Signal processing challenges for digital holographic video display systems
Blinder, David; Ahar, Ayyoub; Bettens, Stijn; Birnbaum, Tobias; Symeonidou, Athanasia; Ottevaere, Prof. Dr. Ir. Heidi; Schretter, Colas; Schelkens, Peter

Published in:
Signal Processing: Image Communication

DOI:
10.1016/j.image.2018.09.014

Publication date:
2019

License:
CC BY

Document Version:
Final published version

Link to publication

Citation for published version (APA):
Signal processing challenges for digital holographic video display systems

David Blinder a,b,∗, Ayyoub Ahar a,b, Stijn Bettens a,b, Tobias Birnbaum a,b, Athanasia Symeonidou a,b, Heidi Ottevaere c, Colas Schretter a,b, Peter Schelkens a,b

a Dept. of Electronics and Informatics (ETRO), Vrije Universiteit Brussel (VUB), Pleinlaan 2, B-1050 Brussels, Belgium
b imec, Kapeldreef 75, B-3001 Leuven, Belgium
c Dept. of Applied Physics and Photonics, Brussels Photonics Team (B-PHOT), Vrije Universiteit Brussel (VUB), Pleinlaan 2, B-1050 Brussels, Belgium

ARTICLE INFO

MSC:
00-01
99-90

Keywords:
Digital holography
Computer generated holography
Holographic visualization
Holographic imaging
Data compression
Quality assessment

ABSTRACT

Holography is considered to be the ultimate display technology since it can account for all human visual cues such as stereopsis and eye focusing. Aside from hardware constraints for building holographic displays, there are still many research challenges regarding holographic signal processing that need to be tackled. In this overview, we delineate the steps needed to realize an end-to-end chain from digital content acquisition to display, involving the efficient generation, representation, coding and quality assessment of digital holograms. We discuss the current state-of-the-art and what hurdles remain to be taken to pave the way towards realistic visualization of dynamic holographic content.

1. Introduction

Holography is a technique that can record and reconstruct the full wavefield of light. Its invention in 1948 is attributed to Denis Gabor, who initially developed the technique for improving the quality of electron microscopes [1]. At the time, no adequate coherent light source was available, forcing him to arrange everything along one axis, known today as in-line holography. Because image quality was poor, holography remained obscure. With the invention of the laser and Leith and Upatnieks’ off-axis holography [2] in the 1960s, high-quality holograms became possible.

However, holography was still purely analogue, since everything had to be recorded on photographic film. This did change in the late 1980s: with the emergence of Spatial Light Modulator (SLM) technologies, Charged-Coupled Device (CCD) image sensors and increasingly powerful computers, it became finally possible to digitally capture, process and display holograms. Digital holography is used today for various purposes, such as microscopy [3], interferometry [4], surface measurements [5,6], storage [7] and three-dimensional (3D) display systems [8]. In this paper, we particularly focus on the realization of the latter.

3D displays enable depth perception, thereby showing a scene in its 3 dimensions. Contrary to conventional two-dimensional (2D) displays or stereoscopic displays, which are only providing a monoscopic or a stereoscopic view, 3D displays provide different visual information depending on the viewer’s eye position and gaze. This technology is an essential component in many visualization applications in entertainment, medicine and industry.

Current solutions for 3D displays like those based on optical light fields [9,10] can only provide a subset of the required visual cues due to inherent limitations, such as a limited amount of discrete views. By contrast, digital holography [11] is considered to be the ultimate display technology since it can account for all visual cues, including stereopsis, occlusion, non-Lambertian shading and continuous parallax. Representing the full light wavefront avoids eye vergence–accommodation conflicts [12].

However, current holographic displays do not yet offer the resolutions and angular Field-of-View (FoV) required for acceptable visual quality — but this may change soon. Although significant technological hurdles still need to be overcome, steady developments in photonics, microelectronics and computer engineering have led to the prospect to realize full parallax digital holography displays with acceptable rendering quality [13,14]. The extreme required image resolutions, paired with the video frame rates needed for dynamic video holography could lead up to terabytes-per-second data rates in its plain form [15].

Much of prior research efforts on holographic display systems focused on the hardware challenges, namely designing and developing the relevant optical and electronic technologies. Examples include high-bandwidth Spatial Light Modulators (SLM) and specialized optics [16–18]. Thorough overviews of the latest developments of 3D displays can be found in recent papers [8,14,19–21].
An aspect which has received much less attention by comparison to hardware designs is how the potentially huge data volumes should be handled for efficient generation, storage and transmission. The signal statistics of a hologram differ considerably from regular natural image and video. Hence, conventional representation and encoding algorithms, such as the standard JPEG and MPEG families of codecs, are suboptimal. When generating synthetic holographic content by deploying Computer Generated Holography (CGH) techniques, algorithms are required that differ considerably from those needed for classical synthetic image rendering that are based on ray-space physics and a particle model for light transport. Indeed, holography adheres to a wave-based transport model of coherent light instead.

This overview paper focuses on the signal processing challenges needed for the realization of holographic video systems from source signal generation to display. We identify challenges in hardware and software and collect a set of important research targets to address them. The steps that need to be undertaken are illustrated in Fig. 1; a global comprehensive approach is required for designing a consistent holographic video display system.

The body of this document is organized as follows:

- **Digital holographic display systems** are currently very premature and heterogeneous in design, so there exists no established standard that specifies how to supply holographic data to the display. Hence, it is important to have a signal processing framework that is sufficiently generic as to allow for compatibility across display systems (see Section 2).
- **Recording digital holograms** at high resolutions is difficult. Not only do these acquisitions need specialized optical setups, but they require expertise to build and operate. Furthermore, the size of objects that can be recorded is limited (see Section 3).
- **Computer-generated holography (CGH)** is much more calculation-intensive than classic image rendering: every point in the scene can potentially affect every hologram pixel. This many-to-many relationship compounded with the large hologram resolutions is too costly to compute by brute-force (see Section 4).
- **Novel transforms and coding technologies** are needed for digital holograms. Conventional transforms and representations perform suboptimally because they do not match the statistical properties of holographic signals. Standard codecs such as JPEG and MPEG are solely designed for natural photography and video but operate poorly on holographic content; novel transforms and motion estimation & compensation techniques need to be integrated (see Section 5).

- **Modelling perceptual visual quality** for holographic content requires accurate models for the subjective quality assessment. These metrics will be needed to steer the other components of the holographic pipeline (CGH, coder/decoders, displays) as to optimize the global quality of experience (see Section 6).

It is important to understand that all challenges outlined above are further complicated by the large computational load and memory/signalling bandwidth associated with the realization of full-parallax, wide viewing angle dynamic digital holographic display systems [22]. Before digging deeper into the physics of holography (Section 3), we first address the high-level properties of holographic displays to better position the challenges that are faced by the modules that are processing the holographic signal prior to display.

### 2. Holographic displays

At present, most commercially available 3D displays are stereoscopic: two different 2D images are fed to the viewer, one per eye (see Fig. 2(a)). This is either achieved with auto-stereoscopic displays or with specialized eye wear, e.g. with passive glasses (anaglyph, polarized), active glasses (shutters), or Head-Mounted Displays (HMD). However, the 3D effect is still limited, because several cues of the human visual system are not accounted for. Such displays lead to eye strain and nausea after prolonged wear. A mismatch between the vergence and accommodation of the eyes during viewing leads to the “vergence-accommodation conflict”, which results in discomfort, see Fig. 3.

