Simple Morphological Filters in One Dimension

W. Harvey Holmes

Abstract

Morphological operations open many new ways to obtain novel heuristic solutions to a range of signal processing problems. This introduction to morphological filtering in one dimension also includes a number of new techniques based on morphological considerations. Applications include removal of baseline wander, sophisticated clipping methods, removal of impulsive noise, and isolation of R or QRS complexes in ECGs.

1 Introduction

Morphology is the study of shape. Morphological filters are essentially nonlinear operations on a signal that aim to change its shape in desired ways.

As an example, the morphological opening operation (see below) is widely used to flatten the components of a 2D greyscale image that are brighter than the background. It can also be used to remove (or separate out) objects that are higher than the surrounding ground.

Other uses of morphological filters include the removal of baseline wander or trends, the selection of significant features (e.g. peaks or edges) of a signal for further analysis, and the elimination of some undesirable forms of noise or signal corruption. The application areas include ECG analysis, among many others.

Clearly, nonlinear morphological operations would not be suitable if it is important not to distort the signal. Instead, they are used to change it in such a way that it is more suitable for further analysis. One example is pitch estimation, where the aim is to remove certain effects such as baseline wander that make pitch estimation difficult.

Much of the literature is either very abstract (mathematical morphology), or it concentrates on 2D applications. However, here we consider only 1D signals. Also, these are only sampled data (digital) signals. This makes the mathematics much simpler, compared for example to the standard references Serra and Vincent 1992; Maragos and Schafer 1987a; Maragos and Schafer 1987b; Maragos 2005 - there is no set theory, no topology, no lattice theory etc. (though many morphological operations have been motivated by set theory, in particular).

2 Classic morphological operations

All morphological operations are traditionally based on two basic morphological operations, erosion and dilation. Erosion obtains the smallest weighted value within a window, also called a structuring element (SE), whereas dilation acquires the largest weighted value within the SE.

Typically, the SE will be either a finite length window B or a finite length signal g_n (which will often be symmetrical and non-negative, with a peak in the middle and tapering towards the ends). See Section 9 for a discussion of the choice of $g\left(n\right)$.

The following operations are realized in various Matlab functions (`dilation.m`, `erosion.m`, etc.). These handle both set-based definitions using only B, and function-based defintions using g_n. These operations are evaluated by the scripts `setops.m.` and `funcops.m` (which differ mainly in the figure details).

These operations could be applied to any function, but to illustrate the six classic morphological operations, as well as some derived operations, the examples in Figures 1-7 below all use the signal

$$f_n = \sin(17\omega_n) + 0.7\cos(19\omega_n) - (\omega_n - \frac{\pi}{2})^2, \;\; \omega_n \in [0, \pi] \tag{1}$$

The structuring element in the examples has support $B = 41$ for set definitions and $B = 81$ for function definitions (see below). For function operations there is a choice in the software `structfunc.m` of several different structuring functions, such as the quadratic, with $B = 81$:

$$g_n = 1 - \left(\frac{n}{41} - 1\right)^2, \; n = 1 : 81 \tag{2}$$

Note that, in some cases to follow, improved results could be obtained by changing the SE, depending on the objective of the operation. But for consistency, the same SEs have been used throughout.

All of the following operations are implemented in functions called by the scripts setops.m and funops.m. See also

3 Dilation and erosion by a set

These basic operations (also called flat dilation/erosion) for a 1D signal $f(n)$ are defined (Maragos 2005; Maragos and Schafer 1987a) by

$$\begin{aligned}
\text{Dilation} : (f \oplus B)(n) &\triangleq \max_{m \in B} f(n - m) && (3)\\
&\equiv \max_{n-m \in B} f(m) && (4)\\
\text{Erosion} : (f \ominus B)(n) &\triangleq \min_{m \in B} f(n - m) && (5)\\
&\equiv \min_{n-m \in B} f(m) && (6)
\end{aligned}$$

where B is a finite length window. In words, dilation is a moving local maximum and erosion is a moving local minimum. (Note that the generalized operations in the next subsection are equivalent, apart from a dc shift, to these definitions if we choose $g(n)$ to be a constant signal within its finite length, hence the adjective "flat".)

