attraction, $r_3 \in \mathbb{R}^n$ is a normally distributed random sequence drawn from $\mathcal{N}(0,1)$, X_{best} indicates the position of current best global solution updated after each function evaluation, and $\hat{X}_p^k$ is the position of the p^{th} bacterium at the k^{th} iteration after the tumble-run process.

If quorum sensing occurs between disparate species, it may cause virulence between each species, which also avoids pre-mature results. In BSA, a small number of the bacteria are randomly selected to be repelled. The repelling rate is denoted by R_{r}. If the p^{th} bacterium turns into the repelling process, a random angle in the range of $[0, \pi]$ is generated. The bacterium is thereby 'moved' to a random position following this angle in the search space, which can be described as:

$$X_p^{k+1} = \hat{X}_p^k + r_3 l_{\text{range}} D_p^k (\hat{\varphi}_p^k + r_2 \cdot \pi/2) \tag{5.3.8}$$

where l_{range} is the range of the search space. The pseudo code of BSA is listed in Table 5.5.

5.3.2 Criteria

The objective of optimisation is to obtain a filter that rejects noise to the greatest extent, and PSNR is the index to show the quality of an image. As stated in section 5.2.2, a larger value of PSNR indicates a higher quality. Therefore, the optimisation is subject to the largest PSNR and the fitness function is selected as 1/PSNR.

To evaluate the quality of the filtered images quantitatively, three criteria are employed in the experimental studies. The first two criteria are PSNR and shape error, as stated in section 5.2.2. The third one is the speckle index,

Table 5.5: Pseudo Code of BSA

Set $k := 0$;
Randomly initialise bacteria's positions;
WHILE (termination conditions are not met)
  FOR (each bacterium p)
    Tumble: Generate a random rotation angle by (5.3.3). Set $h := 1$;
    Run:
      FOR (each run step h)
        Calculate the bacterium's position after the h_{th} run step, $\hat{X}_p^k(h)$; if $h = 1$, use (5.3.4); if $h > 1$, use (5.3.6). If the fitness value at current position is less than the the value at previous position, the bacterium will move towards the angle until it reaches the the maximum step, n_c; otherwise, the bacterium will stop at current position. Increase h by 1;
      END FOR
  END FOR
  Quorum Sensing: Most of the bacteria are attracted to the global optimum by (5.3.7); a small number of bacteria are repelled by (5.3.8);
  Set $k := k + 1$;
END WHILE

termed by S. The definition of S is:

$$S = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{\sigma(i,j)}{\mu(i,j)} \tag{5.3.9}$$

where $\sigma(i,j)$ is the standard deviation of pixel (i,j) within the neighbourhood of a 3×3 window, and $\mu(i,j)$ is the mean value. The smaller S is, the better

quality the image has.

5.3.3 Optimal Combined Filter for Compound Noise Removal

Optimisation process

In image processing, filtering is a technique for modifying or enhancing an image. It applies a certain algorithm to the values of the pixels of the input image within a neighbourhood to calculate the value of the corresponding pixel in the output image. Literally, the algorithm of linear filtering combines the input values linearly. The most common linear filter is the FIR filter, which corresponds to convolution in the space domain. Denote the image and the convolution kernel by I and f, respectively, the definition of FIR filter is given as follows.

$$I \circledcirc f(i,j) = \sum_{s,t} f(s,t) \cdot I(i-s, j-t). \tag{5.3.10}$$

On the other hand, MM is a nonlinear approach for image processing. The pixels within the neighbourhood interact with an SE, a set of the same size as the neighbourhood. The dilation and erosion operators of 2-dimensional grey-scale version can be derived from (2.1.17) and (2.1.18), respectively. The definitions are:

$$I \oplus g(i,j) = \max_{s,t}\{I(i-s, j-t) + g(s,t)\} \tag{5.3.11}$$

$$I \ominus g(i,j) = \min_{s,t}\{I(i+s, j+t) - g(s,t)\} \tag{5.3.12}$$

where I and g denote the input image and the SE, respectively.

Instead of searching for the maximal or minimal element, a more general strategy is used, which returns the r^{th} largest element in the neighbourhood. This filter is referred to as the ranking filter [72], and is defined as:

$$I \overset{r}{\ominus} g(i,j) = R_r\{I(i-s, j-t) + g(s,t)\} \tag{5.3.13}$$

where $R_r(x)$ sorts the elements of x and returns the r^{th} largest one. For binary ranking filters, in which the SE g comprises 0 and 1 only, the definition is

altered to be:

$$I \overset{r}{\oplus} g(i,j) = R_r\{I(i-s, j-t)\}, \text{ if } g(s,t) = 1. \tag{5.3.14}$$

The combined filter calculates the weighted average of the outputs of the FIR filter and the ranking filter, which can be expressed by

$$f_{cf} = \lambda \cdot I \odot f + (1-\lambda) \cdot I \overset{r}{\oplus} g, \ 0 \leq \lambda \leq 1. \tag{5.3.15}$$

In order to design an optimal combined filter, the following parameters should be considered.

- The size of the convolution kernel, f;
- The values of the elements of f, and they should suffice the condition of $\sum_{s,t} f(s,t) = 1$;
- The size of g;
- The shape of g. For gray scale version, the values of the elements of g can be random integers; for binary version, they can be either 0 or 1;
- The ranking coefficient, r, which should be no more than the number of the elements of g involved in (5.3.13) or (5.3.14);
- The weight coefficient, λ, and $0 \leq \lambda \leq 1$.

Simulation results and analysis

Parameter setting. BSA is applied to optimise the parameters of the combined filter depicted by (5.3.15). The problem is slightly simplified by assuming that the sizes of convolution kernel f and SE g are restrained between 2×2 and 5×5. Also, the elements of f are symmetric around the centre. In addition, since the test images are 8-bit bitmaps, the SE follows the gray scale pattern, but the values of its elements are chosen from $\{-1, 0, 1, 2, 3\}$ only. When $g(s,t) = -1$, the corresponding pixel of the image, $f(i-s, j-t)$, is excluded from the ranking filter, R_r. In all experiments, the initial population

(a) Original image (b) Noisy image

Figure 5.9: Add Gaussian white noise and salt & pepper noise to Image Pepper.

size is selected to be 100 and the maximal iteration is 1000. In the tumble-run process, the maximal number of the run steps along an angle is 4 steps. The repelling rate R_r is set by trial. The same tests as described in [65] have been carried out to find a suitable value for R_r, and the result is it is set to be 0.2, *i.e.* a portion of 20 percents of the bacteria will be repelled in each iteration. The coefficient δ is set in the same way and $\delta = 2$.

Design OCF. An original noise-free image shown in Fig. 5.9(a) is given as a reference. Both Gaussian white noise $\mathcal{N}(0, 0.01)$ and salt & pepper noise N(0.05) are added to this image, as shown in Fig. 5.9(b). Here, N(d) denotes salt & pepper noise of density d. BSA is then employed to optimise the parameters of the combined filter, f_{cf}, subject to the highest PSNR. To diminish the computation burden, only a small part of the image (36×36) is used for optimisation. The combined filter optimised by BSA, which is referred to as the BSA filter in this section, is therefore applied to other noise-corrupted images to test its performance.

As stated in section 5.3.3, 6 types of parameters need to be optimised. The curve of convergence process is illustrated in solid line in Fig. 5.10, which is an average of 30 runs. The figure shows that BSA has a fast convergence rate. One of the optimisation results is listed as follows. It should be noted that although the optimisation results vary slightly each time, the corresponding

BSA filters have a highly similar performance, which means BSA has a stable performance. This can be proved in Table 5.7.

$$f = \begin{bmatrix} 0.0411 & 0.0318 & 0.0634 & 0.0318 \\ 0.0318 & 0.0682 & 0.0554 & 0.0682 \\ 0.0634 & 0.0554 & 0.0424 & 0.0554 \\ 0.0318 & 0.0682 & 0.0554 & 0.0682 \\ 0.0411 & 0.0318 & 0.0634 & 0.0318 \end{bmatrix} \tag{5.3.16}$$

$$g = \begin{bmatrix} 2 & 2 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 1 \end{bmatrix} \tag{5.3.17}$$

$$r = 5 \tag{5.3.18}$$

$$\lambda = 0.2659. \tag{5.3.19}$$

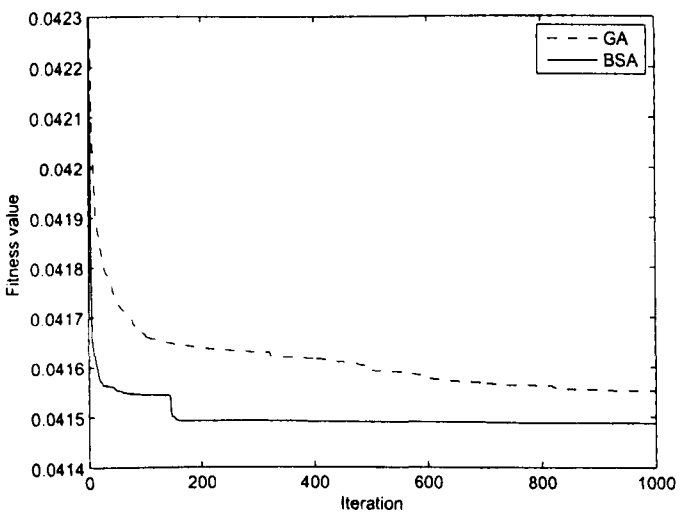


Figure 5.10: Comparison of the convergence process between BSA and GA.

OCF on Image Pepper. The BSA filter whose parameters are set by (5.3.16)~(5.3.19) is therefore used to filter the noisy image of Fig. 5.9(b), and the result is shown in Fig. 5.11(a). For the purpose of comparison, the combined filter is also optimised by GA. The publicly accepted GA toolbox is

employed and its parameters are set as follows. The population size and maximum iteration are set to be 100 and 1000, respectively, which are the same as BSA. The selection function, crossover function and mutation function are set to be roulette, single point and uniform with a mutation rate of 0.01. The convergence process of GA, which is also the average of 30 runs, is plotted in Fig. 5.10 in dashed line. Apparently, BSA converges faster than GA and achieves a better result. One of the optimisation results obtained by GA is listed below. For the convenience of description, the combined filter optimised by GA is called GA filter in this section. The parameters of one of GA filters are set as below.

$$f = \begin{bmatrix} 0.0290 & 0.0004 & 0.0290 \\ 0.0605 & 0.1087 & 0.0605 \\ 0.1087 & 0.2064 & 0.1087 \\ 0.0605 & 0.1087 & 0.0605 \\ 0.0290 & 0.0004 & 0.0290 \end{bmatrix} \tag{5.3.20}$$

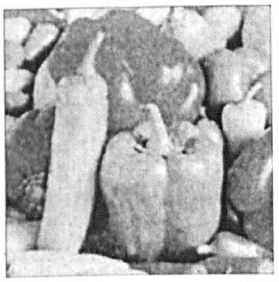
$$g = \begin{bmatrix} -1 & 3 \\ -1 & 1 \\ -1 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{5.3.21}$$

$$r = 4 \tag{5.3.22}$$

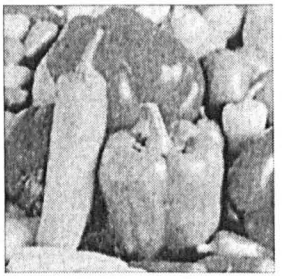
$$\lambda = 0.5006. \tag{5.3.23}$$

Using the GA filter to remove the noise from Fig. 5.9(b), the result is shown in Fig. 5.11(b), which reveals that a certain amount of noise still remains in the output image. The FIR filter and median filter are involved in the simulation studies as well. Their performance is demonstrated in Figs. 5.11(c) and 5.11(d), respectively. The convolution kernel of the FIR filter is:

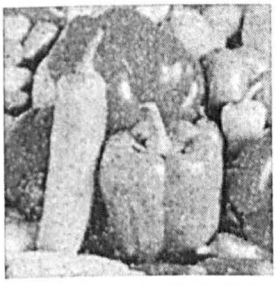
$$f_{\text{FIR}} = \frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}. \tag{5.3.24}$$



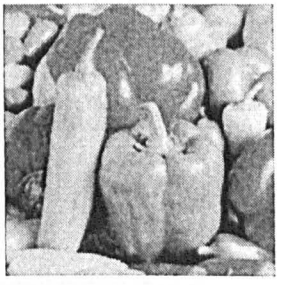
(a) BSA filter



(b) GA filter



(c) FIR filter



(d) Median filter

Figure 5.11: Performance of the four filters on Image Pepper.

The median filter is 4-connected and its size is 3×3.

Intuitively, the FIR filter blurs the edge details and is useless in suppressing the salt & pepper noise. On the contrary, the median filter is effective in dealing with salt & pepper noise as well as keeping the shape information, as it should be. However, it makes the whole image brighter, which is an evidence that the median filter can not remove Gaussian noise effectively. As for BSA filter, it removes both types of noise to a great extent meanwhile keeps more detail information than the other filters.

To quantitatively analyse their performance, the three criteria mentioned in section 5.3.2 are employed and the evaluation results are listed in Table 5.6. To make the results more comprehensible, the percentage of improvement is

also demonstrated. From the table it can be seen that BSA filter achieves the best performance under all the criteria. The results for BSA filter and GA filter are the average results of 30 runs. To show the robustness of the algorithms, the variance of the results is also calculated, which is listed in Table 5.7.

Table 5.6: Quantitative performance of the four filters on Image Pepper subject to the criteria of PSNR, shape error (e_s) and speckle index (S)

Filter	PSNR	e_s	S
Before filtered	17.049	10.708	0.109
BSA filter	27.846(63.33%)	1.283(88.01%)	0.060(44.95%)
GA filter	27.153(59.26%)	1.292(87.93%)	0.065(40.37%)
FIR filter	24.607(44.33%)	1.651(84.58%)	0.071(34.86%)
Median filter	27.677(62.34%)	1.566(85.37%)	0.070(35.78%)

Table 5.7: Variance of the results

Filter	PSNR	e_s	S
BSA filter	0.0449	2.0221×10^{-6}	3.7767×10^{-5}
GA filter	0.0490	1.4302×10^{-5}	5.2397×10^{-5}

OCF on Image Lena. The 30 BSA filters and GA filters are also used to filter other images corrupted by both Gaussian and non-Gaussian noise. An example is given in Fig. 5.12, where the original image of Lena, the contaminated image by Gaussian white noise $\mathcal{N}(0, 0.01)$ and salt & pepper noise N(0.05), the filtering results of BSA filter, GA filter, FIR filter and median filter are illustrated respectively. The quantitative results are listed in Table 5.8, where the results of BSA filter and GA filter are the average value. From the figures and the table, similar conclusion can be drawn that the BSA filter outperforms the other three filters.

5.4 Conclusion

This chapter discusses OSMF for periodic noise removal and OCF for both Gaussian and non-Gaussian noise removal. Both filters are obtained from the

(a) Image Lena

(b) Noisy image

(c) BSA filter

(d) GA filter

(e) FIR filter

(f) Median filter

Figure 5.12: Add Gaussian white noise and salt & pepper noise to Image Lena and the performance of the four filters.

Table 5.8: Quantitative performance of the three filters on Image Lena subject to the criteria of PSNR, shape error (e_s) and speckle index (S)

Filter	PSNR	e_s	S
Before filtered	16.524	8.501	0.096
BSA filter	26.479(60.24%)	1.617(80.98%)	0.053(44.79%)
GA filter	25.901(56.75%)	1.899(77.66%)	0.094(2.08%)
FIR filter	24.655(49.21%)	1.899(77.66%)	0.063(34.38%)
Median filter	25.432(53.91%)	1.895(77.71%)	0.064(33.33%)

optimisation of a generic framework. The optimisation approaches proposed in this chapter are also able to design filters for other signal/image processing applications.

OSMF is optimised from the framework of soft morphological filter using PSOPC. The optimisation process has been carried out in two stages. In the first stage, OSMF only involves soft dilation and soft erosion, while in the second stage, the general formula is extended to include other soft operators. In each stage, once the original image and the noisy image are given, the optimisation process is carried out without requiring the knowledge of the frequency of periodic noise. However, if the noise frequency changes, the optimisation process should be carried out again to update the OSMF parameters.

Simulation studies are carried out to remove periodic noise of various frequencies. The performance of OSMF has been presented in comparison with spectral median filter, a frequency filter that is very powerful in the reduction of periodic noise. When pure periodic noise is added to the original image, OSMF achieves better performance than spectral median filter in the elimination of the noise in high frequency conditions, although is not as good as the latter in low frequency conditions. But taken the shape error and the computation time into consideration, OSMF outperforms spectral median filter greatly in both high and low frequency conditions.

When pure sinusoidal noise is added together with Gaussian white noise, which is more practical, the ability of spectral median filter to reduce noise decreases dramatically. On the contrary, the ability of OSMF to remove the

compound noise is not affected in this case and is much better than that of spectral median filter. Besides, the interference of Gaussian white noise makes spectral median filter introduce a significantly larger shape error to the filtered image, while the shape error caused by OSMF remains at a low level. In conclusion, the simulation results demonstrate OSMF is more effective and less time-consuming in reducing both pure periodic and compound noise meanwhile preserving the details of the original image.

OCF is another effective optimal filter for image processing presented in this chapter. The filter combines both linear and morphological filtering techniques so as to remove Gaussian and non-Gaussian noise. Afterwards, the the parameters of the combined filter are optimised using BSA to obtain optimal filtering results.

BSA is inspired by the underlying mechanisms of bacterial foraging behaviours – chemotaxis and quorum sensing. Chemotaxis is based on the gradient searching behaviour, which ensures that the bacterium always moves to a better position than the previous step. Two versions of the quorum sensing behaviour are introduced in BSA. When quorum sensing happens inside a single bacterial species, bacteria are attracted to global optimum, which accelerates the convergence speed. When quorum sensing happens among disparate species, bacteria are randomly replaced in the search space, which prevents bacteria from being trapped into local optima.

Simulation studies of using BSA and GA to optimise the combined filter, respectively, have shown that the convergence speed of BSA is faster than GA and the combined filter optimised by BSA achieves a better performance than the GA filter, FIR filter and median filter. Implementing these filters to other noise corrupted images, the same conclusion can be drawn.

Chapter 6

Embedding-based Technique

6.1 Introduction

This chapter introduces a novel signal processing technique that is based on the embedding theorem. It is believed that the features of a signal can be more clearly revealed if it is embedded to the phase space through a proper embedding strategy. The embedding theorem [84][85] shows that a time series can be mapped to a higher dimensional space, which is the so-called phase space, through embedding [86]. Therefore, a sampled signal can be transformed to the phase space so that its features can be more clearly viewed. As the theoretical basis of the proposed scheme, the embedding theorem is introduced in section 6.2.

The embedding-based technique is applied to three applications: feature waveform detection of ECG signals, phasor measurement of power system signals and disturbance detection. For the last application, two schemes based on Gustafson-Kessel (GK) clustering and projection, respectively, are proposed.

6.2 The Embedding Theorem

The embedding theorem was originally proposed for dynamical systems [84][87]. In mathematics and physics, a dynamical system is usually described

by a state space, each orthogonal coordinate of which represents one of the variables needed to specify the instantaneous state of the system [88][89]. Hence, all the possible states of the dynamical system are represented in the state space and each possible state corresponds to a unique point [90]. However, in practice, it is usually impossible to measure all the variables of a dynamical system. Fortunately, Takens has proved in [84] that it can be reconstructed from a time series of a collection of the states using the embedding theorem. The following paragraphs briefly explain how to map a d-dimensional dynamical system to a d_{E}-dimensional phase space, where $d_{\mathrm{E}} \geq 2d + 1$.

A d-dimensional dynamical system can be expressed by d first-order differential equations. The solution of these equations, $\mathbf{s} \in \Re^d$, is a state in the corresponding state space, with $\Re$ denoting the Euclidean space. Function $h : \Re^d \rightarrow \Re$ converts a collection of states $\mathbf{s}$ to a scalar time series, *i.e.* $x = h(\mathbf{s})$, where h is called the measure function. The 'delay' of the time series is denoted by a positive number τ. The evolution of the state $\mathbf{s}$ at time i is defined by the function $F_\tau(\mathbf{s}_i) = \mathbf{s}_{i+\tau}$.

