

Using disparity in digital holograms for three-dimensional object segmentation

Tomi Pitkääho^a and Thomas J. Naughton^{b,a}

^a University of Oulu, RFMedia Laboratory, Oulu Southern Institute Vierimaantie 5, 84100 Ylivieska, Finland

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

ABSTRACT

Digital holography allows one to sense and reconstruct the amplitude and phase of a wavefront reflected from or transmitted through a real-world three-dimensional (3D) object. However, some combinations of hologram capture setup and 3D object pose problems for the reliable reconstruction of quantitative phase information. In particular, these are cases where the twin image or noise corrupts the reconstructed phase. In such cases it is usual that only amplitude is reconstructed and used as the basis for metrology. A focus criterion is often applied to this reconstructed amplitude to extract depth information from the sensed 3D scene. In this paper we present an alternative technique based on applying conventional stereo computer vision algorithms to amplitude reconstructions. In the technique, two perspectives are reconstructed from a single hologram, and the stereo disparity between the pair is used to infer depth information for different regions in the field of view. Such an approach has inherent simplifications in digital holography as the epipolar geometry is known a priori. We show the effectiveness of the technique using digital holograms of real-world 3D objects. We discuss extensions to multi-view algorithms, the effect of speckle, and sensitivity to the depth of field of reconstructions.

Keywords: digital holography, 3D image processing, 3D segmentation, 3D displays

1. INTRODUCTION

Holography makes it possible to record and reconstruct real-world three-dimensional (3D) objects.¹ Quite recently it has become reasonable to record holograms in a digital format. Digital recording is possible because of the recording sensors have evolved to their current state. Image segmentation is an image processing operation where image can be divided into regions based on some attribute.^{2,3} We are interested to segment objects from the background regions in a digital holographic reconstruction. Our combining attribute is the depth of the object. Different methods have been introduced to do segmentation from digital holograms. Segmentation has been reported in digital holographic microscopy⁴ and with macroscopic objects.⁵ 3D segmentation is an important phase in the object recognition, measurement and noise reduction. Also 3D segmentation is needed to process digital holograms for some of the conventional 3D displays.

In the second section of this paper we discuss about the important properties of digital holograms. Third part introduces our used method, disparity. Fourth section describes how disparity can be used in segmentation, which is followed by the conclusions.

2. DIGITAL HOLOGRAM PROPERTIES

In this section some of the unique properties of digital holograms are described. Described properties are encoded perspectives and depth of field. These properties are important for our disparity based 3D segmentation.

2.1 Encoded perspectives

A single hologram encodes multiple perspectives of the scene, which is a unique character of a hologram. Different perspectives of the scene can be reconstructed by using different parts of hologram data.⁶ In the reconstruction step a portion of the hologram data is used to get different perspective of the scene using

$$\Lambda[H(x, y)] = \begin{cases} 0, & \text{if } x \geq 0 \\ H(x, y) & \text{otherwise} \end{cases} \quad (1)$$

and

$$P[H(x, y)] = \begin{cases} 0, & \text{if } x < 0 \\ H(x, y) & \text{otherwise} \end{cases} \quad (2)$$

From these subsets of a whole hologram the perspectives can be reconstructed using

$$L = F_{\Lambda, z} \quad (3)$$

and

$$R = F_{P, z} \quad (5)$$

where F is the Fresnel approximation at any distance z given by

$$F_{H, z}(x, y) = \left[\frac{-i}{\lambda z} H(x, y) \bullet \exp \left[i\pi \frac{x^2 + y^2}{\lambda z} \right] \right]^2 \quad (6)$$

where the full hologram H could be replaced by Λ or P . Here, λ is the wavelength of the light and \bullet denotes a convolution operation. When using Λ and P , the equations above show how to get two horizontal perspectives (Figure 1).

To get different perspective of the scene different parts of the hologram can be used in the reconstruction step as shown in Figure 2. Not all digital holograms have the possibility to reconstruct different perspectives. Different perspective reconstruction is dependent of the hologram capture setup. Distance between camera and object has to be large enough for light reflected from the object to hit camera sensor wide enough area. This is the main principle for encoded perspectives and is illustrated in the Figure 3.

