Automatic depth map retrieval from digital holograms using a depth-from-focus approach

Nabil Madali, Antonin Gilles, Patrick Gioia, Luce Morin

Automatic depth map retrieval from digital holograms using a depth-from-focus approach

Nabil Madali[1,3*] Antonin Gilles[1] Patrick Gioia[1,2] Luce Morin[1,3]

[1] IRT b<>com Cesson-Sévigné France
[2] Orange Labs Cesson-Sévigné France
[3] INSA Rennes Rennes France

Abstract

Recovering the scene depth map from a computer-generated hologram is a problem that remains unsolved, despite the growing interest in the subject. In this article, we propose to study the application of depth-from-focus (DFF) methods to retrieve the depth information from the hologram. We discuss the different hyperparameters that are required for the application of the method and their impact on the final result. The obtained results show that DFF methods can be used for depth estimation from the hologram if the set of hyperparameters are well chosen.

1 Introduction

Holography, originally invented by Dennis Gabor in 1948 [1], is an immersive technology that is one of the most promising for reproducing a natural and realistic depth illusion. This technique records and reproduces the light waves emitted by an illuminated scene, giving the illusion that the object is physically present in front of the viewer. Holography provides all the depth perception cues of the human visual system [2] without the need to wear any type of glasses, which are necessary in stereoscopic displays. A single hologram contains more information than a traditional bi-dimensional image and requires a high storage cost, which becomes increasingly important when recording a video sequence. Unfortunately, this cost cannot be significantly reduced using conventional inter/intra-frame redundancy compression methods designed for 2D videos [3], due to significant differences between holographic signal and natural 2D images [2].

First, a natural image has the largest part of its energy concentrated on the low frequencies, and only a limited number of low frequency coefficients are needed to effectively compress the image. On the other hand, a hologram has a wider frequency spectrum evenly distributed on a restricted frequency band from the low to the high frequencies. As a result, eliminating high-frequency coefficients leads to high distortions and thus to poor compression performance. Second, during hologram recording, the light waves scattered by each scene point contribute to every pixels in the hologram. Thus, the holographic signal scrambles the 3D scene information, and there is no spatial localisation of the scene objects in the hologram plane. This is critical as localization is an important assumption for traditional motion estimation and compensation algorithms. Third, two consecutive holographic frames usually show random phases, which results in a non-zero residual when motion compensation is performed in the spatial domain.

To overcome these issues, Raees et al.[4] used the rotational transformation properties of the angular spectrum to design a comprehensive holographic video compression framework with motion compensation formulas in Fourier Space. Their method was able to handle a simple rigid motion of the scene between two consecutive frames. The residual after motion compensation was compressed in phase space with optimized distortion rate. They later extended their work to handle multiple objects with distinctive rigid motion [5], adding a segmentation step in phase space to extract the holograms of each individual object before motion compensation. In these studies, each object is assumed to have known rigid 3D motion and minimum bounding box. Indeed, motion estimation in holography is still an open problem in the general case.

To efficiently estimate the motion matrix of individual scene objects, the scene geometry has to be recovered from the hologram first. To this end, Depth from focus methods (DFF) [6], which have been successfully applied in holographic microscopy, stand as plausible choice. DFF methods compute RGB-D images that completely describe the scene geometry from a reconstruction volume of holographic reconstructions. We believe that from

*This work has been achieved within the Research and Technology Institute b<>com, dedicated to digital technologies. It has been funded by the French government through the National Research Agency (ANR) Investment referenced ANR-A0-AIRT-07. Authors can be reached at {nabil.madali, antonin.gilles, patrick.gioia, luce.morin}@b-com.com.

the RGB-D estimates of two consecutive frames, the bi-dimensional motion vectors on the XY axis could be estimated using traditional optical flow methods [7] on the RGB part, and the depth part could be used to estimate additional motion along the Z axis. Therefore, associating to each pixel of the reconstruction plane the relative three-dimensional motion vector would solve the motion estimation problem. Although the DFF approach is well documented in the literature, its extension to computer-generated holograms (CGH) is less studied. When extending the approach to CGH, many hyper-parameters such as the hologram generation method, resolution, pixel pitch, and reconstruction method, need to be taken into account and controlled to ensure relevant results. In this paper, we review the application of commonly used DFF methods for CGH, and a modified optimal focus plane selection criterion is introduced to improve their performance.

The remaining of the article is organized as follows: Section 2 reviews previous work that extend DFF for digital holographic microscopy data, and their limitations. In Section 3, we present our DFF methodology and modified optimal focus plane selection criterion for CGH data. In Section 4, a series of experiments are conducted to validate the proposed approach. In Section 5, we discuss the advantages and limitations of the proposed approach.

2 Related work

Depth from focus methods have been widely used in Digital Holographic Microscopy (DHM). They consist in reconstructing the hologram at different distances and in calculating for each of the reconstructions a Focus Measure (FM) operator. The maximum of the operator indicates the distance at which the microscopic object (cell, molecule, etc.) is from the hologram plane, such that

$$d = \underset{i=1\dots N}{\arg\max} \left\{ FM(I_i) \right\}, \tag{1}$$

where I_i is the amplitude of the i-th reconstruction plane. The overall performance depends both on the chosen reconstruction distances and on the focus operator. The focus operator must be independent of the image content; it must have a variation that depends on the amount of blur, and the function must be monotonic and uni-modal, with a maximum value reached when the focus is optimal. Finally, the operator must be robust to the different types of noise that may be present in the image. Several operators have been proposed in the literature for natural images, and an exhaustive review of commonly used operators can be found in [8].

Three types of focus operators can be distinguished in DHM; we briefly describe them in the next subsections.

2.1 Sparsity metrics

The sparsity metrics are indicators of the level of sparsity of a given signal. A signal is said to be sparse if a limited number of coefficients in the chosen basis are required to efficiently represent the signal. The simplest sparsity metric is the ℓ_0 norm, which computes the number of non-zero elements of the signal. However, it remains rarely used in practice, since the observed values are often noisy, making the measure inappropriate [9]. Several extensions of this metric have been proposed, including the ℓ_ϵ^0 [10], $-\ell^p$ [11], and $\frac{\ell^2}{\ell^1}$ [12] norms to alleviate limitations of the ℓ_0 norm.

In [13], a study of 16 commonly used sparsity operators in DFF has been performed; a further extension of this study is completed in [14], introducing 16 additional operators such as the $tanh_{a,b}$ and log operators, the kurtosis measure, Hoyer measure, and the Gini Index. These operators have been applied to both the intensity and amplitude of numerical reconstructions of holograms recorded from different diatom cells. The authors highlighted that the operators exhibit a global minimum at the optimal focal plane. However, extracting the optimal focal plane can be tedious since some operators can produce several local minima. Therefore the authors proposed a high-pass filtering step using Savitzky–Golay filter [15] before selecting the optimal focus plane. The obtained results showed that the used operators exhibit similar behavior and can be classified into four different groups according to their reliability. From the set of initially used operators, the authors retained six operators ($\frac{\ell^2}{\ell^1}$, pq-mean, Shannon entropy, modified Shannon entropy, and Gini index) as the most reliable ones.

Liebling et al. [16] proposed to measure the sparsity level in the Fresnelet basis. At each reconstruction distance a Fresnelet decomposition is performed, and the decomposition that requires the fewest coefficients is selected as the optimal focal plane. The experimental results showed that the proposed approach has better performance and is more robust to noise than the Laplacien and squared intensity operators. However, a large number of reconstructions can lead to computation overhead.

2.2 Gradient-based methods

The gradient-based operators [17] are a set of operators which use either the first or the second derivative of the image to detect the amount of contours and edges that are present. An in-focus image is expected to have

a high spatial derivative, therefore, the maximum value of the metric curve can be considered as the optimal focus plane.

Tamamitsu *et al.* [18] investigated the performance of two sparsity metrics (Gini index and Tamura coefficient) applied to spatial gradient modulus of a stack of holographic reconstructions extracted from an in-line hologram of an isolated Giardia lamblia cyst. The authors applied the sparsity metrics on a variable size crop of the obtained spatial gradient modulus. As the crop size increases, the metric curve of the Gini index-based operator changes polarity, i.e. the optimal focus plane is reached at the peak or valley of the focus operator depending on the chosen crop size. Both operators exhibit similar behavior for small crop sizes. However, for large crop sizes, the operator based on the Tamura coefficient showed higher performance.

In [19], the authors compared four focusing operators including the spatial gradient, the Laplacian operator, the local gray level variance and the summation of logarithmically weighted power spectra on DHM with Epifluorescence microscopy. The authors highlighted the high sensitivity to noise of the gradient-based operators and they retained the local gray level variance [20] as the most suitable metric for their experiments. Some of the selected operators exhibit several local minima and therefore require iterative search methods to efficiently retrieve the optimal focus plane, increasing the computational complexity and preventing real-time processing.

2.3 Transform-based methods

Transform-based methods include operators that transform the image into the frequency or space / frequency domain, such as classic discrete cosine transform [21, 22, 23, 24], or wavelets transforms[25, 26]. The transformed domain allows to measure the change between the high and low frequencies in the image, an in-focus image containing more high frequency coefficients than out-of-focus ones.

