IMAGE-QUALITY PREDICTION OF SYNTHETIC APERTURE SONAR IMAGERY

David P. Williams

NATO Undersea Research Centre
Viale San Bartolomeo 400, 19126 La Spezia (SP), Italy

ABSTRACT

This work exploits several machine-learning techniques to address the problem of image-quality prediction of synthetic aperture sonar (SAS) imagery. The objective is to predict the correlation of sonar ping-returns as a function of range from the sonar by using measurements of sonar-platform motion and estimates of environmental characteristics. The environmental characteristics are estimated by effectively performing unsupervised seabed segmentation, which entails extracting wavelet-based features, performing spectral clustering, and learning a variational Bayesian Gaussian mixture model. The motion measurements and environmental features are then used to learn a Gaussian process regression model so that ping correlations can be predicted. To handle issues related to the large size of the data set considered, sparse methods and an out-of-sample extension for spectral clustering are also exploited. The approach is demonstrated on an enormous data set of real SAS images collected in the Baltic Sea.

***Index Terms*—** Image-Quality Prediction, Synthetic Aperture Sonar (SAS), Gaussian Process Regression, Spectral Clustering, Sparse Methods, Out-of-Sample Extension, Variational Bayesian Gaussian Mixture Models, Unsupervised Seabed Segmentation, Large Data Sets.

1. INTRODUCTION

Synthetic aperture sonar (SAS) provides high-resolution imaging of underwater environments by coherently summing sonar ping-returns. The correlation between successive pings at a given range (*i.e.,* distance from the sonar) provides a measure of the success of the SAS processing, as this quantity is directly proportional to the signal-to-noise ratio (SNR) [1].

Processing sonar returns into SAS imagery is a computationally intensive procedure. The ability to do this processing in near real-time onboard a sonar-equipped autonomous underwater vehicle (AUV) is a considerable challenge. Therefore, when tasked to survey a large area of seabed, an AUV will follow a pre-programmed route, with the SAS image-processing done post-mission. The route specified for the AUV assumes that a sufficiently high level of image quality will be attainable up to a certain range.

However, the ability to successfully reconstruct SAS images — and in turn, the achievable area coverage — depends on the motion of the sonar-equipped platform and on the environmental characteristics of the area (such as the seabed-type and the water properties). Excessive platform motion will limit the range for which the coherent summation of pings is possible. Similarly, if the seabed is soft (*e.g.,* muddy), less sonar energy will return to the receivers, also limiting the range to which a SAS image can be reconstructed successfully.

The objective of this work is to predict the SAS image-quality — in terms of the correlation of sonar pings — as a function of range from the sonar. The goal is to make this prediction — without performing the computationally-intensive SAS processing — by using measurements of platform motion and estimates of environmental characteristics. If the range to which the resulting SAS imagery will be of sufficient quality can be predicted, the route of an AUV can be adapted in-mission to both maximize its coverage rate and prevent missing areas of coverage. Moreover, the study can be used to understand the conditions for which SAS processing fails.

To achieve the stated goal, environmental characteristics are estimated by effectively performing unsupervised seabed segmentation, which entails extracting wavelet-based features [2], performing spectral clustering [3, 4], and learning a variational Bayesian Gaussian mixture model [5]. Motion measurements collected onboard the platform and environmental features are then used to learn a Gaussian process regression model [6] so that predictions of the ping correlation can be made. To handle issues related to the large size of the data set considered, sparse methods [7] and an out-of-sample extension [8] for spectral clustering are also exploited. The approach is demonstrated on an enormous measured data set of real SAS images spanning a total area of approximately 44 km^2 in the Baltic Sea.

The remainder of the paper is organized in the following manner. Sec. 2 briefly reviews the aspects of SAS processing relevant for the problem under study, while Sec. 3 describes the process by which features are extracted. Sec. 4 shows experimental results obtained on a data set of real SAS images. Concluding remarks and directions for future work are given in Sec. 5.

2. SYNTHETIC APERTURE SONAR (SAS)

The high-resolution imaging of underwater environments provided by synthetic aperture sonar (SAS) can be used for applications such as mine detection, seabed classification, and the laying of gas or oil pipelines.

