Axial sub-nanometer accuracy in digital holographic microscopy

J Kühn¹, F Charrière¹, T Colomb², E Cuche³, F Montfort³, Y Emery³, P Marquet² and C Depeursinge¹

¹ Ecole Polytechnique Fédérale de Lausanne (EPFL), Imaging and Applied Optics Institute, CH-1015 Lausanne, Switzerland
² Centre de Neurosciences Psychiatriques, Département de psychiatrie DP-CHUV, Site de Cery, CH-1008 Prilly-Lausanne, Switzerland
³ Lycée Tec SA, PSE-A, CH-1015 Lausanne, Switzerland

E-mail: jonas.kuehn@a3.epfl.ch

Received 31 August 2007, in final form 24 January 2008
Published 23 May 2008
Online at stacks.iop.org/MST/19/074007

Abstract
We present state-of-the-art dual-wavelength digital holographic microscopy (DHM) measurement on a calibrated 8.9 nm high chromium thin step sample and demonstrate sub-nanometer axial accuracy. By using a modified DHM reference calibrated hologram (RCH) reconstruction method, a temporal averaging procedure and a specific dual-wavelength DHM arrangement, it is shown that specimen topography can be measured with an accuracy, defined as the axial standard deviation, reduced to at least 0.9 nm. Indeed for the first time to the best of our knowledge, it is reported that averaging each of the two wavefronts recorded with real-time dual-wavelength DHM can provide up to 30% spatial noise reduction for the given configuration. Moreover, the presented experimental configuration achieves a temporal stability below 0.8 nm, thus paving the way to Angström range for dual-wavelength DHM.

Keywords: digital holography, phase imaging, optical metrology, non-contact, real-time

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Digital holographic microscopy (DHM) is an imaging technique that allows measuring quantitatively the wavefront transmitted through or reflected by a specimen seen through a microscope objective (MO) [1, 2]. A hologram, composed by the interference of the wave coming from the object with a reference wave, is recorded with a digital camera and then numerically processed to extract both the wavefront amplitude and phase information. Thanks to its interferometric nature, DHM provides phase images with sub-degree accuracy, corresponding to a nanometric accuracy along the optical axis, revealing detailed information about the specimen.

DHM has proven its efficiency on numerous applications in metrology, going from micro-optics testing [3–6], MEMS-MOEMS characterization [7–10], microstructures investigation [11] or even roughness measurements [12]. The applicability of DHM to metrology is facilitated thanks to the remarkably high measurement stability and robustness of DHM, which results from the off-axis configuration allowing all the necessary information to be recorded with a single hologram acquisition of very short duration (down to few µs). Compared to phase-shifting interferometric techniques, DHM offers similar performance in terms of resolution, precision, repeatability and field of view, but can be considered as an attractive solution presenting specific and unique advantages over other methods: the acquisition rate is higher, because a complete description of the complex wavefront is obtained out of a single hologram, and sensitivity to external vibrations is reduced since the acquisition time can be reduced to a few tens of microseconds. Moreover, the vertical calibration only depends on the wavelength, which ensures an accuracy that is intrinsically not limited by the precision of the control of moving parts, such as piezoelectric transducers or mechanical stages. Finally, DHM permits so-called ‘digital focusing’ by treating the complex wavefront as a single entity, thus extending virtually the depth of field.
Different averaging techniques are treated in the present paper and topographic measurements are presented, obtained by a dual-wavelength DHM setup on a 8.9 nm calibrated step height. At first, single-shot spatial standard deviation measurements on assumed-flat surfaces are obtained with a specific reference hologram reconstruction algorithm (see below), this for each individual wavelength. This reference hologram is acquired on a flat sample region due to FOV constraints, but it should ideally be averaged on different places across the specimen to completely remove object-dependent contributions (e.g. roughness and non-planarity). As a matter of fact, the following results contain an object-dependent ‘bias’ but the measurements should be interpreted in terms of noise reduction, not absolute values converging to zero. Second, the effect of temporal averaging on the reduction of the shot-noise contribution [13] is illustrated by considering again the spatial standard deviation values. Third, the improvement of axial accuracy by dual-wavelength DHM is demonstrated by averaging both wavelength topographic maps, enabling sub-nanometer accuracy. This dual wavelength averaging takes advantage of non-correlated noise between the two wavelengths used in the setup. Finally, the temporal stability of the present DHM configurations is evaluated for a 15 s sequence, by the evaluation of the temporal standard deviation pixel by pixel. In metrology this last parameter is often directly referred to as the precision, as it is usually assumed that a calibration reference image can be subtracted. Spatial averaging will also be considered in the temporal stability study, down-sampling the reconstructed phase image to a pixel size corresponding to the Airy disc defined by the microscope objective.