In [27], the performance of three focus operators, including the gray variance, the Weighted Fourier spectral function, and the standard deviation correlation function, are evaluated. To reduce the computation time by a half, the authors applied the operators on a cropped part of the whole reconstructed image. The focus curves of the three operators exhibit a global maximum at the optimal focus plane with a stronger peak for the gray variance and standard deviation correlation functions. The authors retained the Fourier spectral evaluation function as the optimal focus operator for its shortest computation time.

In [28], the authors investigated the effect of the chosen reconstruction method on the overall performance, and fifteen transform-based and sparsity-based focusing operators were evaluated on phase-shifted synthetic and experimental holograms. The synthetic holograms have no additional uncertainty factors, resulting in better performance. Depending on the used operators, either the global maximum or minimum of the metric curve is selected as the optimal focus plane. The experiments showed that the Angular Spectrum propagation method (ASM) enabled the depth estimation error to be divided by two compared to the Fresnel transform. The authors highlighted the importance of adding a noise-filtering step before applying the focus operators, and that gradient-based operators are more sensitive to noise than other operators. The authors also pointed out the high computational complexity of transform-based measures over the remaining set of focus measures.

Additional focus operators based on statistical methods [29, 30, 31, 32] and energy-based methods [33, 34, 35, 36, 37] can be found in the literature. They are beyond the scope of this article and therefore will not be discussed here.

Although focus operators have been successfully applied in DHM, some peculiarities of computer-generated holograms (CGH) make the direct use of these operators inefficient. CGH synthesis methods allow the calculation of holograms from synthetic scenes with a large depth extent, unlike holographic microscopy where holograms are acquired optically from shallow cells or microscopic objects. Computer-generated holograms can be synthesized from point-clouds, polygon meshes, or planar layers. An exhaustive review of these methdos can be found in [38]. Therefore, while in DHM, the scenes have a small depth extent and can be approximated by a single focus plane per hologram, in CGH the scene can be very deep and even a small change in the spatial coordinates along X, Y can cause a significant depth variation. In order to efficiently compute a depth map with a value per pixel of the hologram, the decomposition of the reconstruction volume into sub-regions must be performed, and the focus operators must be applied independently to each region in order to extract the optimal focus plane. In view of these limitations, the state-of-the-art methods used for DHM are not applicable for CGH. This is why we propose a new method of DFF adapted to CGH. In particular, this method allows to compute a local depth estimate and thus provides a depth value for each pixel of the reconstruction volume. Moreover, since the light diffraction formulas used to numerically reconstruct the holograms simulate the use of a monochromatic laser, and therefore produce speckle noise in the reconstructed images, a generalized selection criterion allows to take into account the speckle noise [39] characteristic of holograms, and the polarity inversion according to the focus operator.

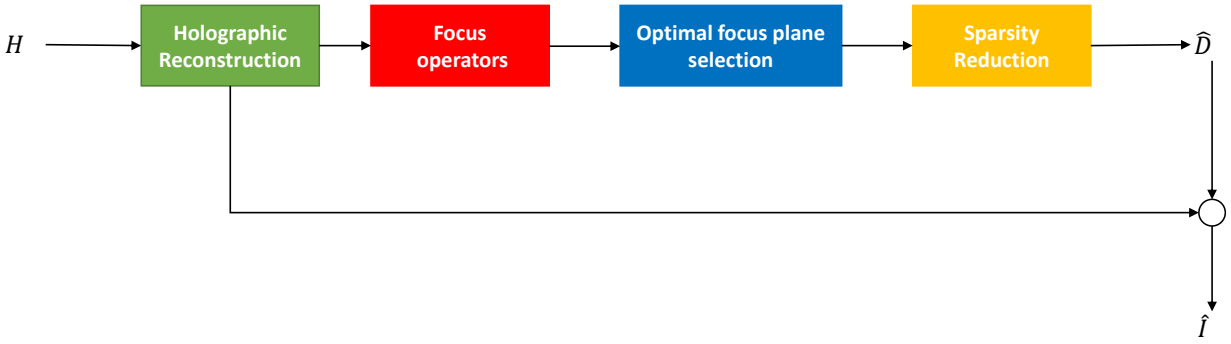

Figure 1: Block diagram representing the different steps of the proposed method.

3 Methodology

In this section, we present our proposed DFF methodology for computer-generated holograms. Our approach comprises four steps, depicted in Figure 1. First, the input hologram H is reconstructed at different focus distances. Then, each reconstructed image I_i is decomposed into overlapping rectangular patches $R_{i,m,n}$. A focus measurement operator FM is computed on each patch, and the optimal focus depth $d_{m,n}$ is selected for this patch. Finally, a voting agreement method is applied to obtain the pixel-wise estimated depth map $\hat{D}$. From the estimated depth map and reconstruction volume of reconstructions, an in-focus image $\hat{I}$ can be calculated. In the following, these different steps are further described.

3.1 Numerical reconstruction

First, given an input hologram denoted by H, a set of holographic reconstructions are performed at N uniformly sampled points in the interval $[z_{\min}, z_{\max}]$, where $z_{\min}, z_{\max}$ are the minimal and maximal depth of the scene, respectively. The holographic reconstruction at a sampled distance z_i with $i \in [1, N]$ is given by the Angular Spectrum Method [2] as

$$\mathcal{P}_{z_i}\{H\}(x,y) = \mathcal{F}^{-1}\left\{\mathcal{F}(H)e^{j2\pi z_i\sqrt{\lambda^{-2}-f_x^2-f_y^2}}\right\}(x,y), \tag{2}$$

where $\mathcal{F}$ denotes the 2D Fourier transform, f_x and f_y are the spatial frequency along the X and Y direction of the hologram plane, λ is the acquisition wavelength, and z_i is the reconstruction depth, given by

$$z_i = \frac{z_{\max} - z_{\min}}{N} i + z_{\min}. \tag{3}$$

Only the amplitude of these holographic reconstructions is kept to constitute the reconstruction volume denoted by $I_1, I_2, ..., I_N$. One of the advantages of holographic reconstructions over natural images is that the reconstruction volume is already aligned; therefore, there is no need to have a registration step to a reference image as for natural pictures, saving computation time.

3.2 Reconstruction volume sampling

To effectively estimate the depth with small numerical error, the number of reconstructions and their distance to the hologram need to be chosen wisely. An under-sampling of the reconstruction distances will induce a high estimation error since the sampled reconstruction distance can be expected to be far away from the optimal focus plane. As the sampling rate increases, the distance between the reconstruction planes and the optimal focus plane is reduced, therefore a lower numerical error can be expected. However, performing a large number of holographic reconstructions can induce a significant computational cost and many reconstruction planes will contain no relevant information.

The sampling rate is also related to the length of the reconstruction interval which is limited by two extreme points at the front and back of the scene. A large scene will have a large reconstruction interval and therefore a finer sampling rate is necessary to obtain a relevant depth estimate. On the other hand, a smaller scene is encapsulated in a smaller interval, so fewer reconstructions are needed to correctly recover the scene depth. In this work, the sampling rate N is fixed at twice the number of acquisition planes [40] of the largest scene of our dataset. Therefore, for each scene in the dataset, the reconstruction volume will contain enough discriminating information to correctly estimate the depth over the entire reconstruction image.

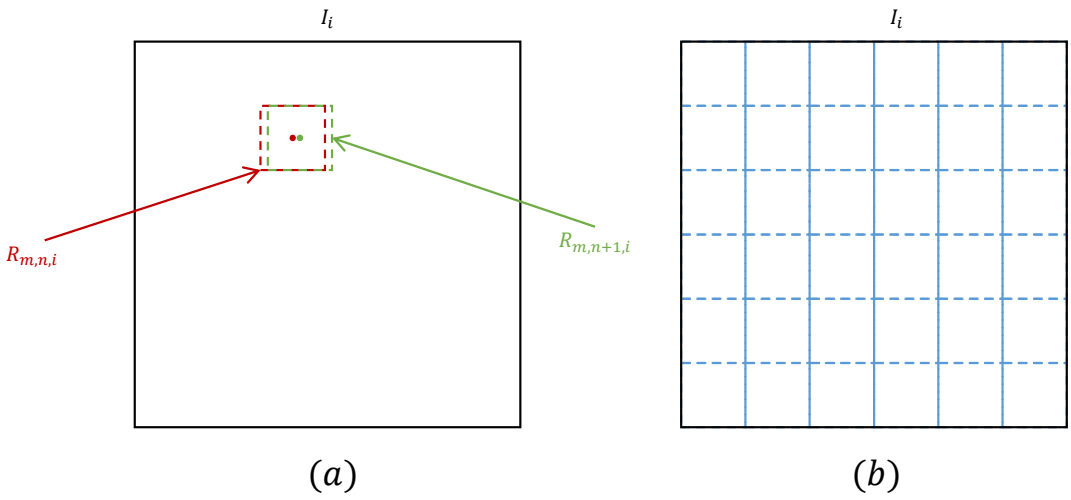

Figure 2: Illustration of two types of reconstruction volume decomposition. In (a) the overlapping patch-based decomposition used in the present work, where the level of focus of each pixel (m, n) in the i-th image I_i of the reconstruction volume is evaluated locally using a centered patch around the pixel $R_{m,n,i}$. In (b) the non-overlapping patch-based decomposition, where each image of the reconstruction volume is decomposed into non-overlapping patches and the level of focus is evaluated for each patch.

