UW-SVC: Scalable Video Coding Transmission for In-network Underwater Imagery Analysis

Mehdi Rahmati and Dario Pompili

Department of Electrical and Computer Engineering, Rutgers University–New Brunswick, NJ, USA

Emails: {mehdi.rahmati, pompili}@rutgers.edu

Abstract—Underwater imagery has enabled numerous civilian applications in various domains, ranging from academia to industry, and from industrial surveillance and maintenance to environmental protection and behavior of marine creatures studies. The accumulation of litter and plastic debris at the seafloor and the bottom of rivers are extremely harmful for the aquatic life. We propose a solution for monitoring this problem using a team of Autonomous Underwater Vehicles (AUVs) to exchange the recorded video in order to reconstruct the map of regions of interest. However, underwater video transmission is a challenge in the harsh environment in which radio-frequency waves are absorbed for distances above a few tens of meters, optical waves require narrow laser beams and suffer from scattering and ocean wave motion, and acoustic waves—while long range—provide a very low bandwidth and unreliable channel for communication. In our solution, the scalable coded video of each vehicle is shared in-network with a selected group of receiving vehicles through the underwater acoustic channel. Presented evaluations, including both simulations and experiments, confirm the efficiency and flexibility of the proposed solution using acoustic software-defined modems.

Index Terms—Underwater Networks; Acoustic Communications; Broadcasting; Scalable Video Coding (SVC).

I. INTRODUCTION

Overview: Marine litter and debris, including both beached and floating objects, is one of the most serious and fast growing environmental threats in the oceans and seafloors. The negative impacts of litter accumulation on the aquatic life are unquestionable. Litter is spread widely throughout the seafloor, but its distribution is usually patchy with densities from 1 item up to around 200 items per each 10 m, as reported for Messina Strait's channels (one of geologically active areas of the Central Mediterranean Sea) [1]. Rivers are one of the main sources of entering litter to the seas, since they carry the litter with their currents to the sea or ocean. Deploying a team of Autonomous Underwater Vehicles (AUVs), equipped with down-looking cameras, can help in detecting these objects on the seafloor and riverbed, build a map of the pollution, and therefore, can issue early warnings so to reduce the damage to human and aquatic life. However, coordination among multiple AUVs is a challenge [2], specially when video is the subject of data exchange. AUVs should be able to encode the video, and to transmit it to other vehicles (generally to heterogeneous dynamic nodes) efficiently [3]. There are still open problems in near-real-time underwater video processing and transmission.

To achieve these goals, novel efficient mechanisms and hardware should be utilized to make the video transmission feasible for underwater scenarios. Boosting the data rate and system reliability is possible if all the available domains are exploited in an efficient manner [4]. To stream and transmit underwater video, we require reliable and robust techniques in an environment, in which Radio Frequency (RF) waves are absorbed for distances above a few tens of meters, optical waves require narrow laser beams and suffer from scattering and ocean wave motions, and acoustic waves—while being able to propagate up to several tens of kilometers—lead to a communication channel that is very dynamic, prone to fading, spectrum limited with passband bandwidths of only a few tens of kHz due to high transmission loss at frequencies above 50 kHz, and affected by the colored ambient noise.

Motivation: Traditional commercial acoustic modems with their fixed-hardware designs hardly meet the required data-rate and flexibility to support the futuristic underwater multimedia applications. Over the past few years, novel solutions based on adaptive and reconfigurable architectures—i.e., Software Defined Acoustic Radios (SDAR)—have been proposed. Using SDAR helps the scientists and engineers to explore different protocols and techniques on a single hardware, perform in-network analysis, and transmit the high-volume data, such as video, to a remote node depending on environment and system specifications. This concept is changing the business model of commercial acoustic modems in a near future since they are focusing more on efficient hardware/architectures and proprietary high-performance algorithms [5].

Furthermore, using conventional video compression/encoding techniques will not meet the requirements for these futuristic underwater video transmissions due to the need for higher data rate and more reliability. Therefore, more reconfigurable and flexible techniques should be utilized to address this problem. In practice and in many underwater imagery/streaming applications, since the visual depth of the camera is limited in the water, the vehicle should get close enough to the target to be able to detect it, therefore, usually a single vehicle/camera can not cover the whole scene (because of the limitation in the field of view and visual depth) and can not create the global map of the environment. We will also address the challenge of coordination among the underwater vehicles in this paper.

Our Vision: We propose a solution to encode and share the video among AUVs until the global information/reconstruction of the region of interest is achieved. Scalable Video Coding (SVC) [6], as the extension of H.264/MPEG-4 AVC, offers the required flexibility by encoding the chunks of video
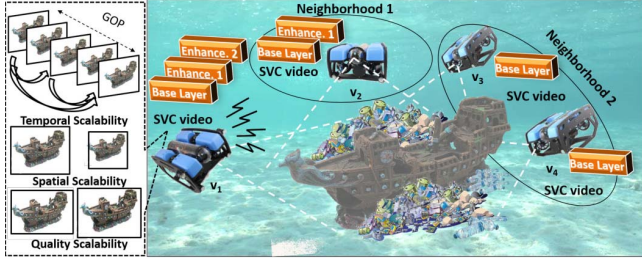
IEEE computer society

Fig. 1. System model of the proposed Scalable Video Coding (SVC)-based video transmission among a team of underwater vehicles (with the help of high-performance modified vehicles, BlueROV2 [7]) which are used for video capturing in marine litter detection missions. Video is encoded via a base-quality layer and l enhancement layers and is shared separately with each neighborhood using temporal, spatial, and quality scalabilities in SVC.

into a base layer and multiple enhancement layers given the requirements of the underwater channel. Fig. 1 shows our vision including multiple vehicles around a pile of objects. SVC base layer provides the minimum required quality, while enhancement layers offer a more enhanced quality based on different modalities——temporal scalability (frame rate), spatial scalability (frame size), and quality scalability (fidelity or SNR)——which makes this encoding a good choice for lossy video compression and erroneous transmission environments such as underwater. Here, a group of independent frames in the video structure is represented by a Group of Pictures (GOP) in the figure. Efficient video coding and reliable communications solutions are demanded for the coordination and communications among the vehicles. The reconstructed map can be used for in-network decision among the vehicles or can be transmitted to the buoy for further considerations.

Our Contributions: In many applications, more than one vehicle, due to the limited field of view and the visual depth of camera in the water, are needed to merge the video from different angles so as to reconstruct the map of region of interest. In this paper, we focus on in-network *scalable* underwater video sharing between AUVs and offer these contributions:

- A framework for underwater imagery analysis using partial information collected by various vehicles around the scene;
- An optimized solution to provide the maximum possible Quality of Service (QoS) via a proposed multicasting scalable coded video, while achieving the maximum Quality of Experience (QoE) for the scene reconstruction;
- Performance evaluation of this system with comprehensive simulations under different scenarios using real videos captured from the Raritan river-New Jersey and through an SDAR testbed.