A SAS system transmits a broad-band signal such that each location on the seafloor is insonified by multiple pings. The ping returns are recorded onboard the sonar-equipped platform, such as an autonomous underwater vehicle (AUV). In order to reconstruct a SAS image, the returns are coherently summed by accounting for the time delay of signals at different ranges from the platform. The correlation between successive pings (which is directly proportional to the SNR [1]) at a given range can be computed, with this quantity providing a measure of the success of the SAS processing.

The displaced phase-center antenna (DPCA) method [1] is a popular data-driven approach used to reconstruct SAS imagery. The algorithm reconstructs a SAS image from N_i^p collected pings by block-processing in the range direction at N_i^r adaptively-determined DPCA range centers and windows. (For the data set considered in this work, the mean number of DPCA ranges per image was

12.2416.) A byproduct of the processing is the correlation of each pair of consecutive sonar pings at each DPCA range.

For the i-th SAS image, there will be N_i^p correlation values at each of the N_i^r DPCA ranges. To obtain a more robust summary measure of the correlation, we compute the mean correlation value (over the N_i^p correlation values) at each of the N_i^r ranges. It is these N_i^r mean correlation values of the i-th SAS image that we wish to predict in this work.

3. FEATURES FOR PREDICTING PING CORRELATION

The ability to successfully perform SAS processing depends not only on the motion of the sonar platform (*e.g.,* AUV), but also on the environmental characteristics of the seabed and the properties of the water through which the signals propagate [9].

In this work, platform motion measurements and seabed-type estimates are used as features to predict the ping correlation as a function of range.

3.1. Motion Features

Platform motion is recorded onboard the vehicle via an inertial navigation system (INS), and hence readily accessible. The $d_m = 5$ motion measurements used in this study are the roll, pitch, and yaw of the vehicle (*i.e.,* rotations of the vehicle about three orthogonal axes), and the speed of the vehicle in the longitudinal and transverse directions.

3.2. Environmental Features

The environmental features used in this study are based on an unsupervised seabed segmentation algorithm, and are related to the proportion of seabed area that belongs to each of k different seabed-types at each range. The manner in which these features are extracted is described below.

In this work, the "atomic" unit for seabed segmentation is assumed to be a 2 m $\times$ 2 m area of seabed. That is, each 2 m $\times$ 2 m area of seabed corresponds to one data point. This particular size was chosen as a compromise among several factors. The larger the area chosen, the more likely that a single data point will have the unfavorable property of containing multiple types of seabed. However, if the area is too small, the distinguishing characteristics of the seabed that indicate a certain seabed-type may be lost.

The proposed unsupervised seabed segmentation algorithm consists of three main steps. First, as in [2], a vector of $d_w = 16$ features derived from the coefficients of a wavelet decomposition are extracted for each 2 m $\times$ 2 m area of seabed.

Spectral clustering [3, 4] is then applied, which transforms the data into a new, lower-dimensional space via an eigendecomposition. In this work, modifications to the standard spectral clustering approach are taken to address certain aspects encountered in the task at hand. Among these enhancements are the "sparsification" of key matrices to handle very large data sets [7], an automatic self-tuning of a parameter in the requisite affinity matrix [10], an out-of-sample extension to embed data points not present when the eigendecomposition is performed [8], and a decision to distinguish between the number of eigenvectors retained and the number of clusters desired. We integrate these various extensions into the standard spectral clustering algorithm of [4] in this work. Space constraints preclude showing the details of the algorithm here.

The last step of spectral clustering is to cluster the normalized (in-sample) embedded data, into k clusters via a clustering algorithm. In this work, a variational Bayesian Gaussian mixture model (GMM) referred to as *component splitting* [5] is used as the clustering algorithm. This method is employed because it can, in a single run, learn a GMM and determine the appropriate number of mixture components that are represented in the data.

Seabed segmentation is effected in this step by assigning each data point to the mixture component that maximizes its posterior probability. Environmental features are then extracted from the result of the seabed segmentation as follows.

Let the number of 2 m $\times$ 2 m seabed blocks in the i-th SAS image that are within the j-th DPCA range window be $N_b^{(i,j)}$. Suppose each seabed block has been assigned to a component of the k-component GMM. Let $N_b^{(i,j,k)}$ be the number of seabed blocks within the j-th DPCA range window of the i-th SAS image that was assigned to the k-th GMM component. The k-th environmental feature for the j-th DPCA range window of the i-th SAS image is then defined to be the fraction $\boldsymbol{x}^e_{(i,j)}(k) = N_b^{(i,j,k)}/N_b^{(i,j)}$.

