

Speed up of Fresnel transforms for digital holography using pre-computed chirp and GPU processing.

Nitesh Pandey*^a, Damien P. Kelly^a, Thomas J. Naughton^{a,b} and Bryan M. Hennelly^a

a)Department of Computer Science, National University of Ireland,
Maynooth, Ireland

b)University of Oulu, RFMedia Laboratory, Oulu Southern Institute,
Vierimaantie 5, 84100 Ylivieska, Finland

ABSTRACT

We show how the reconstruction of digital holograms can be speeded up on ordinary computers by precomputing the chirp factor in the Fresnel transform for a given detector array size. The speedup in time is shown for various hologram sizes. We also run the same algorithm on a Nvidia GPU. The speedup and the error introduced due to quantizing to different levels is investigated. Additionally a variance based autofocus algorithm is implemented using the quantized chirp reconstructions.

1. INTRODUCTION

Digital holography¹ is a fast growing field with many applications including Microscopy², Metrology³ and 3D information processing^{4,5} and display⁶. In digital holography, an interference pattern of coherent light reflected off a 3D object and a plane 'reference' wave is recorded by an image sensor like a CCD or CMOS array. This interference pattern, called the digital hologram (DH), captures the true 3D information of the illuminated part of the object. The image of the object can then be reconstructed from the digital hologram by numerically propagating the wavefield back to the plane of the object using the theory of diffraction⁷. Due to the nature of the interference, an out of focus twin image noise and a zero order noise appears in the reconstruction along with the real image. These can be removed using an off-axis recording geometry of recording holograms⁸ or by capturing a number of phase shifted holograms⁹. Since the reconstruction is done by propagating light numerically any known systematic aberrations in the optical setup can also be numerically compensated. Another advantage of digital holography is that the holograms can be transmitted to another place to be reconstructed remotely and studied^{10,11}. The resolution of the reconstructed object primarily depends on the sensor area size, the total number of pixels in the camera and the bit-depth of the camera¹². Modern CCD camera are in the megapixel range and this makes the DH memory size large (~4MB for a 2 megapixel, 16 bit camera). The corresponding reconstruction algorithms are time consuming and computationally intensive for real time application and for use in simple hardware on portable devices.

It is useful to reduce the complexity of reconstruction of holograms for applications like real time digital holographic microscopy of growing cell cultures and digital holography of fast changing events¹³. Recently the use of graphics cards for general purpose computing has become popular. These methods exploit the parallel many-core capability of the GPU to offer a significant speedup over CPUs. GPGPU (General-purpose computing on graphics processing units) has been used to show the speedup of reconstruction (up to 180x faster) of holograms on the GPU architecture¹⁴⁻¹⁶.

*npandey@cs.nuim.ie

These hardware approaches require extra hardware and programming effort. Another way to decrease processing time of the DHs is to reduce the complexity of the algorithm by reducing the memory usage of the variables. The quantization of the hologram and the result on the reconstructed images has been studied before. Quantization techniques have been applied to compress the digital holograms to reduce the memory size of large digital holograms.¹⁷⁻¹⁹. In Ref [20], the authors reduced the quantization bit-depth of the recovered phase from phase stepped interferograms and showed that reconstructed images are perceptible even at 1 bit of quantization (2 levels). In Ref [21], the authors studied the effect of quantization in phase shifting digital holography and it was shown that the use of 4 bits for hologram quantization is adequate for visual recognition of the reconstructed images. Display devices like LCDs (Liquid crystal displays) and SLMs (Spatial light modulators) have an addressing resolution of typically 8-12 bits. As imaging sensor technology grows towards larger CCDs/CMOS sensors and higher bit-depth cameras and digital holography finds more practical applications, such studies are very useful.

In this paper, we aim to i) reduce the numerical complexity of the reconstruction of the digital hologram on general computer architecture by precomputation and quantization of the quadratic phase factor ('chirp') used in the Fresnel transform. and ii) We investigate the impact of quantization in the numerical computation on the quality of the reconstructed digital holograms.

In section 2, we review the Fresnel transform algorithm and the experimental setup used for recording the digital holograms. In section 3, we show how some precomputations of the chirp can speed up the computation on CPUs and GPUs. In section 4, we subject the chirp to quantization to compress it and reduce the memory size and study the effect this has on quality of the reconstructed image intensity and finally in section 5 we apply these precomputed and compressed matrices in a variance based focus search algorithm in which many reconstructions are required at various distances. We note that the results in this paper can be applied to other numerical transforms.

