Research Article | Vol. 26, No. 20 | 1 Oct 2018 | OPTICS EXPRESS 25524
Optics EXPRESS

Global motion compensation for compressing holographic videos

DAVID BLINDER,[1,2,*] COLAS SCHRETTER[1,2] AND PETER SCHELKENS[1,2]

[1] *Dept. of Electronics and Informatics (ETRO), Vrije Universiteit Brussel (VUB), Pleinlaan 2, B-1050 Brussels, Belgium*
[2] *imec, Kapeldreef 75, B-3001 Leuven, Belgium*
[*] *dblinder@etrovub.be*

Abstract: Large high-resolution digital holographic displays may become feasible in the near future, and they will need considerable amounts of data. Handling this bandwidth is particularly challenging for dynamic content operating at video rates. Conventional motion compensation algorithms from classical video coders are ineffective on holograms because, in contrast to natural imagery, each pixel contains partial information from the whole scene. We propose an accurate motion compensation model predicting how hologram content changes with respect to 3D rigid-body motion that arises in natural scenes. Using diffraction theory, we derive tractable closed form expressions for transforming 2D complex-valued holographic video frames. Our experiments use computer generated hologram videos with known ground truth motion. We integrated the proposed motion compensation model into the HEVC codec. We report Bjøntegaard delta-PSNR ratio gains of 8 dB over standard HEVC.

OCIS codes: (090.1995) Digital holography; (090.2870) Holographic display; (100.2000) Digital image processing; (100.6890) Three-dimensional image processing.

References and links

1. V. M. Bove, "Display holography's digital second act," Proc. IEEE **100**, 918–928 (2012).
2. C. Slinger, C. Cameron, and M. Stanley, "Computer-generated holography as a generic display technology," Computer **38**, 46–53 (2005).
3. J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer, and T. Wedi, "Video coding with H.264/AVC: tools, performance, and complexity," IEEE Circuits Syst. Mag. **4**, 7–28 (2004).
4. G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," IEEE Transactions on Circuits Syst. for Video Technol. **22**, 1649–1668 (2012).
5. Y. Xing, B. Pesquet-Popescu, and F. Dufaux, "Compression of computer generated phase-shifting hologram sequence using AVC and HEVC," Proc. SPIE **8856**, Applications of Digital Image Processing XXXVI (2009).
6. F. Dufaux, Y. Xing, B. Pesquet-Popescu, and P. Schelkens, "Compression of digital holographic data: an overview," Proc. SPIE **9599**, Applications of Digital Image Processing XXXVIII (2015).
7. E. Darakis and T. J. Naughton, "Compression of digital hologram sequences using MPEG-4," Proc. SPIE **7358**, Holography: Advances and Modern Trends (2009).
8. H. Yoshikawa and J. Tamai, "Holographic image compression by motion picture coding," Proc. SPIE **2652**, Practical Holography X (1996).
9. Y.-H. Seo, H.-J. Choi, and D.-W. Kim, "3d scanning-based compression technique for digital hologram video," Signal Process. Image Commun. **22**, 144–156 (2007).
10. Y.-H. Seo, Y.-H. Lee, J.-S. Yoo, and D.-W. Kim, "Scalable hologram video coding for adaptive transmitting service," Appl. Opt. **52**, A254–A268 (2013).
11. Y.-Z. Liu, J.-W. Dong, Y.-Y. Pu, B.-C. Chen, H.-X. He, and H.-Z. Wang, "High-speed full analytical holographic computations for true-life scenes," Opt. Express **18**, 3345–3351 (2010).
12. T. Senoh, K. Wakunami, Y. Ichihashi, H. Sasaki, R. Oi, and K. Yamamoto, "Multiview image and depth map coding for holographic tv system," Opt. Eng. **53**, 112302 (2014).
13. A. Gilles, P. Gioia, R. Cozot, and L. Morin, "Fast generation of complex modulation video holograms using temporal redundancy compression and hybrid point-source/wave-field approaches," Proc. SPIE **2652**, Applications of Digital Image Processing XXXVIII (2015).
14. J. J. Healy, M. A. Kutay, H. M. Ozaktas, and J. T. Sheridan, *Linear Canonical Transforms* (Springer, 2016).

