Research Article Vol. 32, No. 5 / 26 Feb 2024 / Optics Express 7495
Optics EXPRESS

3D object tracking using integral imaging with mutual information and Bayesian optimization

PRANAV WANI, KASHIF USMANI, GOKUL KRISHNAN, AND BAHRAM JAVIDI*

Electrical and Computer Engineering Department, University of Connecticut, 371 Fairfield Road, Storrs, Connecticut 06269, USA
Bahram.javidi@uconn.edu

Abstract: Integral imaging has proven useful for three-dimensional (3D) object visualization in adverse environmental conditions such as partial occlusion and low light. This paper considers the problem of 3D object tracking. Two-dimensional (2D) object tracking within a scene is an active research area. Several recent algorithms use object detection methods to obtain 2D bounding boxes around objects of interest in each frame. Then, one bounding box can be selected out of many for each object of interest using motion prediction algorithms. Many of these algorithms rely on images obtained using traditional 2D imaging systems. A growing literature demonstrates the advantage of using 3D integral imaging instead of traditional 2D imaging for object detection and visualization in adverse environmental conditions. Integral imaging's depth sectioning ability has also proven beneficial for object detection and visualization. Integral imaging captures an object's depth in addition to its 2D spatial position in each frame. A recent study uses integral imaging for the 3D reconstruction of the scene for object classification and utilizes the mutual information between the object's bounding box in this 3D reconstructed scene and the 2D central perspective to achieve passive depth estimation. We build over this method by using Bayesian optimization to track the object's depth in as few 3D reconstructions as possible. We study the performance of our approach on laboratory scenes with occluded objects moving in 3D and show that the proposed approach outperforms 2D object tracking. In our experimental setup, mutual information-based depth estimation with Bayesian optimization achieves depth tracking with as few as two 3D reconstructions per frame which corresponds to the theoretical minimum number of 3D reconstructions required for depth estimation. To the best of our knowledge, this is the first report on 3D object tracking using the proposed approach.

© 2024 Optica Publishing Group under the terms of the Optica Open Access Publishing Agreement

1. Introduction

Two-dimensional (2D) object tracking aims at estimating bounding boxes and the identities of the objects in video or image frame sequences. Object tracking has a wide range of applications. Most state-of-the-art methods incorporate a tracking-by-detection framework [1]. It involves an independent detector that is applied to each frame to obtain likely detections. Next, a tracker (motion predictor) attempts to perform data association, where the aim is to associate detections across frames. Here one bounding box is selected out of many for each object of interest. To aid the data association process, trackers use various methods for modeling the motion [2,3] and appearance [4,5] of objects in the scene. One of the most commonly used tracking algorithms, simple online and real-time tracking (SORT) [1], uses a lean implementation of the tracking-by-detection framework. Objects are represented only by bounding boxes, ignoring other visual appearance features beyond the detection components. Its tracker uses detections from previous and current frames to track multiple objects in real time. It uses the bounding boxes' position and size for motion estimation and data association through frames. A recently established multi-target tracking benchmark [6] suggests that detection quality plays the most significant role in overall object tracking accuracy. Accordingly, SORT uses recent advances in

#517312 https://doi.org/10.1364/OE.517312
Journal © 2024 Received 28 Dec 2023; revised 4 Feb 2024; accepted 5 Feb 2024; published 15 Feb 2024

Optics EXPRESS

visual object detection to solve the detection problem directly rather than aiming to be robust to detection errors. It uses the Faster Region CNN (FrRCNN) [7] detection framework along with the Kalman filter [8] and Hungarian method [9] for motion prediction and data association components of the tracking problem respectively. This method achieves 74.6 multiple object tracking accuracy on the MOT17 dataset with 30 frames per second (FPS). Despite good overall performance, SORT fails in many challenging scenarios like occlusions. To overcome these limitations DeepSORT [10] replaces the bounding box-based association metric with a more informed metric that combines motion and appearance information. It uses a convolutional neural network to obtain a feature vector that can be used to represent a given image. This slightly increases the robustness of the algorithm against misses and occlusions. In recent years several new methods have been proposed with a similar framework. Some of the most prominent are FairMOT [11], TransMOT [12], and ByteTrack [13]. These methods achieve 25.9, 9.6, and 30 frames per second on the MOT17 dataset. For all these methods, object detection accuracy plays the most important role in the final tracking accuracy [10–14].

3D integral imaging [15,16] is a prominent imaging technique that captures angular information about the object scene by recording multiple 2D elemental images from diverse perspectives [17–27]. It has several advantages over other imaging modalities for object detection in adverse environmental conditions like low illumination, partial occlusion, or fog [28,29]. Additionally, integral imaging's depth sectioning ability also aids in object detection and data association tasks. The use of integral imaging for object detection requires the knowledge of the object's depth that can be acquired using different methods such as minimum variance estimate. A recent study [30,31] uses mutual information for passive depth estimation in degraded environments. In this manuscript, we advance this method, by using Bayesian optimization for sample-efficient dynamic tracking of object's depth. The scope of this manuscript is limited to demonstrating the integral imaging-based dynamic depth tracking of an object using Bayesian optimization and mutual information. A rigorous study of its performance or applications is not considered here. We postulate that the adoption of integral imaging in place of traditional 2D imaging systems can enhance the performance of object-tracking methods in adverse environmental conditions. A change to the 3D imaging modality can work in conjunction with the numerous algorithmic advances mentioned previously to provide enhanced performance.

Mutual information-based depth estimation gives mutual information values that vary with depth. Their maxima are located at the objects' true depths. These objective functions have some common underlying structures that lend them as good candidates for Bayesian optimization. Bayesian optimization is a global optimization methodology for sample-efficient evaluation of expensive-to-evaluate objective functions [32–34]. It iteratively builds a statistical model of an objective function according to all the past observations and selects the next evaluation point by maximizing some acquisition function. It has been successfully used in several domains like simulations [35,36], machine learning [37–39], and reinforcement learning [40]. Several other methods exist for multi-modal optimization like gradient descent, quasi-Newton [41,42], or simplex method [43]. However, they require an analytical form of an objective function and tend to get trapped in a local optimum. Evolutionary optimization methods like genetic algorithms [44–46], clonal selection algorithms [47], or artificial immune networks [48,49] can be used in domains with no available analytical expressions. However, these methods rely on heuristic approaches and require many expensive function evaluations. This prevents their application in our use case.

In this paper, we consider the problem of 3D object tracking using integral imaging. Data is captured as video or image frames. We alternatively track the object's lateral and depth locations in each image frame. We use a deep convolutional neural network to track the object's 2D spatial location within a 3D reconstructed image frame of a scene and use mutual information and Bayesian optimization to track the object's depth. We experimentally evaluate the performance

Optics EXPRESS

of our approach for the task of depth tracking of an object moving with bounded velocity and acceleration. Our experimental setup contains a laboratory scene with an object free to move in three dimensions. In our experiments, we achieve depth tracking with as few as two 3D integral imaging reconstructions per frame. It equates to the minimum number of 3D reconstructions required theoretically for depth estimation with our approach as one 3D reconstruction leads to a correspondence problem. A rigorous study of the performance or applications of our proposed method is not considered here, as it is outside of the scope of this manuscript.

2. Integral imaging-based depth estimation

2.1. Integral imaging

Integral imaging is a prominent passive 3D imaging approach. It gains information about the scene's light field by multiplexing the diverse perspectives of the 2D elemental images. These 2D elemental images are recorded by using either a camera array, a single camera mounted on a translation stage, or a single imaging sensor with a lenslet array [17–27]. 2D elemental images are back-propagated through a virtual pinhole to computationally re-compute 3D scenes. Faithful 3D reconstruction can be obtained at any depth within the depth of fields of the 2D elemental images. Multiple perspectives from the 2D elemental images help mitigate the effects of partial occlusion in the 3D reconstructed scene. 3D integral imaging is optimal in the maximum likelihood sense for read-noise dominant images that can occur in photon-starved conditions [50–53]. This enables the 3D reconstructed scenes to have a better signal-to-noise ratio. An overview of recent advances in integral imaging can be found in [15,16].

Our experimental setup consists of an integral imaging system [54] with an image sensor array for image capture. The pickup stage of integral imaging is shown in Fig. 1(a). Once the 2D elemental images are captured, the 3D reconstruction of the scene can be achieved computationally as shown in Fig. 1(b). Figures 1(c) and (d) show the integral imaging pickup and reconstruction process with a single sensor and a lenslet array.

3D reconstruction is achieved by back-propagating the captured 2D elemental images through a virtual pinhole. Reconstructed 3D scene intensity $I_z(x,y)$ is computed as [54]:

$$I_z = \frac{1}{O(x,y)} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left[I_{mn} \left(x - \frac{m \times L_x \times p_x}{c_x \times \frac{z}{f}}, y - \frac{n \times L_y \times p_y}{c_y \times \frac{z}{f}} \right) + \varepsilon \right] \tag{1}$$

where (x, y) is the pixel indices, $O(x, y)$ is the number of overlapping pixels in (x, y). (c_x, c_y), (p_x, p_y), and (L_x, L_y) represent the sensor size, the pitch size between cameras, and the resolution of the camera sensor, respectively. I_{mn} is a 2D elemental image, with (m, n) representing its indices, and (M, N) representing the total number of elemental images. f is the focal length of the camera lens, and z is the reconstruction distance of the 3D object from the camera array. ε is the additive camera noise. Assuming that the incoming rays originate from a lambartian surface, 3D reconstruction at the true depth will minimize the variation of these rays [55].