2. RECONSTRUCTION USING FRESNEL TRANSFORM

The propagation of a wave field can be approximated by the Fresnel transform. This is an approximation of the Fresnel-Kirchhoff integral. Consider the Fresnel transform that describes the relationship between the wavefield at two planes $h(x,y)$ and $\Gamma(\xi,\eta)$ separated by a distance d

$$\Gamma(\xi,\eta) = \frac{ia}{\lambda d} \exp\left[-i\frac{\pi}{\lambda d}(\xi^2 + \eta^2)\right] \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y) \times \exp\left[\frac{-i\pi}{\lambda d}(x^2 + y^2)\right] \times \exp\left[\frac{i2\pi}{\lambda d}(x\xi + y\eta)\right] dx dy \quad (1)$$

where a is a constant, λ is the wavelength. In digital holography, we deal with discrete representations of this transform since the hologram is discretized at the CCD into $N \times M$ samples at intervals of Δx and Δy in the x and y directions. Δx and Δy being the pixel pitches of the sensor. Direct discretization of the Fresnel integral gives the following

$$\Gamma(m,n) = \exp\left[\frac{-i\pi}{\lambda d}(m^2 \Delta \xi^2 + n^2 \Delta \eta^2)\right] \times \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} h(k,l) \exp\left[\frac{-i\pi}{\lambda d}(k^2 \Delta x^2 + l^2 \Delta y^2)\right] \exp\left[i2\pi\left(\frac{km}{M} + \frac{ln}{N}\right)\right] \quad (2)$$

Here $\Gamma(\xi,\eta)$ is the matrix of $N \times M$ points which describes the complex amplitude distribution of the real image and $\Delta \xi$ and $\Delta \eta$ are the sampling periods in the reconstructed images. For a thorough investigation of the resulting numerical algorithm, and the range of d over which it is useful, the reader can consult Ref [22]. We assume we have available to us, a hologram, H , of $N \times M$ size of an object placed at a distance d from the camera. A practical implementation of the direct Fresnel transform to calculate the reconstructed image intensity is as follows.

- Calculate the grid of the X-Y plane where each element represents the X,Y co-ordinates of the pixel. $(m,n) = (m\Delta x, n\Delta y)$ for $m=-M/2$ to $M/2-1$ and $n=-N/2$ to $N/2-1$
- Calculate the chirp $C = \cos\left[\frac{\pi}{\lambda d}\left((m)^2 + (n)^2\right)\right] - i \sin\left[\frac{\pi}{\lambda d}\left((m)^2 + (n)^2\right)\right]$ at all the points on the grid.
- Perform element to element multiplication of C and H , $P=C.H$
- Perform 2D Discrete Fourier transform (DFT) on P . This is fastest using the FFT algorithm (fast Fourier transform)²¹, $F=2DFFT(P)$.
- The reconstructed image intensity is given by $|F|^2$

We focus our attention to the intensity of the reconstructed image and thus the second phase factor is not calculated here.

The camera pixel pitch is Δx and Δy in x and y directions. H is quantized according to the quantization level of the camera. And C is calculated in the computer with user defined precision. The DFT is calculated using the FFT²³ algorithm and this makes the calculation much faster ($O(k \log k)$ where k is the size of the matrix).

1.1 Experimental setup

The experimental setup used to capture our digital holograms for this study is shown in Fig.1.

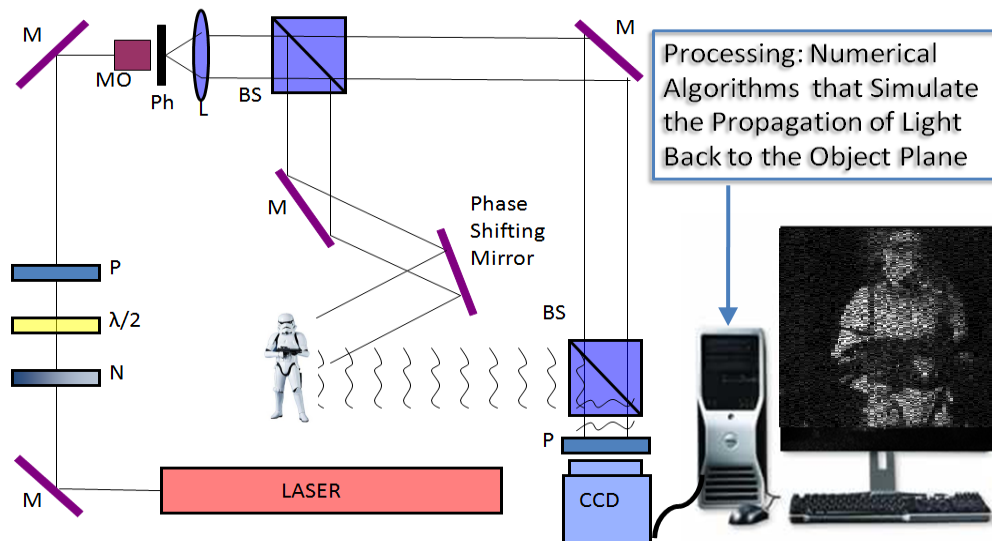


Fig.1 Experimental setup

We use a Mach-Zehnder interferometer to record inline digital holograms of macroscopic objects both in the reflection and transmission mode (for transmissive objects). Our object for this study is a 5cm long stormtrooper

figurine. We use a 12-bit camera (AVT-Dolphin 145-B) for recording and our setup is capable of performing phase shifting digital holography. A mirror is attached to a piezo actuator (Piezosystem Jena, PZ38CAP) that can be moved in a perpendicular direction to its plane in steps of 10nm by applying a voltage to the piezo controller (NV401CLE by Piezोजना) and phase shifting in the object beam path can be realized.