#331626 https://doi.org/10.1364/OE.26.025524
Journal © 2018 Received 4 Jun 2018; revised 22 Jul 2018; accepted 1 Aug 2018; published 17 Sep 2018

Optics EXPRESS

15. M. de Gosson, *The Wigner Transform*, Advanced textbooks in mathematics (World Scientific Publishing Company Pte Limited, 2017).
16. D. G. Abdelsalam, R. Magnusson, and D. Kim, "Single-shot, dual-wavelength digital holography based on polarizing separation," Appl. Opt. **50**, 3360–3368 (2011).
17. L. Zhao, J. J. Healy, and J. T. Sheridan, "Two-dimensional nonseparable linear canonical transform: sampling theorem and unitary discretization," J. Opt. Soc. Am. A **31**, 2631–2641 (2014).
18. "High Efficiency Video Coding (HEVC) reference software," HM-15.0 (Revision 4879), https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/branches/HM-dev/. [retrieved 31 March 2017].
19. G. Bjontegaard, *Calculation of average PSNR differences between RD-curves*, ITU-VCEG (2001).

1. Introduction

Digital holography is a technology that captures and reconstructs the complex-valued amplitude of a wavefront of light. Holographic display systems can account for all visual cues [1], such as depth perception, continuous parallax and cause no accommodation-vergence conflict. By displaying holograms in rapid succession, we get a holographic video. High-quality holograms require resolutions surpassing 10^8 pixels for displaying big scenes and large viewing angles [2]; these resolutions multiplied with video rates result in overwhelming data volumes. That is why adequate compression techniques need to be developed to address this limitation.

A core building block of video coders is the motion compensation: successive frames are strongly correlated, which can be compensated by accounting for the present motion in the recorded scene. This is generally done in modern video coders such as AVC and HEVC by dividing the source frames into blocks, and optimizing for the choice of motion vectors that minimize the difference between the target frame and the compensated source frame blocks [3, 4]. Mostly, only purely translational motion in the frame's (x,y)-plane is accounted for.

Unfortunately, this approach cannot readily be applied on holographic frames: due to the nature of diffraction, any motion that is not parallel to the hologram plane will result in pervasive changes of the holographic signal. This is mainly because the light of localized features will propagate, spreading and redistributing the information over the whole hologram. That is why conventional motion estimation and compensation for video coding will generally fail for this medium. This fact was *e.g.* noticeable in the work of Xing et al. [5] where holographic videos were compressed with a standard HEVC codec: inter-frame decorrelations brought little benefit to the compression performance. We therefore propose to use a motion compensation technique that is adapted to holographic video. In doing so, we demonstrate its suitability for integration with existing modern video codecs.

The field of digital hologram compression has mostly started to develop since the early 2000's [6]. However, the focus has been primarily on static holograms. In comparison, the literature on coding dynamic holograms has been rather limited. Holographic video sequences were compressed with standard MPEG-4 and HEVC architectures [5, 7], although with limited to no inter-frame compression performance. Other techniques involve the use of "scanning" methods [8, 9], where the hologram frames are converted to a representation resembling a light-field by dividing the holograms into small apertures. This approach was later refined into a scalable video coding solution [10]. These resulting views are then reordered and compressed with a conventional video coder. Unfortunately, this approach can cause speckle artefacts due to the small apertures, and only scarcely exploits inter-frame correlations. Another approach is generating holograms on the fly by compressing textures, depth maps, 3D models or multi-view representations [11] of the scene instead [12, 13]. This results indeed in high compression rates, but requires high computational resources to compute the holograms at the decoder side and often limit the type of scenes that can be represented efficiently. Moreover, this approach cannot encode optically acquired holograms.

In Section 2, we explain the proposed motion compensation model for holographic video. Then, we create a dynamic computer-generated hologram of a moving 3D model in Section 3 and

Optics EXPRESS

test our proposed video coding framework on it in Section 4. We discuss some of the limitations and remaining challenges in Section 5, and finally we conclude in Section 6.

2. Methodology

The goal of motion compensation is to predict a video frame as accurately as possible given past and/or future frames to account for the present motion. We propose to do motion compensation by modeling signal changes in the hologram by representing motion in the 3D scene instead of in the 2D hologram.

We will model the set of proper rigid transformation in 3D space called the special Euclidean group. It is the set of isometries that preserves handedness, consisting of all combinations of translations and rotations. These movements can be modeled in signal space via a superset of the Linear Canonical Transforms (LCT) [14], which we call Affine Canonical Transforms (ACT). Thereafter, this notation allows us to elegantly define transforms matching motion in holograms.