Therefore, the embedding $\Phi : \Re^d \rightarrow \Re^{d_{\mathrm{E}}}$, which is called the delay-coordinate embedding here, is defined as:

$$\Phi(h, F, \tau)(\mathbf{s}) = \{h(\mathbf{s}), h(F_\tau(\mathbf{s})), \ldots, h(F_{(d_{\mathrm{E}}-1)\tau}(\mathbf{s}))\}. \tag{6.2.1}$$

For a certain $\mathbf{s}_i$, the above equation has the form of:

$$\begin{aligned} \Phi(h, F, \tau)(\mathbf{s}_i) &= \{h(\mathbf{s}_i), h(\mathbf{s}_{i+\tau}), \ldots, h(\mathbf{s}_{i+(d_{\mathrm{E}}-1)\tau})\} \\ &= \{x_i, x_{i+\tau}, \ldots, x_{i+(d_{\mathrm{E}}-1)\tau}\} = \mathbf{x}_i. \end{aligned} \tag{6.2.2}$$

Therefore, a phase space matrix X of dimension d_{E} and delay τ can be constructed in the following way:

$$\begin{aligned} X &= \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_M \end{bmatrix} = \begin{bmatrix} x_1 & x_{1+\tau} & \cdots & x_{1+(d_{\mathrm{E}}-1)\tau} \\ x_2 & x_{2+\tau} & \cdots & x_{2+(d_{\mathrm{E}}-1)\tau} \\ \vdots & \vdots & \ddots & \vdots \\ x_M & x_{M+\tau} & \cdots & x_{M+(d_{\mathrm{E}}-1)\tau} \end{bmatrix} \\ &= \begin{bmatrix} \mathfrak{x}_1 & \mathfrak{x}_2 & \cdots & \mathfrak{x}_{d_{\mathrm{E}}} \end{bmatrix} \end{aligned} \tag{6.2.3}$$

where the column vectors $\mathfrak{x}_i$ $(i = 1, \ldots, d_E)$ form the coordinate of each dimension and the row vectors $\mathbf{x}_i$ $(i = 1, \ldots, M)$ represent individual points in the phase space. For a time series $x = \{x_1, \ldots, x_i, \ldots, x_N\}$, it can be embedded directly from $\Re$ to $\Re^{d_E}$ by (6.2.3). In this case, X is a $M \times d_E$ matrix, $\mathbf{x}$ is a $1 \times d_E$ vector and $M = N - (d_E - 1)\tau$. The M points form the d_E-dimensional embedded signal in the phase space. Matrix X is also called the trajectory matrix.

6.2.1 Determination of the Embedding Dimension

Two parameters are required by the delay coordinate embedding: the embedding dimension and the time delay. These parameters should be properly chosen so that the feature hidden in the time series can be presented in the phase space. According to [91], a suggested embedding dimension is $d_E = \ulcorner 2 \cdot \text{boxdim}(A) + 1 \urcorner$, where A is the attractor of the dynamical system, boxdim(A) is the system dimension, and $\ulcorner x \urcorner$ denotes the minimum integer larger than or equal to x. Note that boxdim(A) may be fractional and d_E must be integral. The correlation dimension [92][93] is used as boxdim(A) in this thesis to calculate d_E.

The correlation dimension is determined from the correlation integral defined as:

$$C(r) = \lim_{N\to\infty} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \theta(r - |\mathbf{x}_i - \mathbf{x}_j|) \quad (i \neq j) \tag{6.2.4}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are two arbitrary points, and $\theta(x)$ is the Heaviside step function:

$$\theta(x) = \begin{cases} 0, & \text{if } x \leq 0 \\ 1, & \text{if } x > 0 \end{cases} . \tag{6.2.5}$$

Equation (6.2.4) calculates the number of pairs $(\mathbf{x}_i, \mathbf{x}_j)$ satisfying $|\mathbf{x}_i - \mathbf{x}_j| < r$. $\mathbf{x}_i$ and $\mathbf{x}_j$ are two arbitrary points in the 2-dimensional phase space: $\mathbf{x}_i = \{x_i, x_{i+1}\}$. Here, $d_E = 2$ and $\tau = 1$ are selected as the initial condition for the calculation of the real correlation dimension, d_c. Assume that for small r, $C(r)$

behaves as $C(r) \propto r^{d_c}$. d_c is then estimated by:

$$d_c = \lim_{r \to 0} \frac{\lg[C(r)]}{\lg[r]}. \tag{6.2.6}$$

In other words, d_c is the slope of the curve of $\lg[C(r)]$ against $\lg[r]$.

The correlation dimension depends on the value of r. When r is small, the behaviour of the correlation dimension is dominated by the characteristics of noise, which has infinite dimension [94]. In practice, a string of d_c are calculated with various r. If in a range of $r_L \leq r \leq r_U$, $d_c(r)$ is a constant within some tolerance, the correlation dimension is chosen as the average of $d_c(r)$ over $[r_L, r_U]$ [93]. Finally, the embedding dimension is $d_E = \lceil 2d_c + 1 \rceil$.

6.2.2 Determination of the Delay Constant

The delay constant should neither be too short to include unnecessary calculation, nor too long to miss any useful information. If τ is too small, each coordinate is almost the same and the trajectories of the phase space are squeezed along the identity line. On the other hand, if τ is too large, in the presence of chaos and noise, the dynamics at a time instant become effectively and causally disconnected from the dynamics at a later time instant. Hence, even simple geometric objects look extremely complicated. A natural choice of τ is the first minimum of the autocorrelation function such that each coordinate is linearly independent.

According to [95], τ equals to the first minimum of $\lg C_1^{d_E}$, where

$$C_1^{d_E} = \lim_{q \to 1} C_q^{d_E} \tag{6.2.7}$$

$$C_q^{d_E} = \left(\frac{1}{\hat{l}} \sum_{i=1}^{\hat{l}} (P_r(\mathbf{x}_i(\tau)))^{q-1} \right)^{\frac{1}{q-1}} \tag{6.2.8}$$

$$P_r(\mathbf{x}_i(\tau)) = \frac{1}{\hat{l}} \sum_{j=1}^{\hat{l}} \theta(r - |\mathbf{x}_i(\tau) - \mathbf{x}_j(\tau)|) \tag{6.2.9}$$

$$\hat{l} = l - (d_E - 1)\tau$$

$$\mathbf{x}_i(\tau) = \{x_i, x_{i+\tau}, \ldots, x_{i+(d_e-1)\tau}\}.$$

6.3 Feature Waveform Detection and Classification of ECG Signals

6.3.1 Parameter Setting

ECG signals used in this section are free from impulsive noise and baseline wander, and are pre-processed to get rid of high frequency noise. The filtering techniques are explained in chapters 3 and 4, respectively. For such a de-noised test ECG signal of length 1000, the relationship between $\lg r$ and $\lg C(r)$ is plotted in Fig. 6.1. In this case, $d_c = 0.4390$. Hence, $d_E = 2$. For the same test ECG signals of length 3000, the embedding dimension is $d_E = 3$. Thus, the choice of d_E depends on the length of the signal under processing.

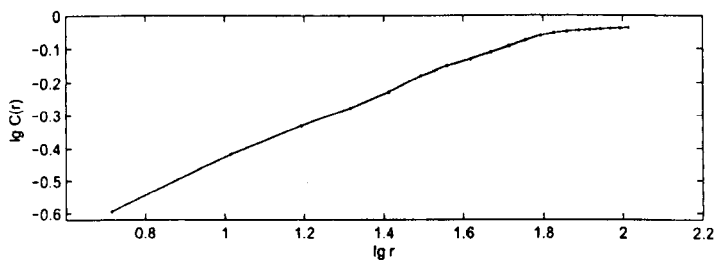


Figure 6.1: $\lg C(r)$ against $\lg r$ for a de-noised test ECG signal when initial conditions are $d_E = 2$ and $\tau = 1$.

For the de-noised test ECG signal of length 1000, the relationship between $\lg C_1^{d_E}(\tau)$ and τ is given in Fig. 6.2 and the minimum of $\lg C_1^{d_E}$ occurs at $\tau = 15$.

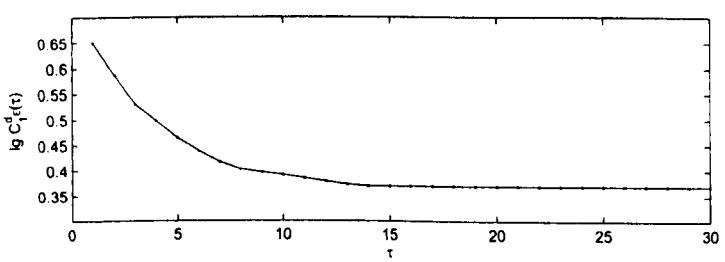


Figure 6.2: $\lg C_1^{d_E}(\tau)$ against τ of a de-noised test ECG signal.

Figure 6.3 shows the embedding of the de-noised test ECG signal when $d_E = 2$ and $\tau = 15$. The white pixels in the figure are enlarged to give a clearer view. As the magnitude of the test signal is digitised, the magnitude of the de-noised test signal is also regulated to integers. Thus, the phase space can be presented by an image and the embedded signal can be presented by the white pixels of the image. The dense cluster corresponds to samples belonging to the segments, and the three orbits correspond to the P waves, T waves and QRS complexes, respectively, according to the size. For the original test signal, the embedding parameters are $d_E = 2$ and $\tau = 8$. The smaller value of τ shows that the signal contains a greater amount of redundant information, which is introduced by the noise, and more unnecessary calculation is included. Thanks to the noise removal procedure, the redundant calculation is avoided.

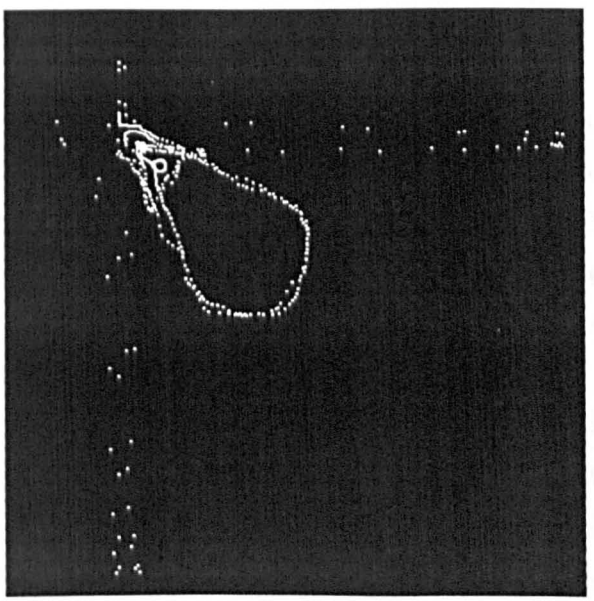


Figure 6.3: Embedding of a de-noised test ECG signal when $d_E = 2$ and $\tau = 15$.

6.3.2 Detection of QRS Complexes

As explained in section 3.1.2, the QRS complexes are distinctive in an ECG signal due to the high magnitude of the R waves. In order to locate the QRS complexes with the tolerance of the baseline wander, the magnitude difference between every τ samples are calculated:

$$d_r(t) = f(t) - f(t+\tau). \tag{6.3.1}$$

The difference signal of d_l is referred to as the right-hand difference. Local minima of d_l with an absolute magnitude value over a pre-set threshold are recorded as the Q points, and local maxima with a magnitude value over the pre-set threshold are recorded as the R points. Similarly, the S points can be located from the left-hand difference of f:

$$d_l(t) = f(t) - f(t-\tau). \tag{6.3.2}$$

Figure 6.4 demonstrates the detection of the QRS complexes of a test signal. Figure 6.5 shows the detection result during a 3-beat ventricular tachycardia. The algorithm has been tested on a group of ECG signals and the detection results are listed in Table 6.1. The ECG signals contains only records in the first five minutes. The number of the QRS complexes detected by the algorithm, *i.e.* the detected beats, is compared with the actual number provided by the database. The errors mainly occur during fusion, which changes the QRS. Take record 208 for example. In the first five minutes, the ECG signal contains 72 fusions, which reduces the accuracy of the detection by 2.7%. On the identification of the R points, an RR interval is considered as a 'cycle' of the signal.

6.3.3 Detection of P Waves and T Waves

The detection of P waves and T waves is based on the embedded signal in the phase space. Samples belonging to the segments form a dense cluster, which can be extracted from the image using morphological reconstruction by

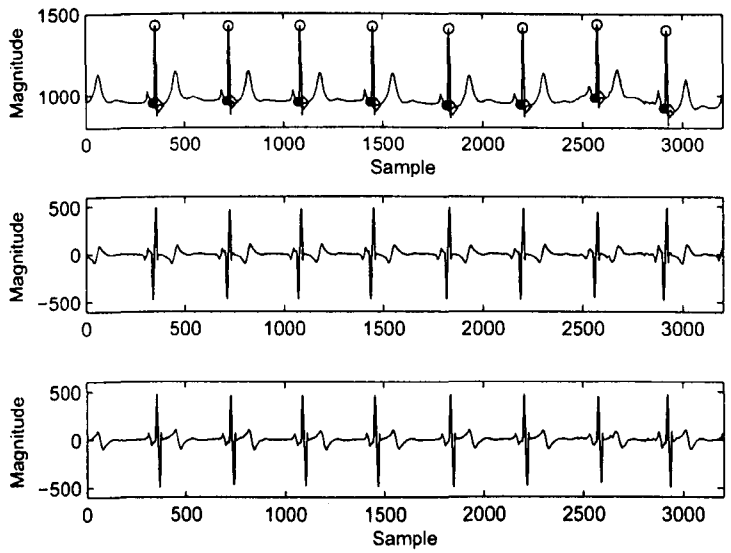


Figure 6.4: (a) Location of the Q points (in •), R points (in ○) and S points (in ◇). (b) Right-hand difference. (c) Left-hand difference.

dilation. Morphological reconstruction requires two images: a mask image and a marker image. The algorithm of morphological reconstruction by dilation of a mask image f_1 from a marker image f_2 ($f_2 \leq f_1$) is described as follows [20].

- Step 1: Set $k := 1$. Set $f_2^{(0)} = f_2$.
- Step 2: Dilate f_2: $f_2^{(k)} = f_2^{(k-1)} \oplus g$, where g is the pre-defined SE.
- Step 3: Calculate the point-by-point minimum of $f_2^{(k)}$ and f_1: $\delta^{(k)}(t) = \min\{f_2^{(k)}(t), f_1(t)\}$.
- Step 4: If $\delta^{(k)} = \delta^{(k-1)}$, terminate. Otherwise, set $k := k + 1$ then go to Step 2.

For a binary image, the intersection of $f_2^{(k)}$ and f_1 is used in step 3. Here, the mask image is the embedded signal, X, and the marker image is the erosion of

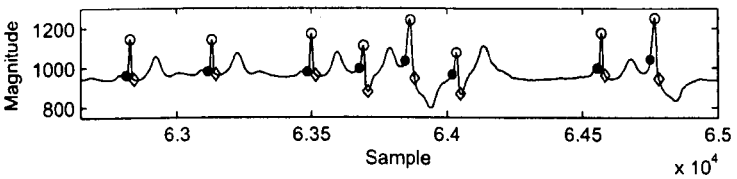


Figure 6.5: Location of the Q points (in •), R points (in ○) and S points (in ◇) during a 3-beat ventricular tachycardia.

Table 6.1: Results of QRS detection

ECG record	Total beats	Detected beats	Accuracy rate
105	417	417	100%
109	433	432	99.77%
200	433	433	100%
208	518	504	97.30%
212	463	463	100%
217	363	354	97.52%
221	407	407	100%
228	350	350	100%

X by an SE of

$$g = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

as shown in Fig. 6.6(a). The SE can also be set as $[0,1,0;1,1,1;0,1,0]$, $[0,1;1,1]$, $[1,1;1,0]$ or $[1,0;0,1]$ to match the shape of the object. The result of morphological reconstruction by dilation is given in Fig. 6.6(b). Denote it by X_b. Apparently, $X_b \subseteq X$.

To separate a P or T wave from the segment, the distance from each objective pixel to the set of X_b is calculated. The distance from a pixel $\mathbf{x}$ to a set

Table 6.2: Measurement of the detected P waves

ECG	Width of P waves (ms)		Peak value of P waves (mV)	
record	Average	Variance	Average	Variance
105	172.83	12.40	0.21	0.0025
109	338.95	11.62	0.29	0.0178
200	100.75	3.09	0.42	0.0146
208	132.62	10.21	0.14	0.0112
212	119.50	2.18	0.24	0.0014
217	43.94	0.43	0.15	0.0115
221	180.38	2.10	0.94	0.0056
228	298.67	7.74	0.22	0.0854

A is defined as the shortest distance from $\mathbf{x}$ to a pixel in A:

$$d(\mathbf{x}, A) = \min_{p}\{\| \mathbf{x} - \mathbf{a}_p \|\}, \ \mathbf{a}_p \in A. \tag{6.3.3}$$

From a detected Q point leftwards, a string of pixels whose distance to X_b is larger than zero are recorded as the embedding of the P wave, as shown in Fig. 6.6(c). From the detected S point rightwards, a string of pixels whose distance to X_b is larger than zero are recorded as the embedding of the T wave, as shown in Fig. 6.6(d). To view it more clearly, the white pixels in Figs. 6.6(c) and (d) are enlarged. Therefore, samples in the time domain that comprise the embedded P wave and T wave are detected to form the feature waveforms of P wave and T wave. The result is given in Fig. 6.7. In clinic, doctors are interested in the width and peak value of feature waveforms. Table 6.2 shows the measurement of the width and peak value of P waves of eight ECG signals, which are the same as the ones used in the previous table. Data of P and T waves can be classified using the geometric information in the phase space, such as centroid, barycentre, area, *etc.*

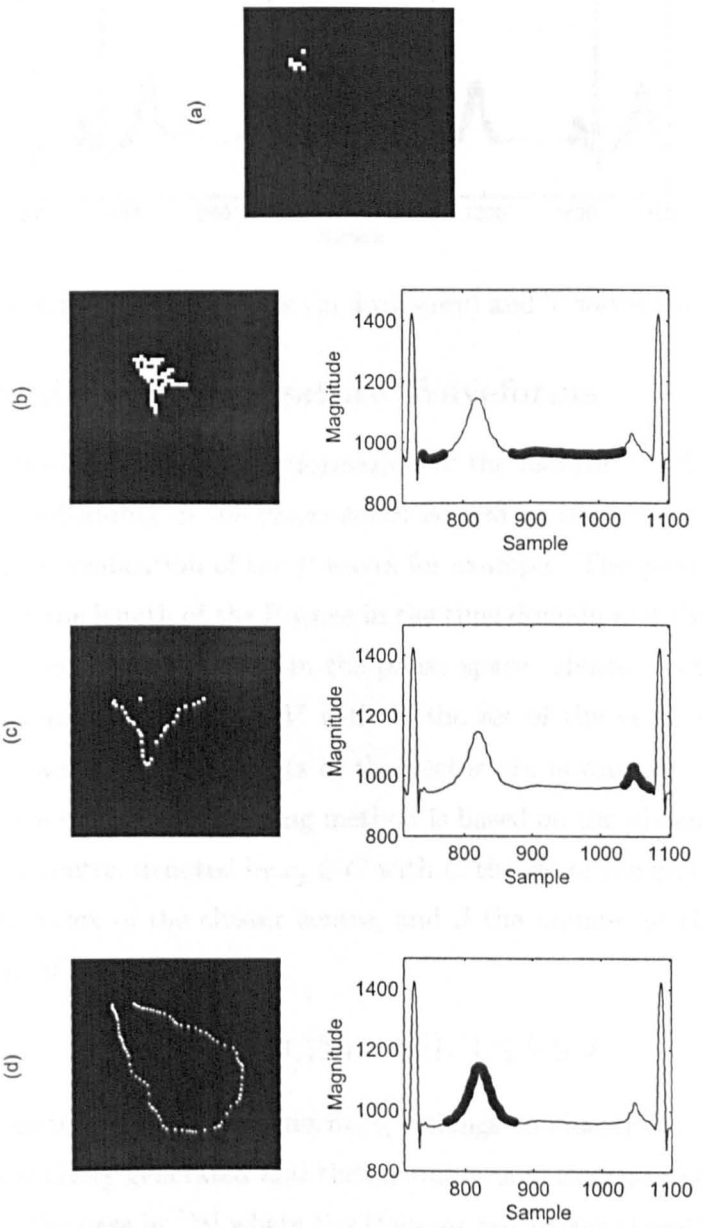


Figure 6.6: (a) The marker image. (b) Morphological reconstruction by dilation and the extracted baseline in the time domain. (c) Extracted P wave in the phase domain and time domain, respectively. (d) Extracted T wave in the phase domain and time domain, respectively.

