Wave atoms for digital hologram compression

Birnbaum, Tobias; Ahar, Ayyoub; Blinder, David; Schretter, Colas; Kozacki, Tomasz; Schelkens, Peter

Published in:
Applied Optics

DOI:
10.1364/AO.58.006193

Publication date:
2019

License:
CC BY

Document Version:
Accepted author manuscript

Citation for published version (APA):
Wave atoms for digital hologram compression

Tobias Birnbaum1,2,*, Ayoub Ahar1,2, David Blinder1,2, Colas Schretter1,2, Tomasz Kożacki3, and Peter Schelkens1,2

1 Vrije Universiteit Brussel (VUB), Dept. of Electronics and Informatics (ETRO), Belgium. 2 imec, Belgium. 3 Warsaw University of Technology, Institute of Micromechanics and Photonics, Poland.

* tobias.birnbaum@vub.be

Abstract: Compression of macroscopic digital holograms is a major research problem, which if unresolved would continue limiting the possible applications of holography in multimedia contexts. The quest of searching for the most suitable representation for compression is still an open problem. In this work, we study sparsification by the wave atom transform, introduced in 2006 by Demanet et al., and experiment on four large-scale representative diffuse macroscopic holograms while testing compressibility in the object-plane, Fourier-plane, and defocused plane representations, respectively. We demonstrate that it is a suitable non-adaptive, sparsifying transform for Fourier or defocused content and by integration into the Wave Atom Coding (WAC) method, we sketch a full fledged codec for the compression of macroscopic holograms. WAC is compared to two variants of JPEG 2000, with equal complexity of coding tools, and the more recent H.265/HEVC. For Fourier and defocused holograms WAC outperforms the JPEG 2000 variants by $0.9 - 7.9$ dB BD-PSNR, especially in the former case, while it is as good as or better than even H.265/HEVC for very deep computer-generated holograms, thus improving on existing approaches.

© 2019 Optical Society of America

1. Introduction

Holograms capture a cross-section through the complete incident light wavefield and account for all visual cues [1], such as depth perception and continuous parallax, without exhibiting visual resolution loss or accommodation-vergence conflicts. When holography became digitally accessible, it was no longer restricted to optically recorded holograms (OH) but could also be extended to computer-generated content such as 3D rendered scenes [2]. One major issue that keeps digital holograms from being found in consumer electronics, is that they amass vast amounts of data. For example, a plane-wave recorded, static, digital hologram of a 10 cm object with field of view as large as $\pm 36^\circ$, recorded with visible light requires $\geq 100$ Mpixel resolution.

In holography, visual information is encoded through a superposition of interference fringes. Each fringe is the result of a spherical wave emitted from one of the illuminated points in the recorded scene whose footprint extends across the whole hologram. Any local change of an emitting point will therefore affect the entire hologram in a predictable manner. The position of each fringe on the hologram plane encodes the lateral spatial position of the emitter and its phase encodes the depth within the scene.

One emergent property of point-sets will be the diffusiveness of surfaces, which is often modeled in physics as a 4-dimensional bidirectional reflectance distribution function (BRDF). In digital holography, a diffuse surface can be modeled by assigning random phases to scene points of a flat surface [3], thus introducing a jittering with respect to depth on the order of $\pm \lambda/2$. From a signal processing perspective, this jittering leads to the presence of a uniform spectral response which causes a near-constant light emission in all directions, thereby making holograms of diffuse objects observable also under large viewing angles. This is because local spatial frequencies $\nu \leq (2p)^{-1}$ of fringe patterns are linked to viewing angles $\theta$, the wavelength
Fig. 1: Typical spectral amplitude distributions of holograms from (a) a diffuse and (b) a specular reflective object, as well as of (c) a natural image.

of the illuminating light $\lambda$ and the pixel pitch $p$, by the grating equation, $\lambda \nu = \sin(\theta)$.

Henceforth, the spectral amplitude distribution of holograms of diffuse objects is much denser than that of specular objects, see Fig. 1b, and is an almost homogeneous probability distribution over spatial frequencies as is shown in Fig. 1a. This strongly contrasts with natural image compression which fundamentally relies on the spectral amplitude distribution decaying as $\sim 1/f^2$ with spatial frequency $f$, see for example [4].

The sheer amount of pixels, as well as the processing capacities required by the non-locality of the data push current high-end computer systems to their limits. The development of visually near-lossless to acceptable lossy compression techniques for static as well as dynamic holographic content becomes a necessity [5,6]. In this paper, we focus on the compression of static, planar, monochromatic on-axis digital holograms of macroscopic, gray-scale scenes with diffuse surfaces represented as a complex-valued matrix.

We distinguish four fundamentally prominent approaches for hologram compression [6]:

1. methods that rely on hologram generation at rendering time and solely encode the required source data (point clouds, meshes, . . .);

2. methods that encode the complex-valued hologram in a general defocus (also: hologram/detector) plane;

3. methods that use numerical propagation schemes to refocus the hologram to the object-plane, where the parts in focus will then appear as natural image contents;

4. methods that use scene-aware representations, which combine knowledge about the scene geometry with the holographic principle of capturing 3D information in 2D.