2. Experimental configuration

The DHM setup used in the present study allows imaging simultaneously at two different wavelengths. In previous works, this configuration was used to allow dual-wavelength DHM measurement with a single acquisition, in order to extend the phase dynamic range in the micrometer range for configurations where unwrapping procedures fail to solve the phase ambiguity [14]. While the dual-wavelength approach has already been intensively applied in digital holography in a sequential way for nearly a decade [24–28], the method of [14] enables for the first time true real-time dual-wavelength imaging. Practically, this is achieved by subtracting two different wavelength phase maps acquired during the same acquisition time to retrieve a so-called beat-wavelength phase image, the beat frequency being the frequency difference between both lasers, typically in the 5–20 µm range. As explained above, here another application of the dual-wavelength operation mode is exposed, taking advantage of the non-correlation between these two available wavefronts. The optical arrangement is depicted in figure 1, with the laser sources consisting of two temperature-stabilized semiconductor laser diodes at λ₁ = 657 nm and λ₂ = 680 nm. The diodes were tested for wavelength stability with a wavemeter and showed a wavelength deviation smaller than 10 pm over 8 h. The diodes also emit a low coherence (coherence length about 0.3 mm) linearly polarized light.

The principle of the setup is to separate each wavelength beam pair in different reference arms, while aligning and combining them in an achromatic object arm. After reflection on the sample, both collinear object wavefronts are collected by the MO (infinity-corrected), and the object images are formed by the tube lens (focal 150 mm) about 50 mm behind
the CCD plane. A low magnification $\times 3$ MO, with a NA of 0.1 (corresponding to 4.4 $\mu$m lateral resolution), is used in the present case to achieve large field of view (FOV), $1 \times 1$ mm$^2$ with 512 $\times$ 512 pixels, and a large depth of field of more than 50 $\mu$m. The CCD camera is a standard 8 bits black and white CCD camera with 6.45 $\mu$m pixel size. Each reference arm comprises a delay line (DL) adjusted to match the optical path length of the corresponding object beam, in order to create an interference on the CCD for both wavelengths. By tilting the pair of mirrors M1 and M2 for the first wavelength reference beam, respectively M3 and M4 for the second one, one can finely tune each $k$-vector incident upon the CCD camera. In other words, each wavelength interferogram fringes can be independently tuned both in spatial frequency and orientation.

The actual configuration is the one depicted in the inset of figure 1(a), with orthogonal spatial frequencies for each wavelength interferogram, because it ensures an optimum repartition of the interference terms in the hologram Fourier domain, minimizing overlaps. The orthogonal repartition of the fringes spatial frequencies in figure 1(c) enables us to filter (select) separately the spatial frequencies of each wavelength in the Fourier spectrum of the hologram [14]. This experimental configuration for recording holograms is very similar to the one in [15] and permits us to encode both wavefronts with only a single acquisition thanks to orthogonal carrier frequencies. A transverse resolution of 4.4 $\mu$m has been measured with a USAF 1950 resolution test target and the system is diffraction limited according to the NA of the MO.