2.2. Experimental setup

Our experimental synthetic aperture integral imaging uses 25 cameras in a 5×5 configuration as shown in Fig. 1(a) and (b). The camera array pitch size is 50 mm in both lateral directions. The object to be tracked can move from 1000 mm to 8000 mm distance from the integral imaging setup along the optical axis. It can also move 1500 mm perpendicular to the optical axis on all sides. We restrict the movement of the object within the field of view of all the integral imaging cameras. The focal length of each camera lens is 50 mm and the lens diameter is 40 mm. The sensor has 2048×2048 pixels with each pixel being 6.5×6.5 micrometers. These system parameters can affect the performance of integral imaging in a complex manner. As an example, more cameras can increase integral imaging's depth sectioning ability, albeit, with

Optics EXPRESS

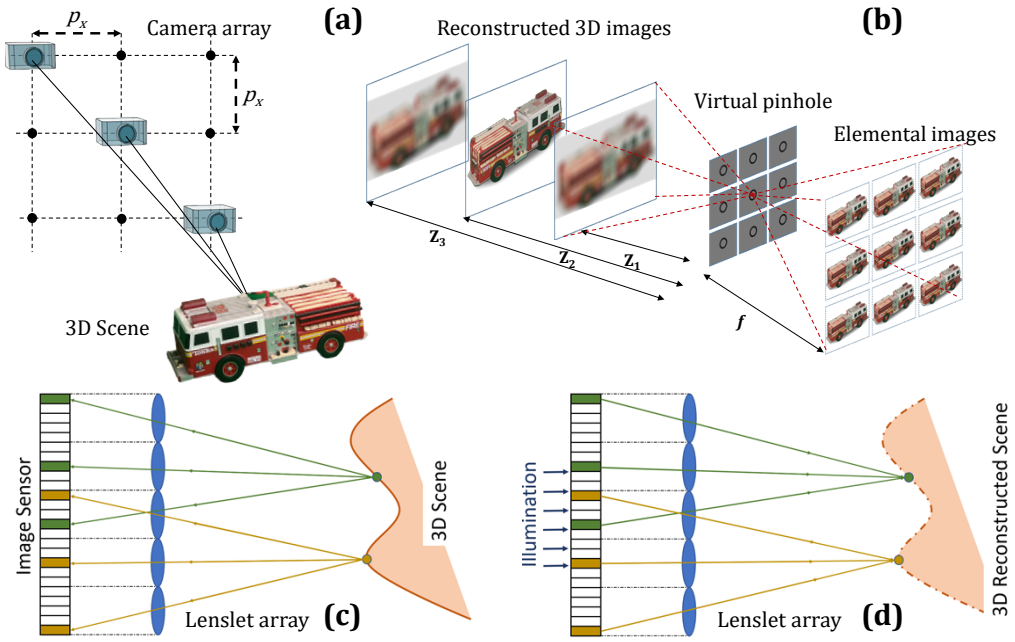

Fig. 1. (a) Integral imaging setup using a camera array for the image pickup process. (b) The reconstruction process of the integral imaging setup of (a). (c) Integral imaging setup using a lenslet array and a single imaging sensor. (d) The reconstruction process of the integral imaging setup of (c).

a corresponding increase in image processing time. Analysis of the effects of these integral imaging system parameters is not considered here. Figure 2(a) shows a sample 2D elemental image of our experimental scene. Its corresponding 3D reconstructions at the plane of the truck and soccer balls are shown in Fig. 2(b) and (c), respectively. Objects in focus are shown in the red box.

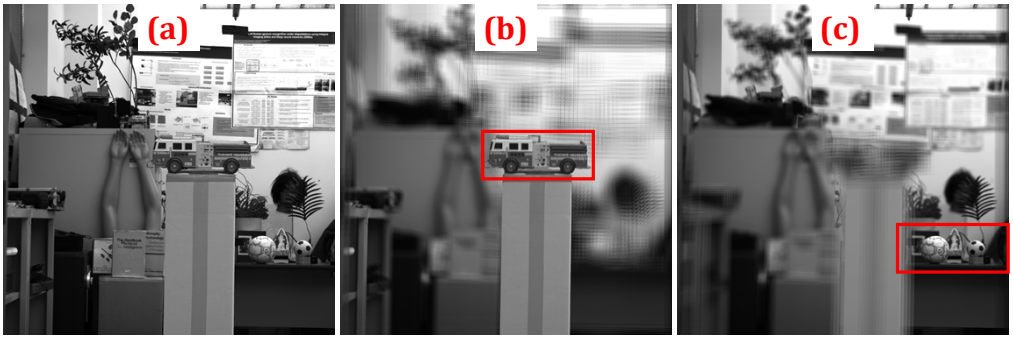

Fig. 2. Integral imaging experimental scenes. (a) Sample 2D central elemental image of one of the scenes. (b) 3D integral imaging reconstruction at the plane of the truck (highlighted in red colored box). (c) 3D integral imaging reconstruction at the plane of the soccer balls.

Optics EXPRESS

2.3. Mutual information

Mutual information between two random images $\boldsymbol{X}$ and $\boldsymbol{Y}$ is defined in terms of the probability density function of the pixel values [56]:

$$MI(X;Y) = \sum_{g_1 \in I} \sum_{g_2 \in I} f_{xy}(g_1, g_2) \log \frac{f_{xy}(g_1, g_2)}{f_x(g_1) f_y(g_2)} \tag{2}$$

Here g_1 and g_2 are the pixel intensity values in images $\boldsymbol{X}$ and $\boldsymbol{Y}$. These variables are free to take any value from a set of available pixel intensity values denoted by I. However, this formulation (pixel-to-pixel correspondence) fails to capture the spatial information that exists in an image. Several investigations on image registration found that lack of spatial information leads to poor robustness to environmental degradations like noise and experimental factors like misalignment errors [57]. Attempts were made to rectify this by incorporating additional spatial information like gradients [58] or by using higher-order mutual information [59]. However, these lead to an exponential rise in data and computational requirements. Some authors tried to mitigate it using dimensionality reduction techniques like principle component analysis or independent component analysis [57]. A more promising approach for incorporating spatial information without an exponential rise in data or computational requirements relies on graph theory [60]. It uses the Gibbs random field formulation which states that the conditional probabilities of a site's gray level corresponding to its neighborhood are proportional to the exponential sum of the potentials of its associated cliques. Thus, different pixel intensity neighborhood configurations that produce the same potential $U(x)$ can be grouped as a single state α. Mutual information between two images is then given as [60]:

$$MI(X;Y) = \sum_{g_x \in I} \sum_{g_y \in I} \sum_{\alpha_x} \sum_{\alpha_y} f_{xy}(g_x, \alpha_x, g_y, \alpha_y) \log \frac{f_{xy}(g_x, \alpha_x, g_y, \alpha_y) f_x(\alpha_x) f_y(\alpha_y)}{f_{xy}(\alpha_x, \alpha_y) f_x(\alpha_x, g_x) f_y(\alpha_y, g_y)} \tag{3}$$

Here I is the set of pixel intensities. α_x and α_y are the unique states corresponding to different neighborhood configurations that produce the same potential. g_x and g_y are the intensity values of pixels. This approach was adopted by [30,31,61] for 3-bit images with one adjacent neighborhood used for spatial information. This gives $I = \{0,1,2,3,4,5,6,7\}$ and the number of α equals nine. Thus, the total combination of the pairs (α, g) is 72. We use this formulation of mutual information henceforth.

We obtain mutual information curves for objects as a function of depth by computing mutual information between object bounding boxes in the 3D integral imaging reconstructed scenes

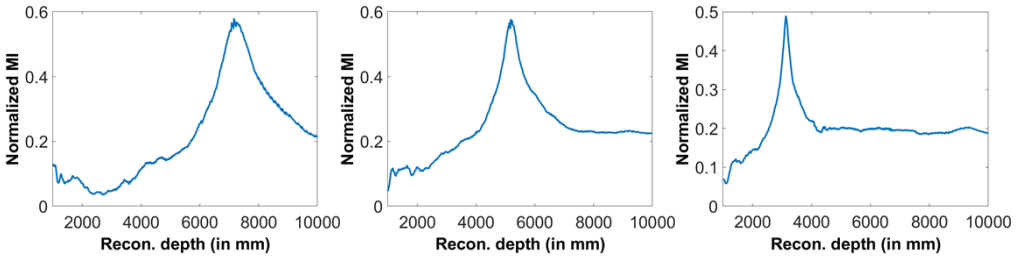

Fig. 3. Passive depth estimation with integral imaging using normalized mutual information (MI). Sample mutual information curves vs. object reconstruction depth (recon. depth) for the truck (see Fig. 2(b)) as it is placed at different depths from the integral imaging setup. The curves are obtained by computing normalized mutual information between the object's bounding box in the 3D reconstructed scene and its corresponding box in the 2D central perspective. Peaks correspond with the true depth of the object. These curves correspond to objective functions in an optimization context.

Optics EXPRESS

and corresponding bounding boxes in the 2D central perspective (central elemental image). The maximum of these curves corresponds to the true depth of the objects. Figure 3 shows sample mutual information curves for the truck (see Fig. 2(b)) as it is placed at different depths. These curves correspond to objective functions in an optimization context.

