Review Article

ADVANCED IMAGING

Underwater optical imaging and sensing in turbidity using three-dimensional integral imaging: a review

Alex Maric, Gokul Krishnan, Rakesh Joshi, Yinuo Huang, Kashif Usmani, and Bahram Javidi*
Department of Electrical and Computer Engineering, University of Connecticut, Storrs, USA

Abstract. Three-dimensional (3D) integral imaging (InIm) is an active research topic in underwater optical imaging and sensing systems. 3D InIm is a passive method to visualize the 3D information of a scene. This method is attractive for imaging in degraded environments and can benefit short to long range applications. In this paper, we present an overview of previously published works on underwater optical imaging and sensing systems showcasing improved visualization and detection capabilities in turbid water using 3D InIm. A set of image sensors capturing images or video sequences were digitally reconstructed using a multidimensional InIm computational reconstruction algorithm for tasks such as visualizing the 3D information of a scene at a particular depth, signal detection, or decreasing the effect of turbidity and partial occlusion. 3D InIm sensing and dedicated computational approaches such as statistical image processing, correlation-based filtering, neural networks, and deep learning make it possible to improve the recovery of information and detection of scenes degraded by turbidity and partial occlusion.

Keywords: 3D integral imaging; underwater optical imaging; scattering; turbidity; deep learning.

Received May 8, 2024; revised manuscript received Oct. 4, 2024; accepted Oct. 17, 2024; published online Nov. 20, 2024.

[DOI: 10.3788/AI.2025.20002]

1. Introduction

Underwater imaging is challenging as visibility is reduced due to the scattering and absorption of light by suspending particles in water. As light propagates in water, it scatters in different directions, which results in the degradation of the captured images. Light propagation in water is wavelength-dependent, as different wavelengths experience varying absorption levels, thus limiting the distance of propagation and depth at which we can visualize the objects. To relate the effects of attenuation with distance, Beer's law may be used to measure the attenuation of light in a scattering medium, that is, $I = I_o e^{-\alpha d}$[1], where I_o is the intensity of light in free space, I is the intensity of light in a scattering medium propagated at a distance d, and α is Beer's coefficient. A mathematical model that may compensate for the scattering and low visibility in underwater images is the Jaffe model[2]. This model summarizes the components of scattering that contribute to image degradation in a scattering medium, which are the forward-scattered component, direct component, and backscattered component. Forward scattering occurs when the light reflected from the object is scattered at a small angle due to particles, whereas the backscattered component arises when the light that reaches the camera sensor is not reflected from the object. The direct component occurs when the light is reflected from the object and reaches the camera sensor without any scattering.

One popular underwater sensing technology is light detection and ranging (LiDAR), an active device that uses laser pulses to measure the distance or the range of the object[3]. In underwater applications, bathymetric LiDAR is the most common LiDAR sensing technology[3–7]. Bathymetric LiDAR is usually mounted on an aircraft or a boat and uses a green laser to penetrate deep in water. However, as the laser propagates deeper underwater, it loses energy due to scattering and absorption. Thus, active sensors can benefit from powerful light sources for longer range operations. However, active sensors such as LiDAR may become more complex and expensive as they require lasers and time-gated detectors.

*Address all correspondence to Bahram Javidi, bahram.javidi@uconn.edu

One promising technique using multi-view imaging is three-dimensional (3D) integral imaging (InIm)[8], which captures images from different perspectives of the scene to reconstruct the scene at a particular depth plane[9]. The technique is passive and is ideal for imaging under degradations. 3D InIm has been used in underwater imaging applications for optical signal detection[10–16], object visualization[17–21], object classification[22], and object detection[16]. These works have implemented 3D InIm systems in various modalities such as with and without lenses, with diffusers, polarimetric systems, etc. In addition, computational algorithms such as the convolutional neural network (CNN)[23], generative adversarial network (GAN)[24], recurrent neural network (RNN)[25], statistical image processing[26], statistical optics[27], active polarization descattering[28], and correlation-based filters[29–31] were used to overcome the effects of turbidity and/or partial occlusion. Additionally, they have potential in many underwater applications including automated detection, classification of objects in turbidity, and marine research.

This paper presents an overview of recent advances in underwater optical imaging and sensing systems using 3D InIm in turbid conditions. The paper is organized into five different sections. Section 1 describes underwater imaging and 3D InIm in underwater imaging applications, and briefly discusses LiDAR. Section 2 describes the basics of 3D InIm and reconstruction algorithms, lensless imaging, and polarimetric InIm. Section 3 summarizes various applications of 3D InIm such as underwater object visualization[18,19,21], underwater signal detection[12–16], underwater object classification[22], and underwater object detection[16]. Section 4 summarizes the comparisons of the published works discussed in Sec. 3. Section 5 concludes the paper.

2. Methodology

2.1. Three-Dimensional Integral Imaging

3D InIm is a multidimensional imaging technique that can reconstruct a 3D scene at a user-defined depth using the angular and intensity information of the captured perspective images, called two-dimensional (2D) elemental images[8–13,16–22]. Conventional 2D imaging, which only collects the intensity of a scene, superimposes the background and the object. However, in 3D InIm, the angular information captured using the perspective images allows the 3D reconstruction at different depths in the scene, which gives the observer a 3D view of the scene. By reconstructing the object at the depth of interest, 3D InIm can segment the object out of the background. This technique is optimal in the maximum likelihood sense in terms of Gaussian noise for read-noise dominant images[32,33] to provide a higher signal-to-noise ratio (SNR) and has depth-sectioning capabilities to segment in-focus planes from out-of-focus planes[9]. Additionally, 3D InIm can visualize through partial obscurations by using parallax to capture non-occluded perspective images and may remedy moderate scattering conditions[13,16,20–22]. During the image acquisition stage [see Fig. 1(a)], the optical rays are projected onto 2D elemental images and recorded using a camera array or a single camera on a moving platform. InIm systems can be implemented using a one-dimensional (1D) array of cameras[12,14,15] or a 2D array of cameras[10,11,13,16–22]. With a 1D array, we only have parallax in one direction (i.e., x-axis or y-axis), whereas, with a 2D array, we have parallax in two directions (i.e., x-axis and y-axis). When using a single camera sensor or a camera array, there are tradeoffs to consider, and scenarios for which these approaches are appropriate. A camera array-based image acquisition is useful in both dynamic (e.g., turbulence, fast-moving objects, and signal detection) and static environments (that is, stationary objects). All cameras in the array are synchronized to record all perspective images or videos simultaneously, resulting in faster data acquisition compared to a single camera on a moving platform. However, camera calibration[34] is required as a preprocessing step for 3D InIm reconstruction to correct any possible misalignment. Another drawback is that the physical size of the cameras limits the minimum possible camera pitch value, that is, how closely the cameras can be placed next to each other. A single camera sensor is beneficial for user flexibility in choosing the camera pitches and the number of perspective images. However, this approach may be only suitable in static environments.