3. PRE-COMPUTATION OF CHIRP MATRICES

In order to speed up the Fresnel reconstruction, we note that the chirp is independent of the input hologram field. The second chirp and the constant factors are also independent of the input. These matrices are composed of the exponential phase factors which are time-consuming to compute. If we precalculate a range of these matrices for different distances and load the corresponding data from memory during reconstruction, we can save $2NxM$ computations of sines and cosines and thus reduce the time for reconstruction. This is advantageous for real time high resolution digital holographic systems running on nominal hardware. The computation reduces to two vector multiplications and the FFT algorithm. Table 1 shows the speedup on different hologram sizes achieved by precomputing.

Hologram size	No loading (time in seconds)	With loading (time in seconds)	Speedup
100x100	0.0613	0.04731	1.30
200x200	0.1027	0.07023	1.46
500x500	0.4390	0.26449	1.66
1000x1000	1.7360	0.99297	1.74
2000x2000	7.1904	4.22684	1.70

Table 1: Speedup on CPU

We perform the reconstructions using preloading on a NVIDIA Geforce 8800 GTX graphics card (with 768 Mb of onboard memory) on a AMD Athlon 64 X2, 2.31Ghz, 2 GB RAM. We use the Jacket²⁴ engine for Matlab to run our code on the GPU. The results are shown in Table 2. A large number of these chirp matrices which cover a large distance can be stored permanently in memory and can be accessed by the program whenever the reconstruction is demanded for a particular distance.

Hologram size	No loading (time in seconds)	With loading (time in seconds)	Speedup
100x100	0.0629	0.0555	1.13
200x200	0.0646	0.0522	1.23
500x500	0.0854	0.0672	1.27
1000x1000	0.1819	0.0775	2.34
2000x2000	0.5336	0.2514	2.12

Table 2: Speedup on GPU

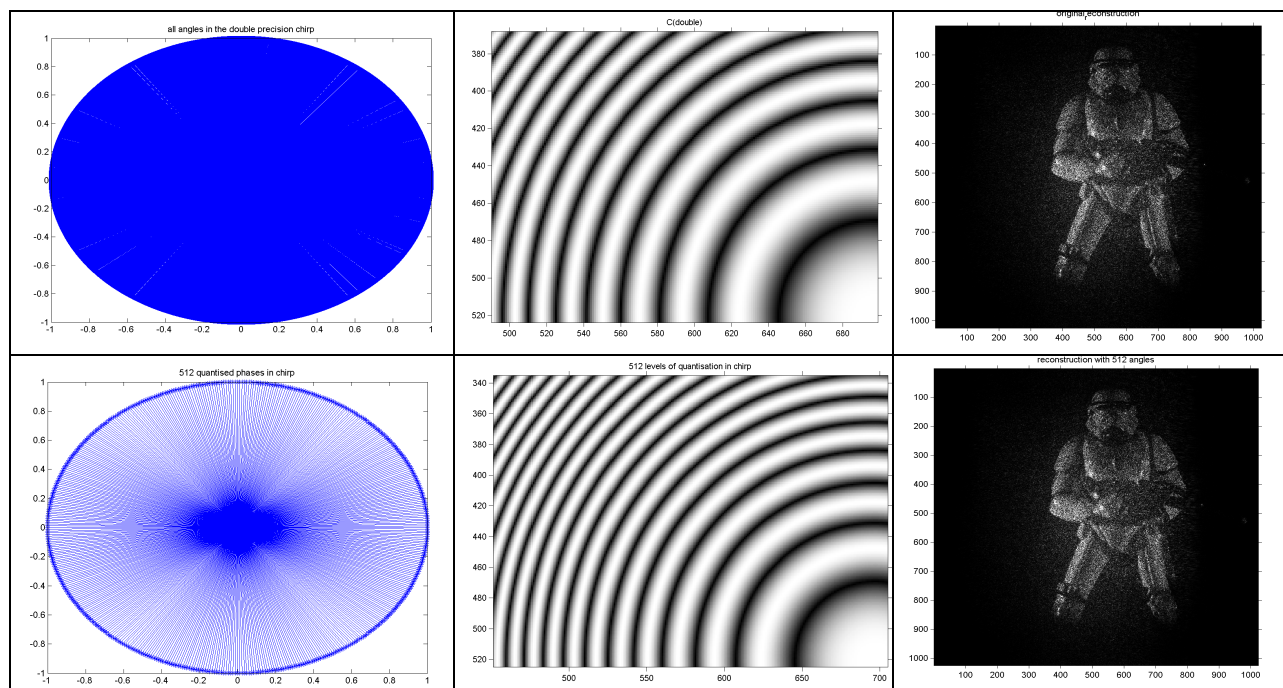
In section III, we study how these matrices can be quantized with reduced number of unique elements and the effect on reconstructed visual quality. Quantization reduces the compressed file size and the memory storage of the chirp. They also contain lesser number of sines and cosines are thus faster for multiplication operations employing lookup tables in FPGAs.