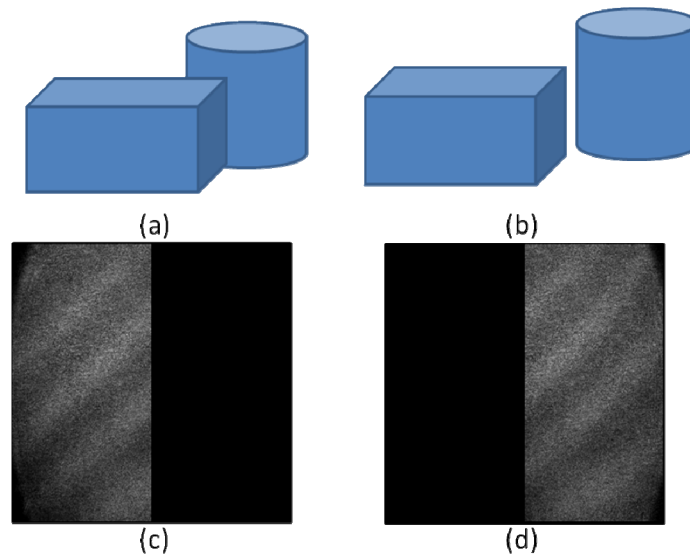


Figure 1. Different perspectives: (a) showing the scene from left perspective, (b) showing the scene from right perspective, (c) hologram data of left perspective, (d) hologram data of right perspective.

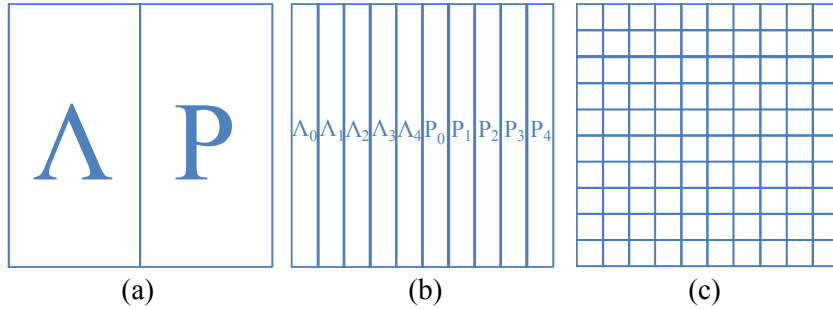


Figure 2. How to get different perspectives: (a) by using hologram data Λ and P two horizontal perspectives can be reconstructed, in the same manner vertical perspectives are possible, (b) by using hologram data 'slices' multiple horizontal perspectives can be reconstructed, in the same manner vertical perspectives are possible, (c) using any part of the hologram data, multiple horizontal and vertical perspectives can be reconstructed as long as used amount of data is sufficient.

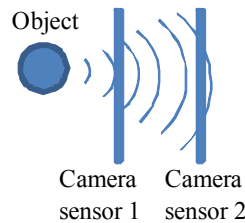


Figure 3. If light reflected from the object hits Camera sensor 1, wide angles are not possible in the reconstruction step. If camera sensor is far enough from the object, light reflected from point of object hits the sensor wide area and perspectives are possible (Camera sensor 2).

2.2 Depth-of-field in a digital hologram

Digital holograms have limited depth-of-field. This is not a problem with digital holograms because numerical reconstructions can be done at various depths through the volume by using a single hologram. However it is not always convenient to reconstruct at multiple depths, because this is computationally time consuming. With a single reconstruction depth depth-of-field can be extended by using smaller parts of the hologram data in the reconstruction step (Figure 4). Smaller reconstruction window leads to decreased resolution in the reconstruction and increased depth of field. So there is tradeoff between increased depth-of-field and noise in the reconstruction. Increased depth of field is desirable for many reasons. First of all with larger depth of field more things are visible in the reconstruction. Also using smaller reconstruction window enables wider range of angles.

3. DISPARITY

Stereo disparity is traditionally linked with human depth perception. A human's left and right eye receive different images because of their different positions. This difference between image locations is called a stereo disparity.⁷⁻⁹ We are taking advantage of digital hologram's property of encoded perspectives and use stereo disparity to get depth information of a holographic scene. To use stereo disparity with digital holograms is possible because of the way how digital holograms are recorded and different perspectives can be reconstructed.