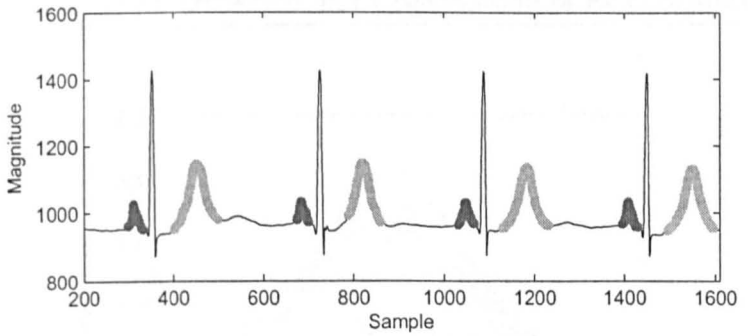


Figure 6.7: Identification of P waves (in dark grey) and T waves (in light grey).

6.3.4 Classification of Feature Waveforms

In this section, the geometric information of the feature waveforms represented by the embedding in the phase space is used as the criteria to classify them. Take the classification of the P waves for example. The geometric information includes the length of the P wave in the time domain and the perimeter and area of the embedded P wave in the phase space. Hence, each P wave is coded as a 3-element vector $v_i \in V$ with V the set of the vectors and i the index of the P waves. The elements of the vector are normalised to $[0, 1]$ for meaningful clustering. The clustering method is based on the distance between v_i and a cluster centre, denoted by $c_j \in C$ with C the set of the cluster centres, $1 \leq j \leq J$ the index of the cluster centre, and J the number of clusters. For an arbitrary v_i, if

$$\|v_i - c_k\| = \min_{1 \leq j \leq J} \{\|v_i - c_j\|\}, \ 1 \leq k \leq J \tag{6.3.4}$$

where $\|\cdot\|$ denotes the Euclidean norm, v_i belongs to cluster c_k. The cluster centres are adaptively generated and their number is not constrained to a preset number, as the case in [28] where the P waves are always classified into two clusters. The classification procedure is described as follows.

- Step 1: Set $J := 1$. Initialise the first cluster centre $c_1 := v_1$. Set $i := 1$.
- Step 2: Calculate the distance $d_j = \|v_i - c_j\|$ for all $1 \leq j \leq J$. If $\min_j\{d_j\}$ exceeds a pre-defined threshold, θ, a new cluster centre is assigned at

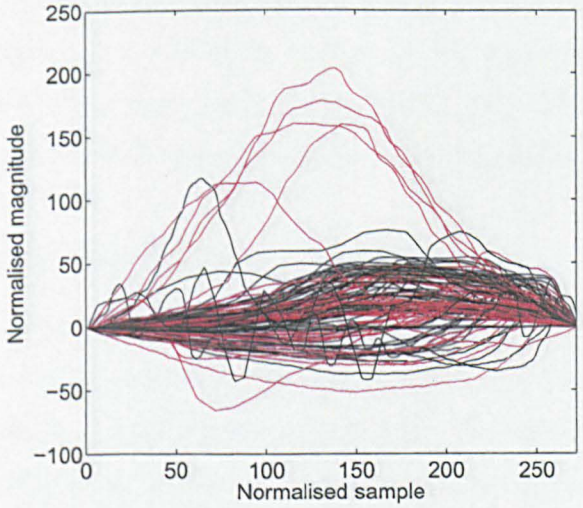


Figure 6.8: Classification of P waves of an ECG signal ($\theta = 1$). Normalised P waves in black are classified to cluster 1 and those in red are classified to cluster 2.

$c_{J+1} = v_i$. Set $J := J + 1$. Otherwise, v_i is classified to cluster c_k if $d_k = \min_j\{d_j\}$.

- Step 3: Update the cluster centres by averaging the vectors belonging to each c_j.
- Step 4: Set $i := i + 1$. Go to Step 2. Terminate when all the vectors are classified.

Figure 6.8 shows the classification result of the P waves of the first five minutes of record 106. To make it more comprehensible, the P waves are normalised taking on zero value at the onset and end samples, as the strategy used in [28], and having the same length. The threshold is set at $\theta = 1$ and under this condition, the P waves are classified into two clusters. As a comparison, when $\theta = 0.8$, the P waves are classified into four clusters, as shown in Fig. 6.9. The same clustering method can also be applied to classify the QRS complexes and T waves.

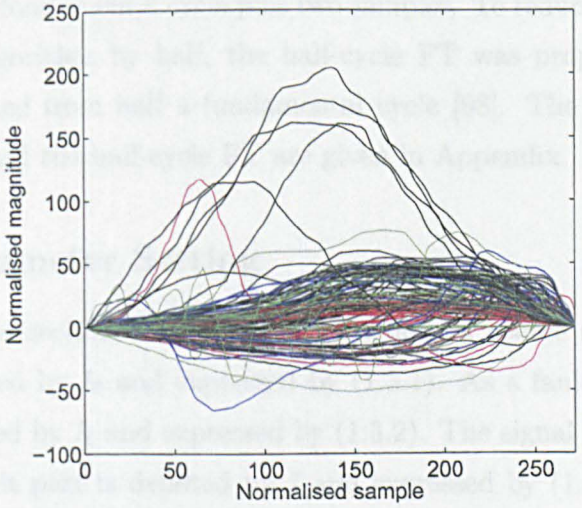


Figure 6.9: Classification of P waves of an ECG signal ($\theta = 0.8$). Normalised P waves are classified into four clusters, plotted in black, red, green and blue, respectively.

6.4 Phasor Measurement of Power System Signals

Phasor measurement measures the amplitude and phase angle of the signal to determine the health of the power system. If a fault occurs on transmission lines, the input current of a relay may contain harmonics and exponentially decaying DC offset. This section proposes an embedding-based scheme to measure the amplitude and phase angle of the fundamental component of the fault signal.

A traditional method for phasor measurement is the Fourier transform (FT). However, the presence of the offset will bring fairly large errors to the measurement result and cause malfunction of relays. Hence, it is necessary to preprocess the current signal to remove the DC offset and keep the fundamental frequency component only. As there is no efficient way to remove the DC offset using an FT-based filter, [96] presented a morphological filter to serve this purpose.

Paper [97] presented an algorithm using a full-cycle FT to identify the exponentially decaying DC offset. The algorithm requires a data window with a length of one fundamental cycle plus two samples. To reduce the computation time of the algorithm by half, the half-cycle FT was proposed, which uses samples obtained from half a fundamental cycle [98]. The algorithms of the full-cycle FT and the half-cycle FT are given in Appendix.

6.4.1 Parameter Setting

As stated in section 1.3.2, in power systems, a source current or voltage signal is denoted by I_0 and expressed by (1.3.1). As a fault occurs, the fault signal is denoted by I_f and expressed by (1.3.2). The signal including a source part and a fault part is denoted by I and expressed by (1.3.6). An example of such a signal is given in Fig. 6.10 in dotted line. In this case, up to 15 harmonics are included in the signal, and a Gaussian noise of a signal-to-noise ratio of 15 dB is added.

Embedded to a 2-dimensional phase space, fault signal I forms the following matrix:

$$\mathbf{I} = \begin{bmatrix} I(t_0) & I(t_0+\tau) \\ I(t_1) & I(t_1+\tau) \\ \vdots & \vdots \\ I(t_n) & I(t_n+\tau) \\ \vdots & \vdots \end{bmatrix} \tag{6.4.1}$$

where $t_n = t_0 + n\Delta t$ with t_0 the beginning and Δt the sampling interval. Matrix $\mathbf{I}$ can be considered as a 2-dimensional signal in the phase space with the left column being its x-values and the right column its corresponding y-values.

6.4.2 Fault Occurrence Detection

Due to the features of circular function, the source signal, I_0, has the following embedding when $\tau = T/4$:

$$\begin{aligned} \mathbf{I}_0 &= [\ I_0(t) \quad I_0(t+\tau)\ ] \\ &= [\ x \qquad y\ ] \end{aligned} \tag{6.4.2}$$

where

$$x = A_0 \cos(\omega t + \phi) \tag{6.4.3}$$

$$\begin{aligned} y &= A_0 \cos(\omega(t+\tau) + \phi) \\ &= A_0 \cos(\omega t + \pi/2 + \phi) \\ &= -A_0 \sin(\omega t + \phi). \end{aligned} \tag{6.4.4}$$

As $x^2 + y^2 = A_0^2$, it shows that the source signal forms a circle whose radius is A_0 and whose centre is at $(0,0)$ in the phase space. In other words, a pair of samples, $I_0(t_n)$ and $I_0(t_n+\tau)$, form a point in the phase space and the Euclidean norm of the point is A_0. However, when the fault occurs, the fault part forms some other shape and the Euclidean norm of the point $(I(t_s-\tau), I(t_s))$ suddenly increases. Therefore, a threshold can be set to check if the fault occurs. The embedding of the signal shown in Fig. 6.10 in dotted line is given in Fig. 6.11. The Euclidean norms of the points are plotted in Fig. 6.10 in solid line, and the estimated fault occurrence point is highlighted by a dot. In this case, it is simulated that the fault occurs at $t_s = t_{87}$ and the detection result is also $\hat{t_s} = t_{87}$.

The morphological filter proposed in [96] is employed to remove the DC offset. As the filtering technique is introduced in detail in [96], it is omitted in this thesis. In the following subsections, the estimation of amplitude and phase angle is based on signals whose DC offset is removed by the morphological filter. An example is given in Fig. 6.12.

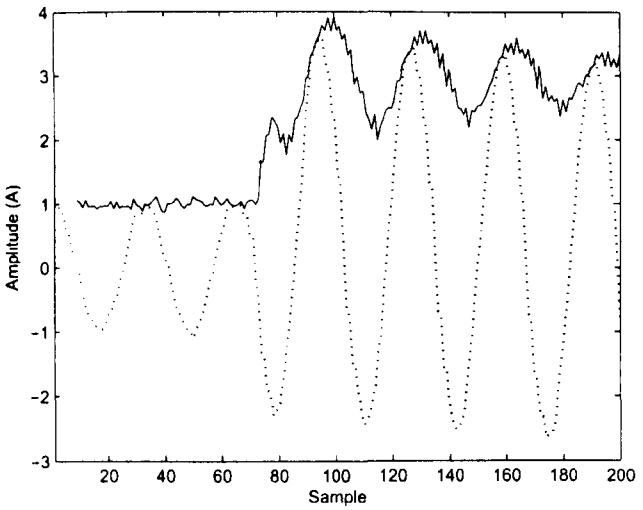


Figure 6.10: Fault occurrence detection. Dotted line: the fault signal. Solid line: the Euclidean norm. Dot: estimated fault occurrence point.

6.4.3 Fundamental Amplitude Estimation

For the fundamental component, $I_1(t)$, its embedding in the phase space can be expressed by:

$$x = I_1(t) = A_1 \cos(\omega t + \varphi) \tag{6.4.5}$$

$$\begin{aligned} y &= I_1(t+\tau) = A_1 \cos(\omega(t+\tau) + \varphi) \\ &= A_1 \cos(\omega t + \varphi + \theta) \qquad (6.4.6) \\ &= A_1 \cos\theta \cos(\omega t + \varphi) - A_1 \sin\theta \sin(\omega t + \varphi) \end{aligned}$$

where $\theta = \omega\tau = 2\pi\tau/T$. From (6.4.5) and (6.4.6), the amplitude of the fundamental component can be calculated from:

$$\hat{A}_1 = \sqrt{(x^2 - 2\cos\theta xy + y^2)/\sin^2\theta} \tag{6.4.7}$$

where $\hat{A}_1$ denotes the estimation of A_1. Equations (6.4.5)~(6.4.7) show that any two samples from the fault current are enough to estimate its amplitude. Usually τ is selected to make θ a common angle. For example, when $\tau = T/8$, $\theta = \pi/4$. In this case, as the signal is digitised at $N = T/\Delta t$ samples per cycle,

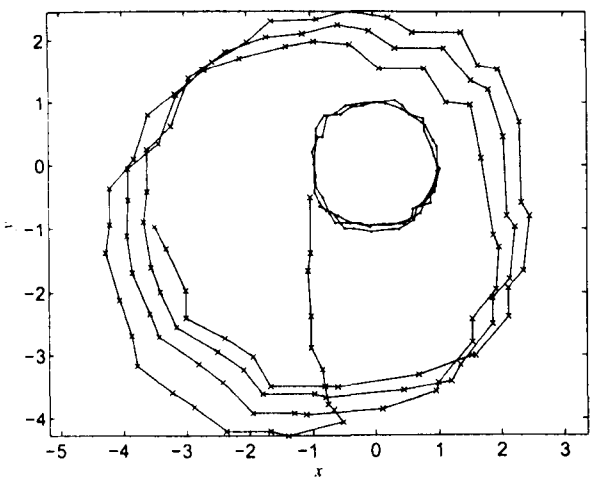


Figure 6.11: Embedding of a fault signal. Dots: embedding of the source part. Crosses: embedding of the fault part.

it uses $N/8$ samples to calculate A_1. When $\tau = T/4$, we have $\theta = 2\pi\tau/T = \pi/2$ and the embedding of I_1 becomes

$$\begin{aligned} x &= I_1(t) = A_1 \cos(\omega t + \varphi) \\ y &= I_1(t+\tau) = -A_1 \sin(\omega t + \varphi). \end{aligned}$$

The estimation therefore becomes:

$$\hat{A}_1 = \sqrt{x^2 + y^2}. \tag{6.4.8}$$

Since

$$(A_1 \cos(\omega t + \varphi))^2 + (-A_1 \sin(\omega t + \varphi))^2 = A_1^2 \tag{6.4.9}$$

the embedded signal of I_1 forms a circle in the phase space, and the centre and the radius of the circle are the origin of the phase space and A_1, respectively.

In practice, due to the presence of noise and harmonics, the estimated amplitude calculated from different pairs of samples varies slightly:

$$\hat{A}_{1n} = \sqrt{x_n^2 + y_n^2} = \sqrt{I^2(t_n) + I^2(t_n + \tau)} \tag{6.4.10}$$

where $\hat{A}_{1n}$ denotes the estimation of A_1 calculated from $x_n = I(t_n)$ and $y_n = I(t_n + \tau)$. Assume that the onset of the fault is detected at time $\hat{t}_s$. From the

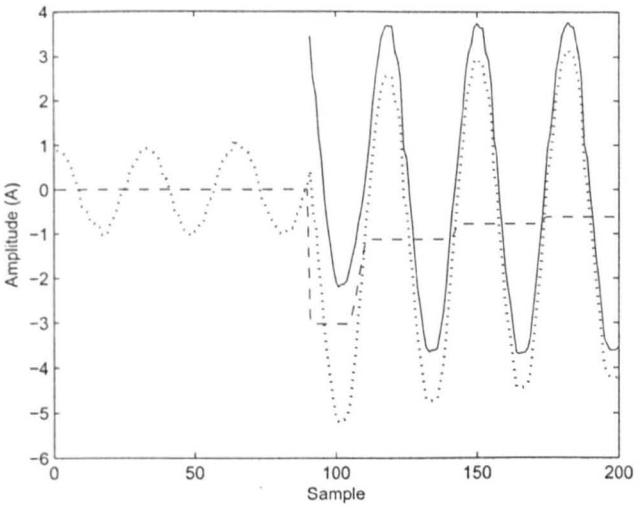


Figure 6.12: Dotted line: the fault current. Dashed line: extracted DC offset. Solid line: the result of removing the DC offset from the fault current.

time instant of $t_n = \hat{t}_s + \Delta t$, the estimated amplitude is calculated using (6.4.7), where $\theta = \omega(t_n - \hat{t}_s)$, until $t_n = \hat{t}_s + \tau - \Delta t$. In this manner, the estimation of the fault current amplitude is not affected by the source current amplitude, except the amplitude at time instant $\hat{t}_s$. In order to reduce the estimation error caused by noise, the average of $\hat{A}_{1n}$ over a user-defined window can be used. An example is given in Fig. 6.13. A fault signal that has its DC offset removed is considered as the input signal, as plotted in dotted line. The signal-to-noise ratio of the fault signal is 10.67 dB. Using the method described above to estimate the amplitude of its fundamental component, the result is given in solid line. As it can be seen, the sudden increase in amplitude has been successfully estimated as the fault occurs. The vibration caused by the noise is avoided by taking the average of the estimated amplitude. As a comparison, the estimation result by the half-cycle FT is also included, which is shown in dashed line. The half-cycle FT uses the samples over a half of a cycle to calculate the amplitude. Hence, it causes a half-cycle delay to accurately estimate the amplitude of the fault signal. On the other hand, the embedding-based method does not introduce such an error in the estimation.

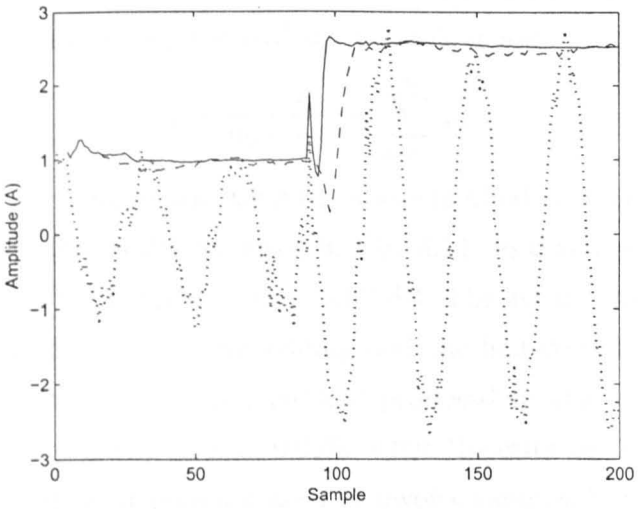


Figure 6.13: Estimation of the fundamental amplitude of the fault current. Dotted line: the fault current with its DC offset removed. Solid line: estimation result by embedding. Dashed line: estimation result by half-cycle FT.

6.4.4 Phase Estimation

As stated previously, when $\tau = T/4$, the embedded fundamental component can be defined by:

$$\mathbf{I}_1 = \begin{bmatrix} A_1\cos(\omega t_0 + \varphi) & -A_1\sin(\omega t_0 + \varphi) \\ \vdots & \vdots \\ A_1\cos(\omega t_n + \varphi) & -A_1\sin(\omega t_n + \varphi) \\ \vdots & \vdots \end{bmatrix}. \tag{6.4.11}$$

When $t_0 = 0$, the phase of the fundamental component can be calculated from the first point of $\mathbf{I}_1$:

$$\hat{\varphi} = \arctan\left(-\frac{\mathbf{I}_1(1,2)}{\mathbf{I}_1(1,1)}\right). \tag{6.4.12}$$

The phase can also be calculated from an arbitrary point of $\mathbf{I}_1$ using:

$$\hat{\varphi}_n = \arctan\left(-\frac{\mathbf{I}_1(n,2)}{\mathbf{I}_1(n,1)}\right) - \mathrm{mod}(\omega t_n, 2\pi) \tag{6.4.13}$$

where $\mathrm{mod}(\omega t_n, 2\pi)$ is the residue of ωt_n divided by 2π. This is to make sure that $\hat{\varphi}_n$ falls in the range of $[0, 2\pi)$. Ideally, for different n, $\hat{\varphi}_n$ should be the

same. In practice, it is more accurate to use the average of $\hat{\varphi}_n$ over a certain window to eliminate the influence of any possible noise:

$$\hat{\varphi} = \frac{1}{n_2 - n_1 + 1} \sum_{n=n_1}^{n_2} \hat{\varphi}_n. \tag{6.4.14}$$

The test is also carried out using the above method and the half-cycle FT, respectively, and the results are given in Fig. 6.14. In this case, the signal-to-noise ratio of the fault signal is also 10.67 dB. The actual phase is 78.7500°, and the estimated results by embedding and the half-cycle FT are 79.4820° and 79.7806°, respectively. The method proposed in this thesis is slightly better with an estimation error of 0.93%, while the error of the half-cycle FT is 1.31%. However, as it does not need to involve samples from a whole cycle in calculation, the proposed method is more computationally efficient.

