Depth-independent segmentation of macroscopic three-dimensional objects encoded in single perspectives of digital holograms

Conor P. McElhinney and John B. McDonald

Department of Computer Science, National University of Ireland, Maynooth, County Kildare, Ireland

Albertina Castro

Instituto Nacional de Astrofísica, Óptica y Electrónica, Apdo. Postal 51, Puebla, Puebla 72000, Mexico

Yann Frauel

Departamento de Ciencias de la Computación, Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas, Universidad Nacional Autónoma de México, México D.F. 04510, Mexico

Bahram Javidi

Department of Electrical and Computer Engineering, University of Connecticut, U-157, Storrs, Connecticut 06269, USA

Thomas J. Naughton

Department of Computer Science, National University of Ireland, Maynooth, County Kildare, Ireland, and University of Oulu, RFMedia Laboratory, Oulu Southern Institute, Vierimaantie 5, 84100 Ylivieska, Finland

Received January 2, 2007; revised February 27, 2007; accepted March 1, 2007; posted March 5, 2007 (Doc. ID 78533); published April 17, 2007

We present a technique for performing segmentation of macroscopic three-dimensional objects recorded using in-line digital holography. We numerically reconstruct a single perspective of each object at a range of depths. At each point in the digital wavefront we calculate variance about a neighborhood. The maximum variance at each point over all depths is thresholded to classify it as an object pixel or a background pixel. Segmentation results for objects of low and high contrast are presented. © 2007 Optical Society of America

OCIS codes: 090.0090, 090.1760, 100.6890, 100.0100, 100.2000.

Holography is an established technique for recording and reconstructing real-world three-dimensional (3D) objects. Digital holography[1–4] and holographic image processing[5,6] have recently become feasible due to advances in megapixel CCD sensors with high spatial resolution and high dynamic range. We use phase-shift interferometry[4] to capture our in-line digital holograms (DHs), which are in an appropriate form for data transmission and digital image processing. We are interested in segmenting object regions from background regions in digital holographic reconstructions. Segmentation is generally the first step to object recognition, but it also has applications in video compression and motion tracking.[7] Segmentation has also recently been applied to 3D data, or layers of 2D reconstructions, in tomography and digital holographic microscopy.[8]

A DH $H_0(x,y)$ contains sufficient information to reconstruct the complex field $U_z(x,y)$ at any distance z from the camera.[2,4] One of several efficient options[9] is the Fresnel approximation[10] as

$$U_z(x,y) = \frac{-\mathrm{i}}{\lambda z} \exp\left(\mathrm{i}\frac{2\pi}{\lambda} z\right) H_0(x,y) * \exp\left[\mathrm{i}\pi\frac{(x^2+y^2)}{\lambda z}\right], \quad (1)$$

where λ denotes wavelength and * denotes convolution. One method for reconstructing a DH at the most in-focus plane is to use a depth-from-focus (DFF) technique. Although there is no definitive criterion for finding the focal plane of a scene or finding the focal distance for a region within a scene, a number of focus metrics have been proposed and demonstrated.[11–15] These employ self-entropy,[11] phase changes,[12] wavelet analysis,[13] gray-level variance,[14] and integrated amplitude modulus,[15] among others. Using these metrics, applications such as the detection of the focal plane[12,13,15] in digital holographic microscopy[16] and the measurement of 3D objects in digitized physical holograms[14] have been demonstrated. These techniques reconstruct over a range of depths and evaluate each 2D reconstruction using a focus metric, and when the focus metric gives a maximum the corresponding depth is returned. This relies on the assumption that a large majority of the scene is in focus at a single depth. This is not the case if there are multiple objects at different depths, or if the physical object itself extends in the z direction. This leads to a reconstruction containing some object regions in focus with the rest out of focus. These blurred out-of-focus regions make it difficult to segment a single reconstruction of a DH based on texture or intensity.