3. Bayesian optimization-based depth tracking

3.1. Background

Bayesian optimization was initially developed by Kushner [62], Zilinskas [63,64], and Mockus [65,66]. It was further adopted for multi-fidelity optimization [67,68] and multi-objective optimization [69–71]. Bayesian optimization aims to achieve sample-efficient optimization of expensive to evaluate objective functions [34,72]. It is especially beneficial when, as in our case, no closed-form representation is available and only noisy point-based observation is possible.

We model the depth-tracking as a spatiotemporal optimization problem, that is, to find $x^*(t) = \arg\max_{x\in \mathrm{X}} f(x,t)$, where $\mathrm{X} \subset \mathbb{R}^d$ is a compact set. Bayesian optimization works well when the domain of x i.e. $\mathrm{X} \in \mathbb{R}^d$ has dimensions less than 20 and the objective function f is continuous. Neither mutual information curves nor their derivatives have closed-form analytical expressions. However, the derivative information, if available, can aid Bayesian optimization [73]. Bayesian optimization performs a sequential search, and at every iteration k selects a new location x_{k+1} to evaluate f and observe its value. Gaussian process regression, the most commonly used surrogate model for Bayesian optimization [74], provides the posterior distribution according to previous observations. The sequential selection is handled by an acquisition function $a : \mathrm{X} \rightarrow \mathbb{R}$, defined over the posterior of the Gaussian process.

A spatiotemporal optimization problem (e.g. depth tracking) requires time-dependent Bayesian optimization. Only limited research exists on this topic. [75] Introduces look-ahead acquisition functions, which are modified acquisition functions designed to predict a time-varying optimum at target horizon T. This formulation does not suit our needs. It tries to optimally predict the optimum at some target horizon T while sacrificing its optimum tracking ability for times before T. However, depth-tracking requires tracking the optimum for every time interval, at least for some restricted environmental conditions like bounded velocity and acceleration assumption. [76] Uses the standard sequential Bayesian optimization framework and models the objective functions with the Gaussian process prior whose evolution follows a simple Markov model. They discard the stale data samples of the time-varying objective function to adapt to its changes. Although this formulation suits our needs, it has a very limited expressiveness as it fails to capture the non-linear evolution of the time-dependent objective function. It also performs poorly on our experimental data. [77] Represents a time-dependent observable as a vector of m components corresponding to different time instances. This vector is modeled by a Gaussian process with m outputs. This formulation is inefficient in terms of the number of required observations (data samples), as high dimensional Bayesian optimization is unstable for less number of observations. [78] Model dynamic objective functions using spatiotemporal Gaussian process priors which capture all the instances of function over time. Information learned from this model is used to guide the tracking of a temporally evolving optimum. We adopt this formulation for our depth tracking problem as it provides the most promising results. Further discussions are provided in subsequent sections.

3.2. Spatiotemporal Bayesian optimization

Spatiotemporal Bayesian optimization relies on spatiotemporal Gaussian process priors as surrogate functions. Gaussian process [79,80] is a collection of random variables $\{F_{x1,t1}, F_{x2,t2}, \ldots\}$ for which any finite subset has a joint multivariate normal distribution. Thus, for any finite length vector $\mathbf{x} = [\{x_1, t_1\}, \{x_2, t_2\}, \ldots, \{x_n, t_n\}]^T$ its corresponding observation values

Optics EXPRESS

$\mathbf{F_x} = [F_{x1,t1}, F_{x2,t2}, \ldots, F_{xn,tn}]$ are jointly normally distributed:

$$\mathbf{F_x} \sim \mathrm{N}\{\mu_0(\mathbf{x}), k(\mathbf{x},\mathbf{x})\} \tag{4}$$

Here elements of $\mu_0(\mathbf{x})$ are given by a prior mean function $\mu_0(\{x_i, t_i\})$, and k is the kernel function. for k to be a valid kernel $k(\mathbf{x},\mathbf{x})$ needs to be a square, positive semi-definite matrix for any $\mathbf{x}$[81]. Values $\mathbf{F_x}$ are obtained by noisily observing the function $f(x,t)$ at indices $\mathbf{x} = \{x_i, t_i\}$, i.e. $F_{x,t} = f(x,t) + \varepsilon$, where $\varepsilon \sim \mathrm{N}(0, \sigma_\varepsilon^2)$ is independent and identically distributed (i.i.d.). The spatiotemporal Gaussian process regression infers the posterior of f given the observations $\mathbf{F_x}$. The posterior distribution at some new point $z \in \{X, T\}$ is Gaussian with mean and variance [79,80]:

$$\mu(F_z|F_\mathbf{x} = \mathbf{f}) = \mu_0(z) + k(z,\mathbf{x})(k(\mathbf{x},\mathbf{x}) + \sigma_n^2 I)^{-1}(\mathbf{f} - \mu_0(\mathbf{x})) \tag{5}$$

$$\sigma^2(F_z|F_\mathbf{x} = \mathbf{f}) = k(z,z) - k(z,\mathbf{x})(k(\mathbf{x},\mathbf{x}) + \sigma_n^2 I)^{-1}k(\mathbf{x},z) \tag{6}$$

The kernel matrix $k(\mathbf{x},\mathbf{x}) + \sigma_n^2 I$ depends only on the observed values and is Cholesky factored instead of inverted. In the absence of observation noise σ_n, a small number must be to the diagonal of $k(\mathbf{x},\mathbf{x})$ to prevent the eigenvalues from approaching zero. The posterior mean is a linear combination of n kernel functions, each one centered at an observed data point.

The kernel function k dictates the structure of the response functions that we can fit. For example, a periodic kernel function is good for a periodic response function. We assume the kernel function to be stationary, i.e. $K(\{x_1,t_1\},\{x_2,t_2\}) = K(\{x_1 - x_2, t_1 - t_2\})$ and full symmetric, i.e. $K(\{x_1,t_1\},\{x_2,t_2\}) = K(\{x_2,t_2\},\{x_1,t_1\})$. Additionally, we assume the spatiotemporal kernel function to be separable, i.e. decoupled into purely spatial and purely temporal factors, i.e. $K(\{x_1,t_1\},\{x_2,t_2\}) = K_S(x_1,x_2) \odot K_T(t_1,t_2)$, where $\odot$ is the element wise or the Hadamard product. We adopt the commonly used squared exponential kernel:

$$k_{\text{squared\_exponential}}(x_i,t_i,x_j,t_j|\theta) = \sigma_f^2 \exp\left(-\frac{1}{2}\frac{(x_i - x_j)^T(x_i - x_j)}{\sigma_{l_x}^2}\right)\exp\left(-\frac{1}{2}\frac{(t_i - t_j)^T(t_i - t_j)}{\sigma_{l_t}^2}\right) \tag{7}$$

Here σ_f is the signal standard deviation. σ_{lx} and σ_{lt} are the spatial and temporal characteristic length scales respectively. These together form the hyperparameter vector of the kernel function denoted by θ. Squared exponential kernels give rise to a Gaussian process whose samples are infinitely differentiable. The kernel function is differentiable with respect to its hyperparameters θ. The marginal likelihood of the data can thus be optimized to compute a maximum likelihood estimate of its hyperparameters.

3.3. Depth tracking

We use Bayesian optimization to build a posterior mutual information distribution using the existing data samples of the latent time-varying mutual information curves (see Fig. 3) and the spatiotemporal Gaussian process prior. This posterior is used to construct an acquisition function that leads the search for a time-varying maximum, which corresponds to the object's depth, by exploring and exploiting the objective function. We cannot make predictions more than one spatial length scale (σ_{lx}) away. Thus, the maximum time gap between frames, or the minimum effective frame rate, will be determined by the object's speed along the optical axis of the integral imaging setup and the temporal length scale of the spatiotemporal kernel function.

In its standard form, Bayesian optimization aims to strike a balance between exploration and exploitation. This, however, does not work well for our problem. Tracking requires a different exploration-exploitation tradeoff than the standard form. Since the goal is to track the maximum, certain exploration steps could heavily penalize the algorithm's performance albeit the importance of exploration in the learning process. For a known budget of Bayesian optimization steps (number of observations), one solution is to allocate a few of those to obtain

Optics EXPRESS

initial samples to learn the function. These "throwaway" number of steps can aid better tracking in future steps. Determining when the learned Gaussian process model of the underlying latent function is good enough to make faithful predictions can be achieved by looking at the rate of change of the characteristic length scale in each iteration [78]. Refrence [78] uses this heuristic criteria to switch between learning, exploring and exploiting, and purely exploiting stages.

Although the hyperparameters (spatial and temporal length scales, and signal standard deviation) can be learned online using the strategy described above and with one of several available methods like maximum likelihood or maximum a posteriori, their accuracy deteriorates significantly for only a few data samples (observations) [31]. We, thus, learn the kernel hyperparameters during a training phase by observing time-varying mutual information curves for multiple objects and multiple trajectories. The spatial characteristic length scale (σ_{lx}) varies with the object's depth from the integral imaging setup, and the temporal characteristic length scale (σ_{lt}) varies with the speed of the object along the optical axis of the integral imaging setup. Thus, a dictionary of spatial and temporal length scales is learned during the training phase. In the tracking phase, we use this learned dictionary of kernel hyper-parameters to construct the spatiotemporal Gaussian process. We then solely focus on exploitation, i.e. to track the maximum without learning the hyper-parameters or searching for other maxima.

