



**HAL**  
open science

# Transmission of images by unmanned underwater vehicles

Alice Danckaers, M.L. Seto

► **To cite this version:**

Alice Danckaers, M.L. Seto. Transmission of images by unmanned underwater vehicles. *Autonomous Robots*, 2020, 10.1007/s10514-019-09866-z . hal-02434610

**HAL Id: hal-02434610**

**<https://hal.science/hal-02434610>**

Submitted on 4 Jun 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Transmission of images by unmanned underwater vehicles

Alice Danckaers<sup>1</sup> · Mae L. Seto<sup>2</sup>

## Abstract

As an acoustic communications medium, water is characterized by frequency dependent attenuation, short range, very low bandwidth, scattering, and multi-path. It is generally difficult to acoustically communicate even terse messages underwater much less images. For the naval mine counter-measures mission, there is value in transmitting images of mine-like objects, acquired by side-scan sonar on-board unmanned underwater vehicles, to the above-water operator for review. The contribution of this paper is a methodology and implementation, based on vector quantization, to compress and transmit snippets of side-scan sonar images from underway unmanned underwater vehicles to an operator. The work has been validated through controlled indoor tank tests and several at-sea trials. The fidelity of the received images is such that trained operators can recognize targets in the received images as well as they would have in the original images. Future work investigates machine learning to improve the compression basis and psycho-visual studies for the specialized skill of feature recognition in sonar images.

**Keywords** Underwater communications · Sonar image transmissions · Unmanned underwater vehicle · Naval mine counter-measures

## 1 Introduction

One application that this reported work contributes to is naval mine counter-measures (NMCM) with multiple collaborating unmanned underwater vehicles (UUVs) and unmanned surface vehicles (USVs). A task within the NMCM mission is an underwater survey of an area with side-scan sonars (SSS) integrated on the UUVs. Collaborating UUVs make the survey more efficient due to the increase in coverage rate and provides redundancy in the event an UUV is unable to complete its survey. The USV can also survey if it is in shallow water or more often, it relays UUV findings to above-water operators. The important outcome of the survey is the geo-referenced SSS data (images).

The payload autonomy (automated target detection, ATD) on-board the UUVs perform in-situ processing to detect targets or mine-like objects (MLO) in their geo-referenced sonars' images. This processing yields geo-referenced image snippets (or mugshots) where each snippet (~ hundreds of kB in size) contains a target. These 8-bit/pixel monochrome images are not usually enhanced. The geo-referenced locations are needed to revisit the targets to confirm their locations for subsequent actions (<http://acomms.who.edu/wp-content/uploads/sites/20/2014/09/401040-SIG-Micromodem-Software-Interface-Guide.pdf>). The sonar images are typically evaluated by the operator after the UUV is recovered and the images are downloaded.

Some NMCM operations require an operator to evaluate the target prior to subsequent actions like further classification or disposal. The UUV recovery, sonar data download and UUV re-launch can take a fair amount of time depending on the sea state and the area to survey. Meanwhile the UUV cannot be re-tasked to continue the survey (or revisit a target) and the target information it contains cannot be transmitted to inform the operator or to re-task another UUV or USV to investigate the target. This reduces the value of the UUV collaboration to just increasing the coverage rate instead of engaging other UUVs and USVs to accelerate the NMCM

---

This is one of the several papers published in *Autonomous Robots* comprising Special Issue on Robot Communication Challenges: Real-World Problems, Systems, and Methods.

---

✉ Mae L. Seto  
mae.seto@dal.ca

Alice Danckaers  
alice.danckaers@gmail.com

<sup>1</sup> ENSTA Bretagne, Brest, France

<sup>2</sup> Defence R&D Canada, Dartmouth, Canada

mission. With modern NCM missions, the timeliness is important.

*This paper reports on communication challenges that were overcome to transmit SSS image snippets from a UUV, as soon as they are available, through the underwater path to an USV which relays to an operator.*

This has not been previously done as underwater communications between robots and other submerged systems is challenging. It is generally difficult to acoustically communicate even terse status and command messages underwater between the collaborating robots/operators much less, images. The communication challenges are briefly explained next.

### 1.1 Underwater acoustic channel and SSS images

Above-water, electromagnetic and light waves provide more than adequate bandwidth for the transmission of sonar image snippets that are  $\sim$  hundreds of kB in size. But, such signals are rapidly attenuated underwater. Acoustic signals are the most practical in terms of range and bandwidth for underwater communications. However, even with acoustic signals the underwater bandwidth and range are still relatively limited as typical center carrier frequencies are only tens of kHz ( $< 50$  kHz) compared to the MHz and GHz for in-air communications with above-water or terrestrial robots.

The underwater acoustic channel is also undermined by low sound speed (high latency), scattering and multi-path as well as local and temporal variations in the sound speed profiles. Consequently, SSS images suffer degradation due to blurring and Doppler effects (low sound speed), low contrast, non-diffuse reflections from the target and interference between the emitted and returned sonar signals. Even artificial targets, like MLOs, that are sought may not stand-out strongly against the background seabed. It is currently impossible to transmit information directly from underwater through the water-air interface. Proposed, is to use USVs as communication relays that receive underwater acoustic messages and translate them to in-air radio waves.

The underwater acoustic channel qualities drive the communications internet protocol (IP) to use—whether it be TCP (transmission control protocol) or UDP (user datagram protocol).

### 1.2 Underwater transmission requirements

With TCP IP, packet transmissions are reliable and ordered at the cost of more overhead (tracking, setup for socket connections, etc.). A packet is continually re-broadcast until there is acknowledgement it was received. Acknowledgements can take seconds or more—especially if propagation conditions are poor. However, this is needed to transmit say a zipped file as it cannot be reconstructed unless every bit was received.

Alternatively, with UDP IP, there is no acknowledgement of received transmissions and no ordering (later transmitted packets can arrive before previous ones) consequently, it is comparably lightweight and fast.

Usually, UDP IP message quality is checked on the receive end. UDP acoustic modems use additional encoding on their binary packets to limit the data loss. For example, checksums are used to indicate transmission errors. The choice to use UDP implies that message acknowledgement is a lower priority. For example, UDP IP can be used for video imagery transmission as a dropped packet means a poorer quality image, but an image still could be reconstructed.

The advantages of UDP IP are that it can broadcast simultaneously to multiple receivers and uses limited bandwidth more efficiently. The choice of UDP-like communications underwater is motivated by the high latency and low bandwidth in the underwater acoustic channel. UDP IP was used in this project.

To adapt to the difficult underwater acoustic communications channel the strategy is terse communications as with the Woods Hole Oceanographic Institution (WHOI) underwater acoustic modem (WHOI 2014). These acoustic modems are used on UUVs, USVs, and stationary seabed sensors. They have carrier frequencies around 25 kHz with a bandwidth of 4 kHz. All data transmitted acoustically is executed as binary fixed-length packets. The data rates for these modems are shown in Table 1. The reliability of the transmission varies inversely with its packet size, therefore, the lowest transmission rates are most reliable. This was quantified by the authors in experimental studies of packet transmission success rates as functions of packet size and range (Nams and Seto 2015). This means smaller messages are more likely to be successfully transmitted and received.

Therefore, the image compression methodology must accommodate small packet sizes to transmit at low bandwidths. PSK (phase-shift-keying) rates 1 and 4 (Table 1) provide a reference starting point for how small. In terms of a performance goal, the smaller the packet size the better and the interest is in what is possible over a 200 m range for NCM purposes. With the transmission constraints and goals defined, image compression was considered next.

### 1.3 Compression of side-scan sonar images

Generally, digital imagery including those from SSS have coding, inter-pixel and psycho-visual redundancies which could be removed for more compact transmission.

Coding redundancy arises when more code symbols are allocated than are needed to represent the monochrome levels in the image. Given an 8-bit representation, side-scan sonar images, at least of MLOs, do not have many high intensity pixels since the image is low contrast. There is a distinctive tail in the pixel intensity distribution which tapers to near-

**Table 1** WHOI Modem Data Rates (WHOI 2014) (bps = bits/s)

Data rate type	Frame size	Frames/packet	Packet size (B)	Packet rate (bps)
FSK-0	32	1	32	80
PSK-1	64	3	192	498
PSK-4	256	2	512	1301
PSK-5	256	8	2048	5388

zero values in the upper 20% of an 8-bit representation (e.g. Fig. 10), that is, most of an MLO SSS image is ‘dark’.

Spatial inter-pixel redundancies come from the correlation between adjacent pixels as they are not statistically independent and contain information about neighbouring pixels. With images from side-scan sonar insonification, this will be especially true with MLOs as opposed to stretches of featureless seabed.

Temporal inter-pixel redundancies come from the statistical correlation between pixels from successive frames as in a video sequence or repeat UUV dives in an NMCM mission. This project compares single images against a reference image so this is not as applicable.

Psycho-visual redundancy exists as human perception does not quantitatively evaluate each pixel’s attributes in visual processing. This means relatively insignificant information could be eliminated without much compromise to the reconstructed image. To address trained operators evaluating SSS images of MLOs requires a user study in a controlled environment which is beyond the scope of the present work.

The additional difficulty with side-scan sonar imagery is that the target’s appearance, in terms of highlights and shadows against this background, is dependent on the insonification direction. Side-scan sonar images will be poorer quality (and their compressed versions, poorer still) compared to camera and video pictures. Consequently, operators must be trained to recognize features like MLOs in SSS imagery from all insonification directions.

Quality metrics of similarity between two images, based on pixel attributes, are described next. The full reference image (as opposed to just the image statistics) is the original image snippet extracted in post-processing on-board the UUV. This would be compared against the reconstructed image which was compressed for transmission.

### 1.3.1 Quality metrics for compressed images

In image analysis the peak signal-to-noise ratio,  $PSNR$ , is a measure of the reconstructed image’s representation relative to the reference image.  $PSNR$  approximates human perception of reconstruction quality. Acceptable  $PSNR$  values for camera pictures and video are 20 dB. The higher the  $PSNR$ , the better the reconstruction. For a reference  $m \times n$  monochrome side-scan sonar image,  $f$ , and its approximate representation,  $g$ , the mean squared error ( $MSE$ ) is defined:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [f(i, j) - g(i, j)]^2. \quad (1)$$

The  $PSNR$  in decibels (dB) is then defined as:

$$PSNR(dB) = 10 \log_{10} \left( \frac{PP_{max}^2}{MSE} \right) \quad (2)$$

where  $PP_{max}$  is the image’s possible pixel maximum value. As the images were 8 bits/pixel,  $PP_{max} = 255$ .

The structural similarity,  $SSIM$ , index is another measure of similarity (perceived quality) between images. Unlike  $PSNR$ , it is based on visible structures in the image and captures perceptual differences (Wang et al. 2004).  $SSIM$  is calculated on sliding square windows of an image. The measure between 2 windows  $x$  and  $y$  of the same  $n \times n$  size is defined as:

$$SSIM(x, y) = \frac{(2\mu_x + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (3)$$

such that  $\mu_x = mean(x)$  and  $\mu_y = mean(y)$  relate to luminance;  $\sigma_x = var(x)$  and  $\sigma_y = var(y)$  relate to contrast and structure and  $\sigma_{xy} = covar(x, y)$  relates to structure.  $c_1 = (0.01PP_{max})^2$  and  $c_2 = (0.03PP_{max})^2$  help  $SSIM$  avoid instabilities when the denominator is small.  $SSIM$  lies in the range of:  $-1 \leq SSIM \leq 1$  where 1 means the two images are identical and 0 means there is no structural similarity between them. Unlike  $MSE$  and  $PSNR$ ,  $SSIM$  better quantifies human visual perceptual quality which is relevant as a measure of what operators recognize from reconstructed images. A comparison of pixel intensity distribution shape and relative magnitudes is a visualization of these similarity measures.

The overview of a solution is presented next given the difficult underwater acoustic channel and the challenges in compressing SSS images inherently and further, to accommodate underwater transmissions.