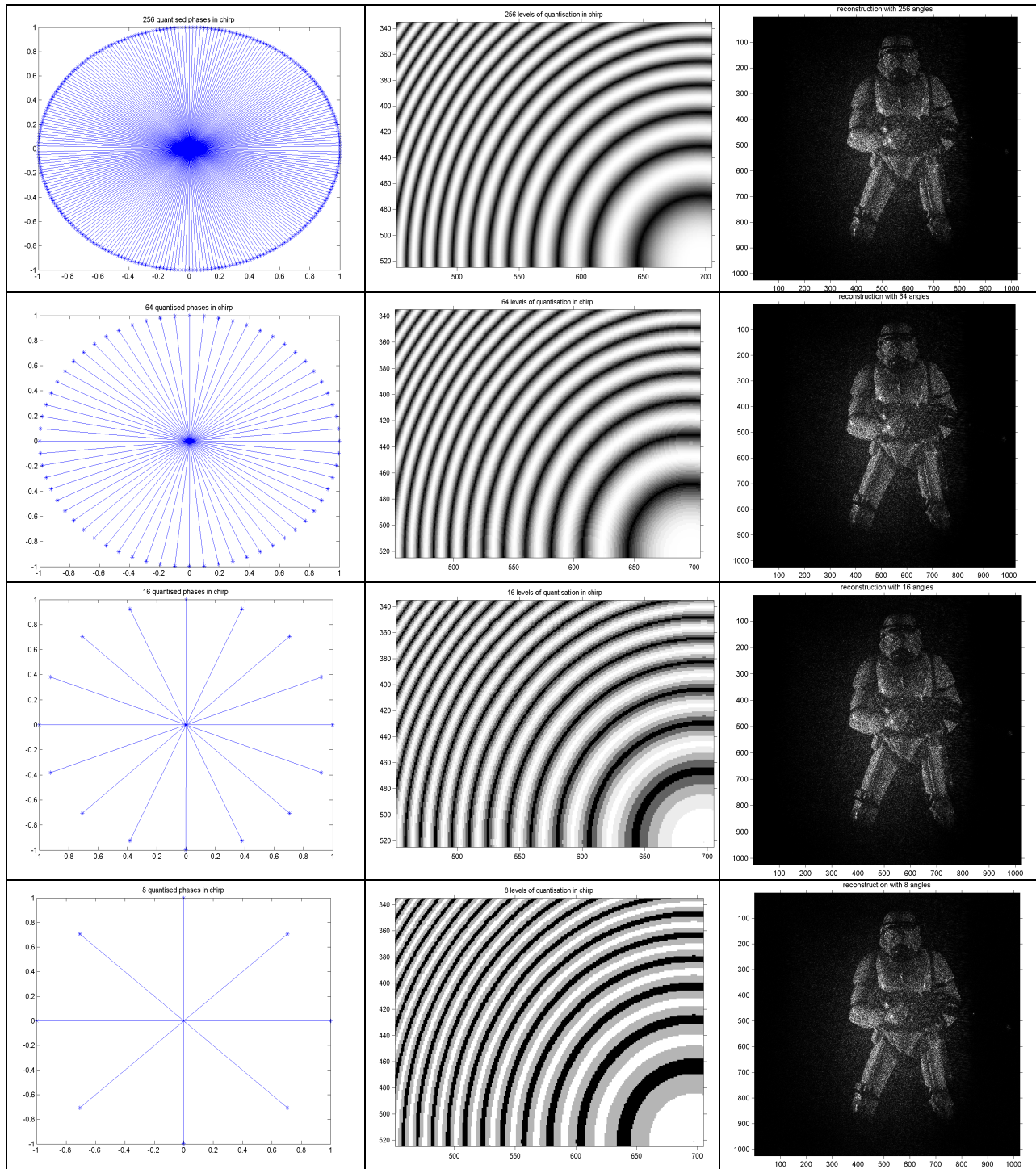
4. QUANTIZATION OF THE CHIRP MATRICES

In this section we reduce the quantization of the chirp phase factor and study the effect on the reconstruction. We do so by restricting the total number of angles used to calculate the sines and cosines to by the chirp to a fixed number. Thus the chirp matrix can be represented using fewer bits. This reduces the total filesize of the data file and thus allows the possibility to load a large number of matrices in the program memory space. The quantized chirp can further be subjected to a lossless or lossy compression to reduce the file size further of the precomputed matrices. The compressed files will then have to decompressed during execution of the Fresnel integral which takes much less time than the full computation.