Exploration-exploitation tradeoff is handled by acquisition functions that sequentially probe the objective function to get point estimates. The most commonly used are the probability of improvement [62], expected improvement [64], upper confidence bound [82], entropy search [83], predictive entropy search [84], and max-value entropy search [85]. As exploration is not of significance to us, we use the upper confidence bound (UCB) acquisition function $UCB(x) = \mu(x) + \beta\sigma(x)$[82]. It works on the principle of selecting an optimistic point under uncertainty. For every query point x, it uses a fixed-probability best-case scenario according to the underlying probabilistic model. β controls the exploration-exploitation tradeoff. A high value of β enables more exploration while a lower value leads to more exploitation. In our experiments we select β as 0.6, however, we did not observe a significant effect of this parameter on the overall performance of the system.

4. Object tracking experiments

4.1. Methodology

We use the 'You Look Only Once v2' (YOLOv2) neural network for object detection [86,87]. Although newer versions of YOLO deep neural networks exist like YOLOv8, these only provide incremental performance improvement, especially for smaller objects and more complex environments. Additionally, these incremental improvements come at the cost of processing speed. As this manuscript presents only a proof of concept on a single object tracking problem, we use the YOLOv2 network which has a good balance of detection accuracy and speed. However, any state-of-the-art detector can be used for this purpose. The YOLOv2 deep neural network simultaneously locates and classifies objects within a scene. Its architecture is inspired by GoogleNet [88] and has 24 convolution layers with two fully connected layers. YOLOv2 has a high-resolution classification capability. It also utilizes the concept of anchor boxes which enables it to detect multiple objects centered at one grid cell. As a rigorous experimental analysis is outside the scope of this manuscript, we use only a limited laboratory-generated dataset to train the YOLOv2 neural network. We assume a near-constant orientation of the object. However, with more training data, YOLOv2 can be trained to detect objects with random orientations. We start the object tracking by first detecting an object of interest and estimating its depth. In each iteration (or image frame), we 3D reconstruct the scene at the predicted depth using integral imaging and utilize YOLOv2 for object detection and tracking. We then estimate the object's depth by using the 2D bounding box provided by the detector and the depth tracking method described in the previous section. We alternatively keep applying object detection and depth

Optics EXPRESS

tracking in each image frame to achieve full 3D tracking of the object. Figure 4 shows a flowchart summarizing this method. In our approach, the 3D tracking of one object is independent of the 3D tracking of other objects. As such, the time complexity of our approach grows linearly with the number of objects under consideration.

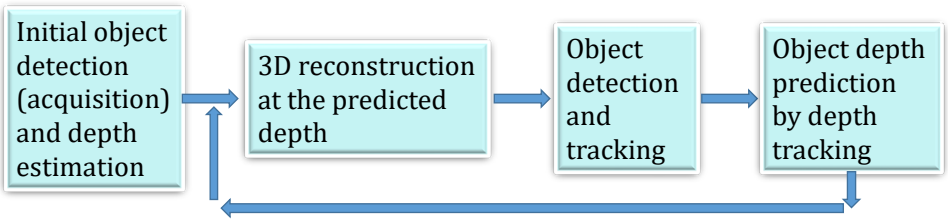

Fig. 4. Flowchart summarizing the 3D object tracking framework using integral imaging. YOLOv2 deep neural network is used for object detection. Object depth tracking is achieved using mutual information and Bayesian optimization.

4.2. Experimental results

Figure 5 shows the depth tracking results for scenes shown in Fig. 2. In accordance with the frame rates of commonly used tracking methods (see Sec. 1), we capture the motion of our object (truck, see Fig. 2) in 30 frames per second (fps). This, however, is not the running speed of our current experimental system as our computational system is not yet optimized. We use two 3D integral imaging reconstructions per frame (rpf) to track the object's depth. 3D object tracking is achieved by alternatively switching between 2D object tracking and depth tracking (see Fig. 4). Figures 5(a) and (d) show the true depth and the tracked depth of the object with time for different motion profiles. Figures 5(b) and (e) show the axial speed profiles of the object corresponding to the depth profiles shown in Fig. 5(a) and (d), respectively. Figures 5(c) and (f) show the true and predicted lateral positions of the object. The spatial location of the object is represented by the object's mid-point pixel coordinates in the captured 2D central elemental image.

As discussed earlier, we cannot make predictions more than a few length scales away. Thus, the maximum time gap between frames, or the minimum effective frame rate, will be determined by the object's speed along the optical axis of the integral imaging setup. Length scale is one of the parameters of the Gaussian process kernel and it signifies the correlation between two points separated in space or time. Length scales for the mutual information curves are proportional to their full widths at half maxima (FWHMs). These depend on integral imaging system parameters, object characteristics, environmental conditions, and depth (axial distance) of the object from the integral imaging setup. Figure 6(a) shows length scales for our experimental mutual information curves as a function of object depth. For example, see Fig. 2 for sample experimental scenes and Fig. 3 for sample mutual information curves. Figure 6(b) shows the maximum allowed axial speed of an object as a function of operational frames per second for a few object depths. These speeds represent the limit wherein depth tracking is more efficient than depth estimation, i.e. previous depths can aid in predicting the current depth.

Maximum axial speeds shown in Fig. 6(b) represent a theoretical upper limit. Figure 6(b) shows that more frames per second are needed for tracking as the object's speed increases. These results are shown for ideal conditions, that is, no environmental degradations. However, several environmental or system parameters affect the depth tracking performance. For example, the number of 3D integral imaging reconstructions allowed per frame has a significant impact on the maximum trackable axial speed. Figure 7 provides an example of the effect of the number of 3D integral imaging reconstructions per frame on depth tracking. Figure 7(a) shows a sample fast axial speed depth profile. The object moves sinusoidally along the optical axis with an average

Optics EXPRESS

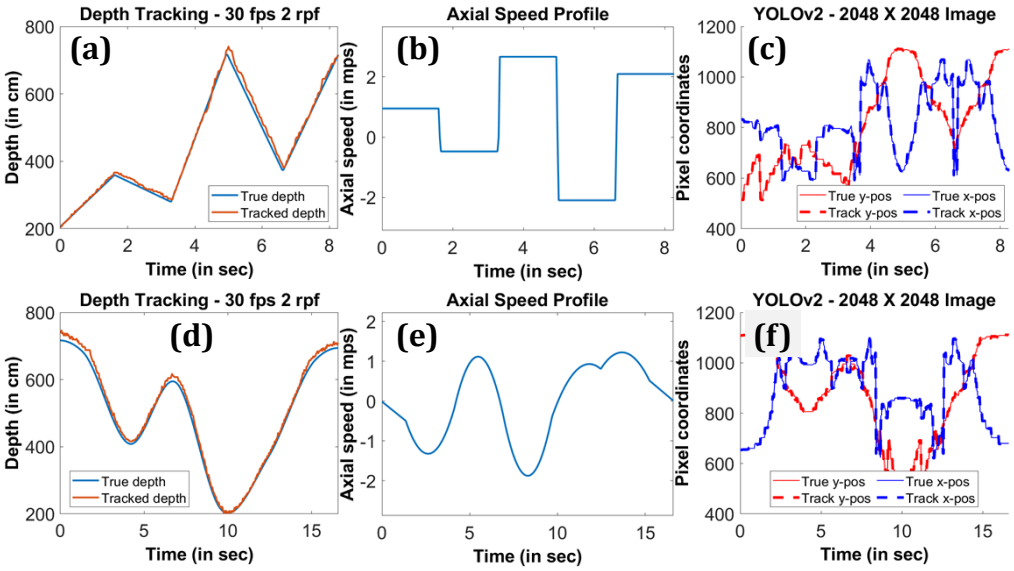

Fig. 5. Experimental depth tracking results for scenes shown in Fig. 2. Motion of the truck is captured with 30 frames per second (fps). Two 3D integral imaging reconstructions per frame (rpf) are used for depth tracking. (a) True and tracked depths of the truck for a sample motion profile. The corresponding depth error has a mean of 9.08 cm and standard deviation of 8.10 cm. (b) The axial speed (measured in meters per second – mps) of the truck corresponding to the depth profile shown in (a). We represent an object moving away from the imaging system with a positive speed and moving closer to the system with a negative speed. (c) True and tracked lateral x and y positions (x-pos, y-pos) of the truck for the same motion profile. (d), (e), and (f) show similar results as (a), (b), and (c) respectively for a different motion profile. The depth error for curve in (d) has a mean of 11.56 cm and standard deviation of 6.75 cm. YOLOv2 is you look only once v2 deep neural network.

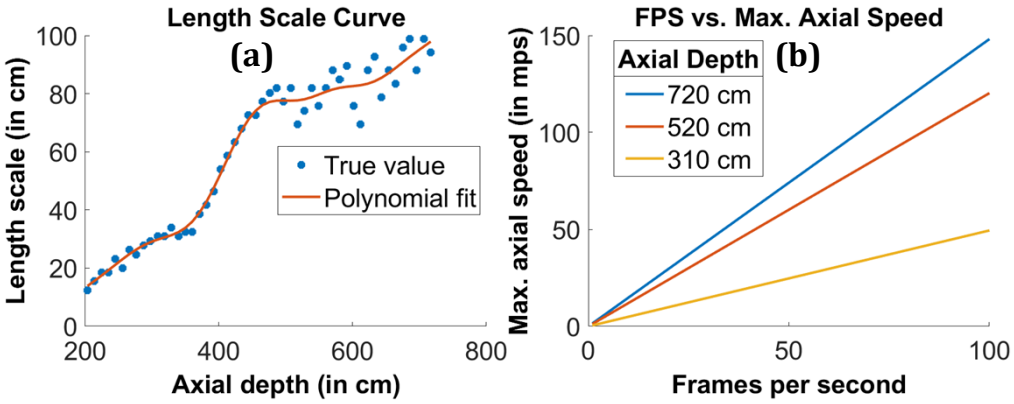

Fig. 6. (a) Length scales for our experimental mutual information curves as a function of object depth. See Figs. 2 and 3 for sample experimental scenes and mutual information curves. (b) Maximum allowed axial speed of an object as a function of effective frames per second (fps) for a few object depths. Mps: meters per second.