The important cues needed for 3D perception [19,23] can be divided into two categories: monocular and binocular. Monocular cues are:

- **Accommodation** — controlled by muscle’s strain for adjusting the eye’s focal length;
- **Motion parallax** — the observed relative velocity of objects with respect to the eye’s position;
- **Occlusion** — indicate the relative distance of objects to the viewer due to the observed partial overlap;
- **Shading** — providing visual information on the shape and 3D position of objects.

Binocular cues are:

- **Vergence** — the angular difference between the gaze direction of the eyes as a function of viewing depth;
- **Stereopsis** — emerging from correspondences between the projected image on the retina of both eyes.
Fig. 2. Different 3D display types: (a) stereoscopic display, (b) multi-view or light-field display, (c) holographic display.

Fig. 3. Illustration of the vergence–accommodation conflict. In (a), e.g. in holographic displays, the accommodation and vergence always match up. In (b), e.g. in stereoscopic displays, accommodation is constrained at a fixed focal distance yielding a mismatch.

Many types of 3D display technologies exist that attempt to better address the important cues with varying degrees of efficacy. 3D displays come in many forms, such as volumetric displays [24] and multi-focal displays [25,26]. One of the most popular are multi-view (or light-field) displays. They generalize the stereoscopy approach by increasing the amount of 2D views from two to many more (Fig. 2(b)). These displays [27,28] discretize the light field into many views emitted in multiple directions.

Because of the current physical limitations in terms of screen resolution, and consequently on how many views can be displayed at once, this will degrade the viewing experience. The main noticeable deficits are the step-wise changes in the motion parallax, and accommodation–vergence mismatches. Furthermore, typically only horizontal parallax is supported.

Multi-view displays come in many forms, including 360° tabletops [28] and HMDs [29–31]. It has to be noted though that in principle, future light field displays which exhibit sufficiently high spatial and angular resolutions enable the super multi-view principle [32,33]: this happens when the angular sampling pitch is small enough so that multiple rays emanating from the same screen point enter the viewer’s pupil, hence alleviating the above deficits. Nonetheless, the ray bundle size dictates a trade-off between the maximal spatial and angular resolution, which is not present in holography [34].

However, holographic displays have the promise to account for all visual cues. In principle they can display virtual objects indistinguishably from real ones. But, realizing high-quality holographic display systems comes with many software and hardware challenges.

2.1. Types of holographic displays

We will focus on dynamic digital pixel-based holographic displays, also known as “electro-holographic displays”. Contrary to e.g. printed or analogue film holograms, they can be electronically updated at video rates. Further guidance on holographic displays and alternatives can be found in the introductory paper from Dodgson [35], the recent summary [34], or the elaborate review [20].

One of the first digital holographic displays was developed by MIT Media Lab [36]. It utilizes bulk acousto-optical modulators to create successive horizontal-parallax-only line holograms for bandwidth reduction. From then on, electro-holographic displays have come in many shapes and sizes, varying widely in resolution, FoV and image size. Today’s SLM market is dominated Liquid-Crystal On Chip (LCOS) and Micro-ElectroMechanical Systems (MEMS) elements [17].

Although a true holographic display should modulate both the amplitude and phase of a wavefront, full complex modulation in a single device is still hard to achieve, but some examples exist [37]. Most SLMs can either modulate amplitude or phase [17]. The latter is often preferred, since phase-only SLMs have the advantage of high diffraction efficiency, which theoretically can reach 100%. The phase component of a hologram is also generally more important for image quality than the amplitude.

Current commercially available SLMs typically have resolutions of up to $1920 \times 1080$ (Full HD) or $3840 \times 2160$ (4K UHD), which do not yet reach the resolutions needed for large, full-parallax realistic displays: depending on the screen dimensions and viewing angle requirements, the needed resolution can surpass $10^{12}$ pixels/frame. Moreover, pixel pitches of commercial solutions are currently limited to approx. 4 μm, cf. Fig. 4; the pixel pitch being the distance between the centres of two adjacent pixels. More details are given in Section 2.3. Hence, engineers need to be resourceful with the available bandwidth. This has led to several types of holographic displays, which we can categorized in three groups (see Figs. 5 and 6). The three following sections describe each of them.
2.1.1. Head-mounted displays (HMD)

HMDs have applications in personalized virtual and augmented reality. Holographic head-up displays have already been proposed early on for aircrafts [45], cars [46] or to aid medical diagnostics and surgery [47]. More recent usage scenarios [48] involve human–machine interfaces for visual content creation, computer games, CAD/CAE design, training simulators, education and 3D-TV. Virtual environments can be rendered with high fidelity, and virtual objects can be seamlessly integrated with the real environment, providing an immersive experience. Usually these display types use two monoscopic holographic displays [34,41,49,50]. Because only a small subset of the information needs to be transported per eye to the user’s eyeball, the required bandwidth is drastically reduced. Head-mounted holographic displays will thus most likely be the first kind of holographic displays that will be available to consumers in the near future.

2.1.2. Single-user desktop displays (SUDD)

These displays are similar to the previously mentioned class. These devices are either too small in display size or field-of-view to be effectively used simultaneously by multiple users [38,44,51]. Some displays even steer the content solely to the eyes of a viewer by tracking his/her gaze in real-time [42,52]. Due to screen size limitations and/or tracking area those displays are meant to be operated by a single user with a rather limited motion range, e.g. when seated or standing in a specific area. The range of applications extends most usages of head-mounted displays by scenarios where the display does not have to be wearable, so larger optical setups (and bandwidth) can be allowed for; examples include high-end 3D-TV or simulations needing advanced calculation-intensive CGH algorithms making the HMDs portability requirement a hindrance. This class is projected to mature shortly after head-mounted displays are commercially available, because many of the challenges overlap.

2.1.3. Multi-user displays (MUD)

These displays can present full-parallax holograms at extreme resolutions to multiple viewers at once. They are highly useful when multiple users need to simultaneously observe and interact with 3D models, especially over long periods of time, enabling applications in design, manufacturing, medicine, sports, etc. [53]. Currently horizontal-parallax-only tabletop displays such as [43,54,55] can give a first impression of this ultimate class of holographic displays. However, present implementations are still limited in resolution and can often only display pre-rendered holograms of small objects, a few centimetres across. They are mostly based on many multiplexed SLMs [40,56,57], or ultra-fast SLMs and rotating optical elements [39].

2.2. Limits of holographic displays

Although steady progress is made in SLM technology, satisfactory holographic displays are still not realizable. The large amount of information contained in a hologram brings current hardware to its limits: Giga/Tera-pixel resolutions, pixel sizes approaching the wavelength of light (i.e. < 1 μm), high interconnection bandwidths and sufficient switching speeds in the case of time multiplexed methods (requiring up to 10 kHz refresh rates) are not yet supported.

The relationship between the pixel pitch p and the supported angular FoV is given by the grating equation: for every spatial frequency ν, the associated diffraction angle θ is given by:

$$\lambda \nu = \sin(\theta)$$

(1)
would produce a 10 Terapixel holographic display with a maximum

e.g. 2D 4K colour television screen, while a small pixel pitch of
e.g. 230 nm basically results in a 2D 4K colour television screen, while a small pixel pitch of e.g. 230 nm would produce a 10 Terapixel holographic display with a maximum angular FoV of 180° (see Fig. 4).