Unfortunately the problem of compression has currently still not been solved up to satisfaction and no favorable wide-spread solution is available. While category 2) entails potentially the lowest amount of computations, the best compression may be obtained with conventional image codecs in combination with methods of category 3). Given an appropriate basis, the Fourier domain, which is the Fourier conjugate of the object plane, should offer as good compression-performance as object-plane compression, while having the same low number of computations as methods of category 2). Indeed, the Fourier plane representation has received considerable attention in the past because it can be directly recorded in optical synthetic aperture Fourier setups, which leverage its efficient space-bandwidth use [7]. The question of which plane to choose best for compression, is still ongoing in the scientific community. But until now, no efficient compression mechanism operating in Fourier domain has been established. By presenting a study of the signal characteristics of holograms, we will be able to connect all three representations and provide an efficient compression method of Fourier holograms. For deep holograms, we will recognize there biggest common challenge as that of finding a more efficient sparsifying transform of defocused content - initially motivating the present contribution.
A great deal of work on hologram compression in a general defocus plane has been based on state-of-the-art standards of image and video compression, and their extensions, applied to the real and imaginary parts or amplitude and phase separately [8, 9]. Leaving the design of the image coding techniques unaltered, \(i.e\). adhering to the \(\sim 1/f^2\) decay of spectral amplitude with spatial frequency \(f\), will in general be suboptimal on the holograms under consideration, which possess a rather homogeneous spectral amplitude distribution. For distinct hologram types most of the common non-adaptive wavelet bases have been investigated [10–13]. Some attempts to use full-packet decompositions and directional wavelet transforms [14] intrinsically lifted the former assumption of the \(\sim 1/f^2\) decay. Whilst this added directionality improved on the frequency localization, the localization remains rather poor for larger frequencies due to the phase-space tiling imposed by wavelet transforms. The idea of directional transforms has recently been picked up again in the formulation of an extension of the H.265/HEVC codec in [15]. Thereby the best suited directional discrete co-/sine transforms are learned on a training data set and then used in place of the separable 2D transforms. For specular computer generated holograms (CGH) no improvement was seen over vanilla H.265/HEVC unless training and compression data coincided for the most efficient data representations - which were the decompositions in real/imaginary parts and phase shifted distances. For a decomposition into amplitude and phase, considerable gains were achieved.

Other designs do not rely at all on sparsification by filterbank based wavelet or Fourier transforms. Instead they aim at reducing the dynamic range through quantization either non-uniform [16] or adaptive [17] thereby ignoring the special phase-space structure of holograms. Yet another different set of methods leverages exactly this and is based on Gabor frames and derivatives [13], which are both optimal localized in space-frequency with respect to the Heisenberg-Weyl uncertainty principle. These methods find often use only in view-dependent compression [18, 19], where only the apertures relevant for a specific perspective are encoded, due to their implicit over-completeness. Unfortunately, over-completeness is required for exact and numerically stable reconstruction from optimally localized complex-valued Gabor-like representations and hence necessitating further subsampling schemes for efficient compression.

With the current work, we investigate the wave atom basis [20] for the purpose of holographic data compression, which circumvents over-completeness by being only nearly optimally localized. We detail the transform, the integration into a full-fledged compression codec, and evaluate its performance on holographic test data. Wave atoms are very attractive due to their critically sampled representation at arbitrary scale and being computable using fast Fourier transforms (FFTs).

The remainder of this paper is organized as follows: We introduce wave atoms in Section 2, focusing on their useful properties in the present context. We subsequently introduce different hologram representations and explain why wave atoms perform well in Section 3. The proposed compression pipeline for holographic content is detailed in Section 4. In Section 5, we benchmark the proposed method against two variants of JPEG 2000 and against H.265/HEVC configured in intra-coding mode. Finally, a conclusion is drawn in Section 6.

2. Wave atoms

2.1. General theory

Wave atoms [20, 21] are a non-adaptive, non-standard Villemoes [22] construction of wave packets, that can be defined at any scale and are well localized in space and frequency domains as we shall show later on. In a 1D case at scale \(j\), centered around \(\pm \omega_{j,m} = \pm \pi 2^m\) in frequency domain and around \(x_{j,n} = 2^{-j}n\) in space with indices \(m, n\) respectively, the real-valued wave atoms are given as:

\[
\psi_{m,n}^j(x) = 2^{j/2} F^{-1}(\hat{\psi}_0)(2^j x - n)
\]  

(1)
with the symmetric, 2 bump “mother” wave atom is given as its scaled Fourier transform $\mathcal{F}$

$$\hat{\psi}_m^0(\omega) = e^{-i\omega/2} \left[ e^{i\alpha_m g(\epsilon_m(\omega - \pi(m + 0.5)))} + e^{-i\alpha_m g(\epsilon_{m+1}(\omega + \pi(m + 0.5)))} \right]$$

for $j \geq 0, n \in \mathbb{Z}, 0 \leq C_1 2^j \leq m \leq C_2 2^j$, arbitrary positive constants $C_1, C_2, \alpha_m = \pi/2(m + 0.5)$, and $\epsilon_m = (-1)^m$. $g$ is any real-valued $C^\infty$ bump function with compact support on an interval of length $2\pi$ that satisfies

$$\sum_m |\hat{\psi}_m^0(\omega)|^2 = 1,$$

and preserves ortho-normality via an asymmetric support condition

$$\forall \omega \in [-\pi/3, \pi/3]: \quad g(-2\omega - \pi/3) = g(\omega + \pi/2).$$