Paper Organization: In Sect. II, we go over the state of the art in underwater video transmission. In Sect. III, we present our solution and discuss scalable video coding and the required optimizations. In Sect. IV, we evaluate our solution via the experiments and simulations, and then scale the results via simulations. Finally, in Sect. V, we draw the main conclusions and present the future work.

II. RELATED WORK

Underwater Video Transmission: There are several unique characteristics of underwater wireless networks that make Quality of Service (QoS) delivery of video content—ranging from delay sensitive to delay tolerant, and from loss sensitive to loss tolerant—a challenging task due to underwater acoustic frequency-dependent transmission loss, colored noise, multipath, Doppler frequency spread, high propagation delay as discussed in [3], [8]. The multiview video transmission in underwater acoustic path is discussed in [9] in which the authors propose time-shifted transmission slots to the encoder and other nodes to exchange control and video packets. The feasibility of transmitting video over short-length underwater links is investigated in [10], [11], where MPEG-4 video compression and a wavelet-based transmission method are tested on the coded Orthogonal Frequency Division Multiplexing (OFDM). Despite all these works, the problem of robust video transmission is still unsolved, and achieving high video quality is still a challenge when we consider the limited available bandwidth along with the harsh characteristics of the underwater acoustic channel, *which calls for novel high-spectral-efficiency in-network collaborative methods.* In the area of underwater video, [12] shows the feasibility of video streaming using currently commercially available hardware defined modems. The reconstructed objects can be used in Simultaneous Localization And Mapping (SLAM). SLAM is a widely used technique in ground robots, but less feasible in underwater environment specially in high turbidity situations and in the absence of reliable static landmarks. Some underwater visual SLAM solutions, such as in [13], create a sparse map for the navigation and localization in clear water.

Scalable Video Coding (SVC): SVC [6] outperforms the regular H.264 encoding when more flexibility and adaptation to the channel's condition are required [14]. In the area of SVC, previous papers have touched on video sharing/multicasting in terrestrial context. A method for adapting the number of layers based on a fixed time allotment is proposed in [15], This link-level method does not explore a multicast scenario. The authors in [16] explore dynamic layer adjustment in a content-delivery context where a direct-download system is paired with peer-to-peer. This sharing is top-down content delivery, rather than a scheme for in-group video sharing where each consumer is also a producer. A method for SVC video transmission is proposed in [17] using transmitter-side distortion estimates based on the channel state information. However, *none of these methods tackle the unique challenges faced in an underwater acoustic channel.*

An adaptive distortion-rate tradeoff for underwater video transmission using a Multi-input Multi-output (MIMO)-based SDAR system is proposed in [18]. The scalability of the system is fulfilled using SVC compression standard. In [4] a new signaling for SVC-encoded underwater videos is proposed based on using non-contiguous OFDM and beamforming techniques with the help of Acoustic Vector Sensors (AVSs).

III. OUR SOLUTION

In this section, we present our solution for in-network video sharing and coordination among multiple AUVs. In Sect. III-A, we discuss the construction of SVC-encoded video streams and the proposed strategy to estimate the optimal parameters given underwater acoustic channel constraints as it will be explained in the optimization problems. In Sect. III-B, we present our SVC-based multicasting solution to increase the overall quality of video. In Sect. III-C, the proposed protocol will be presented for an efficient map reconstruction while multiple vehicles are involved in the merging process.

A. Construction of SVC-encoded Video Streams

Encoding the original video into several layers using SVC discards the need for transcoding or re-encoding the video. However, an efficient strategy is required to leverage the scalibiliteis of SVC and adapt the encoder to the receiver's status as well as the quality of acoustic channel.

Video Sharing Setup: Assume V vehicles are deployed around a scene, as shown in Fig. 1, at time slot t and form a wireless network of (V, H), where H stands for the point to point link between two vehicles, when vehicles are in the communications range of each other. Vehicles encode the initial video using SVC, and make it ready for broadcasting. To facilitate the communications, vehicles set up a basic Time Domain Multiple Access (TDMA) system and assign a time slot to each vehicle since the network size is small in underwater scenarios and the nodes are usually close together. The underwater acoustic channel presents problems for a coordinated and synchronized system such as TDMA, but due to the severe bandwidth constraint, it is important to use a Medium Access Control (MAC) that does not constrain vehicles to an even smaller slice of bandwidth, such as FDMA. Authors in [19] show that even in the underwater acoustic environment, and specially for multicast transmissions, TDMA can allow for efficient and collision-free communications. Other random- and controlled-access MAC solutions such as Carrier-sense Multiple Access (CSMA) transmit multiple packets through the same underwater channel, which might lead to packet collisions at the receiver [2]. To address the synchronization problem in TDMA (as the main weakness of using TDMA underwater), we use an unsynchronized MAC protocol, e.g., Tone Lohi (T-Lohi) [20], especially in sparse networks with limited number of nodes. The vehicles start contending any time they realize the channel is not occupied.

Base-layer Video Sharing: Assume each vehicle records the scene from its own angle and possibly it has an overlapping coverage with other vehicles. SVC-based video is segmented into C chunks in each vehicle $j \in \{1, ..., V\}$ with a base layer b_j (layer 0) with the rate $R(b_j)$ and $l_j \in 1, 2, ..., L_j$ enhancement layers with rate $R(l_j)$. Each node broadcasts the chunks of its base layer video through an acoustic channel. When a vehicle i receives the base layer data of chunk $c \in C$ in time slot t from transmitting vehicle $j \in \{1, ..., V\}$ and $j \neq i$ in the communication range, the received signal can be expressed as $y_i^c(t) = h_{ij}^c(\tau_{ij}; t) * b_j^c(t) + z_i(t)$, where

$h_{ij}^c(\tau_{ij}^c; t)$ stands for the channel coefficient with delay τ_{ij}^c between vehicles i and j, $y_i^c(t)$ represents the received signal, $*$ stands for the convolution operation, and $z_i(t)$ shows the background underwater colored noise. For a band-limited non-ideal underwater channel with the frequency response of $H_{ij}^c(f)$ and a Gaussian noise with the power spectral density of $S_i(f)$, the capacity $\mathcal{C}$ of each channel can be expressed as [21],

$$\mathcal{C}_{ij}^c = \frac{1}{2} \int_{-\infty}^{\infty} \log\left(1 + \frac{P_j^c(f)\mid H_{ij}^c(f)\mid^2}{S_i(f)}\right) df. \quad (1)$$