Consider the algorithm described in section 1 to compute the chirp phase factor for a hologram of an object placed at a distance of 285mm of size 1024x1024, with a wavelength of 785nm and with a pixel pitch of 6.45 μ m in both directions; Using the direct reconstruction, the computed chirp matrix has a total of 104265 unique angles at which the sines and cosines have to be calculated. If we 'bin' the angles into a smaller number of levels, we will not only reduce the number of computations required to generate the full chirp matrix but also reduce the total number of bits per pixel required to 'represent' this matrix. (For example using 512 angles we can represent using 9 bits per pixel).

A digital hologram of the object (a storm trooper figurine), placed at a distance of 28 cm from the CCD is recorded with our setup and the numerical reconstruction is performed using different chirp matrices. A 2D highpass filtering operation is performed on the hologram to remove the zero-order term. Table 3 below shows the quantized values of the angles and the resulting phase distribution and the reconstructed image.





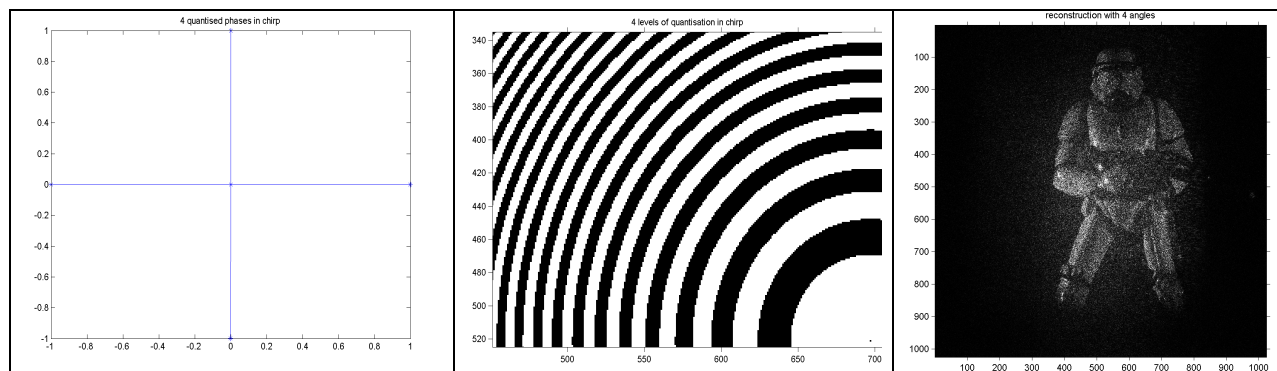


Table 3

First Column: Number of angles representing the chirp (top to bottom) full, 512, 256, 64,16,8 and 4 angles. **Second column:** Image of the real part of the chirp showing the subsequent quantization. **Third column:** Intensity of the reconstructed image.

As it can be seen limiting the number of angles to just 4 gives perceptually satisfactory images.

To quantify the difference, we calculate the NRMS error between the reconstructed intensities and the original intensity by using the following equation where I_{fc} is the reconstructed intensity using a complete chirp calculated at all the required angles and I_{qc} is the reconstructed intensity of the quantized chirp

$$\text{NRMSSE} = \frac{\sqrt{\sum \sum |I_{qc} - I_{fc}|^2}}{\sqrt{\sum \sum |I_{fc}|^2}} \quad (3)$$

Fig.1 shows the NRMS error obtained from the reconstructions. The error is at its highest for 4 angles in the chirp (0.241) and decreases with increasing the number of angles. It can be debated whether the NRMS error metric is a reliable metric when comparing images from a visual perceptive point of view. It has been shown that the MSE is not always good metric for such studies²⁵ and variance of the image is a better metric. Additionally the fact that popular video display units have 8 bits of quantization for each color, and that most high end SLMs have an addressing mode of 8 bits, means that the reconstructed images will have to be subjected to a rescaling and quantization during display. In Fig.2 we show the percentage variance of the reconstructed image versus the number of bits required for the angle in the chirp. The variance reduces to less than 30% for the 4 level representations.

