The Fusion of Large Scale Classified Side-Scan Sonar Image Mosaics

Scott Reed, Ioseba Tena Ruiz, *Member, IEEE*, Chris Capus, *Member, IEEE*, and Yvan Petillot, *Member, IEEE*

***Abstract*—This paper presents a unified framework for the creation of classified maps of the seafloor from sonar imagery. Significant challenges in photometric correction, classification, navigation and registration, and image fusion are addressed. The techniques described are directly applicable to a range of remote sensing problems. Recent advances in side-scan data correction are incorporated to compensate for the sonar beam pattern and motion of the acquisition platform. The corrected images are segmented using pixel-based textural features and standard classifiers. In parallel, the navigation of the sonar device is processed using Kalman filtering techniques. A simultaneous localization and mapping framework is adopted to improve the navigation accuracy and produce georeferenced mosaics of the segmented side-scan data. These are fused within a Markovian framework and two fusion models are presented. The first uses a voting scheme regularized by an isotropic Markov random field and is applicable when the reliability of each information source is unknown. The Markov model is also used to inpaint regions where no final classification decision can be reached using pixel level fusion. The second model formally introduces the reliability of each information source into a probabilistic model. Evaluation of the two models using both synthetic images and real data from a large scale survey shows significant quantitative and qualitative improvement using the fusion approach.**

***Index Terms*—Classification, fusion, Markov random fields, mosaicing, registration, side-scan sonar (SSS), simultaneous localization and mapping (SLAM).**

I. Introduction

RECENT advances in the fields of underwater technologies and robotics have led to the development of Autonomous Underwater Vehicles (AUVs) and stable platforms [1]–[3], fitted with high resolution sonars, and capable of swimming close to the seabed have opened up the oceans to rapid and high resolution mapping, generating large volumes of data. These data have many practical scientific uses including reef management [4], oyster management [5], trawling impact assessment [6] and mine-counter measures [7]. This paper tackles the difficult problem of generating large scale accurate maps of the seabed from sonar imagery obtained using these platforms.

A large body of work on image based classification of sonar data exists, but the methods used have been based on single sonar images. The generation of large scale maps raises the problem of image registration, the difficult related problem of autonomous navigation in an environment deprived of global positioning systems (GPS) and the generation of fused maps from multiple overlapping classified images which may contain contradictory information.

A. Sensors

In water, acoustics provides the main sensing modality. Electromagnetic waves attenuate rapidly and their operating range is limited to well below 5 m in normal operating conditions. Sound navigation and ranging (sonar) offers a good alternative providing the user with accurate resolution and long ranges of hundreds of meters. For imaging, side-scan sonar (SSS) and the emerging synthetic aperture sonar (SAS) provide very high resolution images of up to centimetric accuracy at up to 300 m. These systems use the principle of a long antenna to generate a narrow acoustic beam [8]–[10] which illuminates a narrow stripe of the seabed at any one time. As the system moves through the water, towed by a survey vessel or mounted on an AUV, it generates a wide-area image of the sea bed (obtained as a concatenation of successive stripes) as shown in Fig. 2. The main parameters affecting the resolution of the generated images are the length of the antenna and frequency of the acoustic wave used (these determine the across track resolution) and the speed of the platform (AUV or towed body) which determines the along track resolution.

B. Image Formation Process and Pre-Processing

The interpretation of side-scan imagery is a skilled procedure [8], [9]. There are many parameters of the image formation process contributing to intensity variations in recorded data which are quite separate from the influences of variations in seabed properties and textures. These are generally well understood and good models exist which can aid the process of automated seabed classification [11], [12]. In this paper, the sonar data are preprocessed to correct for the influences of the sonar beam pattern and time-varying gain (TVG) [13], [14]. This enables the use of simpler and faster classification algorithms which is particularly beneficial for application to large area surveys.

C. Classification

Given the vast quantities of data produced, fast classification algorithms are required to produce seabed class maps representing textural variations and areas of scientific interest. Many supervised techniques have been developed to tackle this problem. Neural networks and parametric statistical classifiers have dominated the scene [15]–[20] and have been coupled with feature extraction measures including one-dimensional

Manuscript received March 31, 2005; revised July 14, 2005. This work was supported in part by the 5th Framework Program of Research of the European Community under Project AMASON (EVK3-CT-2001-00059). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Robert P. Loce.

The authors are with the Ocean Systems Laboratory, School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, U.K. (e-mail: scott.reed@seebyte.com; i.tena_ruiz@hw.ac.uk; c.capus@hw.ac.uk; y.r.petillot@hw.ac.uk).

Digital Object Identifier 10.1109/TIP.2006.873448

cepstral and spectral features [21]–[23], fractal analysis [24], spatial point processes [25], gray level run-length measures [18], [26], and co-occurrence matrices [27], [28]. Fuzzy logic analysis [29] and autoregressive models [30] have also been investigated for seabed classification.

These image-based techniques consider each image extracted from the side-scan data in isolation. In this paper, a variant of the power spectrum feature set [22] is used with a simple parametric classifier for rapid supervised classification of the side-scan data. However, unlike previous techniques, image segmentation is improved by fusing classification results for each image with the results derived from all of the images covering the same area of seabed.

D. Registration and Mosaicing

In order to produce large scale classified maps of the seabed, it is first necessary to register the individual SSS images. Given the position of the sensor in the world for each side-scan beam, it is possible to produce a geo-referenced image of the seabed. This process is called mosaicing [31], [32]. The quality of the mosaics produced will ultimately depend on the precision of the position information of each sonar beam, i.e., the vehicle's navigation precision. Underwater, navigation is a difficult problem as conventional GPS receivers do not operate. Underwater platforms therefore rely on dead-reckoning sensors which drift over time. To fix this drift, we use a simultaneous localization and mapping (SLAM) technique based on the stochastic map developed for indoor robotics [33], [34] and adapted to sonar imagery [32], [35]. This is a solution, based on the use of landmarks detected in the sonar imagery to help the navigation, which does not require additional sensors and does not interfere with the data acquisition constraints (stable platform, fixed altitude).

E. Fusion

During a typical survey of the seabed, multiple views of the same area are normally collected from different view points. The fusion of these views enables the generation of improved large scale classified mosaics from the individual classified sonar images. To date, very little work has been done in image fusion in the underwater domain [36], [37] and it has been limited to simple multisensor fusion. To the authors' knowledge, there have been no publications concerning the fusion of underwater imagery to produce large scale classified mosaics of the seafloor.

Fusion of multiple sources of information is a well established research field. When the information sources produce the same type of measurements, standard fusion techniques such as Bayesian theory [38], fuzzy logic [39] and Dempster–Shafer theory [40]–[42] can be used. However, for classifiers which consider different types of input measurements or features, it is often not possible to consider the computed output measurements to be estimates of the same posterior probability [43]. This makes fusion more difficult. When little is known about the information sources, or when they produce information at a high level of abstraction, voting schemes can be successfully used [26], [44] and this is particularly appealing for underwater imaging systems which currently favor "black box" approaches to classification.

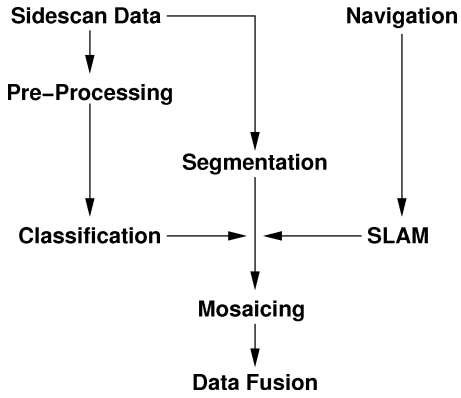


Fig. 1. Data flow for formation of fully classified fused side-scan mosaics.

Fusion for classification within the image domain allows contextual information to be considered and to date has been mainly applied to remote sensing [38], [39], [45]. The fusion of multiple images is generally performed at the pixel level [42], [46]–[49] allowing information from the surrounding area to be considered when classifying each pixel. An effective method for incorporating this spatial information is the use of Markov random fields (MRFs) [50]–[52].

This paper details a fusion model for registered, classified SSS images of the seafloor. Fusion is conducted at the pixel level where each classifier outputs a single class decision for each pixel. This maintains the generality of the fusion scheme and allows classified images from different classifiers to be fused together.

Two separate models are presented. The first uses a voting scheme to initialize the fused class map and a MRF model to incorporate contextual information, smooth the final result and "inpaint" regions of pixels which are *unclassified* following the voting process. This ensures all pixels within the image, for which there are data, are successfully classified within the final fused result. The standard Markovian prior term is altered so that pixels labeled as *unclassified* from the voting scheme do not contribute to the Markovian probability. This ensures that only pixels which have been successfully fused and classified contribute to the fusion process for the other pixels. The second model considers the case when the reliability of the classifier is known [53]. This information is stored using confusion matrices and allows the voting scheme from the first model to be replaced by a probabilistic, Markovian framework.