To digitally reconstruct at a given range (depth) z [see Fig. 1(b)], we backpropagate the optical rays through a virtual pinhole array. The other depth planes are out of focus when the scene is reconstructed at a particular depth. Mathematically, 3D InIm reconstruction of a scene[21] [see Eq. (1)] or video sequences converted to image frames[13] at time frame t [see Eq. (2)] is a superposition of transversally shifted elemental images depth sliced at a user-defined plane z_r:

$$I(x,y,z_r)=\frac{1}{O(x,y)}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}EI_{mn}\left(x-m\frac{N_x p_x f}{c_x z_r},y-n\frac{N_y p_y f}{c_y z_r}\right), \tag{1}$$

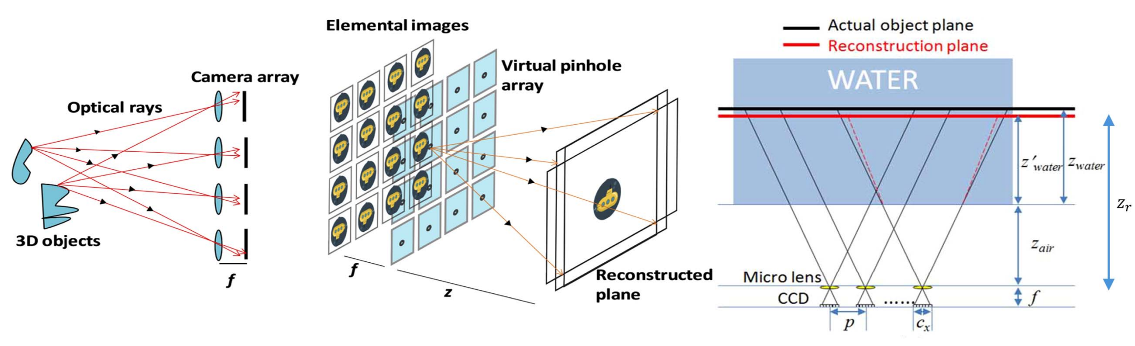

Fig. 1 3D integral imaging. (a) Pickup stage[12]. (b) Reconstruction stage[12]. (c) 3D underwater computational reconstruction[18]. Reprinted with permissions from Refs. [12] and [18].

$$I(x,y,z_r;t) = \frac{1}{O(x,y;t)} \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} EI_{mn}\left(x - m\frac{N_x p_x f}{c_x z_r}, y - n\frac{N_y p_y f}{c_y z_r}; t\right), \quad (2)$$

where I is the intensity of the reconstructed 3D image at the pixel index (x, y) or the reconstructed 3D image at pixel index (x, y, t) at time frame t, EI is the 2D elemental image, $O(x, y)$ and $O(x, y; t)$ are the overlapping number matrices, and M and N are the numbers of elemental images acquired along the x-axis and y-axis, respectively. N_x and N_y are the total numbers of pixels of each 2D elemental image, p_x and p_y are the camera pitches, and c_x and c_y are the horizontal and vertical image sensor sizes, respectively. f is the lens focal length. In free space imaging, z_r is the distance to the actual object plane. According to ray optics in Fig. 1(c), when light rays reach the air–water interface, the rays will refract in water, so the true distance will be shorter than the reported distance. As a post-processing step, we modify z_r in Eqs. (1) and (2) to account for the refractive index of water: $z_r = z_{\text{air}} + \frac{z_{\text{water}}}{n_{\text{water}}}$[17]. z_{air}, a constant value, is the distance from the camera lens to the aquarium tank glass, z_{water} is the object distance in water, and n_{water} is the refractive index of water (≈1.33).

2.2. Lensless Imaging

Advances in optical instruments have provided users with access to better camera sensors to capture high-quality images. However, despite their improvements, camera sensors have a limited field of view due to their sensor size and focal length. To capture more photons, one would need a bigger, bulkier lens, which is expensive. An alternative solution is to use a lensless-based imaging system, which is inexpensive, less bulky, has a larger field of view, and is more compact than their lens-based counterparts[35]. In Ref. [15], a 1D lensless camera array underwater sensing system using diffusers was proposed to classify pseudorandom patterns in turbidity and partial occlusion. It was shown that this approach works well compared with its lens-based counterpart for underwater optical signal detection[14].

2.3. Underwater Polarimetric Integral Imaging

Polarimetric imaging is an imaging technique that captures the polarimetric information of a scene. Polarimetric imaging introduces an additional degree of information, enhancing the data for analysis and enabling a more comprehensive understanding of the observed scene or object. The polarimetric information of a scene contains the surface features of an object as well as the scattered light to enhance the contrast of an object scene in scattering media[36]. In polarimetric imaging, the polarimetric information of the scene can be captured by using either a polarization camera or by placing a polarization filter in front of a camera sensor. Polarimetric imaging can also be performed either using active (i.e., light source) or passive illumination (i.e., natural light) depending on the environment. In Ref. [12], an active polarimetric signal detection system using polarization-difference imaging[37,38] was proposed to transmit a polarized optical signal in turbid water to reduce the effects of turbidity. Polarization-difference imaging was applied to the captured videos and then processed using 3D InIm reconstruction [see Eq. (2)]. In Ref. [20], polarization-based image recovery using active polarization descattering and 3D InIm reconstruction [see Eq. (1)] was proposed to reduce the effects of turbidity for underwater scenes with partial occlusion.

3. Applications

In Sec. 3.1, we briefly review object visualization using statistical image processing (Sec. 3.1.1), peplography (Sec. 3.1.2), and deep learning using GAN (Sec. 3.1.3). In Sec. 3.2, we briefly review signal detection using correlation-based filtering (Sec. 3.2.1) and deep learning using (1) CNN combined with RNN (Sec. 3.2.2) and (2) CNN (Secs. 3.2.3 and 3.2.4). In Sec. 3.3, we briefly review object classification using a neural network (Sec. 3.3.1) and a dual-purpose system using deep learning (Sec. 3.3.2) with (1) CNN and (2) CNN combined with RNN.

3.1. Underwater Object Visualization

Object visualization under degradation was one of the first underwater applications reported using 3D InIm[17]. Due to difficulty visualizing the objects under high turbidity, image restoration techniques were also used along with 3D InIm to mitigate the effects of turbidity. We briefly review InIm underwater object visualization using statistical image processing[18], peplography[19], and physics-based deep learning[21]. For visualization tasks, we record images and perform 3D InIm reconstruction using Eq. (1).

3.1.1. Statistical image processing

The first object visualization report on 3D InIm reconstruction of underwater objects in turbidity was presented in Ref. [18]. A multi-class underwater turbid scene in Fig. 2(a) was degraded by turbidity. As a result, the objects could not be visualized in turbidity; thus, the authors used statistical image processing algorithms[26] along with 3D InIm reconstruction for improved visualization. The authors assumed that turbidity was caused by light scattering, which they statistically modeled as a Gaussian distribution. While this statistical model of turbidity may not be very precise, it allows for the application of statistical approaches to reduce the effects of turbidity. The authors used maximum likelihood estimation[39], gamma correction, histogram matching, and histogram equalization to enhance the visibility of the objects. Finally, the grayscale perspective images were preprocessed using 3D InIm reconstruction. Figure 2 shows an example of this approach where one object is reconstructed in focus at a particular depth plane and the other planes are blurred out.