3.3. Discussion

Spectral clustering is well-suited for the task of seabed segmentation because an area of seabed corresponding to a particular seabed-type (*e.g.,* sand ripples) often undergoes a gradual change in appearance in SAS imagery. Spectral clustering will recognize that such a "chain" of data points should belong to the same cluster [11].

In the standard formulation of spectral clustering, rigorous theoretical underpinnings justify that the number of eigenvectors retained, m, is set equal to the number of clusters, k to be found [4,11]. However, we choose to go against this convention for our particular application. As noted in [10], when dealing with noisy data, the "ideal" block-diagonal affinity matrix is not attained. As a result, using the eigen-gap [4] to determine the number of clusters in the data is unreliable, because the progression of the eigenvalues may not exhibit a distinct jump in magnitudes.

For our application, the number of clusters (*i.e.,* seabed types) is not known *a priori*. Therefore, rather than adopting a questionable approach with spectral clustering to determine the number of clusters, we transfer this burden onto the GMM. This decision makes sense for several reasons.

By doing so, we can choose to retain a very small number of eigenvectors (*e.g.,* $m = 2$), which accelerates the eigendecomposition computation. Choosing the number of clusters based on the eigen-gap would instead require a complete eigendecomposition (or at least many more eigenvectors to be computed), which is computationally-intensive for the very large data set considered in this work.

By embedding the data into such a low-dimensional space via spectral clustering, the complexity (*i.e.,* number of parameters to estimate) of the subsequent GMM is also greatly reduced, which simplifies learning. In addition, the variational Bayesian GMM approach adopted in this work naturally determines the number of mixture components in a principled way. Moreover, the algorithm can ascertain this number in a single run, which is again valuable with such a large data set.

4. EXPERIMENTAL RESULTS

4.1. Data Set

In April-May 2008, the NATO Undersea Research Centre (NURC) conducted the Colossus II sea trial in the Baltic Sea off the coast of Latvia. During this trial, high-resolution sonar data was collected by

the MUSCLE autonomous underwater vehicle (AUV). This AUV is equipped with a 300 kHz sonar with a 60 kHz bandwidth that can achieve an along-track image resolution of approximately 3 cm and an across-track image resolution of approximately 2.5 cm. The sonar data was subsequently processed into synthetic aperture sonar (SAS) imagery.

The data set consists of 8,097 SAS images. Each image covers 50 m in the along-track direction. The vast majority of images cover 110 m in the across-track (*i.e.,* range) direction, typically from 40 m to 150 m away from the AUV. A small number of the images cover ranges as close as 20 m or as far away as 200 m. The total area spanned by the entire data set of images is approximately 44 km^2.

In this work, the $N_s = 8,097$ SAS images contain a total of $N_b = 10,984,025$ 2 m $\times$ 2 m seabed blocks. There are $N_r = 99,120$ image-range pairs for which we possess the (mean) ping correlation value (*i.e.,* the quantity we wish to predict using motion and environmental features).

4.2. Training Stage

For the experiments conducted here, $N_s^{tr} = 200$ of the images were randomly selected to be used as training images, with the remaining $N_s^{ts} = 7,897$ images treated as testing images. Among the training images, there were a total of $N_b^{tr} = 271,250$ seabed blocks and $N_r^{tr} = 2,437$ image-range pairs.

For each of the N_r^{tr} training data points, motion and environmental features were extracted as described in Sec. 3. Specifically, for each of the N_b^{tr} seabed blocks, wavelet features were extracted and spectral clustering (with $m = 2$ eigenvectors retained) was applied. The variational Bayesian component splitting method was then used to learn the requisite GMM, which resulted in $k = 14$ mixture components. For each seabed block, assignment to the GMM component for which the posterior probability was a maximum was then made, which effected a segmentation of the seabeds of the training images. The $k = 14$ environmental features were then calculated for each of the N_r^{tr} image-range pairs.

Three different GP regression models [6] were then learned so that the ping correlation value could be predicted as a function of range and other input features. In all three models, the target output was the mean correlation value of each "data point" (*i.e.,* image-range pair). The three models differed in the input features that were employed. All three models included the scalar range as a feature, while the first model also included the vector of d_m motion features, the second model also included the vector of k environmental features, and the third model included both motion and environmental features.