Fig. 1 illustrates the progression from the raw sensor and navigation data through to the fully classified fused side-scan image mosaics resulting from the processing steps outlined in this paper.

F. Layout

The image formation mosaicing and pre-processing of the SSS data are discussed in Section II. Supervised classification using features derived from the sonar swath power spectra is described in Section III. Section IV describes the SLAM-RTS technique for improving and smoothing the navigation solution. Section V details the two models presented for fusion of the classified SSS images. Results are presented on simulated and MeasTex sample data to illustrate the differences between

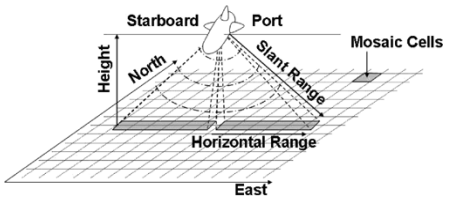


Fig. 2. Diagram showing the assumed SSS geometry.

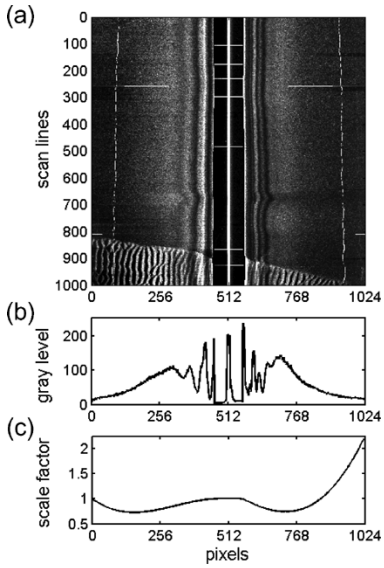


Fig. 3. (a) Sample raw image, 1000 scan lines; (b) beam pattern estimate; and (c) residual TVG estimate.

the two models and quantitatively measure the effectiveness of these approaches. Section VI contains the results of the Fusion model on real, classified SSS mosaics. Section VII concludes the paper.

II. Side-Scan Sonar Image Formation and Data Preprocessing

A. Image Formation

The following results were obtained by processing data gathered during the BP'02 experiments carried out by the SACLANT Undersea Research Centre in La Spezia, Italy. The system used a Marine Sonics SSS operating at 600 kHz on board a REMUS AUV [3]. The sensor covered a swath of 60 m in slant range. The survey was conducted in shallow coastal waters with an average altitude of around 3.45 m.

As noted above, the SSS produces a narrow acoustic beam along the vehicle track, and a wide beam across track. As the sensor is moved forward, ideally in a straight line, the beam generated covers the seabed uniformly. The geometry of the side-scan image formation process is represented in Fig. 2. A sonar image is generated by concatenating these beams, or scan lines, into a 2-D image as seen in Fig. 3(a). It is important to understand that such images are not in Cartesian coordinates but in time coordinates. The vertical axis corresponds to the time at which the beam was emitted from the sonar while the horizontal axis corresponds to the time of flight of the pulse in the across track direction. Converting these time-time images to images in real-world coordinates is called geo-referencing or mosaicing

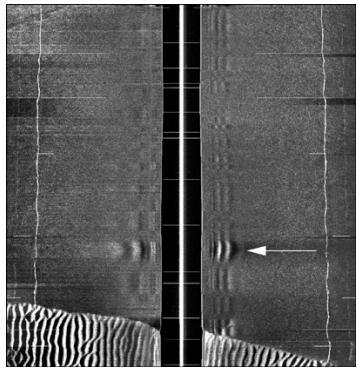


Fig. 4. Corrected image. The correction algorithm cannot compensate for changes in sensor attitude, such as roll on trajectory corrections; see arrow above.

and is critical to our applications. A common feature in all raw side-scan imagery is the largely black region in the center of the image. This corresponds to the transit time of the acoustic wave through the water before reflection from the seabed and indicates the altitude of the sensor. More details on the image formation process can be found in [9].

B. Preprocessing

In many emerging applications, low altitude surveys are required. In such cases, even quite small changes in vehicle altitude can affect the sonar image dramatically. Prior to classification the image data used here have been preprocessed using an advanced radiosity correction algorithm [13]. This is useful because it treats purely range-dependent artefacts, such as residual TVG effects separately to angular effects such as the influence of the sonar beam pattern. Separate correction factors are calculated for each. Whilst this gives better performance than standard radiosity correction algorithms in the presence of sensor altitude changes, platform stability is still assumed with respect to pitch and roll.

A sample raw image and the estimated beam pattern and residual TVG profile for these data are shown in Fig. 3. The complexity of the beam pattern is apparent with four significant lobes in the port channel and as many as six in the starboard channel. The differences in the scales of the y-axes for the beam profiles and TVG estimates result from their methods of calculation and application within the radiosity correction algorithm [13].

The corrected image is shown in Fig. 4. In some places, the beam pattern correction has failed, as indicated by the white arrow. This arises from the behavior of the vehicle, which rolls on turns. With each course adjustment, the small degree of roll affects the symmetry of the beam pattern on the seabed, so that it is poorly compensated near the water column. At these points, classification accuracy is affected. However, where there are overlapping images, the data fusion techniques described below compensate well for any resulting misclassification.

C. Mosaicing

The sonar mosaic algorithm used in this paper assumes the geometry shown in Fig. 2. Under this assumption each sonar channel (port and starboard) insonifies a rectangular area on the

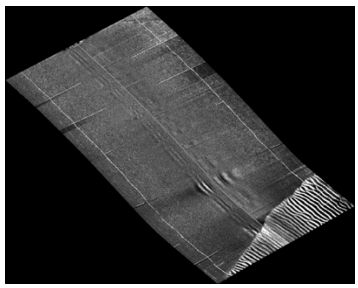


Fig. 5. Sample side-scan mosaic.

sea floor. The length of the rectangle is determined by the slant range of the sonar (the maximum range of the sonar) and the height of the vehicle, assuming a flat seabed. The equation for this length, referred to as horizontal range r_h is

$$r_h = \sqrt{r_s^2 - h^2} \tag{1}$$

where the slant range r_s is obtained from

$$r_s = T \times \frac{c}{2} \tag{2}$$

with c the speed of sound in the water, assumed constant, and T the time of the last return on each beam. The height of the vehicle h is found using

$$h = t_b \times \frac{c}{2} \tag{3}$$

where t_b is the time to the first return of the sonar beam. The width of the beam is determined by the spacing between two successive beams, this guarantees that all the mosaic cells between beams are filled.

The seabed is represented as a flat two-dimensional grid of mosaic cells (see Fig. 2). The resolution of the mosaic cells can be set by the operator. Using simple geometry both channels can be superimposed if the orientation and position of the sonar is known. Each mosaic cell will take the value of the intensity cell on the beam closest to it. In this paper if a mosaic cell takes more than one value then, in the case of the pre-processed image mosaics, the data will be averaged or, in the case of the classified data, the cell will be left as unclassified.

An example output from the mosaicing algorithm is given in Fig. 5 which shows a mosaic obtained from geo-referencing the data from Fig. 4.

III. Classification of Side-Scan Data

Three seafloor textures have been identified for segmentation of the side-scan imagery, defining three classes: flat sediments, sand ripples, and complex regions. In this context, classification refers to the process of assigning each image pixel to one of these texture classes. Segmentation refers to the process of producing a new image in which each pixel is remapped to a new gray level determined by the assigned class. Suppression of the beam pattern effects and some of the residual TVG effects, as described in Section II-B, improves the images to the point where a fast supervised classification scheme can be combined with a relatively simple, small and easily generated feature set.

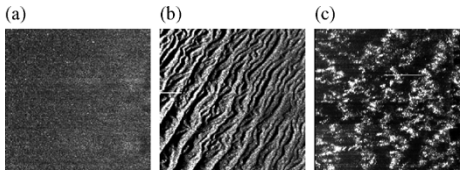


Fig. 6. Training data. Three small images extracted from the full dataset. Classes are (a) flat sediment; (b) sand ripples; and (c) complex texture.

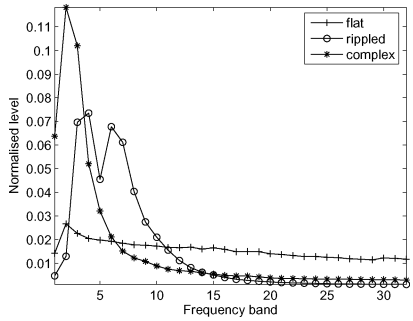


Fig. 7. Averaged normalized power spectral densities for the three training sets. Three features were defined with band limits at 1–4; 4–12; 16–32. The rippled texture is characterized by two dominant peaks corresponding to large and small scale ripples which dominate this region.