2.1. Affine canonical transforms

n-dimensional signals can be represented using a $2n$-dimensional time-frequency (TF) representation, depicting time and frequency domains simultaneously; this is also known in optics as "Phase space" or "Wigner space". A particular class of unitary operators of interest are the LCT: they can be expressed as linear transformations of TF space that preserve the Heisenberg uncertainty principle. The LCT are isomorphic to the symplectic matrix group $\mathrm{Sp}(2n, \mathbb{R})$, which is a subset of the special linear group $\mathrm{SL}(2n, \mathbb{R})$. It is defined by the equivalence

$$\mathbf{S} \in \mathrm{Sp}(2n, \mathbb{R}) \subseteq \mathrm{SL}(2n, \mathbb{R}) \iff \mathbf{S}^{\mathrm{T}}\mathbf{J}\mathbf{S} = \mathbf{J} \tag{1}$$

with the matrices

$$\mathbf{J} = \begin{pmatrix} 0_n & I_n \\ -I_n & 0_n \end{pmatrix} \quad \text{and} \quad \mathbf{S} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \tag{2}$$

where 0_n is a $n \times n$ matrix of zeros, I_n is a $n \times n$ identity matrix, and A to D are $n \times n$ matrices.

There is a bijective mapping between the double covering of $\mathrm{Sp}(2n, \mathbb{R})$ and the metaplectic group $\mathrm{Mp}(n, \mathbb{R})$ [15], which are composed of unitary operators; namely, every symplectic matrix is mapped to a pair of unitary operators differing only by a sign. The mapping operator $\mathcal{M}_{\mathbf{S}} \in \mathrm{Mp}(n, \mathbb{R})$ operating on a signal f for the typical use case ($\det B \neq 0$) is defined as [14]:

$$\mathcal{M}_{\mathbf{S}}[f](\tilde{\mathbf{x}}) := |iB|^{-\frac{1}{2}} \cdot \int_{\mathbb{R}^n} e^{i\pi(\tilde{\mathbf{x}}^{\intercal} D B^{-1} \tilde{\mathbf{x}} - 2\mathbf{x}^{\intercal} B^{-1} \tilde{\mathbf{x}} + \mathbf{x}^{\intercal} B^{-1} A \mathbf{x})} f(\mathbf{x}) d\mathbf{x} \tag{3}$$

The LCTs comprise among others the Fourier transform (a 90° rotation matrix), fractional fourier transforms (arbitrary rotation matrices) and the Fresnel transform (shearing matrix). However, the LCTs cannot account for translations in space or modulations in frequency, which both map to translations of TF space. Fortunately, we can generalize the symplectic matrix group by combining it with the (phase space) translation group, which is isomorphic to $\mathbb{R}^{2n}$:

$$\mathfrak{A}(2n, \mathbb{R}) = \mathrm{Sp}(2n, \mathbb{R}) \ltimes \mathbb{R}^{2n} \tag{4}$$

where $\ltimes$ denotes the semidirect product. We then obtain the group $\mathfrak{A}$, consisting of affine symplectic mappings in the TF domain, which we will refer to as ACT. Elements $\{\mathbf{S}, \mathbf{b}\} \in \mathfrak{A}(2n, \mathbb{R})$ transform every point $\mathbf{p}$ as follows:

$$\{\mathbf{S}, \mathbf{b}\} : \mathbb{R}^{2n} \to \mathbb{R}^{2n} \quad \text{by} \quad \mathbf{p} \mapsto \mathbf{S}\mathbf{p} + \mathbf{b}, \tag{5}$$

where every point $\mathbf{p} = (\mathbf{x}; \boldsymbol{\omega})$ has coordinates in space $\mathbf{x} = (x_1, ..., x_n)^{\intercal}$ and in frequency $\boldsymbol{\omega} = (\omega_1, ..., \omega_n)^{\intercal}$. We refer to them as ACT to differentiate them from the well-known LCT

Optics EXPRESS

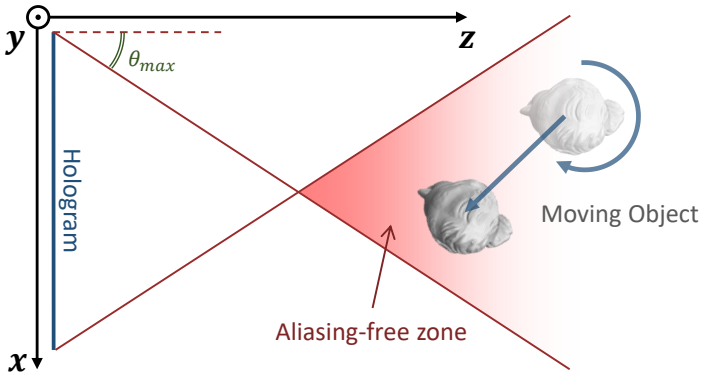


Fig. 1. Illustration of the virtual setup for generating the video of the moving Venus head object, intended for display with a planar reference wave.