Here, $P_j^c(f)$ stands for the power spectral density of b_j^c from transmitting vehicle j in chunk c. We drop time index t for the sake of simplicity and present our analysis for the time length of chunk c. Assume Channel State Information (CSI) is available at the transmitter and the channel is constant during broadcasting of a video stream in chunk c and B_W represents the channel bandwidth, which is assumed to be the same for all the users. The base layer data rate $R_{ij}(b_j^c)$ can be expressed as $R_{ij}(b_j^c) = B_W \mathcal{C}_{ij}^c$. We consider the tradeoff between the transmit power and data rate for a fixed bandwidth B_W in each vehicle j such that the outage does not occur. Since we assume each vehicle j broadcasts its data to all other vehicles in its neighborhood through *independent channels*, the broadcast data rate $\boldsymbol{R}_j(b_j)_{BC}$ for all chunks can be bounded as follows.

$$\boldsymbol{R}_j(b_j)_{BC} = \{R_{ij}(b_j^c) : R_{m,j}^*(b_j) < R_{ij}(b_j^c) < \mathbb{E}[\mathcal{C}_{ij}]\}. \quad (2)$$

In this equation, $\mathbb{E}[.]$ represents the expectation operator, $\boldsymbol{R}_j(b_j)_{BC}$ stands for the practical transmission rate for broadcasting, and $R_{m,j}^*(b_j) \in \boldsymbol{R}_j^*(b_j) = [R_{1j}^*(b_j), ..., R_{V-1j}^*(b_j)]$ shows the minimum rate required in all fading situations [22] for $V - 1$ receiving vehicles to avoid an outage.

In practical scenarios, in which the CSI is not fully known at the transmitting vehicle and channel gains are not known in advance, we assume that the transmitting vehicle j statistically knows the ordering of the other vehicles for each chunk c in time slot t in terms of their instantaneous channel gains, i.e., $\mid h_{1j}^c \mid < \mid h_{2j}^c \mid < ... < \mid h_{3j}^c \mid$, for receiving vehicles $1, ..., V - 1$, from weak to strong. The broadcast channel can be considered as a multiple-component channel such that a weaker component is a degraded version of the other component in a symmetric broadcast channel. It can be proved that the vehicles have the same channel quality and hence could decode the broadcast data. Here, the fading statistics are assumed to be symmetric. Considering the principle of ergodicity, if an arbitrary user k can decode its data reliably, then we can conclude all the other users should be able to decode the broadcast data in the same way. This assumption breaks in the asymmetric fading case in which the users have different fading distributions. Therefore, sorting is not possible which leads to a non-degraded channel [23, Ch. 6].

We optimize the total rate for broadcasting from vehicle j

382

to other vehicles via the following optimization problem.

$$\underset{p_j}{\text{maximize}} \quad \mathbb{E}\left[\sum_{\substack{i=1 \\ i \neq j}}^{V} \alpha_i \log\left(1 + \frac{p_j^c |h_{ij}^c|^2}{s_i}\right)\right], \tag{3a}$$

$$\text{s.t.} \quad p_{th} \leq p_j^c \leq p_{max}, \quad \forall j \in \{1, ..., V\}, \tag{3b}$$

$$\boldsymbol{R}_j(b_j)_{BC} \succeq R_j^*(b_j)\mathbf{1}, \tag{3c}$$

$$\boldsymbol{R}_j(b_j)_{BC} \preceq \mathbb{E}[\boldsymbol{\mathcal{C}}_j]. \tag{3d}$$

Here $\alpha_i \in \{0, 1\}$ is the weighting factor, which is defined in the multicasting strategy, p_{th} and p_{max} show the minimum and maximum transmit power, respectively. $\mathbf{1}$ stands for an all-one vector, i.e., a vector whose entries are all equal to one, $\succeq$ and $\preceq$ represent the component-wise inequality. The capacity $\boldsymbol{\mathcal{C}}_j$ stands for the vector of all capacities to the receiving vehicles.

The optimization problem presented in (3) is a convex problem, since the objective function and the constraints are convex/concave; $\log(1 + p_j^c |h_{ij}^c|^2/s_i)$ is concave because it is the composition of a concave function ($\log$) with an affine mapping of p_j^c. Moreover, the non-negative weighted sum preserves the convexity (concavity) and the expectation of a convex (concave) function is convex (concave) [24]. Furthermore, the constraints are all affine.

In a broadcast scenario, each transmitting vehicle propagates its base layer video to all the receiving vehicles, since decoding the base layer is independent of other enhancement layers. However, the optimized data rate, calculated in (3), might not be sufficient for a higher quality video through the enhancement layers. Each enhancement layer l_j with a defined encoding rate of $R(l_j)$ can be decoded when firstly it is received reliably and secondly its lower layer $l_j - 1$ is successfully decoded, i.e., in other words, unsuccessful decoding of the lower layers leads to a failure in decoding the current layer.

B. Multicasting for Enhancement-layer Video Sharing

In a multicast scenario and due to heterogeneity of underwater nodes, we assume the nodes with poor channel quality are able to decode the video with the base layer (as discussed in Sect. III-A), while the nodes with a better communications channel quality can be served by a scalable video with a higher quality, i.e., with more enhancement layers. To be able to send the enhancement layers, we propose a broadcasting strategy in which the vehicles with the worst channel are shut down in the broadcasting, i.e., $\alpha_i = 0$ in (5a), in order to increase the total transmission data rate. Therefore, a pseudo-multicasting network is created. Apparently, the more vehicles with impaired channels are shut down, the more enhancement layers can be transmitted to the remaining vehicles and therefore video QoS increases.

On the other hand, since the vehicles are at different locations around the scene with different viewpoints (as it is depicted in Fig. 2), shutting them down, leads to lack of observation and so it results in losing some information while the map is reconstructed. Map reconstruction requires a good amount of Fields of View (FoV) overlap among the vehicles.
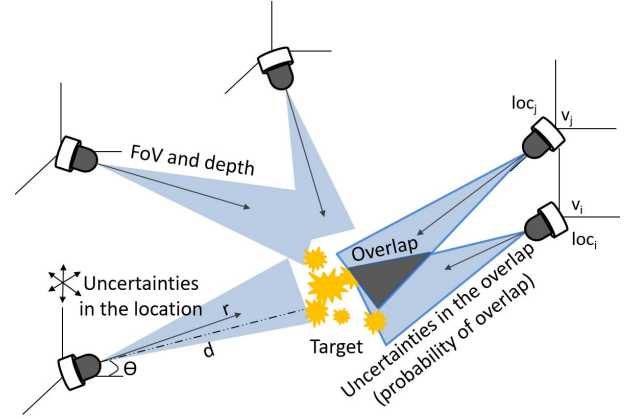

Fig. 2. Schematics of the potential overlap between the vehicles considering the uncertainties in the location of vehicles.

Assume the vehicles' cameras have some degrees of spatial correlation, as shown in Fig. 2, which is identified via the vehicles' configuration, i.e., area of overlap between FoVs of two cameras [25]. The FoV of cameras is limited to the area they observe, therefore, the information they get is directly related to the directional sensing and configuration of the vehicle. This overlap is used by the algorithm as a measure to shutdown the redundant vehicles if there exists a sufficient overlap for map reconstruction.