The features used are derived from Pace and Gao's frequency based sediment classification scheme [22]. In the current case features are calculated directly from the image intensity data after the suppression of any DC component. The classes correspond to relatively large scale image textures, which confines much of the discriminatory content to relatively low frequency bands. Overlapping 64-sample Gaussian windowed FFTs are used to generate the one-dimensional power spectra and this allows for identification of changes in texture across the sonar swath.

If $s_j(t)$ represents a single line of sonar data and $W_i(t)$ represents the Gaussian window centered at position i, the normalized power spectrum can be defined by

$$\begin{aligned} P_{i,j}(f) &= \left|F\left[W_i(t)s_j(t)\right]\right|^2 \\ \tilde{P}_i(f) &= \frac{1}{n}\sum_{j=1}^{n} P_{i,j}(f) \\ P_{iN}(f) &= \frac{\tilde{P}_i(f)}{\int_0^{f_{\max}} \tilde{P}_i(f)df}. \end{aligned} \tag{4}$$

Spatial frequency bands within the normalized power spectra are identified which give a good separation between the classes. For these large sonar data sets, emphasis has been on processing speed for the classification phase and the training set used for these data comprises three small exemplar images of 200×200 pixels, one for each texture. Running on a standard 3.2-GHz Pentium IV computer under Windows XP, the processing time for the entire dataset, is around 90 min. The acquisition time for these data, covering some 720 000 m^2, was nearly 3 h. The training images extracted from the full dataset are shown in Fig. 6.

The averaged normalized spectra for the three training samples are presented in Fig. 7. The bimodal nature of the spectrum

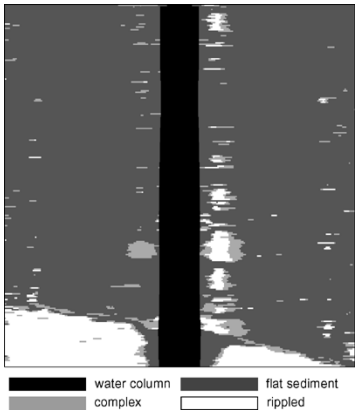


Fig. 8. Classmap generated from the image introduced in Fig. 3. Water column—black; flat sediment—dark grey; complex—light grey; rippled—white. Errors are noted where the correction algorithm has failed due to vehicle attitude changes during course corrections and in the transition regions between textures.

for the sand ripples class occurs because there are two dominant scales for the sand ripples. The training image used has been chosen to be representative of both the large and small scale ripple textures. It is also possible with this data set and classification scheme to separate these two ripple textures and define a four-class training set.

Note that supervised classification of the side-scan data is based on "spatial" frequencies within the sonar image. In common with many textural features, these frequencies will be affected by many factors including sensor attitude and altitude, sensor type, operating frequencies, and range settings. Retraining would be required if any of these were changed significantly.

Returning to the three-class problem, three features are defined using the averaged normalized spectra derived from the training data. These give the proportion of the spectrum lying in sample bands 1–4, 4–12, and 16–32 and each identifies a spectral region in which one of the three classes is dominant; see Fig. 7

$$D_{f1} = \frac{\int_1^4 P_{iN}(f)}{\int_1^{32} P_{iN}(f)}$$
$$D_{f2} = \frac{\int_4^{12} P_{iN}(f)}{\int_1^{32} P_{iN}(f)}$$
$$D_{f3} = \frac{\int_{16}^{32} P_{iN}(f)}{\int_1^{32} P_{iN}(f)}. \quad (5)$$

In classifying a complete sonar image, the three features are generated from the averaged normalized spectral density formed from four successive lines of data. The same 64-sample sliding Gaussian windowed FFT is used and boundary problems between sonar channels are minimized by closing up the water column. This is done simply by shifting the scan lines on the assumption that there will generally be continuity in seabed textures between sonar channels.

Fig. 8 shows a segmentation result for the image introduced in Fig. 3. Misclassifications are greatest near the water column where the correction algorithm has failed. There are some boundary errors, with pixels classified as complex texture in the transition region between flat sediment and sand ripples. Misclassification further from the water column is due primarily to incomplete elimination of the influence of the surface return and crosstalk from other sensors. These effects are particularly prominent in this data set and add to the difficulty of the classification task. These misclassifications can frequently be rectified by the proposed fusion scheme.

IV. Simultaneous Localisation and Mapping (SLAM)

In order to create an accurate mosaic of the classified maps, good navigation is crucial. Conventional GPS and differential GPS (DGPS) receivers do not work underwater. Therefore, when submerged, the side-scan sensor must be localized using dead-reckoning with depth sensors, Doppler velocity logs (DVL), inertial navigation systems (INS), and/or compasses [54]–[56]. To correct drift on the dead-reckoning, the sonar must either be equipped with a GPS/DGPS system and surface intermittently to get a new fix, thus perturbing the data acquisition process, or it must be equipped with acoustic receivers capable of triangulating the position with respect to either acoustic beacons on the seabed, known as long base line (LBL) navigation, or to acoustic beacons on a support vessel, known as short base line (SBL), or ultrashort base line (USBL) [57]. The costs associated with high performance INS systems and with setting up acoustic nets or mobilizing a vessel are considerable and new techniques have been sought to localize positions underwater. Terrain matching methods will use known maps of the environment and data from payload sensors to find the dead-reckoning drift [58], [59]. The purpose of SLAM is to build a map of the environment and use that same map to localize [33], [34]. Recently, SLAM techniques have been developed to work with a SSS [32], [35]. This paper uses this method in order to geo-reference classified side-scan images. Previous work has demonstrated the potential of this method when fusing nonclassified data using Gabor wavelets [60].

The data from the navigation sensors are fused in order to localize the SSS. The technique used to fuse the navigation data in this paper is the stochastic map smoothed using a Rauch–Tung–Striebel (RTS) fixed-interval smoother. It will be referred to as SLAM-RTS. The stochastic map keeps the estimates of the position and creates a map of landmarks to represent the environment. These landmarks are then used to aid localization of the vehicle. It is a SLAM method that works iteratively to provide an estimate of the position at the latest iteration. In order to improve the accuracy of the solution and to smooth it, post-processing is required. The next two sections provide a detailed look at the algorithms.

A. Stochastic Map

The stochastic map is an augmented state extended Kalman filter (EKF) [61], [62]. It adds new states to the state vector to accommodate new landmarks as they are observed [63]. A typical stochastic map state vector is of the form

$$\mathbf{x} = [\mathbf{x_v} \mathbf{x_1} \ldots \mathbf{x_n}]' \quad (6)$$

where $'$ is the transpose of a vector or matrix, $\mathbf{x_v}$ holds the state of the SSS and $\mathbf{x_1}, \ldots, \mathbf{x_n}$ hold the state of the n landmarks in the map.

The stochastic map also stores and maintains all the covariances and correlations between the states. Furthermore, it has been proved in [64] that, in the limit as the number of observations increases, the covariance associated with any single target location estimate is determined only by the initial covariance in the vehicle location estimate and, in the limit, all the target estimates become fully correlated. These properties make the stochastic map highly desirable. With fully correlated landmarks, an observation of any of the landmarks will help correct the whole map. It can also take advantage of the wealth of literature published on Kalman filters. The update equations of the stochastic map are the familiar EKF update equations. To propagate the state

$$\hat{\mathbf{x}}_\mathbf{v}(k) = \mathbf{f_v}\left[\hat{\mathbf{x}}_\mathbf{v}(k-1), \mathbf{u}(k), \mathbf{0}, k\right] \tag{7}$$

where $\hat{\mathbf{x}}_\mathbf{v}(\cdot)$ is the SSS's estimated state and $\mathbf{f_v}[\hat{\mathbf{x}}_\mathbf{v}(\cdot), \mathbf{u}(k), \mathbf{0}, k]$ is its dynamic model, and its associated covariance will be propagated, thus

$$\mathbf{P}(k) = \mathbf{F_{x_v}}\mathbf{P}(k-1)\mathbf{F}^\mathbf{T}_\mathbf{x_v} + \mathbf{F_{w_v}}\mathbf{Q}(k)\mathbf{F}^\mathbf{T}_\mathbf{w_v} \tag{8}$$

where $\mathbf{F_{x_v}}$ is the Jacobian of the dynamic model with respect to the SSS state, used to linearise the state of the SSS error $\tilde{\mathbf{x}}_\mathbf{v}(k-1)$, and $\mathbf{F_{w_v}}$ is the Jacobian of the dynamic model with respect to the process noise. The corrected state estimate becomes

$$\hat{\mathbf{x}}(k+1) = \hat{\mathbf{x}}(k) + \mathbf{K_i}(k)\upsilon_\mathbf{i}(k) \tag{9}$$

where $\mathbf{K_i}$ is the gain of the filter and $\upsilon_\mathbf{i}$ is the innovation. Its associated covariance is updated according to

$$\mathbf{P}(k+1) = \mathbf{P}(k) - \mathbf{K_i}(k)\mathbf{S_i}(k)\mathbf{K}^\mathbf{T}_\mathbf{i}(k) \tag{10}$$

where $\mathbf{S_i}(k)$ is the innovation covariance.