3.1.2. Peplography

As described in Sec. 3.1.1, in Ref. [26], the authors used statistical image processing algorithms, including gamma correction, histogram equalization, and histogram stretching, that are subject to estimation errors. Therefore, it results in artificial color pixel intensities. To alleviate this issue, the authors in Ref. [19] proposed peplography (see Fig. 3), a passive technique that relies on detecting ballistic photons—photons that travel through scattering media without significant scattering—and filtering out scattered photons using statistical algorithms[27]. This technique combines photon counting with statistical algorithms to estimate the turbid medium and extracts the ballistic photons for image reconstruction. This helps to distinguish ballistic photons from scattered photons, allowing for a less noisy visualization of objects. The detected photons are then used to

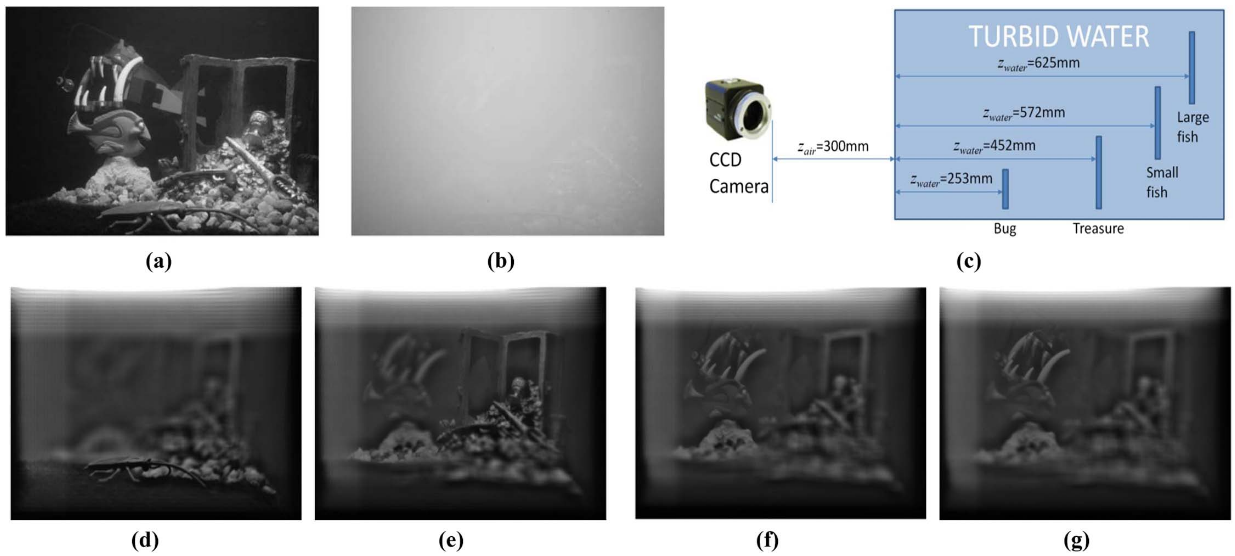

Fig. 2 (a) Clear water 3D scene used in the experiments. (b) The scene in turbid water. (c) Diagram of the 3D underwater imaging system in turbid water. (d)–(g) 3D integral imaging reconstruction. (d) 3D reconstructed bug focused at $z_r = 490$ mm, (e) 3D reconstructed treasure focused at $z_r = 640$ mm, (f) 3D reconstructed small fish focused at $z_r = 730$ mm, and (g) 3D reconstructed large fish focused at $z_r = 770$ mm. Reprinted with permission from Ref. [18].

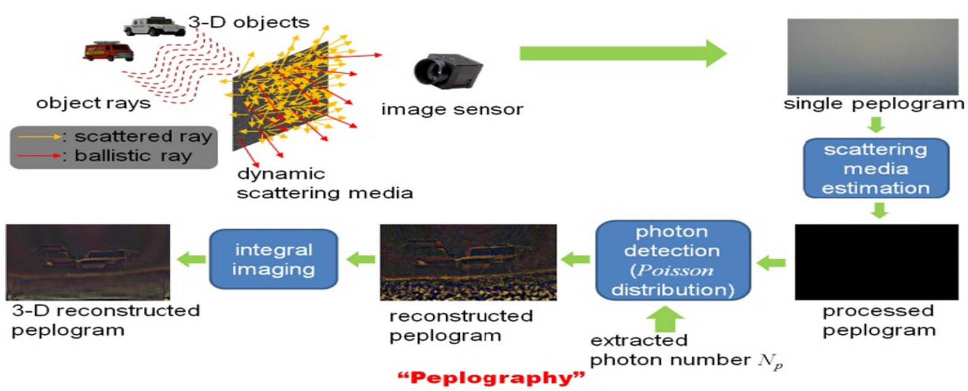

Fig. 3 Flow chart of peplography for imaging in scattering media. Reprinted with permission from Ref. [19].

reconstruct a 3D image of the scene using InIm, which involves capturing multiple perspectives of the scene and processing them to create a 3D representation of the object. However, the expected number of photons is arbitrarily chosen for ballistic photon counting due to fluctuations in their intensity. If the expected value is too low, the ballistic photons may not be captured to recover the 3D object information and to restore the scene in scattering media. 3D InIm reconstruction also uses averaging, as shown in Eq. (1), to reduce noise.

3.1.3. Physics-based deep learning

Deep learning is a computational technique that uses multilayer neural networks to learn representations of features directly from data[40]. Conventional deep learning-based approaches are purely data-driven. However, for improving the performance and interpretability/reliability of deep learning models, incorporating the physical laws in the deep learning models would be beneficial as was proposed in Ref. [21]. In this work, 3D InIm with the physics-informed cycle generative adversarial network (CycleGAN)[41,42] for image recovery was proposed. The network was trained on unpaired underwater datasets in turbid water with external lighting and tested in different turbidities with external lighting with and without partial occlusion. The turbidity was varied from $\alpha = 0.0025\ \mathrm{mm}^{-1}$ to $\alpha = 0.040\ \mathrm{mm}^{-1}$, where α is Beer's coefficient. The degradation parameters were estimated using the dark channel prior[43] and Bayesian optimization[44] to estimate the ground truth values, the atmospheric light, and the depth. The backbone of the approach is the encoder–decoder model, where the encoder takes the clean image as input, along with the estimated ground truth parameters, and outputs a degraded image. A physical model is used to synthetically create degraded images using the same set of input parameters to the encoder, and the loss is computed between the output of the encoder and the synthetically generated images. The output of the encoder is then fed into a decoder to recover the clean images. Examples of

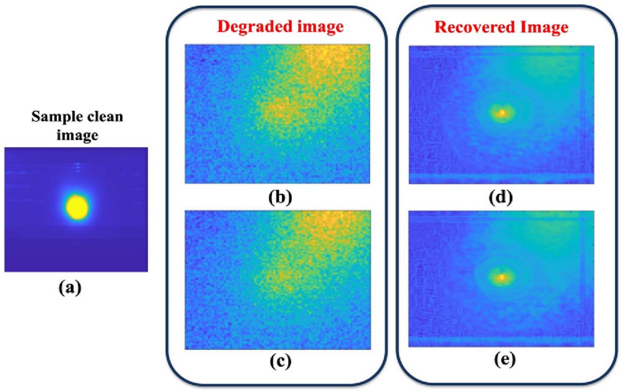

Fig. 4 (a) Sample clean image of a light source ($\alpha = 0.0025\ \mathrm{mm}^{-1}$). (b)–(c) Degraded images in high turbid water ($\alpha = 0.040\ \mathrm{mm}^{-1}$). (d) Recovered clean image from (b). (e) Recovered clean image from (c). α is Beer's coefficient. Reprinted with permission from Ref. [21].