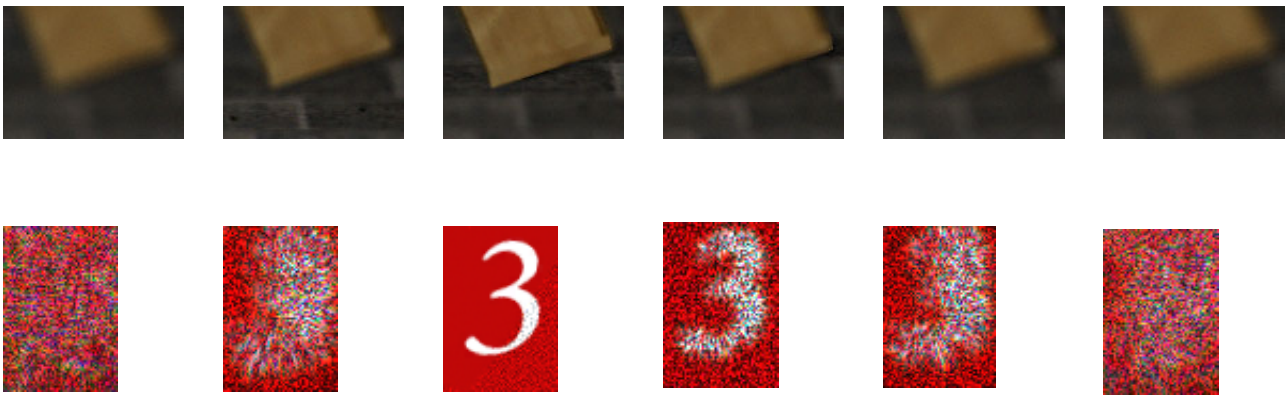

Figure 3: Comparison of focus change in natural images (top) and holographic reconstructions (bottom).

3.3 Pixel-wise depth estimation

Contrarily to DHM, in CGH the reconstructed scene can have a large depth extent, with objects located at different focus distances. As a consequence, it is necessary to decompose each reconstruction into either non-overlapping or overlapping patches with a given stride as shown in Figure 2. Since decomposing the reconstruction image in patches of fixed depth is not easily feasible, we choose to estimate the depth for each pixel of the reconstruction image.

For each pixel (m, n) in the reconstructed image I_i a rectangular patch $R_{i,m,n}$ of resolution $(s \times s)$ centered around the pixel (m, n), is defined such that:

$$R_{i,m,n}(u, v) = I_i(m + u - s/2, n + v - s/2). \tag{4}$$

The focus metrics are then applied to each of these patches independently, as detailed in the following:

$$d_{m,n} = \underset{i=1\ldots N}{\arg\max} \left\{ FM(R_{i,m,n}) \right\}, \tag{5}$$

which results in a depth map containing for each pixel of the reconstruction image, the depth at which the focus is locally optimal. Note that the patch-based decomposition of the reconstruction volume is applied independently of the focus operator. Therefore, the segmentation algorithm has no impact on the performance of the focus operator.

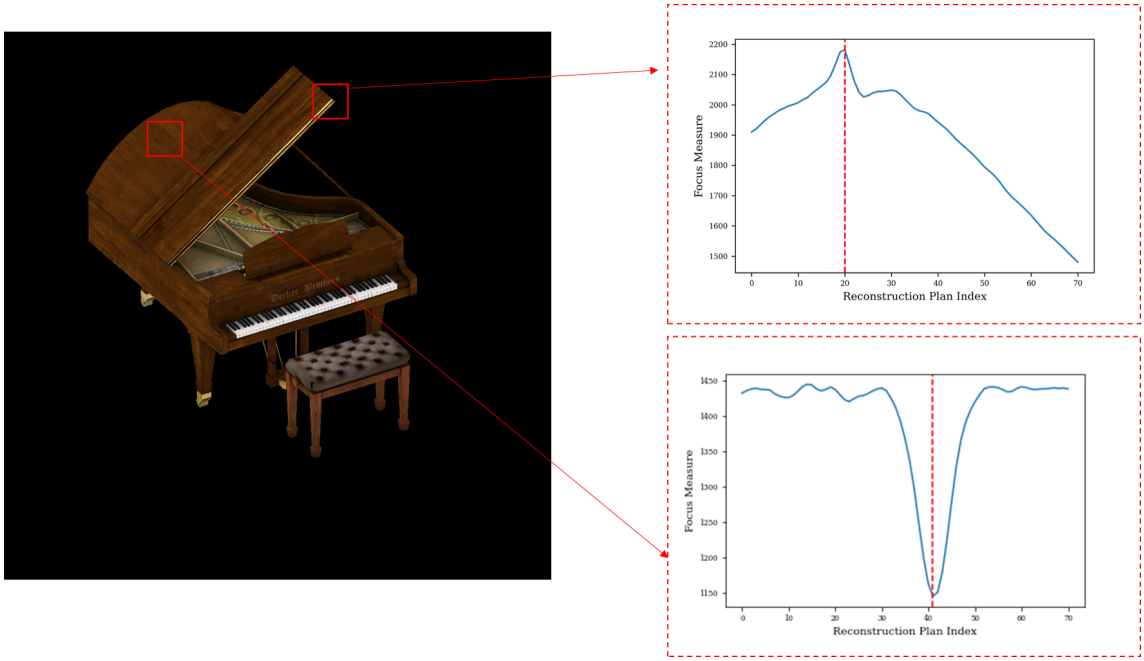

Figure 4: Illustration of polarity change on two different patches for gray-scale focus measurement.

3.4 Auto-switch criterion for optimal focal plane selection

Depending of the chosen focus metric, estimating the depth as the index of the global maximum can result in poor performance due the low depth-of-field and speckle noise in holographic reconstructions. Figure 3 compares the focus change in natural images and in holographic reconstructions. For natural images, as the relative distance from the optimal focal plane increases, the patch of interest becomes increasingly blurred while edges and contours are preserved to some extent. A completely different behavior is noticed for holographic reconstructions, where moving away from the optimal focus plane results in important contamination by speckle noise, and edges as well as contours disappear due to the low depth-of-field.

Therefore, the depth estimation can be reformulated as detecting at what distance the patch of interest is the least contaminated by speckle noise. For example, when choosing the gray-level variance as focus metric, a constant value can be expected outside of the optimal focus plane, since the selected patch of the reconstructed image is heavily contaminated with constant-variance speckle noise. When the optimal focus plane is reached, the focus function exhibits either a global minimum when processing a smooth region, or a global maximum for highly textured region as shown in Figure 4. The same analysis can be extended to other focus metrics.

By making the assumption of a constant operator value starting from a distance ϵ from the optimal focal plane, an efficient criterion for optimal focus plane selection can be devised that we call "auto-switch selection". The depth estimation problem can be formulated as detecting the index at which the sharpness function differs greatly from the focus metric observed in the reconstruction volume. Only the reconstructions in a neighborhood of ϵ around the optimal focal plane are sharp. Apart from this neighbor, the patch of interest is considered to be completely contaminated by speckle noise and therefore gives a constant value with respect to the operator. In order to effectively estimate the depth, it is necessary to automatically detect when the operator deviates from the observed mean value, either by reaching a global minimum or maximum. More formally,

$$d_{m,n} = \arg\max_i \left\{ \left| \bar{d} - FM(R_{i,m,n}) \right| \right\}, \qquad \bar{d} = \frac{1}{N} \sum_{i=1}^{N} FM(R_{i,m,n}). \tag{6}$$

Introducing this new auto-switch formulation allows to automatically handle the polarity change either induced when switching from one focus operator to another, or when changing the region of interest. In addition to the optimal focal plane section criterion, it is important to consider other hyper-parameters that may affect the final result, which we will discuss in the next subsection.