For more details on this implementation of the stochastic map, the interested reader should refer to [35].

B. SLAM-RTS

The Kalman filter and EKF use all measurements up to the last iteration to estimate the state at that last iteration. The RTS smoother uses all measurements *before* and *after* each iteration to estimate the state at each iteration [65]. It is a post-processing filter that works on the stored outputs of a Kalman filter by re-processing it. The smoother works by combining a forward pass Kalman filter with a backward pass filter. It was originally designed to work with fixed size state vectors. However, the stochastic map adds new states to the state vector as it observes new landmarks. The SLAM-RTS algorithm adapts the RTS fixed-interval smoother to work with the stochastic map by fixing the size of the state vector to the size of the stochastic map on the last iteration. The SLAM-RTS algorithm ensures numerical stability in matrix operations by adjusting the estimates of the landmarks' states and covariances at all iterations before they have been observed to the values when they are first observed. The output of the SLAM-RTS has been shown to improve the accuracy of the stochastic map solution [35], as well as providing trajectories more suitable for creating and superimposing mosaics [32], [60].

V. Fusion of Mosaiced Classification Data

This section presents two pixel level models for the fusion of multiple classified SSS mosaics. The first model fuses classified maps using a simple voting scheme while the second model integrates class reliability indexes in the fusion scheme. Both models are formulated within a multisource Markovian framework to take advantage of contextual information and improve classification accuracy. The Markovian model is presented first as it is common to both approaches. The details specific to each of the two fusion models are presented later.

A. Markov Model for Image Fusion

Let us assume first that each of the input class maps is defined on a lattice S where label s specifies a specific pixel location. Two random fields $\mathbf{X}$ and $\mathbf{W}$ are defined. $\mathbf{X} = \{X_s, s \in S\}$ describes the classification field provided by each input map and $\mathbf{W} = \{W_s, s \in S\}$ describes the final fused classification map. For K input class maps, $X_s = (X_s^1, \ldots, X_s^K)$ takes its values from the finite set of classes $\Omega = \{\omega_1, \ldots, \omega_M, \alpha, \gamma\}$. The set Ω contains M recognized seafloor classes, the *unclassified* label α and the *unmeasured* label γ. Label α is allocated to X_s^j, $j \in \{1, \ldots, K\}$ when data is received regarding pixel s in image j but a classification based on the data provided is not possible. Label γ, *unmeasured*, is used when no data is received regarding pixel s, ensuring it is not possible to provide a classification X_s^j.

The distinction between classes α and γ is important during the fusion process. If all the input class maps provide input $X_s^j = \gamma, \forall j$, then there has been no data retrieved by the sensor regarding pixel s. Therefore, pixel s should not be classified by the fusion process. If all the input class maps provide input $X_s^j = \alpha, \forall j$ then while none of the input images have provided a classification for pixel s, the sensor has received data regarding this region of seafloor. In this instance, the Fusion model will return an output classification for pixel s by considering the pixel classifications within the neighboring region.

The fusion problem consists of estimating the true classified map $W = w$ from the individual classified maps $X = x$ where $x = x_s$, $s \in S$ are K classified maps of the same scene. The field $W = \{W_s, s \in S\}$ is said to be Markovian with respect to neighborhood $\eta = \{\eta_s, s \in S\}$ if its distribution can be written as

$$P_{W_s}(W_s = w_s | W_r = w_r, r \neq s) = P(W_s = w_s | W_r = w_r, r \in \eta_s). \tag{11}$$

This formalizes that $P(W)$ is a local probability and that the fused class label w_s for pixel s is dependant only on the class labels of the pixels within its neighborhood η_s. For simplicity, the fusion model described in this paper assumes a second order isotropic neighborhood. This neighborhood η_s for pixel s can be seen in Fig. 9. Further reading regarding MRF models can be found in [51], [52], [66], [67].

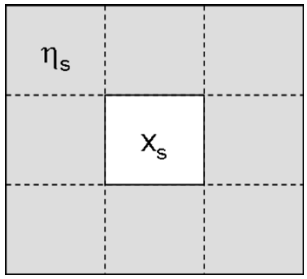


Fig. 9. Markovian prior assumed a second order isotropic neighborhood. The class of pixel s is therefore only dependant on its 8 nearest neighboring pixels.

The Markovian property field W allows Prior Probability $P(W)$ to be written in the form

$$P_W(w) = \frac{1}{Z}\exp\left(-E(w)\right). \tag{12}$$

Where Z is a normalizing constant and $E(w)$ is an energy term.

For the Fusion model, the problem of maximizing probability $P_W(w)$ can be re-cast to the local problem of maximizing energy

$$U(w_s) = \sum_{t\in\eta_s} \beta\delta(w_s, w_t)\left[1-\delta(w_t,\alpha)\right]\left[1-\delta(w_t,\gamma)\right] \tag{13}$$

for pixel s. In (13), $\delta(.)$ is the Kronecker Delta symbol and β controls the importance of the Markovian prior. For all cases in this paper, $\beta = 1.0$. As can be seen from (13), neighboring pixels labeled as *unclassified* or *unmeasured* (and therefore do not contain any useful information to aid the fusion process) do not contribute to the Markovian prior. The minimization of $P(W)$ is performed using the Iterated Conditional Modes method [52]. In this method, a raster scan is used to iteratively visit all the pixels in field W. If $w_s = \gamma$ *unmeasured*, the pixel is not considered further and the pixel remains *unmeasured*. Otherwise, w_s is allocated to the class which locally maximizes $U(w_s)$. This method of segmentation produces a local maximum of $P(W)$. The ICM procedure is iterated until there are no pixel changes within a full image scan. The details of how the energy term described in (13) is applied within each of the two Fusion models is provided in the following sections.

B. Voting/MRF Fusion Model

The voting/MRF fusion model assumes that each image j provides a classification result for each pixel label x_s^j. The fusion field W is initialized by using an adaptation of the generalized majority voting [44]. In this model, a summed binary function $T_s(\omega_i)$ for pixel s, and each recognized seafloor classes ω_i, $1 \le i \le M$ is specified as

$$T_s(\omega_i) = \sum_{j=1}^{K} \delta\left(x_s^j, \omega_i\right) \;\; for\, 1 \le i \le M \tag{14}$$

where the sum is over all the inputted class images and as before, $\delta(.)$ is the Kronecker Delta function. This function is not specified for the *unclassified* or *unmeasured* classes.

The initial fusion Field W is then specified as

$$\begin{aligned} w_s &= \gamma \; if \sum_{j=1}^{K} \delta\left(x_s^j, \gamma\right) = K \\ &= \tau \; if \; T_s(\tau) = \max T_s(\omega_i) \ge \frac{2}{3}K^s \\ &= \alpha, \;\; \text{otherwise} \end{aligned}$$

where K^s is the number of images which do not provide an *unclassified* or *unmeasured* classification for pixel s.

In this model, the adapted voting function first checks to see whether all the images provide an *unmeasured* classification. If they do, the pixel is classified as *unmeasured* within the Fusion Map. For w_s to be allocated a class which is not *unmeasured* or *unclassified*, class $\tau \in \{\omega_1, \ldots, \omega_M\}$ must have the largest summed binary function with a value greater than or equal to $(2/3)K^s$. If this rule is not met, the pixel is labeled as *unclassified*.

Once the voting rule specified in (15) has been used to initialize fusion field W, the Markov energy term $U(w_s)$ in (13) is considered. The ICM technique described in Section V-A is used to locally maximize $U(w_s)$ and complete the Fusion process.

Using the above fusion approach ensures several points.

- Pixels allocated as *unmeasured* in all the images are left so in the fused image as no measurements have been obtained regarding these pixels. They are not allocated a valid seafloor class.
- Pixels initialized as *unclassified* in fusion field W as a result of the voting scheme are always allocated a seafloor class during the ICM process. In effect, the *unclassified* regions are eroded away by the Markovian prior (inpainting). The Markov model allows these regions to be classified by considering the surrounding classified regions.
- The ICM process produces a smoothed version of the initialization produced by the Voting process.

C. Probabilistic Fusion Model

The voting scheme used to initialize the fusion class map in the voting/MRF model described in Section V-B assumes that each input source is equally reliable. A more balanced fusion decision should also consider the reliability of each source. One possible representation of source reliability is to use confusion matrices.