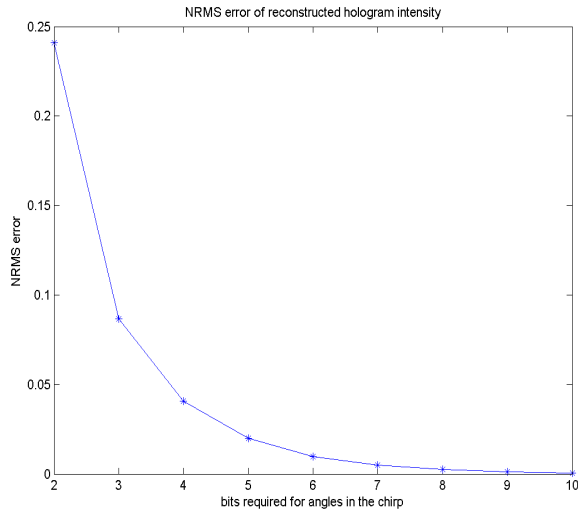


Fig.1 NRMS error for the reconstructed images

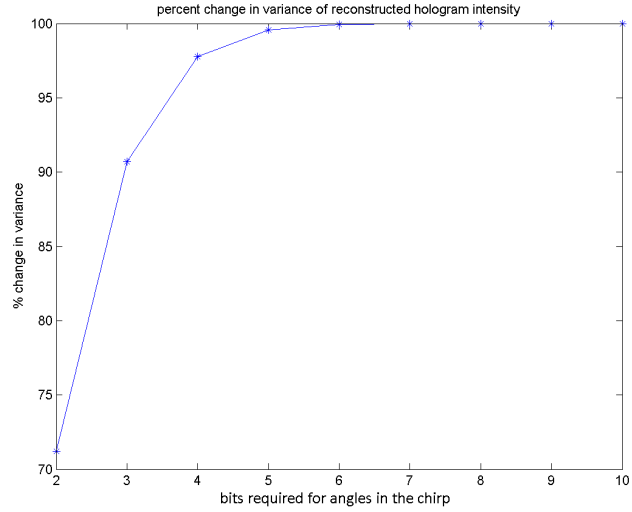


Fig.2. Change in variance in % with respect to original.

Because of the decreased randomness of the data, the quantized chirp matrices can be compressed effectively using lossless compression techniques. We use Matlab's built-in compression scheme used in saving MAT format files to compress and store the chirp matrices. Matlab's compression scheme is based on gzip²⁶⁻²⁸ (DEFLATE algorithm). Fig.3 (a) and (b) show the reduction in filesize of the compressed files. The compressed file size of the un-quantized chirp matrix is 10MB. The file size of the quantized chirp and the average time taken to load and decompress the file decreases along with the quantization level.

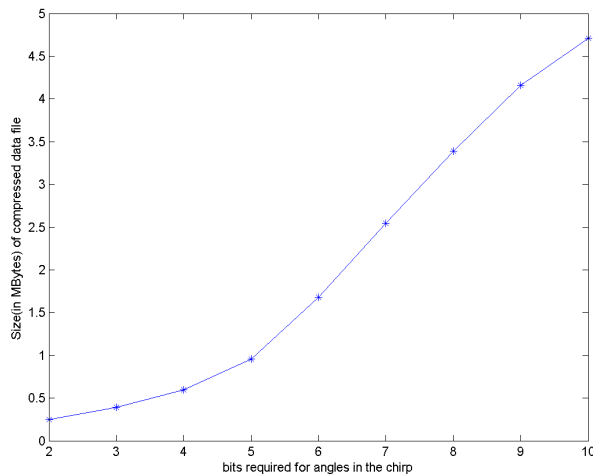
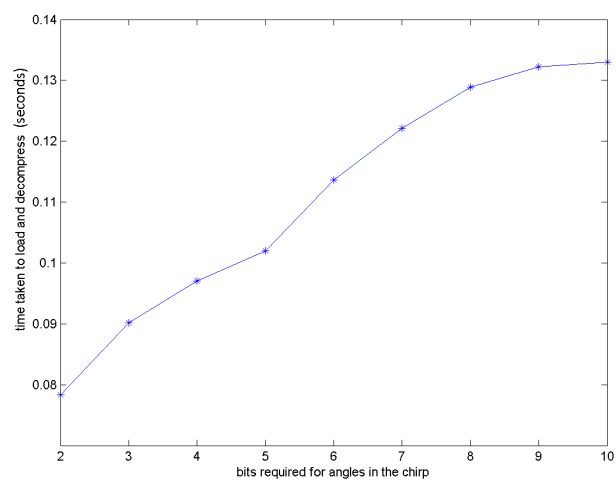


Fig. 3(a) Size of compressed chirp



(b) time taken to load and decompress the chirp

5. VARIANCE BASED FOCUS CALCULATIONS USING QUANTIZED CHIRP

In previous sections, we have shown that even with calculating the chirp function using just 4 angles, the reconstructed object can be perceptible. In this section we investigate the application of our faster Fresnel algorithms to a variance based focus search algorithm that identifies the correct focusing distance. A focused image of an object

has sharp features and the focusing distance can be found by using variance as a measure of the sharpness of the image^{29, 30}. A section of the reconstruction plane was taken and the variance of the pixel values of that section measured at different reconstruction distances. The variance was calculated for the chosen section of the reconstruction plane. The focusing distance is 285mm as checked manually. The variance of a reconstructed object in general varies smoothly as a function of reconstruction distance. The jitter introduced due to the quantized chirp can be removed by a simple mean filter. Similarly this could also be applied for 3D object recognition in digital holographic reconstruction⁴.