4.3. Testing Stage

For the experiments conducted here, the remaining $N_s^{ts} = 7,897$ images were used as testing images. Among these testing images, there were a total of $N_b^{ts} = 10,712,775$ seabed blocks and $N_r^{ts} = 96,683$ image-range pairs.

For each of the N_b^{ts} seabed blocks, wavelet features were extracted, the out-of-sample extension for spectral clustering was applied, and assignment to the GMM component for which its posterior probability was a maximum was made; this last step effected a segmentation of the seabeds of the testing images. The $k = 14$ environmental features were then calculated for each of the N_r^{ts} image-range pairs. Finally, the correlation value for each of the N_r^{ts} image-range pairs was predicted using each of the three learned GP regression models.

4.4. Results

In all results reported here, the term "error" refers to the absolute error of the ping correlation values.

To compare the three GP regression models, the proportion of the $N_r^{ts} = 96,683$ image-range pairs (*i.e.,* "data points") for which each model obtained a lower error than the other models was computed. Because a set of image-range pairs are associated with each image, one can also compute the mean error (averaged across the image-range pairs belonging to an image) for each of the $N_s^{ts} = 7,897$ testing images. This method of assessment is more realistic than comparing individual image-range pairs because in practice, one wishes to make predictions for all ranges of an image jointly, not just at one particular range. Prediction results in terms of the proportion of image-range pairs and also the proportion of images for which each model obtained a lower error than the other models are presented in Table 1 and Table 2, respectively.

Table 1. Proportion of **image-range pairs** for which each model obtained a lower prediction error than each competing model.

	COMPETING MODEL		
MODEL	M	E	M+E
MOTION (M)	———	0.4268	0.4015
ENVIRONMENT (E)	**0.5732**	———	0.4860
M+E	**0.5985**	**0.5140**	———

Table 2. Proportion of **images** for which each model obtained a lower mean prediction error than each competing model.

	COMPETING MODEL		
MODEL	M	E	M+E
MOTION (M)	———	0.3586	0.3109
ENVIRONMENT (E)	**0.6414**	———	0.4541
M+E	**0.6891**	**0.5459**	———

A main motivation for conducting this study was to predict the maximum range to which a SAS image could be successfully reconstructed to a sufficiently high level of quality. Therefore, the performance of the three GP regression models as a function of range is shown in Fig. 1. (Each range bin spans 10 m, with the exception of the last bin, which runs from 150 m to 200 m since there were considerably fewer data points in that range of ranges.) The mean error (averaged across the image-range pairs that fell within each given range bin) is shown for each of the three GP regression models.

As can be seen from Tables 1 and 2 and Fig. 1, the models that included the environmental features consistently perform the best. Of particular interest to our study, however, is prediction performance at long ranges. In this regime, the model incorporating both motion features and environmental features was most accurate.

The results presented above are for a very large data set and are only mean results. To more closely examine performance at an individual image level, we show the GP regression correlation predictions for an example SAS image, along with the actual SAS image, in Fig. 2. (Space constraints preclude showing more example images and results here.)

The SAS image in Fig. 2(a) is notable because the near-range contains relatively flat seabed, while the far-range contains seabed structure (such as rocks and ridges). The presence of seabed structure counteracts the natural correlation degradation that typically

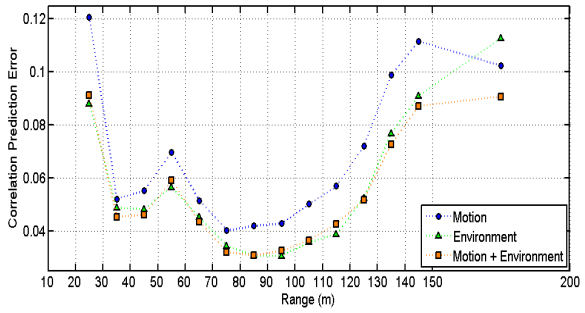
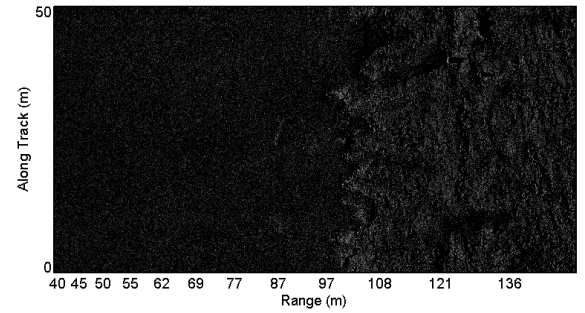
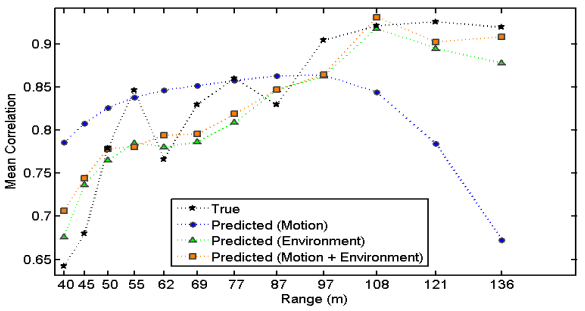