Assume each source $j \in \{1, \ldots, K\}$ produces an input class map X^j as well as a Class Confusion matrix C^j. This matrix is obtainable through training (using a supervised system) and provides a measure of the likelihood term $P(x_s|w_s)$ where it is assumed that random variable X_s is independently conditional on W [66]. This assumption is used extensively in image-based Markov solutions. Using training data, these matrices can be estimated, specifying the probability of source j providing a $x = \tau$ classification decision given that it is known that the actual classification of the decision is $w_s = \epsilon$. For pixel s and input source j

$$P\left(x_s^j = \tau | w_s^j = \epsilon\right) = C_{\epsilon\tau}^j. \tag{15}$$

We assume that the Confusion Matrix C considers only recognized seafloor classes $\omega_1,\ldots,\omega_M$. Assuming the input sources are independent (we assume the random variables $X_s^1,\ldots,X_s^K$ are independent conditionally on W [48]), we can write

$$P\left(x_s^1=\tau,\ldots,x^K=\sigma|w_s=\epsilon\right)=C_{\epsilon\tau}^1\times\ldots\times C_{\epsilon\tau}^K. \quad (16)$$

Using the likelihood term expressed in (16), it is possible to initialize the Fusion Field W. Instead of initializing the field using a voting scheme where each source is considered equally reliable, the model now considers each inputs reliability. As before, if all the sources provide a $x_s=\gamma$ or α decision, w_s is initialized as *unmeasured* or *unclassified* respectively. Otherwise, the label field W is initialized by considering (16). Ten samples are drawn from the class $w_s=\epsilon$ probability distribution which maximizes the likelihood value described in (16). Pixel s is initialized with class w_s if the sampled probability compares favorably to a random number in the range [0,1] for a majority of the samples. If this criterion is not met, the pixel is labeled as *unclassified*. This initialization technique is analogous to the generalized majority technique used in the voting/MRF fusion model.

After initialization, an iterative process is again carried out to complete the fusion. Pixels are visited randomly and allocated a classification which minimizes posterior energy

$$U^{\text{post}}\left(w_s,x_s^1,\ldots,x_s^K\right)=z_s\left[1-\delta(w_s,\alpha)\right]+\beta\sum_{t\in\eta_s}\delta(w_s,w_t)\left[1-\delta(w_t,\gamma)\right]\left[1-\delta(w_t,\alpha)\right]$$

whereas, in the voting/MRF model, pixels initialized as $w_s=\gamma$ are not considered or changed. As before, $\delta(.)$ is the Kronecker Delta function and

$$z_s=\ln\left[P\left(x_s^1,\ldots,x_s^K|w_s\right)\right]. \quad (17)$$

As can be seen from (17), the likelihood term is considered only if the pixel is not currently labeled as *unclassified*. The Markovian term is the same as the one discussed in (13) for the Voting/MRF model. Pixels are visited randomly rather than using the more deterministic raster scan to allow more mixing between the classes. The number of pixel visitations will affect the quality of the final result. The results provided here used $4N$ pixel visitations where N is total number of pixels in field W.

The probabilistic fusion process is completed by applying energy term $U(w_s)$ in (13) to field W in an iterative manner, again using a raster scan. This was done to again ensure the final Fusion Classification contained no *unclassified* pixels. Unlike the raster scan approach, randomly visiting the pixels does not ensure that all initially *unclassified* pixels are classified in the final result. The final raster scan considering only $U(w_s)$ ensures that any remaining *unclassified* pixels are given a recognized seafloor classification in the final, fused map.

D. Results

1) Comparison of the Models on MeasTex Images: This section evaluates the Voting/MRF and Probabilistic fusion model on an image comprising of MeasTex [68] texture samples. Classification results have been obtained using five different classifier combinations. The first is a parametric linear discriminant classifier using co-occurrence matrix features (C1). The second and third consider fractal features using a nonparametric kNN classifier (C2) and a parametric linear discriminant classifier (C3), respectively. The fourth and fifth classifiers consider frequency based features, again with the kNN classifier (C4) and the linear discriminant classifier (C5), respectively. All are supervised systems and the confusion matrices required for the Probabilistic model can be obtained from the training data. A sample test image and the corresponding ground truth image can be seen in Fig. 10.

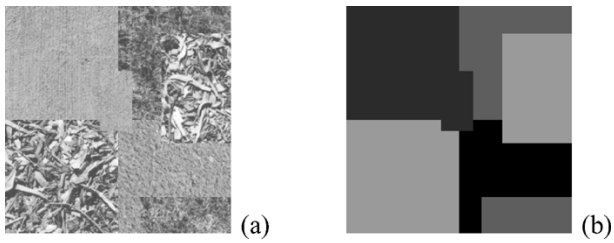


Fig. 10. (a) Texture image composed of 4 MeasTex textures. (b) Ground truth of the textured image displayed in (a).

TABLE I
CLASSIFICATION ACCURACY OF THE 5 CLASSIFIERS USED TO CLASSIFY THE MEASTEX TEXTURE IMAGE SHOWN IN FIG. 10

Classifier	% Classification Accuracy
C1	88.65
C2	75.90
C3	74.91
C4	48.66
C5	49.94

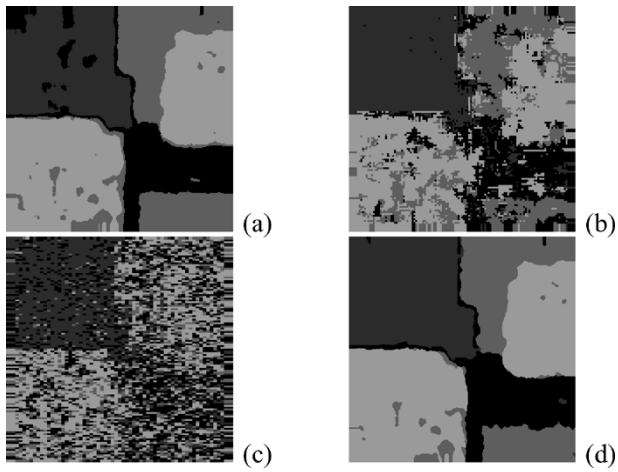


Fig. 11. (a) Classification result from (a) classifier C1, (b) classifier C2, (c) classifier C3, and (d) fusing C1, C2, and C4 using the probabilistic model.

The performance accuracies for each of the five classifiers operating in isolation are given in Table I.

The classification results for three of the classifiers (C1, C2 and C4) are shown in Fig. 11 (C3 and C5 are visually quite similar to C2 and C4 respectively). The figure also contains the classification result obtained from fusing the C1, C2, and C4

TABLE II
CLASSIFICATION ACCURACY OBTAINED FROM FUSING VARIOUS CLASSIFIERS USING BOTH THE VOTING/MRF AND PROBABILISTIC MODELS

Classifiers Fused	Voting/MRF Accuracy	Probabilistic Accuracy
C1 C2 C3 C4	87.05	92.13
C2 C3 C4 C5	81.87	82.66
C1 C2 C4	89.08	92.72
C1 C2 C3	85.58	91.72
C1 C3	89.51	92.09
C2 C3	80.86	82.47
C3 C4	79.40	82.90
C4 C5	62.94	63.50

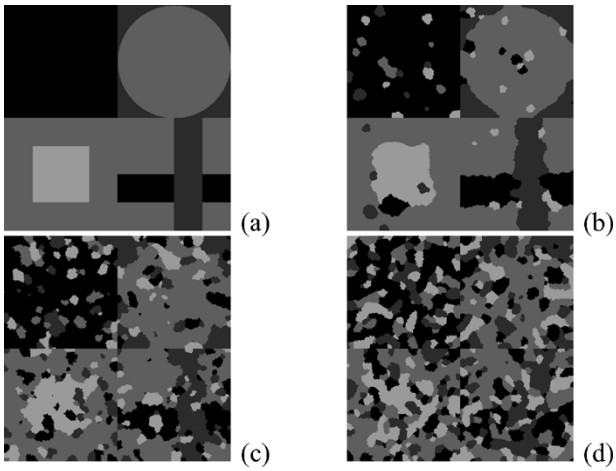


Fig. 12. (a) Ground truth of the simulated image used to test the fusion models. (b)–(d) Simulated classification results from a 90%, 70%, and 50% accurate classifier, respectively.

classifier results using the Probabilistic model. This has been included for comparison purposes.

The classification accuracy obtained from fusing the different classifiers using both the voting/MRF and the Probabilistic model can be seen in Table II.

Table II shows the Probabilistic model outperforming the MRF/Voting model in all cases as expected. The fusion operations containing C2 and C3 together can be seen producing lower classification results than is perhaps expected from considering two good classifiers. This is because C2 and C3 use the same features and so generally classify and misclassify the same image regions. Any misclassifications present in either of these results are simply reinforced by the second. This lends strength to the argument that fusion is most effective when considering results obtained from different sources. This can be seen in the high classification accuracy obtained from fusing C1, C2 and C4.

