Image Processing Techniques for Sector Scan Sonar

by

Lukas Anton Hendriks

Thesis presented at the University of Stellenbosch in partial fulfilment of the requirements for the degree of

Master of Science in Engineering

Department of Electrical Engineering
University of Stellenbosch
Private Bag X1, 7602 Matieland, South Africa

Study leader: Mr J Treurnicht

October 2009

Chapter 4

Image Analysis

Chapter 3 provides a good background of the technology used and the current applications of sector scan sonar. However, at this point it is necessary to take a more detailed look at the characteristics of the sonar images as well as the environment in which they are taken. Two very distinct real world examples of sonar images are examined, each providing its own difficulties. Firstly however, the characteristics of the sonar used for taking these images are examined, to provide some insight into the limitations of the technology and possible further processing.

4.1 Micron DST Sonar

All images were taken using a *Micron DST Sonar* manufactured by *Tritech International*. This sector scan sonar operates at a centre frequency of 700 kHz and uses a digital CHIRP with a bandwidth of 100 kHz. Using Eq. 3.2.2 and taking the speed of sound in water as 1500 m/s, the sonar has an attainable range resolution of 0.0075 m. It is however limited to a maximum of 800 range bins and to a maximum range of 75 m. It produces a single beam with a width of 30° in the vertical and 3° in the horizontal planes. This beam can be mechanically rotated in both direc-

tions and can survey a sector of up to 360°. Because of its small size and weight, 78.5 mm × 56 mm and 180 g (in water), it is ideal for use on any size AUV.

The following sections cover important attributes of the sonar. These attributes have a significant effect on the resulting images, and since the manufacturer provides little information about expected results, they were evaluated through experimentation.

4.1.1 Step Angle

The step angle refers to the angle between each sonar ping and can be set to one of four settings: 0.225°, 0.45°, 0.9° and 1.8°. The choice of scanning step has a direct influence on the speed of the scan. As should be apparent, a sweep at 0.9° intervals will take four times less than a sweep at 0.225°. A faster update rate can be beneficial for the tracking of objects or path planning. A larger step angle can result in a loss of angular resolution, however.

With a horizontal beamwidth of 3°, a small step angle will offer little improvement to the image detail, since the volume illuminated by each ping will largely overlap. Fig. 4.1 shows an example of three images taken of the same scene but at different step angles. The images were taken in a shallow water channel, or flume, containing standing water, which provides very little random scattering and therefore noise effects. The near identical results may be explained by the lack of noise in the system, but this does provide an example of how little influence the choice step angle can have on the resulting image. In contrast, the images of Fig. 4.2 were taken in a narrower flume with flowing water and randomly placed objects. This environment provides a lot of scattering and produces impulse-like noise, most prominent in the image taken with the smallest step angle (left). In this case the difference resulting from an decrease in step angle becomes more pronounced, resulting in a apparent loss of detail. This apparent loss of detail is however rather a

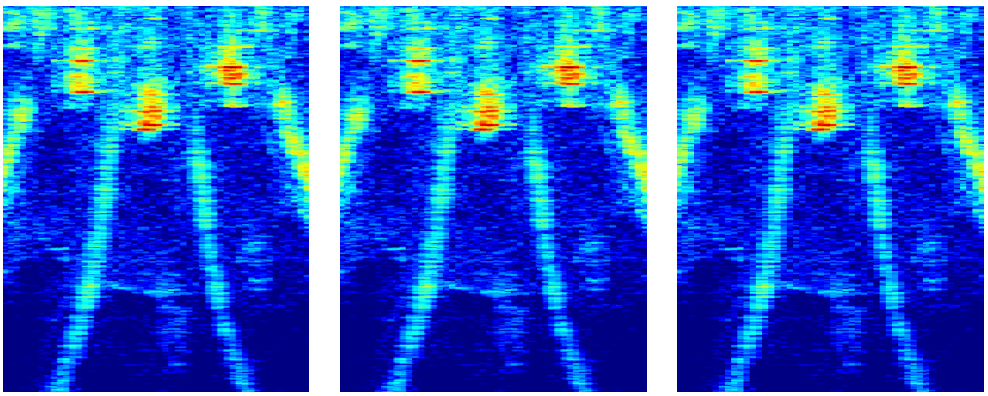


Figure 4.1: Three images of the same object, taken in a low noise environment at 0.225°, 0.45° and 0.9° step angles (from left to right).

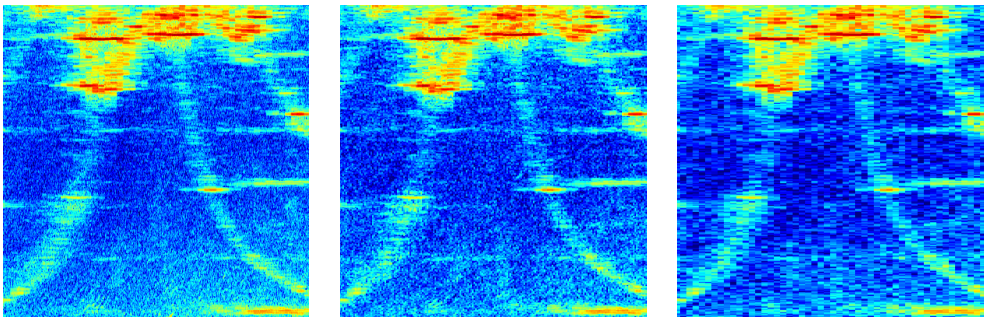


Figure 4.2: Three images taken in a high noise environment at 0.225°, 0.45° and 0.9° step angles (from left to right).

result of the decreased amount of data used to fill the same image size. In fact, on further inspection it seems that using a larger step angle actually reduces the amount of noise visible in the image, whilst the most apparent objects are equally visible in all three images.

From the first set of data it can be concluded that where the step angle is much smaller than the beamwidth, because of the great amount of overlap, it provides no apparent improvement. From the second set it should also be noted that a larger step angle and in effect taking a smaller number of samples (scans) can provide improved noise rejection.

It can therefore be concluded that as long as the step angle remains smaller than half the beamwidth, using a smaller step angle holds no apparent benefit to image quality.

4.1.2 Image Gain

The sonar *image gain* can adjusted at setup and provides a means to control the target strength needed for an object to appear in the image. Setting the gain too low will result in only the strongest targets being visible, basically setting the threshold for detection too low. Unfortunately, setting the gain too high will result in unwanted clutter appearing in the image. Fig. 4.3 shows an example of four images taken at increasing gain levels (the maximum intensity of each image is scaled to a value of one). As can be seen, with an increase in gain, the background level increases. The right-hand column shows the row of pixels containing the peak value of each image. What should be noted, is that a increase in gain does not result in clipping of the peak, but it does result in a loss of *contrast*. This loss in contrast can provide some difficulties in the segmentation process, and as such is an undesirable effect. It is therefore important to keep the gain as low as possible.

4.1.3 Horizontal Beamwidth and Automatic Gain Control

During the scanning process, as the beam is moved across an object, an increasing amount of energy is detected, resulting from a larger part of the beam covering the object. This *convolution-like* procedure is the reason why the attainable angle resolution is determined by the horizontal beamwidth of the sonar, as explained in Section 3.2.2. As a result a certain amount of object spread is expected, increasing the size of a detected object. The manufacturer provides no further information about the beam pattern beyond the main beam width. To investigate the

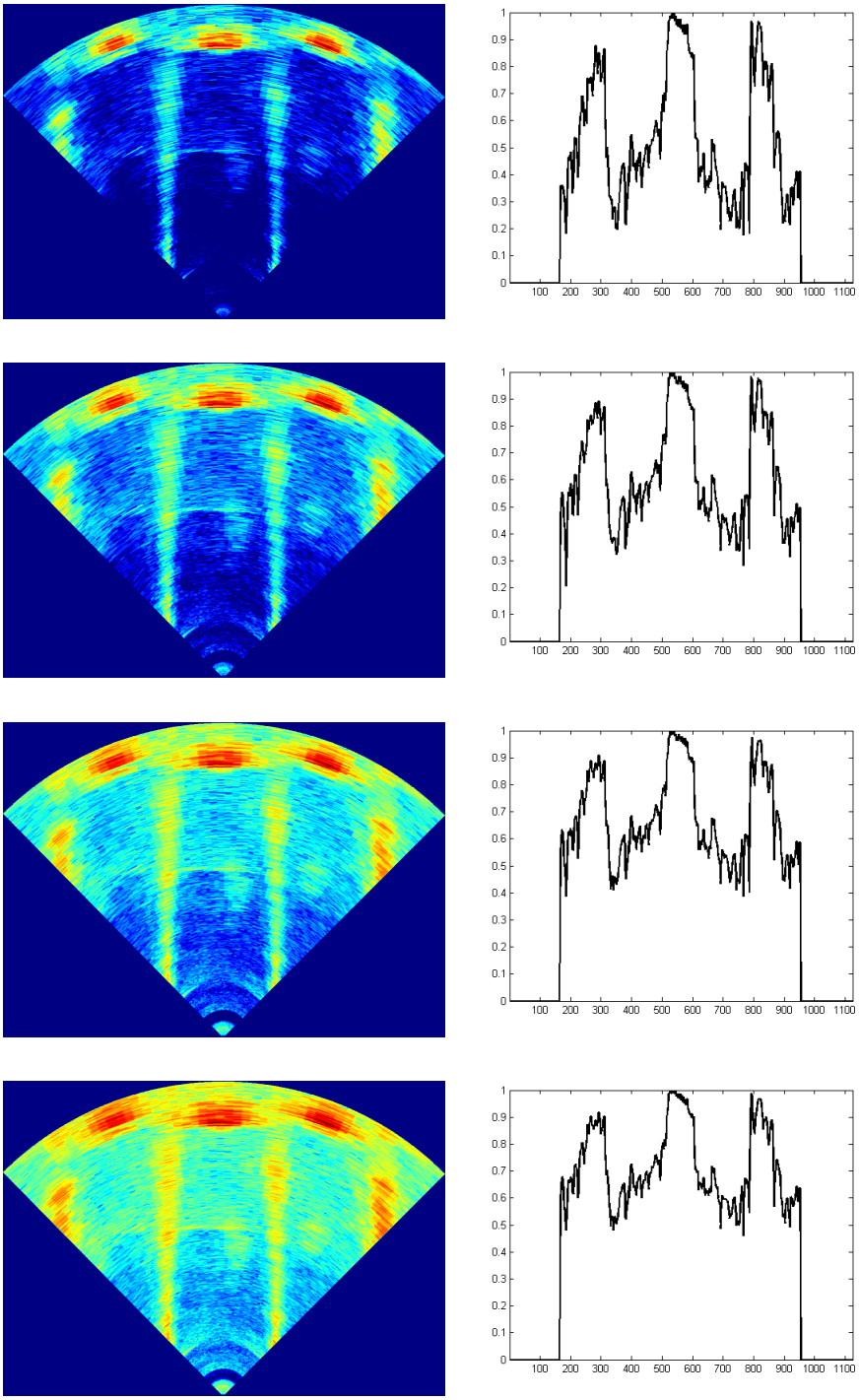


Figure 4.3: Set of images taken at increasing gain values.

amount of spread that can be expected and to get a better idea of the beam pattern, a set of experiments was conducted.

Perfectly spherical objects, which act as perfect reflectors in the direction of incidence, were used as targets. It is therefore expected that the returns should be a good approximation of the physical beam pattern. Although the size of the object should have very little influence on the reflected pattern, a larger object does reflect more energy, since it provides a greater area of reflection. A metal sphere with a diameter of 60 mm was used and the tests were conducted at a range of 5 m in a close to noiseless tank.

Fig. 4.4 shows a set of scans taken at exactly the same settings and at time intervals such that most of the energy from a previous scan has sufficiently dissipated. The images were scaled relative to the maximum value of image (a), for proper comparison. The images give an indication of the beamwidth of the main lobe as well as the first, second, third and in (a) and (f), the fourth sidelobe. Under these somewhat practical conditions, it seems that the 3° beamwidth, as specified by the manufacturer, is closer to a value of 5°. The slight difference in strength of sidelobe pairs, can be attributed to the sphere being slightly offset with respect to sonar centre axis. The beam pattern seems to indicate an asymmetrical third sidelobe, that could cause one side of an object return to undergo more spread than the other. Examination of further collected images, did not completely confirm this result, with both confirming and contradicting evidence appearing.

When comparing the cross-sections of the images, the effect of variable gain, as shown in the previous section, is clearly noticed. This is most visible in (c), where the third pair of sidelobes are just visible. When further comparing (b), (d) and (f) it should be noticed that although they have comparable maximum values, (f) clearly shows an extra sidelobe not visible in (b) and (d). This phenomenon is attributed to the *automatic gain control* of the receiver system. This is a standard part of any telecommunications receiver and therefore not completely unexpected. These

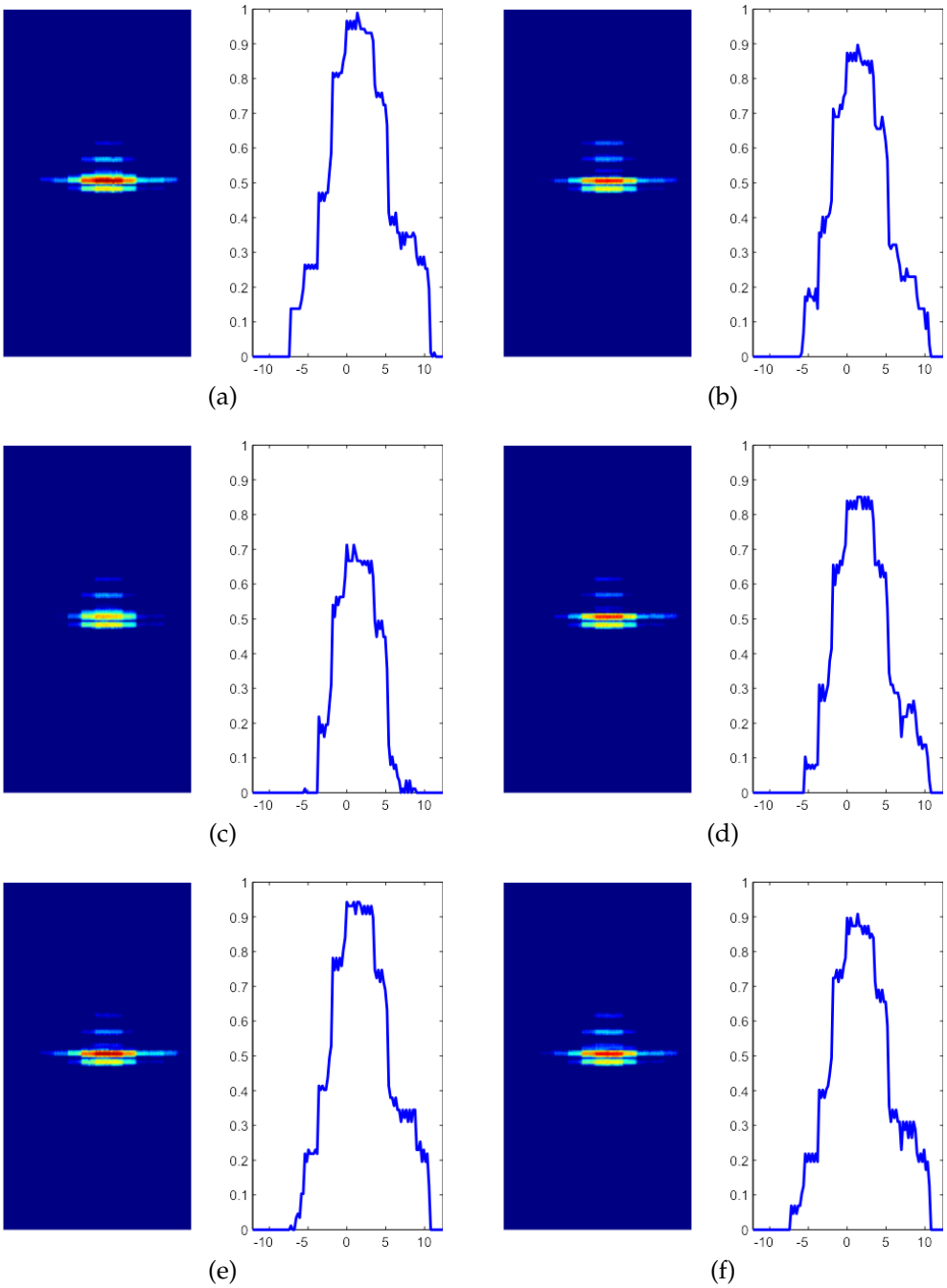


Figure 4.4: Six consecutive scans of a 60 mm metal sphere. Each scan is represented by the complete image and a cross-section containing the maximum value.

images were however, taken under very ideal circumstances, where there was almost no noise or reverberation present, and such variable gain is therefore an unpredicted complication. In most cases the maximum value of an image is scaled to one and such variation does not show up as readily, especially when further coupled with noise and reverberation.

