Ambient Noise Imaging; enhanced spatial correlation algorithms and a way to combine independent images for improved stability and false-alarm rejection.

Choon Kiat Lim & John R. Potter.
Acoustic Research Laboratory, Tropical Marine Science Institute, NUS, 12a Kent Ridge Crescent, Singapore 119223. http://www.arl.nus.edu.sg Email: johnp@arl.nus.edu.sg

Acoustic daylight, second-order temporal and second-order spatial imaging are three algorithms that have been devised for ambient noise imaging. This paper proposes two enhancements (enhancement I and II) to improve the image quality obtained by spatial imaging. Enhancement I repeats the original spatial imaging algorithm for a limited number of times and achieves an improvement of up to 10.5dB. Enhancement II uses linear programming to maximize the separation between the target and non-target sets, achieving up to 2 dB benefit. Enhancement I and II are combined to produce a better result than enhancement I alone with its highest improvement at a further 7.5 dB. We also propose a fusing algorithm, using K-mean clustering with a validity measure, to combine images produced by different algorithms so as to improve the stability and false alarm rejection of the images. An average additional improvement of 2 dB is obtained when compared to direct fusing without using clustering. Better results are usually obtained if the features (i.e. results from different algorithms at a particular pixel) at one channel are sorted prior to clustering.

Keywords: ambient noise imaging, acoustic daylight, second-order statistics, spatial imaging, image fusion, clustering.

INTRODUCTION.

Active and passive acoustic imaging are used extensively underwater. Both types of systems try to minimise noise to improve performance. A revolutionary idea, called 'Acoustic Daylight' using ambient noise to form images of submerged objects emerged a decade ago [1]. This concept is analogous to optical vision, where a diffuse set of light sources are replaced by an ambient noise field produced, for example, by snapping shrimp in tropical waters. The advantage of using ambient noise is that silent targets can be detected without insonifying them with active sonar.

The first ambient noise imaging system, the Acoustic Daylight Ocean Noise Imaging System (ADONIS) consisted of 126 receivers sensitive over 25-85 kHz, arranged in an elliptical pattern at the focus of a spheroidal reflecting lens. ADONIS yielded excellent results at ranges up to the maximum tried, 38m [2]. Following this success, several new imaging algorithms have been developed for second-generation digital ambient noise imaging systems. The Acoustic Research Laboratory (ARL) at the Tropical Marine Science Institute, National University of Singapore is currently building a second-generation ambient noise system, the Remotely Operated Mobile Ambient Noise Imaging System (ROMANIS).

There are, so far, three main statistical algorithms that have been tested successfully numerically [3]. The first is 'Acoustic Daylight' which images the mean intensity, analogous to optical vision. The second is based on a second-order intensity statistic, the variance of the acoustic intensity and has no visual analogue. The third relies on second-order spatial intensity and is even more removed from intuitive imaging ideas. We present investigations to improve this algorithm.

The algorithms are independent in the sense that the images formed in the absence of a target are statistically uncorrelated. There is therefore also a potential to fuse images produced by different algorithms to obtain a more reliable image with improved false alarm rejection.

I. SECOND-ORDER SPATIAL CORRELATION IMAGING

The idea behind this is that the intensity time series of target channels (i.e. channels receiving reflected energy from the target) will cross correlate with higher coefficients than correlations with non-target channels. To test this hypothesis, a normalized correlation coefficient matrix (i.e. 126 x 126 matrix) for ADONIS data is computed. The diagonal elements of this matrix are unity and the non-diagonal elements are the 0-lag cross correlations between two different channels. Remember that the channels are already beamformed. The two channels associated with the minimum coefficient are taken to be reference target and non-target channels and are used as seeds to form target and non-target sets. The extent to which other channels correlate to these two seed channels can be used to form an image. This can be done by calculating the normalized distance of each of the remaining channels from the two seed channels in the correlation space and assigning a value between 0 and 1 that represents the relative distance to each of the two seeds. The pseudocode is shown below, where channel A and B are the seed channels.

```
for (i = 1 to 126)
        c1(i) = normalised distance of channel i from channel A
        c2(i) = normalised distance of channel i from channel B
end
 for (i = 1 to 126)  either
        result(i) = c1(i) / (c1(i) + c2(i))
        or
```

result(i) = c2(i) / (c1(i) + c2(i))

end

From the pseudocode, two possible images are produced. One of the images is arbitrarily chosen. One simple way is to assign the seed channel with higher variance to have higher pixel intensity, an approach that will normally result in 'frontlit' target images.