0146-9592/07/101229-3/$15.00

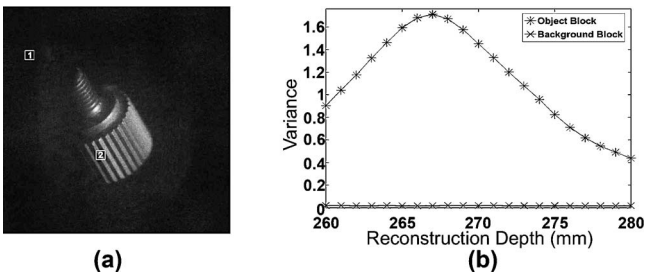

Fig. 1. Screw object DH: (a) numerical reconstruction, (b) variance plot for background and object regions.

An approach for the recovery of 3D shape information from digitized physical holograms was proposed by Ma *et al.*[14] By calculating variance on nonoverlapping blocks from reconstructions of a DH at different depths they recovered depth information from a lower-resolution version of the sensed object. We choose to extend this variance-measurement approach to classify each 1D vector (x,y) in the reconstruction volume (each line of pixels parallel to the optical axis) as either belonging to the object or belonging to the background. The decision is taken as follows: if vector (x,y) contains an in-focus pixel from the object at any depth z then (x,y) is an object pixel; otherwise it is a background pixel.

Each reconstruction $I_z=|U_z|^2$ is of size $M\times N$ pixels. We first apply a Fourier filtering speckle reduction technique to each reconstruction.[17] We then calculate variance for each pixel by calculating variance on $n\times n$ pixel overlapping blocks approximately centered on each pixel and address each block with (k,l), where $k\in[0,M-1], l\in[0,N-1]$. Variance of each overlapping block at each depth z is calculated with function $V_z:\mathbb{R}^{n\times n}\rightarrow\mathbb{R}^+$ defined by

$$V_z(k,l)=\frac{1}{n^2}\sum_{x=k-[(n-1)/2]}^{k+[(n-1)/2]}\sum_{y=l-[(n-1)/2]}^{l+[(n-1)/2]}[I_z(x,y)-\overline{I_z(k,l)}]^2, \tag{2}$$

where $\overline{I_z(k,l)}$ is defined as

$$\overline{I_z(k,l)}=\frac{1}{n^2}\sum_{x=k-[(n-1)/2]}^{k+[(n-1)/2]}\sum_{y=l-[(n-1)/2]}^{l+[(n-1)/2]}I_z(x,y), \tag{3}$$

and where any indexes (x,y) that go outside the extent of I_z evaluate to 0. V is therefore a volume storing a 2D variance image for each depth z. A location with high variance indicates the nearby presence of an object. We find the maximum value in each (k,l) vector with

$$V_{\max}(k,l)=\max_z[V_z(k,l)]. \tag{4}$$

Where the maximum variance is low, this indicates a background region. A threshold τ is chosen and $V_{\max}$ is transformed as

$$\mathrm{SMask}(k,l)=\begin{cases}1, & \text{if } V_{\max}(k,l)\geq\tau\\ 0, & \text{if } V_{\max}(k,l)<\tau\end{cases}, \tag{5}$$

where 0 denotes a background pixel and 1 denotes an object pixel. The binary image SMask is our segmentation mask. Finally, we apply a mathematical morphology erosion operation (with neighborhood $\lceil n/2\rceil\times\lceil n/2\rceil$) to SMask to shrink the boundaries of the object; our use of overlapping blocks uniformly enlarges the mask.