Note, that we will for the ease of speaking keep on referring to the wave atoms as being composed of symmetric bumps with respect to the frequency axes despite this slight and mandatory asymmetry. Typical candidates for $g$ are based on local cosine bases and one example of $g$ is

$$g(\omega) = \begin{cases} \left| s(\omega) - \frac{3\pi}{2} \right|, & \omega \in \left[ \frac{2\pi}{3}, \frac{4\pi}{3} \right] \\ 0, & \text{otherwise} \end{cases}$$

with

$$s(\omega) = \begin{cases} \frac{1}{\sqrt{2}} h\left(\frac{\omega}{4} + \pi\right), & |\omega| \in \left[ \frac{2\pi}{3}, \frac{4\pi}{3} \right] \\ \frac{1}{\sqrt{2}} h\left(\frac{\omega}{4}\right), & |\omega| \in \left[ \frac{4\pi}{3}, \frac{8\pi}{3} \right] \\ 0, & \text{otherwise} \end{cases}$$

$$h(\omega) = \begin{cases} \sqrt{2}, & t(\omega) \leq \frac{\pi}{2} \\ 0, & t(\omega) \geq \frac{3\pi}{2} \\ \sqrt{2} \cos \left( \frac{\omega}{4} - \frac{3\pi}{4} \cos (3t(\omega) - \pi) \right), & \text{otherwise} \end{cases}$$

$$t(\omega) = |((\omega + \pi) \mod 2\pi) - \pi|$$

To obtain the most general formulation for holography, we choose a uniform tiling of space and frequency domains in this work, which makes the concept of wave atoms very similar to Wilson
bases. Any scale might be selected, we select the golden middle between time and frequency, setting $j = \lfloor \log_2(\sqrt{N}/2) \rfloor$ for a signal of length $N$, whereas the factor $1/2$ accounts for the fact that there are two bumps, see Fig. 2b. In the discrete case this construction results in $2^{j-1}/N$ spatial translations and $2^j$ frequency modulations per dimension. This construction ensures good localization in space and frequency at any scale, as we shall detail in the following sub-section.

So far we considered the wave atom only in 1D. In the case of a uniform tiling, higher-dimensional transforms are obtained through standard tensor products of the 1D transforms, i.e. sequential applications of the 1D transforms per dimension.

Aside from the critically sampled ortho-basis defined above, other variants are derived by splitting up the pairs of bumps, in the frequency domain, into different elements. In 2D this results in: a complex-valued tight-frame with over-completeness factor 4, obtained by separating all 4 bumps into different frame elements; a real-valued tight-frame with over-completeness 2, whose elements each oscillate in only one direction and which is obtained by pairing bumps diametric to the origin; and the real-valued orthonormal transform, whose basis elements are a superposition of oscillations into two different directions and which is obtained by superposition of 4 bumps per element, see Fig. 2d.

2.2. Time-frequency localization

In the following, we will precise the vague notion of "good time-frequency localization". When choosing a properly normalized definition of the Fourier transform, the Heisenberg-Pauli-Weyl uncertainty principle [23] states that no function can exist that occupies less than 1 unit in time-frequency domain (TF unit). A single bump of a wave atom occupies at any scale $j$ about 1.22 TF units. While this is quite involved to show analytically due to the chosen piece-wise definition of the bump function $g$, it is easy to verify numerically. For each of the elements of the 1D orthonormal basis this gives about 2 TF units. For the 1D directional, or in this case equivalently complex representation, i.e. individual bumps, the over-completeness of the frame will be 2 while the TF units per element are halved. See also Fig. 2a and Fig. 2b for exemplary wave atoms in 1D next to an optimally localized Gaussian. In dimension $N$ analogous arguments hold and the elements of the orthonormal basis will occupy $\sim 2^N \cdot 1.22$ TF units, whilst the complex variant has $2^N$ as many frame elements of minimal time-frequency volume. The spatial and frequency footprints of an orthonormal wave atom in 2D are shown in Fig. 2, clearly visible in frequency domain are the 2 localized areas per dimension.

The provided freedom of access in the time-frequency domain guaranteed by this time-frequency localization might be visually interpreted as: wave atoms can sense directions by means of oriented oscillations on spatially localized areas.

As concluding remark, note that there exist more than one notion of uncertainty, and strict optimality according to one criterion does not automatically carry over to other notions. What is more, it is yet to be determined which criterion best reflects the perceptual quality as recognized by the human eye [24] and should therefore be used to gauge, for example holographic data during compression. Loosely speaking however, "good" according to one criterion remains "good" in another, as these are all different measures trying to capture the same general idea. We therefore do not need to worry too much about the 2D orthonormal wave atom transform being not strictly optimal with respect to the Heisenberg principle, but attempt to leverage its good localization and its lack of redundancy.

2.3. Coefficient structure & interpretation

Having dealt with theoretical properties of wave atoms, it is mandatory for coding purposes to have some understanding about the arrangement of the individual wave atom coefficients. Here we will outline the arrangement of the 2D orthonormal wave atom transform with a uniform space-frequency partition. Its coefficients are arranged by default in spatial clustering mode as
Fig. 3: Shows the wave atom coefficient organization of a 2D uniform, orthonormal wave atom transform in spatial clustering mode, by highlighting their structure in space and frequency.

shown in Fig. 3. The 2D arrangement has the same size as the original data and is divided into $2^j$ hogels per dimension, each of size $2^{-j_l} N_l$ with $N_l$ being the data-size in respective dimension $l \in \{1, 2\}$. In spatial clustering mode, the various hogels correspond to different frequency modulations. The lowest frequencies sampled in the top-left hogel, whilst moving down or to the right by one hogel, increases the sampled frequencies (symmetric to the respective frequency axes). Within one hogel the spatial support of the corresponding wave atoms varies, mapping to its location within the data on a $2^{-j_1} N_1 \times 2^{-j_2} N_2$ grid of disjoint blocks.