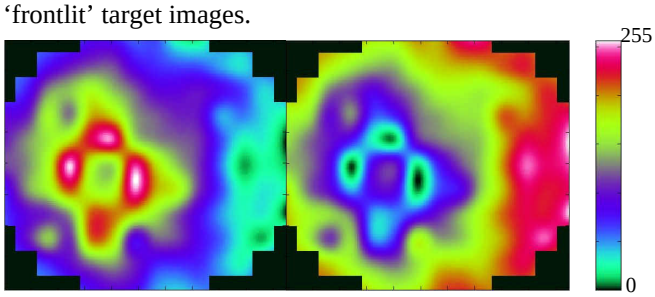


Figure 1. Two possible images produced by spatial imaging algorithm

II. ENHANCEMENT I

This enhancement works by iterating the algorithm, re-selecting the seed channels each iteration to be the ones with the next-most minimum cross-correlation:

$$\min(n) < A_{i,j} \quad i,j \ \ S \cap T'$$

where min(n) is the minimum coefficcient at the n^{th} iteration, S is the set of all possible channel pairs & T is the set of channel pairs giving the previous few minimum coefficients.

The number of images formed is equal to the number of iterations. A simple way to combine the images is to multiply the linear pixel values at each channel together. This is a dB averaging technique that results in target regions showing up more clearly. Similarly, non-target regions tend to be less noisy. Thus an image with higher contrast is anticipated.

There is, however, a polarity problem in combining these images. Images produced at subsequent iterations may be of opposite polarity (i.e. some may have target pixels having higher values than non-target pixels and vice versa for others). We need to ensure that images are of similar polarity before fusing. This can be achieved by correlating the selected seed channels at each iteration with the first pair of seeds to establish a uniform polarity prior to fusing.

A. Limitations to enhancement I

The primary limitation for this enhancement is that it obviously cannot be iterated indefinitely. The absolute limit is 63, but in practice the minimum correlation coefficient increases at every iteration. The algorithm is instructed to stop iterating when the minimum correlation coefficient exceeds a given bound. One simple and intuitive way to set this bound is to take a threshold equal to the average of the highest and lowest cross correlation coefficient from the matrix and set the upper bound equal to the average of the cross correlation coefficients that are lower than the threshold.

B. Data set and performance measures

A subset was selected from the original ADONIS data to test the performance of the enhanced algorithm. The data set consists of 934 frames taken while oriented towards a fenestrated cross target that remains stationary at 38 m range throughout [3]. The actual locations of the target and non-target pixels are known. Using this knowledge, we devise two performance measures for the evaluation of the enhanced algorithm.

The first is the ratio of the mean target pixel value to the mean non-target pixel value. This is termed 'target to non-target ratio' (TNR). Higher TNR indicates that targets are seen clearly against the background.

The second is to use K-mean clustering to separate the image into two groups (target and non-target sets). We then estimate the total number of misclassified (non-target pixels misrepresented as target pixels and vice versa). Fewer misclassified pixels mean that the outline of the target is more clearly defined and false targets reduced.

C. Performance analysis for enhancement I

Figure 2 shows a comparison of the original spatial correlation and enhanced algorithms from frame # 200-350 of the data set with the length of the time series to be correlated set to 10 frames. The plot shows the increase in TNR (in dB) if enhancement I is used instead of the original algorithm. It also shows the decrease in the total number of misclassified pixels.