As part of this experiment, the use of deconvolution using a estimate of the beam pattern, as described by [4, 7] was investigated. The idea is that since the image is the result of a linear convolution it should be possible to produce the inverse of this process. This should reduce the spreading effect and better resolve surveyed objects. The process is however extremely sensitive to noise and variations between the estimated and physical beam pattern. Experiments with this technique proved somewhat useless, mostly because any estimate of the beam pattern would depend on the amount of gain present in the specific image. A general estimate was therefore not possible and deconvolution therefore not applicable.

The test did however show how much spreading can be expected. A further test was done using a sphere with a diameter of 23 mm, which showed a beam pattern with the same size mainlobe and first sidelobe as the 60 mm sphere. It can therefore be concluded that the amount of spread is dependent on the amount of energy reflected or the size of the target. This is not a very unexpected result, but it can therefore be concluded that the return of an object will generally be larger than the physical object.

4.1.4 Vertical Beamwidth

Fig. 4.5 gives a representation of the area illuminated as result of the vertical part of the beam. This diagram illustrates the effects of range, floor depth and beamwidth on the area being illuminated. Depending on the range and depth, two distinct areas are visible in most images: the first without and the second with bottom reflections. This is visible in the

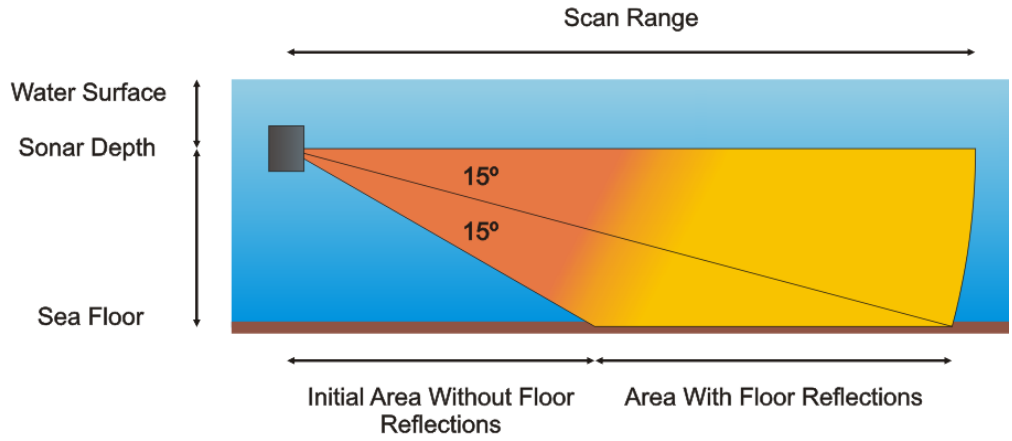


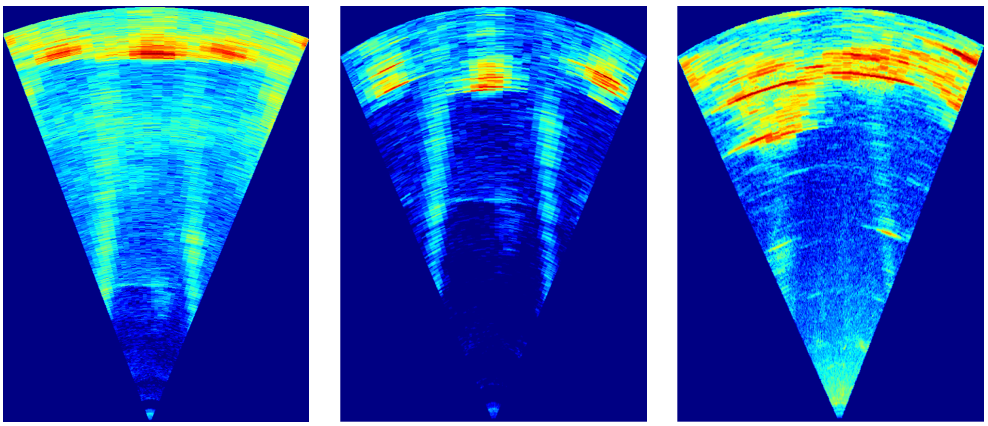
Figure 4.5: Diagram with a side on view of the area illuminated by the sonar beam.

images of Fig. 4.3, as the bottom of the image has a dark background and then a sudden change to a lighter one. If the range is great enough, another area of no bottom reflections appear at the far end of the image.

The bottom reflections are an unwanted effect with navigation in mind, since the AUV is only interested in objects directly in front of it. Strong reflections from the bottom could be classified as objects that could cause collisions, where, in reality, they offer no possible threat. It also adds to the background intensity level, and as a result, objects with a low target strength may be classified as background rather than a target. This is an unavoidable shortcoming of the sector scan sonar, but can be diminished by changing the mounting angle of the sonar as well as using a shorter range.

4.2 Two Data Sets

From the previous sections one can see that sector-scan sonars are far from perfect instruments. Working with the images they produce can result in quite a few interesting and challenging problems. The type of problems that are encountered tend to be greatly dependent on the spe-



(a) (b) (c)

Figure 4.6: Three examples of images taken in a water flume. (a) Image taken with a range of 10 m. (b) Image taken of an object at a range of 6 m. (c) Image taken in a narrow flume with random objects.

cific data being used, which is influenced by the environmental conditions and the type of objects being surveyed.

In order to identify a range of problems that could benefit from further processing, two very different types of images were analysed. The first set of data was taken in very shallow water flumes, at relative short ranges. The second was taken with the sonar on board an AUV in a harbour. Each of these provide a different set of challenges representative of what can be encountered in a variety of real-world applications.

4.2.1 Flume Images

A water flume is a shallow, narrow and long water channel. The use of sonar in shallow water produces a great deal of scattering from both the water surface and the floor, and images taken in this environment provide a good test bed for processing techniques, since it is somewhat of a worst-case scenario.

Fig. 4.6 shows three different images, each with slightly different char-

acteristics, taken in a water flume. Images (a) and (b) were taken in the same flume with a water depth of .41 m and width of 2 m, and concrete walls and floor with a width of 0.25 m. Half an oil drum was used as a target in both images at a range of 8.3 m and 5.1 m respectively and is visible as the area of high intensity at the top of each image. The most notable difference between the two images is the increase in clutter when the object distance is greater. Since the target strength of the object stays constant the amount of energy reflected gets smaller as the distance is increased. The reflections from the floor therefore becomes more pronounced and leads to the greater amount of clutter. Two shadow objects can also be seen to each side of the centre object. These are caused by multi-path effects, resulting from the confined environment. Fig. 4.6(c) was taken in a narrower flume, with a width of 1 m, and with a water depth of 0.7 m. Whereas the flume used for (a) and (b) was made of concrete, the flume used for (c) had glass walls and a metal floor. As a result the reflections from the walls are rather faint when compared with the previous images. This image was also taken in the presence of flowing water, and as a result a lot of impulse noise is visible which is not present in (a) and (b). These images are representative of the effects that can be expected when working in shallow water. It also provides an controlled, but real world, environment in which to study the return of a known target.

Fig. 4.7 shows a further analysis of Fig. 4.6. In the ideal case the image should contain only the object and the two walls, since those are the only obstacles present in the environment. This expected return is shown with the overlay of two vertical lines and half circle at the top of image (a). Fig. 4.7(b)–(c) shows the original image in (a) segmented using a global threshold increased by 10 % in each image. The threshold uses the intensity value of each pixel to determine whether it is a valid return or not. Global thresholding, as used in [19], [16] and [17], is a common way of segmenting sonar images. As used in Fig. 4.7 without any prior processing, is a good starting point for identifying problem areas. As such, (a)

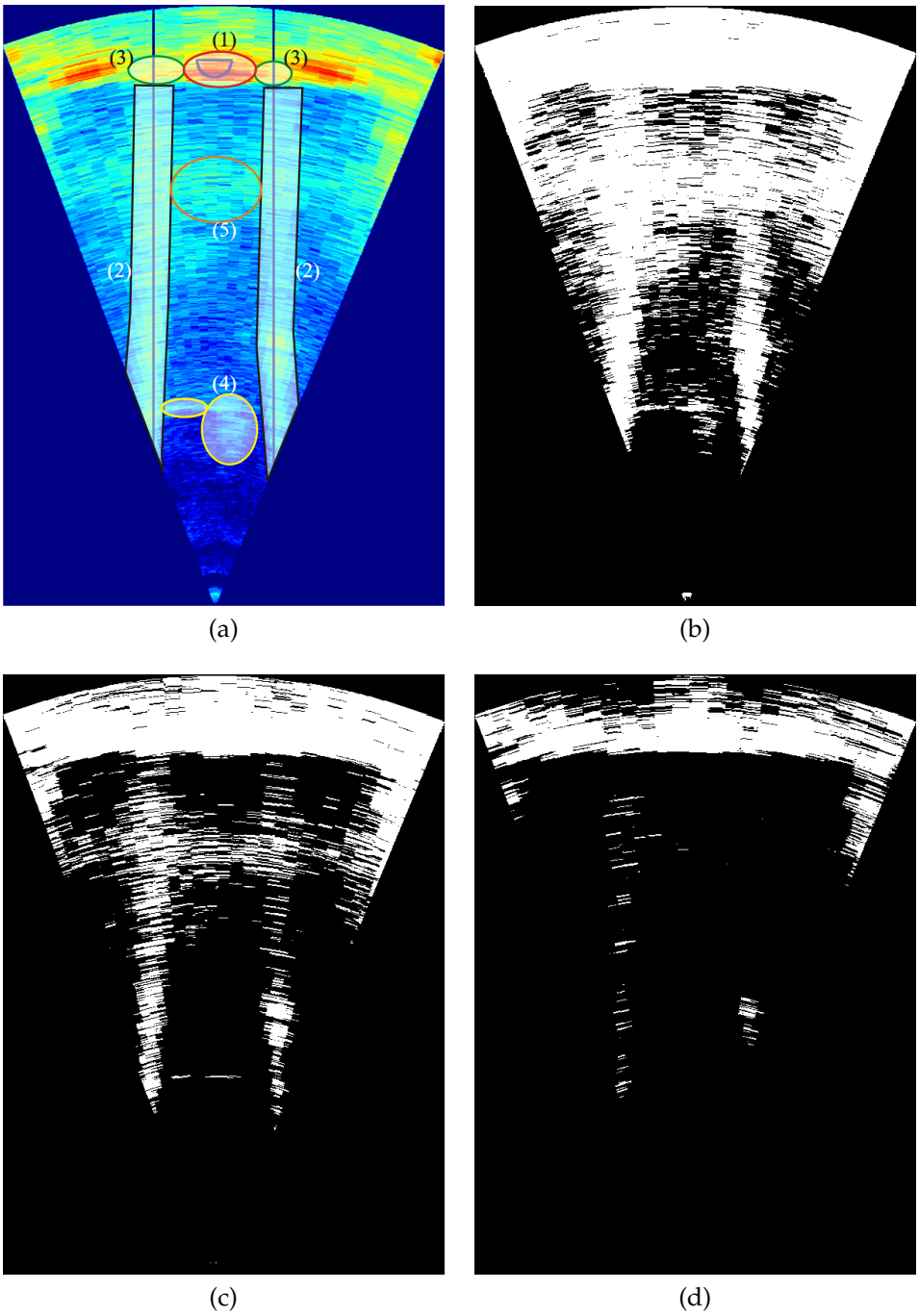


Figure 4.7: (a) Original flume image with highlighted areas of interest. (b)–(d) The image in (a) thresholded at 30 %, 40 % and 50 %.

identifies five areas of interest that will benefit from further processing:

1. Return from target object.
2. Returns from flume walls.
3. Spreading of target due to beam sidelobes.
4. Strong reflections from imperfections in flume floor.
5. Reflections from the flume floor with comparable intensity to the walls.

When comparing the area of highest intensity in (1) with the expected shape of the overlay it is clear that it is quite a bit larger than expected. Together with the further spreading in (3), the object appears even larger. As in the thresholded images, the area of high intensity caused by the target completely spans the width of the flume. The target has a physical width of 0.57 m and compared to the 2 m width of the flume, the representation could be greatly improved.

The flume walls in (2) correspond to physical obstacles in the water and as such need to be properly identified. The sheerness of the wall and the angle of the incident beam result in a small amount of reflected energy. When compared to its immediate background, the walls are visible. It is has a significantly lower level than the target and as shown in the thresholded images, the walls disappear when the threshold becomes too great. In (c) the walls are completely visible. The wall on the right does get broken up into pieces, but if these are connected it could be identified as a solid object. The bottom reflections of (5) hamper proper identification of the walls and need to be removed.

Further reflections of floor surface roughness or objects on the floor surface, as in (4), are common-place in sonar images. They are usually characterised by small reflections on the range axis, and depending on the target's strength and spreading, covers a greater area in the angle axis. These need to be removed, since they are clearly false targets.

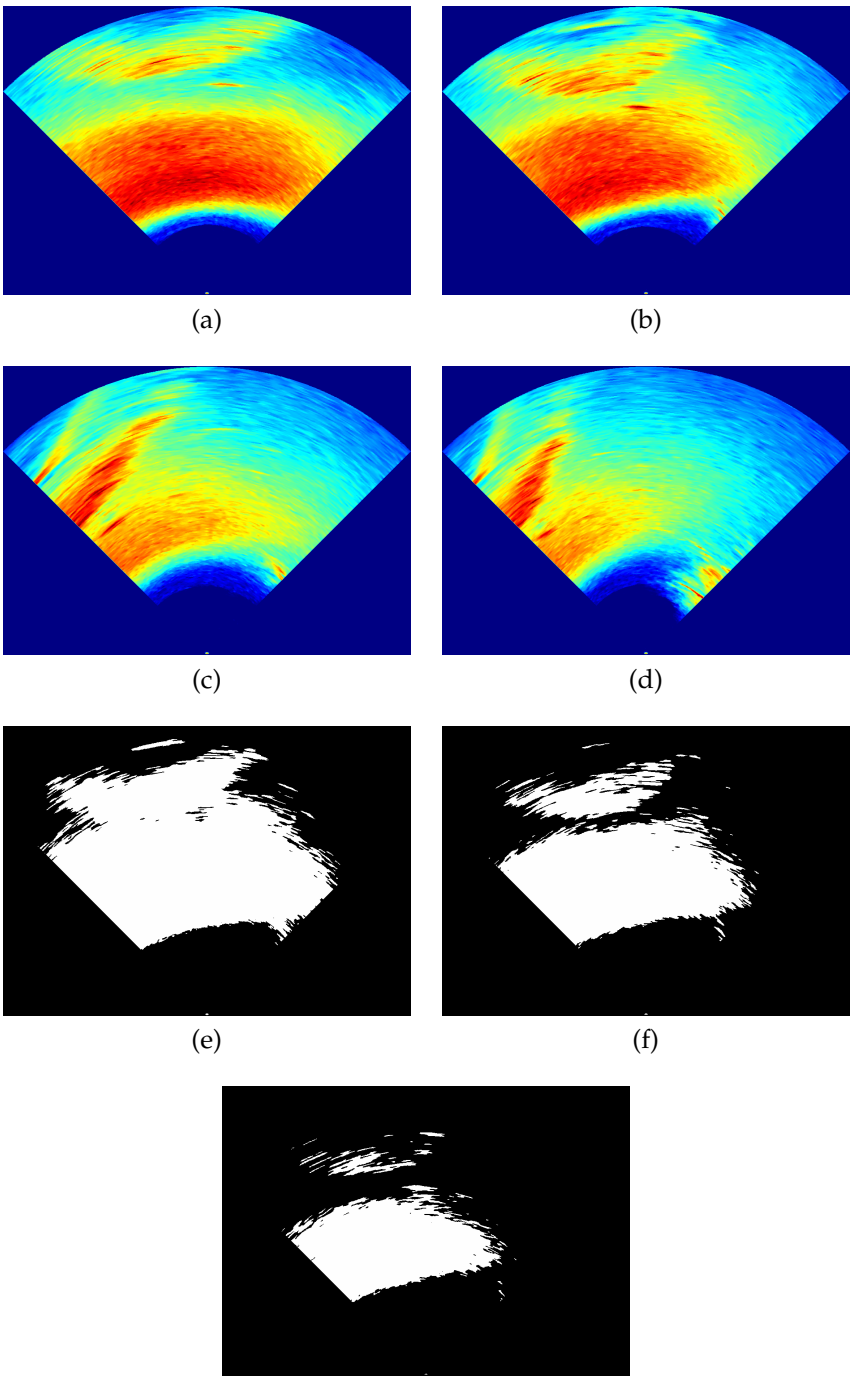
4.2.2 Harbour Images

The second set of images, taken in a harbour, provide a good real-world example of the general type of image that can be expected (the images were provided by IMT and taken at Simon's Town, South Africa).