## 1.4 Underwater robots communicating images above-water

The robot communication challenges to transmit relatively high bandwidth information from underwater to above-water were addressed as follows. The system developed consisted of duplex transmitter (UUV) and receiver (operator) units

(Fig. 3). The transmitter is a payload processor serially interfaced to an underwater acoustic modem on-board the UUV. The autonomy on the transmitter payload processor extracts target image snippets from the SSS data stream, compresses/encodes them and then transmits the ones with high MLO likelihood to the receiver unit. The receiver unit is an underwater acoustic modem serially interfaced to the payload computer on the USV or a deck modem deployed from a small boat. Both could additionally have in-air radios to relay their received underwater messages and subsequent findings to above-water locations.

## 1.5 Novelty and contributions

The contributions from the reported work are as follows.

1. An algorithm to compress and transmit snippets of underwater side-scan sonar imagery above-water for potential in situ operator evaluation. The image compression adapted an existing approach (Murphy et al. 2010) for underwater pictures, to side-scan sonar imagery with similar results to pictures. This means side-scan sonar imagery can compress in the same manner as underwater photographs. Furthermore, sparse encoding methods that learn the basis directly from the imagery are more effective than discrete cosine and wavelet transform approaches.
2. A proof-of-principle of a robotic solution to extract, analyze, compress and transmit small, highly compressed side-scan sonar images in-situ from submerged robots to above-water robots using low-bandwidth acoustic communications. The above-water robots relayed the images to operators for evaluation. This was tested and validated through multiple sea trials.
3. More timely underwater surveys can be performed through a tighter collaboration between the UUV, USV and operator for naval mine counter-measures missions.

While the results ultimately contribute to the timely evaluation of MLOs by operators based on transmitted imagery, and it was initially presented to operators, this work is not a thorough investigation of the psycho-visual or human-robot interaction aspects. The emphasis was on overcoming challenges in communicating complex information between robots in different environments (underwater to above-water to shore).

## 1.6 Paper organization

The rest of this paper is organized as follows. To start, Sect. 2 has a description of the image compression requirements followed by a review of existing image compression algorithms. This is followed by a brief trade-off in simulation

of several possible compression algorithms that could be implemented on an UUV which helped select the vector quantization method pursued. Then, Sect. 3 describes the hardware used and the software developed to realize the method pursued.

This is followed by Sect. 4 which reports the in-water testing. Initially, the realized algorithm was tested in a controlled environment in a small (2 m) indoor tank, followed by at-sea trials to test for range and image database sufficiency. Then, advanced at-sea tests were undertaken where the on-board ATD extracts the images from the side-scan sonar stream to be vector quantized and transmitted from the acoustic modem. The paper concludes with some remarks and future work to address the specialized psycho-visual aspects specific to this problem.

## 2 Related work

Algorithms for compressing above-water camera pictures and videos may not be as effective for sonar images as the noise in underwater imagery is higher and the resolution is lower. The possibilities to compress SSS imagery snippets to the point where they could be transmitted underwater was explored as well as metrics to assess the compression quality.

There is a fair bit of work on processing side-scan sonar images with the intent to enhance it for better object detection. Typically, as mentioned earlier for NMCM and other surveys, the side-scan sonar imagery acquired by the UUV is downloaded and analyzed after the UUV is recovered. This is especially so since the SSS data is not typically analyzed on-board the UUV for several reasons. SSS imagery is memory intensive (30 min survey mission generates  $\approx 80$  MB of data). Prior to recent developments in UUV autonomy and embedded processors, it was not even possible to analyze the data on-board much less transmit information from it. Compressing SSS imagery was motivated by the authors' objectives to transmit snippets via low bandwidth transmissions to exploit adaptive robotic mission-planning and collaboration, sometimes with the operator in-the-loop, towards timely underwater surveys. There has been no motivation, beyond archival ones (Moszynski et al. 2013), to compress sonar data so the topic has not received much attention. The project's image compression requirements are stated next.

### 2.1 Above-water compression applied to SSS images

For the application considered, the image compression assessment will be based on:

- compression ratio which is critical as the image quality can be overlooked if the targets are visually recognizable;

- compressed image distortion as measured through the *PSNR* (not a consistent good measure) as well as the *SSIM* which accounts for some psycho-visual aspects;
- no image processing or augmentation to be considered at this point and
- processing effort to encode and transmit (bits/pixel) images as it is performed on embedded systems

as applicable to underwater images and side-scan sonar in particular.

As described earlier, side-scan sonar images of MLOs have psycho-visual, coding and inter-pixel redundancies that could be reduced without unduly compromising the reconstructed image quality. While there has been little work on SSS images towards this, above and below water pictures and video transmission has a fair amount to offer and is a starting point.

Above-water compression methods are generally divided into lossy and lossless. With lossy methods, information is irrecoverably lost in the compression while lossless methods retain the original information in the reconstructed image. Consequently, lossy methods produce much smaller transmission packets than lossless ones and are commonly used in pictures and videos. Pictures can achieve lossy compression ratios of 10:1 or better. In this regard, SSS images were treated like still pictures to start. Given psycho-visual redundancy some loss may be tolerable for human perception in exchange for larger compression ratios.

Examples of lossless compression methods (Yang and Bourbakis 2005) include the common PNG, GIF, TIF and JPEG2000 (Skodras et al. 2001). However, their compression ratios are insufficient for underwater low bit rate transmissions and are only retained here as a reference.

Methods like Huffman coding (Huffman 1952), discrete cosine transform (e.g. JPEG Skodras et al. 2001; Wallace 1992), vector quantization (Shen and Huang 2010) and wavelet compression (Said and Pearlman 1996) (e.g. SPIHT) are lossy where the reconstructed image only approximates its original but with the advantage of a more favourable compression ratio. This holds potential for compressing SSS images to the point where they might be transmittable from low bandwidth acoustic modems.

Lossy methods address the inter-pixel redundancy by transforming the image then quantizing it to remove the psycho-visual redundancy and finally, encoding it efficiently to remove coding redundancy. The compression performance of these methods will be assessed by applying them to SSS.

### 2.1.1 Discrete cosine transforms

Discrete cosine transform methods like JPEG (Wallace 1992) and even lossless JPEG2000 (Skodras et al. 2001) were not favored for this application as they do not compress well

enough for successful low bit-rate transmissions. JPEG2000 will be retained in the analysis to provide a reliable reference for comparison purposes.

### 2.1.2 Huffman coding

Huffman's method (Huffman 1952) encodes images to compress more efficiently. It is used in JPEG (Wallace 1992) for example. Huffman coding is based on frequently-appearing values given shorter bit representations than less common ones. To achieve this, a tree represents the priority of each value based on its frequency of occurrence. A given value is coded through a traversal of the tree. Therefore, Huffman coding is more efficient than *.zip* or *.rar* compressions. In Moszynski et al. (2013) this was applied to compress multi-beam echo sounder (MBES) images for archiving. MBES data is not unlike SSS data however, this was an application for only archiving.

### 2.1.3 Vector quantization

Vector quantization of above-water pictures can achieve PSNR values up to 30 dB which is acceptable. However, methods that adapt the encoding (Shen and Huang 2010) can substantially improve the quality of vector quantized compressed images still more. With vector quantization, a picture is deconstructed into a mosaic of non-overlapping square tiles. Then, the intensity of each pixel in the tile populates a vector,  $T$ , with a specific ordering. In this way, a constituent tile can be compared against a codebook or database of tiles,  $C_i$ , where  $k$  is the tile vector's dimension. Tile similarity is assessed through the Euclidean distance,  $d$ , between a constituent tile vector  $T$  and database tile vector,  $C_i$ . The smallest distance,  $d$ , is the best fit:

$$d(T, C_i) = \sqrt{\sum_{j=0}^{k-1} (T_j - C_{i,j})^2}. \quad (4)$$

Each constituent tile is compared against tiles in a common database at the transmit and receive end. Once the most similar tile in the database is identified for a constituent tile, that tile's database index is retained (and later transmitted) as representative of that part of the image. Various methods can be used to optimize the training of the database (Bhattacharyya et al. 2014; Ferdowski et al. 2017).

WHOI (Murphy et al. 2010) developed a compression method using vector quantization (VQ, to refer from this point on to the WHOI method), specific to color seafloor pictures. This process is facilitated through a principal components analysis (PCA) (<https://www.quantstart.com/articles/Eigen-Library-for-Matrix-Algebra-in-C>) to determine the Euclidean distance. At the receiver end when all the indices

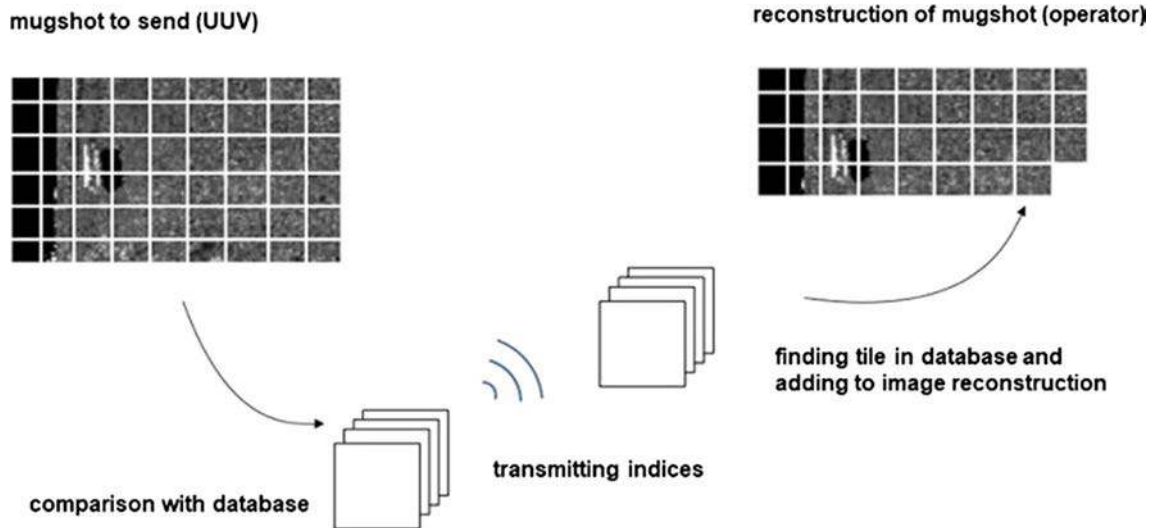


Fig. 1 Steps in the WHOI vector quantization (VQ) compression algorithm

for all constituent tiles are received, the image can be reconstructed from the common database (Fig. 1). It is possible to implement this on the payload processor of an UUV and to transmit images through low bandwidth acoustic modems.

#### 2.1.4 Wavelet-based encoding

SPIHT (Set Partitioning in Hierarchical Trees) (Said and Pearlman 1996) uses wavelet transforms to encode images. The transmission is based on two steps: (1) the wavelet coefficients are ordered by magnitude and (2) to transmit the most significant bits first. The encoder (transmitter end) and decoder (receiver end) use a common ordering scheme so it is not necessary to transmit that. A spatial orientation tree defines the data ordering based on a recursive four sub-band splitting. The tree is defined so each node has either no offspring (the leaves) or four offsprings. Consequently, the compressed image can be truncated at various points and decoded to give a series of progressively refined (higher resolution) versions of the initial image. Bin and Qinggang (2013) implemented an improved SPIHT algorithm, based on adaptive lifting wavelet transforms, and applied it to sonar images. They found it difficult to capture edges well but were able to achieve bit-rates of 0.2 bits/pixel. Others (Liu and Li 2013) also use lifting wavelet transforms with PSNR of  $25^+$  on Lena (people) images, which are high contrast and less noisy, compared to SSS images. They were able to do this at low bit-rates (lowest  $\sim 0.016$  bpp) and with encoding times as low as  $\sim 0.14$  s. However, their method additionally requires more computation resources to achieve this.