The maximum diffraction angle bound will impose limits on the valid viewing zone geometry. This zone is delineated by all positions which are reachable for rays emitted from every hologram point: namely the intersection of all cones placed at every hologram point with an angle \( \theta_{\max} \). Their orientations (and by consequence, the resulting shape of the valid viewing zone) will depend on the shape of the illuminating reference wave. This is shown in Fig. 7 for two typical examples: planar wave and spherical wave illumination. The former will have an infinitely extended viewing zone, but at a larger minimal viewing distance. The latter allows for a wider viewing angle closer to the holographic screen, but generally limits the valid viewing zone to a small, fixed region. That is why spherical wave illumination is better suited for HMDs, while planar illumination might be preferred for full-parallax holographic displays.

2.3. Challenges

The following challenges in the design of holographic display impact signal processing requirements:

Ultra-high resolution. A very small pixel pitch is required for the display of macroscopic objects at a reasonable level of detail. Sub-wavelength pixel sizes enable the display of higher frequencies, i.e. finer resolved fringe patterns, which allow for large viewing angles. At the same time smaller pixel sizes further intensify the resolution problem by increasing the number of pixels required for a hologram of the same physical size (see Fig. 4). This search for high-resolution and wide-angle holograms results in large space-bandwidth products.

Light modulation efficiency. Digital SLMs have a limited number of degrees of freedom per pixel, expressed by the number of bits used per pixel to specify amplitude or phase, i.e. bit-depth, and to what extent phase and amplitude can be modulated, if at all. Furthermore, the diffusion efficiency of a device will affect the display quality. Other issues pop up too when the pixel density increases, such as cross-talk and noise. There is some trade-off between modulation efficiency and pixel count and density. All those aspects will need to be taken into consideration when optimizing these variables for designing the best possible display. Accurate simulations of these variables could help to model their impact on visual quality. Moreover, this may require the application and design of (new) algorithms to post-process the hologram as to obtain better visual quality, such as dithering, or the Gerchberg-Saxton algorithm [60] to generate phase-only holograms.

Multi-colour display. The laws of diffraction are wavelength-dependent and colours are not interchangeable. Blue light has the shortest wavelength, meaning it will have a lower diffraction angle for the same maximal SLM spatial frequency than the other colours, which can lead to problems at large viewing angles. Another consequence is that higher wavelengths can carry more information per unit area. An additional problem is that lasers have (by design) a very narrow bandwidth in the electromagnetic spectrum. This can lead to perceptual colour distortion and to highly visible chromatic aberrations, because the individual colour components will be much better distinguishable than in the incoherent white light case.

Speckle-noise reduction. Highly coherent light sources give rise to speckle noise, which can be mitigated though by various approaches. Partially coherent light sources can reduce speckle, but will cause some blur and some loss of depth perception due to their reduced coherence length. More information on the relation between the coherence and image quality can be found in [61]. Other solutions involve diffusers, temporal modulation of the laser beams, frequency shifting over time, superposition of multiple reconstructions, and quickly repositioning the laser beam.

Embedded signal processing. Due to vast number of pixels that needs to be processed, it is to be expected that internal signalling channels with intrinsic bandwidth constraints will be under high stress. Hence, processing – e.g. decoding and rendering – will need to be parallelized and hierarchically distributed such that data can be locally processed (cf. JPEG Pleno requirements [62]), as it is also the case for high-end light field displays [53].

3. Digital hologram recording

The holograms to be rendered on the holographic display can be obtained either by recording the hologram on an optical capturing setup, either by numerically calculating the hologram using computer-generated holography algorithms. In this section, we will focus on the first approach, Section 4 addresses the latter. To understand the requirements and challenges, we first summarize the underlying fundamental physics. Thereafter, the acquisition of high-resolution holograms and the connected inverse reconstructions methods are discussed.
Holography relies on the wave model of light governed by the laws of diffraction. This representation is dual to the ray (or particle) model of light. The wave nature of light is captured by Young’s famous double-slit experiment, see Fig. 8. Although electromagnetic waves are generally described by Maxwell’s equations, in a sufficiently large, homogeneous, dielectric, isotropic medium, the equations can be simplified to a scalar model.

Holography is thus typically described by scalar diffraction theory: the wavefield of a monochromatic coherent light source is modelled by a scalar field of complex-valued amplitudes, describing any component of the electromagnetic field for every point in space. The magnitude denotes the amplitude, and the angle denotes the (relative) phase delay w.r.t. to some reference wavefield. This contrasts with conventional imaging techniques, which only measure intensity (squared amplitude), but lack phase data to encode depth.

Because holograms can capture all visible features of light waves, they can provide all needed 3D cues. An illuminated object in space will induce interference patterns that can be captured by the hologram to encode the light field emanating from the object.

Mathematically, diffraction can be expressed as a (infinite) collection of point emissions. A single point spread function is given by:

\[
O(\mathbf{p}) = U(\mathbf{x}) e^{i\mathbf{k} \cdot \mathbf{x}} \left\| \mathbf{p} - \mathbf{x} \right\|^{-2}
\]

(3)

where \( \mathbf{p} \) is a point in the wavefield \( O \) on the hologram plane, for a spherical wave centred at a luminous point \( \mathbf{x} \); \( \lambda \) is the wavelength of the light, and \( k = \frac{2\pi}{\lambda} \) is the wave number. For an arbitrary surface, we can integrate over all points using the Huygens–Fresnel principle [59]:

\[
O(\mathbf{p}) = \frac{1}{i\lambda} \int_S U(\mathbf{x}) e^{i\mathbf{k} \cdot \mathbf{x}} \left\| \mathbf{p} - \mathbf{x} \right\|^{-2} \mathbf{n} \cdot (\mathbf{p} - \mathbf{x}) d\mathbf{x}
\]

(4)

where \( \mathbf{x} \in S \) are points on a surface \( S \) over which is integrated and \( \mathbf{n} \) is the surface normal on \( S \) in \( \mathbf{x} \).

Since no physical detector can measure the phase directly, as phase carries no energy, this information needs to be captured indirectly by means of interference. A typical holographic recording setup (Fig. 9) consists of a laser, whose beam is expanded as to be able to fully illuminate the object to be recorded. Next, the beam is split into a known reference beam \( R \) and the sought object beam \( O \), which will illuminate the object and encode its information. These beams are superimposed and we measure the intensity of the two interfering beams on the detector. The measured 2D intensity fringe pattern \( I \) is given as:

\[
I = |R + O|^2 = |R|^2 + |O|^2 + R^*O + RO^*
\]

(5)

where * denotes the complex conjugate. The last term \( R^*O \) contains the sought \( O \). Since \( R \) is known, we can in principle retrieve the complex-valued \( O \). The interferogram further consists of several other undesired terms. For example, since the sought \( O \) is complex-valued, the squared magnitude term \(|O|^2\) will not contain any of the important phase information.

Because of the multiple valid solutions for \( O \) (see Fig. 10), additional steps need to be taken to unambiguously retrieve \( O \). The types of setups can be classified into two categories: single- and multiple-exposure methods. One of the most popular single-exposure techniques is called “off-axis holography” [2] (Fig. 11). By using a tilted plane wave for \( R \), the terms in Eq. (5) will get separated in the frequency domain; the desired term can then be extracted using a bandpass filter. The disadvantage of this technique is that the bandwidth of \( O \) will be only a fraction of the camera bandwidth, because most of it will be discarded by the filter.