Fig. 4.8(a)-(d) shows four consecutive images taken with the sonar mounted on a moving AUV. At the top of the first image the reflection from an anchored boat can be seen. The second image shows the same reflection, but closer to the sonar, as the AUV moves forward. In the third and fourth image the boat's reflection moves to the left as the AUV makes a right turn. The images were taken with a maximum range of 50m and with the sonar 11m-14m above the harbour floor. As a result, the reflections from the harbour floor are very prominent when compared to those of the boat. This is further illustrated in fig. 4.8, where images (e)-(g) show thresholded versions of the second image in fig. 4.8.

These images are good examples of the vertical beam effects discussed in Section 4.1.4, as illustrated by the large increase in intensity at the range where reflections start. This crossover is compounded by the effects resulting from the angle of incidence. As a continuation of the discussion about the vertical beam width, fig. 4.9 shows a diagram of the vertical beam being cut by a horizontal plane. Shown on the diagram are three smaller wedges with the same angle width but rotated at different angles about the origin. From this diagram it should be noticed that although all three wedges cover the same width in angle, when they cut the horizontal plane at the bottom the resulting area covered on this plane increases as the angle of rotation move closer to horizontal.

If it is assumed that the energy contained in each of these wedges are more or less equal it means that the energy per area decreases as the angle of incidence becomes smaller. This effect can really be related back to the fact that as the range increases, the energy decreases. If it is further assumed that the background roughness stays constant, it is a fair assumption that the amount of energy actually being reflected back to the



(a) (b) (c) (d) (e) (f) (g)

Figure 4.8: (a)-(d): Four consecutive images taken with the sonar mounted on a moving AUV. (e)–(g) The image in (b) thresholded at 50 %, 60 % and 70 %.

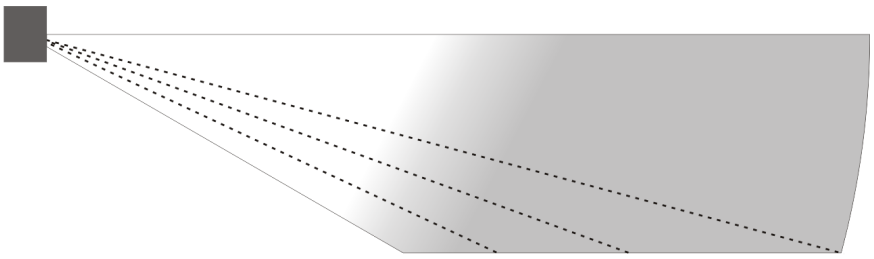


Figure 4.9: Diagram illustrating the effect of the angle of incidence.

source will also become less, as the amount of energy scattered increases as the angle decreases. This is clearly visible in Fig. 4.8(a) where the initial bottom reflections are very strong and then starts to wane rather quickly as the range increases.

As a result, the greatest challenge for further processing is to overcome this initial crossover and successfully suppress the bottom reflections.

Chapter 5

Application of Image Processing

The problems identified in Chapter 4 provide the starting point for examination of further processing techniques. The main objective is to process a single image, without any previous knowledge of the environment, in such a way as to enhance the objects that may lead to collisions, whilst suppressing possible false targets as much as possible.

From the discussion in the previous chapter on the characteristics of sector scan sonar and the images they produce, it should be clear that extracting the correct information from these images is in no way trivial. The lack of resolution and noisy nature of the images makes the use of techniques that are not merely based on intensity values, such as texture based or edge detectors, very difficult. As a result, the possible processing is limited to techniques that can enhance the objects in such a way that make them more discernible in intensity from the background.

The processing is therefore focused on two aspects of the images. The first is noise reduction, and since the flume images are a near worse-case scenario of the influence of noise, the techniques will be evaluated on these images. Secondly, there is the reduction of background clutter. This is a big problem as identified in the harbour images, and one that has not been addressed in the literature.

5.1 Noise Reduction

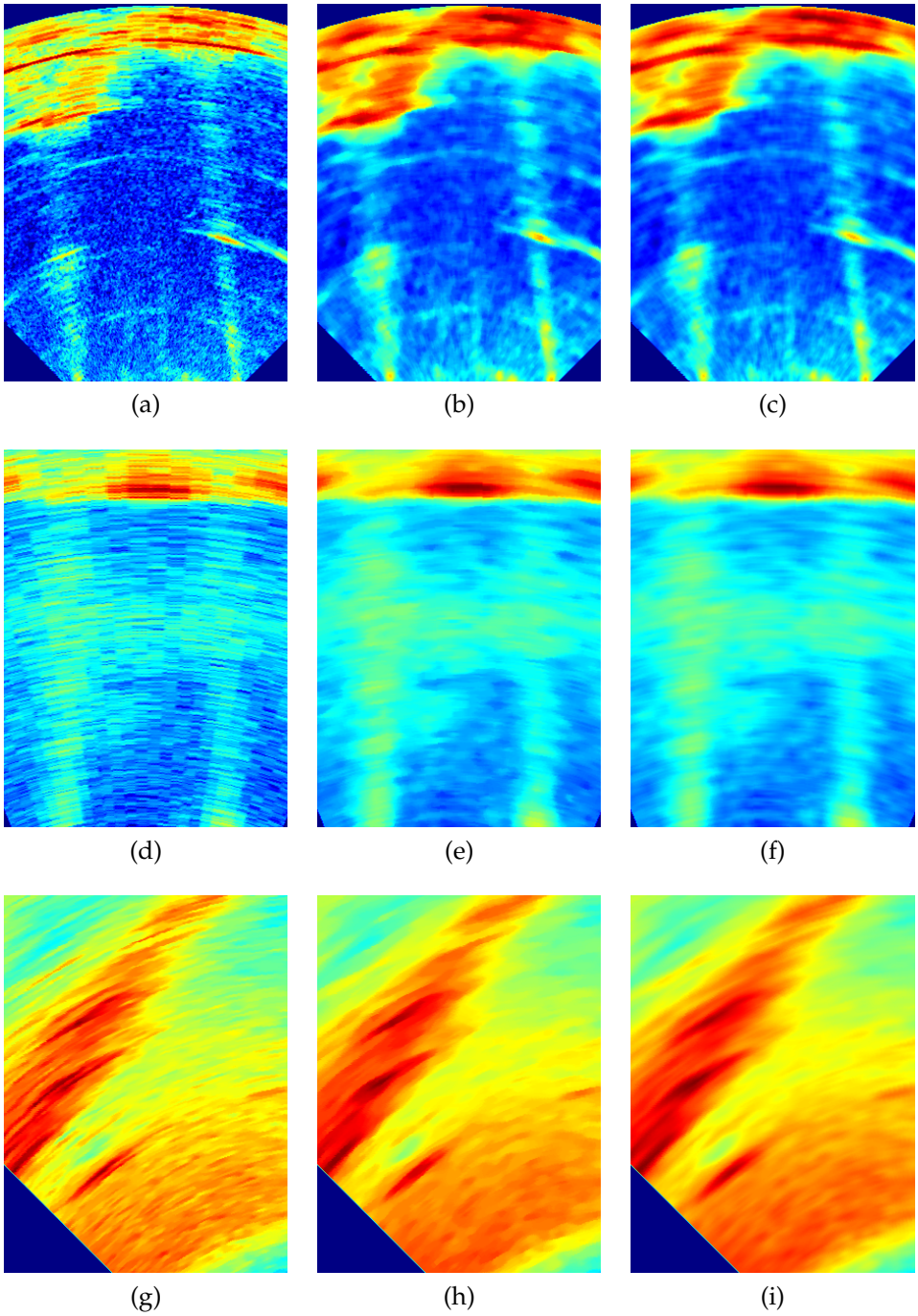
From the analysis in the previous chapter and as stated in the literature, sector scan images are prone to noise. The amount of noise is highly dependent on the environment and can be caused by any discontinuities in the water as well as the surrounding environment. Reducing this noise is the first step in the segmentation process, and can be accomplished by a few different methods.

5.1.1 Spatial Filtering

Commonly, the noise results in a speckle effect, with pixels having either noticeably higher or lower values than its neighbours. This type of noise is encountered in many fields of image processing and the *median filter*, as described in Section 2.2.2.2, provides adequate noise reduction in most cases. It is therefore also the first step in the procedures described by [6], [16] and [17].

As mentioned, the median filter performs well in reducing speckle noise, it does however lose effectiveness where these impulses are not binary in nature, i.e. not the maximum or minimum attainable value. The performance also suffers as soon as the number of noise pixels is greater or equal to half of the number of pixels in the window. This can be seen in fig. 5.1(a) and (b). Although the original image, (a), does contain speckle noise, it is not binary in nature and forms clusters, and as can be seen in (b), the median filter cannot remove the noise cleanly.

Fig. 5.1(c) shows the same image filtered using an *averaging filter*, as introduced in section 2.2.2.1. This filter reduces noise using the same principle as eq. 2.2.2, and increases the signal-to-noise ratio by a factor of N_w, where N_w is the number of pixels in the window [7]. This is however only true when the pixels contained in the window had a constant value before the addition of noise. Since this is rarely the case, the size of window (and as a result the amount of noise reduction) is limited by



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 5.1: Three pairs of denoised subimages. (a)(d)(g): Original Images. (b)(e)(h): Filtered using a 9×9 median filter. (c)(f)(i): Filtered with a 9×9 average filter.

the acceptable amount of distortion, in the form of blurring. This type of filtering can be performed at greatly reduced computational cost to that of an equally sized median filter, and as shown in the image can produce very similar results. A similar result is seen when evaluating fig. 5.1(d) and (g).

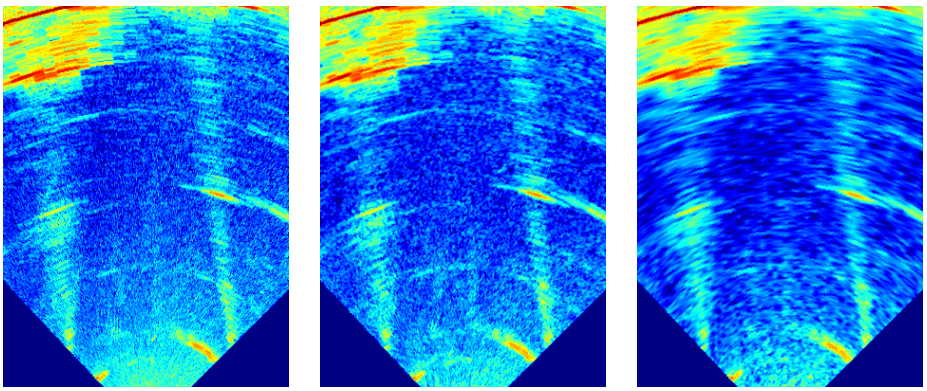
Each of these three sets of images contain different characteristics, as discussed in the previous chapter, and despite this, there is no real benefit from using median filters over the simpler averaging filter. In fact, the averaging filter seems to be better at fusing separated areas with similar intensity values, which can be beneficial to the uniformity of the final object intensities.

5.1.2 Interpolation

As described in section 3.2.4, the transformation from polar to Cartesian space results in a sparsely covered image. It is therefore necessary to either create data to fill the gaps or sample at a smaller step angle. The latter solution is constrained by the capabilities of the hardware and in most cases data needs to be interpolated from the surrounding values.

The method of virtual beams, as proposed by [15] and also covered in section 3.2.4, provides a very geometrically sound solution to the problem. It also has the added benefit of providing some noise rejection, since it uses the weighted average of surrounding pixels to calculate the missing values.

The calculation of the pixel weights from the distances between them increases the computational complexity of this method considerably, since it needs to be calculated for every pixel value to be interpolated. In order to reduce the complexity, the interpolation is performed in polar space. In this case the distances between pixels stay constant and the weights only need to be calculated once. This also results in a fully populated rectangular matrix, unlike the image in Cartesian form, where much of



(a) (b) (c)

Figure 5.2: (a)(b)(c): Images sampled at 0.225°, 0.45° and 0.9° step angles.

the image is populated by zeros denoting the area outside the scan. This reduces the size of the image for further processing and simplifies the implementation of further processing, since the whole image contains scan data.

The interpolation value is calculated from the three sampled values to either side of the empty pixel. The two pixels that are at the same radius are weighted double that of the four pixels that are removed by a single bin. This equation is given by

$$W(p,u) = \frac{k_2[2 \times I(p-k_1,u) + I(p-k_1,u+1) + I(p-k_1,u-1)]}{(k1+k2)\times 4} + \frac{k_1[2 \times I(p+k_2,u) + I(p+k_2,u+1) + I(p+k_2,u-1)]}{(k1+k2)\times 4} \tag{5.1.1}$$

where k_1 and k_2 are the distances in pixels from the pixel to be calculated to the sampled pixels to either side of it. These two values stay constant for the range values belonging to a single angle. Only originally sampled values are used for interpolation, not those previously interpolated.

Fig. 5.2 shows three consecutive images of the same scene taken at increased step angles. Image (b) and (c) are both interpolated to 0.225°

angle steps using eq. 5.1.1. When compared to (a) it is clear that the interpolated images show a reduction in noise. Although these images are taken in a very noisy environment it confirms that there is no real benefit from a too small step angle. In reality oversampling, such as (a), reduces the signal-to-noise ratio and adds no additional information about objects contained in the environment. As such, the amount of noise can be reduced from the beginning by sampling at a larger step angle. This also improves the image update rate which, for the use in navigation, can be very important.

5.1.3 Opening and Closing

Grey-scale *opening* and *closing* are two techniques used to increase the uniformity of local intensity levels by either removing peaks (opening) or removing valleys (closing). These techniques are specifically referred to as *grey-scale* opening and closing, since the derive their names from similar techniques originally used on binary images [21].