The frequency clustering mode is obtained by interchanging the roles of frequency and spatial shifts in Fig. 3 thereby obtaining $2^{-j_1} N_1 \times 2^{-j_2} N_2$ hogels, each corresponding to a fixed spatial region, and being composed of all possible frequency modulations for said region.

3. Various hologram types and their link to wave atoms

After the introduction of the wave atom theory as well as the possible coefficient organizations, we can finally establish the link with holography and show how and why the good time-frequency localization of the transform exactly benefits us.

The general problem, when compressing holograms, is that various representations of holograms differ in their space-frequency localization [6, 25, 26], and parameters such as the recording distance, the pixel pitch, or the wavelength take an influence. This behaviour is illustrated in Fig. 4, where the same hologram of a ball has been propagated to different distances. We refer to the three resulting hologram representations from top to bottom as: defocused hologram (also: Fresnel holograms), Fourier-plane hologram (short: Fourier holograms), and object-plane hologram.

The most general case are defocused holograms and it is the hardest to compress due to its inefficient use of the phase space. To understand this, consider Fig. 4a. The dependence of the slant of the signal depends from the distance $z$ object to hologram plane is $z^{-1}$ within the Fresnel approximation. Backpropagation of the hologram corresponds to shearing the hologram in phase space such that, scene parts in focus appear spatially limited with potentially all frequencies excited, as seen in Fig. 4g. Capturing more of the signal in a given defocused hologram requires an extension of the sampled phase-space both in space (horizontal) and in frequency (vertically). Sampling higher frequencies is, due to Nyquist-Shannon, possible only with a smaller pixel pitch or equivalently by extension in the Fourier domain. If we wish to enlarge the sampled spatial domain, we need an extension in the spatial domain as well, making this representation overall
Fig. 4: Generality of the sparsification of the "Ball" hologram under the wave atom transform at various distances of the hologram plane. The top, middle, and bottom row correspond to the same hologram recorded in a general out-of-focus position, at infinity (Fourier-type), and in the front object plane. The columns correspond from the left to the right correspond to an $x$-$f_x$ slice of phase space, and wave atom coefficients with spatial and frequency clustering respectively.

very inefficient with respect to used signal bandwidth. For very deep scenes or scenes with objects located at multiple depths this representation will however be unavoidable, as parts of the scene will always appear out of focus.

An efficient representation of defocused holograms can be obtained, mathematically speaking, only by automatic support detection in phase space, e.g. through a transform with a uniform tiling of phase space as provided by the proposed variant of the wave atoms. Visually the sparsification of defocused holograms by wave atoms can be understood as follows: with the confined support
of the individual wave atoms acting as small apertures, conceptually similar to pinhole cameras, the depth-of-focus is increased that much that always parts of the scene appear in-focus, which can be leveraged in the frequency clustering mode.

For sufficiently shallow scenes, a more efficient use of the signal space-bandwidth product can be obtained with the other two hologram types and higher compression ratios can be obtained, as empty regions in phase-space could be either manually cropped, or efficiently coded by a reproducing transform in combination with entropy coding.

Fourier holograms, shown in the second row, are focused at a hologram plane that is placed at \( z = \infty \). For the sparsification through wave atoms, the same reasoning as before holds but an even higher sparsification may be achieved by leveraging symmetries within the object through the symmetric frequency bump design of the wave atom transform. As the data itself is, in this case, already present in the Fourier plane, while applying the transform, the scene will be sampled spatially by these symmetric bumps. Due to the clustering by “frequency”, all four quadrants of the actual scene will be in a exactly reversible manner folded on top of each other with the scene center located at the top-left corner. For scenes, such as this ball appearing centered in the scene this will result in extraordinarily high sparsity. For any general scene, the degree of sparsification will be lower but still all possible symmetries will be intrinsically leveraged.

The third type of holograms are object-plane holograms, where the hologram plane is located within the scene at \( z = 0 \), as shown in the bottom row. These holograms appear for shallow scenes similar to images, but remain complex-valued. Wave atoms sparsify these holograms, if coefficients are clustered spatially. This appears as the natural choice, as each hogel will cover all frequencies located within the same spatially confined region. We shall see this representation being the most suitable to combine with conventional image compression as the hologram is well confined in space and spatial sparsity can be leveraged. Naturally, decompositions that focus predominantly on leveraging spatial sparsity will outperform any orthonormal representation well localized in time and frequency, such as wave atoms. Furthermore, any dismissal of high-frequency content, as common for conventional image compression, will not lead to visible artifacts of the hologram, as long as perspective and reconstruction distance remain unchanged. These special properties of this hologram representation will decrease rapidly for scenes that are not spatially sparse in any object-plane anymore due to out-of-focus components. Object-plane holograms differ from Fourier holograms solely by one FFT.