which do not possess a translation $\mathbf{b} \in \mathbb{R}^{2n}$ in most common definitions. The vector $\mathbf{b} = (\mathbf{b_x}; \mathbf{b}_\omega)$ denotes a translation in space and a modulation in frequency, while $\mathbf{S} \in \mathrm{Sp}(2n, \mathbb{R})$.

The ACT generate a group of unitary operators, called the inhomogeneous metaplectic group $\mathrm{IMp}(n, \mathbb{R})$ [15]. We define the transform $\mathcal{M}_{\{\mathbf{S},\mathbf{b}\}} \in \mathrm{IMp}(n, \mathbb{R})$ operating on a signal f as

$$\mathcal{M}_{\{\mathbf{S},\mathbf{b}\}}[f](\tilde{\mathbf{x}}) := e^{2\pi i \mathbf{b}_\omega^\intercal \tilde{\mathbf{x}}} \cdot |iB|^{-\frac{1}{2}} \cdot \int_{\mathbb{R}^n} e^{i\pi(\tilde{\mathbf{x}}^\intercal D B^{-1} \tilde{\mathbf{x}} - 2\mathbf{x}^\intercal B^{-1} \tilde{\mathbf{x}} + \mathbf{x}^\intercal B^{-1} A \mathbf{x})} f(\mathbf{x} - \mathbf{b_x}) d\mathbf{x} \tag{6}$$

when B is nonsingular. When B is singular, we have

$$\mathcal{M}_{\{\mathbf{S},\mathbf{b}\}}[f](\mathbf{x}) := |A|^{-\frac{1}{2}} \cdot e^{i\pi(\mathbf{x}^\intercal C A^{-1} \mathbf{x} + 2\mathbf{b}_\omega^\intercal \mathbf{x})} f(\mathbf{x} - \mathbf{b_x}) \,. \tag{7}$$

Note that when $\mathbf{b} = 0$, the equations reduce to the LCT.

2.2. Global motion model for holography

In this paper, we will describe the hologram in its general form as a complex-valued wavefield $u : \mathbb{R}^2 \to \mathbb{C}$, making our proposed motion model independent of the holographic acquisition method or geometry. Motion in the 3D scene will correspond to transformations of the hologram frames which can be modeled by 4D ACTs, since for holograms $n = 2$. We choose the x-axis and y-axis of the scene to lie in the hologram plane, with the z-axis perpendicular to it, as shown in Fig. 1. We define the 2D Fourier transform $\hat{u}$ of a complex-valued hologram u as follows:

$$\hat{u}(\xi, \eta) = \iint_{\mathbb{R}^2} u(x, y) e^{-2\pi i(x\xi + y\eta)} dx dy \tag{8}$$

For small rotations around the x-axis and y-axis (*i.e.* where $\sin(\theta) \approx \theta$ is valid), we can accurately approximate the tilt by a shift in Fourier space, corresponding to $\mathbf{b}_\omega = (\theta_x, \theta_y)^\intercal$:

$$\mathrm{R}_{\mathrm{x,y}}(\theta_x, \theta_y) : \; \hat{u}(\xi, \eta) \mapsto \hat{u}(\xi - \theta_x, \eta - \theta_y) \iff u(x, y) \mapsto u(x, y) e^{2\pi i(x\theta_x + y\theta_y)} \tag{9}$$

By multiplying with a phasor instead of translating in Fourier space, we can numerically model the tilt more accurately when needing non-integer frequency shifts. This notion similar to the one used in [16] for compensating the tilt of the wavefront around z.