The investigated sample is a thin step height standard from VLSI Standards Incorporated realized by chromium deposition on a quartz wafer. It consists in a precisely etched positive step 100 $\mu$m wide and 750 $\mu$m long, with a step height of 8.9 nm within a uncertainty of $\pm 0.5$ nm, thus permitting measuring resolution-levels down to the nanometer. It is also important to note that this test-target has been certified by the National Institute of Standards and Technology (NIST) to ensure correct measurement comparison with theoretical values. 3D representations of this sample obtained from DHM reconstructed phase images presented later on in this paper are depicted in figure 2. All following measurements in this paper are related to the area of interest corresponding to figure 2(b). Although it is ideally preferable to scan the sample at multiple places for calibration, the specimen does not present a large area without structures for the FOV of the $\times 3$ MO, thus limiting reference hologram (see below) acquisitions to a single reference region. This way the spatial noise data presented here contain some sample-induced ‘bias’ corresponding to the sub-nanometer roughness or non-planarity.

3. Hologram reconstruction

Let us take the example of a hologram acquired by a digital camera, with two collinear object beams $O_1$ and $O_2$ at two different wavelengths $\lambda_1$ and $\lambda_2$ that interfere with two reference beams $R_1$ and $R_2$, emitted by the same pair of laser sources, in an off-axis configuration (slight angle between object and reference beams). The intensity pattern, which results from an incoherent addition of both interferograms at $\lambda_1$ and $\lambda_2$, can be expressed as

$$I_H(x,y) = \left| R_1 \right|^2 + \left| O_1 \right|^2 + \left| R_2 \right|^2 + \left| O_2 \right|^2 + R_1 O_1^* + R_2 O_2^*$$

with $I_H$ being the hologram intensity, $x, y$ the coordinates in the camera plane and * denoting the complex conjugate.

The first four terms in equation (1) correspond to the zero order of diffraction, and can be easily filtered in the Fourier domain [16, 17]. The last four terms correspond to the interference of the object wavefronts $O_i$ (the virtual images), or their conjugate $O_i^*$ (the real images), with the reference waves. These interferences appear as fringes with specific carrier spatial frequencies on the hologram. With $O_i$ parallel to the optical axis, these carrier frequencies are dependent on the $k$-vectors of $R_i$ and $R_2$. Considering different incident angles for the two reference waves, especially the configuration where their $k$-vector projections on the CCD plane are orthogonal, each interference term occupies different positions in the Fourier plane. Provided that there is no overlap between interference terms, a condition which imposes restrictions regarding the spatial frequency content of the object spectrum, it is therefore straightforward to isolate each frequency component by spatial filtering [17]. Then it is possible to numerically propagate in a separate manner the two associated wavefronts. Such a procedure is completely identical to the one for real-time beat-wavelength DHM [14], and very similar to the arrangements for polarization imaging with digital holography [15] or multiple angles endoscopic

Figure 2. 3D perspectives of the investigated thin step height standard obtained by the DHM setup of figure 1. (a) Full-FOV image with the surrounding structures pointing towards the calibrated 8.9 nm step; (b) area of interest defined by the dashed rectangle in (a), where the considered step height is clearly visible.
digital holography [18, 19]. Practically, the two different spatial filters allow us to treat the two interference terms \( R^*_1 O_1 \) and \( R^*_2 O_2 \) independently.

Writing these two filtered holograms as \( I^H_{H,1} \) for \( R^*_1 O_1 \) and \( I^H_{H,2} \) for \( R^*_2 O_2 \) and by using the convolution formulation, we obtain the following expression for the Fresnel propagation:

\[
\Psi_{CF,i}(m, n) = \Gamma^i_i(m, n) \cdot \exp(i2\pi d_i/\lambda_i) \times \text{FFT}^{-1} \left\{ \text{FFT} \left[ \Gamma^H_{RCH}(k, l)I^H_{H,i}(k, l) \right] \right\} \times \exp \left\{ -i\pi \lambda_i d_i \left[ \frac{k^2}{N\Delta x^2} + \frac{l^2}{N\Delta y^2} \right] \right\}, \tag{2}
\]