A perspective reconstruction can be obtained only from Fourier or defocused holograms by masking the respective parts of hologram with an aperture smaller than the hologram prior to backpropagation. For example, using a square aperture of half the hologram size on the bottom left, returns a bottom left view offset from the center by approximately a quarter of the respective horizontal and vertical field of view. It is also worth discussing there respective phase-space characteristics. For defocused on-axis holograms, side or corner views are the result of interferences from the higher spatial frequencies, see Fig. 4a, whereas the center view involves only low spatial frequencies. For Fourier holograms, that have been recorded with respect to a spherical reference point source in the objects middle plane, this assessment is not true anymore. The spherical, or in the near-field approximation quadratic, phase shift typically introduced by propagation is largely annihilated by the curvature of the reference wave, as only differences with respect to the reference are encoded in the hologram [27]. This is the idea of the compact space-bandwidth representation (CSBP) proposed in [26]. On a high level, the modulation with a suitable quadratic phase factor is nothing else than the first step of a conventional Fresnel transform, implemented with the help of the Fourier transform. This results in a near-uniform distribution of frequency content over the whole hologram, see Fig. 4d. Object-plane holograms require aperture masking in an defocused or Fourier plane, because in these holograms the global scene information is not contained in each spatial sub-selection as opposed to Fourier or defocused holograms.
Since the objective of this paper is to study general hologram compression, we are reluctant to observe that the wave atom transform can sparsify holograms recorded at any distance and any perspective part, if either coefficient clustering mode can be accessed. In other words, even complex and deep scenes, where only parts can be in focus, are natively sparsified by this transform unless the entire phase-space is covered and additional thresholding is required. It is only transforms of good time-frequency localization that allow for agnostic detection of the occupied support in phase space for any given hologram.

Notable examples of other optimally localized transforms are almost exclusively windowed short-term Fourier transforms (STFT), including Gabor transforms. These are either poorly localized in time or frequency domain, such as is the case of the regular STFT, which has the binary box filter as spatial window and suffers therefore from a poor localization of the sinc window in the Fourier domain. Or the transforms are required to be over-complete to ensure a stable reconstruction, as is the case with Gabor(-like) transforms, which use Gaussian(-like) windows in either domain. Any critically sampled representation with truly optimal time-frequency localization will be highly unstable (have poorly localized dual windows) as described by the Balian-Low theorem [23]. The wave atom transform circumvents this exactly by the Villemoes construction of joining $2$ bumps per dimension to obtain a good, but not optimal, time-frequency localization and which is similar to the idea of Wilson bases [28].

4. Wave atom coding (WAC)

After we have developed a better understanding of wave atoms and their interplay with holograms in the previous sections, we now detail the proposed coding scheme, termed Wave Atom Coding (WAC) (algorithm 1) which is applied for real and imaginary parts independently. The core coding architecture is based on the wave atom transform and the bit modeling and entropy coding (EBCOT) of a JPEG 2000, part 2, coding engine [29], as illustrated in Fig. 5. The major advantage of the JPEG 2000 part 2 pipeline is that it provides through the bit-plane modeling an easy way to handle input data with $>16$ bits, such as the raw wave atom coefficients.

The encoding process begins with separating the complex-valued hologram into two real-valued channels. Two prominent choices exist to represent the complex valued data: Cartesian decomposition in a real and imaginary part, or polar decomposition in an amplitude and phase part. Although mathematically equivalent, the polar form was found to be severely impacted by already minor distortions such as compression artifacts from lossy compression; with the
Amplitude being less important than phase information. This behavior is already known from image processing [30] and persists for holograms [15], where the phase encodes among others the depth information. The algebraic representation is thus much less prone to error as it avoids these non-linearities, for example due to phase wrappings. Henceforth, we split the holograms in all our experiments into real and imaginary parts, and process them independently further on.

Next, the 2D wave atom transform [31] is applied to each part. If the data has an odd number of pixels per dimension, one zero row or column has to be padded accordingly. If the data has a square aspect ratio the 2D transform from the wave atom package can be used directly, otherwise all our experiments into real and imaginary parts, and process them independently further on.

As mentioned earlier, holograms can consist of $\geq 100$ Mpixel, hence an efficient transform is vital. It is therefore important to note that the wave atom transform can be implemented with FFTs accounting for the main computational complexity. The complexity of the 2D transform is $O(N^2 \log(N))$ with $N$ samples per dimension. In practice the overall calculation effort is about $5 - 10 \times$ that of a full-size FFT. An existing GPU implementation [31] achieves a $> 1000 \times$ speed-up on a Nvidia GTX 1080Ti (tested with random data of size 8192 x 8192).

To ensure best coding efficiency a test is performed to decide whether a reorganization of the coefficients, as detailed in sections 2.3 and 3, can lower the variance per code-block. A smaller variance will result in fewer required bitplanes per code-block and hence higher bitrate distortion ratio. One possible criterion is given in algorithm 1, where the standard deviation of each hogel is calculated and the ordering with the lowest median standard deviation is selected. It does suffice to test only the real part of the hologram, as both parts follow the same statistics.