Dilation flattens the peaks of $f(n)$ and increases the rest of the signal, while erosion fills in its valleys and decreases the rest of the signal. Also, $f(n)$ is bounded by its dilation and erosion:

$$(f \ominus B)(n) \leq f(n) \leq (f \oplus B)(n) \tag{7}$$

These operations can all be realized in Matlab using `movmax.m` or `movmin.m`. They are illustrated in Figure 1.

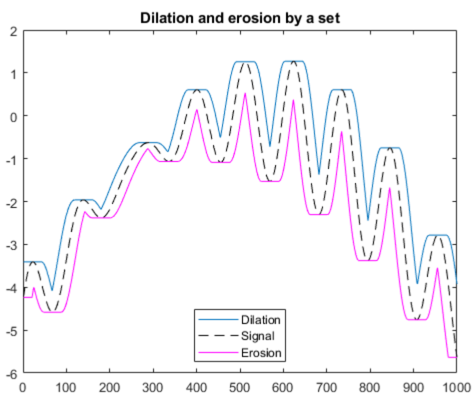

Figure 1: Dilation and erosion of the signal $f(n)$ by a set with support B=41

4 Dilation and erosion by a function

These generalized basic operators for a 1D signal $f(n)$ and structure element $g(n)$ are defined (Maragos 2005; Maragos and Schafer 1987a) by the "additive convolutions"

$$\begin{aligned}
\text{Dilation} : (f \oplus g)(n) &\triangleq \max_{m \in B} [f(n-m) + g(m)] && (8)\\
&\equiv \max_{n-m \in B} [f(m) + g(n-m)] && (9)\\
&\equiv \max_{m-n \in B} [f(n) + g(m-n)] && (10)\\
\text{Erosion} : (f \ominus g)(n) &\triangleq \min_{m \in B} [f(n-m) - g(m)] && (11)\\
&\equiv \min_{n-m \in B} [f(m) - g(n-m)] && (12)\\
&\equiv \min_{m-n \in B} [f(n) - g(m-n)] && (13)
\end{aligned}$$

where B is the support of $g(n)$. However, with these definitions (7) applies only if $\underset{m \in B}{g(m)} \geq 0$, which is usually the case.

Note that, if $\underset{m \in B}{g(m)} = 1$, the these definitions reduce to the set operations in the previous subsection.

> An issue I have had using Matlab is that these definitions don't "hug" the function $f(\cdot)$ in the way I wanted (and expected). So I have shifted g downwards so that $\max(g) = 0$. Then all the results look the way I would like.

See Figure (2) for illustrations of these operations. In my experience with ECG signals, these definitions give better results, in some ways, than the set definitions above.

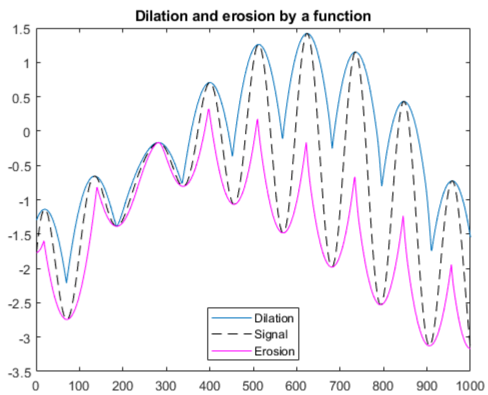

Figure 2: Dilation and erosion of the same signal $f(n)$ by the function $g(n)$ with B=91

5 Opening and closing

Opening and closing are two extended morphological operators based on dilation and erosion.

