



---

Títol article:	Three Dimensional Visualization of Long Range Scenes by Photon Counting Mid-Wave Infrared Integral Imaging
Autors:	Pedro Latorre-Carmona, Bahram Javidi, and Daniel LeMaster
Revista:	IEEE JOURNAL OF DISPLAY TECHNOLOGY, 2015
Versió:	PDF Versió post-print
Cita bibliogràfica (ISO 690):	LATORRE CARMONA, Pedro; JAVIDI, Bahram; LEMASTER, Daniel. Three Dimensional Visualization of Long Range Scenes by Photon Counting Mid-Wave Infrared Integral Imaging. <i>Display Technology, Journal o</i> , 2015, 1-5.
url Repositori UJI:	<a href="http://hdl.handle.net/10234/126126">http://hdl.handle.net/10234/126126</a>

---

# Three Dimensional Visualization of Long Range Scenes by Photon Counting Mid-Wave Infrared Integral Imaging

Pedro Latorre-Carmona, Bahram Javidi, and Daniel LeMaster

**Abstract**—Integral Imaging under photon counting conditions has found different three-dimensional (3D) imaging applications, including 3D image reconstruction and recognition. In this letter, we present the application of the maximum likelihood (ML) estimation method for visualization of 3D scenes in photon starved environments using Mid-Wave Infrared 3D data of real scenes acquired at distances ranging from 50m to more than 2km. To the best of our knowledge, this is the first report on Mid-Wave Infrared 3D photon counting integral imaging of distant scenes.

**Index Terms**—Computational integral imaging (CII), infrared imaging, passive 3-D imaging, photon counting.

## I. INTRODUCTION

INTEGRAL Imaging is a three-dimensional sensing and imaging technique with different applications including TV broadcasting, 3D display, and 3D visualization in different types of media [1]–[6]. Photon counting integral imaging has been proposed to perform 3D image reconstruction [7] and 3D target recognition [8]–[10] in photon starved conditions. In this paper we apply the maximum likelihood estimation (MLE) method for 3D visualization purposes on Mid-Wave Infrared (MWIR) images where photon starved conditions have been simulated by applying a photon counting process on the corresponding elemental images. Reconstruction performance is given in qualitative and quantitative terms with the Peak Signal to Noise Ratio (PSNR). Results show that the ML method provides good visualization capabilities.

## II. THREE-DIMENSIONAL IMAGING AND COMPUTATIONAL RECONSTRUCTION

In Synthetic Aperture Integral Imaging (SAII) [11] (Fig. 1(a)), a single camera is used to acquire the elemental images from slightly different perspectives. This camera may move into positions defined by a grid. Figure 1(b) shows the computational reconstruction methodology. Each elemental image is projected on the desired reconstruction plane and overlaps with

Pedro Latorre-Carmona is with the Institute of New Imaging Technologies, Universitat Jaume I, Campus del Riu Sec sn, 12071 Castellón de la Plana, Spain (e-mail: latorre@uji.es)

Bahram Javidi is with the Department of Electrical and Computer Engineering at University of Connecticut, Storrs, CT 06269-4157 USA (e-mail: Bahram.Javidi@UConn.edu).

Daniel LeMaster is with the Air Force Research Laboratory, AFRL/RyMT, 2241 Avionics Circle, Bldg 620, Wright Patterson Air Force Base, OH 45433 USA (e-mail: daniel.lemaster@wpafb.af.mil).