---

**Algorithm 1** Wave Atom Coding (WAC) encoding procedure

1: procedure WACencode(Re/Im part of Hologram $X \in \mathbb{R}^{N_x \times N_y}$, Target bitrate $b$, Scale $j_1, j_2$)
2:    for $k \in \{1,2\}$ do
3:        $n_T(k) \leftarrow 2^{-j_k} N_k$  \hspace{1cm} \text{Number of different translations}
4:        $n_M(k) \leftarrow 2^{j_k}$  \hspace{1cm} \text{Number of different modulations}
5:    for row in $H$ do
6:        $X(row,:) \leftarrow \text{ForwardWA1D}(X(row,:))$  \hspace{1cm} \text{Forward 1D wave atom transform.}
7:    for col in $H$ do
8:        $X(:,col) \leftarrow \text{ForwardWA1D}(X(:,col))$
9:    if $\text{LocalStdMedian}(X,n_T,n_M)$  \hspace{1cm} \text{Change spatial to frequency clustering}
10:       $X \leftarrow X'$
11:       $X \leftarrow Q(X)$  \hspace{1cm} \text{$Q$ scalar quantization operator...}
12:       bitstream $\leftarrow \text{EBCOT}(X,b)$  \hspace{1cm} \text{RD-opt. handled within JPEG 2000 pipeline.}
13:       return bitstream  \hspace{1cm} \text{JPEG 2000, pt. 2 conform bitstream}
14:    end if
15: end for

16: procedure $\text{ChangeClustering}$(Data $y$, current Block-size $B_s$, target Block-size $B_t$)
17:    for $r = 1 : B_t(1)$ do
18:        for $c = 1 : B_t(2)$ do
19:            for $i = 1 : B_s(1)$ do
20:                for $j = 1 : B_s(2)$ do
21:                    $x(r+(i-1)\times B_s(1),c+(j-1)\times B_s(2)) \leftarrow y((r-1)\times B_t(1)+i,(c-1)\times B_t(2)+j)$
22:                end for
23:            end for
24:        end for
25:    end for
26:    procedure $\text{LocalStdMedian}$(Data $x$, Block-size $B_s$)
27:    for Disjoint block $B$ with id $b$ of size $B_s$ in $x$ do
28:        $\sigma(b) \leftarrow \text{variance}(B)$
29:    end for
30:    return median($\sigma$)
In the final WAC encoding step, the signed float-value data is fed to the JPEG 2000 part 2 pipeline, which quantizes the data and performs entropy coding as well as rate-distortion optimization through bit-plane coding in the EBCOT stage [32]. Because the wave atom transform coefficients are already appropriately weighted, no additional subband weights have been applied, as would be required when deploying a CDF 9/7 wavelet transform. After quantization with a uniform, scalar dead-zone quantizer with optimized step-size, the data is split into code blocks, according to the present hogel size, and entropy coded using context-based binary arithmetic coding per code block. In a penultimate step the bit-plane coding is commenced, which encodes most significant bits first and proceeds to bits of lesser significance. A global rate distortion optimization yields the final code stream.

The main difference related to the default JPEG 2000 encoder is the replacement of the 2D CDF 9/7 biorthogonal wavelet transform by the wave atom transform including a potential reorganization step. The WAC backend and the JPEG 2000 encoding are based on a standard compliant implementation of JPEG 2000 by Bruylants et al. [33]. The low computational complexity and simplicity, of JPEG 2000 compared to H.265/HEVC, enabled an undisturbed comparison of the WA basis with highly spatially localized wavelets. By adding 1 bit to the code stream in our design, WAC could be made to default back to regular JPEG 2000 by merely replacing the sparsifying transform.

5. Numerical experiments

In this section the proposed Wave Atom Coding (WAC) is compared with the conventional image compression codecs JPEG 2000 and H.265/HEVC, as well as a JPEG 2000 adaptation [34] for holography using a full-packet decomposition and directional wavelets. We will study four macroscopic holograms, each compressed in each of the hologram representations discussed in section 3.

5.1. Test data

The four tested holograms have a spatial resolution of 2048 × 16384 pixels and were initially recorded or generated in a Fourier geometry with monochromatic light of 532 nm. The scenes measure 5 – 12 cm in width and height, and have a depth profile according to Table 1. All scenes are gray-scale in object space.

<table>
<thead>
<tr>
<th>Hologram</th>
<th>Ball</th>
<th>Sphere</th>
<th>Chess</th>
<th>DeepChess</th>
</tr>
</thead>
<tbody>
<tr>
<td>central depth ±δ in cm</td>
<td>72.6 ± 2.5</td>
<td>96.0 ± 3.2</td>
<td>64.9 ± 31.6</td>
<td>100 ± 60.0</td>
</tr>
</tbody>
</table>

Table 1: The distance between hologram and center plane of the objects of the tested holograms.

We selected: a computer-generated hologram (CGH) of a perforated Ball of comparatively small depth; an optically recorded hologram (OH) of a physical realization of the same 3D model, hereafter called "Sphere"; as well as a series of two computer-generated chess scenes with moderate and large depths.

The "Ball", "Chess", and "DeepChess" holograms were synthesized from point clouds with the method described in [3]. As a result 16k × 16k plane-wave on-axis holograms of objects with diffuse reflectiveness were obtained. In a subsequent processing step the plane-wave holograms were clipped to 2k × 16k and modulated with a spherical reference wave (with focus at 70cm for "Ball", "Chess"; 100cm "DeepChess") to match the optical test data. The holograms "Chess" and "DeepChess" are sufficiently deep, such that with reasonably large apertures not all rows of chess
figures can be brought into focus simultaneously. We will therefore consider the succession of "Ball", "Chess", "DeepChess" to understand trends arising from increasing depth.