Let the FoV model of vehicle i, after 3-D to 2-D projection and calibration, be described by $(loc_i, r_i, \vec{D}_i, \beta_i)$ as in [26], in which loc_i stands for the location of the vehicle, r_i represents the sensing radius of the camera, $\vec{D}_i$ indicates the sensing direction (i.e., the center line of sight of the camera's FoV), and β_i is the offset angle. A model for the spatial correlation can be derived based on the above parameters as follows. Suppose vehicles i and j are two arbitrary vehicles that observe an overlapped area of interest; their disparity function δ (complementary to the correlation coefficient η as $\delta = 1 - \eta$) is defined as follows [26]:
$\delta = \frac{1}{4}\left(\left|\frac{d\sin\theta}{d+\cos\theta}\right| + \left|\frac{d\sin\theta}{d-\cos\theta}\right| + \left|\frac{d\cos\theta}{d+\sin\theta} - 1\right| + \left|\frac{-d\cos\theta}{d-\sin\theta} + 1\right|\right)$,
where d denotes the camera depth (here, the difference between the loc_i and the target's location assuming the camera sensing direction $\vec{D}_i$ is headed to the target) and θ is the angle between the sensing direction and the x-axis, so that the location loc_i can be expressed by $(-d\cos\theta, -d\sin\theta)$ after the 2-D projection. Specifically, for two vehicles i and j with parameters (d_i, r_i, θ_i) and (d_j, r_j, θ_j), respectively, the disparity between their images can be calculated as follows [25], [26],
$\delta_{i,j} = \frac{1}{4}\left(\left|\frac{-d_i\sin\theta_i - r_i\cos\theta_i}{d_i+\cos\theta_i} - \frac{-d_j\sin\theta_j - r_j\cos\theta_j}{d_j+\cos\theta_j}\right| + \left|\frac{d_i\sin\theta_i + r_i\cos\theta_i}{d_i-\cos\theta_i} - \frac{d_j\sin\theta_j + r_j\cos\theta_j}{d_j-\cos\theta_j}\right| + \left|\frac{d_i\cos\theta_i - r_i\sin\theta_i}{d_i+\sin\theta_i} - \frac{d_j\cos\theta_j - r_j\sin\theta_j}{d_j+\sin\theta_j}\right| + \left|\frac{-d_i\cos\theta_i + r_i\sin\theta_i}{d_i-\sin\theta_i} - \frac{-d_j\cos\theta_j + r_j\sin\theta_j}{d_j-\sin\theta_j}\right|\right)$. However, finding the exact amount of correlation might not be feasible due to the position uncertainty of the vehicles and the effect of currents on the

vehicles due to vehicle's drifting. Therefore, inaccuracies in position estimation increases and it becomes worse over time when the vehicle stays longer underwater, which leads to non-negligible drifts in the vehicle's position and thus making the camera overlap accurate calculations inapplicable.

In [2], an approach has been proposed to estimate vehicles' position through a statistical method based on the vehicles' confidence region. Assume each vehicle i measures N random samples of its location as $\{loc_i^{(n)}\}_{n=1}^N$. The measured locations are samples of a normal distribution $\mathcal{N}(\mu_i, \sigma_i^2)$ with the mean and variance μ_i and σ_i^2, respectively. The samples also follow a normal distribution with mean μ_i' and variance $\sigma_i'^2$. It can be inferred that $\frac{\mu_i' - \mu_i}{\sigma_i'^2/\sqrt{N}}$ is a pivot and it has a student's t-distribution with $N-1$ degrees of freedom. The mean $\mu_i' = \sum_{n=1}^N loc_i^{(n)}/N$ and the variance can be estimated as $\sigma_i'^2 = 1/(N-1)\sum_{n=1}^N \left(loc_i^{(n)} - \mu_i'\right)^2$ [2]. The uncertainty region, i.e., confidence interval, of this vehicle can be derived as $\Pr(L_i \leq \mu_i' \leq U_i) \geq 1-\gamma$. Here γ is the confidence degree, $\Pr(.)$ represents the probability function, and L_i and U_i are the interval boundaries of vehicle i and are estimated as $L_i = \mu_i' - \mathcal{T}_{(N-1,\alpha/2)}\sigma^2/\sqrt{N}$ and $U_i = \mu_i' + \mathcal{T}_{(N-1,\alpha/2)}\sigma_i^2/\sqrt{N}$. Here, $\mathcal{T}_{N-1,\alpha/2}$ is the t-distribution critical value with $N-1$ degrees of freedom. To estimate the amount of overlap between two vehicles i and j, we define the probability of overlap as $\Pr_{i,j}^{(o)} = \Pr(\eta_{i,j} > 0) = \Pr(\delta_{i,j} < 1)$, we have, $\Pr_{i,j}^{(o)} = \int_0^\infty f(\eta_{i,j})d\eta_{i,j} = \frac{1}{\sigma_{i,j}\sqrt{2\pi}} \int_0^\infty \exp\left\{-\frac{1}{2}\left(\frac{\eta_{i,j}-\mu_{i,j}}{\sigma_{i,j}}\right)^2\right\}d\eta_{i,j}$. By defining the auxiliary variable $x = (\eta_{i,j} - \mu_{i,j})/\sigma_{i,j}$, we obtain,

$$\Pr_{i,j}^{(o)} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{(\mu_{i,j}/\sigma_{i,j})} e^{-(x^2/2)}dx = \Phi(\frac{\mu_{i,j}}{\sigma_{i,j}}), \quad (4)$$

where $\Phi(.)$ is the Cumulative Distribution Function (CDF) of the standard normal distribution.

The following optimization problem in (5) justifies the discussion on the number of enhancement layers that the transmitter can handle on the top of the encoded base layer video. This is a knapsack program, which defines the enhancement layers of rate $R_j(l_j)$ that could be transmitted over the underwater channel with maximum achievable communication data rate R_{max},

$$\underset{\lambda_k}{\text{maximize}} \quad \sum_{l=1}^L \lambda_l\lambda_{l-1}R_j(l_j), \quad (5a)$$

$$\text{s.t.} \quad \sum_{l=1}^L \lambda_l\lambda_{l-1}R_j(l_j) \leq R_{max}, \quad (5b)$$

$$\lambda_0 = 1, \; \lambda_l \in \{0,1\}, \forall l \in \{1,...,L\}. \quad (5c)$$