3.5 Voting Agreement

As shown in Figure 5, two challenging scenarios can occur on a given frame of the reconstruction volume: the first case is when no region of the reconstruction image is in-focus, which occurs when no part of the scene is

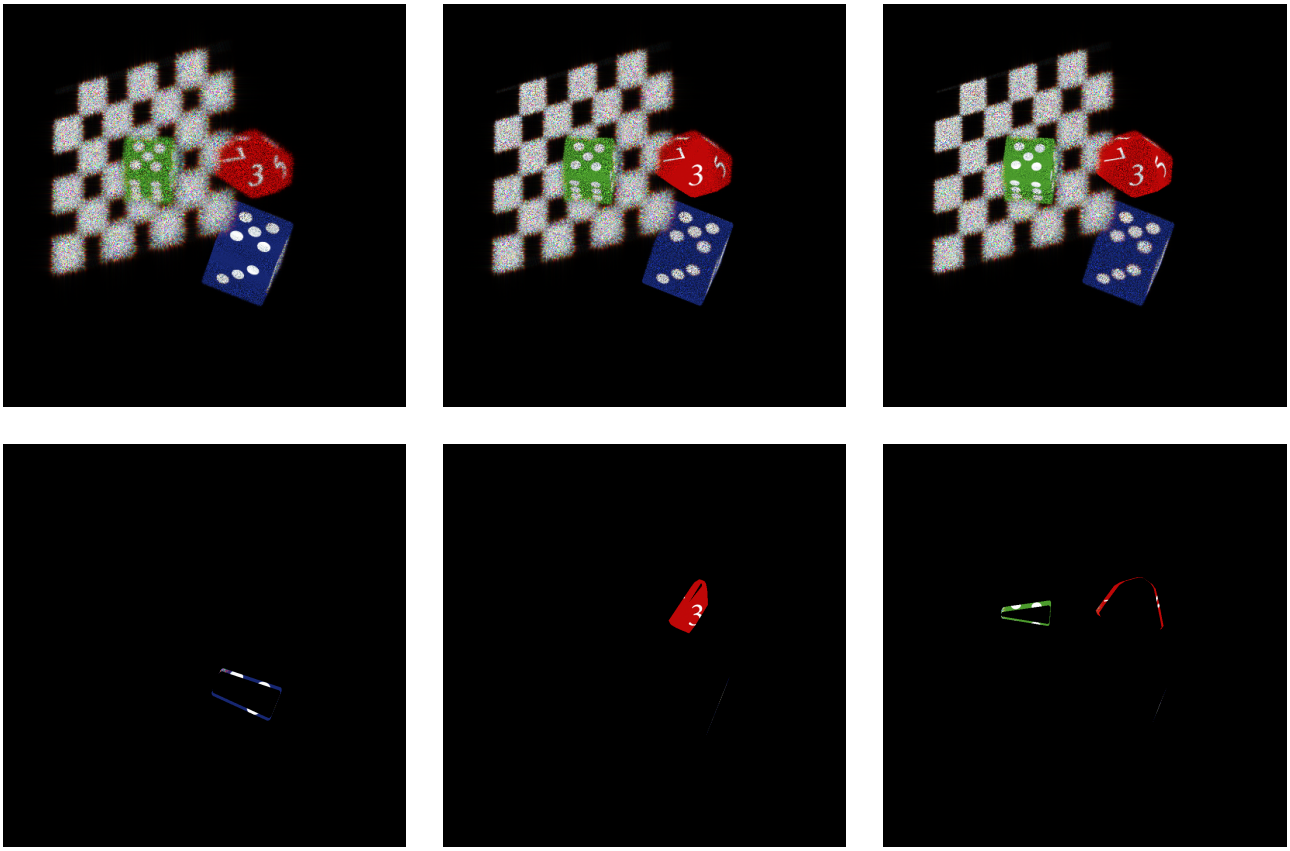

Figure 5: Each column contains a holographic reconstruction performed at a different reconstruction distances. In the top row the reconstruction amplitude is shown, and at the bottom the in-focus regions.

at the same depth as the chosen reconstruction distance. In this case, a good focus operator should produce a low focus value, however, when the sampling rate of holographic reconstructions is high, many reconstruction distances are sampled around the optimal focus plane. Even if they are out of focus, they produce a focus value close to the optimal focus plane, and therefore lead to a noisier estimated depth value.

The second case is when both in-focus and out-of-focus sub-regions are present. The geometries of in-focus sub-regions depending on the initial orientation of the scene, the method, and the primitives chosen for the hologram acquisition. If the depth value is estimated for a pixel (m, n) inside one of the in-focus regions, the cropped patch $R_{i,m,n}$ around the pixel must be perfectly adjusted to the geometry of the primary in-focus region to which the pixel (m, n) belongs, to ensure that the patch $R_{i,m,n}$ contains only one optimal focus plane.

Figure 6 illustrates this situation. The patch $R_{i,m,n}$ is composed of three planar sub-regions which are acquired at three distinct depths. The upper part of the figure illustrates the reconstruction volume resulting from multiple holographic reconstructions. On each reconstruction plane one sub-region at most is in focus, the remaining sub-regions are contaminated by speckle noise. In this particular case, if the patch $R_{i,m,n}$ is used to estimate the depth of the central pixel, the probability of choosing the optimal focus plane is $1/3$ under the optimal conditions. To better estimate the depth of the central pixel, the size of the cropped patch must be reduced considerably. As shown in Figure 6, choosing a smaller patch size (the green dotted line) ensures that only one optimal focus plane can be encapsulated inside the patch.

The patch size thus plays an important role for optimal performance. The larger the patch size, the more important the probability of encapsulating sub-regions with different focus levels. On the other hand, when the patch size is too small, less discriminating information is extracted, reducing the final performance. Moreover when estimating the depth of a pixel located at the junction of two sub-regions, regardless of the size of the chosen patch size, the patch will be split into two sub-regions that are in focus at different reconstruction distances.

To increase the confidence in the prediction, a consensus process between neighbor pixels is introduced. Rather than assigning to each pixel of the reconstruction image the depth estimated from a patch centered on this pixel, a voting process is performed on all the depth estimates extracted using patches that include the targeted pixel. By this way, each time a pixel is included in a patch, an additional vote is counted at the index of the estimated depth for this patch. When all the pixels are processed, the focal plane with the maximum number of votes is chosen. Therefore, even if the estimate from the patch centered on the target pixel

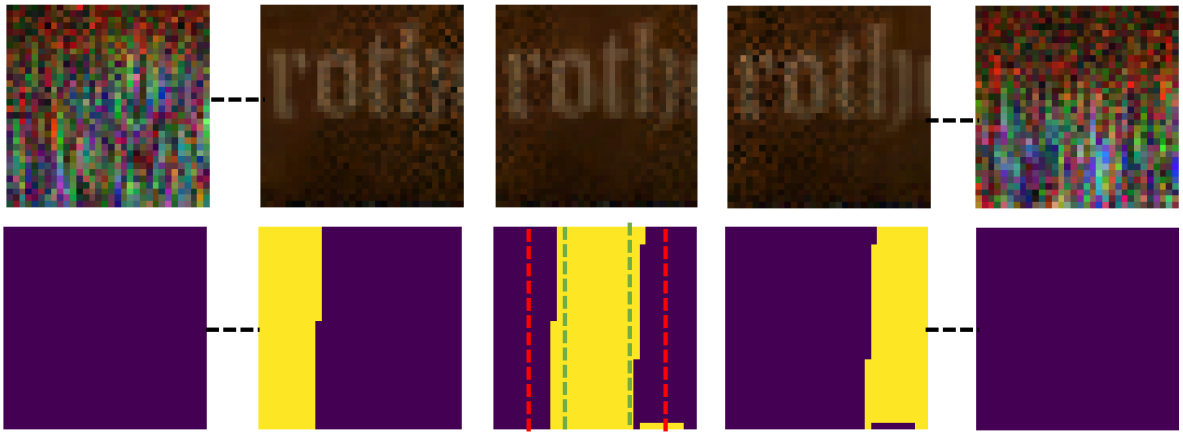

Figure 6: An illustration of a sequence of images constituting a cropped reconstruction volume and the corresponding in-focus map.

is incorrect, the propagation of information from neighbor pixels increases the confidence in choosing a specific plane of focus, while reaching a local consensus, also avoiding sparsity in the final prediction.

The resulting estimated depth map $\hat{D}$ can be viewed as a 2D representation of the 3D scene, where the columns and rows represent the X and Y coordinates, and the intensity value represents the Z coordinate. The depth map can be easily back-projected into a 3D point-cloud, where the color information of each point is assigned to the corresponding intensity pixel value of the reconstructed image at the estimated depth. The RGB-D image thus completely describes the scene geometry. The pseudo code of the method is given in Algorithm 1.

4 Experimental results

4.1 Experimental Setup

To evaluate the proposed approach, 500 holograms divided equally among five different scenes (Piano, Table, Woods, Cars, Dices shown in Figure 10) were acquired using the layer based method proposed in [40], with a resolution of 1024×1024 and a pixel pitch of $6\mu m$. For each hologram, a set of $N = 250$ holographic reconstructions are performed at uniformly sampled distances in the interval $[z_{\min}, z_{\max}]$ as given in Eq. (3), where $z_{\min}, z_{\max}$ are the minimal and maximal depth of the scene set to 0.0049cm and 1.23cm in the experiment. Only the amplitude of each reconstruction is kept, which is further converted into grayscale values before applying the focus operators.

From the set of focus operators available in the literature, 21 operators listed in Table 1 were selected for their reliability and performance. In addition to evaluating the performance of focus operators, the patch size was also examined. A ℓ_1 norm was selected as performance metric for its simplicity, which is defined as

$$\mathcal{L}(d, \hat{d}) = \frac{1}{N_x N_y} \sum_{i,j} \left| d_{i,j} - \hat{d}_{i,j} \right|, \tag{7}$$

where N_x, N_y are the height and width of each hologram, and d and $\hat{d}$ are the ground truth and estimated depth maps. Before computing the performance metrics, the foreground objects have been segmented from background, by thresholding the intensity variation of each pixel of the reconstruction plane, removing all the pixels with low intensity variation. The background points represent the 3D space surrounding the objects which do not actually have a depth information, and therefore should not be evaluated in the final performance score.