The "Sphere" hologram was captured from a 6.5cm object using the green channel of a Basler piA2400-12gm with 3.45µm pixel-pitch in an off-axis lens-less synthetic aperture Fourier holography setup as described in [7, 27]. It was then converted to an on-axis geometry by numerical filtering of the +1 diffraction order. Thereat, the dynamic range was reduced by filtering out all optical zero-order components using Hann windows.

The assembled Fourier holograms were transformed to defocused holograms by demodulation with their respective spherical wave, and converted to object-plane holograms by means of a Fourier transform.

To enable reuse of components from natural image compression codecs, as well as to allow for a fair comparison across the codecs, we further pre-processed the data as follows: (1) removing the DC term from these floating-point holograms; (2) clipping their range at $\approx 3.35\sigma$ with standard deviation $\sigma$; (3) finally the real and imaginary parts were jointly 8-bit quantized using a signed 7-bit representation - yielding 16-bit per complex-valued pixel (bpp). Clipping in (2) is necessary due to an extremely long tail of the coefficient histogram caused by intensity dependent speckle noise. The particular clipping point jointly minimizes the total error induced by both quantization and clipping, assuming Gaussian phase noise on the object surface recorded by the hologram.

5.2. Tested codecs

WAC was largely implemented in Matlab making use of the wave atom toolbox [31]. For the required JPEG 2000 encoding backend, as well as the other tested variants of JPEG 2000, a standard compliant implementation of Bruylants et al. [33] was used. For WAC the code block size was equal to the hogel size, which was $64 \times 256$ in the case of spatial clustering and $32 \times 64$ in the case of frequency clustering. The vanilla JPEG 2000 compression relied on a four-level CDF 9/7 decomposition and a code-block size of $64 \times 64$. The extended JPEG 2000 (FP+D) [14] compression used a 4-level full-packet decomposition and adaptive directional wavelets without any further arguments. H.265/HEVC coding was done with the Fraunhofer implementation (HM 16.19 [35]) and configured as follows: main-RExt profile with level 8.5, intra period of one and 4:0:0 input chroma format, whilst all other options have been left at their default.

On a practical note, WAC implemented in Matlab, is about as fast as vanilla JPEG 2000, both being about $\geq 7 - 10 \times$ faster than H.265/HEVC due to their much smaller number and range of intrinsic parameters.

5.3. Experimental analysis

We tested the bitrate-distortion performance of the named methods for holograms represented at the Fourier-, object-, and a general defocus plane. The analysis will be stated in terms of PSNR scores of the compressed versus uncompressed complex-valued wavefields across the range of $[0.25, 2]$ bpp. All references to bpp are with respect to complex-valued pixels. For example 0.5 bpp correspond to 0.25 bpp per real/imaginary part.

Fig. 6 shows explicit bitrate distortion curves for the considered holograms. We omitted JPEG 2000 FP+D for clarity of presentation but will state the results in terms of Bjøntegaard-Delta PSNR (BD-PSNR) in Table 2. As reference for the BD-PSNR scores (in dB), we will always use WAC in the respective representation, such that positive values show a dominance of WAC, whilst for negative values, the chosen alternative is being more efficient.

For each method and hologram, the most efficient compression is achieved when compressing in the object-plane representation, which is clear as H.265/HEVC, the JPEG 2000 variants, as well as the EBCOT entropy coding stage within WAC (provided the spatial coefficient clustering) were tuned to best leverage any existing spatial sparsity. The second most efficient representation
is in the Fourier-plane, whilst general defocus holograms are the least efficient to encode. The reasons were discussed in section 3.

Evidently, the compression performance of deeper CGH and OHs is on a comparable range, whereas the comparatively shallow CGHs offer itself readily for sparsification, resulting in 5 – 10 dB gains over the OHs for some methods and representations. Whilst for deep CGH scenes the occupied phase-space volume and therefore required bitrates increase, the bitrates of OHs are bloated by incoherent measurement noise evenly spread across the entire phase space - this is completely independent of coherent speckle noise, which affects bitrates only in the object-plane representation.

<table>
<thead>
<tr>
<th>Hologram</th>
<th>Method in OP/FT/DP</th>
<th>H.265/HEVC</th>
<th>JPEG 2000</th>
<th>JPEG 2000 FP+D</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Ball&quot;</td>
<td></td>
<td>-4.2 / 3.9 / 1.7</td>
<td>-1.2 / 7.9 / 4.2</td>
<td>-0.4 / 6.0 / 3.2</td>
</tr>
<tr>
<td>&quot;Sphere&quot;</td>
<td></td>
<td>-1.3 / 1.2 / -0.1</td>
<td>-0.2 / 2.5 / 1.5</td>
<td>0.8 / 1.9 / 0.9</td>
</tr>
<tr>
<td>&quot;Chess&quot;</td>
<td></td>
<td>-2.1 / 1.9 / 0.8</td>
<td>-0.4 / 4.8 / 2.9</td>
<td>0.7 / 4.0 / 2.2</td>
</tr>
<tr>
<td>&quot;DeepChess&quot;</td>
<td></td>
<td>0.0 / 1.3 / 1.1</td>
<td>1.7 / 3.3 / 3.0</td>
<td>1.9 / 2.3 / 1.8</td>
</tr>
</tbody>
</table>

Table 2: BD-PSNR scores with respect to WAC in object-plane (OP) / Fourier-plane (FT) / defocused plane (DP) representations respectively. Positive scores (bold) show where WAC is advantageous.
Upon closer investigation the best overall performance is obtained with H.265/HEVC applied in the object-plane. The gap to WAC used on the object-plane representation declines with increasing depth of the CGH from $-4.2 \text{ dB}$, to $0.0 \text{ dB BD-PSNR}$. That means that the insufficient frequency localization of the local discrete co-/sine transforms employed in H.265/HEVC, see discussion on STFT with spatial binary box filter in section 3, eventually manages to eat up all the gains achieved by the more advanced set of coding tools, such as intra-block prediction, and a context richer entropy coding stage.