Fig. 5.3 shows the two procedures performed in one dimension. When opening is performed, all the peaks with a width of less or equal to the kernel size are removed. In the case of closing, all the valleys of less or equal to the kernel size are filled in. Opening can be used to remove small areas of high intensity, such as corner reflectors or discontinuities

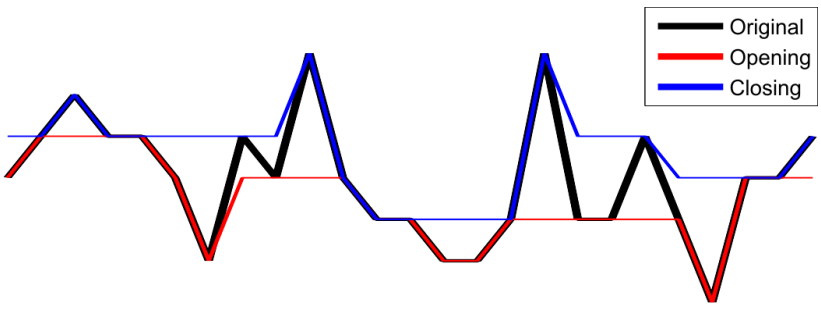
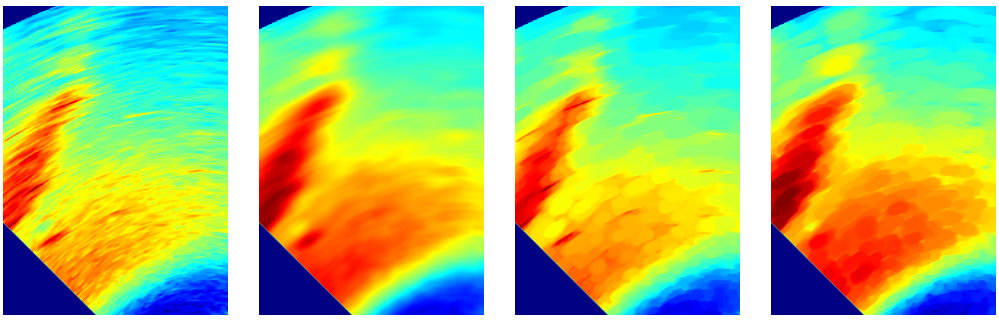


Figure 5.3: The opening and closing procedures performed in one dimension.

on the ocean floor that result in strong reflections. It is also useful for removing speckle noise, like the median filter, but with the advantage that the only criteria for removal is the size of the speckle, regardless of its neighbouring values. As such it can be better tuned to remove specific types of noise. Closing is used to connect areas of similar intensity, albeit, depending on the distance between the peaks, at lower intensity values. This is useful in cases where larger objects do not give a uniform reflection but contain intermittent intensity peaks. In most cases an averaging filter will also be able to connect these peaks, but the closing procedure does not change the original peak values and does not result in the blurring of edges.

Fig. 5.4 shows a comparison of equally sized averaging, closing and opening kernels performed on the harbour images and both operations perform as expected. Comparing (a) and (c), the peaks visible in (a) are still visible and unchanged in (c), whilst the gaps have been filled, resulting in areas of more uniform intensity. By comparison all the peaks visible in (a) have been removed in (d), which also results in uniform areas. It should however be noted that there is a clear difference in contrast as a result of each technique. Since closing does not change the peak values, the peak contrast between the object on the left and the background remains unchanged. The resulting contrast in (b) and (d) are very similar,



(a) (b) (c) (d)

Figure 5.4: (a): Original image. (b): Image smoothed with a 15×15 average filter. (c)(d): Image closed and opened with a 15×15 kernel.

but less than the original and (c). The difference between the object and background affects the thresholding stage of segmentation and needs to be maximised.

Fig. 5.4 shows an example of where opening and closing can be used effectively at the cost of some detail. The two techniques do have some limitations, as shown in Fig. 5.5. Both (a) and (d) contain a high level of noise, and as seen in (b) and (e), when closing is applied the kernel follows the noise and the background level is increased considerably. In (c), opening does well by removing much of the noise, however, since the reflections from the walls are very faint and somewhat intermittent, these are also reduced and almost removed. A similar effect can be seen in (f), although here the walls hold up better and it does show a reduction

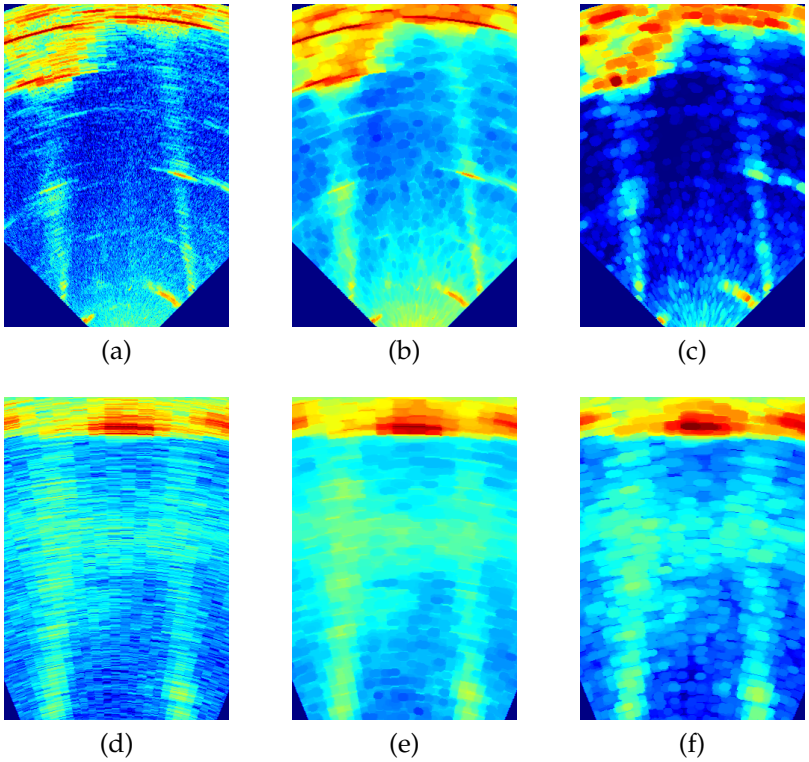


Figure 5.5: Two sets of images containing the original, closed and opened versions.

in the background level.

5.2 Background Suppression

It is true for any sonar system that as the range increases, the received energy decreases, and since the sonar receives an echo from an object this attenuation is inversely proportional to the square of the distance to the object. As such, the further away a target is, the weaker the return from it. Commonly, the sonar receiver does compensate for this attenuation by increasing the gain proportional to the distance. This problem is however further exacerbated with the use of sonar in a shallow water environment, as illustrated in Fig. 4.9. The bottom tends to provide a strong reflection when compared to man-made obstacles that commonly have sheer surfaces and therefore only reflect well when illuminated head-on. This is illustrated in Fig. 5.6 where it is clear that the strongest reflections result from the harbour floor in the middle of the image. As the boat comes into view at the top of the image it is clearly visible when compared to the intensity of its surroundings. It is however not strong enough to significantly dampen the relative intensity of the bottom re-

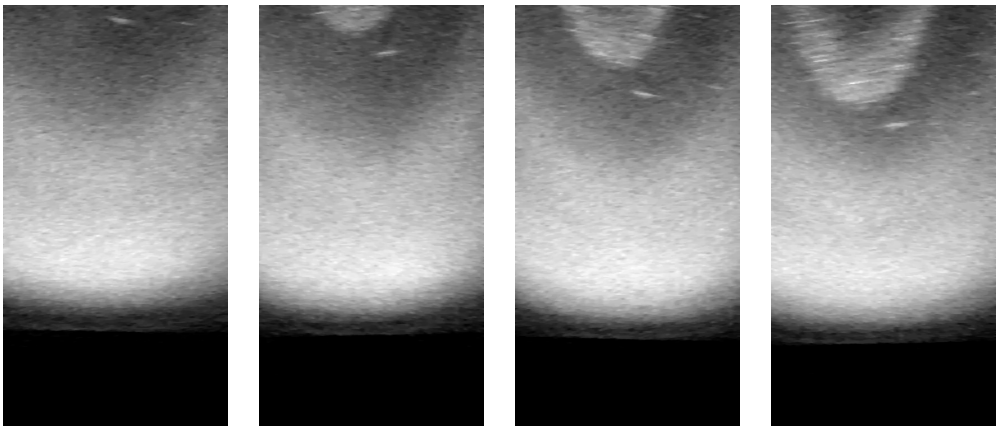


Figure 5.6: Four consecutive images showing the reflection from a docked boat as it comes into view.

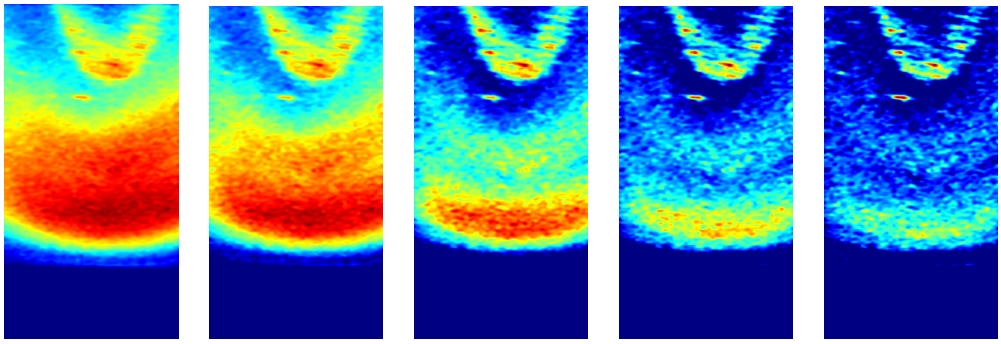
flections.

These background reflections provide the greatest challenge to the segmentation process. The non-uniformity of the background and the fact that it has comparable intensity to that of actual objects, makes it impossible to remove with a simple global threshold. It is therefore necessary to find ways of suppressing the background and as a result enhance the objects contained in the image.

5.2.1 High-pass Filtering

Non-uniform illumination is a problem commonly found in image processing, resulting from environmental factors or the imaging process [21]. It is characterised by variations in the background intensity, such as the shadow caused by an unwanted light source in a photograph. It is a fair assumption that these variations in background are gradual in nature and will therefore be confined to the lower frequencies. It should be possible to gain a reduction of the background by simply high-pass filtering the image.

Fig. 5.7 shows the effect of high-pass filtering with an increasing suppression of the lower frequencies. From the images it is clear that a reduction in the lower frequency components does tend to remove an increasing amount of the background reflections. Unfortunately, since the objects found in the harbour environment can be of relatively large size, such as in fig. 5.7, much of its information is also contained in the lower frequencies. It is therefore not possible to increase the size of the high-pass filter indefinitely, since the object also disappears. This effect starts to show in (d) and (e), where the object starts losing its underlying high mean value and is broken into the peaks that contain the higher intensities. Image (c) is a good example of both the strength and weakness of this technique. From the image it is apparent that the object is much more pronounced in the local sense and that there was a clear reduction of the overall background. Unfortunately, the initial crossover from no



(a) (b) (c) (d) (e)

Figure 5.7: (a): Original image of object coming into view at the far end of the image. (b)-(e): Original image high-pass filtered with a Gaussian filter with a standard deviation of 1,2,3 and 4 pixels respectively.

illumination to full illumination of the sea floor, provides such a sharp increase in intensity, that the high-pass filter is unable to remove it. By the same token, it is unable to produce sufficient range equalisation by suppressing the background and by enhancing the objects at the far end of the range.

Using high-pass filtering can provide good results under very specific circumstances, but does not provide a final answer with the characteristics of the evaluated data. It could however provide satisfactory results, when used in conjunction with other techniques.

It should be mentioned that the Fourier transform is calculated using the *FFT* (Fast Fourier Transform) algorithm. This is standard practice and is an algorithm implemented as a standard function in most signal processing hardware and software packages. Also, because of the periodic nature of the FFT, the transform tries to match the left and right edges and the top and bottom edges of the frequency representation of the image. As a result, the filtered image shows undesirable dark or light areas at the edges after the inverse transformation [21]. This can be avoided by padding the edges of the image and creating a buffer area where the edge effects can occur without influencing the image data. Padding the image

with values mirrored about the edges results in very little variation in the frequency profile, in comparison with, for example, padding with zeros, and is therefore recommended.

5.2.2 Minimum Value Background Estimation

Another way of suppressing or completely removing the background from an image is to calculate some estimate of the background from the data contained in the image. The type of method used greatly depends on the characteristics of the objects contained in the image, such as size and expected intensity. Furthermore, the less uniform the background becomes the more difficult it is to achieve a good estimation. It is therefore necessary to find methods that can cope with variability in local background values, without being influenced by the objects too much. Because of the variability of the observed objects in both size and relative intensity in the images currently under consideration, and since the process needs to be completely automated and repeatable, it leads to very general assumptions about the data and the way a background estimate can be gained.

The *minimum value estimation* method divides the image into equally sized windows and then finds the minimum value from each of these. This results in a smaller image of the local minima contained in the original. This process can be thought of as a subsampling procedure, but instead of the returned values being representative of the values contained in each window, it is merely the lowest value in each.

The reasoning behind this method is that the background will always have a lower intensity than the objects contained in it, and the local minima should therefore provide a good estimate of the background. Also, as long as the window size does not become too small, it should sufficiently reject the objects contained in the image. When the subsampling procedure is completed, this data is used as the seed values for linear interpolation. This results in an image with the same size as the origi-

nal image, and with the same contour as the minimum sampled image. This background estimate can now be used to remove the background by subtracting it from the original.

The biggest problem faced by any estimation technique with the current set of data, is giving a sufficiently close estimation of the crossover from the area of no background reflections to where it does begin. As seen in the previous section, high-pass filtering cannot sufficiently reduce this crossover without completely removing the objects as well. By inspection it is apparent that any window covering this crossover will have a very low minimum value, which will not provide a good estimate of the real background level. The smaller the window the better the estimate should follow the actual background and should provide better rejection of sudden local minima. On the other hand, areas of higher intensity, such as those representing large objects, will also have more influence on the final estimate. It therefore offers better protection against variability in the background but at the cost of greater influence from objects. In the same way, the larger the window, the better it rejects higher intensities from objects, but cannot cope with great variability in the background.

Fig. 5.8 shows the results gained from the minimum value estimate and provides a good example of how the choice of window size influences the resulting image. In (b), with a window size of 10×10, the background is removed very well, but also results in the object being removed. In (c) and (d), with window sizes of 25×25 and 40×40, it is clear that with an increase in window size the object is highlighted much better, but the background is removed to a lesser extent, with areas of higher intensity forming that was not present in (b).