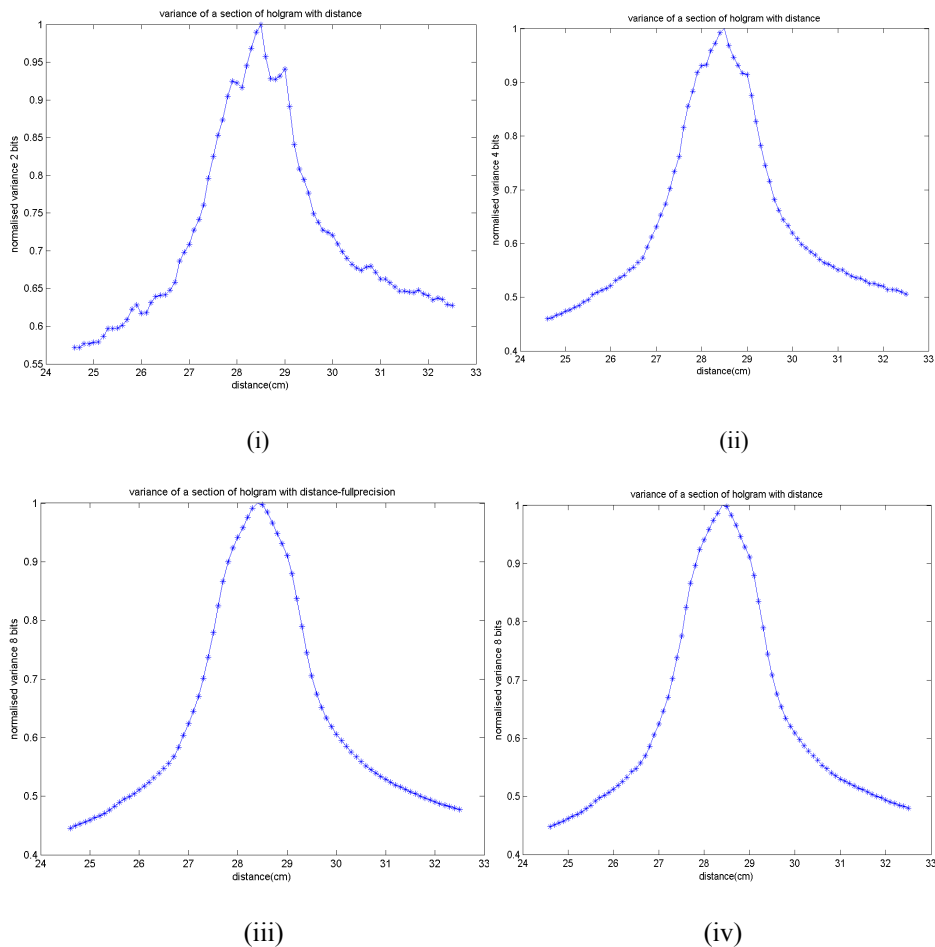


Fig 5. Variance of a section of the hologram as a function of distance of reconstruction. The variance is highest when the object is in focus. (i) 4 level chirp, (ii) 16 level chirp, (iii) 256 level chirp and (iv) full chirp

Fig.5 shows the variance vs. distance plot at different levels of quantization. The correct distance has been unambiguously recovered using a 2 bit representation of the chirp.

6. CONCLUSION

We have shown the benefits of using a table of precomputed chirp phase factor matrix on the speed of digital hologram reconstruction on normal CPUs and GPUs. The speedup improves with larger matrices and occurs due to the fact that loading large data from memory takes very little time compared to calculation 'on the fly'. We have

reduced the quantization of the precomputed phase factors in order for it to be compressed effectively and have shown the NRMSE and variance of the reconstructed image intensity. Additionally we have also shown how this highly quantized chirp can be used in focus search algorithm where a large number of reconstructions at different distances over the search range are required. In this paper we have limited our investigation to the direct method of computing the intensity of the Fresnel transforms which is made up of one chirp multiplication and one FFT. Other methods of reconstruction also require calculation of chirp data which is independent of the input hologram. The slightly more accurate method based on the Fresnel-Kirchoff transformation requires FFT followed by multiplication by a chirp like function followed by a second FFT. Since this chirp like function (which reduces to a chirp in the paraxial Fresnel approximation) is independent of the input hologram field, its precalculation will also result in considerable time savings. These reduced complexity matrices can give perceptually satisfactory reconstruction of 3D objects embedded in holograms and autofocus algorithm based on variance gives the correct focusing distance.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement no. 216105 and also Science Foundation Ireland under the National Development Plan.