Comparing WAC with JPEG 2000, the effects of poor frequency localization are even more pronounced. Whilst the latter only outperforms WAC in the object-plane representation, the BD-PSNR gains are smaller for shallow CGHs and then decline as fast as for H.265/HEVC with depth. This leads to marginal losses of WAC for moderately deep CGH or OH in object-plane representation. The PSNR for JPEG 2000 FP+D in object-plane representation is smaller than that for vanilla JPEG 2000, indicating that the object-plane representation does not warrant the additional overhead. However, although JPEG 2000 FP+D was not designed for diffuse macroscopic holograms, it is still better suited than vanilla JPEG 2000 for any hologram encoded in a defocus or Fourier plane representation, even though remaining considerably inferior to WAC. We may conclude that except for shallow CGH, the partitioning of phase-space by biorthogonal CDF 9/7 wavelets is not advantageous, no matter if directionality and/or full-packet decomposition is employed or not despite the fact that until recently CDF 9/7 wavelets were considered the best-suited filterbank based wavelets for sparsification of smooth wavefields, see [36] and references therein.

This underpins the hypothesis of our paper that for considerably deep scenes CDF 9/7 wavelets or rectangular-windowed local discrete co-/sine transforms should be replaced by spatially better localized transforms such as the wave atom transform. Furthermore, we can clearly see that WAC outperforms any alternative in either a defocused or the Fourier plane representation by considerable margins ($0.8 - 7.9 \text{ dB BD-PSNR}$) with but one exception. The "Sphere" OH in the defocused plane is compressed by H.265/HEVC about as good as with WAC, due to a drastic spread of incoherent measurement noise in phase space in this representation. In case the entire phase-space is occupied, the importance of efficient prediction and entropy coding stages is increased over that of a good sparsifying transform.

While our initial motivation was to find a more efficient transform for defocused holograms, we realized that WAC is extremely efficient in compressing Fourier holograms, outperforming even H.265/HEVC, see 2. In this representation, only for deep scenes or very low bitrates the performance gap between WAC and H.265/HEVC shrinks, due to H.265/HEVC’s more modern entropy coder offering many more coefficient contexts.

We found that PSNR scores on the reconstructions from center and right corner view, differed the most, when compression was done in the object-plane. Thereat the center view showed on average $2 - 5 \text{ dB}$ more than the corner views, whilst for out-of-focus representations the difference was less than $1 \text{ dB}$ on average. Therefore, one might consider compression in the Fourier domain, as the second most effective representation, which at the same time offers similar PSNR scores across all perspectives.

6. Conclusion

In this work we introduced a sparse decomposition of various holograms in wave atoms, prior to compression with the EBCOT encoder. The near-optimal space-frequency localization of wave atoms basis functions allows efficient sparsification of complex-amplitude holograms. We introduced the Wave Atom Coding (WAC), as a combination of wave atom transform with the JPEG 2000 quantization/entropy coding stage. Experiments compared the wave atom transform with CDF 9/7 wavelets, H.265/HEVC, as well as with a full-packet, directional wavelet extension to JPEG 2000. Four typical holograms were studied by compression in object-plane (OP),
Fourier-plane (FT), or defocused plane (DP) representations. We find that compression in the object-plane, which usually entails an expensive numerical backpropagation from the sensor, is best done with conventional codecs such as H.265/HEVC for shallow or moderately deep scenes. In deep scenes, WAC is in par with H.265/HEVC and outperforms JPEG 2000 variants by 1.7 – 1.9 dB PSNR. Optically directly accessible Fourier holograms, which are space-bandwidth efficient for shallow and moderately deep scenes, are best encoded by WAC - no matter the scene depth. We report BD-PSNR gains of 1.2 – 7.9 dB over all tested methods. For defocused holograms, which are the least space-bandwidth efficient, WAC outperforms the JPEG 2000 variants by 0.9 – 4.2 dB BD-PSNR and ranges from −0.1 to 1.7 dB PSNR with respect to H.265/HEVC.

Considering the computational burden, WAC encoding is about as fast as JPEG 2000 and therefore, much faster than H.265/HEVC. Run-time efficiency is important towards real-time compression and transmission of large scale digital holograms. Future extensions of the WAC encoding could trade speed for higher compression rates by using entropy coders such as CABAC used in H.265/HEVC or address color-channel de-correlation for RGB holograms.

Methods achieving a quarter of these bitrates at visually near-lossless quality, i.e., compression of about 16 : 0.25 with respect to the considered 8 + 8 bpp holograms, would be required for future applications. It can be projected that these methods will not entirely rely on a sparsification of the 4D space-frequency domain, but instead involve source data manipulations.

Acknowledgements

We thank Athanasia Symeonidou for providing us the computer-generated holograms used in this paper.

Funding

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

References