The estimates used for these three images were from the unprocessed local minima, and gives a good representation of the procedure in its most basic form. (f) - (h) does however provide an example of how the robustness of the technique can be improved by median filtering the local minima before interpolation. The difference is immediately evident

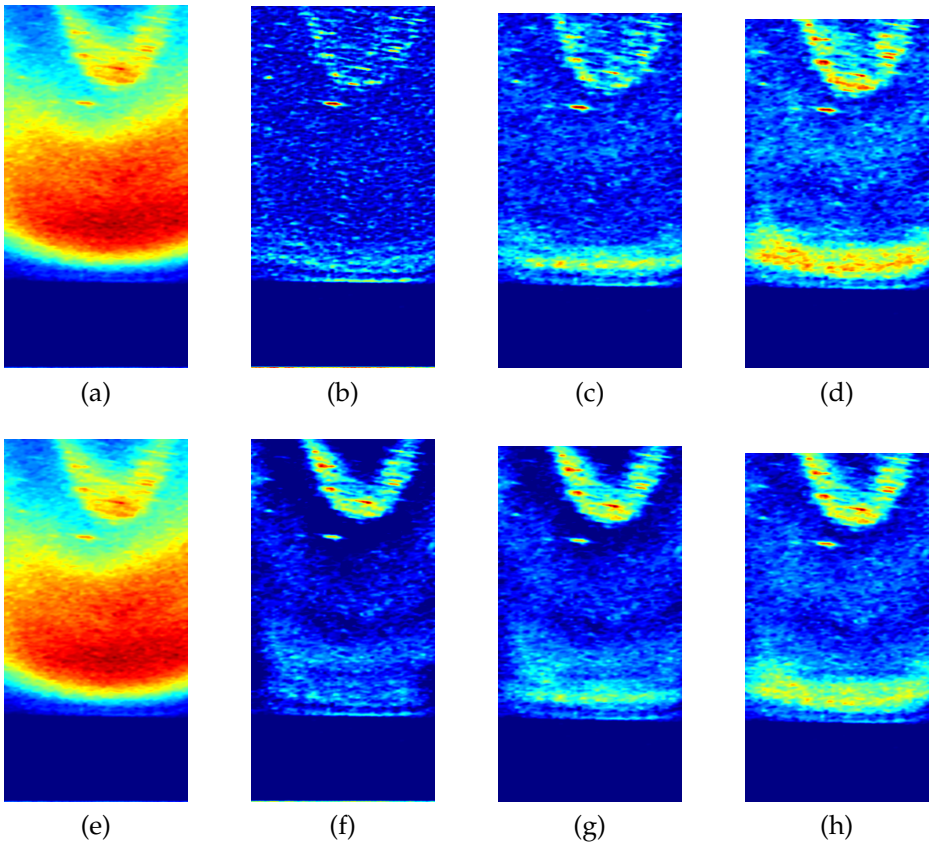


Figure 5.8: (a)(e): Original Image. (c)-(d): Result of subtracting the minimum value estimate from the original image. Estimates taken with window sizes of 10, 25 and 40 pixels respectively. (f)-(h): Result of estimates taken with exactly the same window sizes as (c)-(d), but median filtered before interpolation.

when examining (b) and (f). In (f) the objects are completely visible and with very good contrast. The background is still removed very well, and the median filtering results in better rejection in some areas, as well as a loss in others. In both (g) and (h) the object is identified noticeably better but also at the cost of background rejection. Using a smaller window does result in a larger subsampled image and as such, different size median filters were used for each of the estimates. For (f) a 19×19, for (g) 7×7 and for (h) a 3×3 was used.

From this example it does seem that the minimum value technique

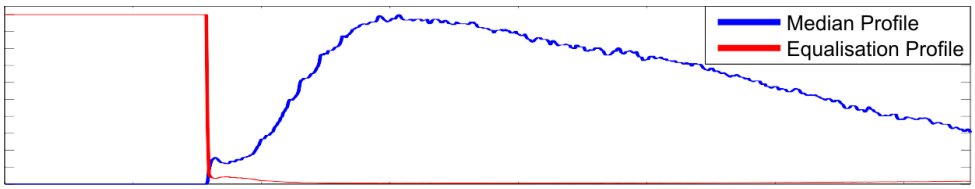
provides a very good estimate of the background and also good enhancement of the objects present in the image. This initial evaluation also shows that the better background rejection of a smaller window size is more desirable than the better object enhancement of a larger window, especially since this shortcoming can easily be overcome with adequate median filtering. This initial result will however be further examined in a later section.

5.2.3 Range Equalisation

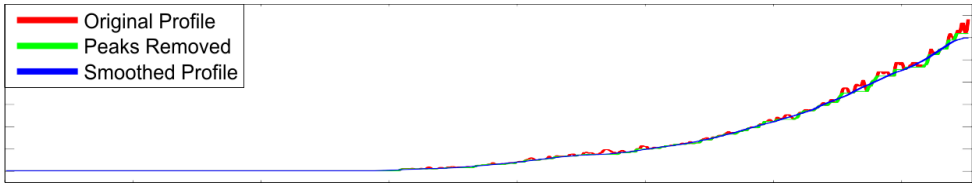
As an alternative to removing the background, a way was sought to equalise the intensity variation found with range. In other words, a function needed to be found to transform the natural illumination profile to a constant value across the image. If successful, the resulting image will have a near constant background and the non-uniformity of the object's intensity due to range dependency should be greatly reduced.

The first step would therefore be to find some estimate of the illumination profile. If the assumption is made that the reflective properties of the background stay more or less constant, the variation in the background reflections can be used to estimate the illumination profile. As such, the estimate needs to provide robustness to the variations found in the background. Furthermore, although some success has been achieved at a three-dimensional estimate of the background, as in the previous section, it was decided to keep the profile estimate a two-dimensional function. This is done to reduce the computational complexity as well as to keep the estimate as robust as possible. From these assumptions, the median value, or statistical centre, of each range row of pixels was chosen to be representative of the illumination at that range. This set of median values then form the estimate of the illumination profile for the entire image.

The illumination profile can now be used to calculate the equalisation function needed. Since the objective is to convert the illumination profile



(a)



(b)

Figure 5.9: (a): The median value profile and the resultant equalisation profile. (b): Estimation profile after being clipped, and the results of peak removal and smoothing.

to a constant value, an equalisation profile is created by calculating the ratio between a constant value and every value in the illumination profile. The equalisation profile then represents the scaling factor needed at each range, and each angle column of the original image is merely multiplied by the profile to complete the equalisation.

Fig. 5.9(a) gives an example of the median profile of an image and the resulting equalisation profile. As with all the techniques thus far, the crossover into background reflections is again a hampering factor. In this case it results in the large initial constant value found in the equalisation profile. Since the values of the equalisation profile is the result of division, this large constant is the consequence of the very small median values found in the initial part of the image. To compensate for this, the median values of this initial region as well as those contained in the crossover region are given the same value as the peak value of the median profile. Beyond just solving the problem, this can be further justified since any object in the main beam of the object at a closer range than that of the background should at the very least receive the same amount of illumination as the peak background reflection.

Alternatively, this means that the initial part of the equalisation profile is floored at its minimum value, as shown in fig. 5.9(b). To keep the equalisation profile as smooth as possible, the peaks are removed with a closing operation and then further smoothed by an averaging window. This is done to remove any undue amplification in the equalised image, caused by sudden and large variations in the original image.

Fig. 5.10 shows the results from implementing the range equalisation procedure. The second column shows the results of applying the equalisation profile. When compared to the original images it is clear that the general illumination pattern is much more uniform. As a result the object data at the far end of the image that was originally locally discernible, has now been scaled to be easily discernible with global thresholding in mind. Since this was the type of result that was sought through range equalisation, it clearly confirms the validity of using the median value to estimate the illumination profile.

The second column shows the results of using the median profile as a background estimate. From the images it is clear that the median profile provides an excellent estimation of the background and does noticeably better at removing the background than high-pass filtering and the minimum value estimation. This is a somewhat unexpected result from the initial premise, but provides further means to improve the dynamic range contained in the image.

In order to further highlight the benefits of range equalisation, the fourth column shows the results of subtracting the median profile from the equalised images. In comparison to the images in the third column the objects contained in the new images have a more uniform intensity with an increase in range, and as expected the detail contained in the farthest ranges of the image is now also clearly visible.

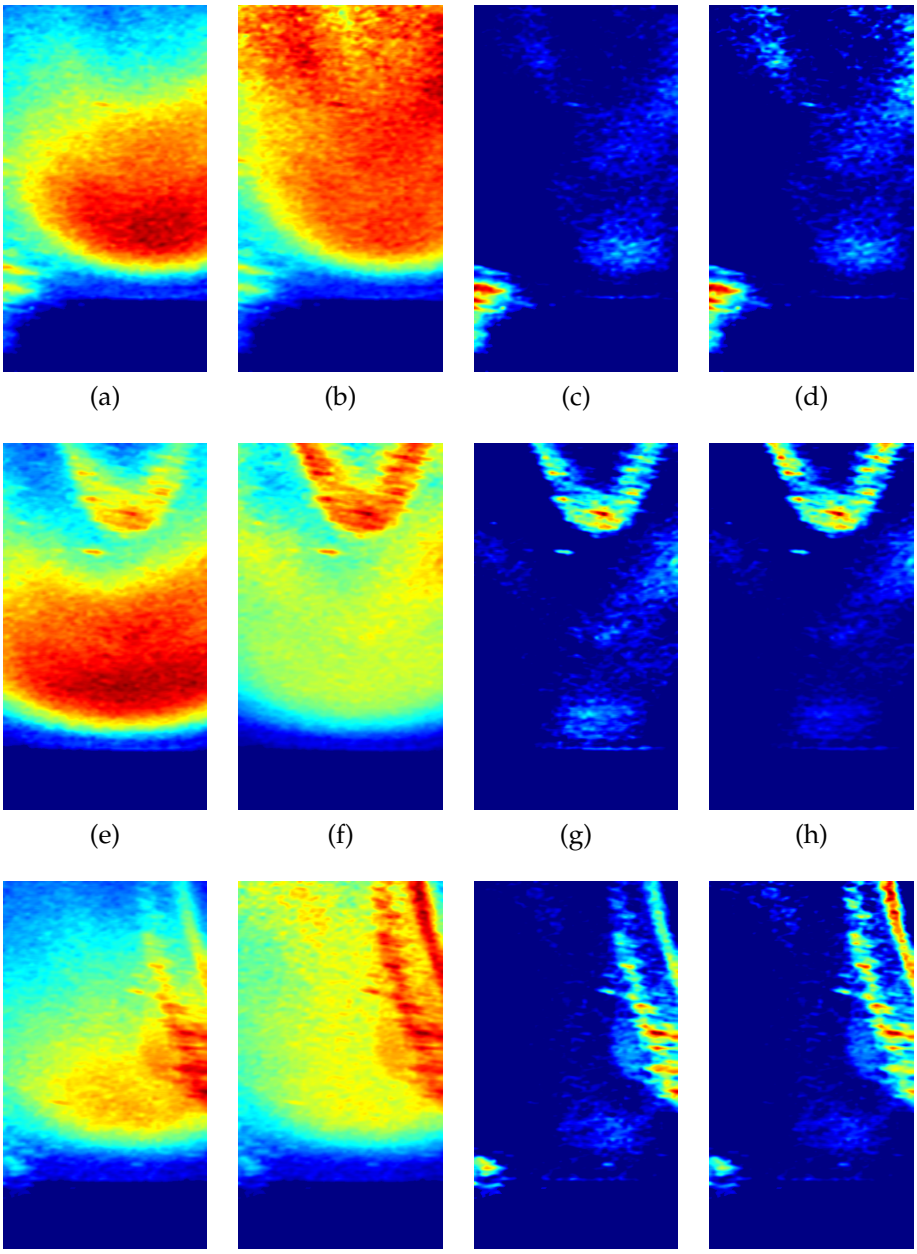


Figure 5.10: Three examples of the result of range equalision. (a)(e)(i): Original Images. (b)(f)(j): Results of range equalisation. (c)(g)(k): Results of subtracting the median value profile. (d)(h)(l): Result of subtracting new median profile from equalised images.

Chapter 6

Comparison of Techniques

6.1 Comparison of Techniques

The evaluation of the techniques thus far have relied mostly on how they influence the images visually. Although this can be somewhat subjective and open to interpretation, it has been sufficient to identify and examine the initial results achieved by each of the techniques. To further frame the problem and compare the effects of each of the techniques, it is however necessary to analyse the data contained in the images.

To successfully set a threshold between the background and object pixels, there has to be a clear enough separation in intensity between these two categories. The simplest way of evaluating this, is by looking at the image histogram, which shows the number of pixels at each intensity value and is an empirical probability density function of the image. The two intensity categories should be visible in the histogram as two discernible modes or groupings of intensity values.

To further minimise the chances of a wrong classification the dynamic range of the image also needs to be evaluated. The dynamic range gives the distance between the two categories of pixels. Although there might be a separation between the modes, if the room for error is too small the

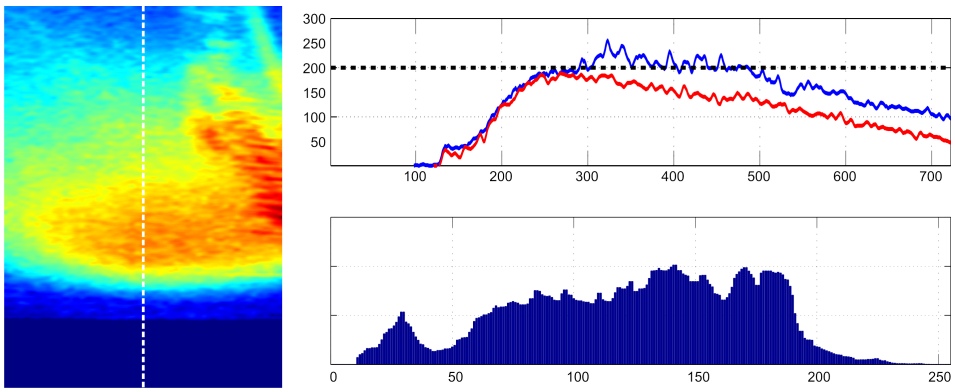


Figure 6.1: Unprocessed image, with its maximum (blue) and background (red) curves and its histogram.

chances of incorrect classification will obviously increase. The dynamic range will also be visible in the histogram, but to a lesser extent. To further illustrate this, the difference between the maximum values and those representative of the background will also be evaluated.

Fig. 6.1 shows an unprocessed image with its histogram and a graph showing the curve of the maximum value at each range and the value of the pixels running up the centre of the image, which is representative of the background. For the purpose of the histogram and the dynamic range graph, the images were converted to 8-bit values, i.e. an integer value between 0 and 255. Therefore, the x-axis of the histogram and the y-axis of the contours correspond to the same range of values. It is therefore possible to quickly compare the range of possible threshold values in both.

From the histogram of the original image, it is clear that there are no apparent modes that would indicate a separation between the background and target. This is partly because the object occupies fewer pixels than the background. If the histogram is analysed to rather find the range of values contained in the background, instead of trying to find two distinct peaks, one could roughly estimate that all the pixels with a value of 200 or less belongs to the background (as indicated by the dotted line). This value is on the final slope and should therefore indicate the end of a

mode.

Using this value as a threshold completely removes the background, but also removes most of the object, as illustrated on the maximum and background contours. Furthermore, it is clear from these graphs that although there is a continual separation between the background and object, no single threshold value will successfully remove the background.

To further compare the techniques, the average dynamic range contained in each resulting image is calculated. This is done by taking the difference between the maximum and background value at each range and calculating the average over all ranges. To make the test as repeatable as possible, the mean dynamic range is calculated in the range of pixels where there is a clear difference between the object and the background. The calculation is started from the pixel number corresponding to the maximum value in the original image, and continued to the final pixel. Using this method, the average dynamic range for the original image is calculated to be 52.

The image in Fig. 6.2 is the result of high-pass filtering the original image with a Gaussian kernel with a standard deviation of 1.5 pixels. As discussed previously, the high-pass filter is not able to remove the crossover region and a single threshold is therefore not possible either.

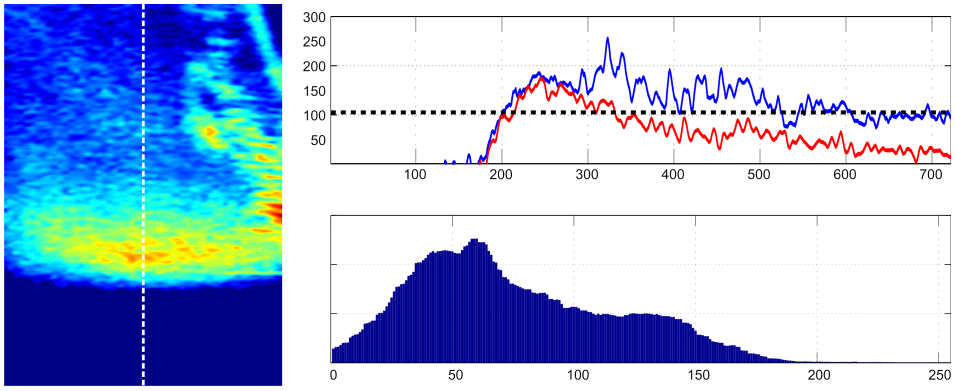


Figure 6.2: The result of high-pass filtering , with its maximum and background contours and its histogram.