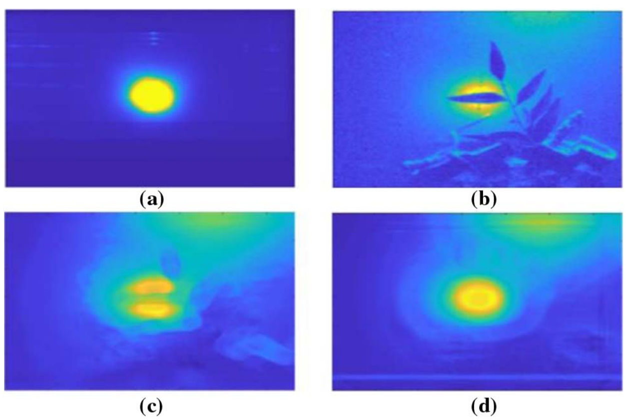

Fig. 5 (a) Sample clean image of a light source ($\alpha = 0.0025\ \mathrm{mm}^{-1}$). (b) 2D degraded central perspective image in turbidity, ambient light, and partial occlusion ($\alpha = 0.008\ \mathrm{mm}^{-1}$). (c) Recovered clean 2D image from (b). (d) Recovered clean 3D reconstructed image from (b). α is Beer's coefficient. Reprinted with permission from Ref. [21].

synthetically generated degraded images of a light source and recovered clean images are shown in turbid water without occlusion and with occlusion in Figs. 4 and 5, respectively. By using a physical model, the deep learning algorithm is limited to generating outputs that are physically consistent. The authors also used an unpaired training set, eliminating the need to collect large, paired datasets. Although physics-based deep learning was applied in turbidity, the method can be extended to other degraded environments (e.g., fog and smoke) using a suitable physical degradation model.

3.2. Underwater Optical Signal Detection

Another application of 3D InIm is optical underwater signal detection, which focuses on transmitting, receiving, and detecting bits in the presence of turbidity and ambient light conditions. Thanks to improvements in computational algorithms, we can enhance the deteriorated signal to improve the detection performance. In this section, we review signal detection papers covering correlation-based filters[12] and deep learning-based signal detection methods[13–15]. The optical signals are temporally encoded, so multi-perspective videos of the transmitted signal are recorded using a camera array. Then, the video frames are converted to image frames for 3D InIm reconstruction [see Eq. (2)][12,13,16] or fed directly into a deep learning model[14,15]. On the transmitter side, a light source such as an LED-generated binary optical signal is coded with a gold sequence[45]. The gold sequence, expressed as “1” (i.e., LED on) and “0” (i.e., LED off), denoting the gold code and the flipped gold code, respectively, is transmitted underwater. At the receiver end, the transmitted signal is decoded, outputting the class conditional probabilities comparing the coded sequence to the transmitted video sequence: “1,” “0,” and “idle” (i.e., neither “1” nor “0”). The “idle” class is included in Refs. [13–16] to better discriminate between “1” and “0.”

3.2.1. Active polarimetric InIm using correlation-based filtering

In Sec. 2.3, we briefly introduced underwater polarimetric InIm. An example was shown in Ref. [12], where the authors proposed an underwater single-shot polarization-difference InIm system for signal detection using four-dimensional (4D) nonlinear correlation[29,30]. The underwater scene was illuminated by an active polarized light source, and a beam splitter divided the transmitted signal into two beams. These beams were then captured by two orthogonally polarized 1×3 camera arrays. The signals were processed using polarization-difference imaging and subsequently reconstructed using 3D InIm to create 4D reconstruction data (x, y, z, t) over time. Lastly, a 4D nonlinear correlation filter was applied for signal detection as depicted in Fig. 6. The system's field of view is limited by the camera array pitch, the size of the beam splitter, and the distance between the beam splitter and the camera arrays. An alternative setup would be to replace the beam splitter and the polarized camera arrays with a polarization camera sensor array equipped with polarization filters. Also, correlation-based filters are effective with smaller datasets, but they may not generalize as well as deep learning techniques.

3.2.2. InIm with deep learning for underwater applications

In this section, we consider a deep learning-based approach for optical signal detection. Partial occlusion was not considered in Ref. [12] but was considered in Ref. [13] in addition to turbidity. In Ref. [13], the authors proposed a signal detection-based deep learning system using a convolutional neural network with bidirectional long short-term memory (CNN-BiLSTM)[46–48]. The classifier learns the spatial information of the data using a pretrained CNN, GoogleNet[49], and learns the temporal information using the BiLSTM. The training data was recorded in clear water without occlusion and evaluated on testing data in clear and turbid water with occlusion from Beer's coefficient of $\alpha = 0.0047\ \mathrm{mm}^{-1}$ to $\alpha = 0.0318\ \mathrm{mm}^{-1}$. After the InIm reconstruction step, the authors performed sliding-window-based temporal segmentation to segment each possible video sequence from the transmitted data sequence. After segmentation, the sequences were input to the CNN-BiLSTM model, which outputs the classification score. Last, a threshold was set to “0” and applied to the classification score to classify the transmitted video sequences (see Fig. 7). The authors assumed the depth information (light source range) of the object was known *a priori* for computational efficiency. However, the

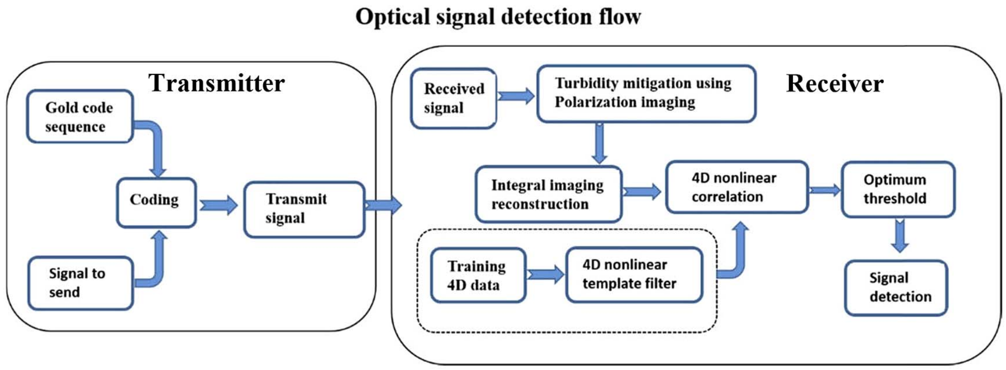

Fig. 6 Flow chart of the underwater optical signal detection pipeline using 4D nonlinear correlation. Reprinted with permission from Ref. [12].

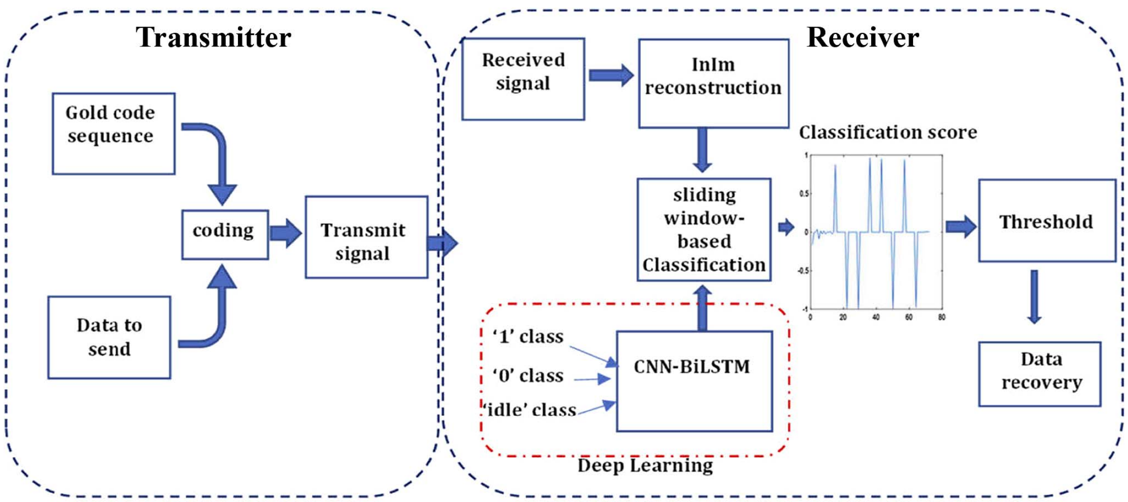

Fig. 7 Flow chart of CNN-BiLSTM-based signal detection using sliding window-based classification. Reprinted with permission from Ref. [13].

unknown source range can be accommodated with additional computation. The proposed system, which involves camera array calibration, 3D InIm reconstruction, depth estimation, and classification in different stages, is time-consuming and may be prone to reconstruction errors, which can influence the detection results. If we devise an end-to-end deep learning architecture, neither of the intermediate steps would be required. Thus, the likelihood of errors would be minimized, and the computational time would be improved.