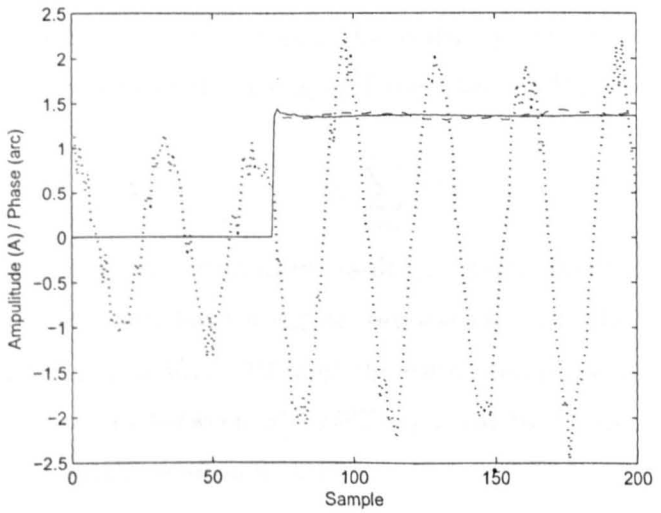


Figure 6.14: Estimation of the phase of the fault current. Dotted line: the fault current with its DC offset removed. Solid line: estimation result by embedding. Dashed line: estimation result by half-cycle FT.

6.4.5 Phase Difference Detection

In some cases, there is a phase difference between the current signal and the voltage signal. The difference can be detected through the embedding of the

two signals to the phase space with $\tau = T/4$. Define the current and voltage signals as follows:

$$f_c = A_c \cos(\omega t + \varphi_c) \tag{6.4.15}$$

$$f_v = A_v \cos(\omega t + \varphi_v) \tag{6.4.16}$$

where A_c and A_v are their amplitudes and φ_c and φ_v are their phases, respectively.

When $\tau = T/4$, an arbitrary point from the embedded current signal has the coordinates of $(x_{cn}, y_{cn}) = (A_c \cos(\omega t_n + \varphi_c), -A_c \sin(\omega t_n + \varphi_c))$ and has a phase angle of $\hat{\varphi}_{cn} = \arctan(-y_{cn}/x_{cn})$. At the same sampling time, the point from the embedded voltage signal has the coordinates of $(x_{vn}, y_{vn}) = (A_v \cos(\omega t_n + \varphi_v), -A_v \sin(\omega t_n + \varphi_v))$ and has a phase angle of $\hat{\varphi}_{vn} = \arctan(-y_{vn}/x_{vn})$. Therefore, the phase difference between the two signals, denoted by $\Delta\varphi$, can be calculated from $\hat{\Delta\varphi} = \hat{\varphi}_{cn} - \hat{\varphi}_{vn}$. Considering the interference of noise, it is more accurate to use the average of the phase differences over a certain window:

$$\hat{\Delta\varphi} = \frac{1}{n_2 - n_1 + 1} \sum_{n=n_1}^{n_2} (\hat{\varphi}_{cn} - \hat{\varphi}_{vn}). \tag{6.4.17}$$

Figure 6.15 shows the estimation result of a test, where the input voltage and current signals both have a signal-to-noise ratio of 10.67 dB. The simulated phase difference is 307.7440° and the estimated phase difference at each sampling instant varies between 306.9407° and 310.1625°, as the figure shows. The average estimation error is 0.12%.

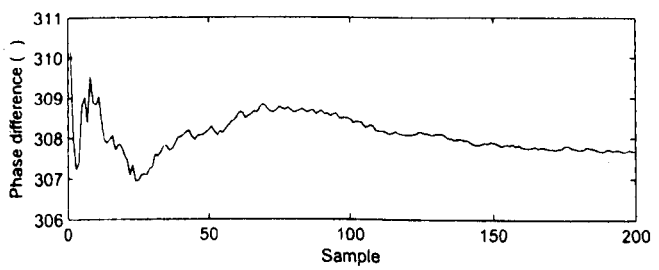


Figure 6.15: Estimation of the phase difference.

6.4.6 Fundamental Frequency Shift Estimation

In some cases, power system signals are influenced by fundamental frequency shift, which can be up to ±5%. A considerable error would be introduced to the result of phasor measurement if the signal is embedded according to its nominal period. To estimate the actual fundamental frequency, a method is proposed as follows. The signal is first embedded with $\tau = T/2$, where T is the nominal period. Hence, the embedding in the phase space is expressed by:

$$x = I_1(t) = A_1 \cos(\omega' t + \varphi) \tag{6.4.18}$$

$$\begin{aligned} y &= I_1(t+\tau) = A_1 \cos(\omega'(t + T/2) + \varphi) \\ &= A_1 \cos(\omega' t + \varphi + \frac{\pi\omega'}{\omega}) \end{aligned} \tag{6.4.19}$$

where ω' denotes the actual fundamental frequency. As the function of

$$F(t) = x + y = 2A_1 \cos(\frac{\pi\omega'}{2\omega}) \cos(\omega' t + \varphi + \frac{\pi\omega'}{2\omega}) \tag{6.4.20}$$

forms a sinusoidal signal, ω' can be calculated from

$$\omega' = 2\omega \arccos(\pm\rho/2A_1)/\pi \tag{6.4.21}$$

where ρ is the amplitude of F and is calculated from

$$\rho = (F \oplus g - F \ominus g)/2 \tag{6.4.22}$$

where $\oplus$ and $\ominus$ denote the operators of dilation and erosion, respectively, and g is a flat structuring element of a cycle long. An example is given in Fig. 6.26 to show the calculation of ρ. In (6.4.21), if $\omega' > \omega$, $-\rho$ is used; otherwise, ρ is used.

To determine if ω' is larger or smaller than ω, the following strategy is used. As stated in section 6.4.4, if there is no fundamental frequency shift, the estimated phase, $\hat{\varphi}_n$, is a flat function. However, if the fundamental frequency has a positive variation, *i.e.* $\omega' > \omega$, $\hat{\varphi}_n$ increases gradually. On the other hand, if $\omega' < \omega$, $\hat{\varphi}_n$ decreases. For the fault current shown in Fig. 6.26, the actual fundamental frequency is $f' = 51.2542$ Hz. Estimating the phase when

the delay coordinate is still set at $\tau = T/4$, where T is the nominal period, the value of $\hat{\varphi}_n$ is given in Fig. 6.17. As $\hat{\varphi}_n$ is basically an increasing function, it is considered that $\omega' > \omega$ and $-\rho$ is used in (6.4.21).

According to (6.4.21), the estimation of ω' also requires the value of A_1. The estimation process is the same as described in section 6.4.3, and the final result of $\hat{A}_1$ is calculated as the average of $\hat{A}_{1n}$ over a cycle. In this example, $A_1 = 3.0102$ and $\hat{A}_1 = 3.0124$. The estimated fundamental frequency is $\hat{f}' = 51.2794$ Hz and the estimation error is 4.92×10^{-4}. On the estimation of the fundamental frequency, the fault current is re-embedded according to the estimated actual fundamental frequency and phasor measurement can be carried out afterwards. As samples from at least a cycle of the fault current are used to estimate the fundamental frequency, the method causes a delay of at least a cycle at the very beginning.

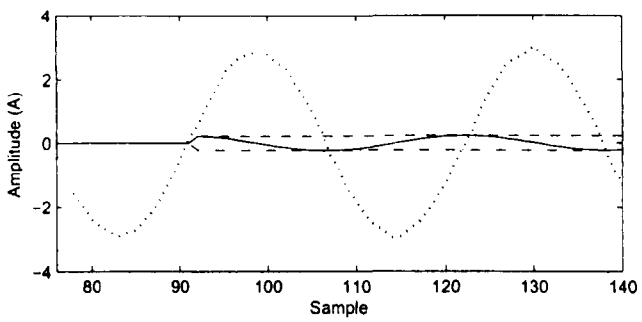


Figure 6.16: Calculation of ρ. Dotted line: the fault current with its DC offset removed. Solid line: the signal of F. Dashed line: dilation of F. Dash-dot line: erosion of F.

6.5 Power System Disturbance Detection

6.5.1 Parameter Setting

Disturbance signals simulated in this section contain up to 40 harmonics and the total harmonic distortion (THD) is around $2 \sim 5\%$. For one of such signals, the relationship between $\lg r$ and $\lg C(r)$ is plotted in Fig. 6.18. Hence,

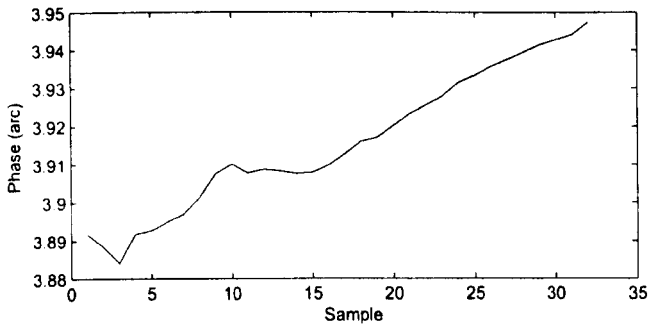


Figure 6.17: Estimation of the phase of the fault current ($f' = 51.2542$) when $\tau = T/4$.

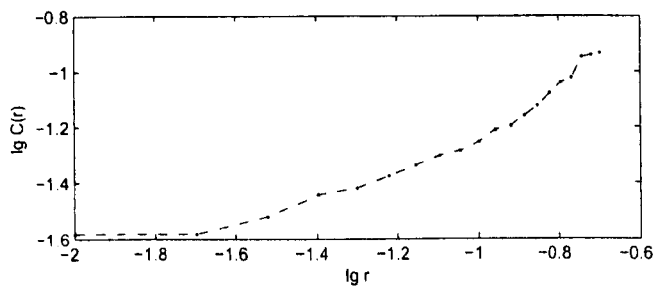


Figure 6.18: $\lg C(r)$ against $\lg r$ for a power system signal when initial conditions are $d_E = 2$ and $\tau = 1$.

the correlation dimension is calculated to be $d_c = 0.7770$ and the embedding dimension is $d_E = 3$.

For a sinusoidal signal $F(t) = A\sin(\omega t + \phi)$, the corresponding sampled signal is given by $F(k) = A\sin(\omega k\Delta t + \phi)$, if it is sampled at a sampling rate of $N_s = 2\pi/(\omega\Delta t)$ samples per cycle. According to the selection criterion described in section 6.2.2, the delay constant, τ, of a disturbance signal should be 1. However, to make full use of the mathematical properties of a sinusoidal signal, the delay constant can be set to $\tau = N_s/4$. In the following subsections, two schemes are proposed for the detection under the two embedding conditions, respectively.

6.5.2 Disturbance Detection and Location through Gustafson-Kessel Clustering

In this subsection, a disturbance signal is transformed to the phase space with $\tau = 1$. Hence, the normal part forms an ellipse and the disturbance forms a shape that deviates from the ellipse. Thus, the two parts are decoupled and distinguished from each other in the phase space. Afterwards, GK clustering [99] is used to distinguish the two clusters and to detect and locate the disturbances accordingly.

GK clustering

The GK clustering algorithm searches in particular for ellipsoidal structures. Assume that a set of data $x = \{x_k | k = 1, 2, \ldots, N\}$ can be partitioned into c clusters. The centres of these clusters are denoted by $\{v_i | i = 1, 2, \ldots, c\}$, respectively. For each cluster, there exists a covariance matrix F_i. Therefore, the following equation

$$(x - v_i)^{\mathrm{T}} F_i^{-1} (x - v_i) = 1 \tag{6.5.1}$$

defines a hyperellipsoid. The length of the j^{th} axis of this hyperellipsoid is given by the j^{th} eigenvalue of F_i, and its direction is given by the j^{th} eigenvector.

The partition of data set x into c clusters is performed by minimising the following objective function:

$$J(X; U, V) = \sum_{i=1}^{c} \sum_{k=1}^{N} (u_{ik})^m \|x_k - v_i\|_{M_i}^2 \tag{6.5.2}$$

where $U = [u_{ik}]$ is the partition matrix that satisfies

$$\sum_{i=1}^{c} u_{ik} = 1, \ 1 \leq k \leq N \tag{6.5.3}$$

and $u_{ik} \in [0, 1]$; equation

$$D_{ikM_i}^2 = \|x_k - v_i\|_{M_i}^2 = (x_i - v_i)^{\mathrm{T}} M_i (x_k - v_i) \tag{6.5.4}$$

describes the distance from a point x_i to a cluster centre v_i; $m \in [1, \infty]$ is the weight exponent that determines the fuzziness of the resulting clusters; $M_i = \det(F_i^{1/n})F_i^{-1}$ is a positive definite symmetric matrix with n denoting the dimension of the data set x.

To perform the GK clustering algorithm, three parameters should be given in advance, which are the number of cluster, $1 < c < N$; the weight exponent, m, mostly $m = 2$; and the termination tolerance, $\varepsilon > 0$. Randomly initialise the partition matrix, $U^{(0)}$, meanwhile it should satisfy (6.5.3). The algorithm can be summarised as follows, where $l = 1, 2, \ldots$ indicates the counter of each iteration [99].

1. Compute the cluster centres

$$v_i^{(l)} = \frac{\sum_{k=1}^{N} (u_{ik}^{(l-1)})^m x_k}{\sum_{k=1}^{N} (u_{ik}^{(l-1)})^m}, \quad 1 \le i \le c;$$

2. Compute the cluster covariance matrices

$$F_i = \frac{\sum_{k=1}^{N} (u_{ik}^{(l-1)})^m (x_i - v_i^{(l)})(x_k - v_i^{(l)})^{\mathrm{T}}}{\sum_{k=1}^{N} (u_{ik}^{(l-1)})^m}, \quad 1 \le i \le c;$$

3. Compute the distances

$$\begin{aligned} M_i &= \det(F_i^{1/n})F_i^{-1} \\ D_{ikM_i}^2 &= (x_i - v_i^{(l)})^{\mathrm{T}} M_i (x_k - v_i^{(l)}) \quad 1 \le i \le c, \ 1 \le k \le N; \end{aligned}$$

4. Update the partition matrix

$$u_{ik}^{(l)} = \frac{1}{\sum_{j=1}^{c} (D_{ikM_i}/D_{jkM_j})^{2/(m-1)}} \quad 1 \le i \le c, \ 1 \le k \le N$$

until $\|U^{(l)} - U^{(l-1)}\| < \varepsilon$.

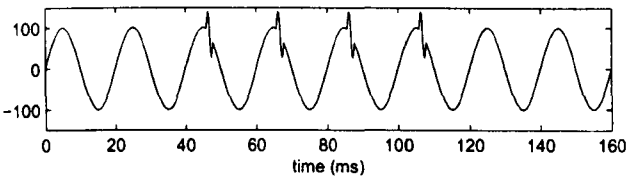


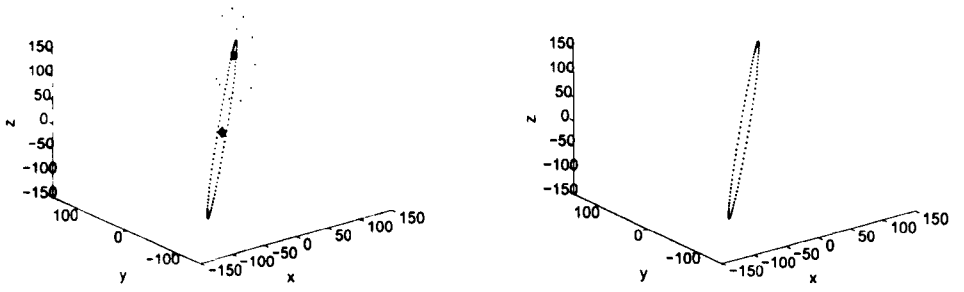
Figure 6.19: A power system signal with periodic notches.

Take a disturbance signal with notching for example. As shown in Fig. 6.19, the fundamental voltage waveform is affected by periodic notches from the third to sixth cycle. The embedded signal is given in Fig. 6.20(a), from which it can be seen that the normal part forms an ellips in the phase space, while the notches form a quasi-ellips, which deviates from the former. Obviously, the GK clustering algorithm is very suitable to distinguish the two waveforms. Applying GK clustering to the embedded signal, the result is shown in Fig. 6.20(a), where the diamond represents the clustering centre of the normal part (v_1) and the square represents that of the disturbance (v_2). According to (6.5.1), samples sufficing $(x_k - v_1)^{\mathrm{T}} F_1^{-1}(x_k - v_1) = 1$ corresponds to the normal part, while those sufficing $(x_k - v_2)^{\mathrm{T}} F_2^{-1}(x_k - v_2) = 1$ corresponds to the disturbance. Figure 6.20(b) shows the samples classified to the cluster of the normal part.

The time-aligned weighted average method

After imposing analysis technique in the phase space, the next step is to transfer the signal back to the time domain. However, once linearly transformed, the resulting trajectory matrix X no longer corresponds to a time delay embedding [100]. In other words, there is no unique map mapping the trajectory matrix back to a one-dimensional signal.

To solve this problem, a time-aligned weighted average method is proposed in [100]. This method gives higher weight to the values in the centre columns of the trajectory matrix and lower weight to the values in the left-most and right-most columns. To describe in detail, the rows of the trajectory matrix are shifted to the right according to the value of τ, which derives the so-



(a) (b)

Figure 6.20: (a) GK clustering of the disturbance signal of notching in the phase space. Diamond: the clustering centre of the normal part. Square: the clustering centre of the disturbance. (b) The samples classified to the cluster of the normal part.

called aligned matrix. Then, the elements near the beginning or end of the aligned matrix are assigned with weights different from those in the middle. To illustrate this process, an example of a trajectory matrix $\hat{X}$, an aligned matrix $\hat{X}_{\text{aligned}}$ and a weighting matrix P are shown in (6.5.5)~(6.5.7), respectively, for the case of $d_{\text{E}} = 3$ and $\tau = 1$.

$$\hat{X}^{\text{T}} = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_M \\ x_2 & x_3 & x_4 & \cdots & x_{M+1} \\ x_3 & x_4 & x_5 & \cdots & x_{M+2} \end{bmatrix} \tag{6.5.5}$$

$$\hat{X}^{\text{T}}_{\text{aligned}} = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_M & & \\ & x_2 & x_3 & \cdots & x_M & x_{M+1} & \\ & & x_3 & \cdots & x_M & x_{M+1} & x_{M+2} \end{bmatrix} \tag{6.5.6}$$

$$P^{\text{T}} = \begin{bmatrix} 1 & 0.5 & 0.25 & \cdots & 0.25 & & \\ & 0.5 & 0.5 & \cdots & 0.5 & 0.5 & \\ & & 0.25 & \cdots & 0.25 & 0.5 & 1 \end{bmatrix} \tag{6.5.7}$$

The output time series is given by:

$$x_{\text{output}}(j) = \sum_{i=1}^{d_{\text{E}}} \hat{X}^{\text{T}}(i,j) \cdot P^{\text{T}}(i,j), \quad j = 1, 2, \ldots, N. \tag{6.5.8}$$

As for the case considered here, since the normal part and the disturbance have been partitioned in the phase space, it is natural to assign zeros to the samples corresponding to the disturbance in matrix X of (6.2.3). It should be noted that a disturbance of length l in the time domain will result in $(l + d_{\mathrm{E}})$ disturbance points in the phase space. Assume that a disturbance occurs at x_5. Hence, in the phase space, points $\mathbf{x}_3 = [x_3, x_4, x_5]$, $\mathbf{x}_4 = [x_4, x_5, x_6]$ and $\mathbf{x}_5 = [x_5, x_6, x_7]$ are the corresponding disturbance points. Assigning 0 to these three points in X, the new matrix, denoted by $\hat{X}_{\mathrm{n}}$, becomes:

$$\hat{X}_{\mathrm{n}}^{\mathrm{T}} = \begin{bmatrix} x_1 & x_2 & 0 & 0 & 0 & x_6 & \cdots & x_M \\ x_2 & x_3 & 0 & 0 & 0 & x_7 & \cdots & x_{M+1} \\ x_3 & x_4 & 0 & 0 & 0 & x_8 & \cdots & x_{M+2} \end{bmatrix}. \tag{6.5.9}$$

Sequentially, the aligned trajectory matrix can be given as follows.

$$\hat{X}_{\mathrm{aligned}}^{\mathrm{T}} = \begin{bmatrix} x_1 & x_2 & 0 & 0 & 0 & x_6 & x_7 & x_8 & \cdots & x_M & & \\ & x_2 & x_3 & 0 & 0 & 0 & x_7 & x_8 & \cdots & x_M & x_{M+1} & \\ & & x_3 & x_4 & 0 & 0 & 0 & x_8 & \cdots & x_M & x_{M+1} & x_{M+2} \end{bmatrix} \tag{6.5.10}$$

In order to transform the aligned trajectory matrix of $\hat{X}_{\mathrm{n}}$ back to the time domain, (6.5.8) is adopted, where the weighting matrix P is the same as in (6.5.7). Denote the result by $\hat{x} = \{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_{M+2}\}$. Obviously, the disturbance occurs at the positions where $\hat{x}_k = 0$. Refer these positions to the input signal x, the disturbance can hereby be extracted.