where \( \Psi_{CF,i} \) is the reconstructed wavefront for wavelength \( \lambda_i \) in the convolution formulation, \( \Gamma^i_i \) is a digital phase mask (DPM) used to compensate for the tilt aberration in the image plane (see [20] for details), \( d_i \) is the propagation distance for wavelength \( \lambda_i \), \( \text{FFT} \) is the fast Fourier transform operator, \( (k, l) \) and \( (m, n) \) are the pairs of integers \((-N/2 < k, l, m, n \leq N/2)\) representing coordinates in the hologram plane, respectively the reconstruction plane, \( N \times N \) is the number of pixels of the CCD camera and \( \Delta x, \Delta y \) are the pixel sizes. \( \Gamma^H_{RCH} \) is a so-called reference conjugated hologram (RCH) used to compensate aberrations—directly in the hologram plane (without propagation)—with a calibration of the system (see below) defined as

\[
\Gamma^H_{RCH}(k, l) = |R(k, l)|^{-1} |O_0(k, l)|^{-1} \exp[-i\varphi_0(k, l)], \tag{3}
\]

where \( |O_0(k, l)| \) is the amplitude of the object wave with an optical flat as specimen in the system (ideally modification of the object wavefront induced by the specimen), \( \varphi_0(k, l) \) is the phase aberration function of the system during calibration and appears with a negative sign in equation (3) in order to cancel the system phase aberration in actual measurement.

The RCH technique was initially proposed by Colomb et al [21]. It is similar to the double-exposure principle introduced by Ferraro et al [22], but instead it operates directly in the hologram plane prior to propagation and allows us to compensate for phase aberrations and image distortion, by first recording a single calibration hologram, which is then used during the reconstruction process as in equation (2). Compared to the initial paper by Colomb et al, the RCH defined here in equation (3) comprises also two amplitude terms \(|R(k, l)|^{-1}\) and \(|O_0(k, l)|^{-1}\), corresponding to the inverse of respectively the reference and the object wave amplitudes of the RCH. The insertion of these two terms in the definition of the RCH permits a full compensation of the amplitude inhomogeneities of the wavefront in the hologram plane. When not compensated, these inhomogeneities may induce some phase aberrations after propagation of the wavefront to the image plane, especially in the low-frequency domain as illustrated in figure 5.

The formulation of equation (2) enables us to propagate each wavefront \( \Psi_{CF,i} \) in an independent manner: the DPM can be adapted to compensate for each individual wavefront’s tilt aberrations with respect to the reference hologram and the propagation distances \( d_i \) could be adjusted differently to compensate for slight chromatic aberrations or specimen displacement with respect to the working distance of the MO. Let us mention that the DPM may also be used as digital magnification lenses to resize the images in the case of stronger chromatic aberrations [23]. The parallel propagations with \( 512 \times 512 \) pixel holograms are achieved with a frame rate of about 7 frames s\(^{-1}\) with a standard dual-core PC at 2 GHz (roughly half the standard reconstruction rate). An a posteriori hologram stack reconstruction can also be performed and a sequence recording becomes this way only limited by the acquisition rate of the CCD camera (here 25 frames s\(^{-1}\)).

### 4. Results

#### 4.1. Dual-wavelength acquisition

In the first step, the two available wavefronts expressed in equation (2) are considered separately, and their corresponding phase images are reconstructed from a single hologram acquisition: this is quite similar to the standard mode of operation of single-wavelength DHM for real-time imaging at 15 frames s\(^{-1}\) with \( 512 \times 512 \) pixels, apart from the simultaneous presence of two interferograms in the same frame. The reconstruction algorithm is the one of equation (2) with a single reference hologram acquisition, in another wafer...
area without structures, and hologram amplitude division as in equation (3). The dual-wavelength hologram and the reconstructed phases at both wavelengths are presented in figure 4. It should be noted that every phase image exposed here is in reality inverted (conjugated phase) in order to render topographic map images, as for reflection configuration positive topographic steps appear as negative phase steps (providing there is no phase ambiguity, which is the case here).