Atallah et al. (2016) applied SPIHT to compress images for underwater transmission. To simulate underwater transmission they added white Gaussian noise and delay paths to

emulate multi-path effects. This does not capture the underwater acoustic channel challenges this project was interested in. Discrete wave transforms were applied to de-noise their received images to get acceptable PSNR values. This work is still at an early stage.

Zhang et al. (2015) presents a seafloor image compression technique based on hybrid wavelet transforms and directional filter banks to underwater color pictures. This is a development beyond discrete wavelet transform approaches as it captures texture and contour structures better. They achieved good compression and reconstruction through high coding efficiency for transmission at low bit rates. The relevant aspects from this for the present application is a novel distortion model to remove psycho-visual redundancies for underwater still images. The psycho-visual redundancies are not a priority for this present project but they are for the next one. This method is computationally intensive and has not been tested for implementation on an embedded system.

Zhang et al. (2016) is a development on the work of Zhang et al. (2015) applied to underwater color video compression. It comprehensively targets the psycho-visual aspects by modelling the human vision system including the spatial-temporal aspects from consecutive images as would be the case in video. Through insight on these psycho-visual redundancies and removing them to achieve lower bit rates they were able to achieve high PSNRs on underwater color images which are more information-rich than underwater side-scan sonar imagery.

It was not clear what the computation effort was for Zhang et al. (2015, 2016) to obtain their results and whether it would be achievable on an embedded system. They achieved low bit-rates (0.1 bits/pixel) for transmitting color video which is comparable to VQ's 0.14 bits/pixel. This methodology mer-

its further consideration when the psycho-visual aspects are addressed in the follow-on project. At that point it is worth testing for encoding on an embedded system. It would likely require a different payload processor than what has been used to date.

Therefore, SPIHT, Huffman coding and vector quantization will be evaluated as the methods that could be implemented with an embedded processor. JPEG2000 will be retained as an example of a lossless method to compare against. The next step was to investigate similarity measures for SSS imagery.

### 2.1.5 Sonar image quality assessment

For sonar image quality assessment, the earlier work of Kalwa and Madsen (2004) used the mean pixel value (noise), entropy (complexity of image) and substance (amount of strong returns in the image). However, these measures do not adequately capture what is needed given the sonar image's nonuniform noise as well as noise that contains strong returns.

Chen et al. (2017) compiled a diverse database of sonar images, initially ground truthed by (subjective) experts, to develop metrics for sonar image quality assessments. They compared images that were degraded by compression and transmission against their full reference image and devised an objective metric that uses the entropy map. They used gradient magnitudes in the sonar intensities to look for targets. This has potential in the highlight regions and shadows that are part of a target. However, there are as many strong gradients in side-scan sonar images from corrupted images and the nadir (blind spots in side-scan sonar) and is dependent on insonification aspect with SSS. In this regard, this is no better than *SSIM*.

Li et al. (2018) devised subjective and objective evaluation measures with sonars and cameras to monitor the density of biological invasive species towards the safe operation of a nuclear power plant. However, their focus was on analyzing sonar (and video) images without references. Their image quality is evaluated as a distance between a no-reference quality aware model and the fit of the distorted image against that. This is not the comparison that was desired.

DeBortoli et al. (2017) applied an atrous convolution architecture as part of a learning approach to extract frames with high quality imagery from a sequence of images. The advantage of this, beyond other convolutional neural network approaches, is the use of more pixels around the target which decreases the dependence on strong local targets. This is extensible to extract high quality frames that are not limited to those the system trained on. Their system selects images based on its usefulness to a human operator. DeBortoli et al. (2017) also targets the psycho-visual aspects which is of interest in the follow-on work. This approach will be pur-

sued in the next project phase as DeBortoli et al. (2017) came out well after our original concept was presented to show that learning the basis of the target was a good approach. Whether DeBortoli et al. (2017) can be implemented on an embedded system (through transfer learning or distillation Hinton et al. 2014 for example) remains to be seen.

Given that there are no image quality metrics that specifically address side-scan, or any other sonar imagery for full reference analysis, the project chose to stay with *PSNR* and *SSIM* as the quality metrics.

Therefore, the image compression method to use was based upon comparing the merits of SPIHT, Huffman coding and vector quantization using *PSNR* and *SSIM*. This is discussed next.

## 2.2 Image compression selection method from simulations

Based on the requirements outlined in Sect. 2.1 for the compression scheme, Huffman coding, vector quantization and SPIHT were further investigated for their appropriateness to SSS images. JPEG2000 was retained as a representative lossless reference. In the evaluation through simulations, the four algorithms were implemented in MATLAB<sup>®</sup>.

An example of an evaluation of these four compression methods against a reference image ( $49 \times 113$  pixels) is shown in Table 2 with a visual example in Fig. 2 from the Klein 5500 SSS sonar.

As shown, JPEG2000 and Huffman coding reduce the file size by less than 10 times and distorts the images as reflected in lower *PSNR* values. The SPIHT algorithm provides either visually satisfactory results with a poor compression ratio (shown) or distorted images with a good compression ratio (not shown). The best results are achieved with the VQ method. The compression ratio is more than satisfactory resulting in a compressed file of only 100 bytes in this case. Furthermore, the image is visually similar to the original image.

Given the complexity and uncertainty in underwater sonar images, a sparse coding method that learns its basis directly from the imagery has greater effectiveness than approaches based on discrete cosine and wavelet transforms. This should at least work where the training images are similar to the test images. Vector quantization is an example of this.






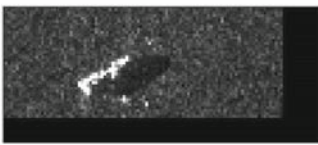
The compression time for VQ takes  $\sim 35$  s on the embedded processor (described later). The other methods took much longer. This relatively short time means it is feasible for the operator to view the images while the UUV is underway and soon after the images were captured. This makes it possible for the NMCM mission to be carried out more rapidly.

The most promising compression scheme, against the requirements in Sect. 2.1, for underwater sonar images from



**Table 2** Relative performance of four compression methods applied to side-scan sonar images

	Method			
	JPEG 200	Huffman coding	Vector quantization	SPIHT
Compress ratio	9.449	8.916	251.682	25.924
PSNR	29.52	21.66	18.19	21.94
Pros	Constant updates	Previous info not required to compress	No distorts fr compression	Image can be truncated to give refined version
	Previous info not required to compress		Quality dep on database	Previous info not required to compress
Cons	Distortions under high compression	Distortions under high compression	Requires a tile database	Distortions under high compression

	<b>original image</b>	
	size:	5.65 kB 49×113
	<b>JPEG 2000</b>	
	size:	588 B
	compress ratio:	9.6
	PSNR:	29.65
	SSIM:	0.54
	<b>Huffman coding</b>	
	size:	671 B
	compress ratio:	8.4
	PSNR:	21.72
	SSIM:	0.40
	<b>SPIHT</b>	
	size:	1119 B
	compress ratio:	7.4
	PSNR:	24.49
	SSIM:	0.66
	<b>vector quantization</b>	
	size:	88 B 10 ×10
	compress ratio:	62.92
	PSNR:	19.24
	SSIM:	0.90
	size:	20 B 20 ×20
	compress ratio:	251.68
	PSNR:	18.19
	SSIM:	0.88

**Fig. 2** Compression algorithm performance on archived Klein 5500 side-scan sonar image

the evaluation is the VQ algorithm (Murphy et al. 2010). This method’s performance depends, not unexpectedly, on the quality of the database. The optimal database would have

many tiles with great diversity that capture a variety of seabed bottom types (gravel, sand, clay, mud, etc.).

Noisier images, like those in SSS, require more representation (bits) to achieve the same fidelity as measured through the *PSNR*. Therefore, they do not compress as well. Vector quantization effectively applies a low pass filter which smooths the high frequency noise yet still captures the main features (in the lower frequencies). This lowers the bit rate and makes it more bandwidth efficient.

Given the VQ database represents 15,000 tiles and only the indices of the constituent tiles in an image are transmitted, this means each  $10 \times 10$  pixel tile can be represented by 14 bits (up to 16, 384 possible tiles), the ‘bit-rate’ is 0.14 bits-per-pixel. This is comparable to other efficient low bit rates at 0.1 bits-per-pixel (Zhang et al. 2016) using wavelet transforms. VQ decoding is fast and straight forward as it re-assembles the image in a look-up table manner. Vector quantization is a method that is resilient against the inherent noise in SSS images.

Therefore, VQ was selected as the image compression method and its implementation is discussed in Sect. 3.3. However, the overall system will be discussed first in more detail.

### 3 System description

#### 3.1 High level systems overview

The systems overview is shown in Fig. 3. The process starts with the UUV acquiring SSS data underwater in the upper left and terminates with the above-water display of the reconstructed transmitted image on the bottom right for operator evaluation. The top blue dashed-line box (transmitter—UIB) graphically shows the steps that occur on-board the UUV prior to the image snippet being transmitted as an acoustic modem message. The bottom blue dashed-line box (receiver—operator) shows the steps to decompress the received image.

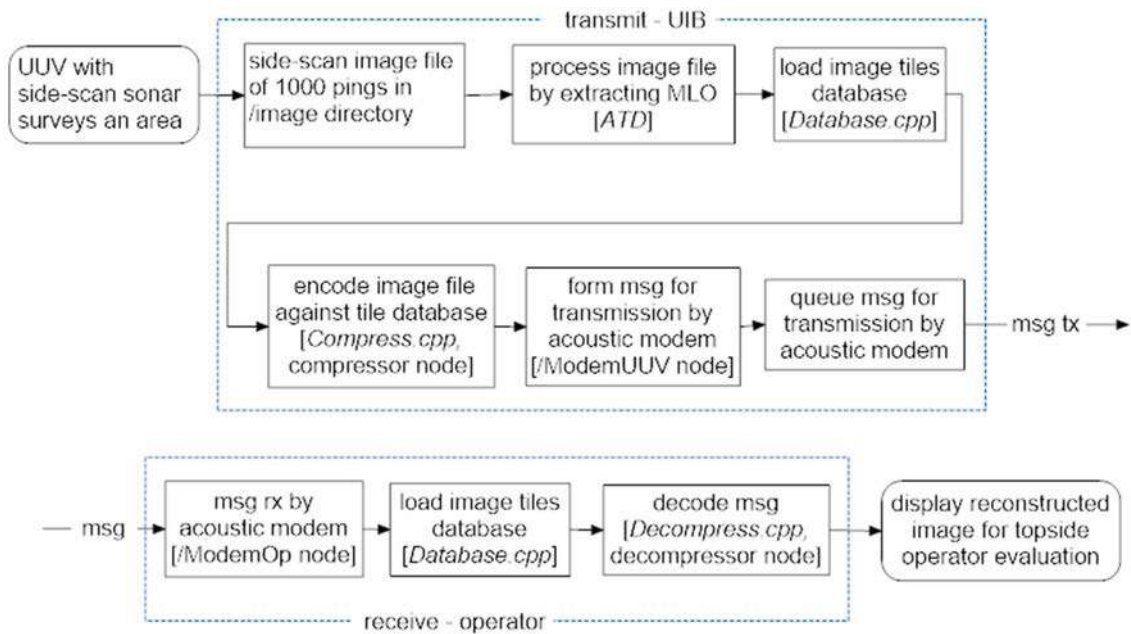


Fig. 3 System overview—hardware components in bolded boxes and names of software components in italics



Fig. 4 IVER3 unmanned underwater vehicle

Items in square brackets identify components realized in software that are described in this Section and the Appendix.

### 3.2 Hardware description

#### 3.2.1 Unmanned underwater vehicle

The work here targets the capabilities of the IVER3 UUVs (Fig. 4) that DRDC owns and uses. One has the Klein 3500 side-scan sonar with an interferometric bathymetry sonar and the other has the MarineSonics HDS side-scan sonar.