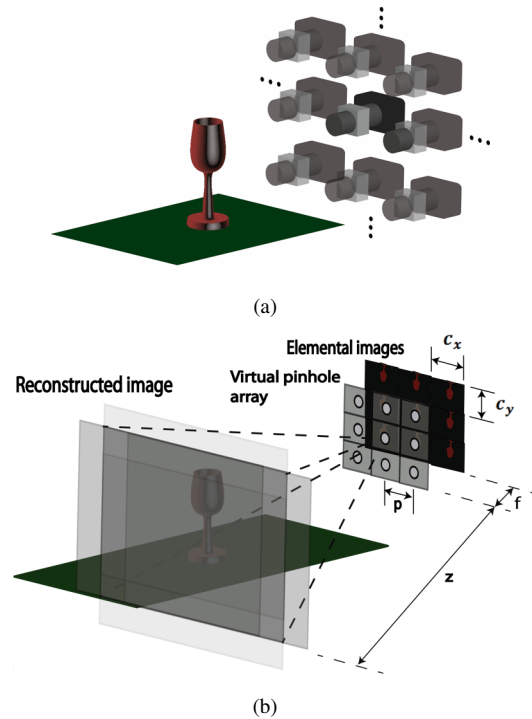


Fig. 1. Synthetic aperture Integral Imaging acquisition and processing. (a) The camera position changes in a grid to acquire images of a scene under different perspectives. (b) The superposition of the ray cones projected from the elemental images reconstructs the 3D scene.

all other back-projected elemental images, as follows [12]:

$$I(x, y, z) = \frac{1}{O(x, y)} \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} E_{kl} \left( x - k \frac{N_x \cdot s}{c_x \cdot M}, y - l \frac{N_y \cdot s}{c_y \cdot M} \right), \quad (1)$$

where  $I(x, y, z)$  represents the intensity of the reconstructed 3D image at depth  $z$ ,  $x$  and  $y$  are the indexes of the pixel,  $E_{kl}$  represents the intensity of the  $k$ th row and  $l$ th column elemental image,  $N_x \times N_y$  is the total number of pixels for each elemental image,  $M = \frac{z}{f}$  is the magnification factor,  $c_x \times c_y$  is the physical size of the camera sensor,  $s$  is the pitch of the camera and  $O(x, y)$  is the overlapping number matrix.

## III. PHOTON-COUNTING MODEL

The integrated irradiance arriving at a camera is proportional to the mean number of photons incident onto the sensor ([7], [13]). On the other hand, it can be shown that the probability

of the number of photons detected in a time interval follows a Poisson density function [13]. If we consider  $I_x$  as the normalized irradiance at pixel  $x$  such that  $\sum_{x=1}^{N_T} I_x = 1$ , where  $N_T$  is the total number of pixels of an image, a Poisson random number with mean parameter  $N_p I_x$  can be generated to simulate an image that has  $N_p$  number of photons in average. In particular, the Poisson distribution can be described as:

$$Pr(C_x|I_x) = \frac{(I_x)^{C_x} \cdot e^{-I_x}}{C_x!}, \quad C_x = 0, 1, 2, \dots \quad (2)$$

where  $C_x$  refers to  $C$  photons at pixel  $x$ .

#### IV. MAXIMUM LIKELIHOOD ESTIMATION FOR VISUALIZATION

As discussed in Section II, the image of an object pixel at  $p \equiv (x, y)$  appears periodically on the elemental images in the following positions:

$$\{(p + \Delta p_{kl})\} \equiv \left\{ \left( x - k \frac{N_x \cdot s}{c_x \cdot M}, y - l \frac{N_y \cdot s}{c_y \cdot M} \right) \right\} \quad (3)$$

It was shown in [7] that the estimation of the irradiance of a reconstructed scene could be obtained from the photon counting elemental images taking into account that the likelihood estimation for the hypothesis of the irradiance inference can be expressed as:

$$\mathcal{L}(I_p^{z_0} | C_{kl}(p + \Delta p_{kl})) = \prod_{k=0}^{K-1} \prod_{l=0}^{L-1} Pr(C_{kl}(p + \Delta p_{kl}) | I_p^{z_0}) \quad (4)$$

Considering the logarithmic expression of Eq. 4 and  $C_{kl} | I_p^{z_0} \sim Poisson(N_p I_p^{z_0})$ , one arrives at:

$$l(I_p^{z_0}) = \sum_k \sum_l [-N_p I_p^{z_0} + C_{kl}(p + \Delta p_{kl}) \cdot \log(N_p I_p^{z_0}) - \log(C_{kl}(p + \Delta p_{kl}))] \quad (5)$$