We determine the minimum number of vehicles to shut down such that we achieve the required QoS in the received video with an acceptable Quality of Experience (QoE) in the reconstructed map of environment based on a defined amount of spatial correlation. Vehicles are eligible to transmit a video with higher enhancement layers while the layers

bellow are successfully received/decoded. In this case, the following optimization problem can be presented for every chunk c of the video, given the optimal power P_j and the data rate R_j calculated from (3),

$$\underset{\alpha_i}{\text{maximize}} \quad \sum_{i=1}^{V-1} \alpha_i \quad (6a)$$

$$s.t. \; \alpha_i \in \{0,1\}, \quad (6b)$$

$$\mathbb{E}\left[\sum_{\substack{i=1\\i\neq j}}^V \alpha_i \log\left(1 + \frac{p_j^c|\,h_{ij}^c\,|^2}{s_i}\right)\right] \geq \text{QoS}_{th}(l_j), \quad (6c)$$

$$D_i < D_{th}, \quad (6d)$$

$$\Pr_{i,k}^{(o)} \geq \Pr_{th}, \quad \forall i,k \in \{1,...,V-1\}, \quad (6e)$$

where the objective function (6a) is the total number of vehicles. Maximizing the total number of vehicles (i.e. minimizing the number of vehicles to shut down) ensures the QoE since more vehicles from different angles are present in the map reconstruction. On the other hand, to satisfy a threshold QoS, the proposed method will shut down the vehicles with the worst channel to keep the average broadcasting rate over a minimum value, as shown in (6c). The other metric for QOS is represented in constraint (6d) which is defined by the SVC encoder and depends on the scalability and the number of enhancement layers that the encoder uses. For an encoded video, we can write [27] $D_i = \hat{\theta}/(R_j - R_0) + D_0$, where D_i represents the distortion of the video at the vehicle i at the time of reconstruction and R_j is the rate of the encoder at vehicle j; the other remaining variables $\hat{\theta}$, R_0, and D_0 depend on the encoded video and on the model, and are estimated empirically. The last constraint (6e) shuts down the vehicles which have a higher probability of overlap with the neighboring vehicles to have the minimum reduction in the QoE in reconstruction from different angles.

C. In-network Marine Litter Map Reconstruction

As it was discussed in the previous sections and due to the limited FoV of each single vehicle, a cooperation among the vehicles is required so that the required map can be reconstructed.

Potential Cooperation Strategies: We propose different strategies based on the exchanged data, acoustic channel requirements, level of complexity (that the vehicles can handle to process the data locally) and the QoS/QoE requirements as follows: (i) Vehicles exchange their local maps after each partial map is created. This strategy requires the minimum amount of data exchange since the merger creates the global map based on only a consensus on the exchanged local maps. (ii) Vehicles exchange the SVC-based channel independent videos, i.e., base layers. (iii) Vehicles exchange SVC adaptive channel dependent video, i.e., base and enhancement layers. This is the most desirable strategy that is also adaptive with the channel quality. (iv) Vehicles exchange the high quality video considering the acoustic channel bandwidth and the channel
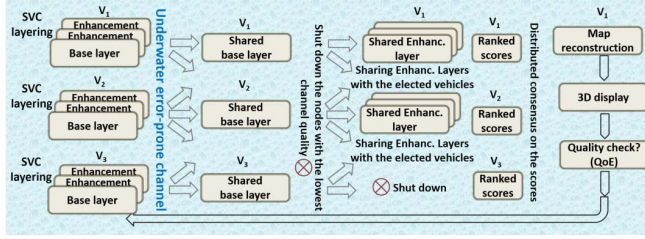
384

Fig. 3. Required steps for a satisfactory map construction. Vehicles V_1-V_3 share their encoded video (base layer) with other nodes. Enhancement layers are shared with the vehicles with a better acoustic channel quality (lower quality channels are shut down). After reaching a consensus, final reconstruction is performed on the highest-rank vehicle after it receives a high-quality video with higher QoS from other eligible nodes. If the QoE is not satisfactory, the process is restarted using a feedback command.

fading. This strategy is usually not feasible underwater due to the bandwidth limitation and time-varying nature of the underwater acoustic channel. Fig. 3 depicts the strategy we choose in this paper. After sharing the base layer, as we discussed in the previous sections, we shut down the vehicles with unreliable channels to be able to reach the required rate for sending the enhancement layers. We lose some part of the scene from those nodes which experience the shut down. Therefore, the vehicles should reach a consensus to decide on the node who finally reconstructs the global map.

Local Map Reconstruction: With the base layer video received at each node, along with that node's own high quality 4K original video, each node can perform a quick attempt at the map reconstruction. First, images are compared pairwise using SIFT/ORB to determine feature matches. Some of these pairwise matches will be false, and will appear in some pairwise comparisons but not in others that show similar perspectives on the scene. Because all nodes have some versions of the video, from different angles, the quality of reconstruction (measured by number of feature matches) should relate to two factors. Firstly, it depends on the amount of error-induced distortion in the base layer videos received from the other nodes. Secondly, it depends on the utility the locally stored 4K quality video on the reconstructing vehicle provides to the map reconstruction. Therefore, a vehicle that makes many feature matches in the intermediate local reconstruction attempt is a good candidate to share its recorded video at a higher quality in the next phase, because its video is a valuable part of the reconstruction and easy to match with the other videos. The underwater environment poses additional challenges in recording good video for the purposes of map reconstruction. While it can be shown that water itself is not a barrier to getting a good reconstruction, there are serious problems with lighting, scattering, turbidity, and clarity when taking underwater video.

Scoring and Sharing: Using the optimizations described in the previous sections, each transmitting node decides on the set of nodes to shut down before broadcasts its higher quality layers, i.e., enhancement layers. Therefore, some nodes miss some portions of video from some other angles since they did

---

Algorithm 1 SVC-based Map Reconstruction.

```
1:  while reconstruction is NOT satisfactory do
2:      Layers = ScalableVideoCoder(localVideo)
3:      EstablishMACchedule()
4:      τb ← allotted time for base layer sharing
5:      while t < τb do
6:          Share(Layers.LayerIndex(0))        % broadcasting
7:          Receive(ExternalVideo)
8:      end while
9:      receivedframes ← extractframes(receivedVideos)
10:     SIFT/ORBmatch(receivedframes)
11:     Reconstruct(matchedframes)
12:     RS ← score(reconstruct)
13:     random_broadcast_max(RS)
14:     τl ← allotted time for enhancement layer sharing
15:     while t < τl do
16:         if v is not FRV then
17:             Shut down the vehicles with the weakest channel
18:             Share(Layers.LayerIndex(L))       % multicasting
19:             Receive(ExternalVideo)
20:         else
21:             Reconstruct(Dataset)
22:         end if
23:     end while
24: end while=0
```

not receive them. We form a Reconstruction Score (RS) which is taken as a metric for how successful this vehicle would be at performing the later final reconstruction, as well as how valuable its local video is. This RS is shared in the following step to elect the Final Reconstructing Vehicle (FRV). Each node will share its RS to the group, such that at the end of this step all vehicles should have a list of each other vehicle's RS. As the process continues, nodes will become more aware of their position relative to other nodes. Since the RS is a very small amount of data, each vehicle can also share in the packet a map of camera positions (past vehicle positions) it has matched with. An average of these maps can be used to inform the vehicle's navigation in the time before the final reconstruction can be performed.