3.2.3. End-to-end integrated 1D InIm convolutional neural network for underwater applications

In contrast to Ref. [13], the approach in Ref. [14] is an end-to-end integrated architecture that could train on the spatial and temporal features of different camera perspectives without the intermediate steps such as camera calibration, depth estimation, and 3D InIm reconstruction of the light source. Thus, several improvements were made at the receiver end to increase the speed and to minimize hardware complexity without comprising signal detection performance (see Fig. 8). The authors used a 1×9 camera array to capture multi-perspective videos of the temporally encoded optical signal for better longitudinal depth resolution[50]. The multi-perspective videos were fed as inputs to a 1D InIm convolutional neural network (1DInImCNN) deep learning model. This minimized the computational load required due to the elimination of the intermediate depth estimation and 3D InIm reconstruction. Furthermore, the spatial and temporal features were learned simultaneously rather than in separate stages. The training data was collected in clear and turbid water without occlusion from $\alpha = 0.0001\ \mathrm{mm}^{-1}$ to $\alpha = 0.0300\ \mathrm{mm}^{-1}$ and tested in clear and turbid water with occlusion from $\alpha = 0.0001\ \mathrm{mm}^{-1}$ to $\alpha = 0.0290\ \mathrm{mm}^{-1}$. The light source was placed in arbitrary locations, assuming that the light source range was unknown *a priori*. In summary, the computational cost, measured using floating point operations, and the detection performance in terms of Matthew's correlation coefficient (MCC)[51] show that the end-to-end integrated 1DInImCNN is substantially better compared to conventional 3D InIm using CNN-BiLSTM.

3.2.4. Lensless imaging using diffusers

In Sec. 3.2.3, we introduced an end-to-end integrated underwater optical signal detection methodology and compared its improvements with the CNN-BiLSTM-based signal detection

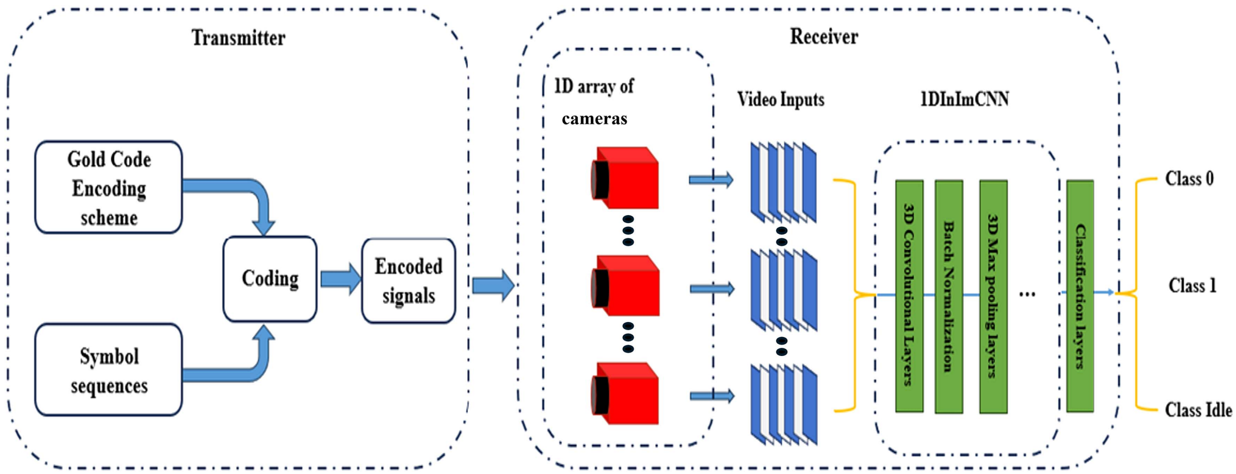

Fig. 8 Flow chart of end-to-end integrated 1DInImCNN-based signal detection model. Reprinted with permissions from Refs. [14,15].

system presented in Sec. 3.2.2. In this section, we extend our discussion of Sec. 3.2.3 to a lensless-based sensing system[15] (see Sec. 2.2), replacing the camera lens with a diffuser. The experiments were performed on a 1×3 lensless camera array as shown in Fig. 8. Using a 1DInImCNN, the model was trained in clear and turbid water without occlusion from $\alpha = 0.0007\ \mathrm{mm}^{-1}$ to $\alpha = 0.0145\ \mathrm{mm}^{-1}$ and tested on clear and turbid water data with occlusion from $\alpha = 0.0010\ \mathrm{mm}^{-1}$ to $\alpha = 0.0215\ \mathrm{mm}^{-1}$ (see Fig. 9). Unlike traditional lenses that form an image on the image sensor, the diffusers scatter light into speckle patterns across the image sensors. The diffuser-based lensless system compared with its lens-based system in Ref. [14] is computationally less demanding in terms of the number of floating point operations. The reason is that the computational requirement could be further reduced by randomly cropping 1D rows or columns of pixels[52] without compromising the signal detection performance significantly. Thus, the authors used a dimensionality-reduced single-camera system with 1 pixel × 50 pixel and 1 pixel × 150 pixel and a dimensionality-reduced 1×3 camera array system with 1 pixel × 50 pixel per camera and 1 pixel × 100 pixel per camera. The authors concluded that the lensless underwater optical signal detection system can outperform the lens-based counterpart under the experimental conditions considered.

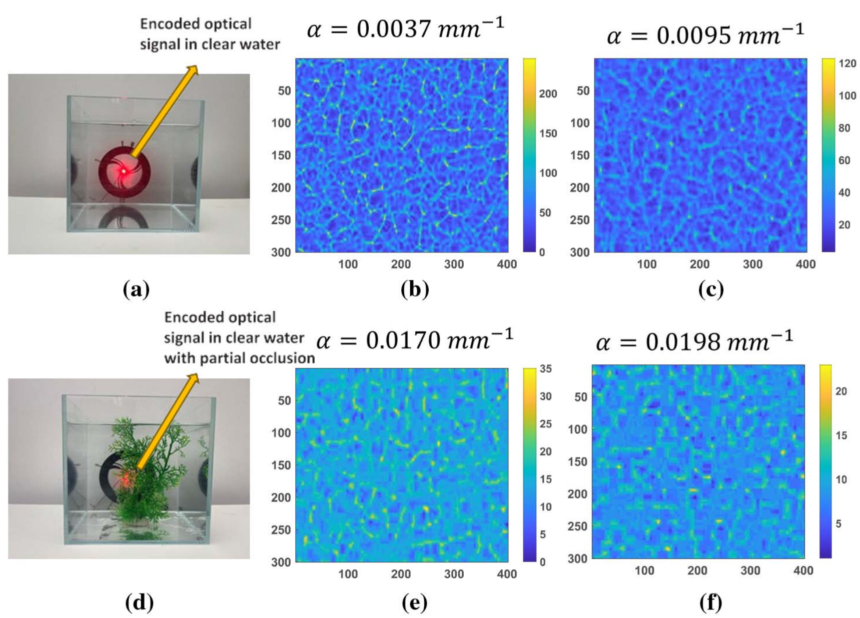

Fig. 9 (a)–(c) Sample training data without occlusion. (a) Central perspective video of the encoded optical signal in clear water. (b)–(c) Sample training data in turbid water from a lensless-based 1D camera array with diffusers at $\alpha = 0.0037\ \mathrm{mm}^{-1}$ and $\alpha = 0.0095\ \mathrm{mm}^{-1}$, respectively. (d)–(f) Sample testing data with occlusion. (d) Central perspective video of the encoded optical signal in clear water with partial occlusion. (e)–(f) Sample testing data in turbid water with partial occlusion from a lensless-based 1D camera array with diffusers at $\alpha = 0.0170\ \mathrm{mm}^{-1}$ and $\alpha = 0.0198\ \mathrm{mm}^{-1}$, respectively. Reprinted with permission from Ref. [15].