The embedded processor on-board the UUVs is an Intel Atom Dual-Core 1.6 GHz N2600 that can access 512 GB of RAM. It has a maximum memory of 2 GB and is capable of 4 threads. This is roughly similar to an Intel I3 processor. There



Fig. 5 HMS (Her Majesty's Ship) Hazzard unmanned surface vehicle

were plans to potentially upgrade to the NVIDIA Jetson TX2 (256 CUDA Cores). This GPU processor is for advanced on-board computer vision processing. Its power consumption however, is twice that of the Intel N2600.

For the advanced in-water trials the underwater messages were sent to unmanned surface vehicles which are briefly discussed next.

#### 3.2.2 Unmanned surface vehicle used

The unmanned surface vehicle that was used as the receiver and relay units in some of the trials is shown in Fig. 5. At a minimum, the USV has an underwater acoustic modem to receive underwater messages and can host the receiver-operator module (Fig. 3). Additionally, the USV must have an in-air radio to relay the reconstructed images to shore.

### 3.2.3 Underwater acoustic micromodems

The underwater acoustic modems on-board the UUVs and USVs were previously described in Sect. 1.2 and WHOI (2014).

## 3.3 Software development

### 3.3.1 Description of software modules

All the software developed is original to the project except for the principal components analysis which was borrowed from the Eigen Library for Matrix Algebra in C++ (<https://www.quantstart.com/articles/Eigen-Library-for-Matrix-Algebra-in-C>). The developed code components, outlined in Fig. 3 in boxes, and described below, is organized as follows:

- **VQScheme** class contains methods and functions for image compression and decompression
- Compression is achieved through *Compress.cpp* (compressor ROS Purvis 2017 node) compiled as an executable.
- Similarly, decompression is achieved through *Decompress.cpp* (decompressor ROS node).
- The VQ database is formed through *Database.cpp* (algorithm in Fig. 6) and support functions (not discussed). The database is formed and compiled ahead of time however, the database in binary form is loaded on-line as shown in Fig. 3.

**VQScheme** has one private function, *matchingTile*, which is defined for vectors of doubles or unsigned integers. This function takes as parameters a 1D-vector of 8 bit unsigned integers that represent a tile and a 2D-vector that represent a database. It is also defined for 1D-vector of doubles that represents a tile and a 2D-vector that represents a database so that it can be used for constituent and database tiles. It compares the constituent tile to those in the database and returns the index of the one that is most similar. This function, *matchingTile*, is used to compress an image by the *encodeCompress.cpp* function. During its execution, this function affects the value of the distance between the closest database tile found and the constituent image tile (Fig. 7).

The results from a sensitivity analysis of parameters that impact the VQ compression are presented and discussed.

### 3.3.2 Vector quantization: database

There are two thresholds in the database creation to determine whether a new tile needs to be added. This, in turn, affects the execution speed and compression quality. To gain insight into these parameters' impact, a sensitivity analysis

```
1: Algorithm create_database (path to database images  
   folder)  
2: IF database core = empty  
3:   tile first image and add to database  
4:   WHILE non-treated image in folder  
4:     read .tiff file  
5:     reconstruct image with current database  
6:     compute FOM  
7:     IF FOM < 39  
8:       tile image  
9:       IF distance to closest database tile > 320  
10:        add tile to database  
11:      END IF  
12:    END IF  
13:  END WHILE  
12: ELSE // database core != empty  
13:   load database from binary file  
14:   WHILE non-treated image in folder  
15:     read .tiff file  
16:     reconstruct image with current database  
17:     compute FOM  
18:     IF FOM < 39  
19:       tile image  
20:       IF distance to closest database tile > 320  
21:        add tile to database  
22:      END IF  
23:    END IF  
24:  END WHILE  
25: END IF  
26: RETURN (decoded image as .tiff in out folder with initial  
   image name)
```

Fig. 6 Database creation algorithm

was performed. The analysis focuses on the 4 parameters deemed most significant:

1. number of dimensions retained in the principal components analysis
2. threshold for a user-defined figure-of-merit, *FOM*, that assesses whether an image reconstruction is acceptable
3. threshold for the Euclidean distance to the database that defines what tiles should be added to the database (if any), and
4. number of images in the database.

The results are evaluated using a set of test images where the following were calculated: the *FOM*; the difference in the ATD confidence that the image contained an MLO before and after compression; the time to compress an image; the size of the database, and the time to create the nominal database (a one-time activity).

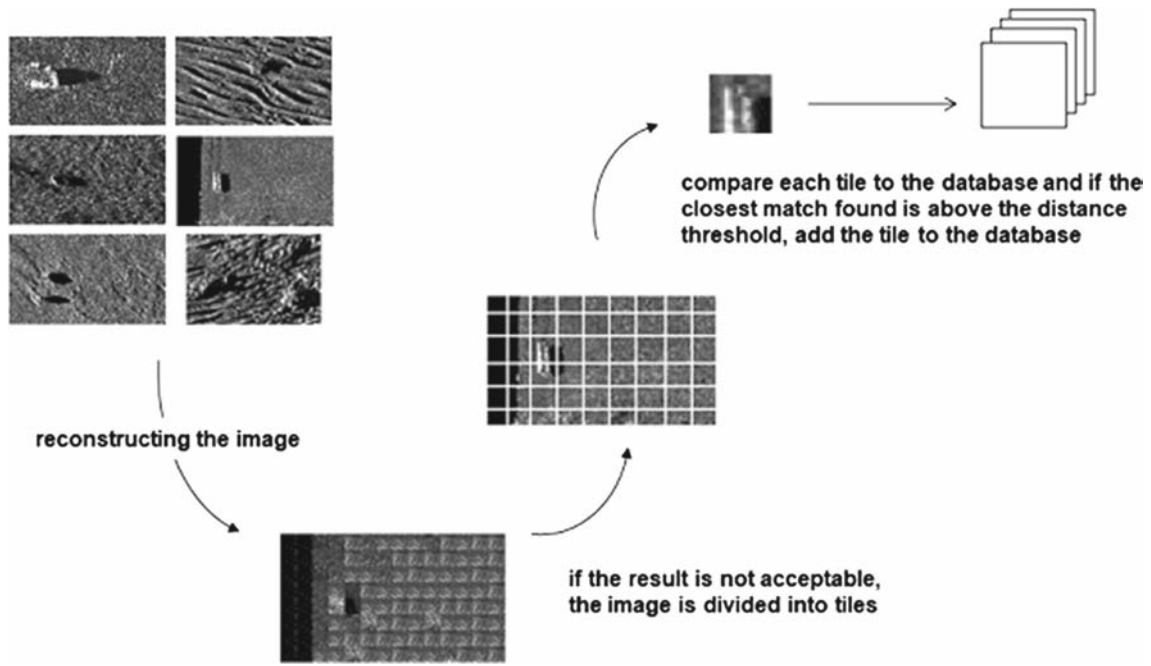


Fig. 7 Illustration of database creation algorithm

The *FOM* pertains to a reconstructed image straight from the database and how similar it is to the original image. It takes as parameters a 2D vector which has on the first row, the reference (database index) to the tiles that constitute the compressed image and on the second row, the distance between each tile of the original image and the tile chosen from the database to encode it. These distances are squared and added to obtain a type of mean squared error,  $MSE_2$ , [to distinguish it from the earlier defined Eq. (1)] for the reconstructed image. Therefore, the *FOM* is a number between 0 and  $\infty$  and is defined as:

$$FOM = 10 \log_{10} \left( \frac{\max(MSE_2)^2}{MSE_2} \right).$$

Defined this way, *FOM* values less than 39 dB produced results that were visually quite similar for the authors - admittedly a bit subjective. This is a similarity measure specific to the database in the VQ implementation for SSS images.

The ATD confidence is obtained through a fusion of several filters and represents the likelihood the ATD detected an MLO. Its value is a rational number between 0.0 to 2.0. The ATD confidence (not shown) was inconclusive as it is mostly constant over the test images examined. However, when the number of tiles used to create the database increases it became more diverse and the maximum error in the ATD confidence diminishes. This proves that diversity in the database ensures better reconstructed images, as expected.

Figure 8 shows that when more dimensions in the PCA are retained, it increases time to create the database and to

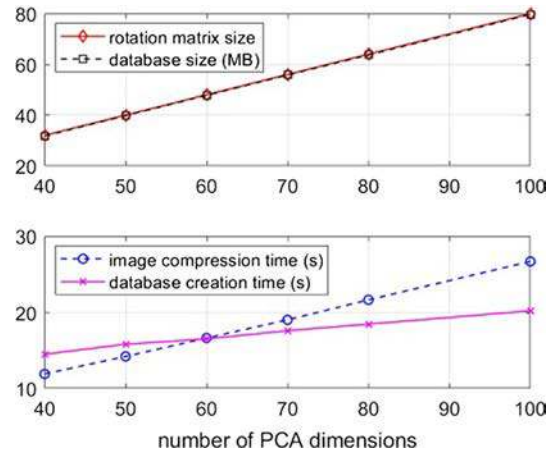


Fig. 8 Sensitivity of VQ parameters to the principal components dimensions retained

compress an image. As well, the size of both the PCA rotation matrix and the database increases. But, the results are visually better even with no significant difference in the *FOM* and the ATD confidence (not shown). Thus, the number of PCA dimensions retained was set to 70, as it is a reasonable balance between reconstructed image quality and computation time.

When the number of images in the database is increased to improve the results, the computation and execution time as well as the memory required, increased quickly. The solution, then, was to build the database incrementally. First, many images are used to generate the nominal database with a low *FOM* and high distance threshold. This means only tiles from

images that were poorly reconstructed and quite diverse from those in the original database were added. Then, the number of images and the threshold distance is reduced and the *FOM* threshold is increased. This process was repeated multiple times to arrive at a database that was a reasonable compromise between computational efficiency and image quality.

To create the VQ database, 500 SSS images from multiple types of SSS were deconstructed into constituent tiles (redundant tiles removed) and inserted into the database. This resulted in a 15,992-tile (88 MB) database which is a respectable size that is manageable by the payload processors on-board the UUVs. This was the training set.

### 3.3.3 Vector quantization: other parameters

The image snippets were automatically cropped to  $80 \times 140$  pixels by the ATD. This size was chosen based on practical considerations. For 900 kHz side-scan sonars the range on each side is about 40–50 m. The altitude of the UUV with the side-scan sonar is typically 10% or less of that range which is 3–5 m. Given that, cropped images of underwater targets are approximately  $80 \times 140$  pixels. This is a size that will encompass the target with enough seabed bottom type visible if bottom type was to be used in the ATD image analysis. This is also approximately the image size that the ATD processes to determine the likelihood a target is an MLO. With the image snippet size determined, the next consideration is the tile size in the database.

The size of each tile in the database impacts the size of the database and the computation speed to determine the best fit for each candidate tile. As shown in Fig. 2, a  $20 \times 20$  pixel tile gives a fourfold increase in the compression ratio over a  $10 \times 10$  pixel tile. However, the *PSNR* degrades, as expected, since the ‘resolution’ of the image drops due to a larger tile size. A  $5 \times 5$  tile size (not shown) needs a larger database and transmitted message to exploit the better resolution. However, it is also more unwieldy for embedded processing and low bandwidth transmission. Based on this, a  $10 \times 10$  pixel size was chosen for the size of the tiles in the database.

The resulting compressed image snippets were larger than the 256 bytes that could be accommodated in 1 frame so two frames were used for PSK-4 transmission. If the full 512 bytes was not needed, the empty bytes were padded with zeroes. The compression ratios (not shown) in all cases were similar to what was shown for VQ in Table 2 for  $10 \times 10$  pixels ( $\sim 60$  or slightly better).

How these nodes were implemented and interfaced with the underwater acoustic modems is described in detail in the Appendix.