Multi-exposure methods on the other hand, consist of taking several measurements in succession. The best known example is phase-shifting digital holography [64], where generally at least 3 exposures are needed to unambiguously retrieve the complex amplitudes (cf. Fig. 10(c)). More exposures can be taken to improve noise robustness. The main drawback is that such methods cannot easily capture dynamic phenomena, reducing their application domain. Combinations such as parallel phase-shifting holography [64] exist as well.

### 3.2. Acquiring high-resolution holograms

One of the current bottlenecks for acquiring high-quality holograms is the limited image sensor resolutions: typical digital sensors consist of multiple megapixels, which is several orders of magnitude below the resolutions needed for large, high angular FoV displays.

A straightforward optical solution is to use multiplexing: by using an array of sensors and/or moving the camera between successive acquisitions, a large synthetic aperture can be created by stitching numerically many sub-holograms together [65,66]. This has applications beyond display, such as in holographic tomography and microscopy [67,68]. However, the setup will become very complex and acquisitions time-consuming, thereby precluding the capture of dynamic scenes. Other optical solutions involve modifications to the capturing system as to extract more information from the recorded intensity pattern, such as utilizing a coded aperture [69,70].

On the other hand, by accounting for the statistical features of holographic signals, it is possible to enhance the resolution of acquired holograms without modifying the recording setup. This can be done with inverse methods, whose goal is to extract a maximum of information from relatively small sets of intensity measurements, thereby improving the resolution without increasing acquisition setup complexity. The number of intensity measurements is generally considered insufficient according to the Shannon–Nyquist sampling theorem, resulting in many holographic signals that can explain the measurements.

In this context, three questions arise. First, we have to identify what constitutes a good set of measurements. Secondly, we have to select one of the holographic signals that best explains the intensity measurements (e.g. the maximum likelihood solution). To make a substantiated choice, prior information about the holographic signal may be used by inverse-problem reconstruction approaches.

Fig. 10. These diagrams illustrate the solution spaces, per pixel, that are consistent with the measurements. (a) For a known value of \( R \) and a measured intensity \( I \), the valid solutions for \( O \) lie on a circle centred at \( R \) with radius \( \sqrt{I} \). (b) When taking two measurements, there is still some ambiguity (two valid solutions), so (c) at least three different \( R \) are needed to find a unique solution for \( O \).

Fig. 11. (a) The intensity of an off-axis hologram of a speckle phantom, (b) its Fourier representation with annotated terms and (c) reconstructed hologram amplitude. Source: [63].
use cases, grams for holographic display might only be suitable for very particular techniques are hardly applicable to dynamic scenes. Hence, recording Tomographic solutions can be utilized for static scenes, but these limitations on the types of objects and scene size that can be captured.

Physical limitations. Aside from resolution these setups have physical limitations. In the case of off-axis geometries, a separation is created in the Fourier domain. Direct image reconstruction approaches will manipulate the frequency domain and apply a spatial filter for recovering only the complex amplitude information of the object beam, discarding the complex conjugate twin image as well as the zero-order illumination. These frequency-domain manipulations techniques are straightforward if the object wavefield is bandlimited [76] and may yield exact reconstruction. However, in practice, the sample contains fine structures and details, independent of the lateral resolution of the detector. Recently, regularized inverse reconstruction approaches have been developed for holographic signals and impressive reconstruction quality may be obtained [77–79], at the price of an increase in computation.

3.3. Inverse reconstruction methods

One example of a class of inverse methods is compressed sensing (CS), which is a generic theory in signal processing that deals with sparse signal recovery from an underdetermined set of linear measurements [71]. Free space diffraction provides a natural signal mixing operator, which makes CS an excellent fit for holography.

CS has been successfully applied for holographic tomography [72], 3D scene segmentation [73] and measuring 3D refractive index distributions of cells. In recent work [74,75], the groundwork was laid on how to determine optimal measurement conditions for holographic recording setups and what CDF wavelet bases are best suited for image reconstruction in compressed sensing.

In the case of off-axis geometries, a separation is created in the Fourier domain. Direct image reconstruction approaches will manipulate the frequency domain and apply a spatial filter for recovering only the complex amplitude information of the object beam, discarding the complex conjugate twin image as well as the zero-order illumination. These frequency-domain manipulations techniques are straightforward if the object wavefield is bandlimited [76] and may yield exact reconstruction. However, in practice, the sample contains fine structures and details, independent of the lateral resolution of the detector. Recently, regularized inverse reconstruction approaches have been developed for holographic signals and impressive reconstruction quality may be obtained [77–79], at the price of an increase in computation.

3.4. Challenges

Optically recording digital holograms still comes with several challenges:

Resolution limitations. The maximum obtainable resolution is limited by the issued detector. It can be extended by deploying inverse methods if a good signal model can be defined, exploiting its sparsity properties. However, results will be still limited by the signal statistics.

Physical limitations. Aside from resolution these setups have physical limitations on the types of objects and scene size that can be captured. Tomographic solutions can be utilized for static scenes, but these techniques are hardly applicable to dynamic scenes. Hence, recording holograms requires specialized equipment and optically captured holograms for holographic display might only be suitable for very particular use cases, e.g. (bio)medical imaging and non-destructive testing.

Optical distortions. Recorded holograms are typically contaminated with speckle noise and by optical aberrations. Hence, additional denoising and aberration compensation techniques need to be deployed.

Fortunately, there is another way of acquiring holographic data, namely by numerically synthesizing holograms from a 3D scene model using computer-generated holography. Its benefits and challenges will be discussed in the next section.

4. Computer-generated holography (CGH)

Computer generated holography (CGH) is an alternative to optical hologram acquisition. It has the important advantage that the object information can be obtained by means of conventional (multi-)camera setups, point cloud data, or even computer graphics. One of the main challenges of CGH is to generate realistic holograms within acceptable computation times. This can be inferred from the Huygens–Fresnel principle in Eq. (4): computing the hologram using brute-force is inefficient, since the integral has to be evaluated over all scene points, for every pixel.

4.1. Types of CGH methods

The computational load of the many-to-many mapping of the diffraction model is compounded with another difficulty: realistic holograms need very high resolutions. This fact can be derived from Eq. (1): suppose we have a green laser (λ = 532 nm), and a desired viewing angle of θ_max = 40°. This amounts to a pixel pitch p of at least 0.414 μm. For a display of 10×10 cm, we then need a massive ultra-high definition display of 58 Gigapixels.

Aside from the valid viewing zone, this also imposes restrictions on the scene geometry and the CGH calculations: incident light at too large angles will otherwise result in aliasing, see Fig. 12. Furthermore, Eq. (4) does only account for free space propagation, but not for effects such as occlusion, shadows, refraction, etc. That is why over the years, many different techniques were developed to accelerate and/or improve the realism of CGH. They can be classified as follows.

4.1.1. Point cloud methods

Point clouds (Fig. 13(a)) may represent objects as a collection of discrete luminous points [80–84]. This will discretize Eq. (4) into a summation over all points. These methods can be augmented using look-up tables with precomputed point-spread functions, or even surface elements with a precomputed light distribution. Occlusion can be cumbersome to handle, but this can be achieved e.g. by attributing a small volume to each point blocking incident light.