3.3. Underwater Object Classification and Object Detection

Another application we present in this review is 3D InIm underwater object classification and object detection. In object classification, given an image, we assign a label to the entire image, whereas in object detection we try to localize the objects present in an image using bounding boxes. A bounding box is only assigned to a target when the confidence score exceeds a predefined threshold and has a maximum score value out of duplicate predictions according to non-maximum suppression[53]. In this section, we review underwater reports using distortion-tolerant object classification[22] and You Only Look Once Version 4 (YOLOv4) object detection[54]. However, more advanced versions of YOLO[55] and other deep learning algorithms may be used.

3.3.1. Object classification

Underwater object classification in turbidity using 3D InIm was investigated in Ref. [22]. This was the first report on using InIm for underwater 3D object classification and as such used the older neural network algorithms available. A microlens array was inserted in front of a stationary single camera sensor to capture the 3D scene on a 39×26 array of elemental images. The training data was collected on six rotation-variable objects in free space and then tested on six rotation-variable objects in clear and turbid water with and without partial occlusion from $\alpha = 3\ \mathrm{m}^{-1}$ to $\alpha = 12\ \mathrm{m}^{-1}$. The training and testing images were preprocessed using 3D InIm reconstruction [see Eq. (1)] and computational algorithms such as gradient map[26], normalized power transform[26], and principal component analysis (PCA)[56], then classified using a neural network. The classification results on the 3D reconstructed underwater testing data in terms of accuracy were 100% for all six object classes in low turbid water ($\alpha = 6\ \mathrm{m}^{-1}$) without and with occlusion. In intermediate turbid water ($\alpha = 12\ \mathrm{m}^{-1}$), the accuracy was 99% for two out of the six object classes and 100% for the other four object classes. In intermediate turbid water ($\alpha = 12\ \mathrm{m}^{-1}$) with occlusion, the accuracy was 98.7%, 99.7%, and 100% for two object classes, one object class, and three object classes, respectively. One limitation of the proposed system is the acquisition of the perspective images using a microlens array, which limits the depth of field and resolution for quality 3D InIm reconstruction. To avoid these limitations and to capture high-resolution perspective images, the microlens array can be removed, and a single camera can be moved laterally on a moving platform, or the imaging system can be implemented with a camera array.

3.3.2. Object detection

Unlike object classification, which ignores the target locations, object detection is a suitable approach for localizing multiple objects. An example was applied in Ref. [16], where a dual-purpose underwater object detection and optical (temporal) signal detection-based system were explored using 3D InIm. The underwater objects were detected using YOLOv4, and the temporally encoded optical signals were classified using the CNN-BiLSTM model in Sec. 3.2.2. Both models were trained on clear and turbid water data without occlusion with Beer's coefficient from $\alpha = 0.0025\ \mathrm{mm}^{-1}$ to $\alpha = 0.0101\ \mathrm{mm}^{-1}$ and tested on clear and turbid water data with occlusion from $\alpha = 0.0027\ \mathrm{mm}^{-1}$ to $\alpha = 0.0391\ \mathrm{mm}^{-1}$. Examples of 3D reconstructed underwater testing scenes of two objects for object detection and a red LED signal for temporal signal detection in partial occlusion are shown in Figs. 10(b), 10(c), 10(e), 10(f), 10(h), and 10(j), respectively. The proposed approach is a comprehensive underwater sensing system that can localize underwater objects and permits optical communication for signal transmission and detection. Furthermore, the system can be installed on underwater robots to navigate, acquire data, and identify objects. Therefore, the system has benefits in various applications such as coastal monitoring, underwater visualization, and underwater signal detection.

4. Results and Discussion

In summary, we compare the previously published InIm-based approaches in underwater imaging scenarios. In Sec. 3.1.1, the turbid images were processed using statistical image processing algorithms, namely, histogram matching, histogram equalization, and gamma correction to remedy the effects of turbidity. However, these algorithms generated color pixel errors. Peplography was introduced in Sec. 3.1.2 to solve the artificial color pixel intensities issue in Sec. 3.1.1 using scattering media estimation and ballistic photon detection. This method works in any homogeneous as well as inhomogeneous media. However, the optimal number of ballistic photons is manually chosen. If this parameter is not chosen correctly, the ballistic photons may not be recovered, or saturation can occur in the reconstructed 3D image.

In Sec. 3.2.1, correlation-based filtering was integrated into an active polarimetric-based InIm system for underwater signal detection. However, correlation-based filters are effective for smaller object variations and are not as effective as deep learning approaches. The earliest underwater InIm report using a neural network was highlighted in Sec. 3.3.1, which used object classification to classify underwater objects. However, the system in Sec. 3.3.1 acquired the perspective images using a microlens array, which limited the depth of field and reduced the resolution, thus, resulting in poor quality reconstructed 3D images. This approach can be improved using a camera array or a single camera sensor and with more advanced deep learning algorithms. For capturing high-quality perspective images, the systems in Secs. 3.1.1–3.1.3, 3.2.1–3.2.4, and 3.3.2 do not use a microlens array. Additionally, deep learning algorithms were applied in object visualization (Sec. 3.1.3), signal detection (Secs. 3.2.2–3.2.4), and object detection (Sec. 3.3.2).

Section 3.1.3 incorporated the physical laws into a deep learning model using an unpaired training set to eliminate the need to collect large, paired datasets, whereas Secs. 3.2.2–3.2.4 and 3.3.2 discussed purely data-driven deep learning approaches. The work in Sec. 3.1.3 can be extended to other scattering media with any other physical degradation model as well. Sections 3.2.2 and 3.3.2 implemented a 3D InIm-based system using a CNN-BiLSTM model. The reconstruction of the image frames, along with camera array calibration, depth estimation, and spatial and temporal feature extraction, increased the computational time and introduced errors, thus, influencing the detection performance. Section 3.2.3 eliminated these intermediate steps by directly feeding the captured multi-perspective images to a deep learning model (1DInImCNN) to lower the computational complexity. Section 3.2.4 replaced the lens-based imaging system in Sec. 3.2.3 with a lensless-based imaging system using diffusers and showed that by reducing the dimensionality of the diffuser data, the computational time of the 1DInImCNN