Optics EXPRESS

axial speed of approximately 50 km per hour and a maximum axial speed of 100 km per hour (see Fig. 7(b)). Figures 7(c), (d), and (e) show the true depth and the tracked depth of the object with 30 frames per second (fps) and two, four, and six 3D reconstructions per frame (rpf) respectively.

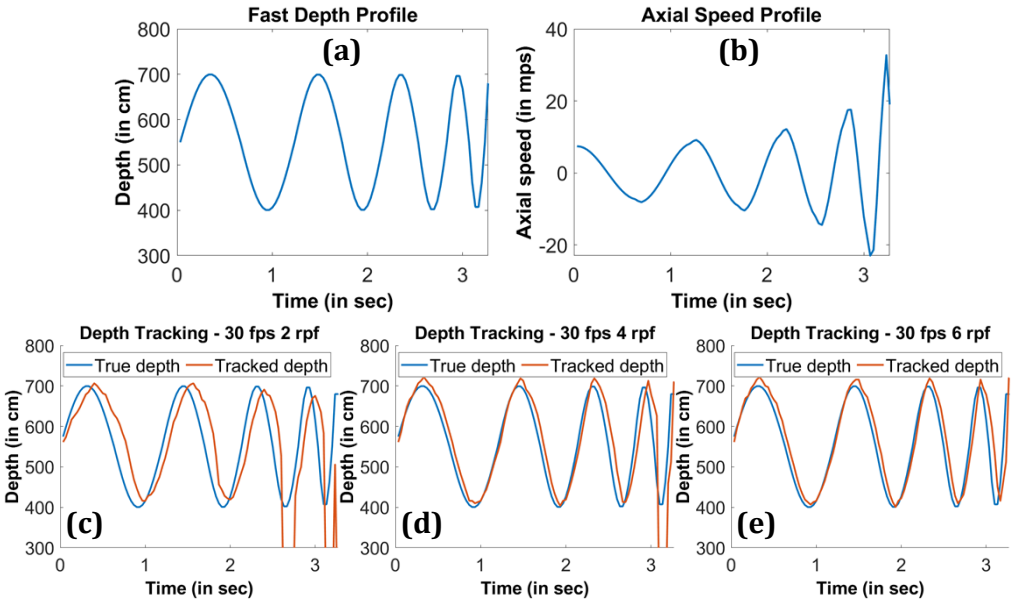

Fig. 7. (a) A sample fast axial speed depth profile. The object moves in a sinusoidal pattern along the integral imaging optical axis. (b) Axial speed profile corresponding to the depth profile shown in (a). We represent an object moving away from the imaging system with a positive speed and moving closer to the system with a negative speed. (c) - (e) True and tracked depths of the truck for the depth profile shown in (a) with two, four, and six 3D integral imaging reconstructions per frame (rpf) respectively. Fps: frames per second.

Figure 7(c)-(e) shows that two 3D integral imaging reconstructions are not enough to provide accurate depth tracking at high axial speeds. Although four 3D reconstructions can track depth with reasonable accuracy for relatively fast speeds, even four reconstructions are insufficient for extremely fast-moving objects.

One of the main applications of integral imaging is to mitigate the effects of degraded environments like partial occlusion. We test our system on two different partial occlusions. Figures 8(a) and (d) show a sample experimental image scene with two different occlusions. Figures 8(b) and (e) show the 3D integral imaging reconstructions of scenes in Fig. 8(a) and (d) respectively at the plane of the occlusion. Figures 8(c) and (f) show the 3D integral imaging reconstructions of scenes in Fig. 8(a) and (d) respectively at the plane of the truck. Figures 9(a) and (b) show sample mutual information vs reconstruction depth curves corresponding to the two occlusions shown in Fig. 8(a) and (d).

Figure 10 shows the depth tracking results for the partially occluded truck corresponding to the scenes in Fig. 8(a) and (d). We use two 3D reconstructions per frame (rpf) for the scenes in Fig. 8(a). However, this is not sufficient for severe occlusion as is the case in Fig. 8(d), and hence we use three 3D reconstructions per frame. More reconstructions per frame are required for depth tracking as the mutual information peak gets less pronounced.

We also test our system for a scene with partial occlusion and low illumination conditions. Figure 11(a) shows a sample experimental image of the scene with low illumination conditions of approximately 6 photons per pixel and partial occlusion. The scenes are captured using a

Optics EXPRESS

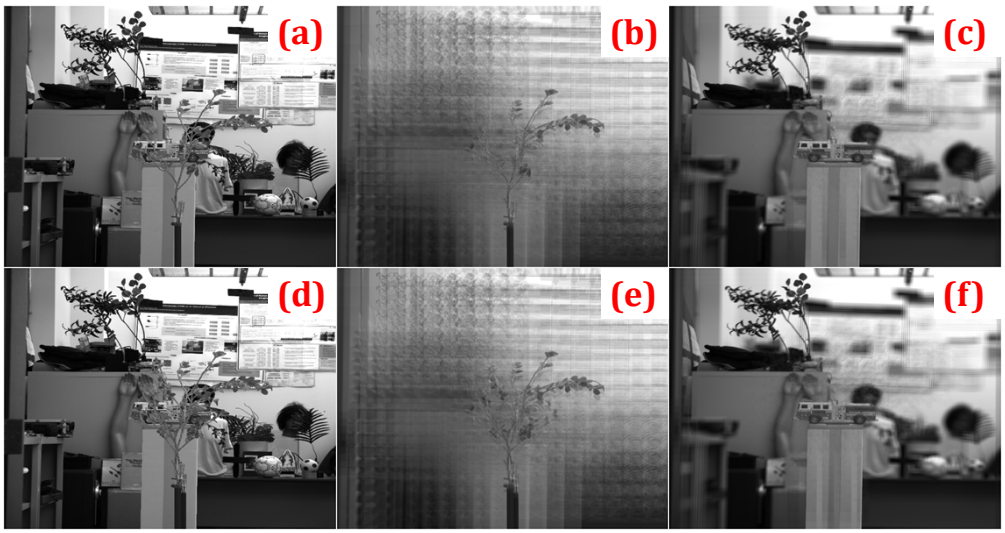

Fig. 8. (a) Sample experimental scene with partial occlusion. (b) 3D integral imaging reconstruction of the scene in (a) at the depth of the occlusion. (c) 3D integral imaging reconstruction of the scene in (a) at the depth of the truck. (d) Sample experimental scene with a more severe occlusion. (e) – (f) 3D integral imaging reconstruction of the scene in (d) at the depths of the occlusion and truck respectively.

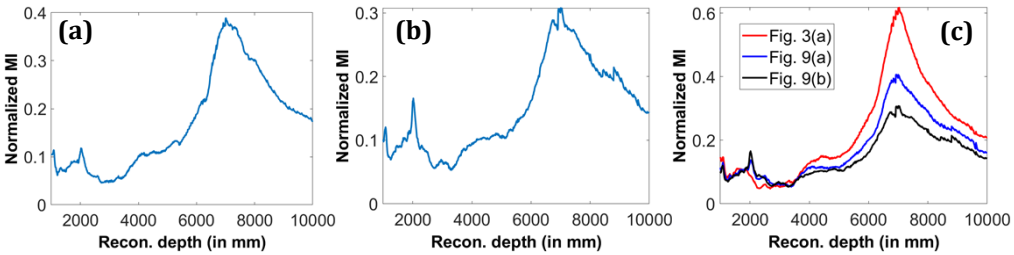

Fig. 9. (a) Mutual information (MI) vs. reconstruction depth (recon. depth) for the truck corresponding to the scene in Fig. 8(a). (b) Mutual information (MI) vs. reconstruction depth for the truck corresponding to the scene in Fig. 8(d). The small secondary peak in (b) at 2020 mm corresponds to the location of the partial occlusion. The same peak is also present in (a) but is not as pronounced. (c) Comparison of mutual information curves in (a) and (b) with that of the mutual information curve of Fig. 3(a) corresponding to the clear scene shown in Fig. 2(a).

low-light camera. For reference, the same scene in high illumination is shown in Fig. 11(b). The partial occlusion used in this scene is the same as that in Fig. 8(a). Figure 11(c) shows the mutual information curve for the truck in a low-illumination and partially occluded scene (Fig. 11(a)).

Figure 12 shows the depth tracking results for the truck corresponding to the scenes in Fig. 11(a). We use three 3D reconstructions per frame (rpf) and 30 frames per second (fps) for tracking.