4.1.2. Polygon methods

Surface elements such as triangles (Fig. 13(b)) leverage the fact that diffraction between planes can be efficiently computed using a convolution [85–87], e.g. the Angular Spectrum Method (ASM). These methods are particularly efficient when the amount of hologram pixels greatly exceeds the polygon count. Generally, the polygons will have an associated complex-valued wavefield, where the amplitude will encode the texture and the phase will encode the light emission distribution (e.g. diffuse/specular). Occlusion is more straightforward, but can be computationally demanding.

4.1.3. RGB+Depth methods

With a depth map (Fig. 13(c)), we can encode holograms with pairs of colour images and per-pixel depth information [88]. They are computationally efficient and can be of high quality, but they are restricted in the supported viewing angles. For example, objects cannot be placed behind others with a depth map. These methods are especially suitable when the viewer should gaze at the display from a fixed position. This shortcoming may be addressed though by using more depth maps for several viewpoints.
4.1.4. Ray-based methods

Ray-space representations (Fig. 13(d)) will approximate the hologram by a discretized light field, which is converted into a hologram by assigning an incident-angle-dependent impulse response to every ray [89–94]. One notable subset are the phase-added stereogram methods: the hologram is subdivided into blocks, each individually represented in its FFT domain (similar to the blockwise DCT). Every frequency component within a single block will approximately correspond to a ray of light. That way, a set of rays can be sampled for every point/coefficient, followed by a block-wise inverse Fourier transform. This mutual conversion between the hologram and discrete light field representation is also known as “ray-wavefront conversion” [95]. The advantage is that those methods are comparatively fast and can use existing ray-tracing software, discretizing the light field will reduce the quality of the generated hologram, degrading some of the 3D cues such as continuous parallax.

4.2. Sparse acceleration techniques

CGH algorithms can be accelerated further using sparsifying transforms. Sparsity is a well-known concept in signal processing literature. It refers to the property that signals drawn from a particular source can be well-approximated by a few coefficients in a certain, sparse basis. This property is very useful for many applications, such as compression, denoising, and compressed sensing. Particularly, the notion of sparsity is highly useful for CGH. The following subsections describe three common accelerations strategies: (1) using an intermediate wavefront recording (WRP) plane, (2) using a stack of WRPs and (3) computing PSFs (Point Spread Functions) in the STFT domain.

4.2.1. Single wavefront recording plane acceleration

Although the wavefield in the hologram plane is very dense, it can become highly sparse when expressed in the right transform domain. This can be leveraged for computational efficiency, since then only a small subset of the coefficients needs to be updated. The most well known example is called Wavefront Recording Planes (WRPs) [82]. It utilizes the property that points close to the hologram plane have a small spatial footprint. By placing a WRP close to the object in 3D space, only a small subset of the WRP pixels needs to be updated (Fig. 14), resulting in high sparsity. After all PSF additions, the WRP can be efficiently propagated to the hologram plane using e.g. the angular spectrum method.

4.2.2. Multiple wavefront recording planes acceleration

The WRP methods can be extended further by using multiple parallel equidistant WRPs, so as to optimize calculation times. An efficient CGH method which exploits multiple WRPs has been proposed in [83]. As shown in Fig. 15 the object is segmented in WRP zones and the points are allocated to their closest WRP. The WRPs are computed sequentially starting from the one furthest to the hologram plane. Furthermore, this method is combined with an occlusion processing technique locally applied per point. The wavefield is progressively processed from the back to the front WRP and finally propagated to the hologram plane [96,97].

Another important acceleration technique relies on look-up tables (LUT). That way, scene elements can be precomputed and stored in any transform domain. For example, PSFs quantized at discrete levels around the WRP can be reused for all the WRP zones. But LUTs are not limited to storing PSFs. This approach has been extended for diffuse surfaces by modulating the PSFs with a random phase factor [98], resulting into a simulated diffuse scattering of the light. More recently, an extension which utilizes the Phong illumination model and shadow mapping has been proposed [97]. Here, each point contributes to the WRP according to the properties of its associated surface, enabling the generation of high quality colour holograms, as shown in Fig. 16.

4.2.3. Wavelet and STFT-based accelerations

A more recent development is to compute PSFs in non-spatial transform domains. PSFs are sparse in the wavelet domain (WASABI method) [100], and even sparser in the Short-Time Fourier Transform...
4.3. Challenges

CGH methods are a compromise between calculation time, realism and supported scene parameters (e.g., dimensions, viewing angles, occlusion, lighting, transparency). Optimizing CGH methods both for offline (video) rendering and for real-time display are still very active domains of research. The main challenges to overcome are:

- **Synthesizing large CGH.** The CGH generation must be effective computationally, especially for full-parallax holographic displays. Generating holograms in real-time at video rates requires specialized software/hardware implementations. Solutions involve adapting the algorithms for multi-threading [103], implementations on Graphics Processing Units (GPUs), [84,91], high-performance computer clusters [104] or FPGA systems [105].

- **Local CGH updates.** Local memory and caching methods are indispensable regarding computational complexity and principles of efficient transforms and sparse representations may be used. This challenge overlaps with the transform design problem for compression (cf. Section 5).

- **Realistic scene rendering.** The aim is to obtain photo-realistic CGH. Although steady progress has been made, the realism is still limited compared to its state-of-the-art counterpart in computer graphics. Attempting to go beyond Phong shading would require adequate models for shadows, transparency and refraction, complex lighting, volumetric light scattering, and many others.

5. Transforms and coding

High-resolution holograms need large volumes of data to be represented, especially for dynamic holography: e.g., the considered 10×10 cm display in Section 4 needs 58 Gigapixels. Supposing we have a frame rate of 30 fps and a phase-only SLM with 1 byte per pixel bit-depth, the required bandwidth would amount to 1.75 TB/s. Because of these data rates, dynamic hologram transmission will be unfeasible with currently available (or near-future) hardware unless we have adequate compression techniques. Efficiently coding holograms will thus be of utmost importance for viewing dynamic content.

Conventional image and video codecs are unfortunately suboptimal for holograms, notably for macroscopic objects; this is because the characteristics of holographic signals differ substantially from natural image content. This becomes immediately clear when one tries to decompose a hologram with wavelets: in Fig. 17, the decomposition is not sparse. This is why novel codec architectures need to be developed to adequately represent holographic data. One of the core codec components that need to be modified is the chosen transform. Furthermore, the relationship between the objective hologram distortion and the perceived subjective distortion is still poorly understood.

5.1. Coding of static holograms

A number of different coding solutions for static holograms have been proposed over the years, and up to recently most of them concerned relatively rudimentary solutions [22]. We can categorize these codecs in four classes, coding: (1) the input content for CGH, (2) the complex amplitude data in the hologram plane, (3) the backpropagated hologram in the object plane and (4) an intermediate representation.
5.1.1. Coding of CGH input
These methods encode the 3D scene representations (e.g. RGB+Depth, meshes, point clouds, light fields) to circumvent the problem of efficiently compressing the holographic signals. Thereafter, CGH will generate the output hologram. For example; the RGB+Depth method can be used with compressed textures and depth maps to generate holograms on-the-fly [107]. The advantage is that the 3D content itself can be efficiently compressed with existing solutions, resulting in small file sizes. The main disadvantage is that the hologram generation is shifted to the display side: CGH has a high computational load, making real-time holography only possible for simple scenes. Many types of holograms (especially optically acquired ones) cannot be encoded with these methods. This solution will thus be impractical for broadcast scenarios. Moreover, the encoded 3D content will still contain redundancies, since only a part of the scene’s information will reach the hologram due to e.g. occlusion.