2) Comparison of the Models on Synthetic Images: This section evaluates the Voting/MRF and Probabilistic fusion models on simulated examples. A simulator model was used to produce classification results of a given classification accuracy, using the same ground truth image. Fig. 12 contains a ground truth image used in this as well as a simulated result from hypothetical 90%, 70% and 50% efficient classifiers.

TABLE III
CLASSIFICATION ACCURACY OBTAINED FROM FUSING FOUR SIMULATED CLASSIFICATION RESULTS OF A GIVEN CLASSIFICATION ACCURACY USING BOTH THE VOTING/MRF AND PROBABILISTIC MODELS

Input Image Accuracy %	Voting/MRF Accuracy %	Probabilistic Accuracy %
100	99.73	99.75
90	96.70	96.97
80	93.51	94.20
70	91.92	91.94
60	84.21	85.32
50	75.11	76.14

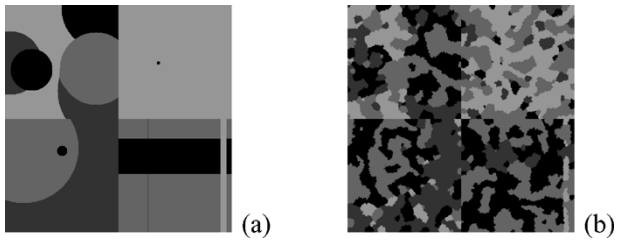


Fig. 13. (a) Ground truth of a simulated class map. (b) The classification result produced from a simulated source.

Simulated classifier output for the ground truth image shown in Fig. 12 was produced for a range of classifier accuracies. Four different classified images for each tested input classification accuracy were produced. These four images were fused using the Voting/MRF and Probabilistic models. The simulator was set up to ensure that any misclassified pixels were spread evenly over the other classes. The results obtained for the two models, fusing four images for each level of classification accuracy, can be seen in Table III.

The table shows that both fusion models generally produce a classification result which is better than any of the input images considered in isolation (the exception is the perfect input classifiers where the MRF component of the models has smoothed the final result). The Probabilistic model can be seen to outperform the Voting/MRF model in all cases, although both provide good results in all of the examples, even when the input classification accuracy drops to 50%.

The fusion example shown in Table III used input images where the misclassified pixels were spread evenly over the other classes. For classifiers where there is strong confusion between some classes, the difference in performance between the MRF/Voting and Probabilistic models can be significant. One such example is highlighted in Fig. 13. This contains a ground truth image and the resultant classification result from a simulated source described by classification confusion matrix $C_{\epsilon\tau}^{\text{sim}}$ where

$$C_{\epsilon\tau}^{\text{sim}} = \begin{pmatrix} 0.50 & 0.30 & 0.20 & 0.00 \\ 0.20 & 0.60 & 0.10 & 0.10 \\ 0.50 & 0.00 & 0.50 & 0.00 \\ 0.05 & 0.10 & 0.35 & 0.50 \end{pmatrix}. \tag{18}$$

When fusing three different sources described by $C_{\epsilon\tau}^{\text{sim}}$, the MRF/Voting model produces a final result with a classification

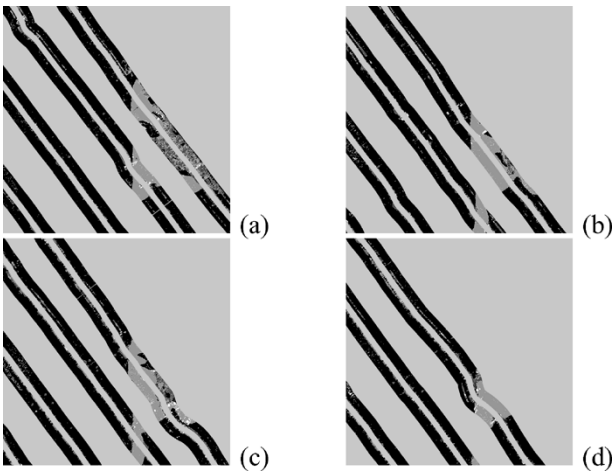


Fig. 14. (a)–(d) Show four mosaics displaying 15 side-scan tracks of the REMUS mission.

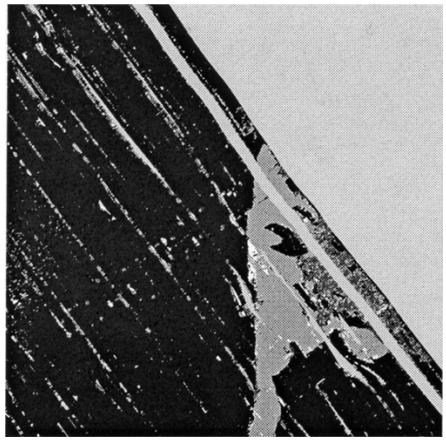


Fig. 15. Initialization of the voting/MRF fusion model after the voting scheme for the sector 2 mosaics. The white pixels describe *unclassified* regions.

Fig. 16. Final fused result obtained from the voting/MRF fusion model for the sector 2 mosaics. The MRF section of the model has ensured no *unclassified* pixels remain.

accuracy of 59.95% while the Probabilistic model achieves an accuracy of 85.23%.

Further analysis of performance using varying numbers of input images and different classification accuracies lies outside the scope of this paper, but the results presented in this section do demonstrate that both of the fusion models will produce a final classification result that is generally higher in accuracy than for any of the input images considered in isolation.

VI. Creating Large Scale Mosaics

We return to the side-scan data first introduced in Section II to show how the fusion approach can improve mapping for a large area survey. The navigation output from REMUS and the side-scan data have already been used to create large scale mosaics of the observed region [69]. The same navigation solution has been used to geo-reference the segmented image data. The complete AUV mission lasted for 2 h, 57 min, and 8 s and followed a set of parallel, regularly spaced and overlapping linear tracks, typical for rapid environmental assessment surveys.

A. Fusion

Since no ground truth information was available on the data set with which to obtain classifier reliability information, the fusion of the classified mosaics was carried out using the Voting/MRF model.

Fig. 14 shows four overlapping mosaics of sector 1 (a predefined region) created by geo-referencing 17 linear tracks. All of the resulting mosaics created using overlapping tracks are geo-referenced to the same reference frame (sector 1) and will constitute an input to the fusion algorithm.

The mosaics in Fig. 14 contain a maximum of five classes. Each pixel is considered to belong to the sand, ripple, complex, *unmeasured* or *unclassified* class. The large light gray regions are *unmeasured* regions over which the AUV has not passed. The white regions are areas which remain *unclassified* following the geo-referencing process. The geo-referencing process sometimes stretches or contracts pixels from the individual input images to allow mosaics of the correct resolution to be created. The initialization from the voting scheme and the final fused result for these mosaics can be seen in Figs. 15 and 16, respectively.

Both the initialization and the final fused results in Figs. 15 and 16 contain much more information than any of the mosaics considered in isolation. The initialization result contains large regions of *unclassified* data where the voting scheme has failed to confidently allocate a seafloor class. These regions have been classified within the final fused result. The regions of seafloor which are classified as *unmeasured* in all the input mosaics have also been left as *unmeasured*. The final output result has fused all the input mosaics to produce a smoothed map where all the pixels where data of some description has been received by the AUV have been classified. The fused map allows a more complete picture of the seafloor to be built up than is possible by considering a single mosaic in isolation.

Overall, the fusion model has produced a more complete and useful picture of the seafloor. All *unclassified* regions from the initialization result in Fig. 15 have been assigned classes by the Markovian aspect of the model. The benefits of the Fusion model can be clearly seen in the example shown in Figs. 16.

VII. Conclusions

An integrated method for creating and fusing classified SSS mosaics has been presented. This utilizes recent advances in side-scan data correction and is applicable to a wide range of classification schemes. The ability to produce good segmentation results from a limited training set is one of the benefits of applying the fusion approach and this can provide additional flexibility in the choice of feature sets and classifiers. Attempts to represent every variation in texture in the training data can lead to "overfitted" solutions. Using the approach described here, robust segmentation is achieved through a fusion scheme which incorporates Markovian spatial priors and information regarding classifier performance with respect to each texture class.

The mosaicing algorithm presented uses SLAM techniques to produce high quality mosaics of the individually segmented images. The generation of these mosaics, where the images are georeferenced in space, provides the basis for multimosaic fusion.

Two models have been introduced for the fusion of the classified mosaics. The first uses voting schemes to initialize the fusion map and a Markov model to both classify regions previously labeled as *unclassified* and smooth the final result. The second model adopts a probabilistic framework and allows the reliability of each source to be considered during the fusion process. The Voting/MRF model has been demonstrated on real classified side-scan mosaics allowing a more accurate picture of the survey region to be built than would have been possible from considering the individual mosaics in isolation.