Therefore, maximizing with respect to  $I_p^{z_0}$  gives:

$$\frac{\partial l(I_p^{z_0})}{\partial I_p^{z_0}} = 0 \rightarrow \tilde{I}_p^{z_0} = \frac{1}{N_p D} \sum_k \sum_l C_{kl}(p + \Delta p_{kl}) \quad (6)$$

where  $\tilde{I}_p^{z_0} \equiv MLE\{I_p^{z_0}\}$  and  $D \equiv \frac{KL}{\log_2 e}$ . Eq. 6 implies that the result of the computational reconstruction for one plane of the scene located at a specific distance using photon-counted elemental images is also the ML irradiance estimate of the scene at that plane. Therefore, the irradiance of the three dimensional objects can be reconstructed by use of the photon counting elemental images. Noise sources may appear as well but if we assume they are statistically independent from the Poisson process that generates the photon counting images (for instance, if noise  $\neq f(I_p^{z_0})$ ) then the probability term in Eq. 4 would become a product of probabilities. In that case, the partial derivative in Eq. 6 would cancel out the terms without dependence on  $I_p^{z_0}$  and therefore the ML estimate would remain the same. This is the assumption we will consider in our case.

#### V. WAVELET SHRINKAGE FOR IMAGE DENOISING

Fourier analysis has been successfully applied to periodic, time-invariant or stationary processes. On the other hand, wavelet analysis was developed in the framework of local, transient, time-varying or non-stationary signals. Wavelets are generated from one single function (basis function) by dilations (scaling) and translations (shifts) in time (frequency) domain. Discrete wavelet transforms (DWTs) are typically associated to Multi-Resolution Analysis (MRA) strategies consisting of the application of low-pass and high-pass filters to give low level (coarse) and high level (detailed) information of the input signal. Multi-Resolution Analysis (MRA) involves the application of two sets of functions, called scaling and wavelet functions. Real DWTs have a series of known drawbacks: (a) they are very sensitive to shifts; their coefficients *appear* only at three spatial orientations (vertical, horizontal and diagonal); and (c) phase information is not considered for processing. The Complex Dual-Tree Discrete Wavelet Transform appeared to try to solve these drawbacks, i. e. (a) it is approximately shift invariant; (b) it has good directional selectivity; (c) phase is taken into account. Consider a complex scaling function and a complex wavelet. The Dual-Tree Complex Wavelet Transform (DT-CWT) [14] uses two Discrete Wavelet Transforms in parallel. The first (upper) tree gives the real part while the second (lower) gives the imaginary part of the Complex Wavelet Transform (CWT). The two trees are in fact real and use two different sets of perfect reconstruction filters. In the particular implementation we used in this paper, we considered Farras filters [15] for the first decomposition stage, and Kingsburry's Q-shift filters [16] for the remaining stages. Farras filters were designed to have two particular important properties for the complex wavelet transforms, i. e., (1) be orthogonal, and (2) have a subset of exactly symmetric coefficients (among all). The Kingsburry's Q-shift filters were designed to eliminate some problems associated to the parallel tree transformation approach [16]. Once this transform has been applied, all high frequency (MRA) coefficients that appear during this decomposition and that are lower than a particular threshold, can be converted to zero. The denoised image is obtained applying the inverse wavelet transform after this thresholding. The reader is referred to [15]–[17] for further details.