Consensus Algorithm on the Scores: To select the vehicle with the highest score for the final reconstruction, vehicles form the communication primitive to their neighboring vehicles. In particular, consensus is an iterative process where the nodes communicate with their neighbors to exchange their scores for a fixed number of iterations or until convergence [25]. As the output of this process, we select the best vehicle for final reconstruction. Asynchronous broadcasting-based consensus method proposed in [25] is to achieve the average value of the initial measurements. However, we wish to sort the scores to find the maximum in each iteration of the process. Each node v broadcasts its own score to its $\mathcal{N}_v$ neighboring nodes within its communication range [28]. The neighbors, such as w, which received the data, update their data according to $y_w(t_c + 1) = \max\big(y_v(t_c), y_w(t_c)\big), \forall w \in$
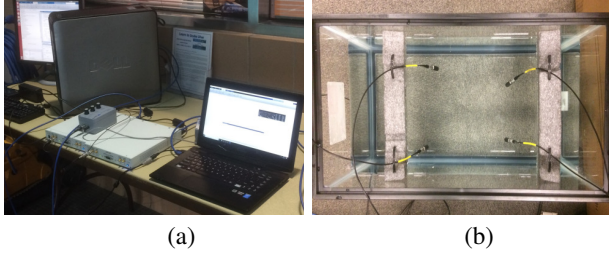
Fig. 4. (a) Software-defined acoustic testbed; (b) Water tank with TC4013 Teledyne transducers.
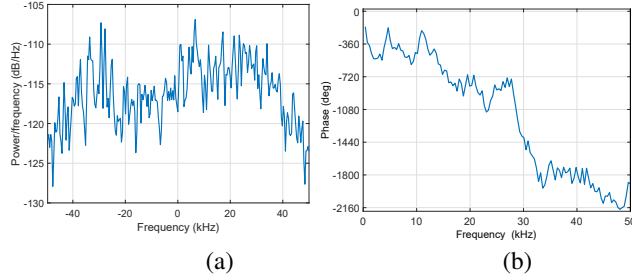

Fig. 5. Experimental channel response in the water tank shows (a) Power spectrum; (b) Phase.

$\mathcal{N}_v$, where N_v stands for the neighborhood of transmitting node v. The remaining nodes in the network update their values as $y_w(t_c + 1) = y_w(t_c), \forall w \notin \mathcal{N}_v$. This algorithm keeps the maximum value and so does not show an undesirable behavior in terms of convergence. After consensus, each vehicle should know the maximum RS among them and the vehicle that has it. The vehicle who has the highest score will transmit a final packet indicating its RS and intent to become the FRV. If there is no reply within the time limit, it is the FRV and the SVC enhancement layer sharing will commence. Algorithm 1 represents the solution in a sequential procedure for a specific coded video while the encoding and reconstruction is performed through the mentioned steps. Vehicles share their encoded base-layer and enhancement layers videos (after shutting down the vehicles with a low quality channel). After local reconstruction, matching and ranking the scores, the node with the highest score will be elected to perform the final reconstruction.

IV. PERFORMANCE EVALUATION

In this section, we present the experimental and simulation results to evaluate the proposed algorithm.

Testbed Setup: We evaluate the proposed approach by conducting preliminary field experiments. A video feed, captured by our underwater vehicles in the Raritan river, New Jersey, is passed to the SDAR and an acoustic transducer in a water tank. A high-performance and scalable platform with a programmable Kintex-7 FPGA, called X-300 designed by Ettus Research Group [29], is exploited as SDAR in this research, as the testbed shown in Fig. 4(a). It contains a mainboard to provide basic functionalities of the modem, while the daughter-boards take care of up/down conversions and
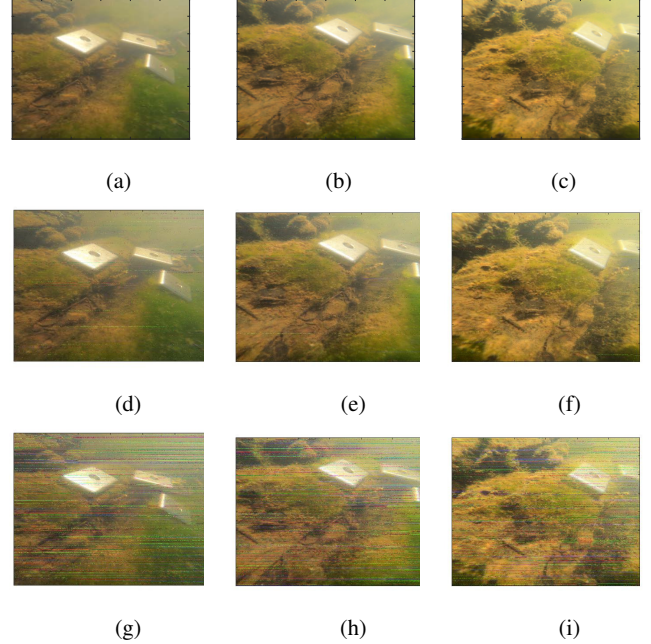

Fig. 6. (a)-(c) Frames from original video; (d)-(f) Frames of video received/reconstructed in a vehicle with a good channel; (g)-(i) Frames of video received/reconstructed at a vehicle with an average to low channel quality.

of the other required bandpass signal processing procedures. Teledyne Marine TC4013 transducers [30] with a frequency range of 170 kHz are used in the proposed testbed, shown in Fig. 4(b). Fig. 5 represents the channel response experienced in this testbed, containing the power spectrum of the channel in 5(a) and its phase in 5(b). The video was collected from the bottom of the Raritan river, New Jersey, using multiple cameras. We use the Joint Scalable Video Mode (JSVM) software as the reference package for implementing SVC. Using the FixedQPEncoder program, test videos were down-sampled and then encoded into multiple layers of different qualities. Each layer has a target fixed bit rate, and the Quantization Parameter (QP) is varied in order to optimize the Peak Signal-to-Noise Ratio (PSNR) metric while staying under the target bitrate.

Results: Fig. 6 shows the effect of the acoustic communication channel on the quality of the received video. The passband channel bandwidth is 100 kHz with carrier frequency of 100 kHz and the sampling rate is 200 kHz. In Figs. 6(a)-(c), the original successive frames are shown, while in Figs. 6(d)-(f) the quality of the received signal through a good channel is compared to the quality of the received signal through a low to average channel in Figs. 6(g)-(i).