It does however start to group the background pixels, as a clear peak is formed in the histogram at around 50. Inspection of the maximum and background curves show that a threshold of around 100 would remove most of the background, excluding the crossover peak. When referring this back to the histogram, this value is situated on the slope between the background and what can be seen as a local maximum for the object. From both the image and maximum and background graphs, it can be seen that except for a few large peaks the resultant dynamic range, although improved, is still not really satisfactory. The average dynamic range is calculated as 76, which is close to the difference between the peaks of the background and object intensity on the histogram. This is more or less expected since the data contained in the lower frequencies provide a rough estimate of the background.

The image in Fig. 6.3 shows the result of the minimum value estimation. It seems that this technique provides a good estimate of the background, as at the far end of the image the objects are much more pronounced. This is further evident from the histogram, where the background values form a clear peak with a greatly reduced width. Following the initial downward slope from this peak an estimated threshold can be set at roughly 80, which, when inspecting the maximum and background graphs, will be able to sufficiently remove the background. The average

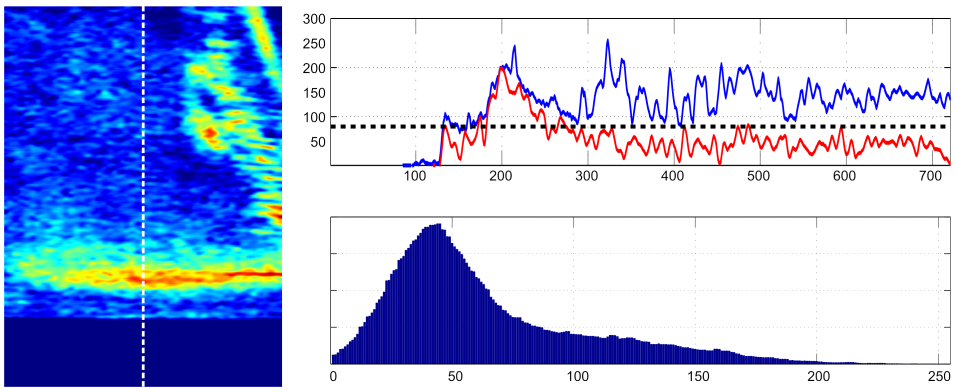


Figure 6.3: Result of the minimum values estimate techniques, with its maximum and background contours and its histogram.

dynamic range is calculated at 111, which as can be expected from both the image and the graphs, is an improvement over that of the original.

The image presented here is the result of using a 40×40 window for the estimate, and as such the estimate cannot completely remove the crossover. Although using a smaller window size does remove the crossover, the global result is not satisfactory. When evaluating the dynamic range, it is clear that the estimate from a smaller window is influenced too much by the large objects contained in these images. The image shown here is the most promising result achieved with this technique, and the presence of crossover effects will be addressed in a following section.

Subtracting the median estimate, as in Fig. 6.4, almost completely removes the background, and confirms the previous results with this technique. The histogram shows a sudden peak and a downward slope, which is expected since the lower-most values have been completely removed, again confirming that the median provides a very good estimate of the background. In this case a threshold anywhere between 50 and 100 will provide adequate separation between the background and objects, as illustrated on the maximum and background graphs. As expected the average dynamic range shows an improvement of three times that of the

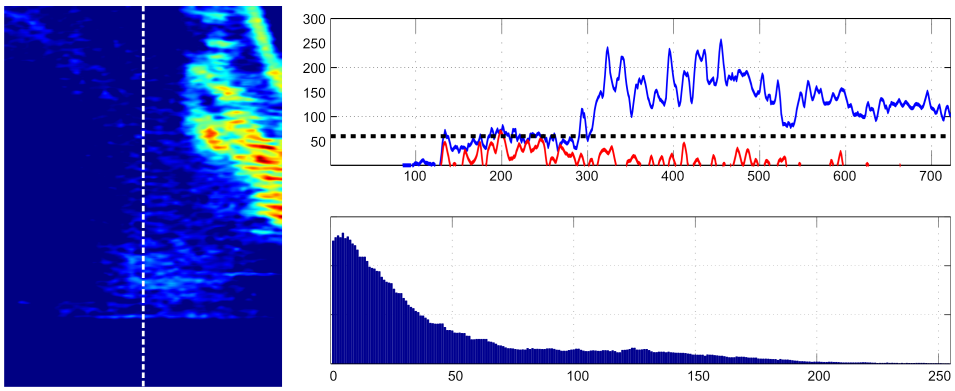


Figure 6.4: Image after subtracting the median value estimate, with its maximum and background contours and its histogram.

original, at 156.

The range equalisation technique, as in Fig. 6.5, again performs as expected. From the image and the graphs it is clear that the technique provides a great improvement in the standard deviation of both the background and object intensities, and comes the closest to grouping the values of the objects. As explained in the previous section, this technique does not dramatically change the dynamic range of the image. This is clear from the maximum and background graphs, where the separation is similar to that of the original image, but they now run near horizontally instead of at a slope. This is also confirmed by the average dynamic range which is calculated at 67. The object can be separated easily, and by following the slope in the histogram from the background peak, a threshold value at roughly 170 will provide adequate separation.

It should however be noted that the intensities at the very far end of the image are enhanced to a greater extent. As a result, the peaks in the background reflections at the far end are also enhanced too greatly and as can be seen in the top left corner of the image, start to have intensities comparable to that of the objects. Although the effect is somewhat limited in this example, in other images where the background reflections are not as uniform, large areas of high intensity appear at the far end of

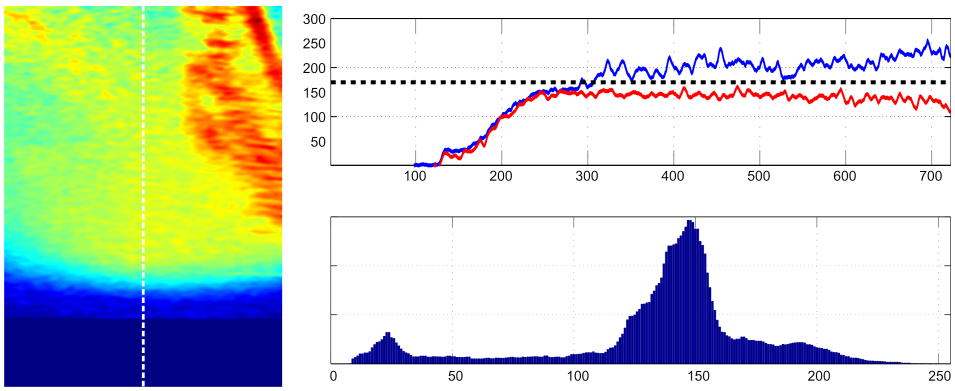


Figure 6.5: The result of range equalisation, with its maximum and background contours and its histogram.

the image after this technique is applied.

This comparison confirms the initial evaluation of each of the techniques, and provide further insight into the influence of each. The high-pass filtering does provide an improvement, but is not really suited to the characteristics of this data set. The minimum-value estimation provides a good estimate of the background and improvement in dynamic range, when the background is near uniform, but where the variation is too great it cannot cope properly. As expected the two median-value based techniques seem to perform the best. Range equalisation does however provide too much enhancement at the far end of the image and depending on the survey range being used, can corrupt the resulting images somewhat. The median estimate removes the background very well and is able to discern the objects from the background in most cases. It also provides the greatest increase in contrast, and as such is the most robust of the proposed techniques.

6.2 Assigning a Measure to Performance

The previous section provides a thorough discussion on the properties of each of the techniques and by introducing the histogram and the maximum and background curves, provides further means to sufficiently compare them visually. It is however necessary to find a more quantitative measure of the performance of each of the techniques, to provide a way for automated comparison.

The *mean dynamic range*, ΔD, provides a good starting point for this global measure. Although it gives insight into the separation between the background and objects it does not give information about the uniformity of the output. Uniformity in the object intensity is an important characteristic, since it reduces the chance of an incorrect classification. This is illustrated in the examples given in the previous section and summarised in table 6.1. When comparing the result of high-pass filter-

Technique	ΔD	ΔM	$\mathcal{P}$	Min. Diff.
Original	52	74	7	-89
HP Filter	74	53	14	-52
Min. Val. Est.	102	62	16	-1
Median Est.	142	68	21	34
Range Eq.	67	34	20	13

Table 6.1: Performance measures of each technique with a scaling factor of ten.

ing and range equalisation, the mean dynamic range gives very similar numbers for both, 74 and 67. However on inspection of the maximum and background curves it is clear that the high-pass result varies considerably with range and does not provide a very good global result, in contrast to the range equalisation which provides good uniformity.

To find this uniformity measure the information contained in the maximum graph was chosen, since it is an estimate of the behaviour of the object reflection. As a starting point the standard deviation was examined as a possible candidate, since it should provide a measure of the variation in intensity values. Unfortunately, this also gives an average measure and the resulting value is not representative of the local behaviour on which judgement is commonly based. It was therefore necessary to move away from a measure influenced by average behaviour and rather focus on a peak value. The only real worry for incorrect classification is the amount of variation below the mean, since maximum peak values will undoubtedly be correctly classified. As such the maximum negative deviation from the mean, ΔM, was chosen as a representative measure of the variability of the contour.

Since you want ΔD as large as possible and ΔM as small as possible, the global performance measure, $\mathcal{P}$, is given by the ratio of the mean dynamic range and the maximum negative deviation or

$$\mathcal{P} = \frac{\Delta D}{\Delta M} \tag{6.2.1}$$

The values for ΔM and $\mathcal{P}$ are also given in table 6.1. These values provide confirmation of the discussion in the previous section, where the best performance is achieved by the median estimate technique, followed by range equalisation. The final column gives the value of the difference between the minimum value contained in the object and the maximum value of the background. It provides a further measure of the achieved global separation, and is added merely to provide further confirmation of the results.

Table 6.2 shows the performance values of the four techniques for a set of ten images with similar object features (see Appendix A). As expected, the median estimate and range equalisation techniques perform, on average, the best.

To draw a better comparison between the results of each test, the performance values given in the table is proportional to the performance value of the original image in each case. The performance values in table 6.2 is therefore given by

$$\mathcal{P}_E = \frac{\Delta D}{\Delta M} \times \frac{10}{\mathcal{P}_o} \tag{6.2.2}$$

	High Pass	Min. Val.	Median Est.	Range Eq.
T1	20	24	33	28
T2	19	20	41	26
T3	20	30	30	25
T4	18	24	26	28
T5	20	25	30	27
T6	21	24	29	41
T7	19	30	32	37
T8	20	29	32	37
T9	20	23	30	28
T10	20	22	30	32
$\mathcal{P}_E$ (Avg)	20	25	31	31

Table 6.2: Performance values, $\mathcal{P}_E$, for a set of ten images.

where $\mathcal{P}_o$ is the performance value of the original image. The scaling factor 10 is applied to improve the separation between the values.

The performance measure provides a means to evaluate the effect of any processing on a given image. As such, it can also be used to measure the performance of a set of techniques, including noise reduction, in order to identify the optimal processing.

6.3 Best Result Processing

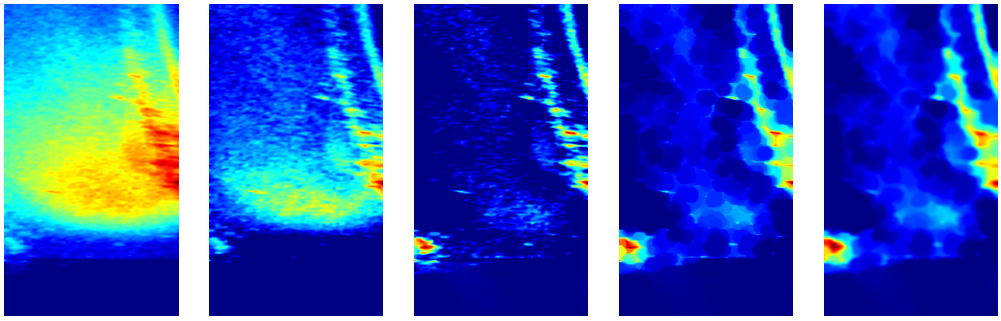
The chapter up to this point has identified and discussed a number of techniques that have the qualities to improve the likelihood of correctly identifying objects. The strength of identifying a number of techniques is that they might be used together and in varied circumstances, rather than trying to find a single answer. It is therefore necessary, as a final discussion, to show how these techniques can be used in sequence to complement each other and overcome their individual weaknesses.

As discussed previously, the first step in any segmentation process is denoising. In the case of the harbour images, which are under consideration here, the original step angle of 0.9° is interpolated to 0.225° followed by a 5×5 mean filter. This provides adequate noise reduction for this type of image and is applied at the start of every sequence. Since the denoising stays constant, the four background removal or equalisation techniques are used as the starting point for each sequence. By adding further processing and using the performance measure of the previous section, the improvement provided by these following stages can easily be evaluated.

High-Pass Filter From the discussion of Section 5.2.1 it was clear that high-pass filtering is best used where the background tends to be more uniform and the objects small. As such, it has been the worst performing of the suggested techniques. It does however provide a measure of range

	High Pass	+ Median Est	+ Closing	+ Mean Filt
$\mathcal{P}_E$ (Avg)	20	24	38	37

Table 6.3: Performance values for the high-pass filtering sequences.



(a) (b) (c) (d) (e)

Figure 6.6: High-Pass Filter. (a): Original image. (b): Result of high-pass filtering. (c): (b) after applying median-estimate technique. (d): (c) after closing. (e): (d) after mean filtering.

equalisation, and may still be of use under the correct circumstances. It is also unable to remove the crossover region, which has really hampered its usefulness. To overcome this problem, the first step in the sequence is to remove the crossover and remnant background by applying the median estimate technique. As seen in table 6.3 this already provides some improvement, but cannot overcome the lack of uniformity produced by the original high-pass filter.

At this point all the low frequency information as well as most of the image that could be considered background have been removed. The resulting image consists almost exclusively of the intensity peaks contained in the original image. These peaks belong mainly to the reflections from the targets and as such need to be connected in some manner to form an uniform object. The closing operation is ideally suited for this type of problem and application using a kernel with a radius of 15 pixels re-

sults in a considerable improvement. As a result, the application of these two further techniques extends the usefulness of high pass filtering by a considerable amount.

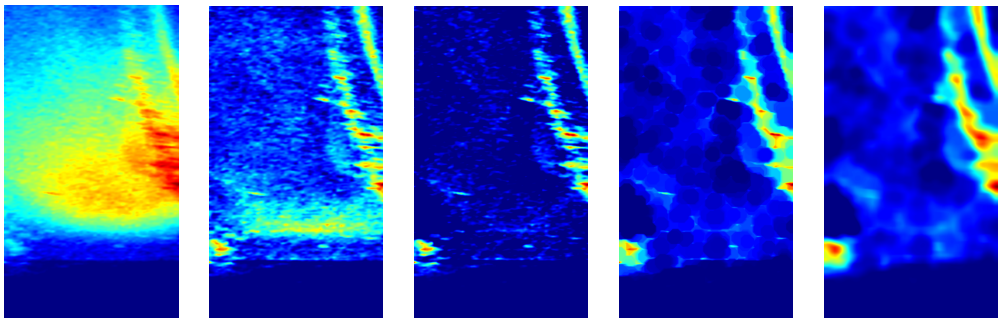
To further try and improve the uniformity of the result, 10×10 mean filter is also applied. This further removes some of the peaks and can also average out the remnant background pixels. It does however also suppress smaller object or areas of lower intensity, and lowers the overall performance somewhat. It is added since it might be a logical further step, but in this case it does not provide a general benefit. Changing the order of the techniques also provide no benefit and are best kept in the order given.

	Min Val Est	+ Median Est	+ Closing	+ Mean Filt.
$\mathcal{P}_E$ (Avg)	25	25	47	48

Table 6.4: Performance values for the minimum value estimate sequences.