Rotations around the z-axis involve simply rotating the hologram itself, or equivalently its Fourier domain. The corresponding LCT can be described by the matrix [17]

$$\mathrm{R}_\mathrm{z}(\theta_z) = \begin{pmatrix} R_2 & 0_2 \\ 0_2 & R_2 \end{pmatrix}, \quad R_2 = \begin{pmatrix} \cos(\theta_z) & \sin(\theta_z) \\ -\sin(\theta_z) & \cos(\theta_z) \end{pmatrix} . \tag{10}$$

Optics EXPRESS

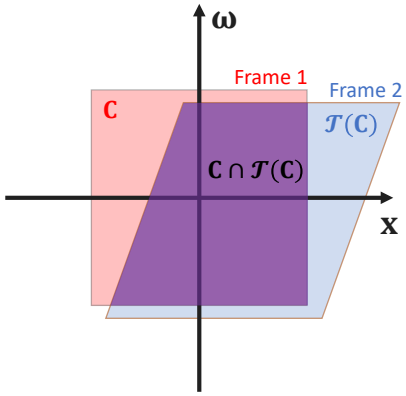


Fig. 2. Simplified diagram of the TF space intersection of a motion compensated 1D hologram (Frame 2) using an ACT w.r.t. another frame (Frame 1). The surface area of the intersection is a good indication of the mutual information between the frames, thereby predicting the efficacy of the proposed motion compensation algorithm.

Translations in the x-y plane correspond to translations of the hologram, *i.e.* $\mathbf{b}_x = (t_x, t_y)^\mathsf{T}$. Similarly to the rotation case, it is preferable to compute translation by multiplying with a phasor in Fourier space, as to accurately model non-integer pixel shifts:

$$\mathrm{T}_{\mathrm{x,y}}(t_x, t_y) : \; u(x, y) \mapsto u(x - t_x, y - t_y) \iff \hat{u}(\xi, \eta) \mapsto \hat{u}(\xi, \eta) e^{-2\pi i (t_x \xi + t_y \eta)} \tag{11}$$

Translations along the z-axis perpendicular to the hologram plane can be modeled by Fresnel diffraction, which is a convolution. This reduces to a multiplication in Fourier space:

$$\mathrm{T}_{\mathrm{z}}(t_z) : \; \hat{u}(\xi, \eta) \mapsto \hat{u}(\xi, \eta) e^{-2\pi i \lambda t_z (\xi^2 + \eta^2)} \tag{12}$$

where λ is the wavelength of the coherent light source. Fresnel diffraction can be expressed as a shear in TF-space by the following LCT [17]:

$$\mathrm{T}_{\mathrm{z}}(t_z) = \begin{pmatrix} I_2 & \frac{2t_z}{\lambda} I_2 \\ 0_2 & I_2 \end{pmatrix} . \tag{13}$$

Since R_z and T_z commute and are both closed under multiplication, the subset $\mathfrak{R} \subset \mathfrak{A}(4, \mathbb{R})$ modeling all possible combinations of rigid body motions can always be written as

$$\tilde{\mathbf{x}} = \mathrm{R}_\mathrm{z}(\psi) \cdot \mathrm{T}_\mathrm{z}(d) \cdot \mathbf{x} + \mathbf{b} \tag{14}$$

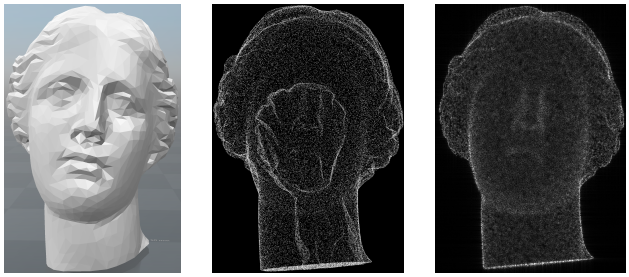
for some $\psi, d \in \mathbb{R}$ and $\mathbf{b} \in \mathbb{R}^4$, thus $\mathfrak{R}$ forms a proper subgroup of the ACTs.

Because all rigid body motions can be combined into a single rotation and translation operator, and because $\mathfrak{R}$ forms a group, we need only 2 FFTs per frame: one for transforming the Fourier space, and one to return to the spatial domain. The operations in both domains will only consist of pointwise multiplications of the signal with phase-only functions and rotations. In summary, the 3D body rotations are applied in the spatial domain, and the 3D translations in Fourier space.

Note that we do not claim to reinvent the laws of diffraction. Instead, we select and combine specific diffraction operators for each of the 6 DoF so to construct the mathematical group $\mathfrak{R}$ isomorphic to small 3D rigid body motions.