We verify our DFF technique using DHs of real-world objects. The first object (screw) is 1 cm^3 and was positioned 268 mm from the camera. Our second object (knight) is 2 cm $\times$ 2 cm $\times$ 0.6 cm and was positioned 371 mm from the camera. Each DH has 2032 $\times$ 2048 pixels. We compute a sequence of reconstructions at different depths from a single perspective with a uniform interval of 1 mm between successive values of z. We apply our DFF technique to this sequence of reconstructions to obtain SMask. A reconstruction of the screw object is shown in Fig. 1(a), where two 81×81 pixel blocks labeled 1 and 2 have been manually selected to indicate example background and object regions, respectively. Plots of variance calculated on these blocks over a range of 21 mm are shown in Fig. 1(b). It can be seen that the variance of the background block is 10^{-2} lower than that for the object block for this hologram. We found this to be true in the general case and allowed us to choose the appropriate normalized values of $\tau=0.02$ and $\tau=4\times10^{-5}$ for the screw and knight, respectively.

In Fig. 2 we compare our results with the ground-truth case, where each pixel is manually classified as background or object, and with the well-known and robust intensity-based segmentation technique known as expectation maximization[7] (EM). The shallow focal range of the screw object DH allows for a comparison between our method and this 2D technique. We use receiver operating characteristic (ROC) analysis[18] to display the relative trade-off between the true positive rate and the false negative rate of our classifiers. In this analysis the DFF block size is the variable in our choice of classifiers. We estimate the true positive (tp) rate of a classifier as

$$\text{tp rate}\simeq\frac{\text{True positives}}{\text{Total number of positives}} \tag{6}$$

and the false positive (fp) rate as

$$\text{fp rate}\simeq\frac{\text{False positives}}{\text{Total number of negatives}}. \tag{7}$$

Classification results are plotted on orthogonal axes defined by tp rate and fp rate, allowing us to choose the block size n that best maximizes the tp rate while simultaneously minimizing the fp rate. For the screw

Fig. 2. Segmentation of screw object DH: (a) manual, (b) our DFF approach, (c) EM approach.

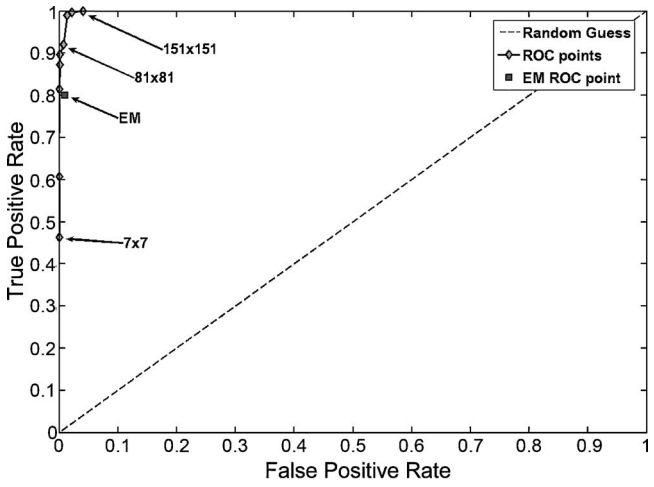

Fig. 3. ROC graph for object segmentation using different block sizes.

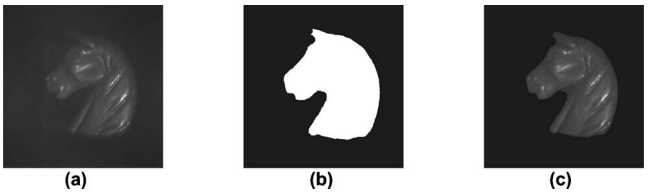

Fig. 4. Segmentation of knight object DH: (a) numerical reconstruction, (b) segmentation mask obtained, (c) segmented reconstruction.

object DH, we used a set of nine different block sizes, ranging from $n=7$ to $n=151$.

The ROC curve of this set of classifiers produced can be seen in Fig. 3(a). As desired, all points are located far from the random guess classifier performance. It is clear from these graphs that a small block size classifies background pixels perfectly at the expense of object pixels, and a large block size classifies object pixels perfectly at the expense of background pixels. Our compromise between perfect background/object segmentation is to minimize the distance between the points in ROC space and the point (0,1), since we regard false positives and false negatives as being equally undesirable. Through experiments with different objects we chose a block size of 81×81 pixels, which has an average tp rate of 98.83% and an average fp rate of 1.04%. Using this block size, we created the segmentation mask shown in Fig. 2(b). By comparison, EM achieved a good fp rate of 1% but a relatively poor tp rate of 80%. The segmentation mask EM is shown in Fig. 2(c).