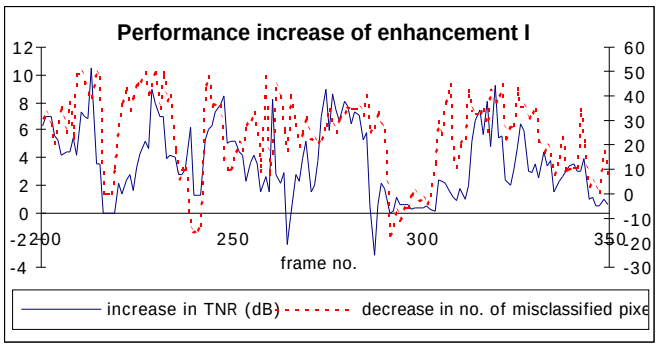
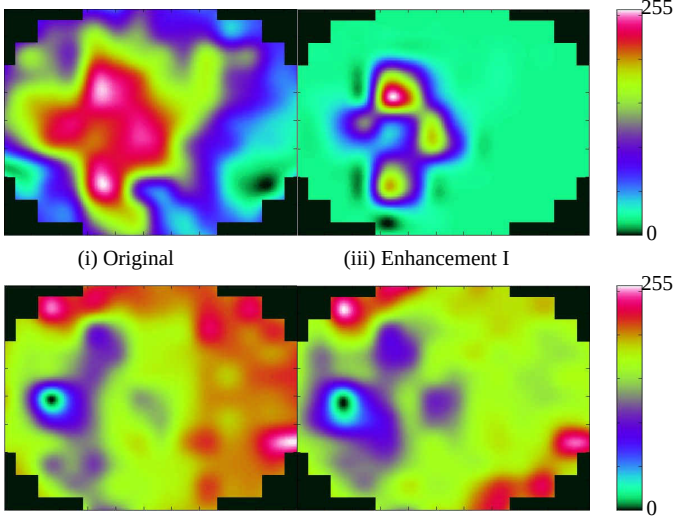


Figure 2. Performance increase of enhancement I on data set A (frame no 200-350).

Figure 2 shows that the performance of the enhanced algorithm is indeed better for this test series, indicated by mostly positive values of the two performance measures. The increase in TNR peaks at 10.5 dB. The maximum decrease in the number of mis-assigned pixels is 52. The improvement in performance is usually seen as an increase in image contrast and the dilution of background noise. However, in some frames the TNR decreases but the number of misclassified pixels still increases. This may be because no target is observed at that particular frame when both algorithms (i.e.

original and enhancement I) are used. Thus the result is irrelevant in comparing the performance of both algorithms.

Figure 3 shows the input and output images of a few frames to show the comparison subjectively.



(i) Original (iii) Enhancement I

(ii) Original (iv) Enhancement I

Figure 3. Images produced by original and enhancement I algorithms

By observing the difference between the images produced by the two algorithms, images on the right (i.e. 3(iii) and (iv)) show a clearer defined outline of the target when compared to the images on the left (i.e. figure 3(i) and (ii)). It is noted that figure 3(iv) shows an enhancement in the reverse polarity. The contrast of the enhanced images is higher.

III. ENHANCEMENT II

This method aims to maximise the separation between target and non-target sets. Firstly, the correlation coefficient matrix and seed channels are found. All channels are then separated into target and non-target sets by means of K-mean clustering. The next step is to cast the problem into two linear programming problems such that the optimised representative signal for each set can be found. For the case of the target set, since a representative signal is to be found to replace the target seed channel, the intra-correlations between the optimised target channel and the rest of the channels in the supposed target set are maximised while the inter-correlations between the optimised target channel and the presumed non-target channels are minimised. A symmetric process applies to the non-target set. The pseudocode for the optimisation of the target set is shown below. Cor(x,y) represents the correlation between channel x and y. It is assumed that channel A is the target seed channel while channel B is the non-target seed channel. {TS = x_1, x_2,… x_T} contains the target pixels where T is the number of target channels while {NTS = y_1, y_2,… y_{NT}} contains the non-target pixels where NT is the number of non-target channels. x_{opt} and y_{opt} are the variables to be optimised. For target channels:

$$\text{Max} \sum_{i=1}^{T-1} Cor(x_{opt}, x_i) \;\&\; \text{Min} \sum_{j=1}^{NT} Cor(x_{opt}, y_j),\ \text{i} \neq \text{A}$$

subject to

$$\sum_{i=1}^{T-1} Cor(x_{opt}, x_i) \geq \sum_{i=1}^{T-1} Cor(x_A, x_i),\ \text{i} \neq \text{A}$$

$$\sum_{j=1}^{NT} Cor(x_{opt}, y_j) \leq \sum_{j=1}^{NT} Cor(x_A, y_j)$$

For non-target channels:

$$\text{Max} \sum_{i=1}^{NT-1} Cor(y_{opt}, y_i) \;\&\; \text{Min} \sum_{j=1}^{T} Cor(y_{opt}, x_j),\ \text{i} \neq B$$

subject to:

$$\sum_{i=1}^{NT-1} Cor(y_{opt}, y_i) \geq \sum_{i=1}^{NT-1} Cor(y_B, y_i),\ \text{i} \neq B$$

$$\frac{1}{T} \sum_{j=1}^{T} Cor(y_{opt}, x_j) \leq \frac{1}{T} \sum_{j=1}^{T} Cor(y_B, x_j)$$

From the above pseudocode, we see that the objective function and the constraints are highly non-linear. In order to convert the problem into a simpler linear programming problem, we begin by normalising all the channels to zero mean and unit variance. Let the normalized signal received by target channel i and non-target channel j be $\mathbf{x_i}$ and $\mathbf{y_j}$ respectively. The optimisation algorithm for the target channel is then reduced to:

$$\text{Maximise} \quad \sum_{i \in TS} \mathbf{x_{opt}}\ \mathbf{x_i},\ \text{i} \neq \text{A}$$

$$\text{Subject to} \quad \sum_{i \in TS} \mathbf{x_{opt}}\ \mathbf{x_i} > thres1,\ \text{i} \neq \text{A}$$

$$\sum_{j \in NTS} \mathbf{x_{opt}}\ \mathbf{y_j} < thres2$$

$$\sum \mathbf{x} = 0 \quad \frac{1}{\text{N} - 1} \sum \mathbf{x}^2 = 1 \ , \text{N} = period$$

where

$$thres1 = \sum_{i \in TS} \mathbf{x_A}\ \mathbf{x_i},\ \text{i} \neq \text{A};\ thres2 = \sum_{j \in NTS} \mathbf{x_A}\ \mathbf{y_j}$$

Although the multi-objective function discussed earlier can be modeled as a weighted sum of both intra and inter-correlation functions, it is simpler to implement a single objective function. Here, only the intra-correlation is maximised.

Note that the solution for the above optimisation will have zero mean and unit variance. If a solution is found, it will replace the signal of the target seed channel. The above procedure also applies to the non-target set. The new correlation coefficient matrix is then recomputed and the procedure is repeated. This causes crossovers to occur

between the two sets. The procedure stops when crossover stops.

A. Performance analysis for enhancement II

The algorithm was run on frames 350-500 of the test data set over an interval of 10 frames. Figure 4 shows the performance increase of enhancement II over the original algorithm.