The opening of a signal is erosion followed by dilation, and closing is dilation followed by erosion:

$$\begin{aligned}
\text{Opening} : f \circ g &\triangleq (f \ominus g) \oplus g && (14)\\
\text{Closing} : f \bullet g &\triangleq (f \oplus g) \ominus g && (15)
\end{aligned}$$

Either definition above of dilation and erosion may be used.

Like (7), these operations also bound the function:

$$(f \circ g)(n) \leq f(n) \leq (f \bullet g)(n) \tag{16}$$

(assuming that $\underset{m\in B}{g(m)} \geq 0$ if the function definitions are used).

Opening and closing also work as morphological filters with clipping effects: opening clips peaks and closing fills valleys. See Figure 3 for examples. (The names opening and closing make more sense in 2D, where they originated.)

The sizes of the objects removed depend on the specified SE. Short SEs can only be used to remove small objects, whereas long SEs are needed to remove large objects. Sometimes a series of different window sizes and varying SEs are used to eliminate objects of various sizes.

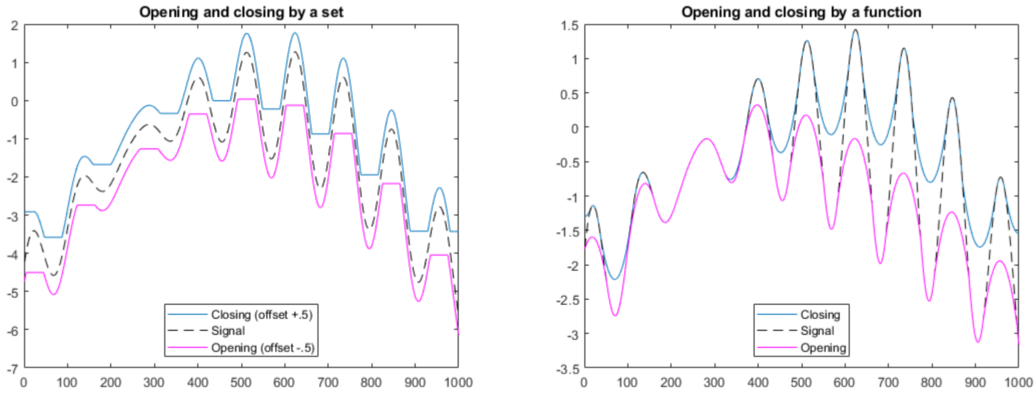

Figure 3: Opening and closing by a set and a function. Closing fills the valleys, erosion clips the peaks. The curves with the set operations are offset so that the differences from the original signal are clearly visible, as otherwise the only differences from the original signal are the top and bottom clippings.

6 Top-hat and bottom-hat operators

Finally, the top and bottom hat operations are

$$\begin{aligned} \text{Top} - \text{Hat} : f_{TH}(n) &\triangleq f(n) - (f \circ g)(n) & (17) \\ \text{Bottom} - \text{Hat} : f_{BH}(n) &\triangleq f(n) - (f \bullet g)(n) & (18) \end{aligned}$$

where the opening and closing could be either set-based or function-based. The top-hat operation essentially picks out parts of the signal that stand above the rest, and the bottom-hat operation picks out valleys that are deeper than the rest, as illustrated in Figure 4.

The bottom-hat operation works on valleys in the same way as the top-hat operation works on peaks.