REFERENCES

- [1] Schnars, U and Juptner, W, "Digital recording and numerical reconstruction of holograms," *Meas. Sci. Technol.* 13, 85–101 (2002).
- [2] Marquet, P., Rappaz, B., Magistretti, P.J., Cuche, E., Emery, Y., Colomb, T. and Depeursinge, C., "Digital holographic microscopy: a noninvasive contrast imaging technique allowing quantitative visualization of living cells with subwavelength axial accuracy," *Opt. Lett.* 30, 468-470 (2005).
- [3] Wagner, C., Osten, W. and Seebacher, S., "Direct shape measurement by digital wavefront reconstruction and multiwavelength contouring," *Opt. Eng.* 39, 79-85 (2000).
- [4] Javidi, B. and Tajahuerce, E., "Three-dimensional object recognition by use of digital holography," *Opt. Lett.* 25, 28-30 (2000).
- [5] Frauel, Y., Naughton, T.J., Matoba, O., Tajahuerce, E. and Javidi, B., "Three-dimensional imaging and processing using computational holographic imaging," *Proceedings of the IEEE*, 94(3), 636-653 (2006).
- [6] Poon, T.C. (Ed.), [Digital Holography and Three-Dimensional Display Principles and Applications], 293-326 Springer (2006).
- [7] Goodman, J.W., [Introduction to Fourier Optics], 63-88, McGraw-Hill, San Francisco, California (1968).
- [8] Leith, E. and Upatnieks, J., "Microscopy by wavefront reconstruction," *J. Opt. Soc. Am.*, 55, 569–570 (1965).
- [9] Yamaguchi, Y. and Zhang, T., "Phase-shifting digital holography," *Opt. Lett.*, 22, 1268-1270 (1997).
- [10] Naughton, T.J., Mc Donald, J.B. and Javidi, B., "Efficient compression of Fresnel fields for Internet transmission of three-dimensional images," *Applied Optics*, 42(23), 4758-4764 (2003).
- [11] Baumbach, T., Osten, W., Kopylow, C.V. and Jüptner, W. "Remote metrology by comparative digital holography," *Appl. Opt.* 45(5), 925–934 (2006).
- [12] Kelly, D.P., Hennelly, B.M., McElhinney, C.P. and Naughton, T.J, "A practical guide to digital holography and generalized sampling," *Proc. SPIE*, Vol. 7072, 707215 (2008).
- [13] Takaki, Y and Ohzu, H., "Fast numerical reconstruction technique for high-resolution hybrid holographic microscopy," *Appl. Opt.* 38(11), 2204-2211 (1999).

- [14] Ahrenberg, L., Page, A.J., Hennelly, B.M., McDonald, J.B. and Naughton, T.J., "Using Commodity Graphics Hardware for Real-Time Digital Hologram View-Reconstruction," *J. Display Technol*, 5(4), 111-119 (2009).
- [15] Ahrenberg, L., Benzie, P.Magnor, M. and Watson, J., "Computer generated holography using parallel commodity graphics hardware," *Opt. Express* 14(17), 7636-7641 (2006).
- [16] Shimobaba, T., Sato Y., Miura, J., Takenouchi, M. and Ito, T., "Real-time digital holographic microscopy using the graphic processing unit," *Opt. Express* 16(16), 11776-11781 (2008).
- [17] Shortt, A.E, Naughton, T.J and Javidi, B., "A companding approach for nonuniform quantization of digital holograms of three-dimensional objects," *Optics Express*, 14(12), 5129-5134 (2006).
- [18] Naughton, T.J., Frauel, Y. and Javidi, B. and Tajahuerce, E. "Compression of digital holograms for three-dimensional object reconstruction and recognition," *Applied Optics*, 41(20), 4124-4132 (2002).
- [19] Darakis, E., Naughton, T.J., and Soraghan, J.J., "Compression defects in different reconstructions from phase-shifting digital holographic data," *Applied Optics*, 46(21), 4579-4586 (2007).
- [20] Matoba, O., Naughton, T.J., Frauel, Y., Bertaux, N. and Javidi, B., "Real-time three-dimensional object reconstruction by use of a phase-encoded digital hologram," *Applied Optics*, 41(29), 6187-6192 (2002).
- [21] Mills, G.A. and Yamaguchi, I., "Effects of quantization in phase-shifting digital holography," *Appl. Opt.* 44(7), 1216-1225 (2005).
- [22] Mas, D., Garcia, J., Ferriera, C. and Bernardo, L.M., "Fast algorithms for free space diffraction patterns calculation," *Opt. Comm*, 164 (4-6), 233-245 (1999).
- [23] Cooley, J.W., and Tukey, J.W, "An algorithm for the machine calculation of complex Fourier series," *Math. Comput.*, 19(90), 297-301 (1965).
- [24] Accelerereyes, www.accelereyes.com
- [25] Buckley, E., "Holographic Laser Projection Technology," *SID 08 Digest*, 70.2, 1074-1079 (2008).
- [26] Mathworks, http://www.mathworks.com/access/helpdesk/help/pdf_doc/matlab/matfile_format.pdf
- [27] Ziv, J. and Lempel, A. "A universal algorithm for sequential data compression," *IEEE Trans.*, IT-23(3), 337-343 (1977).
- [28] gzip, www.gzip.org/
- [29] Groen, F.C., Young, I.T., and Ligthart, G., "A comparison of different focus functions for use in autofocus algorithms," *Cytometry*, 6(2), 81-91 (1985).
- [30] McElhinney, C.P., McDonald, J.B., 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(10), 1229-1231 (2007).