3D Integral imaging improves the detector's performance in adverse environmental conditions like low illumination and partial occlusion. Detector performance is the most significant factor in the overall tracking accuracy. We use a simple motion profile of the truck to evaluate the detector score with and without integral imaging. We compute the detector score on each frame for the 2D scene as well as the 3D integral imaging reconstructed scene as the truck moves from 400 cm to 800 cm axially at approx. 1.8 meters per second. For each frame, the 3D integral

Optics EXPRESS

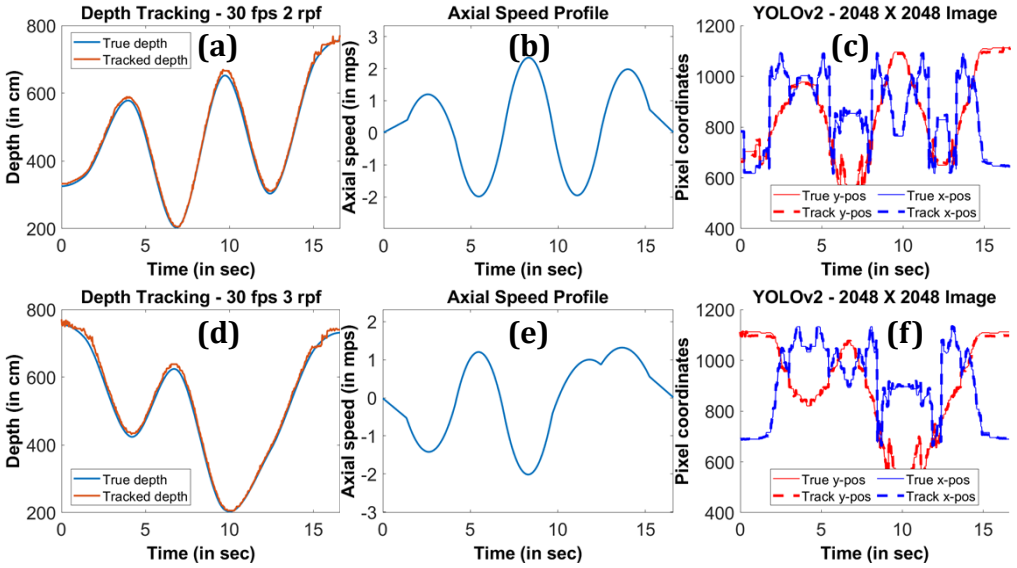

Fig. 10. Experimental depth tracking results for scenes shown in Fig. 8 with partially occluded truck. The motion of the truck is captured at 30 frames per second (fps). (a) True and tracked depths of the truck in the scene shown in Fig. 8(a) for a sample motion profile. The corresponding depth error has a mean of 9.01 cm and standard deviation of 5.09 cm. Two 3D reconstructions per frame (rpf) are used for depth tracking. (b) The axial speed of the truck corresponding to the depth profile shown in (a). We represent an object moving away from the imaging system with a positive speed and moving toward the system with a negative speed. (c) True and tracked lateral positions (*x-pos*, *y-pos*) of the truck for the same motion profile. (d), (e), and (f) show the same results as (a), (b), and (c) respectively for the scene with a partially occluded truck as shown in Fig. 8(d). The depth error for curve in (d) has a mean of 10.30 cm and standard deviation of 5.74 cm. Three 3D reconstructions per frame (rpf) are used for depth tracking as two reconstructions are not sufficient for the severe occlusion present in the scenes. Mps: meters per second.

imaging scene is reconstructed at the tracked depth. Table 1 summarizes the scores for two different partial occlusions (Fig. 8(a) and Fig. 8(d)) and a partial occlusion in low illumination (Fig. 11(a)). Table 2 presents the percent of frames with failed detections – where a failed detection is characterized by a detection score of less than 0.5.

Table 1. Average detector scores for 2D imaging vs. 3D integral imaging in degraded environments[a]

Degradations	Average Detector Score – 2D	Average Detector Score – 3D
Partial occlusion – 1 (see Fig. 8(a))	0.8446	0.9424
Partial occlusion – 2 (see Fig. 8(d))	0.7518	0.9114
Partial occlusion and low illumination (see Fig. 11(a))	0.8265	0.9263

[a]Scene shown in Fig. 11(a) uses similar partial occlusion as that in Fig. 8(a).

Visualization 1 shows a sample video of the experimental results comparing 2D tracking and 3D tracking using the proposed approach. In this video, a green-colored box represents a valid detection using YOLOv2 deep neural network (detection score more than 0.5) and a red-colored box represents the true 2D bounding box of the object corresponding to a failed detection. As

Optics EXPRESS

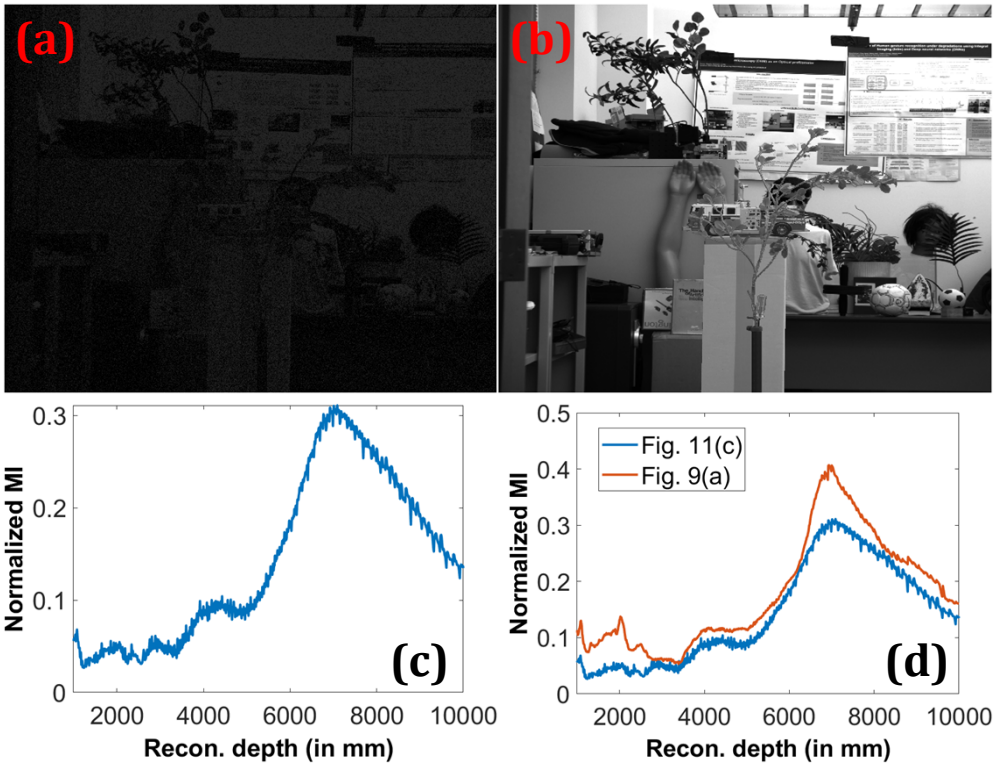

Fig. 11. (a) Sample experimental scene with low-illumination noisy conditions of approximately 6 photons per pixel and partial occlusion. The scenes are captured using a low-light camera. (b) The same scene as in (a) with high illumination. (c) Mutual information (MI) curve as a function of reconstruction depth (Recon. depth) for the truck in scene (a). (d) Comparison of the mutual information curve in (c) for low illumination and partial occlusion with that of the high-illumination scene with partial occlusion as shown in Fig. 8(a). (Visualization 1)

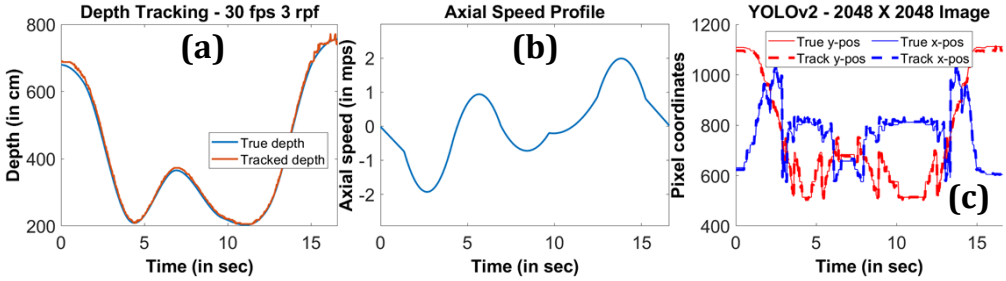

Fig. 12. Experimental depth tracking results for scenes with partial occlusion and low illumination conditions (see Fig. 11(a)). The motion of the truck is captured at 30 frames per second (fps). (a) True and tracked depths of the truck in scenes shown in Fig. 11(a) for a sample motion profile. The corresponding depth error has a mean of 7.62 cm and standard deviation of 5.29 cm. Three 3D integral imaging reconstructions per frame (rpf) are used for depth tracking. (b) The axial speed of the truck corresponding to the depth profile shown in (a). We represent an object moving away from the integral imaging system with a positive speed and moving towards the imaging system with a negative speed. (c) True and tracked lateral positions (*x-pos*, *y-pos*) of the truck for the same motion profile.

Optics EXPRESS

we can see, 2D imaging fails sporadically in tracking the object in degraded environments. In comparison, 3D integral imaging-based tracking performs much better in similar circumstances.

Table 2. Percent of failed detections for 2D imaging vs. 3D integral imaging in degraded environments[a]

Degradations	Percent of Failed Detections – 2D	Percent of Failed Detections – 3D
Partial occlusion – 1 (see Fig. 8(a))	4.54	0
Partial occlusion – 2 (see Fig. 8(d))	10.60	0
Partial occlusion and low illumination (see Fig. 11(a))	6.06	0

[a]Scene shown in Fig. 11(a) uses similar partial occlusion as that in Fig. 8(a).