#### VI. EXPERIMENTAL RESULTS

The elemental images used in this paper were acquired from the 12th floor of the AFRL tower located at Wright Patterson Air Force Base (AFB). They were collected with a Lockheed Martin Santa Barbara Focal plane AuraSR MWIR imager that could move in one fixed direction using a high accuracy rail apparatus. The AuraSR MWIR imager had a StingRay Optics 120mm f/2.3 lens. The pixel size of this camera is 19.5  $\mu$ m. The size of the elemental images is 1024  $\times$  1024 pixels. Two types of scenes were acquired, one of them consisting of a road, a group of trees and a vehicle just behind one of them, and the other scene consisted of a view of the airfield of Wright AFB. For the case of the car and trees scene, 8 elemental images corresponding to an acquisition in a 1  $\times$  8

horizontal grid were considered. For the airfield scene, 10 images in a  $1 \times 10$  horizontal grid were acquired (the reader is referred to [18] for further details). In order to simulate photon starved conditions during acquisition, we assumed that the number of photons detected in a specific time interval follows a Poisson density function. This is also considered as a valid assumption in the case of Infrared detectors [19]. Three noise levels (*layers*) were added to each one of the elemental images for the two scenes in such a way that the total number of noise photons added was:  $N_{dc} = \{10^4, 10^5, 10^6\}$ , simulating the existence of dark current (dc) noise. The proposed dark current noise values are relatively low when compared to real dark current values obtained in some types of Infrared detectors, but these levels might be obtained for the case of some HgCdTe mid-wave infrared detectors cooled to 50–80K (see for instance, Fig. 8 in [20] for details. For example, an HgCdTe sensor cooled at 80K and a cutoff wavelength of  $5\mu\text{m}$ , and assuming a quantum efficiency of 80%, there would be 1 photon per pixel per second of dark current). Figure 2 shows two elemental images corresponding to the two scenes. In Figure 2(a), the trees in front of the vehicle are at a distance of approximately  $z = 210\text{m}$ , and the vehicle is at a distance of  $z = 237\text{m}$ . In Figure 2(b) the hangars are at a distance  $z = 970\text{m}$  and the deposit at a distance of around  $z = 2200\text{m}$ . Figure 3(a) shows the results of the depth reconstruction for the case of the car and trees scene, for  $z = 237\text{m}$  when using the original elemental images. Figure 3(b) shows one of the elemental images of this scene, for  $N_p = 3.0 \times 10^5$  photons (i. e., 0.29 photons/pixel). Figure 3(c) shows the same elemental image, but for  $N_p = 3.0 \times 10^5$  photons, and  $N_{dc} = 10^6$  photons (1.24 photons/pixel in total). Figures 3(d)-(e) show the reconstruction results for  $z = 237\text{m}$  when using the elemental images without and with additional noise, respectively. In order to help improve its visualization, wavelet shrinkage technique for image denoising as explained in section V was applied on the depth reconstructed images. For the car and trees scene, a threshold value of  $T = 4$  was used. We can see in Figure 3(d) that the car shape can be clearly distinguished when only photon counting information is present. The addition of noise (Figure 3(e)) makes it more difficult its visualization. Figure 4(a)-(b) shows the results of the depth reconstruction for the case of airfield scene, for  $z = 960$  and  $z = 2200\text{m}$  when using the original elemental images. Figures 4(c)-(d) show the reconstruction obtained at  $z = 960$  and  $z = 2200\text{m}$ , respectively, for  $N_p = 3.0 \times 10^5$  photons (i. e., 0.29 photons/pixel). Figures 4(e)-(f) show the reconstruction obtained at  $z = 960$  and  $z = 2200\text{m}$ , respectively, for  $N_p = 3.0 \times 10^5$  and  $N_{dc} = 10^6$  photons (1.24 photons/pixel in total). Figures 4(c)-(f) were also post-processed applying the same wavelet shrinkage denoising technique used for Figure 3 (in this case, the threshold was set at:  $T = 20$ ). The Peak Signal to Noise Ratio (PSNR) can be used to compare the computational reconstruction using the original elemental images with the reconstructions using different levels of photon counting and added noise. PSNR is