## 4 In-water tests

### 4.1 Preliminary validation in an indoor water tank

After the ROS nodes were developed to realize the algorithm, they were tested in an indoor water tank with two submerged acoustic modems each serially interfaced to a computer – the transmitter (UUV) and receiver (operator) units. The two acoustic modems were 2–3 m apart in the tank. The transmit (UUV) unit hosted the *talker* node which simulated the ATD detecting and extracting MLOs in the sonar data stream and its modem emulates the on-board UUV modem. The *talker* node is redundant when the ATD is later integrated into the system. The receiver (operator) unit is used by the operator to receive the compressed images via its modem.

Modem transmission rates were trialled at PSK-1 and PSK-4 (Table 1). During these tests the communications was adapted to handle broken packets (a very real situation). Previously, if a packet was received it was considered to contain all the data needed to reconstruct half the sonar image. However, because of the highly reverberant water tank (small size and lack of anechoic cladding) some of the acoustic packets were damaged: the end of a packet might be unintentionally padded with zeros or random numbers. The algorithm was adapted so that if a tile index in the encoded message was greater than the maximum index possible then it was replaced with a black tile to signify a bad packet.

After these adaptations, the in-water tank tests were successful. A few packets were lost or damaged because of the tank reverberations and the PSK-4 rate, as expected, is not as reliable as the PSK-1 rate. These tests validate the implementation of the compression/decompression algorithms and the transmit/receive protocols used. They also suggest that for at-sea conditions, the acoustic modems should rely on the PSK-1 rate though the PSK-4 rate was used as well. The at-sea tests are described next.

### 4.2 Initial tests and validation at-sea

The objective of the at-sea tests was to assess the transmission and reception success over much larger acoustic modem separations, in a body of water with variable sound speeds and over a real seabed. These factors present underwater communication challenges to successful message transmission and reception. Otherwise, the testing and validation in Bedford Basin (to depths of 20 m) were similar to the indoor water tank ones. The same images were transmitted and received to assess the impact of range. PSK-1 and PSK-4 modem modes were used. PSK-1 gave better reliability at longer modem separations.

The transmitter (UUV) unit computer was on-board a small boat with its modem deployed over the side into 6.4 m water depth. On the receiver (operator) unit, its acous-


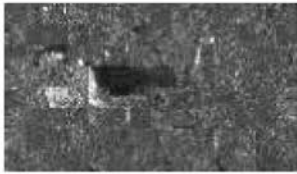
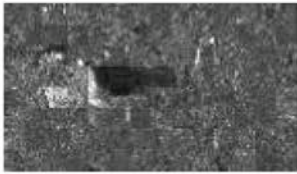
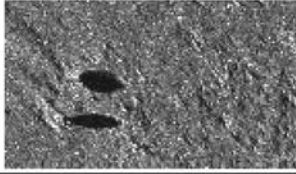
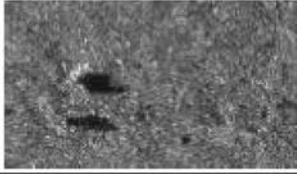


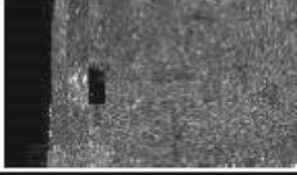

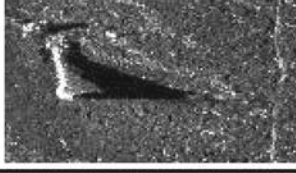


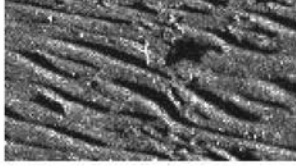
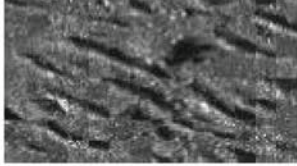
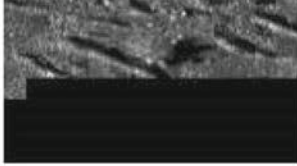
case	characteristics	original image	image received at 20 m	image received at 60 m
1	bright returns and shadows			
2	distinguish 2 objects			
3	bright returns, shadows and holiday saturation			
4	bright returns and shadows			
5	sand ripples			

Fig. 9 Representative results from indoor tank and initial sea trials and validation of underwater transmissions at different ranges

tic modem was tethered to a barge at the same water depth. The underwater transmissions ranges trialed were 2 to 200 m between UUV and operator units. Representative results are summarized in Fig. 9 and analyzed in detail in Figs. 10, 11, 12 and 13.

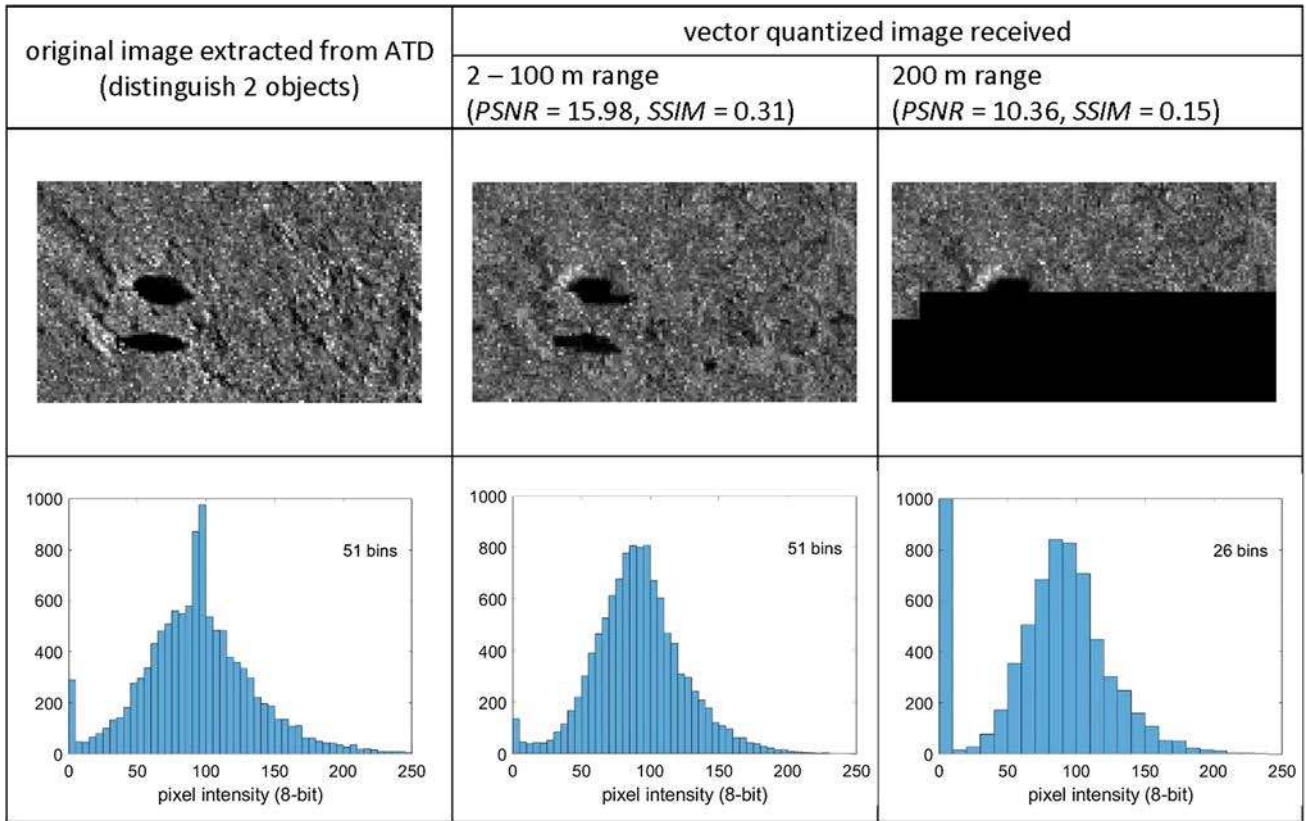
At less than 130 m range between the transmitter and receiver units, given the acoustic propagation conditions that day, most transmitted packets were received. The transmission was 100% successful when the size of the images was  $140 \times 80$  pixels. For smaller images the second packet was missing in some cases. This error was intermittent and due to the second and third frame of the packets at PSK-1 being empty or not full to start, thus resulting in unreliable transmissions. This was fixed for the next set of trials. At greater than 130 m range, most packets were lost. A sound velocity profiler dropped at the beginning of the runs showed the best possible range achieved would have been about 130/140 m. This is consistent with what was observed and thus this is not a reflection on the VQ algorithm.

Of the images (or parts of images) received they were reconstructed with the same fidelity as obtained in the indoor tank tests or they were completely unrecognizable.

#### 4.2.1 Analysis of results

The analysis of the at-sea SSS collected and extracted images (representative results in Table 2) was through calculating the *PSNR* and *SSIM* as well as inspecting pixel intensity distributions for vector quantized images against their reference (original) image.

The way the ATD extracts targets, the target is usually centered vertically in the frame and to the left. Therefore, the decision was made that if an image had to be transmitted over two packets the packets would be divided along the lines of the top and bottom (as opposed to left and right) of the image snippet. Black pixels were used as filler in a decompressed image for a packet that was not received. Consequently, in



**Fig. 10** Comparison of the original image against the received vector quantized version as a function of range between the acoustic modem transmitter and receiver. Fidelity of the reconstructed image is quanti-

fied through the peak signal-to-noise ratio ( $PSNR$ ), structural similarity index ( $SSIM$ ) and distribution of pixel intensities. At 200 m range, 1 of the 2 packets for the image was dropped (case 2 from Fig. 9)

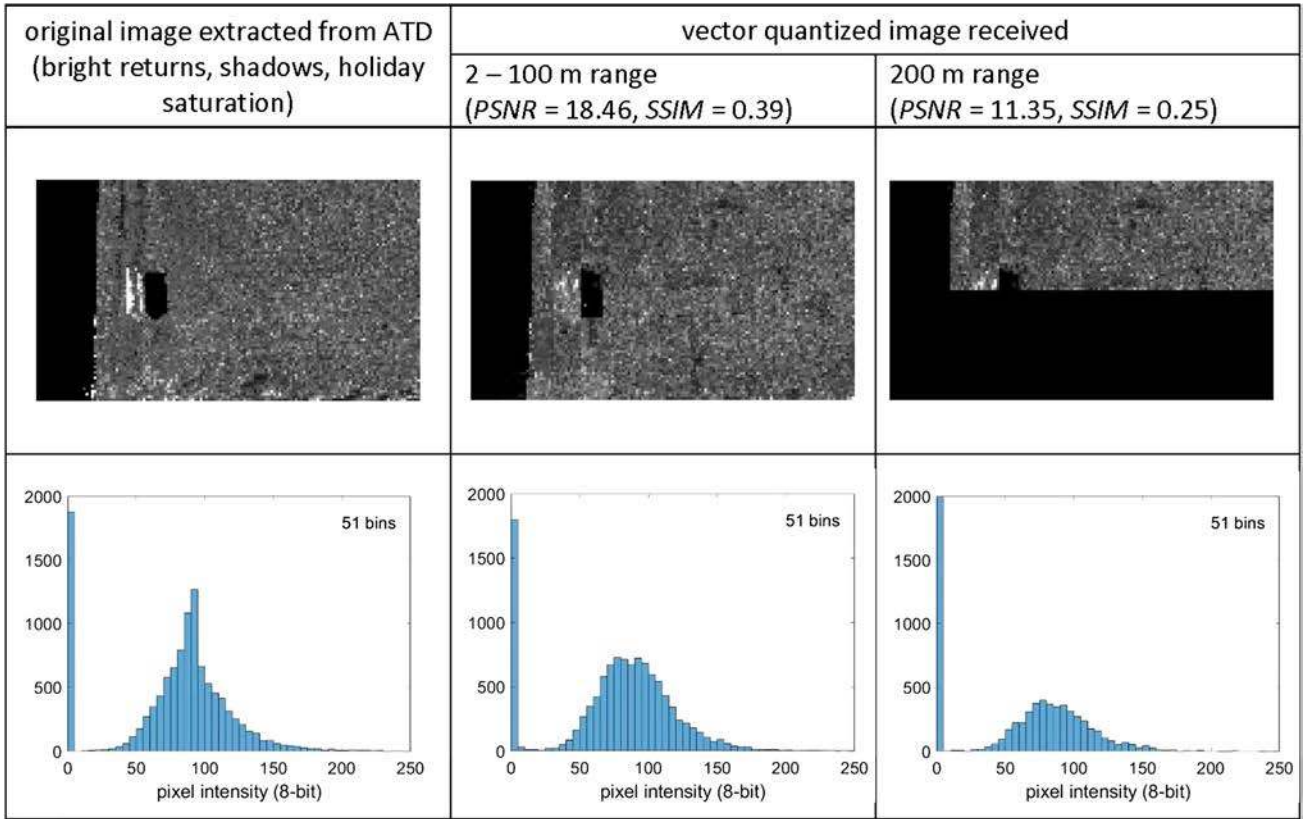
those images, bin 1 in the pixel intensity distribution was artificially inflated.