5.1.2. Coding in hologram plane
These methods on the other extreme treat the raw holograms as a special type of image. These are generally existing image or video codecs, which are modified and/or extended to better match the statistics of holographic signals as to improve their rate-distortion behaviour w.r.t. the default unmodified codec. They mostly involve transforms that are better tailored to handle high frequencies and strong directionality, caused by the interference fringes. For static holograms, various transform extensions were proposed: directional-adaptive wavelets and arbitrary packet decompositions [108], vector quantization lifting schemes [109], wavelet-bandelets [110], the wave atom transform [111], and mode dependent directional transform-based HEVC [112,113]. Another family concerns Gabor wavelets, which obtain an optimal compromise between spatial and angular resolution permitted by the Heisenberg principle [114]. It was shown by Viswanathan et al. [115] that this transform can be used for adaptive view-dependent coding. El Rhammad et al. specified a matching pursuit approach for an overcomplete Gabor wavelet dictionary [116] and further expanded their support for efficient view-point dependent decoding in support of holographic HMDs [117].

Many holographic displays are driven by phase-only SLMs. For these displays, one can discard the amplitude since we need to code only the argument of the complex-valued wavefield, also known as the “wrapped phase”, typically represented by values between 0 and 2\(\pi\). Because the wrapped phase behaves non-linearly, it should not be processed with conventional linear transforms such as the DCT or wavelets, since they will incorrectly treat phase wraps as edges and wrongly model phase distance. For example, for some very small \(\epsilon > 0\), the values 0 and \((2\pi - \epsilon)\) will be treated as highly different intensity values, even though they are very close in phase distance.

One possible solution is to use phase unwrapping [118] to linearize the signal, but this is an ill-posed problem, computationally intensive and can drastically increase the dynamic range of the signal. Other solutions are the use of nonlinear phase-only transforms, such as modulo wavelets [119,120]. The drawback is that lossy coding is more tricky; care has to be taken when quantizing the signal because of the transforms’ non-linearity.

Methods coding the hologram plane data can in principle represent any content and are fast, since they leverage existing coding architectures, and do not involve any generation of holographic data. Their main drawback is that it becomes much harder to efficiently compress data,
primarily because the transforms lack awareness of the 3D scene content. This problem becomes even trickier for holographic video: unlike for regular video content, local (e.g., block-based) motion estimation and compensation techniques are inadequate rendering them ineffective, because small motion in the 3D space will influence all hologram pixels. Another important issue is the rate–distortion behaviour: as the content is modelled indirectly, the relationship between objective and perceived (subjective) distortions is not straightforward.

5.1.3. Coding in object plane

This class of compression methods utilizes a single backpropagation, typically numerical Fresnel diffraction or ASM, which is efficiently computable with 1 or 2 FFTs. When the hologram has a large depth-of-focus (small aperture, large pixel pitch) or a scene which is (nearly) flat, then a backpropagation to the right depth will effectively refocus the entire scene. That way, the refocused hologram will resemble an image, making it subsequently better compressible by conventional codecs [121]. This parallels the use of the WRP in CGH, which utilizes its sparsity in the object plane for simplifying computations.

A notable example of this approach are the Fresnelets [122]: they are a closed-form mathematical expression of Fresnel diffraction combined with B-spline wavelets, with demonstrated use for compression [123]. Object-plane coding was recently investigated by combining it with modern codecs as well [113,124]. However, an important limitation of using a global Fresnel (or ASM) transform is that it is ill-suited for deep scenes: multiple objects or single extended-depth objects cannot all be completely brought into focus simultaneously.

5.1.4. Coding of intermediate representations

Coding in the object plane works well for flat scenes like for example microscopic data; however, when encoding deep scenes, it is impossible to identify one single object plane that is suitable for globally encoding the data. Choosing a particular object plane results in large parts of the scene being out of focus. Consequently, this gives rise to poor encoding performance due to a weak match with natural image statistics – significant fringe patterns are still present – and to loss of the depth information. This is illustrated for the Fresnel backpropagation case in Fig. 18(b) where numerical reconstructions of the decoded holograms, with either focus on back or front planes of the scene, are poor. For comparison purpose, encoding in the hologram plane with plain JPEG 2000 is shown as well.

To discuss the design of an intermediate representation that allows for the preservation of depth information, we need to introduce the time–frequency representation – a.k.a. space–frequency representation, the Wigner representation or phase space – in which the properties of holographic signals can elegantly be represented visually and mathematically. It is a key tool to build a deeper understanding of the characteristics of interferometric holographic signals. For simplicity, we restrict in Fig. 19 the graphical portrayal to 1D signals. Similarly to a music score, time (or space) and frequency are jointly represented on the horizontal (\(x\)) and vertical (\(\omega\)) axes, respectively.

Because frequencies correspond to diffraction angles given by Eq. (1), points in the time–frequency domain will correspond to individual rays. Note that we cannot have perfect localization in both space and frequency simultaneously due to the Heisenberg uncertainty principle, so in practice we cannot extract individual rays from a wavelet-based representation such as holography.

A mathematically elegant solution to address the intermediate representation problem is to use Linear canonical transforms (LCTs). They form a class of unitary transforms, i.e., linear operations on the time–frequency domain. LCTs can be used to model Fresnel free-space back-propagation as a simple global shear in the time–frequency domain. Unfortunately, LCTs cannot adequately represent deep scenes, similarly to the backpropagation case.

To that end, Blinder et al. [101] derived the reversibility conditions on the scene depth map needed for having valid nonlinear canonical transforms. Furthermore, they proposed an efficiently computable subset using a piecewise linear approximation of the depth profile of the scene surface (Fig. 20). This was achieved by segmenting the hologram and applying a series of (complementary) unitary transforms based on the depth profile of the object or scene and as such seamlessly tile the time–frequency domain.

This will result in an intermediate representation where all segments are brought into focus and hence a classic image or video coding strategy can be deployed. These transforms do not exhibit redundancy, but can be more computationally costly and have some restrictions on the depth maps it can represent. The proposed transform shows markedly improved reconstructions and demonstrates the importance of modelling non-linearities for successfully capturing depth information from scenes (see Fig. 18(c)).

5.2. Coding of dynamic holograms

Conventional video codecs will attempt to compensate for motion by using localized prediction schemes (e.g. block matching algorithms).
This is ineffective for holography, because even small motions will cause pervasive changes to the signal because of the nature of diffraction, i.e. for a larger angular FoV, one can expect that every pixel contains information about every point in the scene, and hence also information about all motion present in the scene. Consequently, motion estimation will fail. This fact was e.g. noticeable in the work of Xing et al. [125], where holographic videos were compressed with a standard HEVC codec. Inter-frame decorrelation brought little benefit for compression.