5. Conclusions

We have considered 3D object tracking with integral imaging using mutual information and Bayesian optimization. Integral imaging has several advantages over conventional 2D imaging for object detection in adverse environmental conditions such as low light and partial occlusion. Additionally, its depth sectioning ability also aids in object classification in a multi-object scenario in the presence of 3D background noise. We postulate that object tracking could benefit from using 3D integral imaging instead of conventional 2D imaging in two main aspects – improvement of detector performance in degraded environments and improvement of object association due to integral imaging's depth sectioning ability. The use of integral imaging requires depth tracking in addition to conventional 2D object tracking. A recent study estimates an object's depth by computing mutual information between the object's bounding box in the 3D reconstructed scene and the 2D central image [30]. We have improved upon this method by using Bayesian optimization to continuously track the object's depth. We evaluated our proposed method on laboratory scenes with an object free to move in all three dimensions. Our preliminary results show that 3D integral imaging object tracking outperforms 2D object tracking in degraded environments, and as few as two 3D reconstructions per image frame may be sufficient to track an object's depth. For faster-moving objects, more 3D reconstructions per frame are required. Two 3D reconstructions are the theoretical minimum number of 3D reconstructions required for depth estimation with our approach as one 3D reconstruction leads to a correspondence problem.

This manuscript provided a proof-of-concept for using integral imaging in 3D object-tracking applications. However, a rigorous study of its performance or various applications was not considered here as it is outside the scope of this work. In the future, rigorous benchmarking on multi-object tracking datasets is needed for this approach. However, most of the standard datasets available use traditional 2D imaging techniques to capture image frames. Thus, new datasets also need to be collected using 3D integral imaging. Additionally, this approach also needs to be tested in other degraded environments like under-water and other sources of noise like brownout conditions. We also plan to study the effects of various integral imaging system parameters such as the number of cameras, pitch size, and sensor size [89] on 3D object tracking. We postulate that our approach can be used in many real-world tracking scenarios currently handled by 2D imaging [90]. A few examples of such applications are vehicle tracking for autonomous driving, pedestrian tracking on streets, microscopy [25,91,92], and gesture recognition and tracking [93].

Funding. National Science Foundation (2141473); Air Force Office of Scientific Research (FA9550-21-1-0333); Office of Naval Research (N000142212349, N000142212375).

Disclosures. The authors declare no conflict of interest.

Data availability. Data underlying the results are not publicly available at this time but may be obtained from the authors upon reasonable request.

Optics EXPRESS

References

1. A. Bewley, Z. Ge, L. Ott, *et al.*, "Simple online and realtime tracking," arXiv, arXiv:1602.00763v2 (2017).
2. C. Dicle, M. Sznaier, and O. Camps, "The way they move: tracking multiple objects with similar appearance," in *Int. Conf. on Comp. Vis.* (2013).
3. J. H. Yoon, M. H. Yang, J. Lim, *et al.*, "Bayesian multi-object tracking using motion context from multiple objects," in *Winter Conf. on Appl. of Comp. Vis.* (2015).
4. A. Bewley, L. Ott, F. Ramos, *et al.*, "ALEx-TRACK: affinity learning by exploring temporal reinforcement with association chains," in *Int. Conf. on Robotics and Automation* (2016).
5. C. Kim, F. Li, A. Ciptadi, *et al.*, "Multiple hypothesis tracking revisited," in *Int. Conf. on Comp. Vis.* (2015).
6. L. Leal-Taixe, A. Milan, I. Reid, *et al.*, "MOTChallenge 2015: towards a benchmark for multi-target tracking," arXiv, arXiv: 1504.01942 (2015).
7. S. Ren, K. He, R. Girshick, *et al.*, "Faster R-CNN: towards real-time object detection with region proposal networks," in *Adv. in Neural Inf. Processing Systems* (2015).
8. R. Kalman, "A new approach to linear filtering and prediction problems," J. Basic Eng. **82**(1), 35–45 (1960).
9. H. W. Kuhn, "The Hungarian method for the assignment problem," Naval Research Logistic Quarterly **2**(1-2), 83–97 (1955).
10. N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," arXiv, arXiv: 1703.07402 (2017).
11. Y. Zhang, C. Wang, X. Wang, *et al.*, "FairMOT: on the fairness of detection of re-identification in multiple object tracking," arXiv, arXiv:2004.01888 (2021).
12. P. Chu, J. Wang, Q. You, *et al.*, "TransMOT: spatial-temporal graph transformer for multiple object tracking," arXiv, arXiv: 2104.00194v2 (2021).
13. Y. Zhang, P. Sun, Y. Jiang, *et al.*, "ByteTrack: multi-object tracking by associating every detection box," arXiv, arXiv: 2110.06864v3 (2022).
14. Z. Ge, S. Liu, F. Wang, *et al.*, "YOLOX: exceeding YOLO series in 2021," arXiv, arXiv: 2107.08430v2 (2021).
15. M. M. Corral and B. Javidi, "Fundamentals of 3D imaging and displays: a tutorial on integral imaging, light-field, and plenoptic systems," Adv. Opt. Photonics **10**(3), 512–566 (2018).
16. B. Javidi, A. Carnicer, J. Arai, *et al.*, "Roadmap on 3D integral imaging: sensing, processing, and display," Opt. Express **28**(22), 32266–32293 (2020).
17. G. Lippmann, "Epreuves reversibles donnant la sensation du relief," J. Phys. **7**, 821–825 (1908).
18. N. Davies, M. McCormick, and L. Yang, "Three-dimensional imaging systems: a new development," Appl. Opt. **27**(21), 4520–4528 (1988).
19. H. Arimoto and B. Javidi, "Integral Three-dimensional Imaging with digital reconstruction," Opt. Lett. **26**(3), 157–159 (2001).
20. F. Okano, H. Hoshino, J. Arai, *et al.*, "Real-time pickup method for a three-dimensional image based on integral photography," Appl. Opt. **36**(7), 1598–1603 (1997).
21. M. Martinez-Corral, A. Dorado, J. C. Barreiro, *et al.*, "Recent advances in the capture and display of macroscopic and microscopic 3D scenes by integral imaging," Proc. IEEE **105**(5), 825–836 (2017).
22. A. Stern and B. Javidi, "Three-dimensional image sensing and reconstruction with time-division multiplexed computational integral imaging," Appl. Opt. **42**(35), 7036–7042 (2003).
23. E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision," Computational Models of Visual ProcessingM. Landy and J.A. Movshon, eds. The MIT Press 1, 3–20 (1991).
24. J. Liu, D. Claus, T. Xu, *et al.*, "Light field endoscopy and its parametric description," Opt. Lett. **42**(9), 1804–1807 (2017).
25. G. Scrofani, J. Sola-Pikabea, A. Llavador, *et al.*, "FIMic: design for ultimate 3D-integral microscopy of in-vivo biological samples," Biomed. Opt. Express **9**(1), 335–346 (2018).
26. J. Arai, E. Nakasu, T. Yamashita, *et al.*, "Progress overview of capturing method for integral 3-D imaging displays," Proc. IEEE **105**(5), 837–849 (2017).
27. M. Yamaguchi, "Full-parallax holographic light-field 3-D displays and interactive 3-D touch," Proc. IEEE **105**(5), 947–959 (2017).
28. P. Wani, K. Usmani, G. Krishnan, *et al.*, "Lowlight object recognition by deep learning with passive three-dimensional integral imaging in visible and longwave infrared wavelengths," Opt. Express **30**(2), 1205–1218 (2022).
29. K. Usmani, T. O'Connor, P. Wani, *et al.*, "3D object detection through fog and occlusion: passive integral imaging vs active (LiDAR) sensing," Opt. Express **31**(1), 479–491 (2023).
30. P. Wani, G. Krishnan, T. O. Connor, *et al.*, "Information-theoretic performance evaluation of 3D integral imaging," Opt. Express **30**(24), 43157–43171 (2022).
31. P. Wani and B. Javidi, "3D integral imaging depth estimation of partially occluded objects using mutual information and Bayesian optimization," Opt. Express **31**(14), 22863–22884 (2023).
32. B. Shahriari, K. Swersky, Z. Wang, *et al.*, "Taking the human out of the loop: A review of Bayesian optimization," Proc. IEEE **104**(1), 148–175 (2016).
33. J. Wu, M. Poloczek, A. G. Wilson, *et al.*, "Bayesian optimization with gradients," arXiv, arXiv: 1703.04389v3 (2017).