Of the examples shown in Fig. 9, cases 2 (Fig. 10), 3 (Fig. 11), 4 (Fig. 12) and 5 (Fig. 13) were analyzed in more detail and are discussed next.

Case 2 (Fig. 10) had 2 targets that were close and insonified from an aspect that produced clear shadows. The vector quantized image distinguished between the targets and their shadows with a  $PSNR \sim 16$  and the  $SSIM = 0.31$ . Based on these metrics, the similarity between the two was marginal but acceptable. Overall, the pixel intensity distribution (statistics) in the vector quantized image captured the shape and magnitudes well. What was not captured in the compressed image was the increased number of pixels at the peak of the distribution. At 200 m range, only the packet representing the top of the image was received. Consequently, with half the image missing the number of bins was halved but the shape and the amplitudes were representative of the original image. The large number of pixels at pixel intensity bin 1 is due to the black pixels used as filler for the missing half image.

Case 3 (Fig. 11) shows a target extracted by the ATD with bright highlight pixels as well as a shadow. The interesting aspect of this image was that it was close to the nadir of the SSS (black strip on the left)—not atypical with SSS images. This is the blind spot beneath the sonar that cannot be insonified and was filled in with black pixels. The impact of the nadir on the vector quantized image was more pixels in bin 1 of the pixel intensity distribution. The original and vector quantized images showed this - not surprisingly. As in case 2, the pixel statistics were preserved with the nominally correct shape and magnitudes except at the peak values. The presence of the nadir did not make a difference in the fidelity of the vector quantized image. It also had no impact in the compression since each tile in the original image was represented by the same number of bits. This would not be the case for methods like SPIHT where frequently appearing values have a different representation than less frequent ones which would impact compression.

At 200 m range the bottom half of the image was not received however, unlike case 2, the bin discretization remained the same as the original. Unlike case 2, the magnitudes for other than intensity bin 1 were reduced although the



**Fig. 11** Comparison of the original image against the received vector quantized version as a function of range between acoustic modem transmitter and receiver. Fidelity of the reconstructed image is quantified through the peak signal-to-noise ratio ( $PSNR$ ), structural similarity

index ( $SSIM$ ) and distribution of pixel intensities. The black strip on the left is the typical nadir zone below the side-scan sonar (case 3 from Fig. 9)

shape was nominally correct (which proved our argument that it was better to divide the image along top and bottom). The reduced magnitudes are due to the smaller number of pixels available compared to case 2. Both the  $PSNR$  and  $SSIM$  were larger than case 2's. Lost packets aside, both similarity measures agree that case 3's vector quantized image was a better match than case 2's.

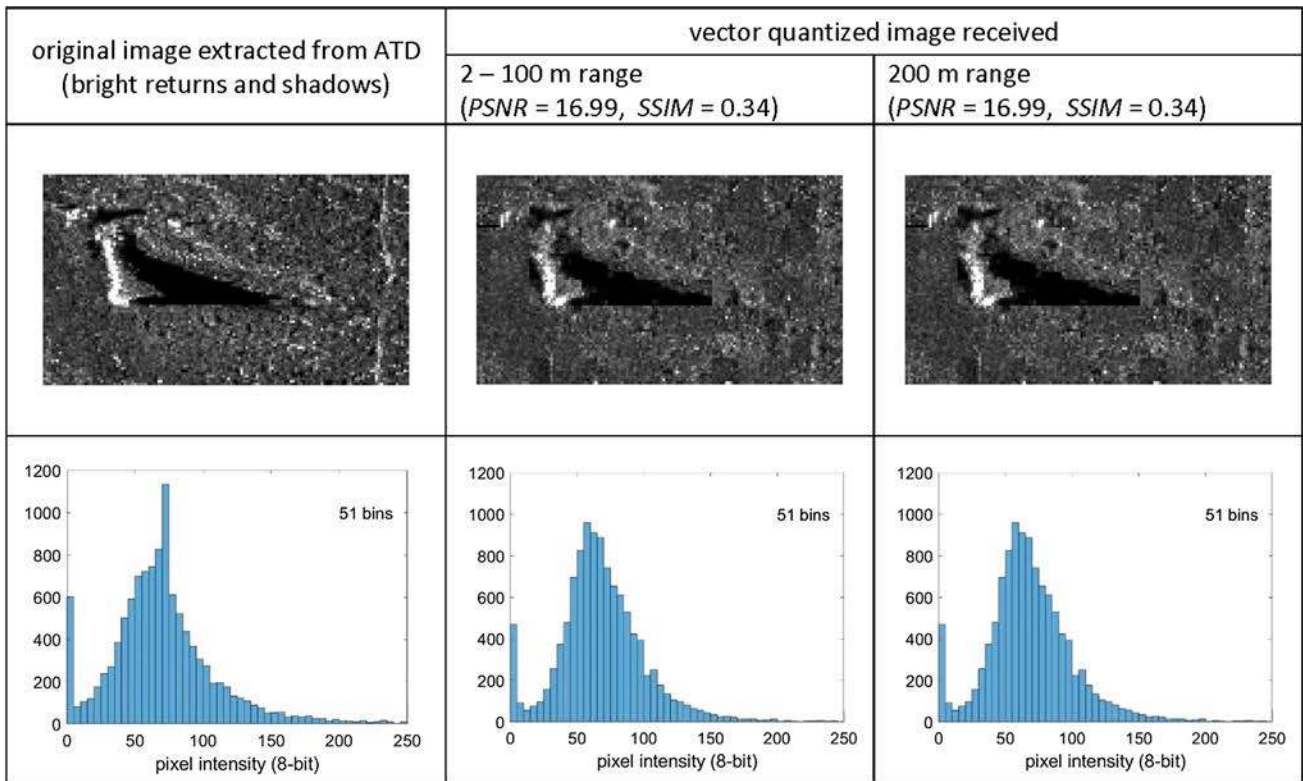
Case 4 (Fig. 12) is similar to case 1 (detailed analysis not shown). Both show highlights and dark features in the target image which speaks to a larger dynamic range. Case 4 appears to have greater diversity in the highlight (brighter) returns. The pixel intensity distribution did not show trends that were different from cases 2 and 3. The only small notable difference is that the tail of the distribution was not as well captured as the other 2 cases. Case 4's  $PSNR$  and  $SSIM$  are midway between case 2 and 3's.

Case 4 suggests the tiles database lacks in diversity for some of the highlight (brighter) returns which were not needed as much to capture cases 2 or 3, well. Consequently, the database was amended to include more tiles with bright returns.

Case 5 shows sand ripples which is not an unusual underwater feature. This case is interesting in that there is no one distinct target that the ATD could extract. Such cases may be completely ignored by the ATD. Sand ripples are a difficult background to identify targets from for ATD based on matched filters. The vector quantized pixel intensity distribution had the correct shape but, unlike the other 3 cases, its amplitudes were diminished and it had more resolution in the bins. It suggests that randomly distributed bright returns are difficult for the tiles database to match well to. For whatever reasons, the full image was only received for the shortest ranges. The case with only half an image still captured the correct shape and intensity albeit with lower resolution (as in case 2, only half the number of bins). The other noteworthy difference is that its  $PSNR$  (15.31) is the lowest of the 5 highlighted cases due to the lack of a singular bright target. Its  $SSIM$  (0.43) however, is the highest. The lack of a distinct target manifests in the  $PSNR$  trends not tracking the  $SSIM$  ones well.

As in the study in the literature survey, the  $PSNR$  values are less than the 20 dB threshold considered for 'good' recon-





**Fig. 12** Comparison of the original image against the received vector quantized version as a function of range (distance between acoustic modem transmitter and receiver). Fidelity of the reconstructed image

is quantified through the peak signal-to-noise ratio ( $PSNR$ ), structural similarity index ( $SSIM$ ) and histogram of pixel intensities. The whole image was transmitted for the entire range tested (case 4 from Fig. 9)

struction in monochrome pictures. This is due to the inherent high noise and low contrast and luminosity in side-scan sonar imagery.

The next step was to integrate these validated ROS nodes into a larger stand-alone node with the ATD. This provided a ROS node for a function that was directly integrated into the UUV payload autonomy. Then, an opportunity was sought to gather more SSS data files over different seabed types, propagation conditions, and sonar types to further test VQ applied to SSS and the amended database in particular.

### 4.3 Advanced tests and validation at-sea

Unmanned Warrior 2016 (UW16) was a Royal Navy (RN) hosted exercise at the British Underwater Test and Evaluation Centre (BUTEC) in Loch Alsh, Scotland. Participants achieved significant milestones and world firsts in collaborative robotics for NMCM. Underwater transmission of sonar images was one such milestone.

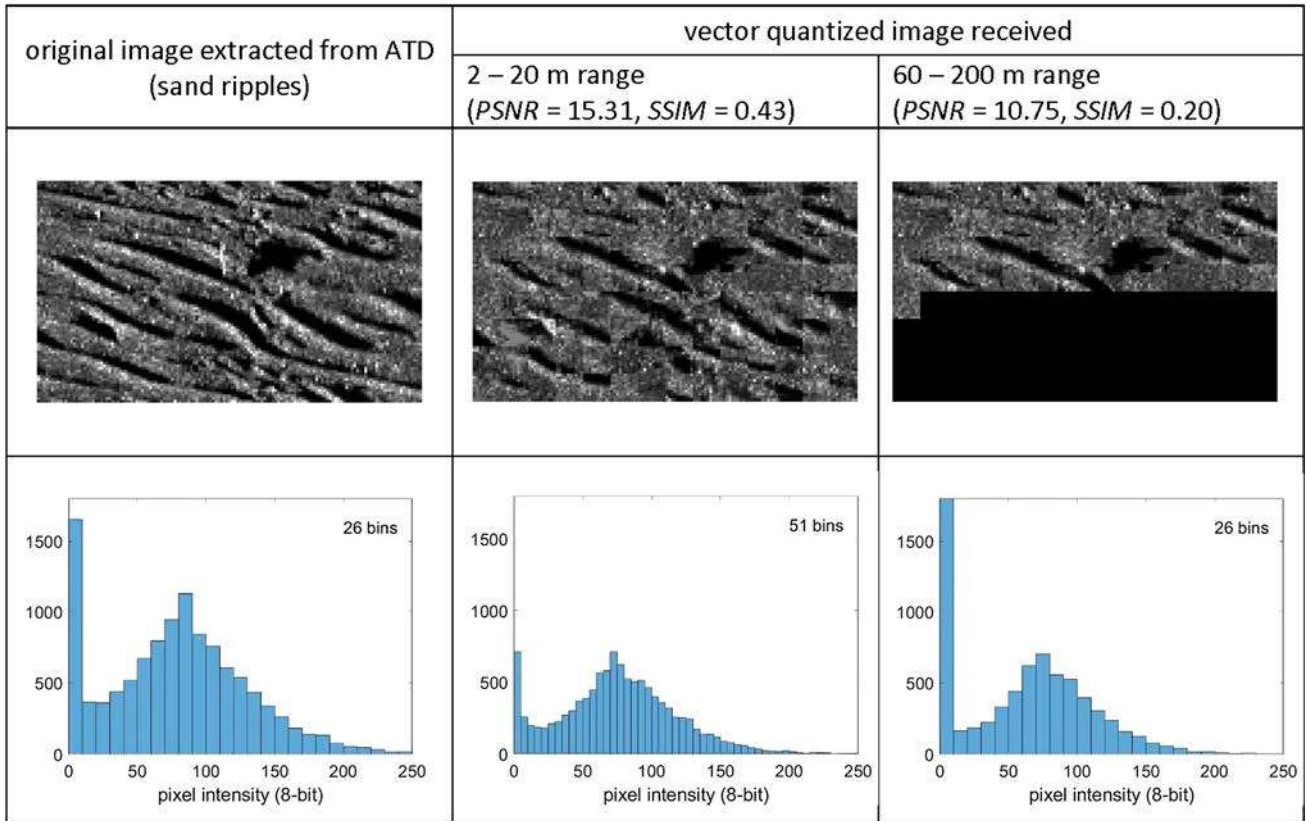
UW16 was a venue to gather side-scan sonar imagery from other types of SSS and test under other underwater acoustic propagation conditions. It was also an opportunity for experienced sonar and UUV operators to view the recon-

structed SSS images for some initial feedback. This part was not intended as a thorough investigation to explicitly address psycho-visual or human-robot aspects in recognizing MLOs.