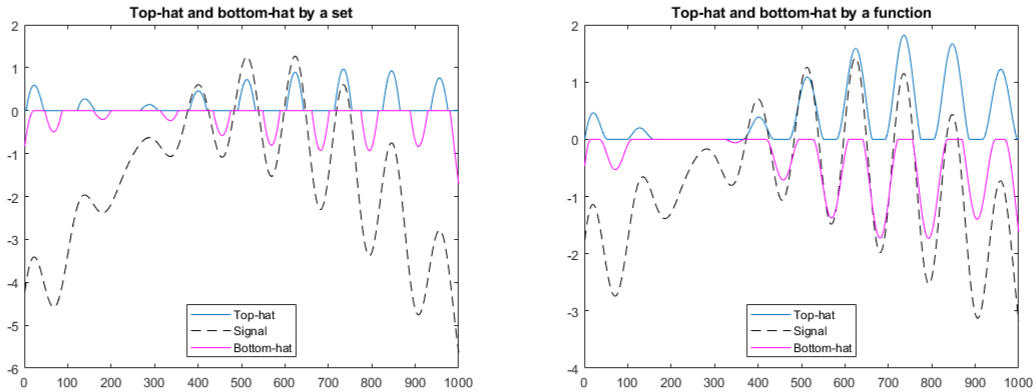

Figure 4: Top-hat and bottom-hat operations using set and function definitions.

7 Non-standard definitions

The above are the standard definitions of the six basic morphological operations.

However, erosion and closing could also have been defined as

$$\begin{aligned}\text{Dilation}: (f \oplus g)(n) &\stackrel{?}{=} \max_{m \in B} [f(n-m)\, g(m)] && (19)\\ \text{Erosion}: (f \ominus g)(n) &\stackrel{?}{=} \min_{m \in B} [f(n-m)\, g(m)] && (20)\end{aligned}$$

These are essentially sliding-window nonlinear convolutional FIR filters. They may not have the nice set-theoretic properties of the classic definitions, but can have similar effects on the signal. The derived operations above (opening etc.) stay the same. These operations are realized, with two different interpretations of the nonlinear convolution, in the function `dilerosconv.m`, which is called by the script `convops.m` .

However, I haven't been able to get good results with them so far, although they do smooth the function in a nice way, albeit with different amplitudes. One problem is that they produce outputs that don't bound $f(n)$ in the way (7) does with the original definitions. (This can be "fixed" by adjusting some signal amplitudes, as done in `convops.m`.)

Many other nonlinear operations can also be used in some cases. For example, the median and rank order operations are often very useful, though computationally intense.

8 Derived operations

There are many derived operations, based on those in Section 2, that have been or could be used, and many of these may be very useful. These are generally targeted at some particular desired outcome. Some of them are as follows.

8.1 Removal of baseline wander

Removing baseline wander is one of the most important operations that morphological filters can do very well. The first thing is to estimate the actual baseline $b(n)$.

Zhang and Bae (2012) and others have used the simple peak extractor

$$\begin{aligned} b &= ((f \oplus g) + (f \ominus g))/2 && (21)\\ c &= f - b && (22)\end{aligned}$$

in order to detect the QRS complex in ECG recordings accurately and quickly, instead of a series of advanced openings and closings as in other literature. The baseline estimate is (21), roughly half way between clipped peaks and filled troughs. This is then subtracted from the signal to give $c(n)$, which is just the smoothed positive and negative departures from the baseline, or the de-trended signal.

I have found that a better measure of baseline wander than in (21) is

$$b = ((f \bullet g) + (f \circ g))/2 \qquad (23)$$

This gives a smoother baseline estimate if the function definitions are used instead of the set definitions. In either case the choice of SE in relation to the signal f is important, especially its support B.

The detrended signal (22), with this estimate of baseline wander, can also be expressed in terms of the top-hat and bottom-hat operations as

$$b = (f_{TH} + f_{BH})/2 \qquad (24)$$

These operations (22) and (23) to estimate and remove baseline wander are realized in the scripts `setops.m` and `funcops.m`. They are illustrated in Figure 5 using the function definitions. If the set definitions are used instead, the result is distorted, but it is still very suitable for further processing to estimate the frequency, for example (it is well known that baseline wander upsets many frequency estimation methods).