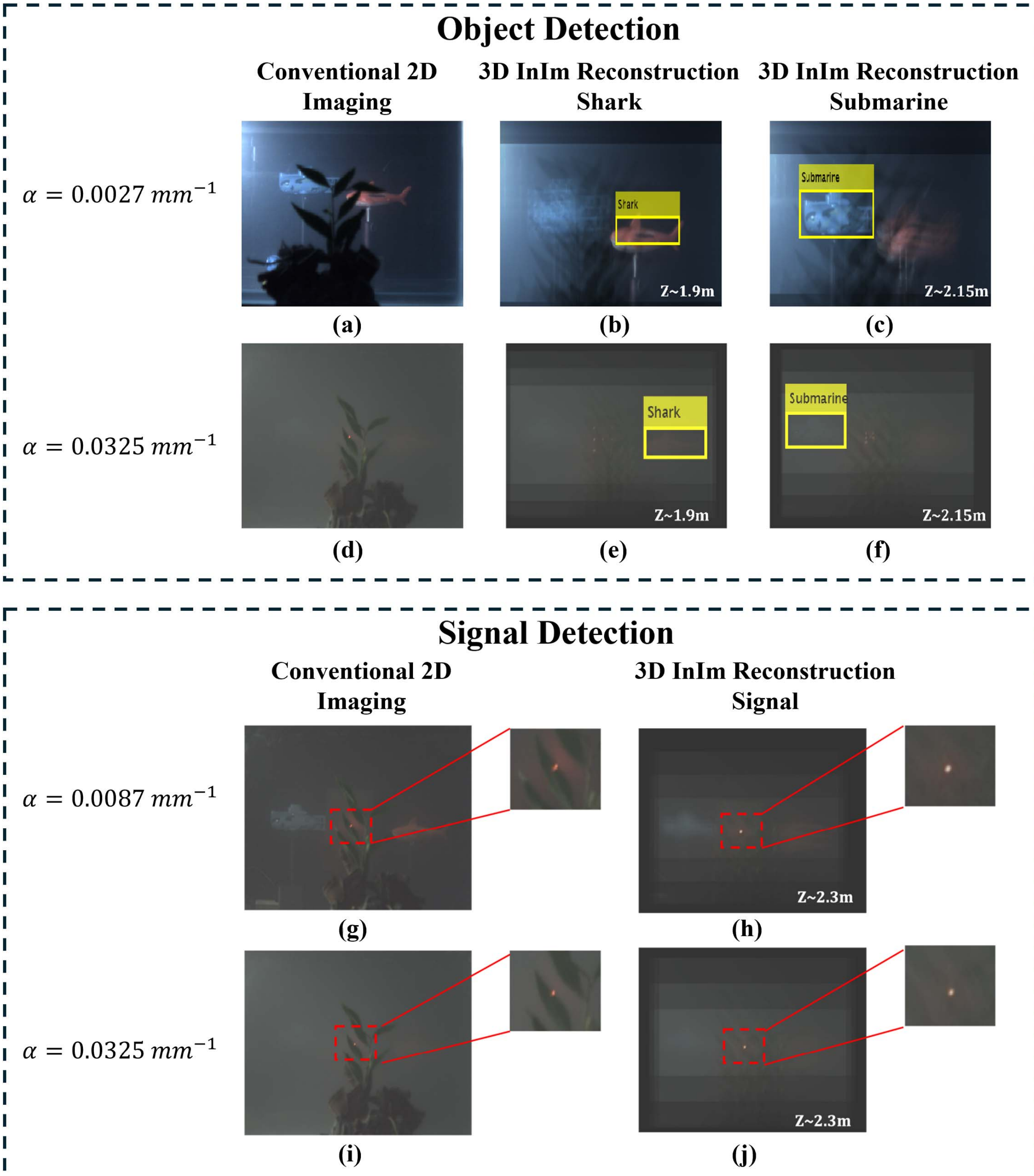

Fig. 10 Sample underwater testing scenes with occlusion applying object detection and signal detection. Reprinted with permission from Ref. [16].

model could be decreased, and the detection performance could be improved substantially. Sections 3.2.1–3.2.4 and 3.3.1 reviewed signal detection-based and object classification-based systems, respectively. Section 3.3.2 reviewed a dual-purpose underwater system that can perform object detection and temporal signal detection simultaneously.

5. Conclusion

We reviewed the published reports on underwater imaging and sensing using 3D InIm for object visualization[18,19,21], temporal signal detection[12–16], object classification[22], and object detection[16]. 3D InIm is shown to be effective in reducing the effect of turbidity, having depth-sectioning capabilities, and working in challenging environments such as partial occlusion. Visualization of objects degraded by scattering and absorption may be remedied using statistical image processing algorithms, peplography, and physics-based deep learning. For signal detection, we reviewed correlation-based filtering using a 4D nonlinear correlation filter, deep learning using a CNN-BiLSTM model, and an end-to-end integrated 1DInImCNN model. We also reviewed YOLOv4 and a distortion-tolerant neural network for object detection and object classification, respectively.

Although LiDAR was briefly discussed, detailed descriptions and reviews of this approach are outside the scope of this manuscript; therefore, we limit our discussions only to 3D InIm-based approaches. Most of our research efforts have focused on 3D InIm in turbidity without occlusion or with partial occlusion, but there are other degraded environments that we have not considered that are of importance in underwater optical imaging and that could be considered in future work such as turbulence and photon sparse environments. All underwater 3D InIm systems covered in this review are lens-based, except for the work in Ref. [15]. To the best of our knowledge, Ref. [15] is the earliest report implementing a lensless-based 3D InIm system using diffusers in underwater imaging. Lensless-based systems are gaining interest in the research community due to their low-cost, compact size, large field of view, and low weight over lens-based imaging systems. However, there is limited research on this subject in underwater imaging. Polarimetric underwater imaging, which uses active illumination, was covered in signal detection using 3D InIm and polarization-difference imaging, object visualization using 3D InIm and active polarization descattering, and in other published reports in turbid environments[57–60]. There are more advanced polarimetric imaging techniques in the underwater literature, which are outside the scope of this paper, such as Stokes-based imaging[61] and Mueller matrix imaging[62,63]. These techniques could be potential future work for underwater 3D InIm-based systems. As in any review papers of this nature, we may have omitted discussions reported in published papers or works by other research groups, for which we apologize in advance.

Acknowledgments

Alex Maric wishes to thank Dr. Timothy O'Connor and Dr. Xin Shen for their discussion on 3D integral imaging. The authors declare no conflicts of interest.