By using these two perspectives L and R we calculate the disparity space image as follows. For each pixel in the left image $L(x, y)$, we estimate corresponding pixel $R(x', y)$ in the right image. In digital holography the epipolar geometry is known a priori, which simplifies the procedure and avoids the rectification stage. We only have to look in the horizontal direction because when computing reconstructions from a single hologram we can ensure that there is a perfectly horizontal baseline between perspectives. The disparity for that pixel (x, y) is the difference in pixel locations $x-x'$, and the set of disparity values corresponding to each pixel (x, y) is the disparity space image DSI. More formally, given left and right 2D perspectives of a 3D scene, L and R , respectively,

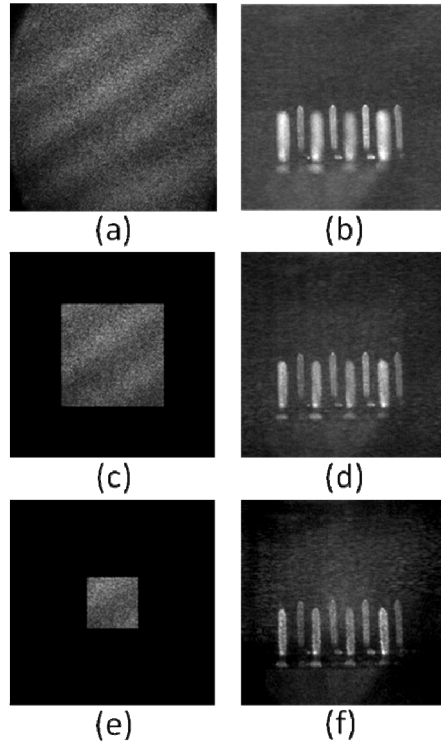


Figure 4. Depth of field with different reconstruction window sizes. Reconstruction distance is 186.25 mm and depth of object is approximately 10 mm. It is clear that smaller reconstruction window increases the depth of field: (a) reconstruction window with full hologram size 2048 x 2048 pixels, (b) numerical reconstruction of a, (c) reconstruction window 1024 x 1024 pixels, (d) numerical reconstruction of c, (e) reconstruction window 512 x 512 pixels, (f) numerical reconstruction of (e).

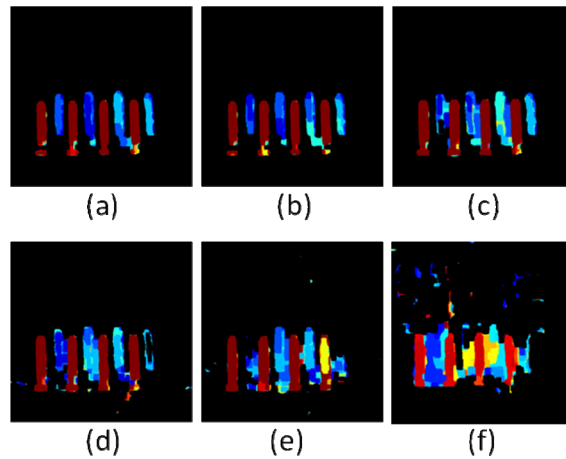


Figure 5. Depth of field effect on functionality of correlation based disparity algorithm. Shows DSIs with different reconstruction windows sizes. Red colour is closer to the camera: (a) DSI with reconstruction window sizes 2048x1024, (b) DSI with reconstruction window sizes 1536x1024, (c) DSI with reconstruction window sizes 1024x1024, (d) DSI with reconstruction window sizes 768x768, (e) DSI with reconstruction window sizes 512x512, (f) DSI with reconstruction window sizes 128x128.

$$DSI(x, y) = \begin{cases} \text{NaN, if } \max(C_{x,y}(u)) \leq \tau \\ \max(C(x, y)(u)), \text{ otherwise} \end{cases} \quad (7)$$

where $C_{x,y}$ is the normalized cross correlation function used to estimate the disparity. Operator max finds the disparity u that maximizes this correlation, and τ is a fixed threshold. For each pixel the normalized cross-correlation is calculated over a rectangular block of size $2t + 1 \times 2t + 1$ as

$$C_{x,y}(u) = \frac{\sum_{x,y} \tilde{L}_{x,y} \tilde{R}_{x,y}(u)}{\left[\sum_{x,y} (\tilde{L}_{x,y})^2 \sum_{x,y} (\tilde{R}_{x,y}(u))^2 \right]^{1/2}} \quad (8)$$