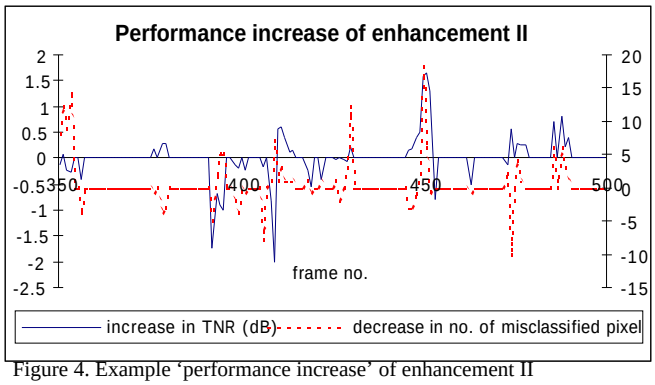
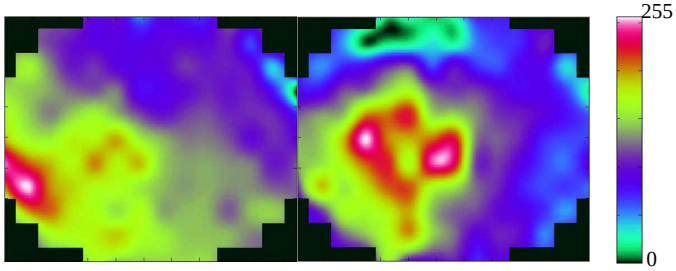
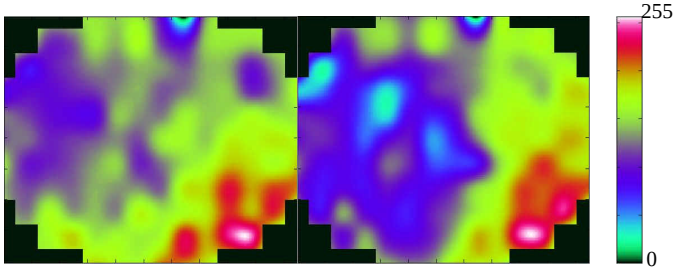


Figure 4. Example 'performance increase' of enhancement II

For these chosen frames, the enhancement algorithm produces exactly the same result as the original one for most of the frames. This may indicate that the original seed channels are very close to ideal target and non-target representations for most frames. The maximum increase in TNR is ~2 dB, considerably lower than results obtained by the enhancement I algorithm. By looking at some example images, the subjective effect can be explored.



(i) Original algorithm (iii) Enhancement II



(ii) Original algorithm (iv) Enhancement II

Figure 5. Original and optimised images

In figure 5(i) and 5(ii), the target cannot be seen using the original algorithm, but iterative optimisation leads to slightly clearer images shown in figure 5(iii) and 5(iv). Note that figure 5(iv) shows the enhancement of the target in the 'reverse' polarity.

IV. COMBINATION OF ENHANCEMENT I AND II

It is possible to combine enhancement I and II. Note that in enhancement II, the signal is normalized to zero mean and unit variance. Since we no longer have the variance information, we cannot use enhancement I directly as we are unable to differentiate between target and non-target seed channels. This problem can be avoided by determining the target and non-target channels from the first iteration as the reference. Variance information can be used in this case as the signal at the first iteration is not yet normalized. Subsequent target and non-target channels are determined from the cross correlation between the seed channels to be determined and the reference channels. This ensures that the images have the same polarity.

A. Performance analysis of combined enhancements I and II

The combined algorithm was applied to the same section of the test data set used in evaluating the performance of enhancement II. The performance is compared with results produced by the enhancement I alone in Figure 6.