Acknowledgment

The authors would like to thank the SACLANT Undersea Research Centre, the U.S. Office of Naval Research, and the Woods Hole Oceanographic Institution for allowing the inclusion of data from the BP'02 experiment.

References

[1] J. G. Bellingham, C. A. Goudey, T. R. Consi, J. W. Bales, D. K. Atwood, J. J. Leonard, and C. Chryssostomidis, "A second generation survey AUV," in *Proc. Autonomous Underwater Vehicle Technology*, Jul. 1994, pp. 148–155.

[2] A. Ishoy, "How to make survey instruments "AUV-friendly"," in *Proc. MTS/IEEE Int. Conf. OCEANS*, 2000, pp. 1647–1652.

[3] C. von Alt, B. Allen, T. Austin, N. Forrester, R. Goldsborough, M. Purcell, and R. Stokey, "Hunting for mines with REMUS: A high performance, affordable, free swimming underwater robot," in *Proc. MTS/IEEE Int. Conf. OCEANS*, 2001, pp. 117–122.

[4] K. J. Cuevas, M. V. Buchanan, and D. Moss, "Utilizing side scan sonar as an artificial reef management tool," in *Proc. MTS/IEEE Int. Conf. OCEANS*, 2002, pp. 136–140.

[5] D. Diaz, K. Cuevas, M. V. Buchanan, S. Gordon, and W. S. Perret, "Side scan sonar in oyster management," in *Proc. MTS/IEEE Int. Conf. OCEANS*, 2002, pp. 141–145.

[6] B. G. Gonzalez, Y. Petillot, and C. Smith, "Detection and classification of trawling marks in side scan sonar images," presented at the Advances in Technology for Underwater Vehicles Conf., Mar. 2004.

[7] S. Reed, J. Bell, and Y. Petillot, "An unsupervised approach to the detection and extraction of mine features in side scan sonar," in CAD/CAC Conf., Halifax, NS, Canada, Nov. 2001, to be published.

[8] R. J. Urick, *Principles of Underwater Sound*. New York: McGraw-Hill, 1975.

[9] C. Mazel, "Side Scan Sonar Record Interpretation," Klein Assoc., Inc., 1985.

[10] J. Chatillon, A. E. Adams, M. A. Lawlor, and M. E. Zakharia, "SAMI: a low-frequency prototype for mapping and imaging of the seabed by means of synthetic aperture," *IEEE J. Ocean. Eng.*, vol. 24, no. 1, pp. 4–15, Jan. 1999.

[11] APL-UW, *APL-UW High-Frequency Ocean Environmental Acoustic Models Handbook*. Seattle, CA: Appl. Phys. Lab., Univ. Washington, 1994.

[12] J. M. Bell, "A Model for the Simulation of Sidescan Sonar," Ph.D. dissertation, Heriot-Watt Univ. , Edinburgh, U.K., Sep. 1995.

[13] C. Capus, I. Tena Ruiz, and Y. Petillot, "Compensation for changing beam pattern and residual tvg effects with sonar altitude variation for sidescan mosaicing and classification," presented at the 7th. Eur. Conf. Underwater Acoustics, Delft, The Netherlands, 2004.

[14] S. Anstee, "Removal of Range-Dependent Artifacts From Sidescan Sonar Imagery," DSTO Aeronautical1 and Maritime Res. Lab., Australia, Tech. Rep. DSTO-TN-0354, Apr. 2001.

[15] R. D. Muller, N. C. Overkov, J.-Y. Royer, A. Dutkiewics, and J. B. Keene, "Seabed classification of the south tasminrise from simrad em12 backscatter data using artificial neural networks," *Austral. J. Earth Sci.*, pp. 689–700, 1997.

[16] W. K. Stewart, M. Marra, and M. Jiang, "A hierarchical approach to seafloor classification using neural networks," in *Proc. IEEE Int. OCEANS Conf.*, Oct. 1992, pp. 109–113.

[17] B. Bourgeois and C. Walker, "Sidescan sonar image interpretation with neural networks," in *Proc. IEEE Int. OCEANS Conf.*, Honolulu, HI, 1991, pp. 1687–1694.

[18] W. K. Stewart, M. Jiang, and M. Marra, "A neural network approach to classification of sidescan sonar imagery from a midocean ridge area," *IEEE J. Ocean. Eng.*, vol. 19, no. 2, pp. 214–224, Apr. 1994.

[19] B. Zerr, E. Maillard, and D. Gueriot, "Seafloor classification by neural hybrid system," in *Proc. IEEE Int. OCEANS Conf.*, 1994, pp. 239–243.

[20] D. Alexandrou and D. Pantzartzis, "Seafloor classification with neural networks," in *Proc. IEEE Int. OCEANS Conf.*, Washington, DC, 1990, pp. 18–23.

[21] Z. Reut, N. G. Pace, and M. J. P. Heaton, "Computer classification of seabeds by sonar," *Nature*, pp. 426–428, 1985.

[22] N. G. Pace and H. Gao, "Swathe seabed classification," *IEEE J. Ocean. Eng.*, vol. 13, no. 2, pp. 83–90, Apr. 1988.

[23] D. Tamsett, "Seabed characterization and classification from the power spectra of sidescan sonar data," *Marine Geophys. Res.*, pp. 43–64, 1993.

[24] D. R. Carmichael, L. M. Linnett, S. J. Clarke, and B. R. Calder, "Seabed classification through multifractal analysis of sidescan sonar imagery," *Inst. Elect. Eng. Radar, Sonar, Navig.*, no. 3, pp. 140–148, 1996.

[25] L. M. Linnett, D. R. Carmichael, S. J. Clarke, and A. D. Tress, "Texture analysis of sidescan sonar data," presented at the *IEE Conf. Texture Analysis in Radar and Sonar*, London, U.K., Nov. 1993.

[26] B. R. Dasarathy and E. B. Holder, "Image characterizations based on joint gray-level run-length distributions," *Pattern Recognit. Lett.*, pp. 497–502, 1991.

[27] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textrural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-2, no. 6, pp. 610–621, Nov. 1973.

[28] S. Subramaniam, H. Barad, and A. B. Martinez, "Seafloor characterization using texture," presented at the *IEEE South East Conf.*, New Orleans, LA, Apr. 1993.

[29] M. Mignotte, C. Collet, P. Perez, and P. Bouthemy, "Fuzzy logic modeling in sonar imagery: Application to the classification of underwater floor," *Comput. Vis. Image Understand.*, pp. 4–24, 2000.

[30] D. T. Cobra and H. A. de Moraes, "Classification of sidescan sonar images through parametric modeling," in *Proc. IEEE Int. OCEANS Conf.*, Brest, France, 1994, pp. 109–113.

[31] D. Gueriot, "Bathymetric and side-scan data fusion for sea-bottom 3-D mosaicing," in *Proc. MTS/IEEE Int. Conf. OCEANS*, 2000, pp. 1663–1668.

[32] I. Tena Ruiz, Y. Petillot, and D. M. Lane, "Improved AUV navigation using side-scan sonar," in *Proc. MTS/IEEE Int. Conf. OCEANS*, 2003, pp. 1261–1268.

[33] J. J. Leonard, R. N. Carpenter, and H. J. S. Feder, "Stochastic mapping using forward look sonar," in *Proc. Int. Conf. Field Service Robotics*, Pittsburgh, PA, Aug. 1999, pp. 69–74.

[34] I. Tena Ruiz, Y. Petillot, D. M. Lane, and C. Salson, "Feature extraction and data association for AUV concurent mapping and localization," in *Proc. IEEE Int. Conf. Robotics Automation*, Seoul, Korea, May 2001, pp. 2785–2790.

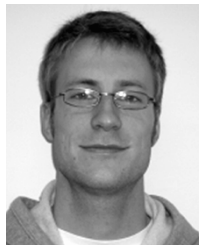
[35] I. Tena Ruiz, S. de Raucourt, Y. Petillot, and D. M. Lane, "Concurrent mapping and localization using side-scan sonar," *IEEE J. Ocean. Eng.*, vol. 29, no. 2, pp. 442–456, Apr. 2004.

[36] H. Singh, C. Roman, L. Whitcomb, and D. Yoerger, "Advances in fusion of high resolution underwater optical and acoustic data," in *Proc. IEEE Int. Symp. Underwater Technology*, May 2000, pp. 206–211.

[37] J. Wright, K. Scott, T.-H. Chao, and B. Lau, "Advances in fusion of high resolution underwater optical and acoustic data," in *Proc. IEEE Symp. Autonomous Underwater Vehicle Technology*, Jun. 1996, pp. 167–175.

[38] A. Solberg, A. Jain, and T. Taxt, "Multisource classification of remotely sensed data: fusion of landsat TM and SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 768–778, Jul. 1994.