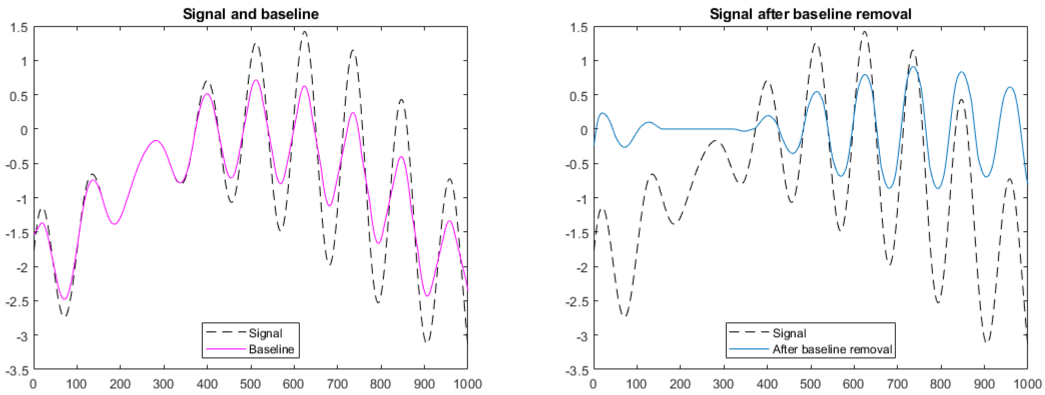

Figure 5: Baseline estimate using (23) and the resulting signal after subtracting the baseline estimate as in (22), using the function definitions.

Another possible nonlinear estimate of baseline wander is provided by the sliding median filter (Maragos 2005), though it is much more computationally intensive than (23). This is defined as

$$b(n) = \underset{m \in B}{\mathrm{median}} \left(f(n-m) \right) \tag{25}$$

A linear version of this would be $b(n) = \underset{m \in B}{\mathrm{mean}} \left(f(n-m) \right)$, but this is not as robust.

8.2 Clipping

Opening gives a signal that is clipped at the top, and closing one that is clipped at the bottom. A signal that is clipped at both top and bottom can be produced by opening followed by closing:

$$f_C = (f \circ g) \bullet g \tag{26}$$

as illustrated in Figure 6. This is quite different from simply clipping the signal at certain upper and lower levels.

It seems that repeated clipping operations hardly alter the clipped signal - i.e. this form of clipping is (almost) an idempotent operation!

Also, with appropriate length of g, this operation might be suitable for baseline estimation in place of the methods in the preceding subsection.

Several other ways of clipping are explored in `funcops.m`, but the above is the best solution I have found.

8.3 Impulse noise removal

Morphological operations are well adapted to the removal of impulsive noise (i.e. loud noise of short duration), provided appropriate SEs are used. An example is in Figure 7, which used the clipping function with B about the size of the zero crossing interval of the interfering riples in the decaying sine wave interference.

8.4 Use as quasi-matched filters to extract ECG features

The R pulse in an ECG recording is of large amplitude and narrow width (c. 10 samples in `1071.mat`). If B is chosen to be small (about the same width as the R pulse), the top-hat operation tends to select the (positive) R pulse and ignore features of longer duration. This is qualitatively similar to the behaviour of an ideal matched filter.

To extract the QRS complex, de-trending (as in the previous subsection) using a small B is suitable, since the durations of the Q and S pulses are normally similar to that of the R pulse and shorter than the P or T pulses (Oweis and Al-Tabbaa 2014), which will be rejected or greatly reduced by this operation.