For the disturbance signal with notching, the detection result is given in Fig. 6.21, where the disturbance-free signal, the input disturbance signal, and the disturbance extracted from the input signal are shown in the subfigures, respectively. As it can be seen from the figure, the disturbance is identically recorded in the output signal; on the contrary, the disturbance-free part maps to 0-value in the output signal. To sum up, the disturbance is precisely extracted from the input signal.

Aside from notching, simulation studies are also carried out on other types of disturbance signals. Figure 6.22(b) illustrates a typical disturbance signal of an impulsive transient, and the extracted disturbance is plotted in Fig. 6.22(c).

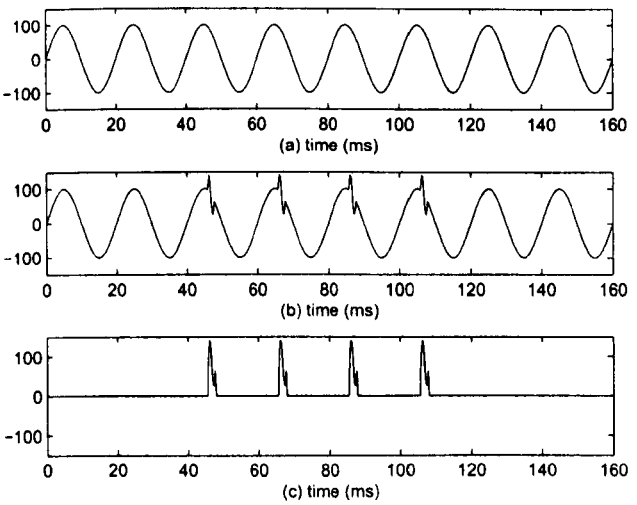


Figure 6.21: Disturbance detection of notching.

As can be seen from the figure, the transient is correctly extracted, while at the places where there is no disturbance, the output is 0.

A voltage swell lasting from the third to the sixth cycle in a fundamental sinusoidal waveform is presented in Fig. 6.23(b), while the disturbance-free signal is given in Fig. 6.23(a) as a reference. The detection result is shown in Fig. 6.23(c), from which it can be seen that the swell is exactly detected at the points where the voltage ascends and recovers.

Figures 6.24(a) and (b) describe a disturbance-free signal and a disturbance signal of momentary interruption, respectively. The interruption begins from the third cycle of the signal and lasts 4 cycles. As in the previous studies, the extracted disturbance is plotted in Fig. 6.24(c), which illustrates that the proposed approach is applicable for this type of disturbance as well.

The detection scheme based on the GK clustering algorithm is more applicable to off-line power quality analysis, as it requires the presence of the disturbance and the clustering procedure is relatively time-consuming. In the following subsection, a scheme involving only three samples to detect and locate the disturbances is presented, and the classification of the disturbances

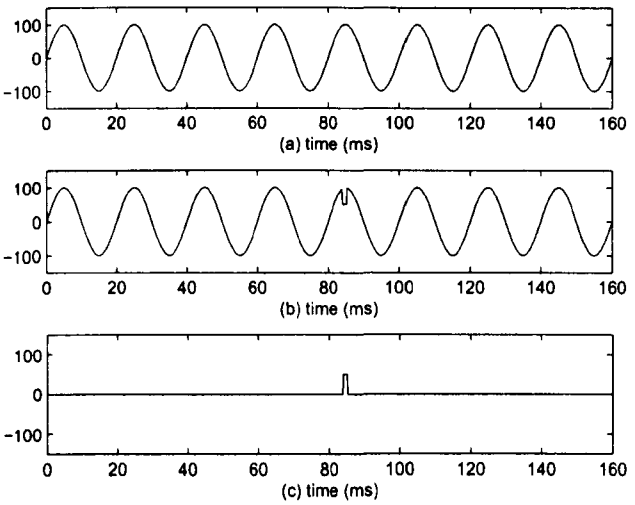


Figure 6.22: Disturbance detection of an impulsive transient.

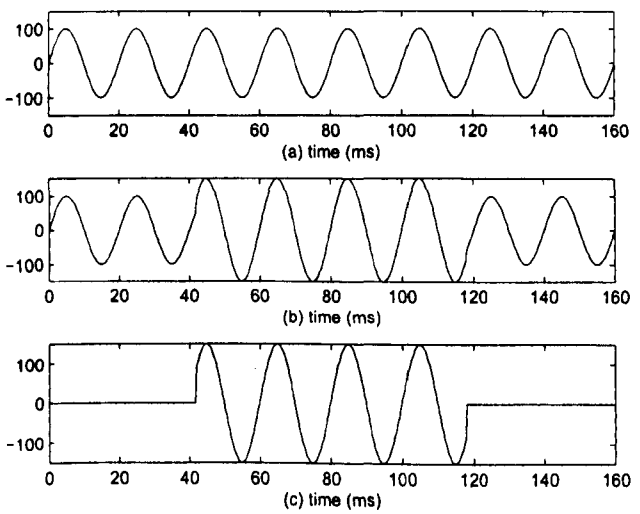


Figure 6.23: Disturbance detection of voltage swell.

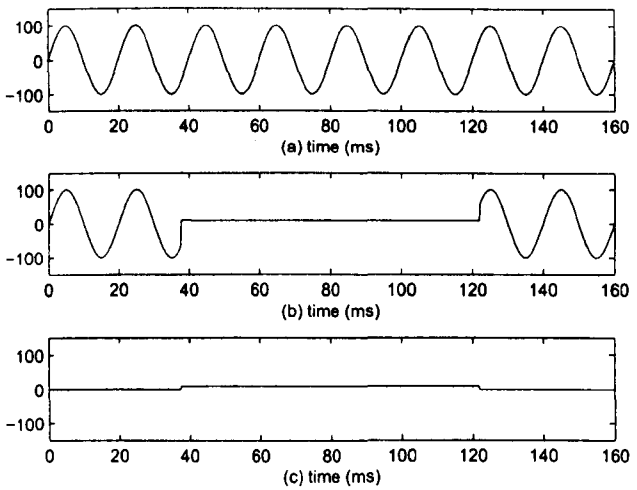


Figure 6.24: Disturbance detection of momentary interruption.

based on the detection result is also available.

6.5.3 Disturbance Detection, Location and Classification through Projection

The embedding strategy

In this subsection, the delay is set to be a quarter of its period, *i.e.* $\tau = N_s/4$. Thus, we have:

$$\begin{aligned}
&\mathbf{x}_k \\
&= \begin{bmatrix} A\sin(\omega k\Delta t+\phi) & A\sin(\omega(k+\tau)\Delta t+\phi) & A\sin(\omega(k+2\tau)\Delta t+\phi) \end{bmatrix} \\
&= \begin{bmatrix} A\sin(\omega k\Delta t+\phi) & A\cos(\omega k\Delta t+\phi) & -A\sin(\omega k\Delta t+\phi) \end{bmatrix} \\
&= \begin{bmatrix} \mathfrak{x}_{1k} & \mathfrak{x}_{2k} & \mathfrak{x}_{3k} \end{bmatrix}.
\end{aligned} \tag{6.5.11}$$

Apparently, $\mathfrak{x}_{1k}^2 + \mathfrak{x}_{2k}^2 = A^2$, $\mathfrak{x}_{2k}^2 + \mathfrak{x}_{3k}^2 = A^2$ and $\mathfrak{x}_{1k} + \mathfrak{x}_{3k} = 0$, which shows that the embedded signal in the phase space is an ellipse. Using x, y and z to represent the coordinates of the phase space, $\mathfrak{x}_1$, $\mathfrak{x}_2$ and $\mathfrak{x}_3$, respectively, the embedded signal in the phase space can be defined by either of the following

equations:

$$\begin{cases} x + z = 0 \\ x^2 + y^2 = A^2 \end{cases} \quad \begin{cases} x + z = 0 \\ y^2 + z^2 = A^2 \end{cases} . \tag{6.5.12}$$

The projection of the embedded signal to the xy-plane and the yz-plane is a circle whose radius is the amplitude of the sinusoidal signal and whose centre is the origin of the plane, while the projection to the xz-plane is a straight line.

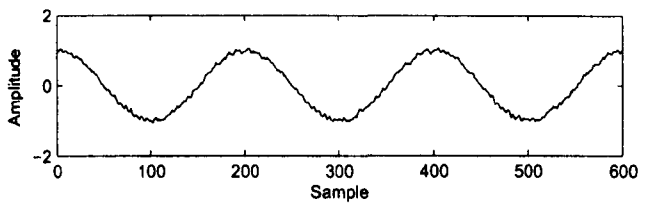
On the other hand, the disturbances will be mapped to some other shapes than the ellipse. This feature can be used to detect the disturbances from the embedded signal. An example is given in Fig. 6.25, which shows a sinusoidal signal $F(k) = \sin(2\pi \cdot 50k\Delta t + \phi)$ ($N_s = 200$) corrupted by harmonic distortion with the THD of 4.94% and white noise with the signal-to-noise ratio (SNR) of 30 dB and its embedding in the phase space. The projections in the xy-plane and the yz-plane have the same characteristic. Therefore, the detection is based on xy-plane and xz-plane projections. The projections on these planes are denoted by $\hat{X} = [\hat{\mathbf{x}}_1\ \hat{\mathbf{x}}_2\ \ldots\ \hat{\mathbf{x}}_M]^{\mathrm{T}} = [\mathfrak{x}_1\ \mathfrak{x}_2] = [x\ y]$ and $\check{X} = [\check{\mathbf{x}}_1\ \check{\mathbf{x}}_2\ \ldots\ \check{\mathbf{x}}_M]^{\mathrm{T}} = [\mathfrak{x}_1\ \mathfrak{x}_3] = [x\ z]$, respectively.

Frequency shift

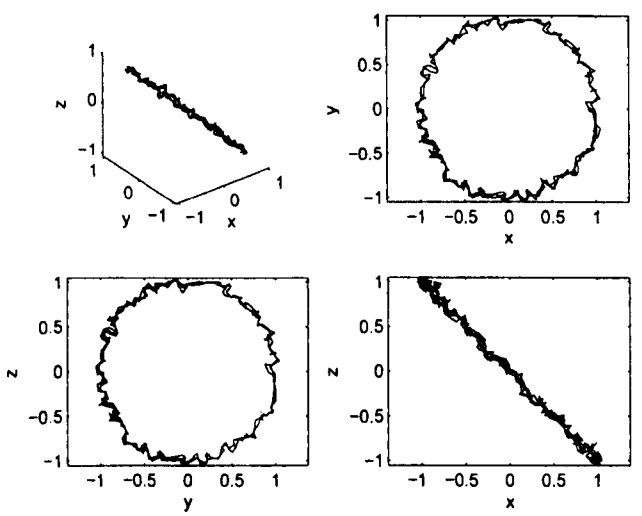
In the proposed scheme, the delay selected as $\tau = 1/(4f\Delta t)$, where f is the nominal fundamental frequency. However, in practice, the fundamental frequency may deviate slightly from its nominal value when the balance between the load and the capacity of the available generation changes. Such a frequency shift distorts the projections in the xz-plane from a straight line segment to an ellipse, as shown in Fig. 6.26. Hence, it is necessary to estimate the actual fundamental frequency, f', and re-embed the signal with $\tau = 1/(4f'\Delta t)$. The strategy of estimating f' is described in section 6.4.6.

Impulsive transients

Figure 6.27(a) shows a test signal with an impulsive transient, which occurs at x_{695} and x_{696}. In this case, the actual fundamental frequency is $f' = 50.5099$ Hz and the estimated one is $\hat{f}' = 50.5215$ Hz. Hence, $\tau = 49$. Embedding the



(a) A noisy sinusoidal signal



(b) The embedded signal in the phase space ($d_E = 3$) and its projection on xy-, yz- and xz-plane, respectively

Figure 6.25: A noisy sinusoidal signal and its embedding in the phase space.

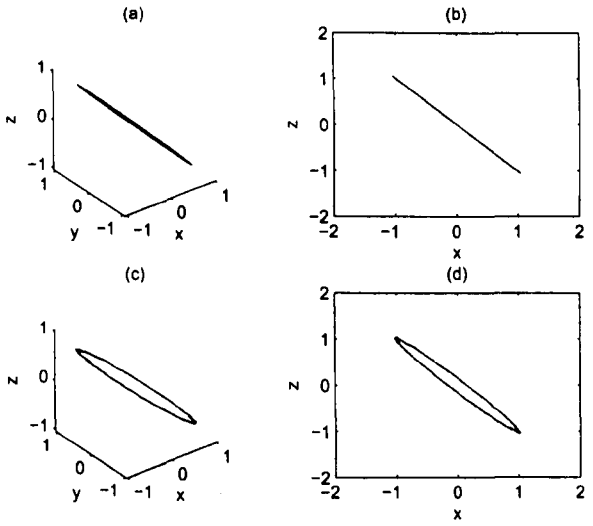


Figure 6.26: Influence of fundamental frequency shift on embedding. (a) and (b): $f' = 50$ Hz. (c) and (d): $f' = 52.5$ Hz.

disturbance signal to the phase space, the result is shown in Fig. 6.27(b). From its projection on xy-plane, it can be seen that points belonging to the normal part of the signal form a circle whose radius is around 1, while points belonging to the disturbance is much nearer to the origin. Since the projection involves two samples from the input signal, points from (x_{646}, x_{695}) to (x_{647}, x_{696}) and from (x_{695}, x_{744}) to (x_{696}, x_{745}) are affected. Hence, two disturbances appear in the phase space.

To view it more clearly, the Euclidean norm of each point on the xy-plane, denoted by E and $E(k) = \sqrt{x_k^2 + x_{k-\tau}^2}$, is plotted in Fig. 6.27(c). E also determines the distance between each point and the origin of the xy-plane. For the points belonging to the normal part of the signal, their Euclidean norm should be within the range of $[A(1-\gamma), A(1+\gamma)]$, where γ is the threshold introduced to tolerate the corruption of noise. For points whose E falls out of this range, they are detected as disturbance. The beginning and ending samples of the disturbance are recorded by the detection scheme. In this case, two pairs of disturbances are located at $p_1 =< 695, 696 >$ and $p_2 =< 744, 745 >$. Noting that one sample in the time domain affects two points in the xy-plane of the

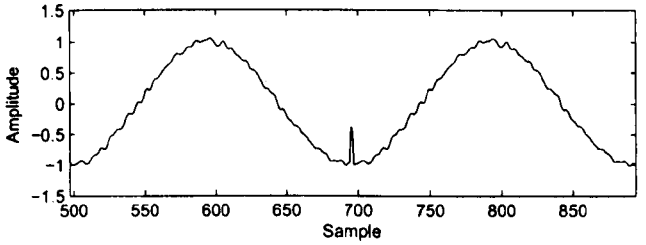
phase space and $p_{2,1} - p_{1,1} = \tau$, $p_{2,2} - p_{1,2} = \tau$, the disturbance is located at from x_{695} to x_{696}. This procedure can be performed in real time, for the embedding requires the current sample and a previous sample.

The disturbance can also be detected from the projection of the embedded signal to the xz-plane, where points belonging to the normal part of the signal satisfy function $f(x, z) = x + z = 0$. For each point on the xz-plane, its value of $f(x, z)$, denoted by ι, is calculated and plotted in 6.27(d). Considering the existence of the noise, points whose ι value falls out of range $[-\gamma, \gamma]$ are detected as disturbance. In this case, points (x_{597}, x_{695}), (x_{598}, x_{696}), (x_{695}, x_{793}) and (x_{696}, x_{794}), which correspond to samples $\{x_{695}, x_{696}\}$ of the test signal, are detected as disturbance. The detection result is the same as the result obtained in the xy-plane. In the following studies, the final detection result is the 'union' of that in the xy-plane and the xz-plane.

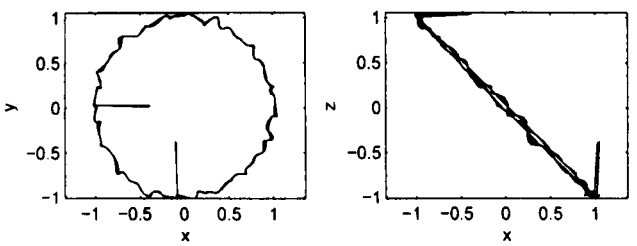
Oscillatory transients

Test oscillatory transients are simulated to contain low frequency (300 $\sim$ 900 Hz) component only, with a duration of 0.3 $\sim$ 50 ms and a magnitude of 0 $\sim$ 4 p.u. One of the test signals is demonstrated in Fig. 6.28(a) and its embedding in Fig. 6.28(b). In this case, $f' = 49.8248$ Hz, $\hat{f}' = 49.8709$ Hz and $\tau = 50$. The simulated oscillatory transient occurs from x_{626} to x_{968}.

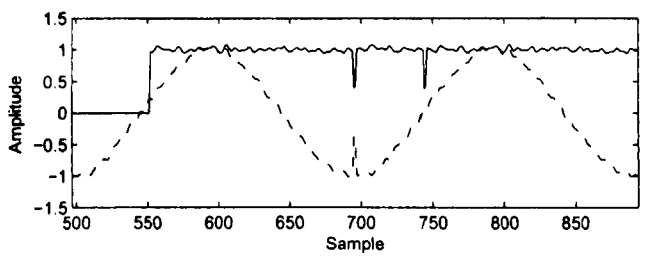
The detection result in Fig. 6.28(c) shows that the disturbance causes a larger fluctuation in the Euclidean norm, which means much more pairs of beginning and ending samples whose Euclidean norm falls out of the range of $[A(1 - \gamma), A(1 + \gamma)]$ will be recorded. Denote these pairs by $p_1, \ldots, p_m$. The detection scheme reckons that if the number of the pairs exceeds one twentieth of the samples involved, *i.e.* $m > (p_{m,2} - p_{1,1}) \times 5\%$, the disturbance is classified as oscillatory transients. The shortcoming of this scheme lies in that it can only determine the location of the transients, but cannot analyse the frequency and amplitudes of the transients separately. In this test, the disturbance is located from x_{627} to x_{966}. The detection result is almost consistent with the simulated disturbance. On the other hand, Fig. 6.28(d) shows that points with a larger



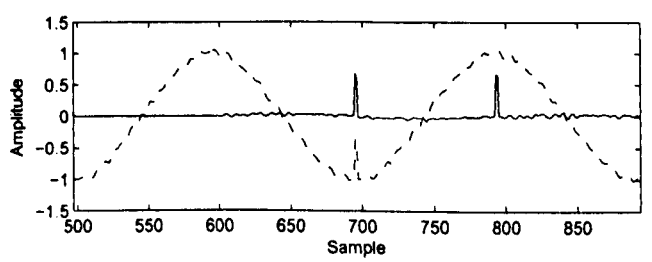
(a) A noisy power system signal with the disturbance of an impulsive transient



(b) xy-plane and xz-plane projection of the embedded signal



(c) Distance between each point and the xy-plane origin



(d) Value of $f(x, z)$ of each point on the xz-plane

Figure 6.27: Disturbance detection of an impulsive transient.

value of $f(x, z)$ than the threshold locate between $\check{\mathbf{x}}_{526}$ and $\check{\mathbf{x}}_{1068}$. Hence, it is determined that the disturbance lasts from x_{626} to x_{968} in the test signal.