The step height is easily recognizable in the phase images of figures 4(b) and (c), where the equivalent topographic spatial standard deviations are given as follows: $\sigma_1 = 1.93$ nm for $\lambda_1 = 657$ nm and $\sigma_2 = 1.62$ nm for $\lambda_2 = 680$ nm. These values are computed by taking into account all the flat areas in the images FOV, the step itself and the surrounding substrate, and represent the usual parameters to estimate DHM axial accuracy. While these values are within the same 1.5–2 nm range, one can note that the image is a bit noisier for $\lambda_1 = 657$ nm than for $\lambda_2 = 680$ nm. This translates some wavelength- or source dependence of the noise, either due to remaining parasitic interferences, non-optimal Fourier filtering or signal-to-noise ratio differences due to degradation of the interference regimes, e.g. by linear polarization instability of the $\lambda_1 = 657$ nm diode or non-optimal incident angle of the reference wave $R_1$ regarding the linear polarization direction of $O_1$ (inset of figure 1(a)).

4.2. Temporal averaging
As previously detailed in [13], shot noise contribution in DHM needs to be accounted for, even in the current optical power configuration. Indeed, with the red-range diodes used in this setup and an integration time on the CCD camera of about 2 ms, the full electron well capacity of the CCD chip (16 000 e$^-$ in our case) is clearly reached, leading to an average of 8000 photons per pixel over the whole hologram per frame acquisition [13]. However, due to the simultaneous acquisition, the available photons for one wavelength should be considered to be about 4000 photons instead: in this regime the shot noise phase perturbation has a typical standard deviation between 0.25$^\circ$ and 0.6$^\circ$ [13]. To remove this contribution to the phase noise, a temporal averaging over a 1 s period is applied by acquiring a sequence of 25 holograms at the maximum frame rate of 25 images s$^{-1}$, and each reconstructed wavefront in the hologram plane [23] is averaged, without individual propagation. This is done first for the reference hologram acquisition to obtain a ‘temporally averaged reference hologram’, then for the sample hologram with the averaged object wavefront in the hologram plane being finally digitally propagated. The resulting phase images and spatial standard deviations for the step height specimen are illustrated in figure 5.

By comparing the results of a 1 s temporal averaging of figure 5 with the single-shot measurements of figure 4, one
can estimate the gain in axial precision as about 0.5–0.6 nm (0.5° in phase) for the standard deviation, in this case corresponding to more than 30% enhancement. These values for the shot noise contribution agree very well with the theoretical estimation given in [13] for the 4000 photons regime under which the dual-wavelength DHM setup is operating. At this point, the nanometer-range axial precision is nearly reached. As mentioned previously, with just one wavelength reconstruction, one additional recommended possibility of reducing the spatial deviation furthermore could consist in spatially averaging the reference hologram at different places on the substrate, in order to discard any remaining fixed pattern due to nanometer-range specimen roughness. Nevertheless, the investigated sample does not offer this option as the available on-site flat areas are too sparse for the setup FOV.

4.3. Dual-wavelength wavefront averaging

The experimental configuration described in figure 1 provides two different wavefronts at different wavelengths during the same camera acquisition. This can be used to achieve a spatial averaging between both wavelength topographic maps obtained above. By making the hypothesis that these wavefronts are not correlated and follow a Gaussian noise distribution with similar topographic standard deviations $\sigma_{x_i}$, one can write the following for the averaged topographic map standard deviation $\bar{\sigma}_x$:

$$\bar{\sigma}_x = \sqrt{\frac{2\sigma_{x_i}^2}{2}} = \sqrt{2}\sigma_{x_i}. \tag{4}$$

Looking at equation (4) it is straightforward to conclude that the spatial averaging process between both wavefronts is theoretically equivalent to a noise attenuation factor of $\sqrt{2}$, numerically a 0.71 multiplying factor or a $-29\%$ reduction. Figure 6 presents the resulting spatially averaged topographic map obtained from the temporally averaged single-wavelength maps of figure 5.