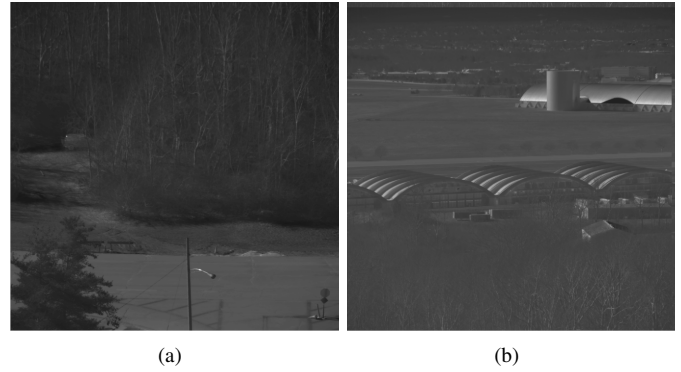


Fig. 2. Elemental images for: (a) the case of the car and trees scene and (b) for the airfield

defined as:

$$PSNR = 10 \log_{10} \left[ \frac{I_{max}^2}{MSE(I, \tilde{I})} \right] \quad (7)$$

where  $MSE(I, \tilde{I})$  is the Mean Square Error which provides an estimate of the average error per reconstructed pixel of the 3D scene.  $I$  is the computational reconstruction made with the original elemental images, whereas  $\tilde{I}$  is the reconstruction made by the elemental images with different levels of photon counts ( $N_{dc}$ ) and dark current photon counts ( $N_{dc}$ ).  $I_{max}$  is the maximum irradiance of the reconstructed scene. Figure 5 shows the PSNR vs  $N_p$  results for the reconstruction of the trees and car scene for a distance of  $z = 237\text{m}$  (distance where the car is in focus), for the case where no additional noise is added, and when  $N_{dc} = \{10^4, 10^5, 10^6\}$  photons are added to the elemental images. In [7] it was also shown that, for the case where only photon counting counts were considered, a dependence of the type  $PSNR \propto \log(N_p)$  followed. Figure 5 also shows the theoretical  $\log(N_p)$  dependence of PSNR, for the case where no additional noise is added, for comparison purposes. This Figure shows, on the one hand, that as  $N_p$  increases, the MSE error decreases and as a result PSNR increases. On the other hand, it also shows that PSNR decreases as we increase the  $N_{dc}$  noise in the elemental images, as expected.

## VII. CONCLUSION

In this paper, we have presented the results of the application of the maximum likelihood (ML) estimation method for visualization of 3D scenes in photon starved environments corresponding to Mid-Wave Infrared 3D data of real scenes. We have shown that visualization improves using multi-perspective photon-counted images, in relation to the information provided by each photon-counted elemental image.

## ACKNOWLEDGMENT

Pedro Latorre-Carmona would like to acknowledge Universitat Jaume I for his financial support through its call for research stays (E–2014–04).