[39] I. Bloch, "Information combination operators for data fusion: a comparative review with classification," *IEEE Trans. Syst., Man, Cybern. A*, vol. 26, no. 1, pp. 52–67, Jan. 1996.
[40] ——, "Some aspects of Dempster-Shafer evidence theory for classification of multi-modality medical images taking partial volume into account," *Pattern Recognit. Lett.*, pp. 905–919, 1996.
[41] S. Foucher, M. Germain, J.-M. Boucher, and G. Benie, "Multisource classification using ICM and Dempster-Shafer theory," *IEEE Trans. Instrum. Meas.*, vol. 51, no. 2, pp. 277–281, Apr. 2002.
[42] L. Fouque, A. Appriou, and W. Pieczynski, "An evidential Markovian model for data fusion and unsupervised image classification," in *Proc. 3rd ISIF Int. Conf. Information Fusion*, 2000, pp. 25–32.
[43] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.
[44] L. Xu, A. Krzyzak, and C. Y. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition," *IEEE Trans. Syst., Man, Cybern.*, vol. 22, no. 3, pp. 418–435, May/Jun. 1992.
[45] S. L. Hegarat-Mascle, I. Bloch, and D. Vidal-Madjar, "Application of Dempster-Shafer evidence theory to unsupervised classification in multisource remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 4, pp. 1018–1031, Jul. 1997.
[46] J. V. Cleynenbreugel, S. A. Osinga, F. Fierens, P. Suetens, and A. Oosterlinck, "Road extraction from multi-temporal satellite images by an evidential reasoning approach," *Pattern Recognit. Lett.*, pp. 371–380, Jun. 1991.
[47] L. Fouque, A. Appriou, and W. Pieczynski, "Multiresolution hidden Markov chain model and unsupervised image segmentation," in *Proc. 4th IEEE Southwest Symp. Image Analysis and Interpretation*, 2000, pp. 121–125.
[48] A. Bendjebbour, Y. Delignon, L. Fouque, V. Samsom, and W. Pieczynski, "Muitisensor image segmentation using Dempster-Shafer fusion in Markov fields context," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 8, pp. 1789–1798, Aug. 2001.
[49] S. L. Hegarat-Mascle, A. Quesney, D. Vidal-Madjar, and O. Taconet, "Land cover discrimination from multitemporal ERS images and multispectral landsat images: A study case in an argicultural area in France," *Int. J. Remote Sens.*, no. 3, pp. 435–456, 2000.
[50] W. A. Wright, "Fast image fusion with a Markov random field," in *Proc. IEE Conf. Image Processing and Its Applications*, 1999, pp. 557–561.
[51] D. Geman, S. Geman, C. Graffigne, and P. Dong, "Boundary detection by constrained optimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 609–628, Jul. 1990.
[52] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc.*, no. 3, pp. 261–279, 1986.
[53] J. A. Benediktsson and I. Kanellopoulos, "Classification of multisource and hyperspectral data based on decision fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1367–1377, May 1999.
[54] N. A. Brokloff, "Matrix algorithm for doppler sonar navigation," in *Proc. IEEE Int. Conf. OCEANS*, 1994, pp. III/378–III/383.
[55] R. Cox and S. Wei, "Advances in the state of the art for AUV inertial sensors and navigation systems," *IEEE J. Ocean. Eng.*, vol. 20, no. 4, pp. 361–366, Oct. 1995.
[56] G. Grenon, P. E. An, S. M. Smith, and A. J. Healey, "Enhancement of the inertial navigation system for the Morpheus autonomous underwater vehicles," *IEEE J. Ocean. Eng.*, vol. 26, no. 4, pp. 548–560, Oct. 2001.
[57] K. Vickery, "Acoustic positioning systems, a practical overview of current systems," in *Proc. Autonomous Underwater Vehicle Technology*, Aug. 1998, pp. 5–17.
[58] O. Bergem, "Bathymetric Navigation of Autonomous Underwater Vehicles Using a Multibeam Sonar and a Kalman Filter With Relative Measurement Covariance Matrices," Ph.D. dissertation, Univ. Trondheim, Norway, Dec. 1993.
[59] M. Sistiaga, J. Opderbecke, and M. Aldon, "Depth image matching for underwater vehicle navigation," in *Int. Conf. Image Analysis Processing*, 1999, pp. 624–629.
[60] E. Coiras, I. Tena Ruiz, Y. Petillot, and D. M. Lane, "Fusion of multiple side-scan sonar views," presented at the IEEE Int. Conf. OCEANS, 2004.
[61] P. Maybeck, *Stochastic Models, Estimation and Control*. New York: Academic, 1982, vol. 141–142.
[62] Y. Bar-Shalom and T. Fortmann, *Tracking and Data Association*. New York: Academic, 1988, vol. 179.
[63] R. Smith, M. Self, and P. Cheeseman, "Estimating uncertain spatial relationships in robotics," in *Autonomous Robot Vehicles*, I. Cox and G. Wilfong, Eds. New York: Springer-Verlag, 1990.
[64] M. W. M. G. Dissanayake, H. Durrant-Whyte, S. Clark, and M. Csorba, "A solution to the simultaneous localization and map building (SLAM) problem," Univ. Sydney, Sydney, Australia, Tech. Rep. ACFR-TR-01-99, Jan. 1999.
[65] A. Gelb, Ed., *Applied Optimal Estimation*. Cambridge, MA: MIT Press, 1974.
[66] M. Mignotte, C. Collet, P. Perez, and P. Bouthemy, "Sonar image segmentation using an unsupervised hierarchical MRF model," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1216–1231, Jul. 2000.
[67] H. Derin and H. Elliot, "Modeling and segmentation of noisy and textured images using Gibbs random field," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no. 1, pp. 39–55, Jan. 1987.
[68] MeasTex Image Texture Database and Test Suite, MeasTex. (1997, May). [Online]. Available: http://www.cssip.uq.edu.au/meastex/meastex.html
[69] I. Tena Ruiz, Y. Petillot, D. M. Lane, and A. Cormack, "Large scale side-scan sonar mosaics," presented at the Maritime Reconnaissance for NATO's Recognized Environmental Picture (MREP), La Spezia, Italy, May 2003.



Scott Reed received the M.Phys. Astrophysics (Hons) degree from Edinburgh University, Edinburgh, U.K, in 1999, and the M.Sc. and Ph.D. degrees from the Ocean Systems Laboratory, Heriot-Watt University, Edinburgh, in 2000 and 2004, respectively. His Ph.D. dissertation examined automated detection, classification, and segmentation techniques for use with side-scan sonar imagery.

He is currently with SeeByte, Ltd., Edinburgh, developing data fusion techniques for the U.K. MoD BAUUV project. His research interests include Markov models, statistical image processing, texture analysis, and data fusion.

Ioseba Tena Ruiz (M'98) received the B.Eng (Hons.) degree in electronic engineering and the Ph.D. degree from Heriot-Watt University, Edinburgh, U.K., in 1996 and 2001, respectively.

He was a Research Associate at Heriot-Watt University in five different research projects: CLASS (CMPT), ARAMIS (EU), AUTOTRACKER (EU), ALIVE (EU), and AMASON (EU). His interests include navigation systems, concurrent mapping and localization, processing sonar returns, classifying sonar returns, and obstacle avoidance systems for autonomous underwater-vehicles. He is currently working on simultaneous localization and mapping across multiple platforms.

Chris Capus (M'05) received the M.Sc. degree and the Ph.D. degree in advanced time-frequency signal analysis from Heriot-Watt University, Edinburgh, U.K., in 1997 and 2002, respectively.

He is a Researcher at the Ocean Systems Laboratory, School of Engineering and Physical Sciences, Heriot-Watt University. His current projects include detection and obstacle avoidance for autonomous underwater vehicles, sonar modeling, seafloor classification for environmental assessment, and marine mammal signal analysis and development of novel biomimetic signalling systems. These projects are supported by the U.K. MoD, the European Union, and the Office of Naval Research.

Yvan Petillot (M'02) received the M.Sc. degree in optics and signal processing and an engineering degree in telecommunications, with a specialization in image and signal processing, and the Ph.D. degree in real-time pattern recognition using optical processors from University de Brewage Occidental.

He is currently a Lecturer at Heriot Watt University, Edinburgh, U.K., carrying out research and teaching activities in image processing including texture and shape analysis, segmentation and classification of sonar and video images, simultaneous mapping and localization and navigation. He also has a particular interest in mine detection and classification using image based statistical techniques. He is the Principal Investigator on two EPSRC Research projects (GR/S68088/01 and GR/S16980/01) on weapons detection using mm-wave imaging (222 K) and on multimodal sensor fusion for seabed classification (129 K), respectively. He is also co-investigator of the Framework V AUTOTRACKER project (obstacle avoidance and acoustic pipe tracking using an AUV) and the Framework V Amason project (data fusion of sonar data products for seabed classification and habitat mapping). Each project has a university budget of around 300 K.