The volume of a part of the TF space corresponds to the spatial bandwidth product (SBP), which is a measure of the information content. This property can be used to model how much redundant

Optics EXPRESS



(a) (b) (c)

Fig. 3. (a) 3D model of the Venus head, (b) the associated point cloud, (c) amplitude of backpropagated hologram generated from the pointcloud with occlusion. The model of Venus is courtesy of Direct Dimensions Inc. (Head sculpture).

information there is between subsequent video frames. The dimensions of the 4D-cuboid $\mathbf{C}$ denoting the signal content of the hologram will depend on the size and sampling rate of the hologram, and its volume will be proportional to the SBP. In the subsequent frame, the signal in the cuboid will be transformed by some $\mathcal{T} \in \Re$. However, this information will only be captured partially due to the bounded size and sampling rate of the hologram. Assuming constant resolution and pixel pitch over time, the proportion of mutual information content shared between two subsequent frames is given by the (hyper)volume ratio

$$I = \frac{\mathrm{Vol}(\mathbf{C} \cap \mathcal{T}(\mathbf{C}))}{\mathrm{Vol}(\mathbf{C})} \in [0, 1]\,, \tag{15}$$

which can serve as a measure for the interframe redundancy. This principle is illustrated on Fig. 2.

3. Computer-Generated Holography

For testing the validity of our model, we generated a digital hologram sequence using a triangle mesh of Venus, *cf.* Fig. 3. We randomly sample the mesh surface to get a point cloud consisting of $N = 10^5$ points. To compute the hologram, we use the ray-tracing equation, evaluating the sum of all point spread functions of the point cloud onto the hologram:

$$U(x, y) = \sum_{j=1}^{N} \frac{a_j}{r_j} \exp(i \frac{2\pi}{\lambda} r_j) \tag{16}$$

where $r_j = \sqrt{(x - x_j)^2 + (y - y_j)^2 + z_j^2}$ is the distance between every object point (x_j, y_j, z_j) for a hologram pixel (x, y). a_j is the associated point amplitude and λ is the wavelength of the coherent light source. As is, this approach does not model occlusion because *e.g.* light emanating from the back of the head will not be blocked. To account for occlusion, we check for all points whether any of the emanated rays to all hologram pixels is blocked by the mesh. We can establish which of the hologram pixels will be visible for every point source by rendering the visibility map as a binary mask to the hologram plane, see Fig. 4. This map will be multiplied with every PSF, so that blocked rays will not contribute to the complex amplitude $u(x, y)$ in Eq. 16. This method was implemented on a GPU using CUDA and OpenGL.

For this experiment, we generated a 64-frame video of Venus (see Fig. 5) moving within the aliasing-free zone, *cf.* Fig. 1. This is the region where point sources do not create frequencies ν

Optics EXPRESS

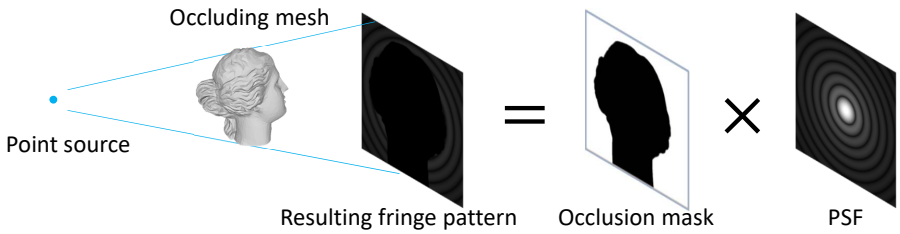


Fig. 4. Diagram of a CGH contribution of a single occluded PSF using ray-tracing. This process is repeated for every PSF and summed over to obtain the hologram frame.

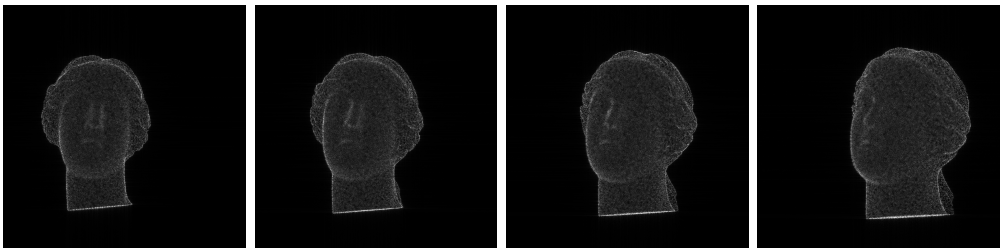


Fig. 5. Amplitude of several reconstructed hologram frames from the tested reference video sequence, backpropagated to the center of the mesh.

surpassing the maximally allowed frequency determined by the hologram pixel pitch p, otherwise causing aliasing. This region is delineated by the maximum diffraction angle θ_{max}, associated to the maximum frequency $\nu_{max} = \frac{1}{2p}$ using the grating equation $\lambda\nu = \sin(\theta)$.