Minimum Value Estimate The minimum value estimate technique provides a very good estimation and removal of the background for much of the image. It is however unable to remove the crossover and as with high pass filtering, it is followed by the median estimate technique. As shown in table 6.4, this does not provide any actual benefit beyond removing the crossover. The minimum value estimate also provides very adequate range equalisation. Since it is influenced by the larger areas of higher intensity contained in the objects, it leaves only the peaks of these areas, and as a consequence the dynamic range contained in the resulting image stays near constant.

This inherent uniformity can therefore be further improved by applying a closing kernel to connect the gaps between the peak values. The



(a) (b) (c) (d) (e)

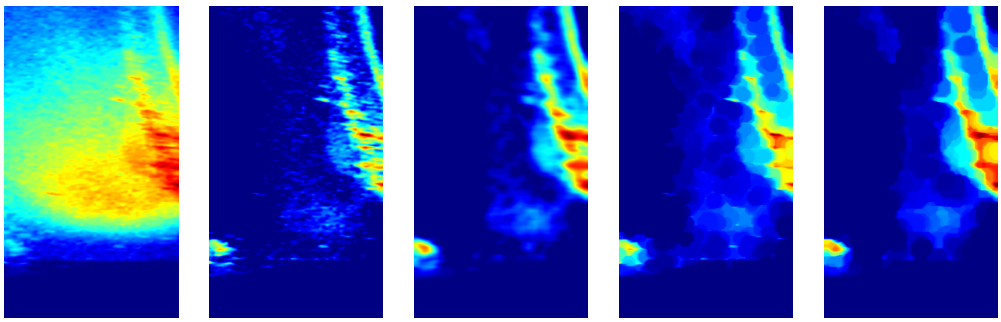
Figure 6.7: Minimum Value Estimate. (a): Original image. (b): Result of minimum value estimate. (c): (b) after applying median estimate technique. (d): (c) after closing. (e): (d) after mean filtering.

result is again a considerable improvement. A closing kernel with a radius of 15 pixels was found to provide the best results. For the sake of comparison, a kernel with a radius of 10 pixels produced a average performance of 37 compared to the 47 of the 15 pixel one. A final averaging filter is again applied, to try and remove the largest peaks. The smoothing does remove the peaks and in most cases does increase the average mean value of the object. Unfortunately it does not increase the dynamic range and the resulting decreased contrast also increases ΔM. The resulting performance figure does not show much of an improvement.

It is therefore apparent that the inherent uniformity produced by the minimum window technique in conjunction with the closing operation provides a considerable improvement to the ease of segmentation in most cases.

	Median Est.	+ Mean Filt	+ Closing	+ Opening Closing
$\mathcal{P}_E$ (Avg)	31	39	38	45

Table 6.5: Performance values for the median estimate sequences.



(a) (b) (c) (d) (e)

Figure 6.8: Median value estimate. (a): Original image. (b): Result of median value estimate. (c): (b) after applying a averaging filter. (d): (b) after closing. (e): (b) after applying opening and closing.

Median Estimate Of the background removal techniques the median estimate is the only one capable of removing the crossover region. As a result is has already been employed in conjunction with the previous two techniques. Before any further processing is added to the sequence, it already performs quite well and provides a good improvement on the dynamic range of the original image.

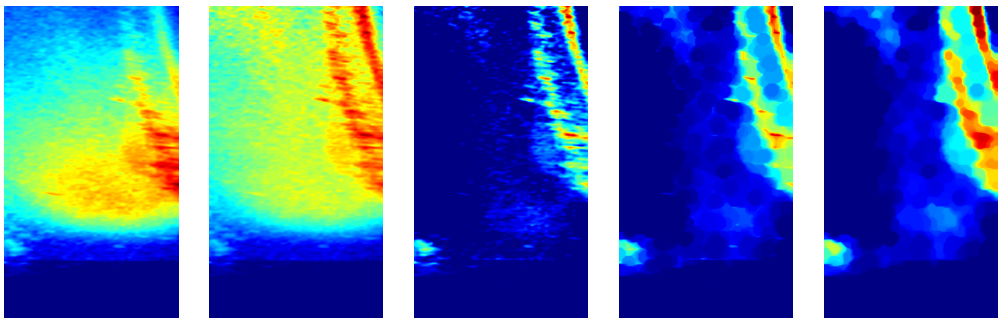
Unfortunately, it does not provide any real range equalisation as the object intensities at the far end of the image remain lower than those in the middle, similar to the original image, as seen in fig. 6.8(b). As a result, even after applying either a large 20×20 mean filter or a closing kernel with a radius of 15, the object uniformity does not increase as dramatically as in the previous two cases. This is confirmed by the performance values of table 6.5. Although a clear improvement is seen, the overall performance is still below that of the minimum value estimate sequences.

The median estimate does however have a rather high mean value with the initial object reflections having large peaks above this mean. To improve the uniformity, these peaks can be removed with an opening procedure, followed by a closing operation to fill the remaining valleys. The opening procedure should however be used with caution since it can completely remove object features. As a result, a small opening kernel

with a radius of 5 pixels is applied, followed by a closing kernel of radius 20. This sequence of processing provides quite good results and is much more comparable to those of the minimum window estimate.

	Range Eq	+ Median Est	+ Closing	+ Opening
$\mathcal{P}_E$ (Avg)	31	29	56	65

Table 6.6: Performance values for the range equalisation sequences.



(a) (b) (c) (d) (e)

Figure 6.9: Range Equalisation. (a): Original image. (b): Result of range equalisation. (c): (b) after applying median estimate technique. (d): (c) after closing. (e): (d) after opening.

Range Equalisation The range equalisation technique produces a very uniform representation of both the background and object intensities. It can however benefit from increased dynamic range to simplify the thresholding process. To improve this, the median estimate technique is first applied. This results in a lower performance value, since the original uniformity is reduced, forming larger peaks and valleys in the intensity profile.

To regain the original uniformity a closing kernel with a radius of 20 pixels is applied, resulting in an immediate and drastic improvement in the performance rating. With the valleys filled the only way to further improve the uniformity and increase the dynamic range, is to remove the peaks still present. To this end an opening kernel with a radius of 8 is applied, again resulting in a considerable improvement.

It was also found that applying the median estimate after the closing and opening techniques results in the same performance value. From both the performance values and visual results of all the techniques, it is clear that the range equalisation technique produces superior uniformity and therefore noticeably better results.

Chapter 7

Conclusions and Recommendations

7.1 Conclusions

As described in chapter 1 the focus of this thesis was the analysis and post processing of sonar images to be used as part of the collision avoidance system of an AUV (Autonomous Underwater Vehicle). These images are noise prone and when taken in shallow water suffer from strong bottom reflections, obscuring returns from targets. As a result, the testing and development of post processing techniques was necessary to ensure proper target extraction. The research done here can be separated into four major topics:

1. The exploration of sonar parameters and their influence on the returned images.
2. The identification and testing of available noise reduction techniques.
3. The identification and testing of available techniques as well as the development of new techniques for the removal of bottom reflections.

4. The development of a quantitative measure of performance and evaluation of the proposed techniques according to this measure.

Sonar Parameters

The images used as part of this process was taken using a commercially available sector scan sonar. This sonar was chosen since it is representative of the current sonar technology used for collision avoidance. The basic principles of the sonar operation and the imaging process are described in section 3.2.

Further analysis of the images taken with this sonar identified the *step angle*, *image gain* and *elevation angle* as important attributes requiring consideration prior to sonar operation.

- As concluded in section 4.1 using a step angle of less than half of the horizontal beamwidth shows no apparent improvement to the image detail and does not warrant the increase in data rates and sweep time.
- The image gain should also be set as low as possible to reduce the intensity of bottom reflections and increase the contrast of the image.
- The elevation angle of the sonar along with the vertical beamwidth influences the amount of energy reflected from the bottom, as illustrated in fig. 4.5. This should be limited by mounting the sonar tilted upward, depending on the depth being worked at.

Noise Reduction

From the review of related literature, section 3.3, and the analysis of example images, noise reduction was identified as the first step in the post

processing cycle. Because of the noisy nature of the images, noise reduction can have a significant improvement on the uniformity of object reflections, whilst reducing the noise floor and improving the likelihood of correct identification.

Reduction of noise in the spatial domain is standard practise in many fields of image processing, and has also been applied to sonar processing. Although other methods exist, the non-linear *median filter* and the linear *averaging filter* are the two most commonly used.

Using two sets of images with varying amounts of noise, as in section 5.1.1, both these techniques were evaluated. In both cases the more computationally expensive median filter provided no apparent improvement over the simpler averaging filter, which provided adequate noise reduction and improvement in uniformity by grouping similar pixels. As such, it is deemed both necessary and adequate as a first stage of processing.

The use of *linear interpolation* was also evaluated, as in section 5.1.2. This interpolation is applied during the image creation stage to augment the lack of data from the scanning process. Using the suggested algorithm of Eq. 5.1.1 also provides a fair amount of noise reduction and improves the overall image quality.

The use of *grey scale opening and closing* was also evaluated as a means to reduce the noise. In this regard both techniques achieved some success. It is however the improvement in uniformity, not found with any other that techniques, that make these techniques useful.

Removal of Bottom Reflections

By examining a real world example of a set of images taken from a AUV in a harbour, the reflections from the harbour floor was identified as a significant problem. These strong reflections have intensities comparable to those from possible obstructions and therefore hamper the correct extraction of these targets. This problem was not identified in the litera-

ture review, and as such possible solutions had to be identified. Due to the lack of a reference to the possible extent of the problem, the proposed solutions had to contain an inherent amount of flexibility.

Since the bottom reflections cover much of the image it was deemed as an analogue to the background in general imaging terms. Due to similar characteristics, removal or suppression was approached as a non-uniform illumination problem, as found in many imaging procedures. From the general image processing literature two techniques was identified as possible solutions. *High-pass filtering* and *minimum value background estimation*, as described in sections 5.2.1 and 5.2.2 respectively, both rely on a certain amount of uniformity in the background, something that the given data does not adhere to at all times, and achieved mixed results with the data set.

High-pass filtering was able to reduce a certain amount of the background, but in general the characteristics of the background is not simply confined to the lower frequencies and this technique does not perform sufficiently for this test case.

The minimum value estimate is more robust to local variation and performed considerably better than high-pass filtering. It does however reduce the contrast contained in the image when large objects are present. These standard methods can provide sufficient background removal when the background is reasonably uniform, but do not perform desirably with the examined data.

To achieve the desired results with the example data, two further techniques were developed. As described in section 5.2.3, both the *range equalisation* and *median value background estimate* rely on the principle that the median value of each row in the image is a good estimate of the background value. For the given data, this principle does hold true and both these techniques provide excellent results. The median estimate provides a closer estimate of the background when compared to the minimum value technique and results in better contrast. The range equalisation

technique provides an significant improvement in the background uniformity and as a result targets are easily discernible solely based on the intensity difference.

Performance Measure and Evaluation

Through further examination of these techniques and the examples given in section 6.1, the *dynamic range* contained in the image and the *uniformity* of the resultant objects were identified as the two determining factors in assigning a performance value to a technique. The dynamic range gives a value to the separation between the background and maximum values of an object, whilst the uniformity gives a value to the variation found in these maximum values. Using these two values as in eq. 6.2 the output from a technique can be evaluated quantitatively and is a valuable tool to assess and compare the performance of possible techniques.

As described in section 6.3, post-processing sequences was constructed from the presented techniques, and evaluated using the performance measure. In each case a different background removal technique was used as the starting point for the processing, and the result was evaluated using a set of 10 similar images. The sequences and the resulting average performance values are presented here in table 7.1.

	High Pass Median Est Closing Mean Filt	Min Val Est Median Est Closing Mean Filt	Median Est Opening Closing	Range Eq Median Est Closing Opening
$\mathcal{P}_E$ (Avg)	37	48	45	65

Table 7.1: Performance values for the processing sequences using bottom reflection removal techniques.

Using the examples in section 6.3 as a reference, a performance value of 35–40 would show an acceptable result, with good background removal and acceptable uniformity. A value 41–50 indicates a good result,

with good background removal and good uniformity. A value of higher than 50 is a very good result and a very uniform target return can be expected along with very good separation between the target and background intensities.

Summary of Conclusions

In all the different background removal sequences considerable improvement in dynamic range was achieved with the median estimate technique and improvement in uniformity with the closing procedure. It was however the use of these two operations in conjunction with the range equalisation technique that provided the best results. The range equalisation technique already provides a considerable improvement in uniformity and by adding the closing and opening procedures this is further accentuated. Adding the improvement in dynamic range of the range equalisation technique, fully complements these techniques and provides considerable better results than any of the other procedures.

In general the removal of bottom reflections and the improvement in target uniformity can be seen as a successful contribution. The problem was identified in real world images and no previous solution could be found in the literature, for either review or comparison. Although, high-pass filtering and the minimum value estimate was identified in the image processing literature as solutions to similar problems, they could not sufficiently be adjusted for the problem. As such, the development of the *median background estimate* technique provides a novel solution to the problem and should be of use in further research on the subject.

As a first investigation, as part of the greater research initiative, into the use of sonar images for collision avoidance, a fair amount of knowledge was gained into the subject. The approach taken does however mirror those taken by previous researchers and different approaches may exist.

7.2 Recommendations

- The techniques represented here fully cover the denoising and target improvement stage of the segmentation process. Along with the work done in the literature there is not enough scope for further investigation. It would therefore not be recommended to re-evaluate the subject, unless the procedures are to be extended to use additional information from, for example, previous images or other sensors.
- The logical next step is the development of thresholding techniques. Using the bottom removal techniques enables the use of a single global threshold and using a constant threshold value may even be possible. This would then be followed by the extraction of targets from the thresholded images and representation of the data in such a way that it can be passed to the collision avoidance algorithms.
- Implementation of the techniques into hardware may require careful thought with regard to memory management and removal of the interpolation stage could be considered to keep the amount of data to a minimum.

Bibliography

[1] Tritech International Limited, "http://www.tritech.co.uk/", 2008.

[2] Imagenex Technology CORP., No.209 - 1875 Broadway St., Port Coquitlam,Canada, *Sonar Theory and Applications*, Excerpt from IMAGENEX model 855 Color Imaging sonar user's manual.

[3] Rafael C. Gonzalez and Richard E. Woods, *Digital Image Processing*, Prentice Hall, second edition, 2002.

[4] William K. Pratt, *Digital Image Processing*, Wiley-Interscience, third edition, 2001.

[5] John G. Proakis and Dimitris G. Manolakis, *Digital Signal Processing Principles, Algorithms and Applications*, Prentice Hall, third edition, 1996.

[6] David M. Lane, Mike J. Chantler, and Dongyong Dai, "Robust tracking of multiple objects in sector-scan sonar image sequences using optical flow motion estimation", *IEEE Journal of Oceanic Engineering*, vol. 23, no. 1, pp. 31–46, January 1998.

[7] Anil K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, 1989.

[8] Sylvie Daniel, Fabrice Le Leannec, Christian Roux, Bassel Solaiman, and Eric P. Maillard, "Side-scan sonar image matching", *IEEE Journal of Oceanic Enineering*, vol. 23, no. 3, pp. 245–259, July 1998.

[9] Robert J. Urick, *Principles of Underwater Sound*, McGraw-Hill, 1975.

[10] I. Quidu, J. Ph. Malkasse, G. Burel, and P. Vilbé, "A 2-d filter specification for sonar image thresholding", Advanced Concepts for Intelligent Vision Systems, August 2001.

[11] Heinz G. Urban, *Handbook Of Underwater Acoustic Engeneering*, STN ATLAS Elektronik GmbH, 2002.

[12] B.W. Flemming, M. Klein, and P.N. Denbigh, *Recent Developments in Side Scan Sonar Techniques*, W.G.A. Russel-Cargill, 1982, Ed. W.G.A. Russel-Cargill.