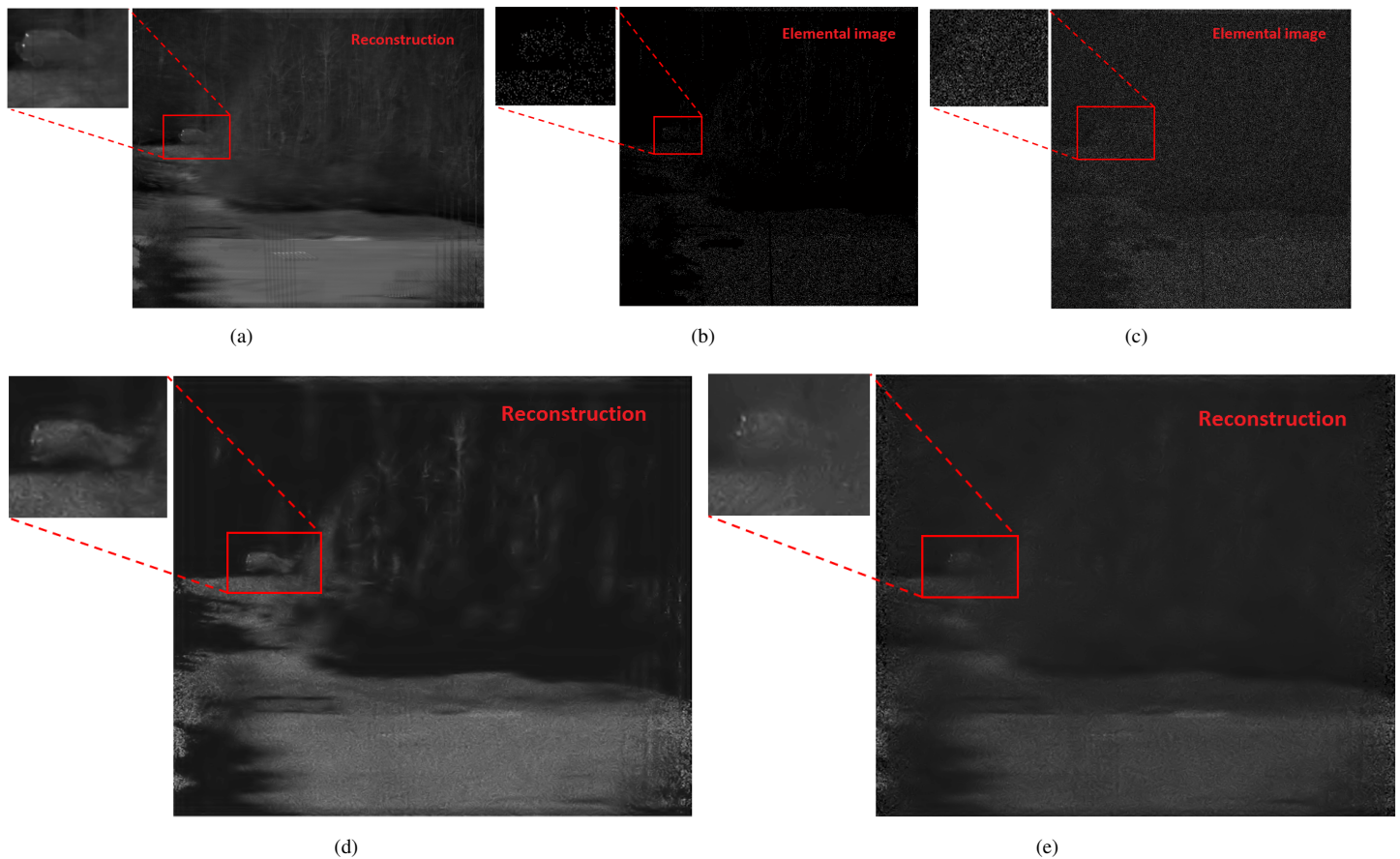


Fig. 3. Reconstruction results for the car and trees scene. (a) Depth reconstruction for  $z = 237\text{m}$  when using the original elemental images. (b) Elemental image, for  $N_p = 3.0 \times 10^5$  photons (i. e., 0.29 photons/pixel), without addition of dark current noise. (c) Elemental image, for  $N_p = 3.0 \times 10^5$  photons, where  $N_{dc} = 10^6$  photons have also been added (1.24 photons/pixel in total). (d)-(e) Reconstruction results for  $z = 237\text{m}$  when using the elemental images without and with additional noise, respectively. Wavelet shrinkage technique for image denoising was applied on Figures (d)-(e) for visualization purposes. Threshold value for the wavelet shrinkage technique:  $T = 4$ .