where $\tilde{L}_{x,y} = L(x, y) - \bar{L}_{x,y}$ and $\tilde{R}_{x,y}(u) = R(x - u, y) - \bar{R}_{x,y}(u)$. In these $\bar{L}_{x,y} = \frac{1}{(2t + 1)^2} \sum_{k=x-t}^{x+t} \sum_{l=y-t}^{y+t} L(k, l)$ is the mean

value of pixels of the left image within the block of pixels around (x, y) and $\bar{R}_{x,y}(u) = \frac{1}{(2t + 1)^2} \sum_{k=x-u-t}^{x-u+t} \sum_{l=y-t}^{y+t} R(k, l)$ is the

mean value of the pixel values of the right image within the offset block around $(x - u, y)$, $u = 0, \dots, T$ is the offset, and T is the maximum search length to be considered. An output of this phase is disparity space image DSI. Morphological image processing is applied to the DSI usually at this stage. Morphological operations are needed because the correlation based disparity algorithm function distorts the shape of the object making it larger than it actually is. Amount of disparity is dependent of the reconstruction depth as well as its direction. Disparity is 0 at the reconstruction depth, positive in front of it and negative behind it.

As mentioned earlier depth-of-field varies depending of the size of the reconstruction window. Because disparity is feature based algorithm, which tries to find correspondence between two images, big depth of field is desirable. Smaller reconstruction window brings more things visible in the reconstruction and correlation based disparity algorithm finds objects through the volume (Figure 5). If depth of the scene is so large that all the objects cannot be seen in the reconstruction, more reconstructions at different depths are needed. Using small reconstruction window verifies that disparity algorithm works well with noisy data, which digital holograms usually present. If part of a hologram is corrupted, lost or not transmitted disparity calculation is still possible. Smaller reconstruction window size also reduces needed calculation time. The disparity algorithm requires an appropriate amount of disparity for the best results. To get enough disparity, reconstruction windows have to be chosen correctly, which is not necessarily an easy task. Many unavoidable things cause noise on the reconstruction and noise overall causes problems in correlation based algorithms. Twin can be seen as a noise on top of the reconstruction. To avoid noise caused by a twin, picking up the window that halves the object decreases disturbance of a twin. Orientation of the object has the biggest effect on the disparity. Larger disparity and more 3D information is revealed when reconstruction windows are chosen perpendicular to the object orientation. The orientation of the object(s) can be calculated using the intensity thresholded reconstruction given by

$$Mask(x, y) = \begin{cases} 1, \text{ if } F_{H,z}(x, y) \geq \tau \\ 0, \text{ if } F_{H,z}(x, y) < \tau \end{cases} \quad (9)$$

where 0 denotes background pixel, 1 denotes an object pixel and τ is a threshold value for intensities. Object(s) in this binary image can be labeled with unique labels and for each object in the labeled image. From this labeled image orientation of the object(s) can be calculated and best reconstruction window positions can be determined (see Figure 6). When analyzing the orientation of the object seen in the Figure 6 horizontal windows should give best disparity result. Result for the DSI with vertical and horizontal reconstruction windows can be seen in Figure 7.

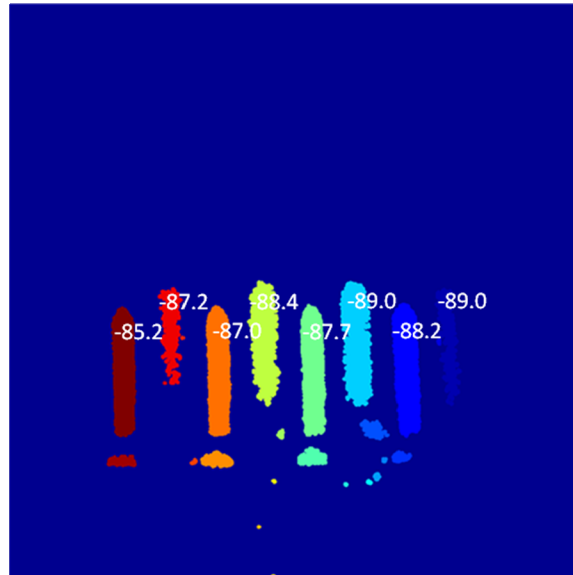


Figure 6. Using object's orientation to determine best possible reconstruction window size. Intensity based thresholding is applied on reconstruction's amplitude and orientation of the objects is calculated. Orientation is calculated in relation to the x axis. Orientation is given here from -90 to 90 degrees. If orientation is less than -45 or more than 45 degrees horizontal windows should give best disparity and with orientation between -45 to 45 choosing vertical windows more disparity and revealed 3D information should be available.