The resulting spatial standard deviation in figure 6(c) is $\sigma = 0.87$ nm, now achieving a sub-nanometer axial precision with DHM. Compared to the theoretical gain of equation (4), the improvement is 28.5%, being in perfect agreement with the predicted value. This validates the hypothesis of uncorrelated wavefronts with two different laser sources at different wavelengths, and noise following a Gaussian relation. Qualitatively, this can also be observed by the apparent ‘granularity’ reduction between figures 6(c), (f) and figures 6(a), (b), (d), (e).

Finally, figure 7 presents 1D profiles extracted from figure 6 and gives measured heights in specific zones.
placed according to ISO-5436-1 standard. The profiles of figure 7 are obtained transversally along the horizontal median of figures 6(a)–(c). One can notice that when the noise standard deviation is too different between the two wavelength results, the relation of equation (4) is no longer valid, the precision gain by averaging the wavefronts thus becomes negligible. Based on the measured heights in figure 7 the thin step can be estimated as 8.65 nm high, within the manufacturer uncertainty range. Moreover the standard deviation on flat areas reaching sub-nanometer values assesses that this resolution level has been reached, given the sample-certified roughness in the ±0.5 nm range.

4.4. Temporal stability (precision)

First, to estimate the temporal stability of our dual-wavelength DHM system, the temporal standard deviation \( \sigma_{t1} \) for one pixel is computed for a 15 s long sequence. The values of \( \sigma_{t1} = 1.26 \) nm for \( \lambda_1 = 657 \) nm, and respectively \( \sigma_{t2} = 1.03 \) nm for \( \lambda_2 = 680 \) nm, are obtained in this way. This tends to reveal some stability issues for the first diode at 657 nm compared to the second source. However, to truly estimate the temporal stability—and only for this purpose—it is necessary to take into account the accuracy of the system and the sampling in the image plane. As illustrated by the hologram spectrum in figure 1(c), the dual wavelength setup off-axis configuration fully uses the CCD sampling capacity. The diameter of the Airy pattern in the reconstructed image, defined by the CCD sampling and the reconstruction algorithm point spread function (filtering), is 4 pixels. Consequently, the temporal standard deviation for the 15 s sequence has also been calculated for a group of 16 pixels \((4 \times 4)\), giving the temporal stability of the system for a phase image down-sampled to a sampling corresponding to the diffraction limit. It should be noted that imaging this way would theoretically not degrade resolution, as the down-sampling would match the experimental diffraction-limited resolution. By using this approach to evaluate temporal stability, the following values are obtained: \( \sigma_{t1} = 0.77 \) nm and \( \sigma_{t2} = 0.69 \) nm, showing that our DHM system precision—in the metrology definition—lies effectively within a sub-nanometer regime. Simply increasing the averaging time can furthermore reduce these values according to a square root law.

5. Conclusion

We demonstrated digital holographic microscopy (DHM) results with an axial topographic resolution reaching the sub-nanometer range. A 8.9 ± 0.5 nm high calibrated chromium thin step sample was successfully investigated and measurements are within manufacturer uncertainty. To achieve sub-nanometric axial accuracy, three successive steps are necessary in this configuration. First, we expose a modified reference-conjugated hologram reconstruction algorithm capable of correcting amplitude-dependent parasitic phase variations. Second, by temporal averaging of both reference and object holograms over a 1 s sequence it is possible to eliminate shot noise contributions, thus reducing noise by about 0.5 nm. Finally, we propose a novel
single-acquisition dual-wavelength DHM approach to achieve wavefront spatial averaging, leading to an additional 30% noise reduction gain and consequently breaking the nanometer barrier. In addition, these results may contain remaining specimen-induced contributions and they can be enhanced with multiple on-sample reference acquisitions, providing the specimen contains large enough flat areas. On top of this, the proposed DHM experimental configuration shows a precision (temporal stability) below 1 nm over a 15 s time lapse, definitely entering a new era for DHM future performance and applications. An overview of the exposed results is summarized in table 1.

References