Fig. 7 depicts different SVC layers of a single selected frame from the captured video. Fig. 7(a) shows the base layer and Fig. 7(b)-(e) represent the base and 1 to 4 enhancement layers of the original captured video. The corresponding frame rates for these layers are $1.8750, 3.75, 7.5, 15, 30$, respectively with the minimum bit rates of $100.9, 179.4, 293.3, 415.3, 517.5$ kbps. The cor-
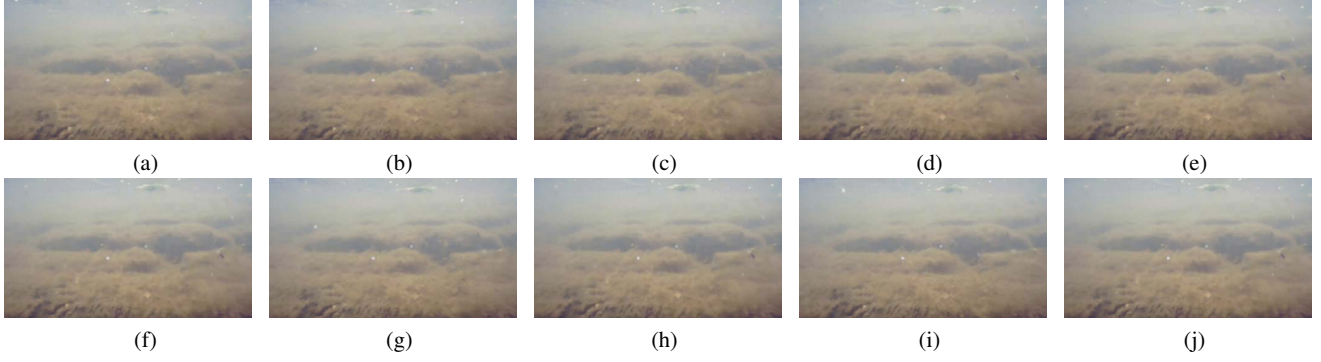
386

Fig. 7. SVC layers for a selected frame; (a) base layer of original video; (b)-(e) base layer and $1-4$ enhancement layers of original video; (f) base layer of received video; (g)-(j) base layer and $1-4$ enhancement layers of received video.
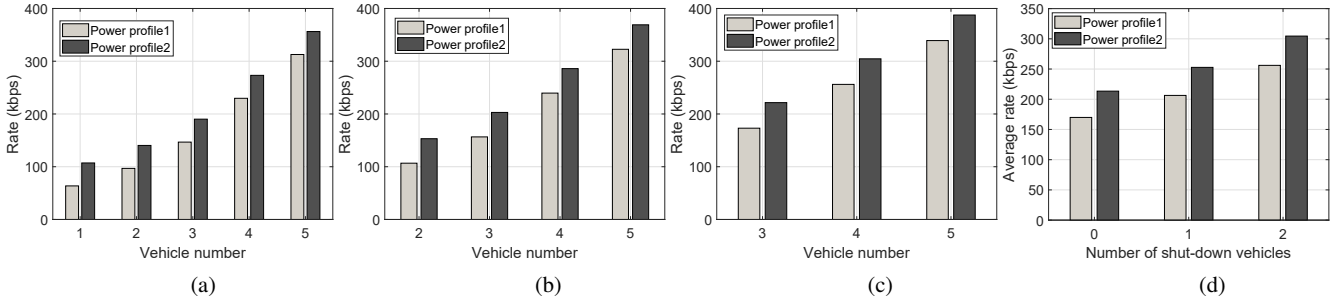

Fig. 8. Optimal received rate at different vehicles which are sorted based on their channel quality for two defined power profiles. In (a) all vehicles are active; (b) the vehicle with the worst channel quality is shut down; (c) two vehicles are shut down; (d) Achieved broadcast rate for two defined power profiles when the number of shut down vehicles changes.
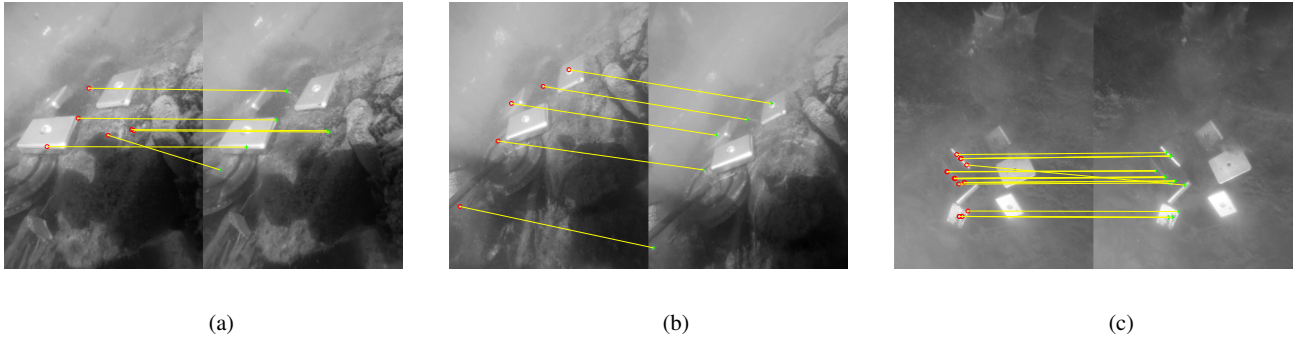

Fig. 9. Feature matching for different vehicles.

responding PSNR values are $45.1, 44.14, 43.31, 42.68$ and 42.19 dB, respectively. Fig. 7(f)-(j) show the associated base and enhancement layers of the same frame, when passed through our testbed. Note that the difference between number of enhancement layers can be distinguished better in the video.

Figs. 8(a)-(c) demonstrate the optimal received rates at different vehicles as a result of solving the proposed optimization problems. The vehicles are sorted based on their channel quality for two different power profiles. In Fig. 8(a), all the vehicles which are able to receive the base layer video are assumed in active mode. The vehicle which experiences a better channel receives the video with a higher rate. Fig. 8(b)-(c) shows the vehicles with the worst channel quality are shut down (one vehicle and two vehicles in these two figures,

respectively). Fig. 8(d) represents the proposed solution for the broadcast rate when variable number of vehicles are shut down. Two different power profiles are considered. By shutting down the vehicles with a low channel quality, the average broadcast rate is improved as shown in this figure. However, QoE in the result decreases since less vehicles are involved in the procedure, as explained in the solution.

Figs. 9(a)-(c) show the output of the feature matching and reconstruction based on the proposed algorithm. As shown in these figures, each vehicle observes the region of interest partially since there are serious problems with lighting, scattering, turbidity, and clarity when taking underwater videos. In Figs. 9(a)-(b), the vehicles detect three objects, while from other perspective, as shown in Fig. 9(c), six objects
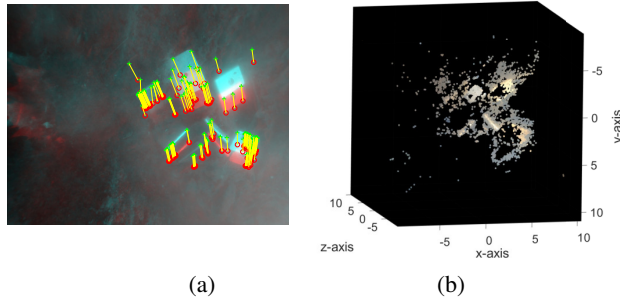
387

Fig. 10. (a) Tracked points; (b) Reconstructed map.

are detected. Fig. 10 shows the final steps towards map reconstruction. Fig. 10(a) represents the tracked features in the shared images and Fig. 10(b) is the reconstructed map of the region. The map can be used as a QoE metric to evaluate how accurate the desired map should be.