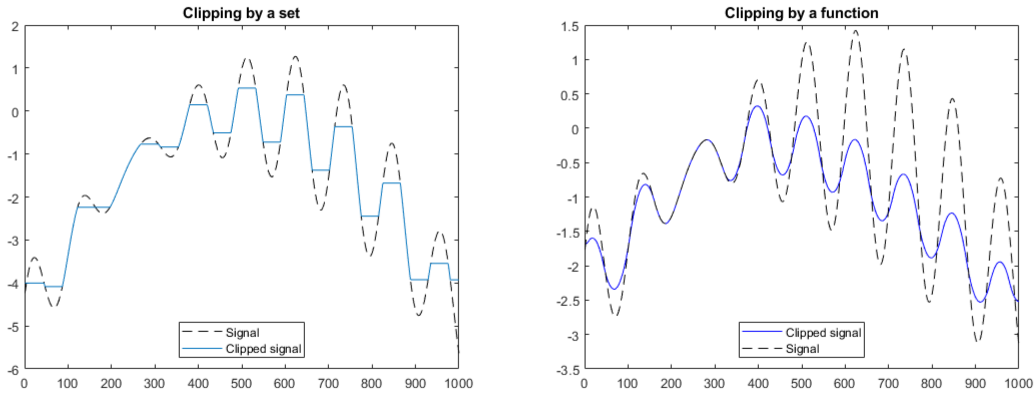

Figure 6: Signal clipped at both top and bottom with set and function definitions.

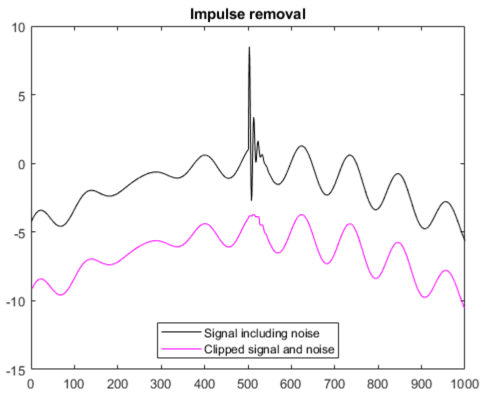

Figure 7: Removal of impulsive noise using the set definitions. In this case the impulsive noise was a decaying sine wave, and $B = 5$.

An example, using a segment from `1071.mat`, is shown in Figure 8. This is an almost noiseless example, but it certainly shows promise. The top-hat operation with a SE about the length of the R-pulse effectively selects them and removes the baseline wander. To obtain the QRS complex, a sum of the top-hat and bottom-hat operations is very effective.

9 Choice of structuring element and morphological operations

There is considerable freedom in using morphological methods. It is important to make these choices based on the desired outcome.

The length of the SE (B, the length of $g(n)$) is a major parameter that affects how they operate. For example, with baseline removal (Section 8.1), signal features shorter than B will tend to pass through the operation almost unaltered, whereas longer features tend to be greatly reduced or even elimininated entirely. The shape of $g(n)$ will also affect how they work.

Then there is the choice of the actual morphological operations themselves, which is almost unlimited. The discussion in Section 2 can serve as a guide to choice of methods. However, the operations there can be combined in many ways to form new methods, and new ones can be invented if necessary, e.g. as in Section 7. The morphological approach opens up many new ways to obtain good heuristic solutions to many problems (see https://en.wikipedia.org/wiki/Heuristic).

One way to get new methods is to repeat the operations with different SEs, and this is often done. For