Voltage sag

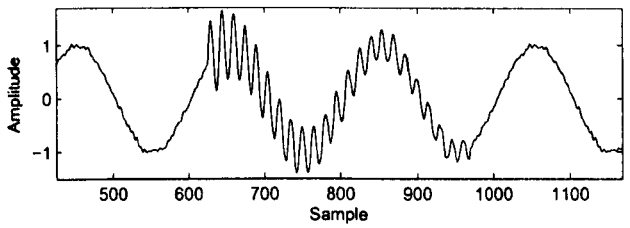
A test signal with a voltage sag is plotted in Fig. 6.29(a). Here, $f' = 48.0213$ and the amplitude of the sag is 0.8 p.u. The sag occurs and ends at the zero-crossing point to simulate a gradual drop of voltage (at x_{832} and x_{1665}), which is more difficult to detect than a sudden drop. The estimated fundamental frequency is $\hat{f}' = 47.9974$, thus $\tau = 52$. The Euclidean norm of the points in the xy-plane gradually decreases from 1 p.u., as shown in Figs. 6.29(b) and (c), and it indicates a disturbance other than transients. The E value exceeds the threshold at $E(845)$ and remains exceeding until $E(1703)$. Hence, the sag is located at from x_{845} to x_{1651}.

The location of the sag obtained from function $f(x, z)$ is shown in Fig. 6.29(b). Figure 6.29(d) shows that two segments from ι_{855} to ι_{912} and from ι_{1692} to ι_{1752} exceed the threshold, which means points $\check{\mathbf{x}}_{755} \sim \check{\mathbf{x}}_{812}$ and $\check{\mathbf{x}}_{1592} \sim \check{\mathbf{x}}_{1652}$ deviate from the line segment defined by $f(x, z) = 0$. The dragging phenomenon is due to the delay of embedding. Hence, the location of the sag is determined at from x_{855} to x_{1652}.

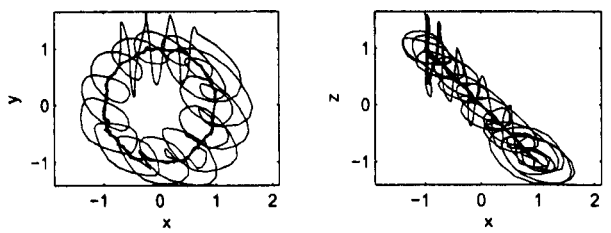
Momentary interruption

A momentary loss of voltage on a power system can be called a momentary interruption. An interruption with a reduced voltage of 0.1 p.u. from x_{804} to x_{1608} is plotted in Fig. 6.30(a). The change of voltage is also simulated in a gradual way. In this case, $f' = 49.7417$ Hz, $\hat{f}' = 49.6698$ and $\tau = 50$. The embedding of this signal is shown in Fig. 6.30(b). The Euclidean norm of each point on the xy-plane is given in Fig. 6.30(c), and the value of $f(x, z)$ of each point on the xz-plane is given in Fig. 6.30(d).

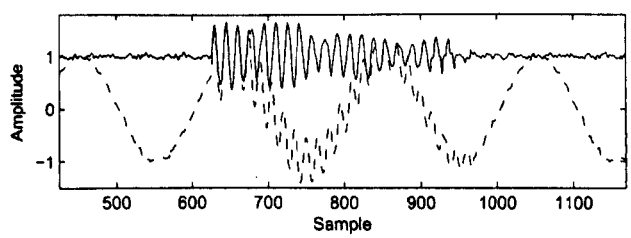
The detection result obtained by measuring the Euclidean norm of the points in the xy-plane is from x_{821} to x_{1592}. Due to the shape feature of the input signal, when the interruption occurs, there is a drop of the Euclidean



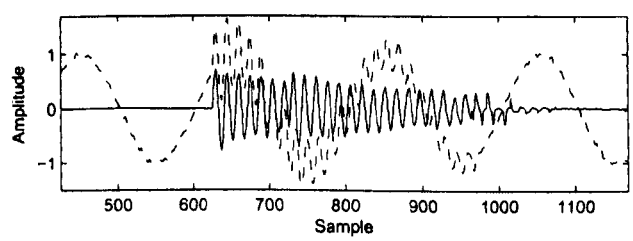
(a) A noisy power system signal with the disturbance of an oscillatory transient



(b) xy-plane and xz-plane projection of the embedded signal

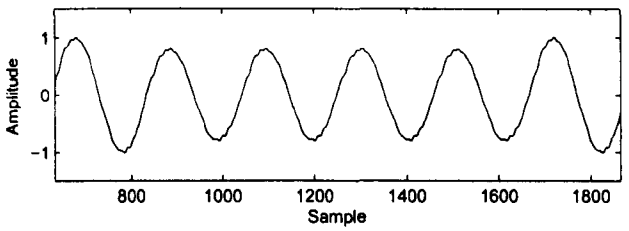


(c) Distance between each point and the xy-plane origin

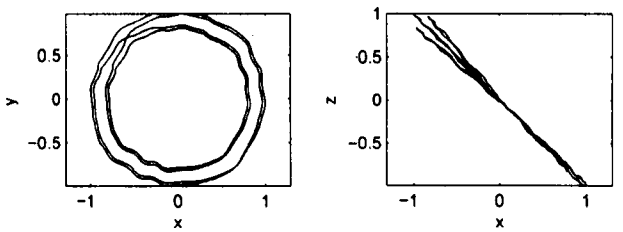


(d) Value of $f(x, z)$ of each point on the xz-plane

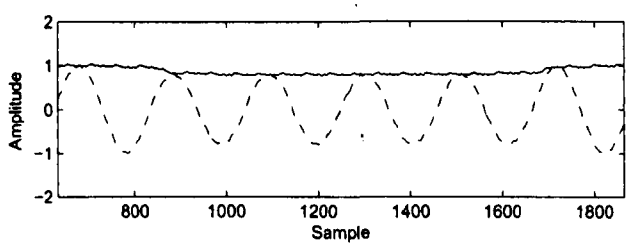
Figure 6.28: Disturbance detection of an oscillatory transient.



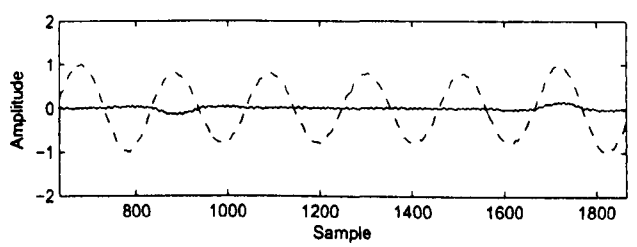
(a) A noisy power system signal with the disturbance of a voltage sag



(b) xy-plane and xz-plane projection of the embedded signal



(c) Distance between each point and the xy-plane origin



(d) Value of $f(x, z)$ of each point on the xz-plane

Figure 6.29: Disturbance detection of a voltage sag.

norm of the corresponding point on the xy-plane. However, as the embedding requires two samples and in this case, one sample comes from the normal part and the other one from the interruption, the points form a horizonal line on the xy-plane and their Euclidean norm gets closer to the radius. This phenomenon lasts a duration of τ samples. Similar situation happens when the signal returns to normal, when a group of τ samples form a vertical line. Obviously, in this case, it is not accurate enough to use the Euclidean norm to locate the interruption. An alternative means is to calculate the distance between every two neighbour points, which is defined as $\mathfrak{D}_i = \| \hat{\mathbf{x}}_i - \hat{\mathbf{x}}_{i-1} \|$. For points corresponding to the normal part and the interruption, their $\mathfrak{D}$ is relatively small, while for the beginning and ending points of the interruption, the value of $\mathfrak{D}$ is much larger. The threshold is set to $T_{\mathfrak{D}} = 2A\sin(2\pi/N_s) \times 2$, which comes from the law of sines. For this test, four points whose distance to its previous neighbour exceeds the threshold are recorded, namely $\hat{\mathbf{x}}_{757} = (x_{757}, x_{807})$, $\hat{\mathbf{x}}_{807} = (x_{807}, x_{857})$, $\hat{\mathbf{x}}_{1561} = (x_{1561}, x_{1611})$ and $\hat{\mathbf{x}}_{1611} = (x_{1611}, x_{1661})$. Hence, the interruption is located at from x_{807} to x_{1611}, which is more accurate than the result obtained earlier.

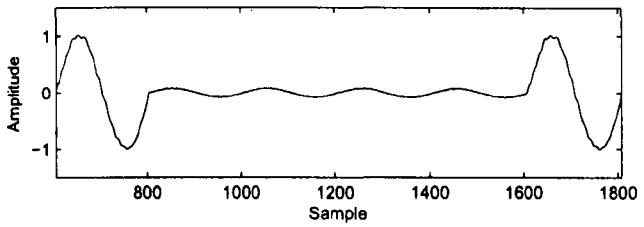
Examining Fig. 6.30(d), it can be seen that there are two groups of successive of samples whose $f(x, z)$ value exceeds the threshold, from ι_{808} to ι_{892} and from ι_{1606} to ι_{1695}. Considering the dragging phenomenon, the interruption is located at from sample x_{808} to sample x_{1606}.

Voltage swell

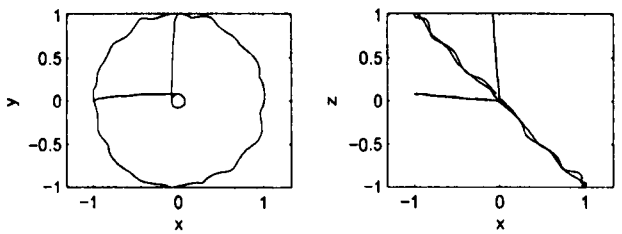
A voltage swell is a short term increase of system voltage. The detection of voltage swell is very similar to that of voltage sag, except that when the swell occurs, the voltage increases rather than decreases.

Notching

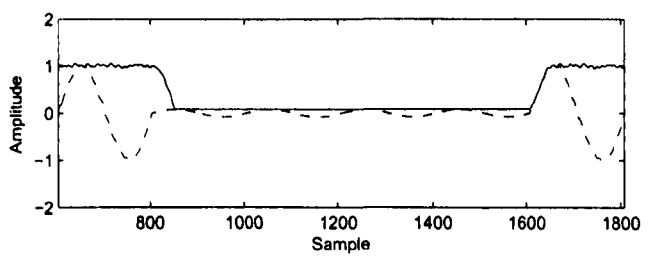
A signal with such a disturbance of notching is shown in Fig. 6.31(a). Four notches are simulated at $x_{831} \sim x_{839}$, $x_{981} \sim x_{989}$, $x_{1131} \sim x_{1139}$ and $x_{1281} \sim x_{1289}$. Its embedding ($f' = 49.6620$ Hz, $\hat{f}' = 49.6666$ Hz and $\tau = 50$)



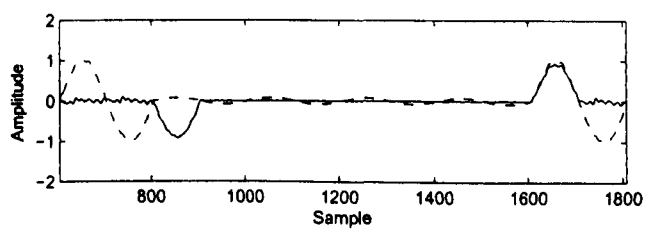
(a) A noisy power system signal with the disturbance of a momentary interruption



(b) xy-plane and xz-plane projection of the embedded signal



(c) Distance between each point and the xy-plane origin



(d) Value of $f(x, z)$ of each point on the xz-plane

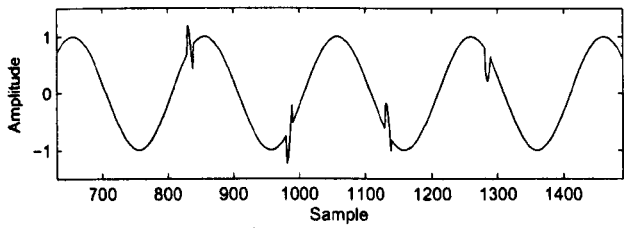
Figure 6.30: Disturbance detection of a momentary interruption.

is given in Fig. 6.31(b), which shows that the notches form a quasi-sinusoidal shape in the phase space.

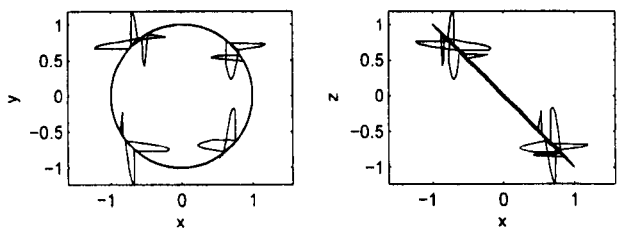
The disturbance location procedure is the same as described previously. Take the first notch for example. As can be seen from Fig. 6.31(c), four groups of E values exceeding the threshold are detected at $E(831) \sim E(833)$, $E(838) \sim E(839)$, $E(881) \sim E(883)$ and $E(888) \sim E(889)$, respectively. The last two pairs are caused by embedding and should not be included in the final detection result. Hence, the disturbance is located at $x_{831} \sim x_{833}$ and $x_{838} \sim x_{839}$. The values of samples $x_{834} \sim x_{837}$ are not affected by the notches, so no disturbance is detected at these positions. The location result obtained from the signal projected to the xz-plane is the same. If the notching occurs simultaneously with short duration variations, the threshold the Euclidean norm compared with is set to $[\tilde{A}(1-\gamma), \tilde{A}(1+\gamma)]$, where $\tilde{A}$ is the amplitude of the short duration variation.

Transients at the beginning and ending of a sag

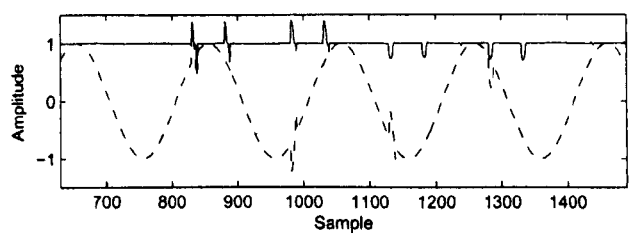
In this scenario, three types of disturbances are simulated in one signal. An impulsive transient ($x_{772} \sim x_{774}$) occurs at the beginning of a sag ($x_{775} \sim x_{1551}$), followed by an oscillatory transient ($x_{1552} \sim x_{1645}$). The signal and its embedding ($f' = 51.5736$ Hz, $\hat{f}' = 51.5930$ Hz, $\tau = 48$) are shown in Figs. 6.32(a) and (b), respectively. From the xy-plane projection and the E values, as shown in Fig. 6.32(c), it can be seen that $E(773) \sim E(774)$ exceed the threshold, and after a gradual decrease from $E(790)$ to $E(820)$, a local maximum of $E(821) \sim E(822)$ is detected before the E value reaches a stable value of around 0.3 p.u. As $p_1 =< 773, 774 >$, $p_2 =< 821, 822 >$ and $p_{2,1} - p_{1,1} = \tau$, $p_{2,2} - p_{1,2} = \tau$, an impulsive disturbance can be located at $x_{773} \sim x_{774}$. The E value remains at around 0.3 p.u. from $E(823)$ to $E(1553)$ before it starts to vibrate and a group of successive Euclidean norms exceeding the threshold (from $p_3 =< 1590, 1595 >$ to $p_{14} =< 1682, 1684 >$) are recorded. Hence, a sag is located at $x_{775} \sim x_{1553}$ and an oscillatory at $x_{1590} \sim x_{1636}$, respectively.



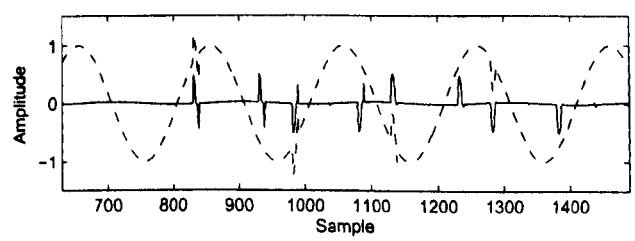
(a) A noisy power system signal with the disturbance of periodic notching



(b) xy-plane and xz-plane projection of the embedded signal



(c) Distance between each point and the xy-plane origin



(d) Value of $f(x, z)$ of each point on the xz-plane

Figure 6.31: Disturbance detection of periodic notching.

The projection on the xz-plane and the ι value are shown in Figs. 6.32(b) and (d), respectively. A group of segments $\iota_{722} \sim \iota_{724}$, $\iota_{781} \sim \iota_{870}$, $\iota_{1553} \sim \iota_{1661}$, $\cdots$, $\iota_{1739} \sim \iota_{1741}$ whose ι value exceeds the threshold are recorded. With consideration of classification criteria described in section 6.5.3, an impulsive transient, a sag and an oscillatory transient are detected at $x_{722} \sim x_{724}$, $x_{781} \sim x_{1553}$ and $x_{1553} \sim x_{1645}$. The final location is the 'union' of the two results, which is $x_{722} \sim x_{724}$, $x_{775} \sim x_{1553}$ and $x_{1553} \sim x_{1645}$. The detection result almost coincides with the simulation.

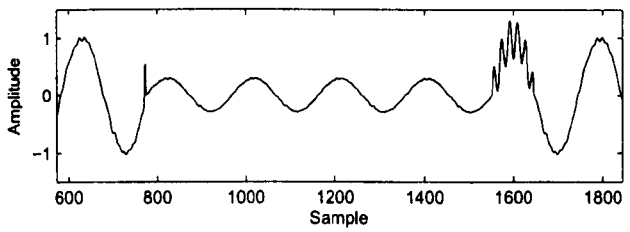
Disturbance signals obtained from PSCAD simulations

In this section, test signals are obtained from PSCAD simulations based on realistic power system circuit parameters. Figure 6.33 shows an oscillatory transient at the beginning of a sag and the detection of the disturbance. The detection procedure is the same as described in the previous section. Both E values and ι values exceed the pre-set threshold as the disturbance occurs.

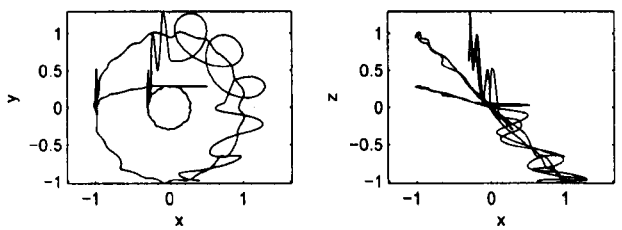
Disturbances classification

The classification is based on the detection results from xy-plane as well as the features of the disturbances. The detection scheme records pairs of the beginning and ending samples of the disturbance. A disturbance is first classified according to its length. Disturbances of sags, interruptions and swells usually last longer than a half of a cycle, while other types of disturbances last much shorter. Therefore, if a pair of samples $p =< i, j >$ has been detected and $j - i$ exceeds a pre-set threshold, the disturbance is classified as a sag/interruption/swell. The E value is used to further distinguish these three types of disturbances, as their amplitude is in the range of $0.1 \sim 0.9$ p.u., $0 \sim 0.1$ p.u. and $1.1 \sim 1.8$ p.u., respectively.