## REFERENCES

- [1] A. Stern, B. Javidi, *3D image sensing, visualization, and processing using integral imaging*, Proc. IEEE, vol. 94, pp. 591–607, 2006.
- [2] B. Javidi, F. Okano, and J. Y. Son, *Three-Dimensional Imaging, Visualization, and Display*, Springer, 2009.
- [3] R. Martinez-Cuenca, G. Saavedra, M. Martinez-Corral, B. Javidi, *Progress in 3-D multiperspective display by integral imaging*, Proc. IEEE vol. 97, pp. 1067–1077, 2009.
- [4] M. Cho, B. Javidi, *Three-dimensional visualization of objects in turbid water using integral imaging*, Journal of Display Technology, vol. 6, no. 10, pp. 544–547, 2010.
- [5] J. Arai, F. Okano, M. Kawakita, M. Okui, Y. Haino, M. Yohimura, M. Furuya, M. Sato, *Integral three-dimensional television using a 33-megapixel imaging system*, Journal of Display Technology, vol. 6, no. 10, pp. 422–430, 2010.
- [6] F. Okano, J. Arai, K. Mitani, M. Okui, *Real-time integral imaging based on extremely high resolution video system*, Proceedings of the IEEE, vol. 94, no. 3, pp. 490–501, 2006.
- [7] B. Tavakoli, B. Javidi, and E. Watson, *Three-dimensional visualization by photon counting computational Integral Imaging*, Optics Express, vol. 16, no. 7, pp. 4426–4436, 2008.
- [8] S. Yeom, B. Javidi, and E. Watson, *Photon counting passive 3D image sensing for automatic target recognition*, Opt. Express vol.13, pp. 9310–9330, 2005.
- [9] S. Yeom, B. Javidi, E. Watson, *Three-dimensional distortion-tolerant object recognition using photon-counting integral imaging*, Opt. Express vol. 15, pp. 1513–1533, 2007.
- [10] M. DaneshPanah, B. Javidi, E. A. Watson, *Three dimensional object recognition with photon counting imagery in the presence of noise*, Opt. Express vol. 18, pp. 26450–26460, 2010.
- [11] J. S. Jang, B. Javidi, *Three-dimensional synthetic aperture integral imaging*, Optics Letters vol. 27, pp. 1144–1146, 2002.
- [12] S. H. Hong, J. S. Jang, B. Javidi, *Three-dimensional volumetric object reconstruction using computational integral imaging*, Optics Express vol. 3, pp. 483–491, 2004.
- [13] J. W. Goodman, *Statistical Optics*, John Wiley and Sons, 1985.
- [14] N. G. Kingsbury: *The dual-tree complex wavelet transform: a new efficient tool for image restoration and enhancement*, Proc. European Signal Processing Conference (EUSIPCO), pp. 319–322, 1998.
- [15] F. Abdelnour, I.W. Selesnick, *Design of 2-band orthogonal near symmetric CQF*, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Salt Lake City, 6, pp. 3693–3696, 2001.
- [16] N. Kingsbury, *A dual-tree complex wavelet transform with improved orthogonality and symmetry properties*, Proc. IEEE International Conference on Image Processing, ICIP, 2, pp. 375–378, 2000.
- [17] M. Vetterli, J. Kovacevic, *Wavelets and subband coding*, Prentice Hall, 1995.
- [18] D. LeMaster, B. Karch, B. Javidi, *Mid-Wave Infrared 3D Integral Imaging at Long Range*, Journal of Display Technology, vol. 9, no. 7, pp. 545–551, 2013.
- [19] E. L. Dereniak, G. L. Boreman, *Infrared detectors and systems*, Wiley Series in Pure and Applied Optics, 1996.
- [20] J. W. Beletic, R. Blank, D. Gulbransen, D. Lee, M. Loose, E. C. Piquette, T. Sprafke, W. E. Tennant, M. Zandian, J. Zino, *Teledyne Imaging Sensors: infrared imaging technologies for astronomy and civil space*, Proc. SPIE 7021, High Energy, Optical, and Infrared Detectors for Astronomy III, 70210H, 2008.