V. CONCLUSIONS AND FUTURE WORK

A novel in-network coordination that employed Scalable Video Coding (SVC) was introduced. Large amounts of data such as videos underwater is not easy to transmit due to the error-prone underwater channel. This paper investigated sharing SVC streams among AUVs in a multicast manner in which the vehicles with different capabilities/channel can be served by a single scalable stream to perform in-network map reconstruction. Performance evaluation was presented based on experiments using video captured from the Raritan River, New Jersey and transmitted through our software-defined acoustic testbed, in addition to simulation. In the future, we will extend our current solution to other efficient encoding schemes such as High Efficiency Video Coding (HEVC) in order to maximize the quality we can achieve under limited bandwidth constraints. Different compression rates will be compared and the effect of lighting, back scattering, and the turbidity will be accounted for and evaluated.

REFERENCES

[1] M. Pierdomenico, D. Casalbore, and F. L. Chiocci, "Massive benthic litter funnelled to deep sea by flash-flood generated hyperpycnal flows," *Scientific Reports*, vol. 9, no. 1, p. 5330, 2019.

[2] M. Rahmati and D. Pompili, "Probabilistic spatially-divided multiple access in underwater acoustic sparse networks," *IEEE Transactions on Mobile Computing, doi:10.1109/TMC.2018.2877683*, pp. 1–13, 2018.

[3] D. Pompili and I. F. Akyildiz, "A multimedia cross-layer protocol for underwater acoustic sensor networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 9, pp. 2924–2933, 2010.

[4] M. Rahmati and D. Pompili, "Ssfb: Signal-space-frequency beamforming for underwater acoustic video transmission," in *Proceedings of the International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*. IEEE, 2017, pp. 180–188.

[5] H. S. Dol, P. Casari, T. Van Der Zwan, and R. Otnes, "Software-defined underwater acoustic modems: Historical review and the nilus approach," *IEEE Journal of Oceanic Engineering*, vol. 42, no. 3, pp. 722–737, 2017.

[6] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the h. 264/avc standard," *IEEE Transactions on circuits and systems for video technology*, vol. 17, no. 9, pp. 1103–1120, 2007.

[7] Blue Robotics. [Online]. Available: http://www.bluerobotics.com.

[8] M. Stojanovic, "On the relationship between capacity and distance in an underwater acoustic communication channel," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 11, no. 4, pp. 34–43, 2007.

[9] T. Fujihashi, S. Saruwatari, and T. Watanabe, "Multi-view video transmission over underwater acoustic path," *IEEE Transactions on Multimedia*, 2018.

[10] J. Ribas, D. Sura, and M. Stojanovic, "Underwater Wireless Video Transmission For Supervisory Control and Inspection Using Acoustic OFDM," in *Proceedings of MTS/IEEE OCEANS Conference*, 2010, pp. 1–9.

[11] L. D. Vall, D. Sura, and M. Stojanovic, "Towards Underwater Video Transmission," in *Proceedings of the International Workshop on Underwater Networks*. ACM, 2011, p. 4.

[12] F. Campagnaro, R. Francescon, D. Tronchin, and M. Zorzi, "On the feasibility of video streaming through underwater acoustic links," in *2018 Fourth Underwater Communications and Networking Conference (UComms)*, Aug 2018, pp. 1–5.

[13] M. Rahmati, S. Karten, and D. Pompili, "Slam-based underwater adaptive sampling using autonomous vehicles," in *Proceedings of OCEANS Conference*. MTS/IEEE, 2018, pp. 1–6.

[14] T. Stutz and A. Uhl, "A survey of h. 264 avc/svc encryption," *IEEE Transactions on circuits and systems for video technology*, vol. 22, no. 3, pp. 325–339, 2012.

[15] Y. P. Fallah, H. Mansour, S. Khan, P. Nasiopoulos, and H. M. Alnuweiri, "A link adaptation scheme for efficient transmission of h.264 scalable video over multirate wlans," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 7, pp. 875–887, July 2008.

[16] M. Wichtlhuber, J. Rückert, D. Winter, and D. Hausheer, "How to adapt: Svc-based quality adaptation for hybrid peercasting systems," in *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, May 2015, pp. 174–182.

[17] M. K. Jubran, M. Bansal, and L. P. Kondi, "Low-delay low-complexity bandwidth-constrained wireless video transmission using svc over mimo systems," *IEEE Transactions on Multimedia*, vol. 10, no. 8, pp. 1698–1707, Dec 2008.

[18] M. Rahmati, A. Gurney, and D. Pompili, "Adaptive underwater video transmission via software-defined mimo acoustic modems," in *Proceedings of OCEANS Conference*. MTS/IEEE, 2018, pp. 1–6.

[19] S. Lmai, M. Chitre, C. Laot, and S. Houcke, "Throughput-efficient super-tdma mac transmission schedules in ad hoc linear underwater acoustic networks," *IEEE Journal of Oceanic Engineering*, vol. 42, no. 1, pp. 156–174, 2017.

[20] A. A. Syed, W. Ye, and J. Heidemann, "T-lohi: A new class of mac protocols for underwater acoustic sensor networks," in *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*, 2008, pp. 231–235.

[21] J. G. Proakis and M. Salehi, *Digital communications*. McGraw-Hill, 2008.

[22] N. Jindal and A. Goldsmith, "Capacity and optimal power allocation for fading broadcast channels with minimum rates," *IEEE Transactions on Information Theory*, vol. 49, no. 11, pp. 2895–2909, 2003.

[23] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[24] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[25] P. Pandey, M. Rahmati, D. Pompili, and W. U. Bajwa, "Robust distributed dictionary learning for in-network image compression," in *Proceedings of International Conference on Autonomic Computing (ICAC)*. IEEE, 2018, pp. 61–70.

[26] R. Dai and I. F. Akyildiz, "A spatial correlation model for visual information in wireless multimedia sensor networks," *IEEE Transactions on Multimedia*, vol. 11, no. 6, pp. 1148–1159, 2009.

[27] K. Stuhlmuller, N. Farber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 1012–1032, 2000.

[28] F. Iutzeler, P. Ciblat, and J. Jakubowicz, "Analysis of max-consensus algorithms in wireless channels," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 6103–6107, 2012.

[29] USRP X Series. [Online]. Available: https://www.ettus.com

[30] RESON TC4013 Hydrophone Product Information . [Online]. Available: http://www.teledynemarine.com/reson-tc4013