Fig. 1. Mean prediction error as a function of range bin for each GP regression model.



(a)



(b)

Fig. 2. (a) A SAS image and (b) its associated GP regression correlation predictions.

occurs at longer ranges, thereby permitting high correlation values from being obtained. The GP regression models that account for environmental features are able to exploit this information and accurately predict the relatively high correlation values at the longer ranges.

5. CONCLUSION

This work was an application paper that exploited several machine-learning techniques to address the problem of synthetic aperture sonar (SAS) image-quality prediction. The problem is important because if the range to which the resulting SAS imagery will be of sufficient quality can be predicted, the route of an AUV can be adapted to both maximize its coverage rate and prevent missing areas of coverage. Moreover, the study can be used to better understand the conditions for which SAS processing fails.

The experimental results suggest that prediction of the SAS image-quality in terms of the correlation values via motion measurements and environmental characteristics is indeed feasible. The results also demonstrated that using both motion features and environmental features together achieved the best prediction performance at the long ranges of particular interest.

Future work will seek to incorporate other environmental measurements (*e.g.,* wind speed, water temperature) as additional features in the regression model.

Often, before collecting any sonar data, one will possess rough *a priori* knowledge of the seabed-types over the area of interest, in the form of a seabed map. However, for the data set considered in this work, such a seabed map was not available. Therefore, to assess the feasibility of the image-quality prediction task, environmental features were extracted from the seabed segmentation that was performed on the data. Admittedly, this approach should produce the optimal performance, since the "*a priori*" seabed knowledge is essentially assumed to be perfect. Future work will examine correlation-prediction performance when the environmental features are based on only *a priori* seabed knowledge possessed before data collection commences (rather than on the segmentation of the actual imagery that would not be available).

6. REFERENCES

[1] A. Bellettini and M. Pinto, "Theoretical accuracy of synthetic aperture sonar micronavigation using a displaced phase-center antenna," *IEEE J. Oceanic Engineering*, vol. 27, no. 4, pp. 780–789, 2002.

[2] D. Williams, "Unsupervised seabed segmentation of synthetic aperture sonar imagery via wavelet features and spectral clustering," in *Proc. IEEE Int. Conf. Image Processing*, 2009.

[3] M. Meilă and J. Shi, "Learning segmentation by random walks," in *NIPS*. MIT Press, 2000, pp. 873–879.

[4] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *NIPS*. MIT Press, 2001, pp. 849–856.

[5] C. Constantinopoulos and A. Likas, "Unsupervised learning of Gaussian mixtures based on variational component splitting," *IEEE Trans. Neural Networks*, vol. 18, no. 3, pp. 745–755, 2007.

[6] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[7] Y. Song, W. Chen, H. Bai, C. Lin, and E. Chang, "Parallel spectral clustering," in *Proc. ECML/PKDD*, 2008.

[8] Y. Bengio, J. Paiement, P. Vincent, O. Delalleau, N. Le Roux, and M. Ouimet, "Out-of-sample extensions for LLE, isomap, MDS, eigenmaps, and spectral clustering," in *NIPS*. MIT Press, 2004, pp. 177–184.

[9] M. Hayes and P. Gough, "Broad-band synthetic aperture sonar," *IEEE J. Oceanic Engineering*, vol. 17, no. 1, pp. 80–94, 1992.

[10] L. Zelnik-Manor and P. Perona, "Self-tuning spectral clustering," in *NIPS*. MIT Press, 2004, pp. 1601–1608.

[11] U. von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007.