We also present the results of applying our DFF technique to a DH of a low contrast object: a knight object whose reconstruction is shown in Fig. 4(a). As part of the speckle reduction measures taken for this DH, the reconstruction size has been reduced to 512×512 pixels. This is one quarter of the resolution of the previous experiments, which leads us to choose a block size of 21×21 pixels (one quarter of the block size chosen as a result of the ROC analysis). The erosion operator has a neighborhood of 11×11 pixels (half the block size used to create the segmentation mask). The resulting segmentation mask is shown in Fig. 4(b). A segmented reconstruction at a single depth z [Fig. 4(c)], obtained from $I_z(x,y)\cdot\mathrm{SMask}(x,y)$, where $\cdot$ means pointwise product, illustrates how the object can be successfully segmented from the background.

We expect our single object segmentation method will be successful for all macroscopic objects recorded by digital holography except pure phase objects. For microscopic objects it is expected that a phase-unwrapping-based approach would be best. The accuracy of our approach is limited by an appropriate choice of block size and threshold value. Also, this method is currently limited to the segmentation of the scene from its background where that background is not itself composed of other objects. Extensions to our technique are currently planned for multi-object segmentation.

The authors acknowledge support from Science Foundation Ireland, Enterprise Ireland, and IRCSET. C. P. McElhinney's e-mail address is conormce@cs.nuim.ie. T. Naughton's e-mail address is tomn@cs.nuim.ie.

References

1. T. Kreis, *Handbook of Holographic Interferometry* (Wiley-VCH, 2005).
2. L. Onural and P. D. Scott, Opt. Eng. **26**, 1124 (1987).
3. U. Schnars and W. P. O. Jüptner, *Digital Holography* (Springer, 2004).
4. I. Yamaguchi and T. Zhang, Opt. Lett. **22**, 1268 (1997).
5. T. J. Naughton, Y. Frauel, B. Javidi, and E. Tajahuerce, Appl. Opt. **41**, 4124 (2002).
6. J. Maycock, C. P. McElhinney, B. M. Hennelly, T. J. Naughton, J. B. McDonald, and B. Javidi, Appl. Opt. **45**, 2975 (2006).
7. D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*, 3rd ed. (Prentice-Hall, 2003).
8. M. DaneshPanah and B. Javidi, Opt. Express **14**, 5143 (2006).
9. B. M. Hennelly and J. T. Sheridan, J. Opt. Soc. Am. A **22**, 928 (2005).
10. J. W. Goodman, *Introduction to Fourier Optics* (Roberts and Company, 2005).
11. J. Gillespie and R. King, Pattern Recogn. Lett. **9**, 19 (1989).
12. P. Ferraro, G. Coppola, S. De Nicola, A. Finizio, and G. Peirattini, Opt. Lett. **28**, 1257 (2003).
13. M. Liebling and M. Usner, J. Opt. Soc. Am. A **21**, 2424 (2004).
14. L. Ma, H. Wang, Y. Li, and H. Jin, J. Opt. A, Pure Appl. Opt. **6**, 396 (2004).
15. F. Dubois, C. Schockaert, N. Callens, and C. Yourassowsky, Opt. Express **14**, 5895 (2006).
16. E. Cuche, F. Bevilacqua, and C. Depeursinge, Opt. Lett. **24**, 291 (1999).
17. J. Maycock, B. M. Hennelly, J. B. McDonald, T. J. Naughton, Y. Frauel, A. Castro, and B. Javidi, "Reduction of speckle in digital holography by discrete Fourier filtering," J. Opt. Soc. Am. A (to be published).
18. T. Fawcett, Pattern Recogn. Lett. **27**, 861 (2006).