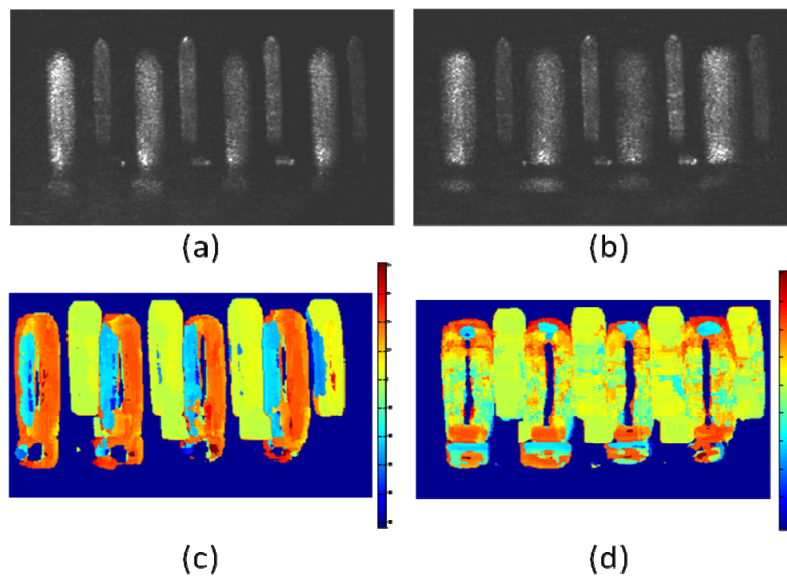


Figure 7. Choosing vertical and horizontal reconstruction windows: (a) left perspective of a digital hologram, (b) top perspective, (c) unprocessed disparity map which is obtained by using reconstructions from left and right perspectives and (d) unprocessed disparity map which is obtained by using top and bottom perspectives. More information about the 3D scene is available in (c) compared to (d).

4. 3D SEGMENTATION

In the DSI, most of the background is usually filtered out because a noise in the reconstructions moves more than our disparity algorithm tries to find. DSI can be considered as a depth map of the scene. From the DSI we can get a binary mask by using a threshold r as

$$SMask(x, y) = \begin{cases} 1, & \text{if } DSI(x, y) \geq r \\ 0, & \text{if } DSI(x, y) < r \end{cases} \quad (10)$$

where 0 denotes background pixel and 1 denotes an object pixel. The binary image $SMask$ is our segmentation mask. Multiplying image, which is an amplitude of reconstruction on optical axis we can segment object from the background. This is the procedure to segment objects as a whole and possible depth differences are ignored. For single objects experimental results show how this works in practice (Figure 8). Object in this case is a screw, which is located approximately 330 mm from the camera.

As experimental results show most of the noise around the object is not present in the DSI. Segmentation mask is easily obtained because of that. Using depth information of the scene object can be segmented individual parts according to the depth. Processing depth map in the Figure 8b furthermore we can combine depth map values and segment the screw object's different parts in individual segments. This is shown in the Figure 9. Real-world scenes have different 3D data available when viewed from different perspectives. Digital holograms offer possibility to reconstruct the scene from multiple perspectives, which can be used for this and furthermore disparity calculations can be extended to more than two perspectives. DSIs calculated from multiple perspectives can be combined and 3D information of the scene can be extracted from multiple perspectives.

5. CONCLUSIONS

We have shown that our method of using a digital hologram's unique property of encoded perspectives with a conventional computer vision approach can be used to obtain a depth map of the 3D scene, which can be used for 3D segmentation. Our approach has the advantage of needing only two reconstructions of a single hologram instead of need to reconstruct at large number of depths (for alternative techniques based on focus detection). Our technique is suitable for different kinds of hologram capture setups, both macroscopic and microscopic as long as distance between the object and the camera sensor is large enough. Also, we have shown some practices of how to choose the best reconstruction window positions when a stereo view of the scene is required.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from Science Foundation Ireland under the National Development Plan, the Academy of Finland, and the European Community's Seventh Framework Programme FP7/2007-2013 under 575 Grant agreement No. 216105 (acronym "Real 3D").