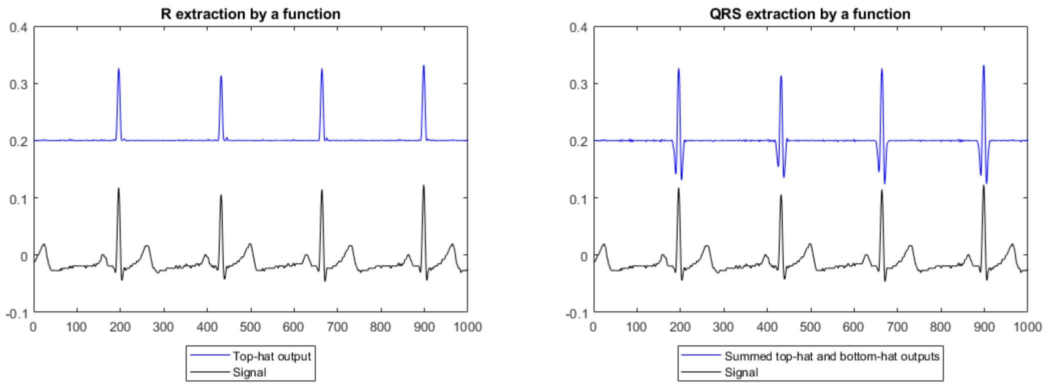

Figure 8: Matched filter behaviour of the top-hat and baseline removal operations to select the R pulse or the QRS complex of an ECG signal using $B = 81$.

example, features of different durations can be targeted one after the other in this way. A simple example is clipping (26).

10 Applications to ECG processing

There have been quite a few papers that apply these concepts for the analysis of ECG signals, but they haven't caught on, probably largely because of the difficulty of understanding the theory.

I would propose morphological operations as part of the pre-processing of an ECG signal. Exactly what it should be would depend on the purpose of the processing.

For IBI analysis, the processing could for example be as follows:

1. The first stage would be linear filtering to remove noise components outside the frequency band of a typical ECG signal, but this should be done so it doesn't significantly distort QRS pulses. A simple highpass filter may be used as well to help reduce baseline wander, but it must also be one that doesn't distort the underlying ECG signal.

2. It may seem attractive to then use noise reduction techniques such as spectral subtraction, but I think the noise is far too non-stationary to give much chance of success with such methods.

3. This filtering would be followed by a series of morphological operations as above. This would have at least three functions:

   (a) To remove baseline wander. This is best done using a relatively long SE compared to the width of the R pulse, but short compared to the pulse period (e.g. $B = 41$), as in Figure 5.

   (b) To do some nonlinear noise reduction, e.g. to remove strong impulsive noise as in figure 7. There may be other noise reduction techniques that could be developed to do this.

   (c) To select significant peaks for further analysis by the IBI estimator to follow. For example, we could follow drift removal and noise reduction by further morphological analysis using a short SE (e.g. $B = 11$) and top-hat and/or bottom-hat operations as in Figure 8.

4. The resulting signal would then go to a good pitch (IBI) detector such as SEEVOC. Alternatively, for detecting arrhythmia, epilepsy or other pathological conditions, it would go to appropriate further stages of processing (though SEEVOC can be modified to give information for these conditions).

The optimum sequence of morphological operations still has to be determined.

References

Maragos, P. (2005). “Morphological filtering for image enhancement and feature detection”. In: *The Image and Video Processing Handbook*. Ed. by A. C. Bovik. 2nd ed. Vol. The Image and Video Processing Handbook. Elsevier Academic Press. Chap. 3.3, pp. 135–156.

Maragos, P. and R. W. Schafer (1987a). “Morphological filters-Part 1: Their set-theoretic analysis and relations to linear shift-invariant filters”. In: *IEEE Transactions on Acoustics, Speech and Signal Processing* ASSP-35.8, pp. 1153–1169.

– (Aug. 1987b). “Morphological filters-Part 11: Their relations to median, order-statistic, and stack filters”. In: *IEEE Transactions on Acoustics, Speech and Signal Processing* ASSP-35.8, pp. 1170–1184.

Oweis, R. J. and B. O. Al-Tabbaa (2014). “QRS detection and heart rate variability snalysis: A survey”. In: *Biomedical Science and Engineering* 2.1, pp. 13–34. DOI: DOI:10.12691/bse-2-1-3. URL: http://pubs.sciepub.com/bse/2/1/3.

Serra, J. and L. Vincent (Jan. 1992). “An overview of morphological filtering”. In: *Circuits, Systems and Signal Processing* 11.1, pp. 47–108.

Zhang, C. F. and T.-W. Bae (Mar. 2012). “VLSI friendly ECG QRS complex detector for body sensor networks”. In: *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 2.1, pp. 52–59.