[13] R. Coetzee, "Signal development for chirp side-scan sonar", Master's thesis, University of Stellenbosch, Desember 1992.

[14] Francois Le Chevalier, *Principles of Radar and Sonar Signal Processing*, Artech House, 2002.

[15] Andrea Trucco, Maria Palmese, and Stefania Repetto, "Image projection and composition with a front-scan sonar system: Methods and experimental results", *IEEE Journal of Oceanic Enineering*, vol. 28, no. 4, pp. 687–698, October 2003.

[16] David M. Lane and John P. Stoner, "Automatic interpretation of sonar imagery using qualitative feature matching", *IEEE Journal of Oceanic Enineering*, vol. 19, no. 3, pp. 391–405, July 1994.

[17] M.J. Chantler, D.M. Lane, D. Dai, and N. Williams, "Detection and tracking of returns in sector-scan sonar image sequences", *Proc. Inst. Elect. Eng. Radar, Sonar and Navigation*, vol. 143, no. 3, pp. 157–162, June 1996.

[18] M. J. Chantler and J. P. Stoner, "Automatic interpretation of sonar image sequences using temporal feature measures", *IEEE Journal of Oceanic Enineering*, vol. 22, no. 1, pp. 47–56, January 1997.

[19] Yvan Petillot, Ioseba Tena Ruiz, and David M. Lane, "Underwater vehicle obstacle avoidance and path planning using a multi-beam forward looking sonar", *IEEE Journal of Oceanic Engineering*, vol. 26, no. 2, pp. 240–251, April 2001.

[20] Stuart W. Perry and Ling Guan, "A recurrent neural network for detecting objects in sequences of sector-scan sonar images", *IEEE Journal of Oceanic Enineering*, vol. 29, no. 3, pp. 857–871, July 2004.

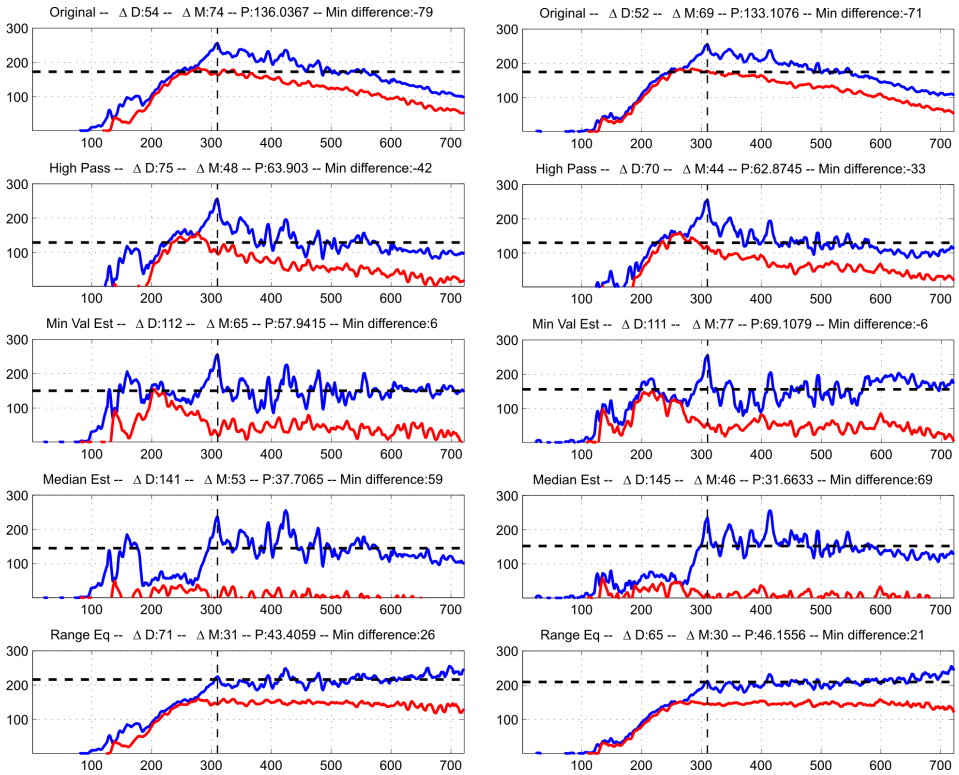
[21] John C. Russ, *The Image Processing Handbook*, CRC Press, second edition, 1995.

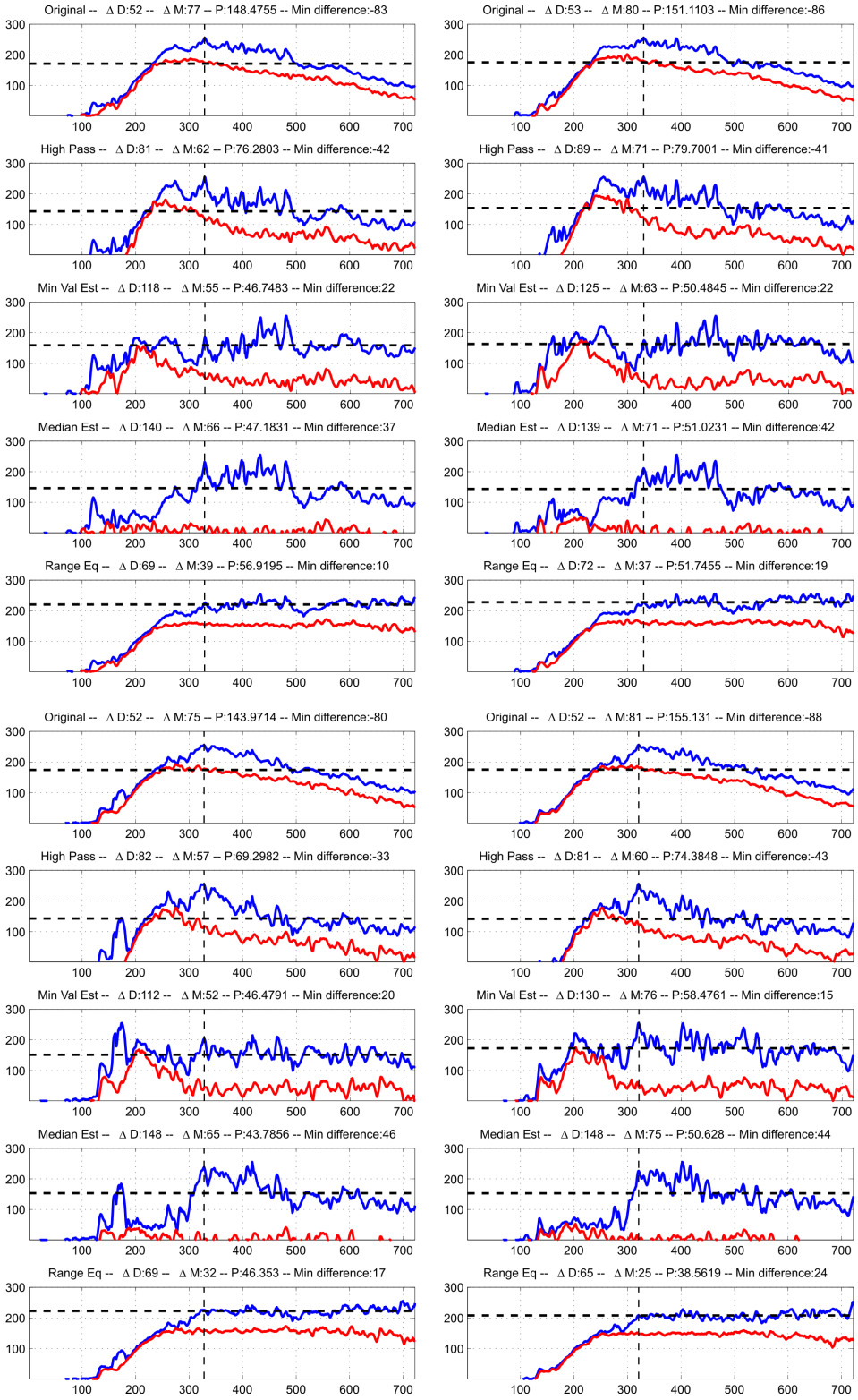
[22] Mark Russell, "Sonar and how it builds a spatial mapping through the use of sound", Tech. Rep., University of Sydney, 2007.

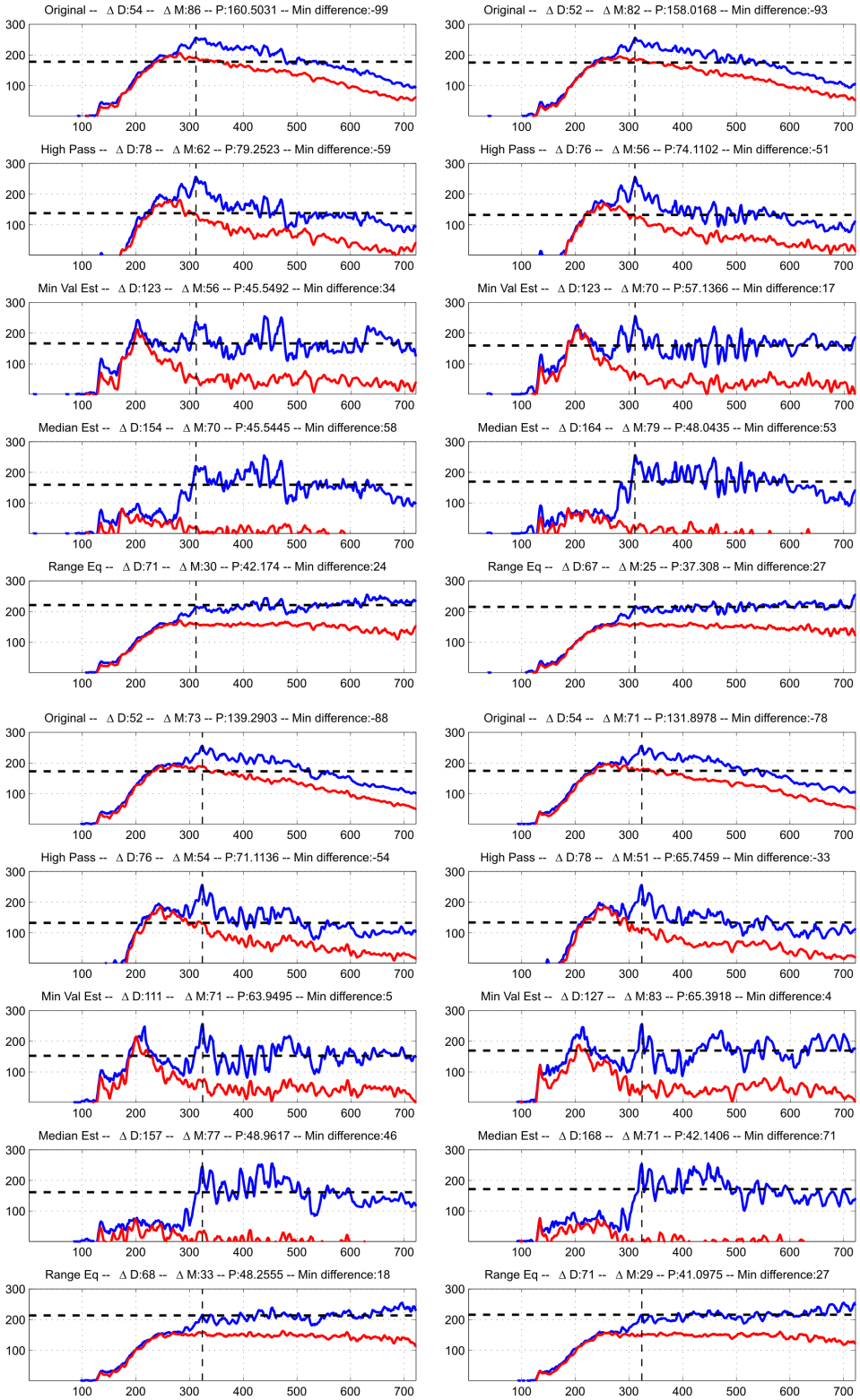
Appendices

Appendix A

Test Data




Original -- Δ D:52 -- Δ M:77 -- P:148.4755 -- Min difference:-83
Original -- Δ D:53 -- Δ M:80 -- P:151.1103 -- Min difference:-86
High Pass -- Δ D:81 -- Δ M:62 -- P:76.2803 -- Min difference:-42
High Pass -- Δ D:89 -- Δ M:71 -- P:79.7001 -- Min difference:-41
Min Val Est -- Δ D:118 -- Δ M:55 -- P:46.7483 -- Min difference:22
Min Val Est -- Δ D:125 -- Δ M:63 -- P:50.4845 -- Min difference:22
Median Est -- Δ D:140 -- Δ M:66 -- P:47.1831 -- Min difference:37
Median Est -- Δ D:139 -- Δ M:71 -- P:51.0231 -- Min difference:42
Range Eq -- Δ D:69 -- Δ M:39 -- P:56.9195 -- Min difference:10
Range Eq -- Δ D:72 -- Δ M:37 -- P:51.7455 -- Min difference:19
Original -- Δ D:52 -- Δ M:75 -- P:143.9714 -- Min difference:-80
Original -- Δ D:52 -- Δ M:81 -- P:155.131 -- Min difference:-88
High Pass -- Δ D:82 -- Δ M:57 -- P:69.2982 -- Min difference:-33
High Pass -- Δ D:81 -- Δ M:60 -- P:74.3848 -- Min difference:-43
Min Val Est -- Δ D:112 -- Δ M:52 -- P:46.4791 -- Min difference:20
Min Val Est -- Δ D:130 -- Δ M:76 -- P:58.4761 -- Min difference:15
Median Est -- Δ D:148 -- Δ M:65 -- P:43.7856 -- Min difference:46
Median Est -- Δ D:148 -- Δ M:75 -- P:50.628 -- Min difference:44
Range Eq -- Δ D:69 -- Δ M:32 -- P:46.353 -- Min difference:17
Range Eq -- Δ D:65 -- Δ M:25 -- P:38.5619 -- Min difference:24


Original -- Δ D:54 -- Δ M:86 -- P:160.5031 -- Min difference:-99
Original -- Δ D:52 -- Δ M:82 -- P:158.0168 -- Min difference:-93
High Pass -- Δ D:78 -- Δ M:62 -- P:79.2523 -- Min difference:-59
High Pass -- Δ D:76 -- Δ M:56 -- P:74.1102 -- Min difference:-51
Min Val Est -- Δ D:123 -- Δ M:56 -- P:45.5492 -- Min difference:34
Min Val Est -- Δ D:123 -- Δ M:70 -- P:57.1366 -- Min difference:17
Median Est -- Δ D:154 -- Δ M:70 -- P:45.5445 -- Min difference:58
Median Est -- Δ D:164 -- Δ M:79 -- P:48.0435 -- Min difference:53
Range Eq -- Δ D:71 -- Δ M:30 -- P:42.174 -- Min difference:24
Range Eq -- Δ D:67 -- Δ M:25 -- P:37.308 -- Min difference:27
Original -- Δ D:52 -- Δ M:73 -- P:139.2903 -- Min difference:-88
Original -- Δ D:54 -- Δ M:71 -- P:131.8978 -- Min difference:-78
High Pass -- Δ D:76 -- Δ M:54 -- P:71.1136 -- Min difference:-54
High Pass -- Δ D:78 -- Δ M:51 -- P:65.7459 -- Min difference:-33
Min Val Est -- Δ D:111 -- Δ M:71 -- P:63.9495 -- Min difference:5
Min Val Est -- Δ D:127 -- Δ M:83 -- P:65.3918 -- Min difference:4
Median Est -- Δ D:157 -- Δ M:77 -- P:48.9617 -- Min difference:46
Median Est -- Δ D:168 -- Δ M:71 -- P:42.1406 -- Min difference:71
Range Eq -- Δ D:68 -- Δ M:33 -- P:48.2555 -- Min difference:18
Range Eq -- Δ D:71 -- Δ M:29 -- P:41.0975 -- Min difference:27