4.2 Results

Tables 2 and 3 show the obtained results using different patch sizes and focus measure operators, without and with the voting agreement process, respectively.

First, we observe that for each focus operator, the patch size has an important impact on the final result. For small patch sizes (under 13×13 pixels), the performance is better on areas with strong spatial gradient

Algorithm 1 The proposed DFF algorithm

Require: Hologram H, number of reconstructions N, reconstruction interval $[z_{\min}, z_{\max}]$, patch resolution s, and focus measurement operator FM.

Ensure: Predicted depth map $\hat{D}$ and in-focus image $\hat{I}$

for $i = 1...N$ **do**
  $z_i = \frac{z_{\max} - z_{\min}}{N} i + z_{\min}$ ▷ Initialize reconstruction distance
  $I_i = \left| \mathcal{F}^{-1} \left\{ \mathcal{F}(H) e^{j2\pi z_i \sqrt{\lambda^{-2} - f_x^2 - f_y^2}} \right\} \right|$ ▷ Angular Spectrum reconstruction
  $I_i = ZeroPadding2D(I_i, [s/2])$ ▷ Zero-padding of the reconstruction
end for
$V = 0_{N_x, N_y, N}$ ▷ Initialize voting scores
for all hologram pixels (m, n) **do**
  for $i = 1...N$ **do**
    for $u = 0...(s-1)$ **do**
      for $v = 0...(s-1)$ **do**
        $R_{i,m,n}(u, v) = I_i(m + u - s/2, n + v - s/2)$ ▷ Initialize rectangular patch
      end for
    end for
  end for
  $\bar{d} = \frac{1}{N} \sum_i FM(R_{i,m,n})$ ▷ Mean focus measure
  $d_{m,n} = \arg\max_i \left\{ \left| FM(R_{i,m,n}) - \bar{d} \right| \right\}$ ▷ Optimal focus plane selection
  for $u = -s/2...(s/2-1)$ **do**
    for $v = -s/2...(s/2-1)$ **do**
      $V(m + u, n + v, d_{m,n}) = V(m + u, n + v, d_{m,n}) + 1$ ▷ Increase voting score
    end for
  end for
end for
for all hologram pixels (m, n) **do**
  $\hat{D}(m, n) = \arg\max_i \{V(m, n, i)\}$ ▷ Predicted depth map
  $\hat{I}(m, n) = I_{\hat{D}(m,n)}(m, n)$ ▷ Predicted in-focus image
end for

Focus operator	Abbr.	Focus operator	Abbr.
Gaussian derivative [41]	GDER	Brenner's measure [42]	BREN
Energy of gradient [43]	GRAE	Image contrast [44]	CONT
Thresholded absolute gradient [42]	GRAT	Graylevel variance [45]	GLVA
Squared gradient [46]	GRAS	Graylevel local variance [47]	GLLV
Tenengrad [45]	TENG	Normalized Graylevel variance [42]	GLVN
Tenengrad variance [47]	TENV	Histogram range [21]	HISR
Energy of Laplacian [43]	LAPE	Steerable filters [48]	SFIL
Modified Laplacian [49]	LAPM	Spatial frequency [46]	SFRQ
Variance of Laplacian [47]	LAPV	Vollath's correlation [42]	VOLA
Diagonal Laplacian [50]	LAPD	Variance of wavelet coefficients [51]	WAVV
DCT energy ratio [52]	DCTE	Ratio of the wavelet coefficients [51]	WAVR
DCT reduced energy ratio [53]	DCTR		

Table 1: Abbreviations of focus measure operators used in experiments.

	5×5	9×9	13×13	17×17	21×21
Gradient based operators					
GRAE	24.14/24.09	11.77/10.24	9.13/7.25	7.91/6.35	7.3/6.1
GRAT	33.7/27.24	17.71/8.34	12.43/5.27	10.73/4.29	9.86/4.01
TENG	25.84/25.79	15.26/14.05	11.27/9.49	9.38/7.77	8.39/7.09
TENV	24.3/24.3	15.02/15.02	10.43/10.43	8.42/8.41	7.49/7.47
Laplacian-based operators					
LAPE	33.9/33.9	18.24/17.9	11.72/10.73	9.11/8.13	7.87/7.07
LAPM	37.09/33.62	20.73/10.75	13.96/6.58	11.67/5.17	10.6/4.66
LAPV	33.92/33.92	18.25/17.9	11.72/10.73	9.11/8.13	7.87/7.07
LAPD	31.73/23.33	16.44/7.57	11.93/4.91	10.63/4.11	9.96/3.89
DCT-based operators					
DCTE	29.53/29.38	14.83/14.11	9.5/9.08	8.23/8.04	7.92/7.8
DCTR	32.03/32.03	16.63/16.49	10.82/10.69	9.01/8.95	8.34/8.3
Wavelet-based operators					
WAVV	34.65/34.65	18.81/18.5	11.99/11.04	9.29/8.33	7.96/7.17
WAVR	15.88/14.5	8.83/6.86	7.68/5.63	7.28/5.34	7.09/5.38
Miscellaneous operators					
BREN	18.49/17.29	11.96/9.09	10.47/6.78	9.6/6.11	9.12/6.1
CONT	15.78/ **5.89**	11.79/ **2.79**	11.0/ **2.25**	10.78/ **2.21**	10.85/ **2.29**
GLVA	15.89/ **6.88**	10.69/ **3.25**	9.23/ **2.98**	8.34/ **3.22**	7.71/ **3.54**
GLLV	21.17/21.17	11.47/11.47	8.56/8.55	7.36/7.35	6.7/6.67
GLVN	15.52/ **13.41**	9.03/ **5.16**	7.99/ **3.94**	7.55/ **3.95**	7.34/ **4.22**
HISR	27.82/20.38	17.36/11.09	12.77/8.92	10.51/8.0	9.16/7.57
SFIL	42.92/33.35	40.06/25.76	36.94/20.49	33.06/16.87	29.63/14.81
SFRQ	38.04/35.26	21.76/11.7	14.65/7.09	12.08/5.52	10.9/4.9
VOLA	62.17/40.84	46.34/27.22	38.65/17.89	40.01/12.4	42.7/9.54

Table 2: L1 error (argmax/ auto-switch) without voting process step.

like contours, edges, and corners, but, as expected, the results are poor on flat areas with very low textures as shown in Figures 7 and 8. As the patch size increases, more discriminating information is encapsulated inside the patch, which helps to improve the final result. However, when using very large patch sizes (over 21×21 pixels), there is a strong probability of observing multiple focus levels inside the same patch, and choosing a single depth value per patch results in poor performance as illustrated in the last columns of Figure 8. Using the voting agreement process significantly reduces the error induced by multi-level focus inside the patch, as shown in Table 3. From these results, we observe that the optimal patch size is 13×13.

Second, a better performance can be noticed when using the automatic switch between the global maximum and minimum, especially for large patches where polarity change may occur. The gain is less noticeable when using the DCT-based operators and operators based on variance. For operators based on variance such as LAPV, a pre-operation is applied (here the Laplacian) and then the variance of the patch is computed. Outside the optimal focal plane, since the patch is completely contaminated by speckle noise, the pixel value variance is relatively small. Therefore keeping only the maximum value can result in relatively high performance. The flat regions will also result in small patch variance, however, the obtained value can be included in a larger interval produced by the remaining value computed on the reconstruction volume. In other words, the focus operator produces an optimal depth value that will be neither a global maximum nor a minimum. The same analysis can be extended to DCT-based operators. In order to effectively use the auto-switch, the operator's value obtained at the optimal focus plane must produce a large deviation from the average value observed in the reconstruction volume.

Finally, introducing the voting system gives a significant performance gain as shown in Figure 10. The obtained gain decreases as the patch size increases, since each pixel is included in a higher number of patches, introducing uncertainty during the selection of the optimal focal plane. A better result might be obtained if a confidence value (such as Softmax value output from neural network) is associated with the increment,