In the test setup, the hologram is centered at the origin of the scene, perpendicular to the z-axis. The used pixel pitch is $p = 4\,\mu$m, the wavelength is $\lambda = 633$ nm and the hologram resolution is 1024×1024 pixels. Venus, with a size of approximately 1 cm, was centered at its initial position in $\vec{x}_0 = (-2, 0, 100)$ mm. The object was translated with a speed of 10 mm/s in the z-direction and 4 mm/s in the x-direction, while being rotated 30°/s around its center along the y-axis.

The motion compensation model is applied on every pair of consecutive frames, where each frame #n was subtracted with a compensated version of frame #$(n-1)$. Data outside of the available TF-domain was assumed to be 0.

4. Experiments

The global motion compensation model can be combined with a video codec to significantly improve the compression performance. We propose an unidirectional reversible compensation scheme, detailed in Algorithm 1. The first frame is compressed independently in a lossy manner, which is then decompressed to serve as an estimate for the next frame. We then use the proposed motion compensation model on the previous decompressed frame to transform the signal to match the translation and rotation parameters of the subsequent frame. We cannot use motion compensation on the original frames, as they won't be available at the decoder side.

For the compression experiments, we evaluate the performance of the High Efficiency Video Coding (HEVC) codec using the reference software HM 15.0 [18]. The complex-valued holograms are quantized to a dynamic range of 8 bits per pixel (bpp) and stored in RGB 4:4:4 format, without

Optics EXPRESS

Algorithm 1 Holographic video compression with global motion compensation.

procedure HOLOGRAPHIC MOTION COMPENSATION
  $bitstream \leftarrow \{\}$
  $comp \leftarrow 0$ ▷ compensated frame
  for $i \in \{1, ..., N\}$ **do**
    $c \leftarrow$ intra-compression$(frame_i - comp)$
    $bitstream$.append(c)
    $d \leftarrow$ decompression(c)
    $comp \leftarrow$ motion_compensation$(comp + d)$
  end for
  return $bitstream$
end procedure

Table 1. Compression results with the BD-PSNR improvements (in dB) w.r.t. to the default HEVC configuration for video compression in the range of 0.125 to 2 bpp.

	no backprop	backprop
HEVC inter	(reference)	4.14 dB
HEVC intra	0.19 dB	3.96 dB
proposed MC	5.33 dB	8.79 dB

subsampling. Only two channels are used: the R-channel contains the real part of the signal, the G-channel contains the imaginary part. The B-channel is not used.

We test three different codec configurations: (1) inter mode: using (forced) inter-frame compression, with GOP size = 8, GOP structure IBBBBBBB. (2) intra-only mode: inter-frame compression is turned off. (3) proposed mode: we use our proposed motion compensation model to predict subsequent frames, combined with intra-mode compression. When configuring HEVC to its default settings, it almost consistently resorts to intra-only mode, confirming the inadequacy of HEVC's motion compensation model for holograms.

Moreover, we can backpropagate the (compensated) hologram frames as well. This will concentrate the dispersed light into a smaller region and make the hologram frames look more image-like, thereby further improving the compression performance. We tested the three configurations with and without backpropagation, resulting in 6 different codec configurations.

We evaluated all codec configurations at compression rates between 0.125 and 2 bpp per frame. At every rate, we obtain a distortion expressed in average PSNR, according to the formula

$$\mathrm{PSNR} = \frac{1}{N}\sum_{i=1}^{N} 10\log_{10}\Big(\frac{\max(|R_i|)^2}{\mathrm{mean}(|R_i - X_i|^2)}\Big) \tag{17}$$

for a video with N frames, pairwisely comparing the ith reference frame R_i with the corresponding distorted frame X_i. The PSNR is calculated in the complex domain. The results are shown in Fig. 6. We also report the average relative gain in PSNR for all algorithms using Bjøntegaard-Delta PSNR (BD-PSNR) [19]. The BD-PSNR results are shown in Table 1. Additionally, we uploaded a video with side-by-side reconstructions of the dynamic hologram to demonstrate the visual gain in quality (Visualization 1).