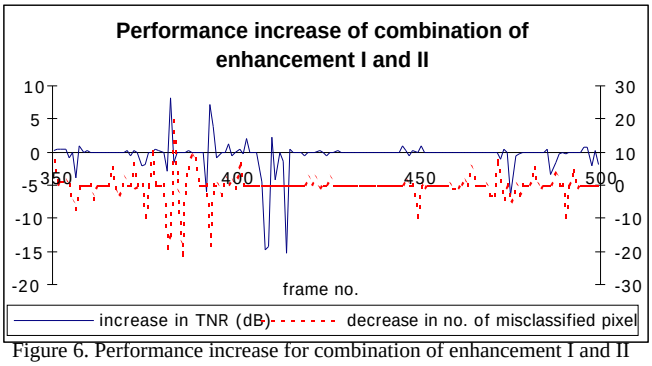
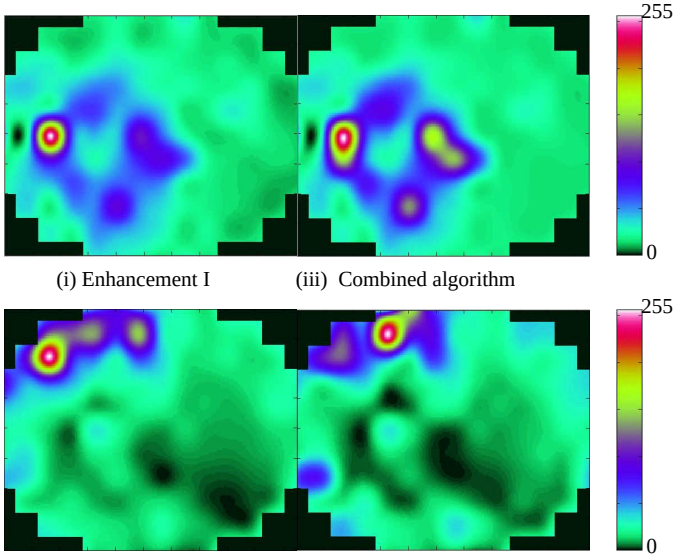


Figure 6. Performance increase for combination of enhancement I and II on section of data set A (i.e. frame no.350-500).

As can be seen from the plot, there are no significant changes in the image quality for most of the frames. This is expected from the performance of enhancement II illustrated in Figure 4, where no significant optimisation is possible for many frames. However, there are also some significant changes in a few frames, apparently both good and bad. The best improvement is ~7.5 dB, while the greatest negative change in TNR is ~-15 dB. An example image corresponding to a positive TNR change is shown in figure 7 (i) and (ii). Negative TNR changes do not necessarily imply that the quality of the image is degraded. Instead, it may mean that the outline of the target becomes more prominent in the opposite polarity. This is illustrated in figure 7 (iii) and (iv) where the target pixels become darker implying that target pixels are intensified by the optimization algorithm.



(i) Enhancement I (iii) Combined algorithm

(ii) Enhancement I (iv) Combined algorithm

Figure 7. Comparison of image produced by enhancement I and combination of both enhancement methods in the reverse polarity

V. FUSING IMAGES

A. Introduction

Each independent imaging algorithm will produce a different image from the same data. Some may show strong signals while others show faint signals in the supposed target region. Furthermore, the non-target region may be noisy for some images but not so or in a different way for others. Therefore a way is needed to fuse independent images to produce a final image with improved stability and better false alarm rejection.

Ultimately, a target area is identified by being anomalous in comparison with the background, which is expected to occupy most of the field of view, being relatively bland and unstructured. The simplest way to fuse images is to average them (first ensuring they are all of the same polarity). Although averaging enhances the images, it is unable to reduce the impact of impulsive noise efficiently. Thus it is proposed that clustering is done prior to averaging. This groups pixels into target and non-target sets. Several clustering techniques have been proposed in the literature. These include K-mean clustering and fuzzy C mean clustering [4-6]. Since the algorithm will be used in a real time application, a reliable and efficient method is needed. We have therefore selected K-mean clustering to provide a fast and simple algorithm. Validity measures [4] are also used to determine the ideal number of clusters.

B. Preprocessing

Consider a scenario whereby images produced by three algorithms (acoustic daylight, variance and spatial correlation) are to be fused. Note that all three algorithms are capable of producing images with targets of either polarity. Thus preprocessing is needed to ensure that all images are of the same polarity. This can be done by finding the dot-product cross-correlation coefficient between all possible pairs of images. There are three cross-correlation values between the three images, all of which should have values of the same polarity if all the images are of the same polarity. If an image has a reversed polarity with respect to the others, there will be one high value and two relatively low values for the coefficients. To check for this, one image is chosen fixed (selected at random) and the polarity of the other two is selected to maximise the sum of the dot-product cross-correlations.