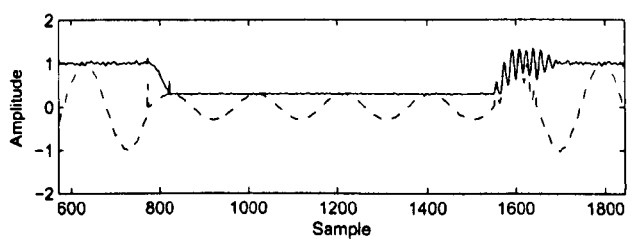
On the other hand, if a pair p_n takes only several samples, *e.g.* $p_{n,2} - p_{n,1} < \tau$, the disturbance is classified as a transient or a notching. According to [33], a low frequency oscillatory transient usually lasts $0.3 \sim 50$ ms, while the duration of impulsive transients and notches has not been clearly defined. However, the



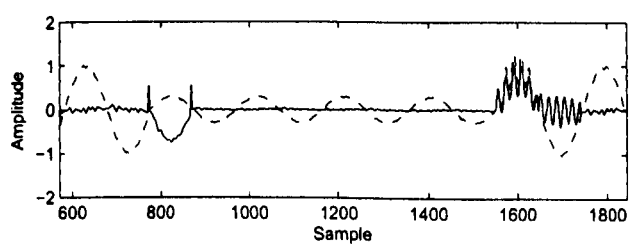
(a) A noisy power system signal with the disturbances of two transients and a sag



(b) xy-plane and xz-plane projection of the embedded signal

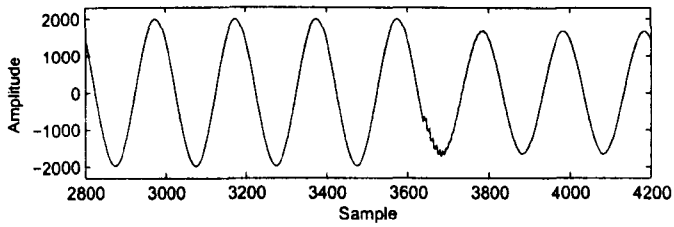


(c) Distance between each point and the xy-plane origin

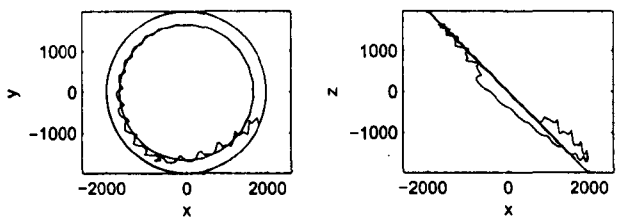


(d) Value of $f(x, z)$ of each point on the xz-plane

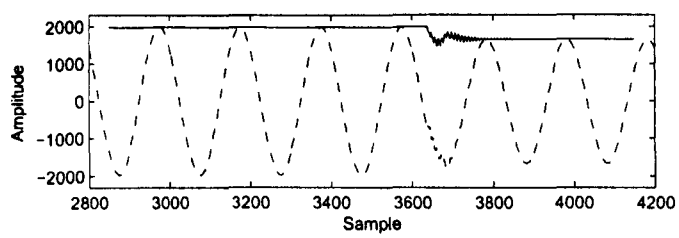
Figure 6.32: Disturbance detection of two transients and a sag.



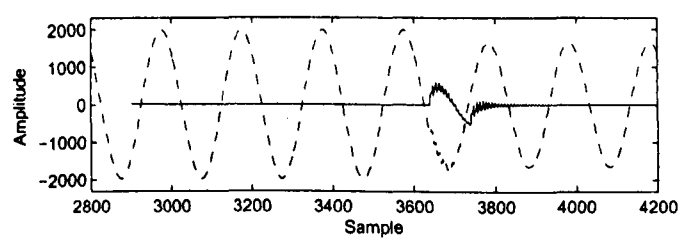
(a) A power system signal with the disturbances of an oscillatory transient and a sag



(b) xy-plane and xz-plane projection of the embedded signal



(c) Distance between each point and the xy-plane origin



(d) Value of $f(x, z)$ of each point on the xz-plane

Figure 6.33: Disturbance detection of a test signal simulated by PSCAD.

impulsive transients and notches considered in this thesis could be as short as 2 ms. Therefore, the duration is a secondary criterion to distinguish oscillatory transients from impulsive transients and notches. The primary criterion is if there is another pair p_m that satisfies $p_{m,2} - p_{n,2} = \tau$ and $p_{m,1} - p_{n,1} = \tau$, it is either an impulsive transient or a notch. Otherwise, if the number of the recorded pairs exceeds a pre-defined threshold, the disturbance is classified as transients, as described in section 6.5.3. Since notches occur continuously, if a group of beginning and ending pairs of short intervals over one quarter of a cycle are detected, they are classified as notches. If the pair detected is isolated, it is considered as noise. The classification process is illustrated in the flowchart shown in Fig. 6.34.

Simulation results

The test disturbance signals are simulated to contain up to 40 harmonics with a THD of $2 \sim 5\%$ and Gaussian noise with an SNR of $30 \sim 60$ dB, and they may also have a $\pm 5\%$ fundamental frequency shift. The values of these parameters are randomly selected with a uniform distribution. As for each disturbance, its location, duration and magnitude are randomly selected within a range with accordance to the parameters given in section 1.3.2 so that they are at different levels.

In order to simulate real PQ events, each test signal may contain up to three disturbances. Forty signals are simulated to contain a disturbance of each type, respectively. Hence, $40 \times 6 = 240$ signals in total contain only one disturbance. For signals containing two or three disturbances, the arrangement is listed in Table 6.3. Altogether, 300 test signals are generated and 240 of them contain one disturbance, 45 contain two disturbances and 15 contain three disturbances. The location, duration and magnitude are randomly generated within the range given in Table 1.1.

The detection results of the 300 tests are listed in Table 6.4, including the rate of correct determination of the existence of a disturbance, the average accuracy $\bar{\sigma}$ of the location of the disturbance and the rate of correct classifi-

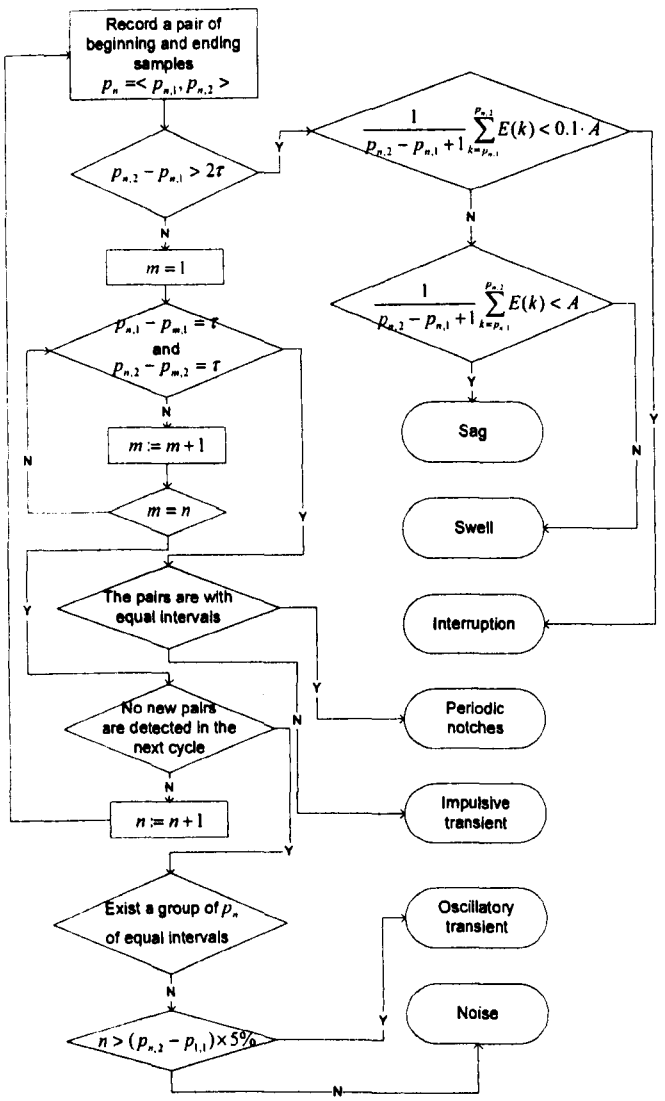


Figure 6.34: The process of disturbances classification.

Table 6.3: Arrangement of the multi-disturbances contained in a signal

Number of signals	Impulsive transients	Oscillatory transients	Voltage sags	Momentary interruptions	Voltage swells	Periodic notches
3	√	√	√			
3	√	√		√		
3	√	√			√	
3		√		√		√
3	√			√		√
5	√		√			
5	√			√		
5	√				√	
5		√	√			
5		√		√		
5		√			√	
5			√			√
5				√		√
5					√	√

cation. As can be seen from the table, the proposed scheme is able to detect the occurrence of disturbances, except that in two tests when the impulsive transient occurs together with an oscillatory transient and a periodic notching, respectively, the impulsive transient has not been detected. The location results of sags, interruptions, swells and oscillatory transients are satisfactory, with an error of only a couple of samples in the location and an accurate result in classification. If more than one notch occurs in the test signal, the average degree of match of all the notches is recorded for statistic analysis. A relative low $\bar{\sigma}$ of impulsive transients and periodic notching is due to its short duration. An error of one sample may cause a significant decrease in the index of degree of match. In general, a mis-location of one sample may cause a $10 \sim 20\%$ fall of σ. The classification result indicates that the classification strategy proposed in section 6.5.3 is effective.

Table 6.4: The detection results of 300 test signals

Power disturbances	Total number of the disturbances	Accuracy rate of determination	Average value of σ: $\bar{\sigma}$	Accuracy rate of classification
Impulsive transients	67	97.01%	90.50%	100%
Oscillatory transients	67	100%	88.62%	100%
Voltage sags	61	100%	90.03%	100%
Momentary interruptions	61	100%	94.13%	100%
Voltage swells	58	100%	96.39%	100%
Periodic notches	61	100%	87.93%	100%

In section 3.3, a morphological gradient wavelet (MGW) is proposed to de-

tect and locate the transient disturbances of power system signals. It has been proved that MGW outperforms Daubechies DB4 wavelet. However, compared with the results presented in this section, it can be seen that the scheme proposed here is more accurate and reliable than MGW and requires less computation. The scheme explores an alternative to detect and locate the disturbances. Moreover, it involves classification, which is also important for power quality analysis.

6.6 Conclusion

The embedding-based technique is applied to various signal processing tasks. For ECG signal analysis, the ECG signals are transformed to the phase space to form a binary image, and the feature waveforms correspond to the objectives of the image. Hence, the identification of the feature waveforms is carried out through the processing of the image. Moreover, the classification of the P waves and T waves can be implemented using the geometry characteristics of the objectives.

For phasor measurement, the method takes advantage of the mathematical properties of a power system signal to transfer it to a 2-dimensional phase space through delay coordinate embedding. In this manner, the amplitude and phase angle of a current or voltage signal and the phase difference of current and voltage signals can be calculated sample by sample. The calculation involves two or four samples only, unlike traditional FT-based methods that uses samples of half a cycle or an entire cycle. Moreover, the method can also be used to estimate the actual fundamental frequency when it deviates from its nominal value.

For disturbance detection, two schemes based on embedding are proposed. The delay constant of the first scheme is selected according to the the general strategy proposed in [95], and the normal part of a disturbance signal forms an ellipse in the phase space while the disturbance forms a shape that deviates from the ellipse. The GK clustering algorithm is therefore used to distinguish

the two shapes due to its ability to search in particular for ellipsoidal structures. The scheme has been applied to detect several types of power disturbances and the simulation results have shown its ability to localise the disturbances precisely. As the scheme requires the presence of the disturbance and the clustering procedure is relatively time-consuming, it is more applicable to off-line power quality analysis rather than on-line monitoring.

The scheme based on projection uses the mathematical properties of sinusoidal signals to determine the delay constant so that the projection of the normal part on the xy-plane forms a circle and the projection on the xy-plane forms a line segment. The scheme extracts the features from the projections and uses the gauges of Euclidean norm and function $f(x.z)$ to determine the location of the disturbance and its classification. Since the embedded signal is constructed using the data that are sequentially sampled within a small sampling window in the time domain, the location of the disturbances is almost real-time and the computation time is greatly reduced compared with methods that process the signal within a much longer sampling window. The proposed scheme has been evaluated on a number of test signals, which are of six different types of disturbances under various conditions. The simulation results have shown that the proposed scheme is able to locate the occurrence of disturbances and can accurately classify them, as long as the disturbances are not buried in the noise.

Chapter 7

Conclusion and Future Work

This chapter concludes the thesis, summarises the major outcome of the research work presented in the thesis and indicates possible directions for further investigation of advanced morphological operators.

7.1 Conclusion

The objective of the research is to develop advanced morphological operators and to apply them to signal processing. MM is a non-linear technique that focuses on the shape information of a signal. This property enables MM to concentrate on characteristic waveforms and to replace traditional techniques based on integral calculation, such as the FT and the WT.

Apart from basic morphological operators, such as dilation, erosion, opening and closing, the thesis also engages the schemes of soft MM, multi-resolution decomposition, multiscale MM and the embedding theorem in the development of novel advanced morphological operators. Several advanced morphological operators have been developed in the research to fulfill the tasks of noise removal of ECG signals, feature waveform identification of ECG signals, disturbance detection, location and classification of power system signals, phasor measurement of power system signals, noise removal of images.

For the three types of noise existing in most ECG signals, three methods have been designed to remove them respectively. The multi-resolution mor-

phological filter is used to remove the impulsive noise, the adaptive multiscale morphological filter is developed to reject high frequency (Gaussian) noise, and a fusion method hybridising EMD and MM for the normalisation of the baseline wander. It should be noted that all the filters are able to preserve the shape of the feature waveforms to the greatest extent while removing the noise.

In order to identify the feature waveforms, namely the QRS complexes, the P waves and the T waves, two methods have been developed respectively. The first one is the multi-resolution morphological filter. In this case, the feature waveforms are considered as 'noise' and they are extracted by rejecting them separately. The method utilises the geometric features of ECG signals, such as the R waves having a remarkably higher amplitude and the width of the QRS complexes being shorter, *etc*, to identify the three types of feature waveforms one by one. The second method is based on the embedded signal and treats it as a binary image. As the feature waveforms form separate objectives in the image, the identification is carried out based on the geometric information extracted from the image.

The core idea of disturbance detection is feature extraction. As the occurrence of disturbances arouses change in the gradient, the MGW is developed to extract the gradient information of the signal. As the disturbances and the normal part of the signal have distinguished geometric characteristics, the disturbance signals are embedded to the phase space to have the characteristics more clearly viewed. Two methods for the separation of the disturbances and the normal part in the phase space have been proposed, based on the GK clustering algorithm and the mathematical information of the projection of the embedded signal, respectively.

The embedding-based method has also been used for phasor measurement. It is able to measure the amplitude and the phase angle of a voltage or current signal, as well as the phase difference between two signals. The method can also be used to estimate the actual fundamental frequency, which is especially useful at the presence of fundamental frequency shift. Being more accurate and much less computational complex, the method can replace traditional methods

such as the half-cycle FT.

The thesis has also proposed a novel way to design an optimal morphological filter for noise removal. The scheme involves two steps. First, a morphological filter or a combination filter is selected according to the characteristics of the noise. Second, the parameters of the morphological filter are optimised by an EA towards the pre-set targets, such as lowest PSNR value. Therefore, whatever the feature of the noise is, an optimal filter can always be constructed to reject it, targeting the criteria defined by the user. The simulation results have demonstrated that the optimally designed filters outperform the traditional filters.

A large amount of simulation studies have been carried out to test the performance of the proposed operators and the results have shown that they are competent to fulfill the tasks. Moreover, as the operators are developed based on a generic framework, they can be used to other applications with minor or even no modification.

7.2 Future Work

The thesis aims at the development of advanced MM and to fully explore its potential for applications in signal processing. It has also been recognised that the lack of MM analysis in the frequency domain is another obstacle that baulks the development of MM. Although the mathematical background of MM implies that MM and frequency analysis are disrelated and hybridising the two techniques faces enormous difficulty, it is worthwhile to fill the gap and the outcome will greatly benefit the realm of signal processing. Future work will concentrate on investigating an effective method that combines MM and frequency-based techniques. It may include studying the frequency response of commonly used morphological operators, such as opening and closing. Based on the study, a generic framework will be constructed to design morphological filters that have specific frequency response. The morphological filters can also be designed to have better performance than traditional frequency filters and

to consider both shape and frequency information while processing a signal. As morphological wavelet fits in the framework of the WT and involves morphological operation, it might perform frequency analysis if the morphological operation uses a series of sine-based structuring elements. The future work will take the following steps.

- On the investigation of the frequency properties of basic morphological operators, a generic mathematical framework will be constructed to design morphological filters that have specific frequency responses. The morphological filters will be designed to have better performance than traditional frequency filters and to consider both shape and frequency information while processing a signal.

- Preliminary study on the morphological wavelet has shown the merits of infusing proper morphological operation in the decomposition and approximation procedure. The future work will investigate the influence of the morphological operators engaged in the morphological wavelet and generalise a strategy to select or design the most suitable morphological operators to solve a specific signal processing problem.

- Some powerful MM-based schemes have been developed in recent literature, such as the slope transform, morphological gradient, morphological pyramid and morphological undecimated decomposition. Yet, they are limited to deal with certain types of signals. The behaviour of these schemes in the frequency domain will be studied, which will lead to the investigation of any underlying linkage among them. These schemes will afterwards be advanced to involve frequency analysis to enhance their accuracy and ability in feature extraction.

- The theoretical achievement will be applied to design a new generation of protection relays for power systems. Tasks may include: to exactly distinguish and extract the faint surge of transient faults for ultra-high-speed relays, to identify the waveforms of a fault voltage/current signal,

and to detect and compensate the distorted waveform caused by CT saturation.

Appendix A

Full-cycle and Half-cycle Fourier Transforms

A.1 Full-cycle Fourier Transform

The full-cycle FT is the most widely used algorithm for extraction of the amplitude and phase angle of a fault signal. It is immune to constant DC offsets and can filter integral harmonics [101][102]. Any measured periodic voltage signal can be expanded into its Fourier series expansion [98] as:

$$v(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos(n\omega_0 t) + \sum_{n=1}^{\infty} b_n \sin(n\omega_0 t) \qquad \text{(A.1.1)}$$

where $\omega_0 = 2\pi f_0$ and f_0 is the fundamental frequency. The coefficients a_0, a_n and b_n can be obtained from:

$$a_0 = \frac{1}{T}\int_{t_0}^{t_0+T} v(t)\mathrm{d}t \qquad \text{(A.1.2)}$$

$$a_n = \frac{2}{T}\int_{t_0}^{t_0+T} v(t)\cos(n\omega_0 t)\mathrm{d}t, \quad n = 1, 2, \ldots, \infty \qquad \text{(A.1.3)}$$

$$b_n = \frac{2}{T}\int_{t_0}^{t_0+T} v(t)\sin(n\omega_0 t)\mathrm{d}t, \quad n = 1, 2, \ldots, \infty \qquad \text{(A.1.4)}$$

where T is the period of the fundamental frequency component of the signal. If the sampled signal is represented in a discrete form with N samples per

fundamental cycle, the real and imaginary parts of the phasor representing the sampled signal are obtained as:

$$V_{\mathrm{Re}}(k) = \frac{2}{N}\sum_{n=0}^{N-1} v(k-n)\cos\left(\frac{2\pi n}{N}\right) \tag{A.1.5}$$

$$V_{\mathrm{Im}}(k) = \frac{2}{N}\sum_{n=0}^{N-1} v(k-n)\sin\left(\frac{2\pi n}{N}\right). \tag{A.1.6}$$

The amplitude and phase angle can be obtained using the real and imaginary components, respectively, as follows:

$$V(k) = \sqrt{V_{\mathrm{Re}}^2(k) + V_{\mathrm{Im}}^2(k)} \tag{A.1.7}$$

$$\phi(k) = \tan^{-1}\left(\frac{V_{\mathrm{Im}}(k)}{V_{\mathrm{Re}}(k)}\right). \tag{A.1.8}$$

A.2 Half-cycle Fourier Transform

To reduce the computation time of the algorithm by half, the half-cycle FT was proposed, which uses samples obtained from half a fundamental cycle. The half-cycle FT is described as follows. Any measured periodic voltage signal can be expanded into its Fourier series expansion [98] as:

$$v(t) = a_0 + \sum_{n=1}^{\infty} a_n\cos(n\omega_0 t) + \sum_{n=1}^{\infty} b_n\sin(n\omega_0 t) \tag{A.2.1}$$

where $\omega_0 = 2\pi f_0$ and f_0 is the fundamental frequency. The coefficients of the Fourier series expansion are expressed by:

$$a_0 = \frac{1}{T/2}\int_{t_0}^{t_0+T/2} v(t)\mathrm{d}t \tag{A.2.2}$$

$$a_n = \frac{2}{T/2}\int_{t_0}^{t_0+T/2} v(t)\cos(n\omega_0 t)\mathrm{d}t, \quad n = 1, 2, \ldots, \infty \tag{A.2.3}$$

$$b_n = \frac{2}{T/2}\int_{t_0}^{t_0+T/2} v(t)\sin(n\omega_0 t)\mathrm{d}t, \quad n = 1, 2, \ldots, \infty \tag{A.2.4}$$

where T is the period of the fundamental frequency component of the signal. The real and imaginary parts of the phasor representing the sampled signal are calculated as:

$$V_{\mathrm{Re}}(k) = \frac{4}{N} \sum_{n=0}^{N/2-1} v(k-n) \cos\left(\frac{2\pi n}{N}\right) \tag{A.2.5}$$

$$V_{\mathrm{Im}}(k) = \frac{4}{N} \sum_{n=0}^{N/2-1} v(k-n) \sin\left(\frac{2\pi n}{N}\right). \tag{A.2.6}$$

The amplitude and phase angle can be obtained using the real and imaginary components, respectively, as follows:

$$V(k) = \sqrt{V_{\mathrm{Re}}^2(k) + V_{\mathrm{Im}}^2(k)} \tag{A.2.7}$$

$$\phi(k) = \tan^{-1}\left(\frac{V_{\mathrm{Im}}(k)}{V_{\mathrm{Re}}(k)}\right). \tag{A.2.8}$$

The half-cycle FT is used as a reference in the simulation studies: The performance of the method proposed in section 6.4 is compared with the half-cycle FT.