As expected, the default motion compensation algorithm in HEVC cannot effectively account for motion in the hologram. Even after backpropagation, the conventional motion prediction

Optics EXPRESS

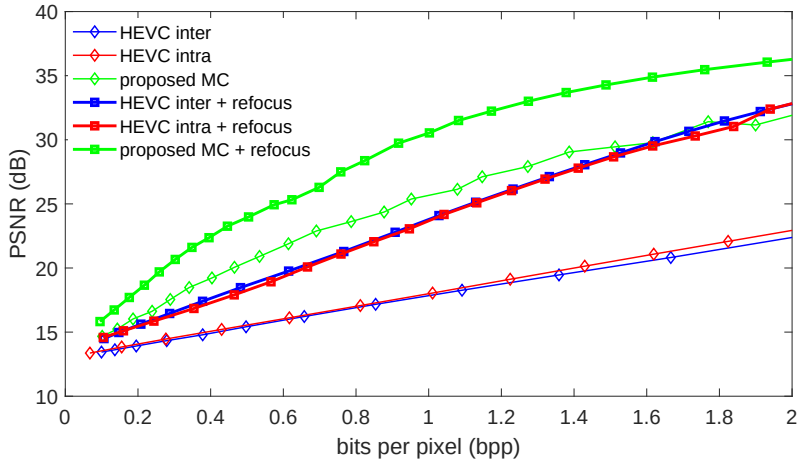


Fig. 6. PSNR as a function of the allocated rate for all six codec configurations. The proposed motion compensation outperforms local in-plane motion compensation.

scheme fails to adequately account for scene motion. The proposed motion compensation model improves the compression performance with more than 5 dB BD-PSNR, both with and without backpropagation. The backpropagation itself independently improves performance with roughly 4 dB BD-PSNR, emphasizing its importance, as it will concentrate the energy into a smaller spatial region and render the hologram to look more image-like, improving compressibility.

Regarding the computational complexity, the main contributing factor is the hologram resolution; the FFT operator has linearithmic complexity, which will largely dictate how the computational cost will scale for high resolutions. The calculation time scales linearly with video length, and is independent of the video content. The motion compensation procedure took 0.47s of calculation time per frame on average, which was computed on a machine with an Intel Core i7 6700K CPU, 32GB RAM, OS Windows 10 using MATLAB R2018a software.

5. Discussion

The proposed motion model is only a first step towards a general holographic video coder. Therefore, we will briefly summarize in this section discussing some of the current limitations of the method and several remaining challenges that need to be tackled.

The current motion model, when directly applied to the hologram frames, can only compensate for the rigid motion of a single object and camera. Generalizing this to multiple independently moving object (pieces) would require some form of signal segmentation before applying the ACT individually on each piece. This segmentation will again not be possible in the spatial domain, because the signals to be separated will strongly overlap due to the nature of diffraction. Potential solutions would involve segmenting the signal in the TF domain instead, perhaps with the use of various non-stationary filters.

Presently, we applied the proposed algorithm on computer-generated holograms with known ground-truth motion. To successfully apply this methodology on optically acquired holograms, one would need to estimate the scene motion in the holographic video as well. Automated general motion estimation in digital holography is a research challenge that has not been solved yet, especially for complicated scenes and arbitrary motion.

Currently, the motion compensation model was integrated in a default HEVC implementation. However, HEVC is highly configurable and can thus be optimized for specific applications,

Optics EXPRESS

potentially improving compression efficiency; this includes parameters such as GOP size and structure, transform settings, configuring the rate-distortion optimizer, etc. Even processing steps such as the deblocking filter should be re-evaluated to establish whether they are suited for hologram coding. Moreover, modifications to existing video codecs will have to be made to enable support for (computationally efficient) encoding/decoding of high-quality holographic video streams surpassing 10^8 pixels.

Furthermore, the proposed motion model could also be used for active global motion compensation, such as for holographic virtual/augmented reality head-mounted displays: the transmitted video stream can be updated and compensated in function of any relevant changes in the user's viewing angle and position (or scenery). This can strongly reduce the required video bandwidth compared to a transmission of the full hologram.

Ultimately, the goal is to build a scalable codec tuned for static and dynamic digital holograms, efficient both in terms of compression performance as well as in calculation time.

6. Conclusion

We derived a motion compensation model for efficiently compressing holographic video, by modeling how holographic signals are transformed by rigid body motion in 3D scene space. We integrated the proposed motion compensation model into a HEVC codec, and report gains exceeding 8 dB BD-PSNR over conventional motion compensation approaches. Future work will tackle motion estimation, tests on optically recorded holograms, enabling local motion models and further codec optimizations.

Funding

European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Grant Agreement n.617779 (INTERFERE).