C. Fusing algorithm

For image fusion, the K-mean clustering is iterated and the validity measure is calculated at each iteration. The ideal number of clusters is chosen based on the validity measure [4]. The maximum number of clusters is arbitrarily limited to 10; the image is not expected to have a high number of distinct target and non-target regions. The algorithm starts by separating the sample points into two clusters. The two initial cluster centers are calculated such that the centers are well separated and also well within each cluster set.

$$\overline{x_i} = \frac{1}{T}\sum x_i, \qquad i = 1,2,\ldots,F$$

$$c_1(1) = (\overline{x_1} - a_1, \overline{x_2} - a_2 \ldots, \overline{x_F} - a_F)$$

$$c_2(1) = (\overline{x_1} + a_1, \overline{x_1} + a_2 \ldots, \overline{x_F} + a_F)$$

where T is the number of samples, F is the number of features and a_1, a_2,…, a_F are the offsets needed for each feature.

The offset for feature j is determined by taking into account the maximum, max_j, and minimum values, min_j, for that feature. Therefore a_j will be half of the smaller of $(\overline{x_j} - \min_j)$ and $(\max_j - \overline{x_j})$. Once the clustering is completed, the validity measure is computed. The clustering procedure is repeated by increasing one cluster at a time until the maximum number of clusters is reached. At each step, the extra cluster is obtained by splitting the cluster with the maximum variance. Two new cluster centers are determined by the offset method described above, except that the sample mean is replaced by the cluster center whose cluster is to be split and the range of offset is now $(c_j - \min_j)$ and $(\max_j - c_j)$ where min_j and max_j are the minimum and maximum values of each feature occurring in that cluster. This ensures a good choice of cluster centers. The variance, σ_j, of cluster, C_j, is determined by

$$\sigma_j^2 = \frac{1}{F}\sum_{i=1}^{F}\frac{1}{M_j}\sum_{x \in C_j}(x - c_{ji})^2,$$

$$i = 1,2,\ldots,F \;\&\; j = 1,2,\ldots,N$$

Once all clustering has been done, the optimal number of clusters is determined by selecting the one with the minimum

validity [4]. Finally, the fused image is obtained by multiplying each individual feature with a factor determined by the clustering algorithm and then summing the feature values together. The factor for a particular pixel is computed from the ratio of the magnitude of the cluster center to which the pixel belongs and the maximum magnitude of the cluster centers.

For three imaging algorithms, the three values for a particular channel form a feature vector. It is anticipated that a better result will be obtained by sorting the values within a feature vector before clustering. This is illustrated by the following example. Let feature vector A be (0.5,-0.1,0.2) and feature vector B be (-0.1,0.5,0.2). If the feature vectors are not sorted, feature A and B will most probably be clustered into two different categories although the fused results for A and B are the same. This is because the value orders for each feature are different. If sorting is done, feature vector A and B become equivalent and they will be grouped in the same category.

D. Performance analysis of image fusion

Figure 8 shows the increase in performance using clustering compared to no clustering before fusion of three images for frames 150-350 of our test dataset. The plot shows that image quality is improved for all frames with an average improvement of ~2 dB.

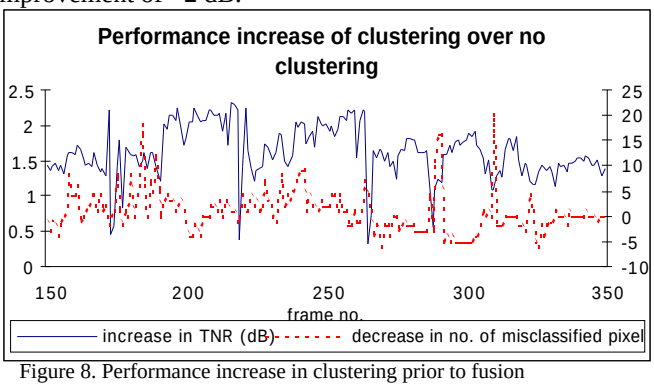


Figure 8. Performance increase in clustering prior to fusion

Figure 9 shows that the performance of sorting before clustering varies considerably, ranging from –0.7 to 1.8 dB, but averaging only about 0.2 dB, indicating that sorting provides a marginal improvement if applied before clustering