	5×5	9×9	13×13	17×17	21×21
Gradient based operators					
GRAE	21.84/21.84	9.7/8.21	7.77/5.99	6.8/5.56	6.34/5.57
GRAT	32.04/23.24	13.53/4.88	9.35/3.13	7.81/2.83	7.22/2.85
TENG	24.35/24.32	12.87/11.9	9.63/8.1	7.98/6.68	7.22/6.43
TENV	23.32/23.32	13.56/13.56	9.33/9.33	7.65/7.65	6.82/6.82
Laplacian-based operators					
LAPE	32.54/32.54	14.94/14.82	9.46/8.6	7.47/6.65	6.74/6.24
LAPM	35.72/31.22	16.31/6.85	10.36/4.0	8.4/3.4	7.53/3.28
LAPV	32.57/32.57	14.95/14.82	9.46/8.6	7.47/6.65	6.74/6.24
LAPD	29.82/19.01	12.74/4.57	9.11/3.0	7.66/2.77	7.13/2.8
DCT-based operators					
DCTE	28.16/28.08	13.35/13.08	8.95/8.84	7.99/7.96	7.91/7.92
DCTR	30.96/30.96	15.1/15.07	10.16/10.15	8.83/8.82	8.33/8.32
Wavelet-based operators					
WAVV	33.5/33.5	15.46/15.36	9.61/8.85	7.69/6.89	6.77/6.31
WAVR	13.68/12.64	7.98/5.85	7.2/4.9	6.98/4.86	6.89/5.06
Miscellaneous operators					
BREN	16.67/15.63	11.11/7.31	9.6/5.43	8.77/5.17	8.14/5.36
CONT	14.24/ **3.77**	10.24/ **1.61**	9.39/ **1.54**	8.76/ **1.63**	8.38/ **1.88**
GLVA	14.31/ **4.15**	10.03/ **1.72**	8.78/ **1.91**	7.92/ **2.38**	7.06/ **2.89**
GLLV	19.39/19.39	10.61/10.61	7.85/7.85	6.66/6.66	6.03/6.03
GLVN	12.82/ **10.94**	8.22/ **3.6**	7.51/ **2.99**	7.32/ **3.29**	7.23/ **3.68**
HISR	26.98/18.82	15.75/9.22	11.39/7.48	9.64/7.29	8.34/6.91
SFIL	42.87/31.31	38.84/22.16	34.84/15.95	31.07/11.94	27.44/9.44
SFRQ	36.99/33.65	17.24/7.67	10.89/4.36	8.75/3.63	7.67/3.43
VOLA	68.81/40.45	28.7/23.44	24.35/13.97	24.67/9.36	26.31/7.22

Table 3: L1 error (argmax/ auto-switch) with voting process step.

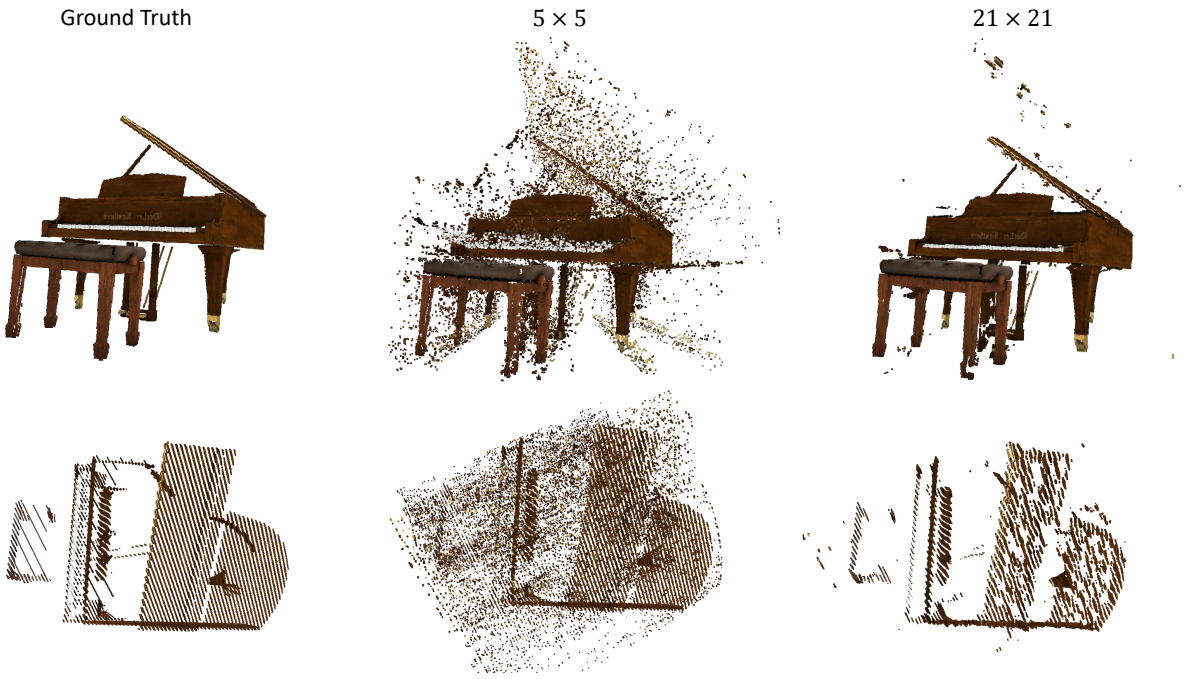

Figure 7: Illustration of the patch size influence on the final performance using the CONT operator with the proposed metric.

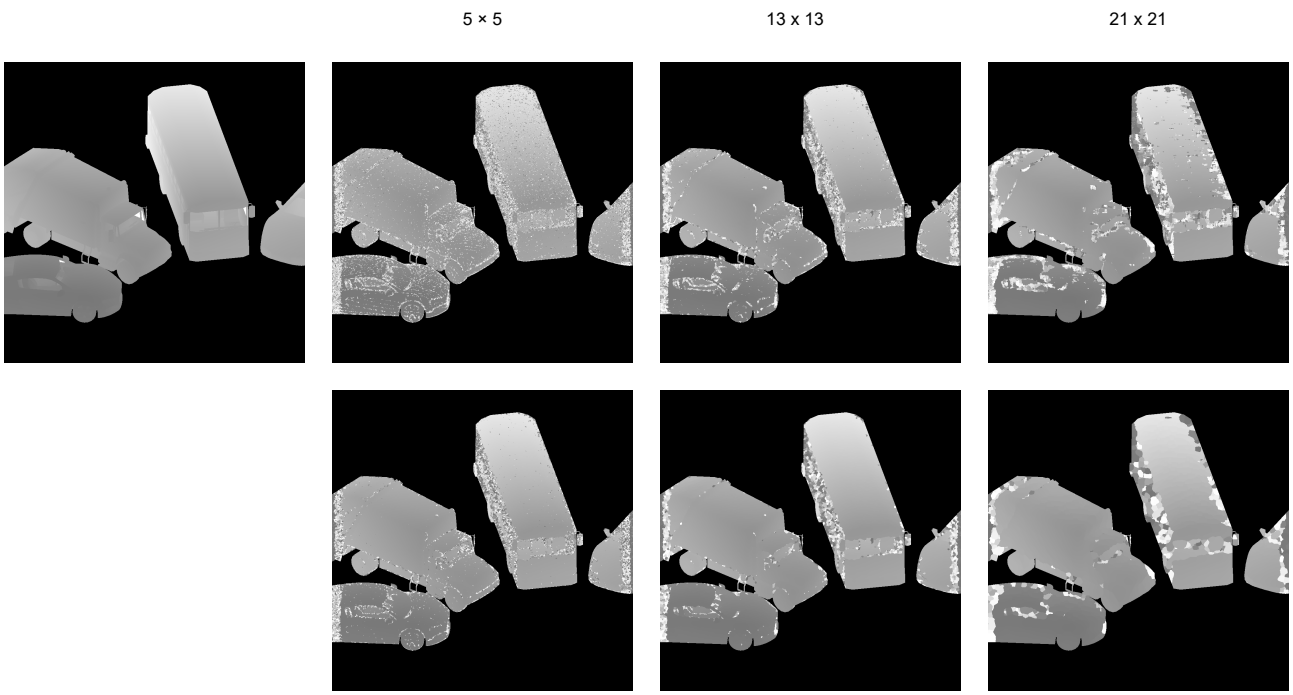

Figure 8: The first column illustrates the ground truth depth map. The remaining columns illustrate in the top row the initial estimate using the GLVA operator with the proposed metric and in the bottom row the results obtained after the voting process with increasing patch sizes ($5 \times 5, 13 \times 13, 21 \times 21$).

therefore balancing between true-positive and false-positive votes. The gain is also less noticeable when the initial prediction is wrong, due to the fact that erroneous information is propagated, leading to a loss of performance.

Most operators converge towards the same results when using a reasonable patch size and the adequate optimal focal plane selection criteria. However, some operators require more computation time [8], so it is important to take this into account when selecting the optimal focus measurement operator. Overall, the most accurate depth prediction was obtained using the Image contrast (CONT) focus metric with a patch size of 9×9 and with the voting system. To verify the validity of the proposed approach and focus metric, we back-projected the retrieved intensity and depth maps to obtain a 3D point cloud for each test scene, shown in Figure 9. As shown in this figure, the estimated 3D point clouds are very close to the ground truth. These experimental results pave the way for all the practical applications requiring the scene geometry to be recovered from the hologram, including holographic data segmentation, classification and edition, as well as holographic video coding with motion estimation and compensation.

5 Conclusion

In this paper, we studied the suitability of DFF methods for the depth estimation from CGH holograms. We proposed a modified "auto-switch" optimal depth selection criterion to handle the polarity change of focus operators, and a voting agreement process to reduce depth estimation errors. Overall, we found that the CONT [44], GLVA [45], and GLVN [42] focus operators gave the most accurate depth prediction results for a patch size of 13×13 with the voting process. The DFF approach may be a judicious choice to recover the geometry of the scene. However, its important to correctly choose the many hyper-parameters which are involved in the different stages of the depth estimation process in order to get an estimated depth map with the least numerical errors.

We believe that the depth maps retrieved with our methodology can be used for many practical applications requiring the scene geometry to be recovered from the hologram, including holographic data segmentation, classification and edition, as well as holographic video coding with motion estimation and compensation.

References

[1] Dennis Gabor. A new microscopic principle. *Nature*, 161:777–778, 1948.