References

[1] G. Matheron. *Random sets and integrated geometry.* Wiley, New York, 1975.

[2] J. Serra. *Image analysis and mathematical morphology.* Academic Press, London, 1982.

[3] J. Serra. *Image Analysis and Mathematical Morphology II: Theoretical Advances.* Academic Press, London, 1988.

[4] J. Serra. Automatic scanning device for analysing textures. *Patent, Belgium, Canada, Great-Britain, Japan, Sweden, U.S.*, 23(273), July 1965.

[5] G. Matheron. *Elments pour une theorie des milieux poreux.* Masson, Paris, 1976.

[6] T. Y. Ji, Z. Lu, and Q. H. Wu. Baseline normalization of ECG signals using empirical mode decomposition and mathematical morphology. *Electronics Letters*, 44(2):82–83, 2008.

[7] P. Sun, Q. H. Wu, A. M. Weindling, A. Finkelstein, and K. Ibrahim. An improved morphological approach to background normalization of ECG signals. *IEEE Transactions on Biomedical Engineering.*

[8] L. Koskinen, J. Astola, and Y. Neuvo. Soft morphological filters. In *Proceedings of SPIE Symposium on Image Algebra and Morphological Image Processing*, pages 262–270, 1991.

[9] A. Gasteratos and I. Andreadis. Soft mathematical morphology: Extensions, algorithms and implementations. *Advances in Imaging and Electron Physics*, 110:63–99, 1999.

[10] T. Y. Ji, Z. Lu, and Q. H. Wu. Detection of power disturbances using morphological gradient wavelet. *Signal Processing*, 88(2):255–267, February 2008.

[11] P. Kuosmanen and J. Astola. Soft morphological filtering. *Mathematical Imaging and Visions*, 5(3):231–262, 1995.

[12] I. Bloch and H. Maitre. Fuzzy mathematical morphologies: A comparative study. *Pattern Recognition*, 28(9):1341–1387, 1995.

[13] A. Gasteratos, I. Andreadis, and P. Tsalides. Realisation of soft morphological filters. *IEE Proceedings: Circuits Devices and Systems*, 145(3):201–206, 1998.

[14] D. Shinha and E. R. Dougherty. Fuzzy mathematical morphology. *Visual Communication and Image Representation*, 3(3):286–302, 1992.

[15] M. Daoudi, R. Benslimane, D. Hamad, and J.-G. Postaire. A new interactive pattern recognition approach by multilayer neural networks and mathematical morphology. *Proceedings of International Conference on Systems, Man and Cybernetics, 'Systems Engineering in the Service of Humans'*, 4:125–130, October 1993.

[16] M. I. Quintana, R. Poli, and E. Claridge. Morphological algorithm design for binary images using genetic programming. *Genetic Programming and Evolvable Machines*, 7(1):81–102, March 2006.

[17] T. Y. Ji, Z. Lu, and Q. H. Wu. A particle swarm optimizer applied to soft morphological filters for periodic noise reduction. In *Lecture Notes on Computer Science, The Ninth European Workshop on Evolutionary Computation in Image Analysis and Signal Processing*, pages 367–374, Valencia, Spain, 11-13 April 2007.

[18] T. Y. Ji, Z. Lu, and Q. H. Wu. Optimal soft morphological filter for periodic noise removal using a particle swarm optimiser with passive congregation. *Signal Processing.*

[19] T. Y. Ji, M. S. Li, Z. Lu, and Q. H. Wu. Optimal filter design using a bacterial swarming algorithm. In *2008 IEEE Congress on Evolutionary Computation, CEC 2008*, pages 452–458, Hong Kong, 1-6 June 2008.

[20] P. Soille. *Morphological Image Analysis: Principles and Applications (second edition).* Springer-Verlag, New York, 2003.

[21] P. S. Addison. Wavelet transforms and the ECG: a review. *Physiological Measurement*, 26:155–199, 2005.

[22] R. V. Andre ao, B. Dorizzi, and J. Boudy. ECG signal analysis through hidden Markov models. *IEEE Transactions on Biomedical Engineering*, 53(8):1541–1549, August 2006.

[23] S. Suppappola and Y. Sun. Nonlinear transforms of ECG signals for digital QRS detection: A quantitative analysis. *IEEE Transaction on Biomedical Engineering*, 41(4):397–400, April 1994.

[24] T. P. Pander. A suppression of an impulsive noise in ECG signal processing. In *Proceedings of the 26th Annual International conference of the IEEE EMBS*, pages 596–599, San Francisco, USA, 1-5 September 2004.

[25] T. K. Mitra. Control of EMI interference for linear ECG recording. In *Proceedings of the International Conference on Electromagnetic Interference and Compatibility*, pages 391–394, 6-8 December 1999.

[26] K. Anant, F. Dowla, and G. Rodrigue. Vector quantization of ECG wavelet coefficients. *IEEE Signal Processing Letters*, 2(7):129–131, July 1995.

[27] C.-S. Chen, J.-L. Wu, and Y.-P. Hung. Theoretical aspects of vertically invariant gray-level morphological operators and their application

on adaptive signal and image filtering. *IEEE Transactions on Signal Processing*, 47(4):1049–1060, April 1999.

[28] A. Herreros, E. Baeyens, R. Johansson, J. Carlson, J. R. Perán, and B. Olsson. Analysis of changes in the beat-to-beat P-wave morphology using clustering techniques. *Biomedical Signal Processing and Control*, 2009.

[29] R. Silipo and C. marchesi. Artificial neural networks for automatic ECG analysis. *IEEE Transactions on Signal Processing*, 46(5):1417–1425, may 1998.

[30] I. Christov, G. Gómez-Herrero, V. Krasteva, I. Jekova, A. Gotchev, and K. Egiazarian. Comparative study of morphological and time-frequency ECG descriptors for heartbeat classification. *Medical Engineering & Physics*, 28:876–887, 2006.

[31] http://www.physionet.org/physiobank/database/mitdb/.

[32] D. D. Sabin and A. Sundaram. Quality enhances reliability. *IEEE Spectrum*, 33:34–41, 1996.

[33] *IEEE Recommended Practice for Monitoring Electric Power Quality.* IEEE Std. 1159-1995., November 1995.

[34] V. E. Wagner, J. C. Balda, D. C. Griffith, A. McEachern, T. M. Barnes, D. P. Hartmann, D. J. Phileggi, A. E. Emannuel, W. F. Horton, W. E. Reid, R. J. Ferraro, and W. T. Jewell. Effects of harmonics on equipment. *IEEE Transactions on Power Delivery*, 8(2):672–680, 1993.

[35] S. Santoso, J. Lamoree, W. M. Grady, E. J. Powers, and S. C. Bhatt. A scalable PQ event identification system. *IEEE Transactions on Power Delivery*, 15(2):738–743, April 2000.

[36] M. Karimi, H. Mokhtari, and M. R. Iravani. Wavelet based on-line disturbance detection for power quality applications. *IEEE Transactions on Power Delivery*, 15(4):1212–1220, October 2000.

[37] C.-H. Lin and C.-H. Wang. Adaptive wavelet networks for power-quality detection and discrimination in a power system. *IEEE Transactions on Power Delivery*, 21(3):1106–1113, July 2006.

[38] Z. Lu, J. S. Smith, Q. H. Wu, and J. Fitch. Identification of power disturbances using the morphological transform. *Transactions of the Institute of Measurement and Control*, 28(5):441–455, 2006.

[39] T. Radil, P. M. Ramos, F. M. Janeiro, and A. Cruz Serra. PQ monitoring system for real-time detection and classification of disturbances in a single-phase power system. *IEEE Transactions on Instrumentation and Measurement*, 57(8):1725–1733, August 2008.

[40] S. Mishra, C. N. Bhende, and B. K. Panigrahi. Detection and classification of power quality disturbances using S-transform and probabilistic neural network. *IEEE Transactions on Power Delivery*, 23(1):280–287, January 2008.

[41] E. Styvaktakis, M. H. J. Bolen, and I. Y. H. Gu. Expert system for classification and analysis of power system events. *IEEE Power Engineering Review*, 22(2):64–64, February 2002.

[42] S.R. Sternberg. Grayscale morphology. *Computer Vision, Graphics, and Image Processing*, 35(3):333–355, 1986.

[43] R. van den Boomgaard. *Mathematical morphology: Extensions towards computer vision.* PhD thesis, University of Amsterdam, 1992.

[44] M. H. Sedaaghi and Q. H. Wu. Real-time implementation of grey-scale morphological operators. *Electronics Letters*, 33(21):1761–1763, July 1997.

[45] A. Gasteratos, I. Andreadis, and P. Tsalides. Fuzzy soft mathematical morphology. *Vision, Image Signal Processing, IEE Proceedings*, 145(1):41–49, 1988.

[46] M.S. Hamid, N.R. Harvey, and S. Marshall. Genetic algorithm optimisation of multidimensional grey-scale soft morphological filters with applications in archive film restoration. *Circuits and Systems for Video Technology, IEEE Transactions*, 13(5):406–416, May 2003.

[47] S. Mallat. *A Wavelet Tour of Signal Processing.* Academic Press, San Diego, CA.

[48] P. K. Ghosh and K. Deguchi. *Mathematics Of Shape Description: A Morphological Approach to Image Processing and Computer Graphics.* John Wiley & Sons, Asia, 2008.

[49] J. Goutsias and H. J. A. M. Heijmans. Nonlinear multiresolution signal decomposition schemes – Part I: Morphological pyramids. *IEEE Transactions on Image Processing*, 9(11):1862–1876, 2000.

[50] H. J. A. M. Heijmans and J. Goutsias. Nonlinear multiresolution signal decomposition schemes – Part II: Morphological wavelets. *IEEE Transactions on Image Processing*, 9(11):1897–1913, 2000.

[51] W. Sweldens. Wavelets and the lifting scheme: A 5 minute tour. *Zeitschrift für Angewandte Mathematik und Mechanik*, 76:41–44, 1996.

[52] W. Sweldens. The lifting scheme: A construction of second generation wavelets. *SIAM Journal on Mathematical Analysis*, 29(2):511–546, 1998.

[53] P. T. Jackway. Morphological scale-space. In *Proceedings of the 11th IAPR International Conference on Pattern Recognition, Conference C: Image, Speech and Signal Analysis*, volume 3, pages 252–255, The Hague, Netherlands, 30 August-3 September 1992.

[54] A. Witkin. Scale-space filtering: A new approach to multi-scale description. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 150–153, March 1984.

[55] F. Meyer and P. Maragos. Morphological scale-space representation with levelings. In *Scale-space Theories in Computer Vision, Lecture Notes in Computer Science*, volume 1682, pages 187–198, Berlin / Heidelberg, January 1999. Springer.

[56] R. van den Boomgaard and H. Heijmans. Morphological scale-space operators.

[57] P. Maragos. Pattern spectrum and multiscale shape representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):701–716, July 1989.

[58] E. S. Cardoso and E. A. B. da Silva. Design of nonlinear multi-resolution decomposition schemes using morphological operations. *Electronics Letters*, 36(9):843–845, Apr 2000.

[59] C. H. H. Chu and E. J. Delp. Impulsive noise suppression and background normalization of electrocardiogram signals using morphological operators. *IEEE Transactions on Biomedical Engineering*, 36(2):262–273, February 1989.

[60] M. Hossein Sedaaghi. *Morphological Filtering In Signal/Image Processing*. PhD thesis, The University of Liverpool, Department of Electrical Engineering and Electronics, 1998.

[61] N. S. D. Brito, B. A. Souza, and F. A. C. Pires. Daubechies wavelets in quality of electrical power. In *8th International Conference on Harmonics And Quality of Power*, volume 1, pages 511–515, 14-16 October 1998.

[62] R. R. Coifman and M. V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38(2):713–718, 1992.

[63] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N-C. Yen, C.C. Hung, and H. H. Liu. The empirical mode decomposition

and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceeding of The Royal Society*, 454:903–995, 1998.

[64] B. Weng, M. Blanco-Velasco, and K. E. Barner. Baseline wander correction in ECG by the empirical mode decomposition. In *Proceedings of the IEEE 32nd Annual Northeast Bioengineering Conference*, pages 135–136, Easton, PA, April 2006.

[65] S. He, Q. H. Wu, J. Y. Wen, J. R. Saunders, and R. C. Paton. A particle swarm optimizer with passive congregation. *Biosystems*, 78:135–147, 2004.

[66] L. Yaroslavsky and M. Eden. *Fundamentals of Digital Optics*. Springer, Boston, 1996.

[67] I. Aizenberg and C. Butakoff. Frequency domain median-like filter for periodic and quasi-periodic noise removal. In *SPIE Proceedings of Image Processing: Algorithms and Systems*, pages 181–191, 2002.

[68] R. Arthur and Jr. Weeks. *Fundamentals of Electronic Image Processing*. Wiley-IEEE Press, January 1996.

[69] M. Shao and K. E. Barner. Optimization of partition-based weighted sum filters and their application to image denoising. *IEEE Transactions on Image Processing*, 15(7):1900–1915, July 2006.

[70] A. S. Liu, R. B. Wu, and Y. C. Yu. A two-phase full-wave GA optimization for w-band image rejection waveguide filter design. In *Proceedings of 2003 IEEE Antennas and Propagation Society International Symposium*, volume 2, pages 60–63, Columbus, USA, 22-27 June 2003.

[71] N. R. Harvey and S. Marshall. The use of genetic algorithms in morphological filter design. *Signal Processing: Image Coummunication*, 8(1):55–71, January 1996.

[72] L. Pessoa and P. Maragos. MRL-filters and their adaptive optimal design for image processing. In *Mathematical Morphology and its Application to Image and Signal Processing*, pages 155–162. Kluwer Academic Publishers, Boston, USA, 1996.

[73] Z. Lu, M. S. Li, W. J. Tang, and Q. H. Wu. Multi-objective optimization of reactive power dispatch using a bacterial swarming algorithm. In *Proceedings of the 26th Chinese Control Conference*, volume 3, pages 460–464, Zhangjiajie, China, 26-31 July 2007.

[74] W. J. Tang, Q. H. Wu, and J. R. Saunders. A bacterial swarming algorithm for global optimization. In *Proceedings of IEEE Congress on Evolutionary Computation*, pages 1207–1212, Singapore, 25-28 September 2007.

[75] J. Kennedy and R. C. Eberhart. Particle swarm optimization. In *Proceedings of IEEE International Conference on Neural Networks*, volume IV, pages 1942–1948. IEEE Press, Piscataway, NJ, 1995.

[76] R. C. Eberhart and J. Kennedy. A new optimizer using particle swarm theory. In *Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, pages 39–43. Kluwer Academic Publishers, Nagoya, Japan, 1995.

[77] M. Clerc and J. Kennedy. The particle swarm - explosion, stability, and convergence in a multidimensional complex space. *IEEE Transactions on Evolutionary Computation*, 6(1):58–73, February 2002.

[78] J. K. Parrish and W. M. Hamner. *Animal groups in three dimensions.* Cambridge University Press, Cambridge, UK, 1997.

[79] P. Kuosmanen, P. Koivisto, H. Huttunen, and J. Astola. Shape preservation criteria and optimal soft morphological filtering. *Journal of Mathematical Imaging and Vision*, 5(4):319–335, December 1995.

[80] K. M. Passino. Biomimicry of bacterial foraging for distributed optimization and control. *IEEE Control Systems Magazine*, 22(3):52–67, June 2002.

[81] H. C. Berg. Motile behavior of bacteria. *Physics Today*, 53(1):24–29, January 2000.

[82] J. Bar Tana, B. J. Howlett, and D. E. Koshland. Flagellar formation in *Escherichia coli* electron transport mutants. *Journal of Bacteriology*, 130(2):787–792, May 1977.

[83] M. B. Miller and B. L. Bassler. Quorum sensing in bacteria. *Annual Review of Microbiology*, 55:165–199, 2001.

[84] F. Takens. Detecting strange attractors in turbulence. *Lecture Notes in Mathematics*, 898:366–381, 1981.

[85] J. D. Farmer and J. J. Sidorowich. Predicting chaotic time series. *Physical Review Letters*, 59(8):845–848, Aug 1987.

[86] P. Bryant, R. Brown, and H. D. I. Abarbanel. Lyapunov exponents from observed time series. *Physical Review Letters*, 65(13):1523–1526, September 1990.

[87] D. S. Broomhead and G. P. King. Extracting qualitative dynamics from experimental data. *Physica D: Nonlinear Phenomena*, 20(2-3):217–236, June 1986.

[88] S. Neil Rasband. *Chaotic Dynamics of Nonlinear Systems.* Wiley-Interscience, Canada, 1989.

[89] G. L. Baker and J. P. Gollub. *Chaotic Dynamics: An Introduction.* Cambridge University Press, U.K., 1990.

[90] Saber N. Elaydi. *Discrete Chaos.* Chapman & Hall/CRC, U.S., 2000.

[91] T. Sauer, J. A. Yorke, and M. Casdagli. Embedology. *Journal of Statistical Physics*, 65(3-4):579–616, November 1991.

[92] P. Grassberger and I. Procaccia. Estimation of the Kolmogorov entropy from a chaotic signal. *Physical Review A*, 28(4):2591–2593, October 1983.

[93] A. M. Albano, J. Muench, C. Schwartz, A. I. Mees, and P. E. Rapp. Singular-value decomposition and the Grassberger-Procaccia algorithm. *Physical Review A*, 38(6):3017–3026, September 1988.

[94] J.-P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Reviews of Modern Physics*, 57(3):617–656, July 1985.

[95] W. Liebert and H. G. Schuster. Proper choice of the time delay for the analysis of chaotic time series. *Physics Letter A*, 142(2-3):107–111, 1989.

[96] Z. Lu, T. Y. Ji, and Q. H. Wu. A morphological transform for exponentially decaying DC-offset removal. *Electronics Letters*, 44(9):595–U104, April 2008.

[97] J. C. Gu and S. L. Yu. Removal of DC offset in current and voltage signals using a novel Fourier filter algorithm. *IEEE Transactions on Power Delivery*, 15(1):73–79, 2000.

[98] E. Kreyszig. *Advanced Engineering Mathmatics.* Wiley, New York, 1995.

[99] N. Lv, X. Yu, and J. Wu. A fault diagnosis model through G-K fuzzy clustering. In *Proceeding of 2004 IEEE International Conference on Systems, Man and Cybemetics*, volume 6, pages 5114–5118, October 2004.

[100] M. T. Johnson and R. J. Povinelli. Generalized phase space projection for nonlinear noise reduction. *Physica D: Nonlinear Phenomena*, 201(3-4):306–317, February 2005.

[101] A. T. Johns and S. K. Salman. *Digital Protection for Power Systems.* Institution of Electrical Engineers, London, 1995.

[102] M. V. V. S. Yalla. A digital multifunction protective relay. *IEEE Transactions on Power Delivery*, 7(1):193–201, 1992.