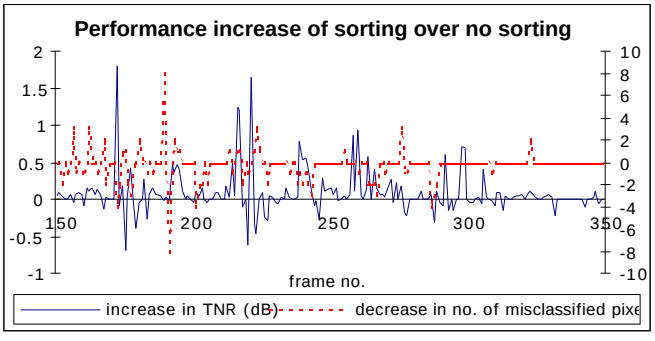


Figure 9. Performance increase due to sorting prior to clustering

VI. CONCLUSION

Two new enhancements to spatial cross-correlation imaging have been developed that yield better image quality. The first enhancement iterates the algorithm with different channel seeds. Testing on a restricted sample of data indicates that the image quality can be improved by as much as 10 dB. The second enhancement uses a linearised optimisation algorithm to maximise the separation between target and non-target sets. The improvement attained by the second enhancement is only moderate, with the best result being 2 dB. Performance can be further improved by combining enhancements I and II, resulting in a further 7.5 dB gain compared to enhancement I alone, a very significant improvement.

A fusing algorithm has also been proposed to combine images from different algorithms. Clustering is performed on the pixel values from each algorithm, regarded as an input feature. Images are combined by multiplying pixel values by their respective factors determined by the clustering algorithm before averaging. The factor for a pixel is determined by the distance to the cluster center to which it belongs. Moderately positive results are obtained, with an average improvement of 2 dB compared to a theoretical maximum incoherent gain of 4.8 dB. Slightly better results are obtained if sorting within the feature vectors is applied prior to clustering.

VII. ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of DSTA and SRN & the provision of ADONIS data by SIO.

REFERENCES

[1] M J. Bukingham, B. V. Berkhout and S. A. L. Glegg, ‘Imaging the ocean with ambient noise, ‘Nature (London) 356, 327-329 (1992)

[2] Michael J.Buckingham, John R.Potter and Chad L.Epifanio, ‘Seeing Underwater with Background Noise’, Scientific American, Feb, 1996, vol.274 no.2

[3] John R. Potter and Mandar Chitre, ‘Ambient noise imaging in warm shallow seas; second-order moment and model based imaging algorithms’, JASA 106(6), Dec 1999, pp 3201-3210

[4] Siddheswar Ray and Rose H. Turi, ‘Determination of number of clusters in K-means clustering and application in colour image segmentation’, (invited paper) in N R Pal, A K De and J Das (eds), Proceedings of the 4th International Conference on Advances in Pattern Recognition and Digital Techniques (ICAPRDT'99), Calcutta, India, 27-29 December, 1999, Narosa Publishing House, New Delhi, India, ISBN: 81-7319-347-9, pp 137-143.

[5] A.W.C.Liew, S.H.Leung, W.H.Lau, ‘Fuzzy image clustering incorporating spatial continuity’, IEE Proc. Vis. Image Signal Processing, Vol. 147, No.2 April 2000, pp 185-192

[6] J. H. Wang and C. Y. Peng, ‘Optimal clustering using neural network’, Proc. IEEE International Conf. on Systems, Man, and Cyber., pp.1625-1630, San Diego, USA, 1998.