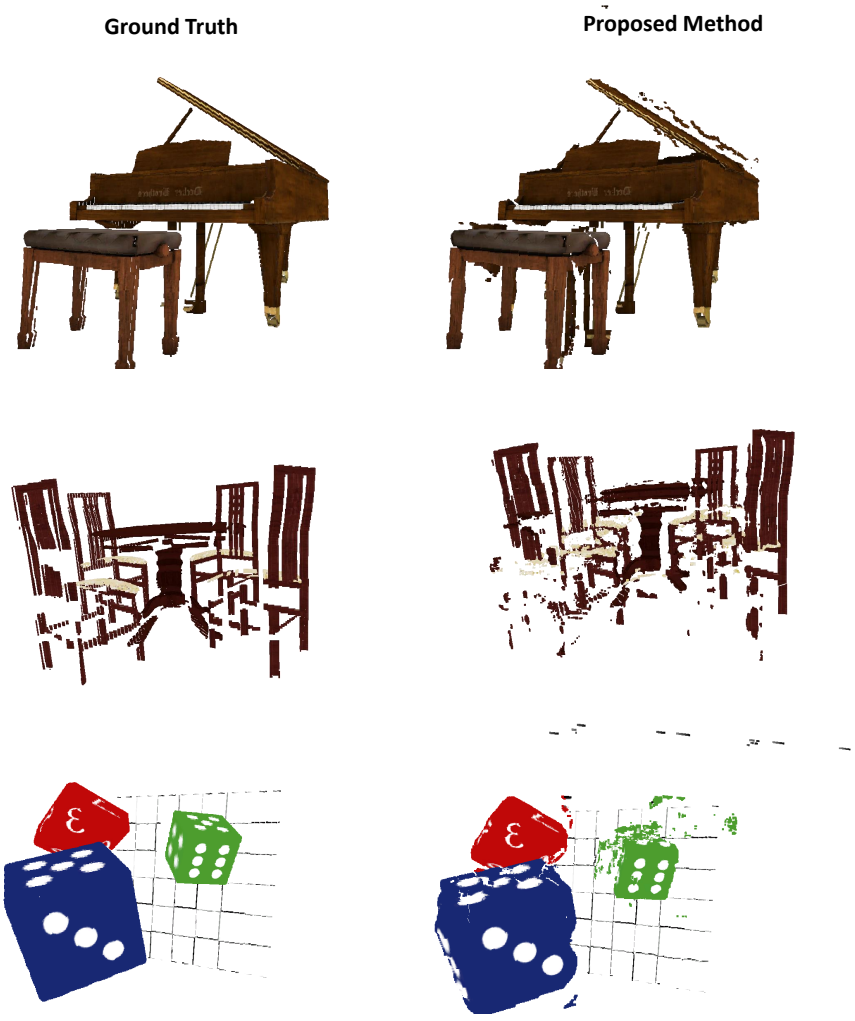

Figure 9: The estimated point cloud obtained using the CONT operator with the proposed metric on three separate scenes.

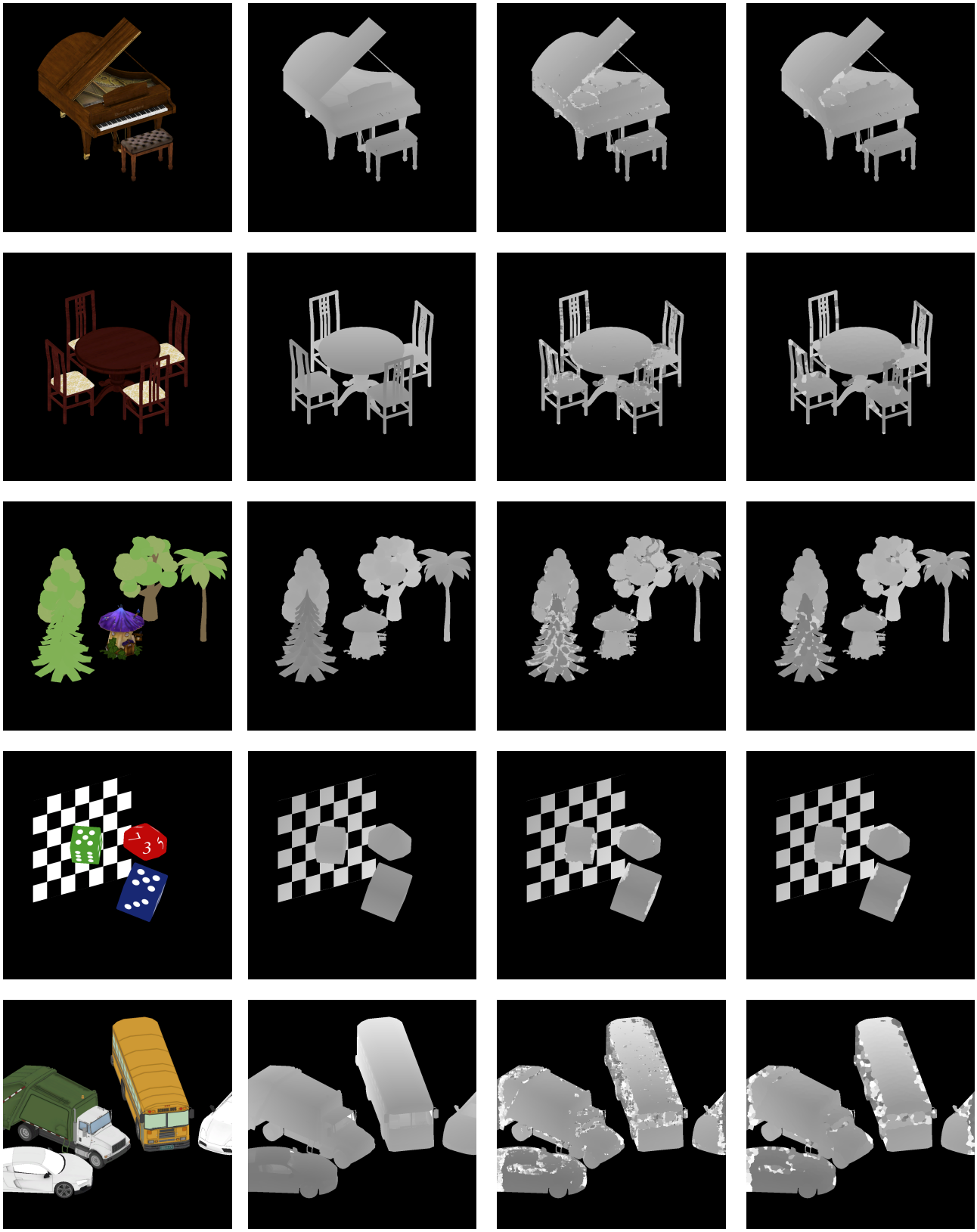

Figure 10: Each row contains a sample taken from our datasets, from left to right, the ground truth all in-focus image, the ground truth depth map, the initial depth map estimate using the GLVA operator with the proposed metric, and the refined estimate using the voting system.

[2] Joseph W Goodman. Introduction to fourier optics. *Introduction to Fourier optics, 3rd ed., by JW Goodman. Englewood, CO: Roberts & Co. Publishers, 2005*, 1, 2005.

[3] Frédéric Dufaux, Yafei Xing, Béatrice Pesquet-Popescu, and Peter Schelkens. Compression of digital holographic data: an overview. In *SPIE Optical Engineering + Applications*, 2015.

[4] Raees Muhamad, David Blinder, Athanasia Symeonidou, Tobias Birnbaum, Osamu Watanabe, Colas Schretter, and Peter Schelkens. Exact global motion compensation for holographic video compression. *Applied Optics*, 58:G204, 11 2019.

[5] Tobias Birnbaum, David Blinder, Raees Muhamad, Colas Schretter, Athanasia Symeonidou, and Peter Schelkens. Multi-object scene segmentation for local motion compensation in digital holography. *Optics Express*, 28, 03 2020.

[6] P. Grossmann. Depth from focus. *Pattern Recogn. Lett.*, 5(1):63–69, jan 1987.

[7] John Barron, David Fleet, and S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12:43–77, 02 1994.

[8] Said Pertuz, Domenec Puig, and Miguel García. Analysis of focus measure operators in shape-from-focus. *Pattern Recognition*, 46, 11 2012.

[9] D.L. Donoho, M. Elad, and V.N. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on Information Theory*, 52(1):6–18, 2006.

[10] Gagan Rath, Christine Guillemot, and Jean-Jacques Fuchs. Sparse approximations for joint source-channel coding. In *2008 IEEE 10th Workshop on Multimedia Signal Processing*, pages 481–485, 2008.

[11] Peng Xu, Yin Tian, Huafu Chen, and Dezhong Yao. Lp norm iterative sparse solution for eeg source localization. *IEEE Transactions on Biomedical Engineering*, 54:400–409, 2007.

[12] Niall Hurley and Scott Rickard. Comparing measures of sparsity. *IEEE Transactions on Information Theory*, 55(10):4723–4741, 2009.

[13] Xin Fan, John Healy, and Bryan Hennelly. Sparsity metrics for autofocus in digital holographic microscopy. page 989619, 04 2016.