This project's UW16 objective, to create a "UUV-in-a-box" (UIB), required the development of a stand-alone ROS node that encompassed all the ROS nodes described earlier including the ATD. Above and beyond what has already been described to this point, the UIB node needed to be developed. The full UIB node (Fig. 3) was finalized, tested and demonstrated at UW16.

The UIB node runs as a stand-alone ROS node that was integrated into the UUV's on-board autonomy (ATD) as a function. The ATD polls a specific directory for a new SSS data file (1000 pings in length). Once a new SSS data file was detected, the ATD processes the file to detect and extract targets. If targets were detected, their images were extracted, compressed and formed into acoustic modem messages. Then, the messages were queued for transmission to the operator.

At UW16 the UIB node read in and processed pre-logged SSS data files collected at UW16. These pre-logged sonar data files were collected over the duration of UW16. The side-scan sonars that contributed images were the MarineSonics



**Fig. 13** Comparison of the original image against the received vector quantized version as a function of range (distance between acoustic modem transmitter and receiver). Fidelity of the reconstructed image

is quantified through the peak signal-to-noise ratio ( $PSNR$ ), structural similarity index ( $SSIM$ ) and histogram of pixel intensities (case 5 from Fig. 9)

High Definition Sonar and the Klein 3500. They were payload sensors on-board REMUS 100 and IVER3 UUVs at UW16.

#### 4.4 Analysis of results

The results from the trial produced a large number of SSS data files and runs with more sonar brands, different acoustic propagation conditions and target sizes as desired.

Figure 14 presents image snippets that were identified and cropped by the on-board ATD from SSS data files, compressed by VQ, formed into a message and transmitted with the acoustic modem. Then, the messages were received and decompressed at the receiver end. The sonar files were automatically deposited in the polled directory at a rate consistent with what occurs on-board the UUV. At UW16, the receiver unit was an underwater acoustic modem deployed from the USV HMS Hazzard (Fig. 5) which had operators on-board. It received the packet, then decompressed and decoded it. Next, on-board operators analyzed the image and decided if it was an MLO. Once the operators on-board Hazzard examined the decompressed images, they were further relayed,

in-air, to a shore-based command and control center. This chain of events (Fig. 15) emulates the situation where the expert opinion of shore-based experts was sought for further target prosecution.

As in the initial at-sea tests, the  $PSNR$  and  $SSIM$  were used as similarity measures between the original ATD cropped images and the compressed ones that were transmitted, received and reconstructed. Also in the results from the initial at-sea tests, PSK-1 and PSK-4 were mainly used for the underwater modem transmission rates.

##### 4.4.1 Similarity across different sonars

Figure 14 shows example targets ordered by their  $PSNR$  values and which coincidentally falls along the lines of decreasingly smaller targets in the snippets. Interestingly enough, the trend in the  $SSIM$  values were inversely related to the  $PSNR$  ones. Not surprisingly,  $SSIM$  tended to be higher if there were larger and more distinct features in the image. With SSS the image size is controlled by the altitude of the sonar/UUV. This is usually set based on operational requirements. If the target of interest is small in the ‘field-

	original image (L)	received image (R)	PSNR	SSIM
1			16.91	0.35
2			17.01	0.34
3			17.53	0.32
4			18.27	0.32
5			18.29	0.32
6			18.44	0.31

**Fig. 14** Comparison of targets across two different sonars. The sonar image snippets are sorted by *PSNR*. Note there is not a lot of variation in the metrics

of-view’ then the *SSIM* value is not the best measure of similarity.

Having said that the variation in *PSNR* and *SSIM* was not large even across different types and resolutions (frequency) of side-scan sonars for this image set. This is likely due to the seabed being the same as the images were from the same area and the low contrast expected of SSS imagery.

#### 4.4.2 Operator evaluation

The decompressed images were initially presented to sonar and UUV operators who inspect SSS images for MLOs. In an average of 18 out of 20 images, 4 operators correctly recognized them as MLO (or not). This was not a user survey. It was a check that the compression algorithm and nominal database had some merit and that it was worth pursuing to the next stage with operators.

#### 4.4.3 VQ compression for transmission

Side-scan sonar imagery is amenable to compression using VQ techniques that were applied to underwater camera color pictures. The *PSNR* for transmittable SSS images were below the 20 dB threshold used for above-water pictures. However, no enhancements were applied to the ATD extracted image prior to compression or to improve the recon-

structed image. Despite that, the images were recognizable to the few (including the authors) who are experienced in inspecting SSS images for MLOs.

With the VQ compression, the results critically depend on the tile database diversity. Consequently, improvements in the database directly improve the transmitted images’ quality. The database can be adapted to target specific seabed types and offer flexibility to the algorithm. Using a very large database impacts the speed of the analysis. The nominal database formed used images from a wide variety of trials and seabed types from places around the world. It performed reasonably well against the seabed type experienced in Scotland without any amendments.

Based on the calculated *PSNR* and *SSIM* values as well as pixel density distributions, side-scan sonar MLO imagery compresses well with VQ (Murphy et al. 2010). With insight from this analysis, the *PSNR* and *SSIM* measures are reasonable measures for capturing side-scan sonar image similarity but there is motivation to consider other similarity measures for monochrome, noisy and low contrast images.

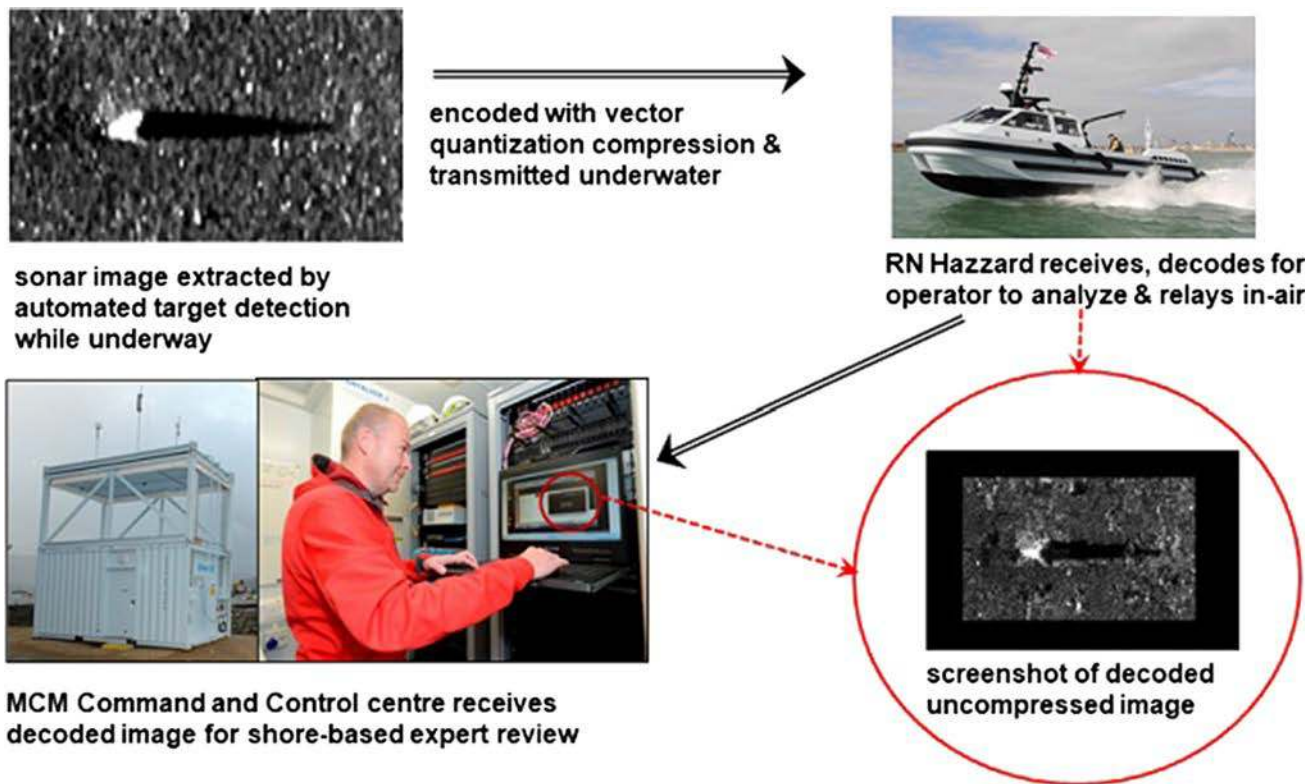
#### 4.4.4 Underwater acoustic propagation

Transmission and reception of the side-scan image snippets works well when the underwater acoustic communications channel was favorable for communications at 25 kHz. On days with high underwater acoustic ambient noise from winds and or motions of the acoustic modem in a sea state, packet loss occurred a little more frequently. At UW16, one of the days had higher winds (15 knots) and more sea state (sea state 2+) than in the initial at-sea tests which had none. Dropped packets were occasionally seen at shorter ranges (< 150 m) than in the initial tests. They are similar in effect to Fig. 9 (case 5) and Fig. 11. The acoustic packet’s form and transmission is otherwise robust and fits within a reasonably sized number of packets.

## 5 Concluding remarks

The focus was on overcoming challenges in communicating complex high bandwidth information across very different environments (under and above water). The work reported show that it is possible for marine robots (UUV, USV) to transmit and receive small sonar images underwater and to further relay them to above-water points Fig. 15.

A compression method appropriate for side scan sonar image snippets was identified, developed and implemented. The compression scheme was tested on side-scan sonar images collected from a variety of trials and sonars. Then, the sonar image transmissions were tested in a real at-sea environment. Finally, the compression and transmission scheme was integrated with ATD tools and tested, evaluated and



**Fig. 15** Example of underwater sonar image transmitted at Unmanned Warrior 2016. Note the similarities between the original sonar image (upper left) and the decompressed and reconstructed version that was transmitted underwater (bottom right). This is case 1 of Fig. 14

demonstrated at the Royal Navy Unmanned Warrior 2016 exercise.

SSS images are amenable to compression using vector quantization methods from underwater color pictures. However, there is degradation as Table 2 demonstrates that SSS imagery can be compressed at higher *PSNR* and *SSIM* that are well above the 20 dB threshold but at a cost of not being transmittable through low bandwidth modems.