References

1. A. Beer, "Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten," *Ann. Phys. Lpz.* **162**, 78 (1852).
2. J. S. Jaffe, "Computer modeling and the design of optimal underwater imaging systems," *IEEE J. Ocean. Eng.* **15**, 101 (1990).
3. M. Massot-Campos and G. Oliver-Codina, "Optical sensors and methods for underwater 3D reconstruction," *Sensors* **15**, 31525 (2015).
4. V. Klemas, "Beach profiling and lidar bathymetry: an overview with case studies," *J. Coast. Res.* **27**, 1019 (2011).
5. C.-K. Wang and W. D. Philpot, "Using airborne bathymetric lidar to detect bottom type variation in shallow waters," *Remote Sens. Environ.* **106**, 123 (2007).
6. M. Shangguan *et al.*, "Day and night continuous high-resolution shallow water depth detection with single-photon underwater lidar," *Opt. Express* **31**, 43950 (2023).
7. G. Mandlburger, M. Pfennigbauer, and N. Pfeifer, "Analyzing near water surface penetration in laser bathymetry—a case study at the river Pielach," *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **II-5/W2**, 175 (2013).
8. G. Lippmann, "La photographie intégrale," *C. R. Acad. Sci.* **146**, 446 (1908).
9. M. Martínez-Corral and B. Javidi, "Fundamentals of 3D imaging and displays: a tutorial on integral imaging, light-field, and plenoptic systems," *Adv. Opt. Photonics* **10**, 512 (2018).
10. S. Komatsu, A. Markman and B. Javidi, "Optical sensing and detection in turbid water using multidimensional integral imaging," *Opt. Lett.* **43**, 3261 (2018).
11. R. Joshi *et al.*, "Optical 4D signal detection in turbid water by multi-dimensional integral imaging using spatially distributed and temporally encoded multiple light sources," *Opt. Express* **28**, 10477 (2020).
12. R. Joshi *et al.*, "Signal detection in turbid water using temporally encoded polarimetric integral imaging," *Opt. Express* **28**, 36033 (2020).
13. G. Krishnan *et al.*, "Optical signal detection in turbid water using multidimensional integral imaging with deep learning," *Opt. Express* **29**, 35691 (2021).
14. Y. Huang *et al.*, "End-to-end integrated pipeline for underwater optical signal detection using 1D integral imaging capture with a convolutional neural network," *Opt. Express* **31**, 1367 (2023).
15. Y. Huang *et al.*, "Underwater optical signal detection system using diffuser-based lensless imaging," *Opt. Express* **32**, 1489 (2024).
16. R. Joshi *et al.*, "Underwater object detection and temporal signal detection in turbid water using 3D integral imaging and deep learning," *Opt. Express* **32**, 1789 (2024).
17. R. Schulein and B. Javidi, "Underwater multi-view three-dimensional imaging," *J. Disp. Technol.* **4**, 351 (2008).
18. M. Cho and B. Javidi, "Three-dimensional visualization of objects in turbid water using integral imaging," *J. Disp. Technol.* **6**, 544 (2010).
19. M. Cho and B. Javidi, "Peplography—a passive 3D photon counting imaging through scattering media," *Opt. Lett.* **41**, 5401 (2016).
20. R. Joshi and B. Javidi, "Three-dimensional integral imaging visualization in scattering medium with active polarization descattering," in *Optica Imaging Congress (3D, COSI, DH, FLatOptics, IS, pcAOP)* (2023), paper JTu4A.39.
21. G. Krishnan *et al.*, "Three-dimensional integral imaging-based image descattering and recovery using physics informed unsupervised CycleGAN," *Opt. Express* **32**, 1825 (2024).
22. R. Schulein, C. M. Do, and B. Javidi, "Distortion-tolerant 3D recognition of underwater objects using neural networks," *J. Opt. Soc. Am. A* **27**, 461 (2010).
23. K. O'Shea and R. Nash, "An introduction to convolutional neural networks," arXiv:1511.08458 (2015).
24. J. Goodfellow *et al.*, "Generative adversarial nets," arXiv: 1406.2661 (2014).
25. R. M. Schmidt, "Recurrent neural networks (RNNs): a gentle introduction and overview," arXiv:1912.05911 (2019).
26. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. (Prentice Hall, 2008).
27. J. W. Goodman, *Statistical Optics*, 2nd ed. (John Wiley & Sons, Inc., 2015).
28. T. Treibitz and Y. Y. Schechner, "Active polarization descattering," *IEEE Trans. Pattern Anal. Mach. Intell.* **31**, 385 (2009).
29. B. Javidi and D. Painchaud, "Distortion-invariant pattern recognition with Fourier-plane nonlinear filters," *Appl. Opt.* **35**, 318 (1996).
30. B. Javidi, "Generalization of the linear matched filter concept to nonlinear matched filters," *Appl. Opt.* **29**, 1215 (1990).
31. B. Javidi, "Nonlinear joint power spectrum based optical correlation," *Appl. Opt.* **28**, 2358 (1989).
32. B. Tavakoli, B. Javidi, and E. Watson, "Three-dimensional visualization by photon counting computational integral imaging," *Opt. Express* **16**, 4426 (2008).
33. A. Markman and B. Javidi, "Learning in the dark: 3D integral imaging object recognition in very low illumination conditions using convolutional neural networks," *OSA Contin.* **1**, 373 (2018).
34. G. Bradski and A. Kaehler, *Learning OpenCV: Computer Vision with the OpenCV Library*, 1st ed. (O'Reilly Media, 2008).
35. V. Boominathan *et al.*, "Recent advances in lensless imaging," *Optica* **9**, 1 (2022).

36. X. Li *et al.*, "Polarimetric imaging through scattering media: a review," *Front. Phys.* **10**, 815296 (2022).
37. M. P. Rowe *et al.*, "Polarization-difference imaging: a biologically inspired technique for observation through scattering media," *Opt. Lett.* **20**, 608 (1995).
38. J. S. Tyo *et al.*, "Target detection in optically scattering media by polarization-difference imaging," *Appl. Opt.* **35**, 1855 (1996).
39. J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. (Springer, 1985).
40. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* **521**, 436 (2015).
41. J.-Y. Zhu *et al.*, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *IEEE International Conference on Computer Vision (ICCV)* (2017), p. 2242.
42. X. Chen *et al.*, "InfoGAN: interpretable representation learning by information maximizing generative adversarial nets," arXiv: 1606.03657 (2016).
43. K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," in *IEEE Conference on Computer Vision and Pattern Recognition* (2009), p. 1956.
44. A. Candelieri, "A gentle introduction to Bayesian optimization," in *Winter Simulation Conference (WSC)* (2021), p. 1.
45. R. Gold, "Optimal binary sequences for spread spectrum multiplexing (Corresp.)," *IEEE Trans. Inf. Theory* **13**, 619 (1967).
46. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.* **9**, 1735 (1997).
47. M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.* **45**, 2673 (1997).
48. J. Y.-H. Ng *et al.*, "Beyond short snippets: deep networks for video classification," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015), p. 4694.
49. C. Szegedy *et al.*, "Going deeper with convolutions," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015), p. 1.
50. S. Kishk and B. Javidi, "Improved resolution 3D object sensing and recognition using time multiplexed computational integral imaging," *Opt. Express* **11**, 3528 (2003).
51. D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics* **21**, 6 (2020).
52. P. M. Douglass, T. O'Connor, and B. Javidi, "Automated sickle cell disease identification in human red blood cells using a lensless single random phase encoding biosensor and convolutional neural networks," *Opt. Express* **30**, 35965 (2022).
53. X. Wu, D. Sahoo, and S. C. H. Hoi, "Recent advances in deep learning for object detection," *Neurocomputing* **396**, 39 (2020).
54. A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: optimal speed and accuracy of object detection," arXiv:2004.10934 (2020).
55. J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A comprehensive review of YOLO architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO-NAS," *Mach. Learn. Knowl.* **5**, 1680 (2023).
56. I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. (Springer, 2002).
57. M. Dubreuil *et al.*, "Exploring underwater target detection by imaging polarimetry and correlation techniques," *Appl. Opt.* **52**, 997 (2013).
58. J. Guan and J. Zhu, "Target detection in turbid medium using polarization-based range-gated technology," *Opt. Express* **21**, 14152 (2013).
59. F. Liu *et al.*, "Deeply seeing through highly turbid water by active polarization imaging," *Opt. Lett.* **43**, 4903 (2018).
60. T. Yu *et al.*, "Underwater polarization imaging for visibility enhancement of moving targets in turbid environments," *Opt. Express* **31**, 459 (2023).
61. H. Hu *et al.*, "Polarimetric image recovery in turbid media employing circularly polarized light," *Opt. Express* **26**, 25047 (2018).
62. J. Guan, M. Ma, and P. Sun, "Optimization of rotating orthogonal polarization imaging in turbid media via the Mueller matrix," *Opt. Lasers Eng.* **121**, 104 (2019).
63. H. Wang *et al.*, "Underwater imaging by suppressing the backscattered light based on Mueller matrix," *IEEE Photonics J.* **13**, 7800106 (2021).