Hence, due to the problematic nature of motion estimation, it makes more sense to inherit motion information from the source data that was used to computer-generate the holograms. The motion vectors can then subsequently be used to steer the motion compensation process. For translational and rotational motion parallel to the hologram plane this can easily be facilitated, but out-of-plane rotations/translations need operations in the frequency domain. Note that since full-parallax holographic data contains information for different viewing angles, inherently information is present to support compensation for rotations around the horizontal and vertical axes defining the hologram plane. Translational motion on the axis perpendicular to the hologram plane results in shearing of the spectrum in the time–frequency domain (cf. Fig. 19).

A solution recently proposed by Blinder et al. [126] is to generalize LCTs using affine symplectic time–frequency transforms, enabling to model all small rigid-body motions (6 DOF) in 3D space as particular unitary transforms. That way, the transform can be used for motion compensation (Fig. 21). This was successfully utilized for holographic video coding, resulting in substantial gains surpassing 5dB over standard HEVC codecs using a simple global motion compensation scheme [126]. Even higher gains are expected when using more advanced motion compensation architectures. Furthermore, the same scheme can compensate for the motion of users when viewing dynamic holographic content with HMDs.

5.3. Challenges

There is as of yet no generic transform or representation that is applicable to holograms with arbitrary scene content. Ultimately, the goal is to build a codec tuned for static and dynamic digital holograms, efficient both in terms of compression performance as well as in encoding/decoding time. To conclude, we list some of the important challenges to be tackled concerning hologram coding:

Efficient representations. Coding strategies based on intermediate representations are a promising approach, particularly those based on the recently proposed nonlinear canonical transforms. However, more research is necessary to lay sound mathematical foundations as well as time–frequency domain segmentation approaches. In addition, further investigation of sparse signal representations is a necessity, and research is welcomed on intra-frame prediction techniques as deployed in many image and video coders but then particularly tuned to holographic signals.

Motion compensation and estimation. Accounting for scene-space motion is needed because every 3D scene point can potentially affect every hologram pixel and classical local motion models on the hologram (such as block-based motion compensation) will be ineffective. Novel algorithms will be needed to compensate for arbitrary motion of multiple independently moving objects. Moreover, optically acquired holograms present a challenge because the ground-truth motion is unknown, and will thus have to be estimated. Automated general motion estimation in digital holography for complex scenes has not been solved yet either.

Coding complex values. Holograms are complex-valued and new approaches are needed for efficiently coding wrapped phase data for phase-only holograms, or for the phase components of complex-valued signals. An additional difficulty will be achieving lossy coding, because of the inherent non-linearity of the phase information.
been undertaken to design subjective quality assessment procedures. Both are also requiring suitable holographic test data. Let us take a closer look at these three aspects.

### 6.1. Holographic test data

Over the past years, several research groups have setup open public databases containing various holograms [106,127,128], see Fig. 22. Their goal is to provide a representative data set of high-resolution holograms, both optically recorded and computer-generated, containing both static and dynamic content. At the time of writing, the databases are still modest in size and content: each one contains maximum a dozen holograms, consists almost entirely of static holograms, and often encode simple scenes with humble resolutions ranging up to 8K and only sporadically up to 16K or larger. An important goal is to expand these databases by adding more holograms with higher resolutions and more heterogeneous and dynamic content. Currently, the JPEG Pleno standardization initiative [129], initiated by the JPEG committee, attempts to setup a database “JPEG PlenoDB” [99], collecting publicly available data sets and to define guidelines for selection of suitable test content.

### 6.2. Objective visual quality measures

Objective visual quality measures are typically classified into two categories: (1) *signal fidelity measures* that have a clear physical meaning, but which are typically poor in terms of predicting the quality perceived by the human visual system and (2) *perceptual visual quality measures*, typically modelled and trained via subjective quality experiments [130].

In digital holography, popular signal fidelity measures are Euclidean distance-based error measures, in particular Mean Square Error (MSE) and Peak Signal-to-Noise Ratio (PSNR). These are used predominantly for both error analysis and as a measure for reconstruction quality. However, MSE technically intends only to measure the global energy variation. Other well-known regular imagery measures like Structural Similarity Index Measure (SSIM) have been used separately on the Cartesian components of the complex wavefield [109,131]. Even tough a polar representation would be mathematically equivalent, directly applying fidelity measures on those channels were found to be less accurate [113] (e.g. due to the non-linear response of wrapped phase to distortions).

This behaviour is already known for regular image processing [132–135] and persists for holograms, where the phase encodes among others the depth information. The Cartesian representation is much less sensitive to compression artefacts as it avoids non-linearities by splitting holographic information into real and imaginary parts [113]. In addition, the quality can be measured in the object plane, hologram plane and after numerical reconstruction. The first two approaches have the advantage that they measure the global quality of the hologram with respect to the reference hologram in terms of PSNR or SSIM. The latter can be deployed if the numerical reconstruction quality for a particular viewing angle and focus plane needs to be examined [131].

Recently in [136], a new versatile similarity measure (VSM) was introduced, aiming to address the mentioned shortcomings for complex valued error measurement. Another novel quality metric, the Sparseness Significance Ranking Measure (SSRM) [137], utilizing sparse coding and a ranking system based on the coefficient magnitudes, was shown to be more effective than the PSNR or SSIM on holographic data as well [138].

To the knowledge of the authors, no perceptual visual quality measures have been proposed yet for digital holography. To construct them, subjective test procedures are a necessity.

### 6.3. Subjective test procedures

A standardized testing methodology to measure the perceptual visual quality – i.e. the visual quality as perceived by the Human Visual System (HVS) – for holography is still an open problem. The challenge is two-fold: (1) the heterogeneity of currently existing holographic display systems, differing in many aspects, such as light source properties, type of SLMs, resolution, pixel pitch, viewing zone geometry, optical aberrations; and (2) the high complexity of 3D input data, with content observable from many angles and at different focal depths. We consider two categories: *direct assessment*, where users directly view and rate holograms on an holographic display, and *indirect assessment* which relies on a substitute display (e.g. a light field display) showing views extracted from the hologram to be analysed.

For direct assessment to be reproducible, one would have to establish what the most relevant parameters are. On top of the existing requirements for quality assessment on regular screens, aspects to consider are: (1) minimal screen capabilities (eyebrow dimensions, viewing angle, bit-depth) and (2) viewer positioning (should the viewer move freely, go to specific positions or stay still?). However, given that holographic displays are evolving rapidly, care should be taken that the requirements are sufficiently general to remain relevant in the years to come.

It is also possible to reliably display high-quality synthetic holograms up to even hundreds of Gigapixels by printing them in e.g. glass plates [139]. Although this provides an experience which is very close to what a future high-end holographic display could show, it becomes impractical and expensive to print a large quantity of instances to compare e.g. various codecs, distortions, CGH algorithms etc. Furthermore, this medium cannot be used for displaying dynamic content.

Alternatively, these obstacles can be partially overcome using indirect assessment techniques, by utilizing conventional 2D displays or even light-field displays to render various reconstructions of the hologram. Nonetheless, this will come at the cost of ignoring several visual cues, especially the ones pertaining to depth perception and parallax. Initial work on this problem was done in [136,140], where a large 4K 2D screen was used to rate the holograms of the public data set found in [127]. In this test, the goal was to rate only the overall quality of central view reconstructions for the holograms. Such a setup allows for adapting to the currently available standard testing methodologies with minimal modifications.