## 6 Future work

To capture an image to the fidelity needed a good approach is to learn the basis from the target itself. Work is underway to apply machine learning to optimize the database quality while building it in the lab (DeBortoli et al. 2017). Additionally, the *FOM* evaluation could be used in-situ to do small incremental on-line builds of the database. The tiles database should adapt, if needed, to a new area by learning from repeat dives in that area. A more powerful payload processor is being considered for the UUV however there are implications on robot endurance from this.

The second objective for the next stage work is to address the psycho-visual aspects. The focus of the work reported was on pixel-based approaches of monochrome atypical tar-

gets being evaluated, without augmentation at the transmit or receive end, by specially trained sonar operators. Their psycho-visual perceptions are different. They are more like those specifically trained for medical ultrasound. This could be better addressed through a targeted user study of sonar operators of all types.

The work of Zhang et al. (2015) may be investigated initially for its ability to be implemented on an embedded processor on-board a robot.

**Acknowledgements** This project is grateful for the timely and helpful advice and assistance of the WHOI Acoustic Communications Group.

## 7 Appendix

### 7.1 Implementation as ROS nodes

ROS (robotic operating system) is a widely supported open source middleware for robotic systems (Purvis 2017). It provides operating system services, including hardware abstraction, low-level device control, implementation of commonly used functionality, message-passing between processes, and package management. It also provides tools and libraries to build, write, and run code across multiple platforms. It is

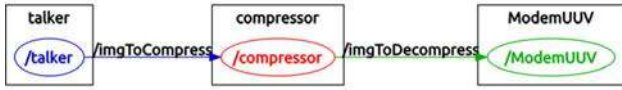


Fig. 16 Interconnection of nodes on the transmit (UUV) unit

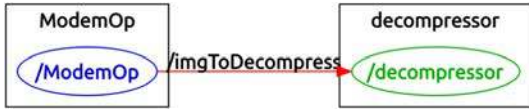


Fig. 17 Interconnection of nodes on the receive (operator) unit

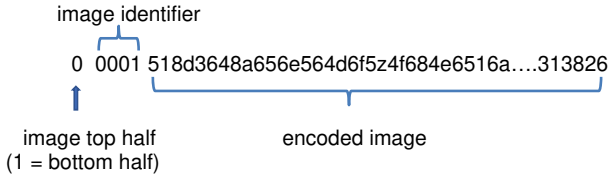


Fig. 18 Example of a message in an acoustic packet as formed by ModemUUV node

light weight enough that it is widely applied in embedded systems. ROS works as a publish-subscribe architecture.

ROS is used by Defence R&D Canada (DRDC) as a middleware for its IVER3 UUVs’ on-board autonomy. Implementation of the compression and transmit algorithm as ROS nodes facilitates rapid integration into this autonomy.

There are nodes on the transmit (UUV) side and the receive (operator) side. On the transmit (UUV) side, the compression algorithm consists of 3 ROS nodes described next and illustrated as a ROS *rqt\_graph* in Fig. 16.

- *talker* Tasked with publishing the path of the images to compress on the dedicated topic *imgToCompress*. Initially, it simulated the ATD algorithm detecting MLO images from the sonar data stream. Later, it was directly integrated with the ATD
- *compressor* Subscribes to the topic *imgToCompress*. Each time it receives a new image path it compresses the image and publishes the encoded image on the topic *imgToDecompress*.
- *uuvModem* This node subscribes to the topic *imgToDecompress*. It is connected to the acoustic modem through a serial port. For each encoded image it receives, it sends it through the underwater acoustic link using the appropriate messages.

On the receive (operator) side (Fig. 17) are two nodes:

- *opModem* This node is connected to the receiving acoustic modem through a serial port. It ‘listens’ to all messages sent by the UUV’s modem to find the packet that contains the encoded image. When that message is found, it

extracts the content of the message and publishes it on the topic *imgToDecompress*.

- *decompressor* This node subscribes to the topic *imgToDecompress* and when a new encoded image is received, it decodes it and saves the reconstructed image to the /out folder.

## 7.2 Underwater acoustic communications

The acoustic modems can transmit/receive at several discrete rates that vary with packet size. For the initial in-laboratory tests, rate 4 (Table 1) was chosen so an encoded image could fit into the minimum number of packets of one frame each. This makes it easier to perform the nodes’ algorithmic proof-of-principle tests at the higher risk of losing packets. Losing the occasional packet was not a primary concern for these tests. Packet loss will be addressed in the at-sea tests. The proof-of-principle tests were performed in an indoor water tank where the transmit and receive acoustic modems were 2–3 m apart which limits packet loss. The final at-sea implementation will be performed at rate 1 which is more reliable albeit at a cost of smaller packets which means an image would be transmitted over multiple frames. This will be explained in a later section. The maximum size of a frame at rate 4, is 256 bytes so an encoded image could be transmitted over two packets. This means half of one encoded image, at a time, would be published on the topic *imgToDecompress*. As shown in Fig. 18, the message published contains a header with the first character identifying whether the packet contains the top or bottom half of an image followed by the image identifier coded on four characters to recognize the origin of the half image. The rest of the message is half of the encoded image.

The *uuvModem* node is connected to the acoustic modem through a serial port, which was configured in write-only mode. At initialization, it gives the acoustic modem an identifier to send messages in the later stages. The syntax to set the identifier to ‘2’ is:

$$\text{\$CCCFG, SRC, 2}\backslash\text{r}\backslash\text{n.} \quad (1)$$

Then, when a message is published on the topic it writes the following instruction to the serial port:

$$\text{\$CCCYC, 0, 2, 1, 4, 0, 1}\backslash\text{r}\backslash\text{n.} \quad (2)$$

This informs the acoustic modem that a packet will be sent from modem number 2 to modem number 1 with acoustic modem rate 4 containing 1 frame. The modem then replies with a query for the packet to send with:

$$\text{\$CADRQ, < Time of request >, 2, 1, 0, 64, 1}\backslash\text{r}\backslash\text{n.} \quad (3)$$

The user then provides the hex-encoded message to the modem:

\$CCTXD, 2, 1, 0, < HexData > \r\n. (4)

On the operator side (Fig. 17), the *ModemOp* node configures the connection with the modem through the serial port to read and write mode. At initialization of the node, it will set the identifier for the operator's modem to '1' using a command similar to (1) above. Then, the node listens to the modem output through the serial port. If it reads a CARXD message it means a packet was received. The content of that packet is read and published on the topic *imgToDecompress*.

The *decompressor* node is subscribing (listening) to the topic *imgToDecompress*. When a message is published on that topic, this node will identify (1) whether it is for the top or bottom half of an image and (2) its unique identifier from the message header. If it previously received the other half of the image, it will concatenate the two and decompress the image using functions from the **VQScheme** class. But, if the other half came from a different image, then the older packet will be decompressed after zero padding its missing half. The last packet received will be stored to wait for its other half for decompression. If no other half was received, the node will wait for the next packet for decompression.

## References

- Atallah, A. M., Ali, H. S. & Abdallah, M. I. (2016). An integrated system for underwater wireless image transmission. In *Proceedings of the 28th international conference on microelectronics* (pp. 169–172).
- Bhattacharyya, P., Mitra, A., & Chatterjee, A. (2014). Vector quantization based image compression using generalized improved fuzzy clustering. In *Proceedings of the 2014 international conference on control, instrumentation, energy and communication* (pp. 662–666).
- Bin, L., & Qinggang, M. (2013). An improved SPIHT wavelet transform in the underwater acoustic image compression. In *Proceedings of 2013 2nd international conference on measurement, information and control (ICMID)* (Vol. 2, pp. 1315–1318).
- Chen, W., Yuan, F., Cheng, E., & Lin, W. (2017). Subjective and objective quality evaluation of sonar images for underwater acoustic transmission. In *Proceedings of the 2017 IEEE International Conference on Image Processing* (pp. 176–180).
- Committee for Mine Warfare Assessment, Naval Studies Board, National Research Council, Naval Mine Warfare: Operational and Technical Challenges for Naval Forces, <http://acomms.whoj.edu/wp-content/uploads/sites/20/2014/09/401040-SIG-Micromodem-Software-Interface-Guide.pdf>. (2001).
- DeBortoli, R., Nicolai, A., Li, F., & Hollinger, G. (2017). Assessing perception quality in sonar images using global context. In *Proceedings of Intelligent Robots and Systems Conference* (pp. 1–8).
- Eigen Library for Matrix Algebra in C, <https://www.quantstart.com/articles/Eigen-Library-for-Matrix-Algebra-in-C>. Accessed December 2016.
- Ferdowski, S., Voloshynovskiy, S., & Kostadinov, D. (2017). A multi-layer image representation using regularized residual quantization: Application to compressing and de-noising. In *Proceedings of the 2017 international conference on Image Processing* (pp. 2697–2701).
- Hinton, G., Vinyals, O., & Dean, J. (2014). Distilling the knowledge in a neural network. In *Proceedings of NIPS Deep Learning Working* (pp. 1–9).
- Huffman, D. A. (1952). A method for the construction of minimum-redundancy codes. In *Proceedings of I.R.E.* (pp. 1098–1101).
- Kalwa, J., & Madsen, A. L. (2004). Sonar image quality assessment for an autonomous underwater vehicle. In *Proceedings World Automation Congress* ( Vol. 15, No. 33–38).
- Li, J., Wang, W., Li, K., Meng, W., & Guo, X. (2018). Application of image quality assessment in marine biological density prediction. In *Proceedings of Chinese Control and Decision Conference* (pp. 6658–6661).
- Liu, P., & Li, G. (2013). An improved SPIHT algorithm for image compression in low bit rate. *Communications and Network*, 5, 245–248.
- Moszynski, M., Chybicki, A., Kulawiak, M., & Lubniewski, Z. (2013). A novel method for archiving multibeam sonar data with emphasis on efficient record size reduction and Storage. *Polish Maritime Research*, 20(1), 77–86.
- Murphy, C., Wang, R. Y., & Singh, H. (2010). Seafloor image compression with large tile size vector quantization. In *IEEE/OES autonomous underwater vehicles* (pp. 1–8).
- Nams, D., & Seto, M. L. (2015). On-line adaptation of underwater acoustic transmission rates to optimize communications for collaborative AUV missions. In *Proceedings of 2015 MTS/IEEE Oceans Conference* (pp. 1–8).
- Purvis, M. (Ed.). Indigo—ROS wiki, edited 2014-08-22, <http://wiki.ros.org/indigo>. Accessed May 2017.
- Said, A., & Pearlman, W. A. (1996). A new, fast, and efficient image codec based on set partitioning in hierarchical trees. *IEEE Transactions on Circuits and Systems for Video Technology*, 6(3), 243–250.
- Shen, J. & Huang, H.C. (2010). An adaptive image compression method based on vector quantization. In *IEEE 1st international conference on pervasive computing, signal processing and applications* (pp. 377–381).
- Skodras, A., Christopoulos, C., & Ebrahimi, T. (2001). The JPG 2000 still image compression standard. *IEEE Signal Processing Magazine*, 18(5), 36–58.
- Wallace, G. K. (1992). The jpeg still picture compression standard. *IEEE Transactions on Consumer Electronics*, 38(1), xviii–xxxiv.
- Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 1–14.
- WHOI Acoustic Communications Group: Micro-Modem Software Interface Guide, Document ID: 401040, Version 3.25. (2014).
- Yang, M., & Bourbakis, N. (2005). An overview of lossless digital image compression techniques. In *IEEE* (pp. 1099–1102).
- Zhang, Y., Li, Q. Z., & Negahdaripour, S. (2015). Seafloor image compression using hybrid wavelets and directional filter banks. In *Proceedings of the MTS/IEEE Oceans Conference* (pp. 1–9).
- Zhang, Y., Negahdaripour, S., & Li, Q. (2016). Low bit-rate compression of underwater imagery based on adaptive hybrid wavelets and directional filter banks. *Signal Processing: Image Communication*, 47, 96–114.