Optics EXPRESS

34. J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," Proc. NIPS 12, 2951–2959 (2012).
35. A. Marco, F. Berkenkamp, P. Hennig, *et al.*, "Virtual vs. real: Trading off simulations and physical experiments in reinforcement learning with Bayesian optimization," Proc. ICRA **1**, 1557–1563 (2017).
36. L. Acerbi and W. J. Ma, "Practical Bayesian optimization for model fitting with Bayesian adaptive direct search," Proc. NIPS 17, 1834–1844 (2017).
37. J. Bergstra, R. Bardenet, Y. Bengio, *et al.*, "Algorithms for hyper-parameter optimization," Proc. NIPS 11, 2546–2554 (2011).
38. K. Swersky, J. Snoek, and R. P. Adams, "Multi-task Bayesian optimization," Proc. NIPS 13, 2004–2012 (2013).
39. C. Thornton, F. Hutter, H. H. Hoos, *et al.*, "Auto-WEKA: combined selection and hyperparameter optimization of classification algorithms," Proc. KDD' **13**, 847–855 (2013).
40. E. Brochu, V. M. Cora, and N. de Freitas, "A tutorial on Bayesian optimization of expensive cost functions, with applications to active user modeling and hierarchical reinforcement learning," arXiv, arXiv: 1012.2599v1 (2010).
41. H. Chen, H. C. Wu, S. C. Chan, *et al.*, "A stochastic quasi-newton method for large-scale nonconvex optimization with applications," IEEE Trans. Neural Netw. Learning Syst. **31**(11), 4776–4790 (2020).
42. A. S. Lewis and M. L. Overton, "Nonsmooth optimization via quasi-newton methods," Math. Program. **141**(1-2), 135–163 (2013).
43. K. Butt, R. A. Rahman, N. Sepehri, *et al.*, "Globalized and bounded Nelder-Mead algorithm with deterministic restarts for tuning controller parameters: Method and application," Optim Control Appl Methods **38**(6), 1042–1055 (2017).
44. C. Stoean, M. Preuss, R. Stoean, *et al.*, "Multimodal optimization by means of a topological species conversion algorithm," IEEE Trans. Evol. Computat. **14**(6), 842–864 (2010).
45. J. P. Li, "Truss topology optimization using an improved species-conversion genetic algorithm," Engineering Optimization **47**(1), 107–128 (2015).
46. Y. Liang and K. S. Leung, "Genetic algorithm with adaptive elitist-population strategies for multimodal function optimization," Applied Soft Computing **11**(2), 2017–2034 (2011).
47. L. De Castro and F. J. V. Zuben, "The clonal selection algorithm with engineering application," Proc. GECCO 2000, 36–39 (2001).
48. L. N. De Castro and J. Timmis, "An artificial immune network for multimodal function optimization," Proc. CEC' **02**, 699–704 (2002).
49. L. N. De Castro and F. J. V. Zuben, "aiNet: an artificial immune network for data analysis," *Data Mining: A Heuristic Approach* (IGI Global, 2002), pp. 231–260.
50. A. Markman and B. Javidi, "Learning in the dark: 3D integral imaging object recognition in very low illumination conditions using convolutional neural networks," OSA Conti. **1**(2), 373–383 (2018).
51. D. Aloni, A. Stern, and B. Javidi, "Three-dimensional photon counting integral imaging reconstruction using penalized maximum likelihood expectation maximization," Opt. Express **19**(20), 19681–19687 (2011).
52. X. Shen, A. Carnicer, and B. Javidi, "Three-dimensional polarimetric integral imaging under low illumination conditions," Opt. Lett. **44**(13), 3230–3233 (2019).
53. B. Tavakoli, B. Javidi, and E. Watson, "Three dimensional visualization by photon counting computational integral imaging," Opt. Express **16**(7), 4426–4436 (2008).
54. J. S. Jang and B. Javidi, "Three-dimensional synthetic aperture integral imaging," Opt. Lett. **27**(13), 1144–1146 (2002).
55. M. Daneshpanah and B. Javidi, "Profilometry and optical slicing by passive three-dimensional imaging," Opt. Letters **34**(7), 1105–1107 (2009).
56. T. M. Cover and J. A. Thomas, *Elements of information theory*, (John, Wiley & Sons, 1991).
57. D. B. Russakoff, C. Tomasi, T. Rohlfing, *et al.*, "Image similarity using mutual information of regions," in *European Conf. on Comp. Vis. (ECCV)*(2004), pp. 596–607.
58. J. P. Pluim, J. B. Maintz, and M. A. Viergever, "Image registration by maximization of combined mutual information and gradient information," IEEE Trans. Med. Imaging **19**(8), 809–814 (2000).
59. D. Rueckert, M. J. Clarkson, D. L. G. Hill, *et al.*, "Non-rigid registration using higher-order mutual information," Proc. SPIE **3979**, 438–447 (2000).
60. E. Volden, G. Giraudon, and M. Berthod, "Information in Markov random fields and image redundancy," *Selected papers from the 4th Canadian workshop on information theory and applications II*, 250–268 (1996).
61. S. R. Narravula, M. M. Hayat, and B. Javidi, "Information theoretic approach for accessing image fidelity in photon-counting arrays," Opt. Express **18**(3), 2449–2466 (2010).
62. H. J. Kushner, "A new method of locating the maximum point of an arbitrary multipeak curve is the presence of noise," J. Basic Eng. **86**(1), 97–106 (1964).
63. A. G. Zhilinskas, "Single-step Bayesian search method for an extremum of functions of a single variable," Cybernetics (Engl. Transl.) **11**(1), 160–166 (1976).
64. J. Mockus, V. Tiesis, and A. Zilinskas, "The application of Bayesian methods for seeking the extremum," *Towards Global Optimization 2* (Springer, 1979), pp. 117.
65. J. Mockus, "On Bayesian methods for seeking the extremum," in *Optimization Techniques IFIP Technical Conference* (1974), pp. 400–404.

Optics EXPRESS

66. J. Mockus, *Bayesian Approach to Global Optimization: Theory and Applications*, (Springer, 1989).
67. D. Huang, T. T. Allen, W. I. Notz, *et al.*, "Sequential kriging optimization using multiple-fidelity evaluations," Struct Multidisc Optim **32**(5), 369–382 (2006).
68. A. Sobester, S. J. Leary, and A. J. Keane, "A parallel updating scheme for approximating and optimizing high fidelity computer simulations," Struct Multidisc Optim **27**(5), 371–383 (2004).
69. A. J. Keane, "Statistical improvement criteria for use in multiobjective design optimization," AIAA J. **44**(4), 879–891 (2006).
70. J. Knowles, "ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems," IEEE Trans. Evol. Computat. **10**(1), 50–66 (2006).
71. J. B. Mockus and L. J. Mockus, "Bayesian approach to global optimization and application to multiobjective and constrained problems," J Optim Theory Appl **70**(1), 157–172 (1991).
72. P. I. Frazier, "A tutorial on Bayesian optimization," arXiv, arXiv: 1807.02811v1 (2018).
73. D. J. Lizotte, "Practical Bayesian optimization," Ph.D. dissertation (University of Alberta, 2008).
74. C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, (MIT Press, 2005).
75. S. A. Renganathan, J. Larson, and S. M. Wild, "Lookahead acquisition functions for finite-horizon time-dependent Bayesian optimization and applications to quantum optimal control," arXiv, arXiv: 2105.09824v1 (2021).
76. I. Bogunovic, J. Scarlett, and V. Cevher, "Time-varying Gaussian process bandit optimization," arXiv, arXiv: 1601.06650v1 (2016).
77. Z. Deng, I. Tutunnikov, I. S. Averbukh, *et al.*, "Bayesian optimization for inverse problems in time-dependent quantum dynamics," J. Chem. Phy. **153**(16), 164111 (2020).
78. F. M. Nyikosa, M. A. Osborne, and S. J. Roberts, "Bayesian optimization for dynamic problems," arXiv, arXiv: 1803.03432v1 (2018).
79. C. K. I. Williams, "Prediction with Gaussian processes: From linear regression to linear prediction and beyond," *Learning in Graphical Models* (Springer, 1998), pp. 599–621.
80. C. E. Rasmussen, "Gaussian processes in machine learning," *Advanced Lectures in Machine Learning: ML Summer School 2003* (Springer, 2003), pp. 63–71.
81. J. S. Taylor, *Kernel Methods for Pattern Analysis*, (Cambridge, 2004).
82. T. L. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," Advances in Applied Mathematics **6**(1), 4–22 (1985).
83. P. Hennig and C. J. Schuler, "Entropy search for information-efficient global optimization," Journal of Machine Learning Research **13**, 1809–1837 (2012).
84. J. M. Hernandez-Lobato, M. W. Hoffman, and Z. Ghahramani, "Predictive entropy search for efficient global optimization of black-box functions," Proc. NIPS 14, 918–926 (2014).
85. Z. Wang and S. Jegelka, "Max-value entropy search for efficient global optimization," Proc. ICML 17, 3627–3635 (2017).
86. J. Redmon, S. Divvala, R. Girshick, *et al.*, "You look only once: unified, real-time object detection," *Proc. IEEE Conf. Comput. Vis. Pattern Recongit. (CVPR)*, 779–788 (2016).
87. J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," *Proc. IEEE Conf. Comput. Vis. Pattern Recongit. (CVPR)*, 6517–6525 (2017).
88. C. Szededy, W. Liu, Y. Jia, *et al.*, "Going deeper with convolutions," *Proc. IEEE Conf. Comput. Vis. Pattern Recongit. (CVPR)*, 1–9 (2015).
89. F. Jin, J. Jang, and B. Javidi, "Effects of device resolution on three-dimensional integral imaging," Opt. Lett. **29**(12), 1345–1347 (2004).
90. Daniel LeMaster, Barry Karch, and Bahram Javidi, "Mid-Wave Infrared 3D Integral Imaging at Long Range," J. Display Technol. **9**(7), 545–551 (2013).
91. J. S. Jang and B. Javidi, "Three-dimensional Integral Imaging of Micro-objects," Opt. Lett. **29**(11), 1230–1232 (2004).
92. M. Martinez-Corral, G. Saavedra, and B. Javidi, "Resolution improvements in integral microscopy with Fourier plane recording," Opt. Express **24**(18), 20792–20798 (2016).
93. Q. Lu, Y. Bar-Shalom, P. Willett, *et al.*,, "Measurement extraction for two closely-spaced objects using an imaging sensor," IEEE Transactions on Aerospace and Electronic Systems **55** 2965–2977 (2019).

68.114.58.156 {ts '2024-05-15 12:04:40'}