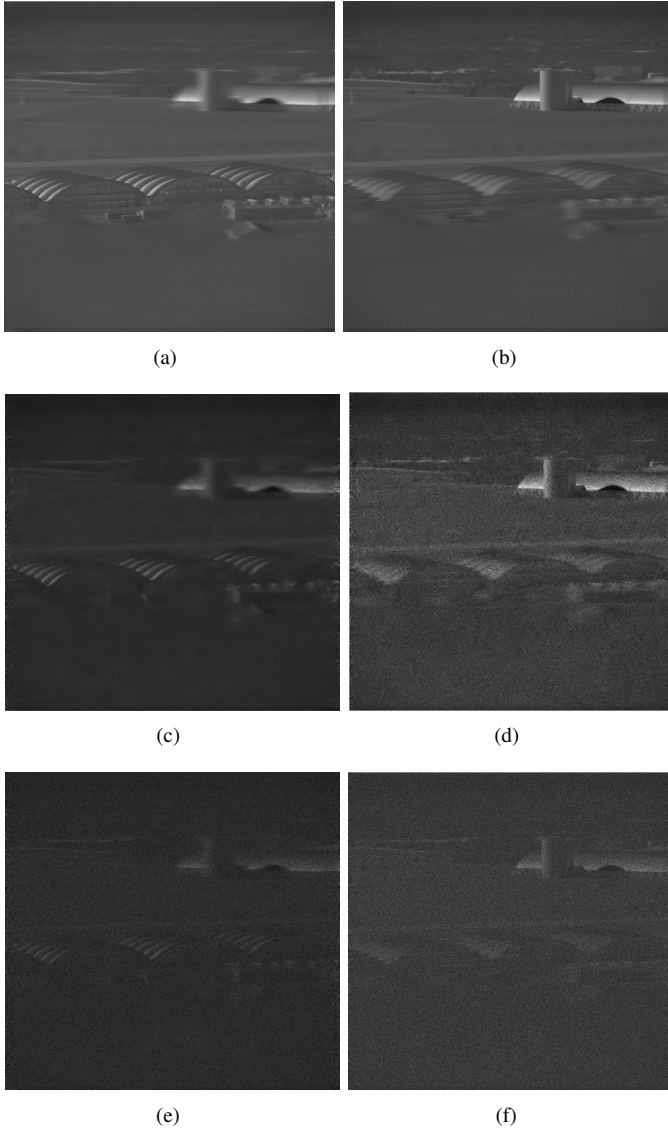


Fig. 4. Reconstruction results for the airfield scene.(a)-(b) Depth reconstruction for  $z = 960$  and  $z = 2200$ m when using the original elemental images. (c)-(d) Reconstruction at  $z = 960$  and  $z = 2200$ m, respectively, for  $N_p = 3.0 \times 10^5$  photons (i. e., 0.29 photons/pixel), without addition of dark current noise. (e)-(f) Reconstruction obtained at  $z = 960$  and  $z = 2200$ m, respectively, for  $N_p = 3.0 \times 10^5$  photons, where  $N_{dc} = 10^6$  photons have also been added (1.24 photons/pixel in total). Wavelet shrinkage technique for image denoising was applied on Figures (c)-(f) for visualization purposes. Threshold value for the wavelet shrinkage technique:  $T = 20$ .

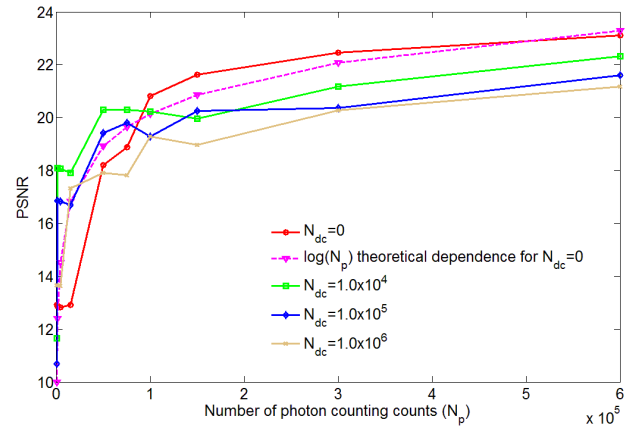


Fig. 5. PSNR vs  $N_p$  results for the reconstruction of the trees and car scene ( $z = 237$ m), for  $N_{dc} = \{0, 10^4, 10^5, 10^6\}$  dark current photons added to the elemental images.