REFERENCES

- [1] Kreis, T., [Handbook of Holographic Interferometry], Wiley-VCH, Berlin, (2005).
- [2] Pratt, W., [Digital image processing], Wiley, New Jersey, (2007).
- [3] Horn, K. B. P., [Robot vision], MIT Press, Cambridge & Massachusetts & London, (1986).
- [4] Daneshpanah, M. and Javidi B., "Segmentation of 3d holographic images using bivariate jointly distributed region snake," *Opt.Express* 14, 5143–5153 (2006).
- [5] Mc Elhinney, C., McDonald J., Castro A., Frauel Y., Javidi B. and Naughton T.J., "Depth independent segmentation of macroscopic three-dimensional objects encoded in single perspectives of digital holograms," *Opt. Lett.* 32, 1229–1231 (2007).
- [6] Frauel, Y., Naughton, T., Matoba O., Tajahuerce, E. and Javidi B., "Three-dimensional imaging and processing using computational holographic imaging," *Proceedings of the IEEE*, 94, 636–653 (2006).
- [7] Marr, D. and Poggio, T., "Cooperative computation of stereo disparity," *Science* 194, 283–287 (1976).
- [8] Barnard, S. and Fischler, M., "Computational stereo," *ACM Comp. Surveys* 14, 553–572(1982).

[9] Scharstein, D. and Szeliski, R., "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," International Journal of Computer Vision 47, 7–42 (2002).

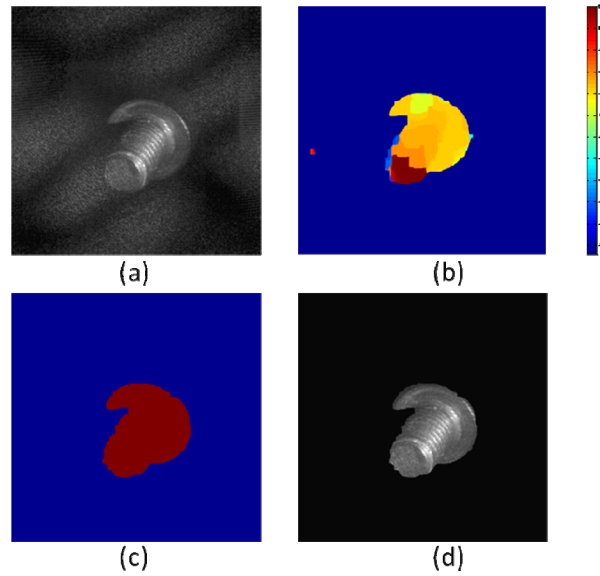


Figure 8. Segmentation of single object while ignoring the depth variance: (a) hologram numerically reconstructed at 333 mm from one perspective, (b) DSI of left and right perspective, (c) segmentation mask, which is obtained by thresholding b, (d) segmented reconstruction.

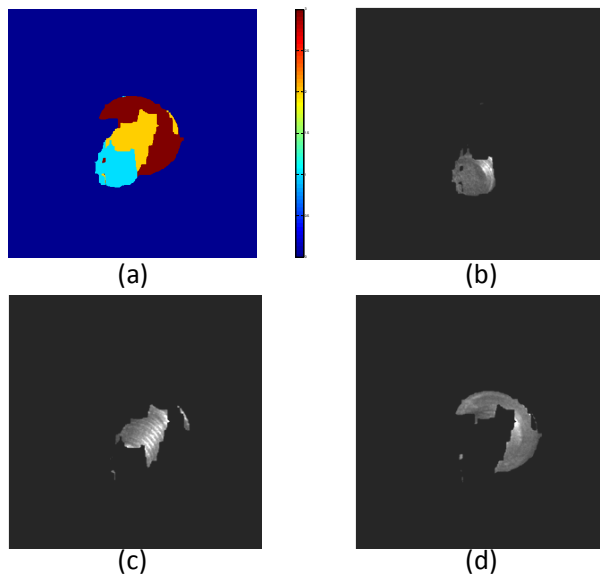


Figure 9. Segmentation of a single 3D object into multiple segments (a) depth map with four different depths, which is used to make segmentation masks, (b), segment 1: front part of the screw, (c) segment 2: middle part of the screw (d) segment 3: rear part of the screw.