[14] Xin Fan, John J. Healy, and Bryan M. Hennelly. Investigation of sparsity metrics for autofocusing in digital holographic microscopy. *Optical Engineering*, 56:053112, May 2017.

[15] Jean Steinier, Yves Termonia, and J J Deltour. Smoothing and differentiation of data by simplified least square procedure. *Analytical chemistry*, 44 11:1906–9, 1972.

[16] Michael Liebling and Michael Unser. Autofocus for digital fresnel holograms by use of a fresnelet-sparsity criterion. *Journal of the Optical Society of America. A, Optics, image science, and vision*, 21:2424–30, 01 2005.

[17] P. Memmolo, Lisa Miccio, Melania Paturzo, Giuseppe Caprio, Giuseppe Coppola, Paolo Netti, and Pietro Ferraro. Recent advances in holographic 3d particle tracking. *Advances in Optics and Photonics*, 7:713, 12 2015.

[18] Miu Tamamitsu, Yibo Zhang, Hongda Wang, Yichen Wu, and Aydogan Ozcan. A robust holographic autofocusing criterion based on edge sparsity: comparison of gini index and tamura coefficient for holographic autofocusing based on the edge sparsity of the complex optical wavefront. In *BiOS*, 2018.

[19] M. Fatih Toy, Stéphane Richard, Jonas Kühn, Alfredo Franco-Obregón, Marcel Egli, and Christian Depeursinge. Enhanced robustness digital holographic microscopy for demanding environment of space biology. *Biomed. Opt. Express*, 3(2):313–326, Feb 2012.

[20] Frans C. A. Groen, Ian T. Young, and G Ligthart. A comparison of different focus functions for use in autofocus algorithms. *Cytometry*, 6 2:81–91, 1985.

[21] Lawrence M. Firestone, Kelly Ann Cook, Kaitlyn Culp, Niharika Talsania, and K. Preston. Comparison of autofocus methods for automated microscopy. *Cytometry*, 12 3:195–206, 1991.

[22] Jaehwan Jeon, Inhye Yoon, Jinhee Lee, and Joonki Paik. Robust focus measure for unsupervised autofocusing based on optimum discrete cosine transform coefficients. In *2011 IEEE International Conference on Consumer Electronics (ICCE)*, pages 193–194, 2011.

[23] Sang-Yong Lee, Yogendera Kumar, Ji-Man Cho, Sang-Won Lee, and Soo-Won Kim. Enhanced autofocus algorithm using robust focus measure and fuzzy reasoning. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(9):1237–1246, 2008.

[24] Xiangfen Zhang, Hongfeng Wu, and Yan Ma. A new auto-focus measure based on medium frequency discrete cosine transform filtering and discrete cosine transform. *Applied and Computational Harmonic Analysis*, 40:430–437, 2016.

[25] Akshay Sharma, Gyanendra Sheoran, Z.A. Jaffery, and Moinuddin. Improvement of signal-to-noise ratio in digital holography using wavelet transform. *Optics and Lasers in Engineering*, 46(1):42–47, 2008.

[26] S. Montresor, P. Y. Quehe, S. Verhaeghe, and P. Picard. Evaluation of denoising algorithms applied to the reduction of speckle in digital holography. In *2015 23rd European Signal Processing Conference (EUSIPCO)*, pages 2316–2320, 2015.

[27] Huaying Wang, Aili Qin, and Min Huang. Autofocus method for digital holographic reconstruction of microscopic object. In *2009 Symposium on Photonics and Optoelectronics*, pages 1–4, 2009.

[28] Elsa S. R. Fonseca, Paulo Torrão Fiadeiro, Manuela Pereira, and António M. G. Pinheiro. Comparative analysis of autofocus functions in digital in-line phase-shifting holography. *Applied optics*, 55 27:7663–74, 2016.

[29] Frank Dubois, Cédric Schockaert, Natacha Callens, and Catherine Yourassowsky. Focus plane detection criteria in digital holography microscopy by amplitude analysis. *Opt. Express*, 14(13):5895–5908, Jun 2006.

[30] Dieter Vollath. Automatic focusing by correlative methods. *Journal of Microscopy*, 147, 1987.

[31] Frank Dubois, Ahmed El Mallahi, Jérôme Dohet-Eraly, and Catherine Yourassowsky. Refocus criterion for both phase and amplitude objects in digital holographic microscopy. *Optics Letters*, 39, 08 2014.

[32] D. Vollath. The influence of the scene parameters and of noise on the behavior of automatic focusing algorithms. *Journal of Microscopy*, 151:133 – 146, 08 2011.

[33] Pietro Ferraro, Giuseppe Coppola, Sergio De Nicola, Andrea Finizio, and Giovanni Pierattini. Digital holographic microscope with automatic focus tracking by detection sample displacement in real time. *Optics letters*, 28:1257–9, 08 2003.

[34] Carlos Trujillo and Jorge Garcia-Sucerquia. Automatic method for focusing biological specimens in digital lensless holographic microscopy. *Optics letters*, 39:2569–2572, 05 2014.

[35] Carlos Trujillo and Jorge Garcia-Sucerquia. Comparative analysis of the modified enclosed energy metric for self-focusing holograms from digital lensless holographic microscopy. *Applied Optics*, 54:5102, 06 2015.

[36] Patrik Langehanenberg, Lyubomira Ivanova, Ingolf Bernhardt, Steffi Ketelhut, Angelika Vollmer, Dieter Dirksen, G. Georgiev, Gert von Bally, and Björn Kemper. Automated three-dimensional tracking of living cells by digital holographic microscopy. *Journal of biomedical optics*, 14 1:014018, 2009.

[37] Patrik Langehanenberg, Gert von Bally, and Björn Kemper. Autofocusing in digital holographic microscopy. *3D Research*, 2:1–11, 2011.

[38] Dapu Pi, Juan Liu, and Yongtian Wang. Review of computer-generated hologram algorithms for color dynamic holographic three-dimensional display. *Light, Science & Applications*, 11, 2022.

[39] Tobias Birnbaum, Ayyoub Ahar, Silvio Montrésor, Pascal Picart, Colas Schretter, and Peter Schelkens. Speckle denoising of computer-generated macroscopic holograms. In *Digital Holography and Three-Dimensional Imaging 2019*, page W3A.1. Optica Publishing Group, 2019.

[40] Antonin Gilles, Patrick Gioia, Rémi Cozot, and Luce Morin. Hybrid approach for fast occlusion processing in computer-generated hologram calculation. *Appl. Opt.*, 55(20):5459–5470, Jul 2016.

[41] Jan-Mark Geusebroek, Frans Cornelissen, Arnold Smeulders, and Hugo Geerts. Robust autofocusing in microscopy. *Cytometry*, 39:1 – 9, 01 2000.

[42] A Santos, Carlos Ortiz-de Solorzano, Juan Jose Vaquero, J Peña, Norberto Malpica, and Francisco Del Pozo Guerrero. Evaluation of autofocus functions in molecular cytogenetic analysis. *Journal of microscopy*, 188:264–72, 01 1998.

[43] Murali Subbarao, Tae Choi, and Arman Nikzad. Focusing techniques. *Journal of Optical Engineering*, 32:2824–2836, 1993.

[44] Harsh Nanda and Ross Cutler. Practical calibrations for a real-time digital omnidirectional camera. *Proceedings of CVPR, Technical Sketch*, 01 2001.

[45] Eric Krotkov and Jean-Paul Martin. Range from focus. *Proceedings. 1986 IEEE International Conference on Robotics and Automation*, 3:1093–1098, 1986.

[46] A.M. Eskicioglu and P.S. Fisher. Image quality measures and their performance. *IEEE Transactions on Communications*, 43(12):2959–2965, 1995.

[47] J.L. Pech-Pacheco, G. Cristobal, J. Chamorro-Martinez, and J. Fernandez-Valdivia. Diatom autofocusing in brightfield microscopy: a comparative study. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, volume 3, pages 314–317 vol.3, 2000.

[48] Rashid Minhas, Abdul Adeel Mohammed, Q. M. Jonathan Wu, and Maher Sid-Ahmed. 3d shape from focus and depth map computation using steerable filters. pages 573–583, 07 2009.

[49] S.K. Nayar and Y. Nakagawa. Shape from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(8):824–831, 1994.

[50] Andrea Thelen, Susanne Frey, Sven Hirsch, and Peter Hering. Improvements in shape-from-focus for holographic reconstructions with regard to focus operators, neighborhood-size, and height value interpolation. *IEEE Transactions on Image Processing*, 18(1):151–157, 2009.

[51] Ge Yang and B.J. Nelson. Wavelet-based autofocusing and unsupervised segmentation of microscopic images. In *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453)*, volume 3, pages 2143–2148 vol.3, 2003.

[52] Yan Zhang, De-Xiang Zheng, Jing-Ling Shen, and Cun-Lin Zhang. 3d locations of the object directly from in-line holograms using the gabor transform. *Proc SPIE*, 02 2005.

[53] Sang Yong Lee, Jae Tack Yoo, Yogendera Kumar, and Soo Won Kim. Reduced energy-ratio measure for robust autofocusing in digital camera. *IEEE Signal Processing Letters*, 16(2):133–136, 2009.