For more advanced and thorough quality assessments, one may take inspiration from quality assessment procedures used for light fields: multiple views and focal depths can be shown in succession in a video format to simulate scene perception from many different viewpoints [141]. The advantage of this approach is that the testing conditions for subjective quality measurements have been meticulously standardized for classical 2D displays [142–144]. Another similar approach would be to allow users to actively explore hologram content. This can be achieved on a screen as well, or could be experienced more naturally with HMDs.

In recent efforts [98], a high-end light field display was used to render a set of colour CGHs. The display provided the possibility of rendering a wide field of view but only with horizontal parallax. Such a display setup can be utilized for subjective tests, where the observers will be able to freely move within the FoV of the screen, observe the reconstructed hologram objects, and score the perceptual visual quality from different points of view.
6.4. Challenges

Consequently, quality assessment of holographic data still faces many challenges to be resolved:

**Public holographic databases.** Extended databases should be created by adding more holograms with higher resolutions and more heterogeneous and dynamic content. These holograms should depict a good utilization of the time–frequency space. If large swathes of the time–frequency domain are zero, this means that the hologram capacity is underused. However, solely looking at the spatial or Fourier representation can be misleading, since both can have a dense signal yet have a very sparse time–frequency representation. This will happen when e.g. an object surface is insufficiently diffuse. Secondly, objects should be placed close enough to the hologram plane. This is needed to enable proper evaluation of depth cues such as accommodation and parallax. Finally, the bounds of the aliasing-free zone should be respected (cf. Fig. 7).

The hologram parameters dictate the allowed scene geometries [145]. Care should be taken to correctly place objects in the scene, otherwise the useable viewing zone will shrink and unwanted artifacts will appear. For example, objects too close to the hologram plane will vanish at steep viewing angles, having the effect as if the viewer is looking at the scene through a keyhole.

These databases can subsequently be deployed to develop quality metrics that are targeted at assessing the perceptual visual quality. These metrics face the following challenges.

**Visualization artefacts.** Visual distortions are entirely different from natural imagery: e.g. because of the non-local nature of the hologram’s information content, the effect of local point defects on the hologram will be less visible than for natural imagery and have a more distributed impact on the reconstructed hologram [15]. This requires an extra effort to study the effect of different distortion types affecting the wavefield on the reconstructed hologram. Examples include: errors induced by lossy codecs, display limitations (bit-depth, phase/amplitude only) and optical aberrations.

**Speckle noise perception.** Speckles arise naturally in holography and will affect the assessment of any reconstructed views, whether optically recorded or synthetic, by reducing the perceived reconstruction quality. Quality metrics should either account for or ignore this noise depending on the use case. As explained in the previous section, two generated holograms can possess highly different signals because of the different random speckle instances for diffuseness, yet be visually nearly identical after reconstruction, making hologram comparison difficult.

**View position and angle dependencies.** Although traditionally a single quality score is expected to represent the overall visual quality, this may not suffice for 3D content. For wide FoV holograms of deep scenes, a human subject may experience a different perceptual visual quality depending on the viewing position, angle and focal depth. For example, standard JPEG compression with perceptual quantization will favour low frequencies, thereby better preserving the centre view than larger viewing angles [123]. Visual quality measures are not only expected to provide overall quality prediction per hologram, but also perspective-dependent predictions as well. As far as concerning depth resolution, compression technologies might affect the depth of field as well, see Fig. 18.

**Subjective quality assessment.** To model and train perceptual quality metrics, extensive subjective testing is required to facilitate their design. New procedures on holographic display should be developed that account for aspects such as minimal screen capabilities and viewing position. An additional challenge would be to design these assessment procedures such that they are sufficiently flexible to support successive generations of holographic displays.

**Alternative displays.** Adapting subjective quality assessment procedures will be required as long as sufficiently mature holographic displays are lacking. These methodologies are based on displaying numerically reconstructed holograms on nowadays available displays ranging from classical 2D displays to light field displays, the latter being capable rendering large, wide angle, numerical hologram reconstructions. Modelling the relationship between quality scores collected via such “pseudo-holographic” setups with quality scores obtained on early holographic displays will be a necessity to validate predictions from these experiments.

7. Conclusions

Realizing digital holographic display systems still faces many challenges on the signal processing side. In this paper, we provided an overview on the state-of-the-art in holographic displays, computer-generated holography, efficient transforms and coding solutions and quality assessment for digital holograms. Because of its multidisciplinary nature, several challenges remain to be addressed:

**Holographic displays** do not yet reach the requisite visual quality, screen sizes and viewing angles needed for truly high quality experiences. Various attributes of holographic displays will need to be leveraged to get the best signal given the technological constraints. Examples are light modulation efficiency, cross-talk, bit-depth, amplitude/phase modulation, colour display accuracy, speckle noise and embedded signal processing within the display.

**Recording digital holograms** at high resolutions especially for large scenes remains cumbersome. Hence, for holographic display, computer-generated holography solutions are preferred. CGH algorithms have to overcome many hurdles for reaching an acceptable computational cost. Numerical evaluations of diffraction is highly computationally expensive for modelling realistic-looking content, all while needing resolutions orders of magnitudes larger than current 2D displays. Good candidates are sparse CGH methods, which exploit spatial, temporal and frequency redundancies in holographic signals without compromising quality.

**Transforms and codecs** require new efficient representations, which remains an open problem for holographic signals with large space-bandwidth products. Optimized transforms and prediction schemes will have to strike a balance between sparsity, computational efficiency, 3D scene-awareness and scalability requirements. Because holographic signals are highly non-stationary, we expect that the best transforms will model scene content in phase space, i.e. the time–frequency domain. Not only will this track the behaviour of diffraction more closely, but this will also allow for selective ROI decoding of the time–frequency space so as to only process the displayable subset of the holographic scene or to extract specific views.

**Quality assessment** is important for designing holographic systems. Adequate quality measures will help to steer the joint design of holographic codecs, CGH algorithms and display systems. The first steps to achieve this goal will be (1) to design large, varied public data sets containing holograms of highly realistic objects that can account for all human visual cues; (2) to design initial quality assessment procedures for numerically reconstructed holograms on regular 2D screens or light field displays at various viewing angles and focal depths; (3) to evaluate the relevant hardware parameters affecting quality perception in holographic display systems aiding the establishment of standardized testing procedures.

In this context, it is encouraging that also standardization bodies are already addressing some challenges. Particularly, the JPEG committee is taking the lead by developing the JPEG Pleno standard that will facilitate the capturing, representation and exchange of not solely light field, but also point cloud and holographic imaging modalities [129,146,147]. The standard aims at (1) defining tools for improved compression while
providing advanced functionalities at system level, (2) supporting data and metadata manipulation, editing, random access and interaction, protection of privacy and ownership rights as well as other security mechanisms and providing methodologies for quality assessment.

In summary, we can state that holographic display systems are steadily coming within reach [8,14,19–21]. However, to enable this type of visualization, significant efforts are still required, not solely from the optics, photonics and nano-electronics communities, but also from the signal processing community. Moreover, all developed algorithms will have to run at an acceptable computational cost, which in the end might be one of the highest hurdles to take. Nonetheless, while accounting for these many challenges, we can still expect on a relatively short-term horizon, (1) providing advanced functionalities at system level, (2) supporting data and metadata manipulation, (3) enabling components and results, Proc. SPIE 9599 (2015) http://dx.doi.org/10.1117/12.2190997.


