SHEARLET-DOMAIN LIGHT FIELD RECONSTRUCTION FOR HOLOGRAPHIC STEREOGRAM GENERATION

Erdem Sahin, Suren Vagharshakyan, Jani Mäkinen, Robert Bregovic, Atanas Gotchev

Department of Signal Processing, Tampere University of Technology, Tampere, Finland

ABSTRACT

Holographic stereograms (HSs) constitute one of the most widely used types of computer-generated holograms. The scene information required to calculate the HSs can be acquired by conventional digital cameras. It is, however, usually required that the scene should be captured from dense set of view points. Therefore, relieving this requirement is critical in the sense of easing the capture process. In this paper, in the capture stage of holographic stereograms, we employ our previously presented light field reconstruction algorithm [1], where we utilize sparse representation of light fields in the shearlet domain and reconstruct dense light fields from their highly under-sampled versions. The simulation results demonstrate that we can relieve the dense view sampling requirement of HSs, e.g. by as high as 8×8 sub-sampling factor, and still keep the perceived image quality of holographic reconstructions at satisfactory levels. This enables, for example, replacing the scanning camera setups with the more convenient multi-camera arrangements.

Index Terms— Holographic stereogram, light field, sparse reconstruction, shearlet

1. INTRODUCTION

An end-to-end holographic capture and display is usually regarded as the ultimate way of 3D scene replication. As it is well-known, however, holographic capture relies on the wave interference principle which requires illumination of the scene with a coherent light source. This in turn makes the recording of real-life scenes difficult for various practical reasons. The endeavor towards solving this problem goes back to 1966 when the first computer-generated hologram was proposed by Brown *et al.* [2]. Computer-generated holography (CGH), in a sense, simulates the optical recording process carried out in holographic capture. In other words, the interference of reference and object waves is calculated numerically. Therefore, it enables obtaining holographic information from a synthetically generated scene or a real-life scene which is illuminated by (incoherent) white light.

Unlike other model-based approaches (which require scene depth information), such as Fresnel hologram [3] and phase-added stereogram [4], holographic stereogram (HS) [5] is an image-based CGH technique which relies only on a set of captured images of the scene. Due to this relieved scene capture requirement as well as their efficient way of calculation, HSs have found various applications, especially, with the recently developed holographic print techniques [6]. The set of captured images required for HSs are usually described using the light field (LF) formalism [7]. The sampling requirements of the LF, which is to be used in HS calculation, is determined based on the human visual system (HVS) as well as the properties of the scene. These requirements mostly result in densely sampled LFs which then set challenging constraints on the practical capture setups. As a consequence of these constraints, the scene is usually captured by a scanning camera which should be moved by an accurate camera positioning system with step sizes in millimeter or sub-millimeter ranges [8]. Relieving the sampling requirements of LF is crucial for HSs, not just to ease the tedious work to be accomplished by such capture systems but also to enable new capture setups. For example, relaxing the camera movement step size to centimeter ranges could make a multi-camera setup useful for capturing, which then removes the static scene assumption of the scanning camera systems and enables dynamic scene capture.

Cao *et al.* [9] and Rivenson *et al.* [10] have used the recently emerged compressive sensing techniques to reconstruct 3D scenes from sub-sampled versions of Fresnel holograms and HSs, respectively. In these approaches, the relation between the scene and the hologram plane is modeled via superposition of 2D-to-2D propagations defined between a set of parallel planes representing the scene and the hologram plane. Thus, independent treatment of different sections of the scene leaves the applicability of such methods questionable for scenes with occlusions. As the HSs actually rely on the captured LF, we aim to solve the sparse capturing problem within the context of LF reconstruction from its sub-sampled versions. By this way, we surpass the above-mentioned modeling problem, which then makes our approach applicable to scenes with occlusions. The novel view synthesis approaches are directly applicable to the problem considered in this paper. Indeed, in [11, 12, 13], depth-image based rendering techniques have been used to reduce the number of captured images required for CGH. The performance of such approaches is dictated by the quality of the depth estimation which is

very much scene dependent. Our LF reconstruction algorithm, which has been previously presented in [1], relies on an image based rendering technique. Thus, it does not require explicit depth information. Furthermore, it is particularly useful for the problem we consider here with its competence in LF reconstruction, i.e. its ability of acceptable quality dense LF reconstructions from highly sub-sampled LFs. The sub-sampling rates we consider in this paper are significantly higher (which results in sparser set of cameras), for example, than those reported in the recently presented LF reconstruction technique [14].

2. HOLOGRAHIC STEREOGRAMS

A holographic stereogram contains the sampled LF information on a certain plane (hologram plane). The LF required to calculate HS can be described using the hologram plane and the capture plane which defines camera locations. Thus, the required LF information is obtained from cameras capturing perspective views as illustrated in Fig. 1. The light rays crossing these two planes can be equivalently parametrized using the hologram plane and two angular coordinates which correspond to positions on the camera plane. Please note that in Fig. 1 and in the following derivations we consider the 2D cross-section of the 3D space for simplicity, the extension to 3D case is straightforward.

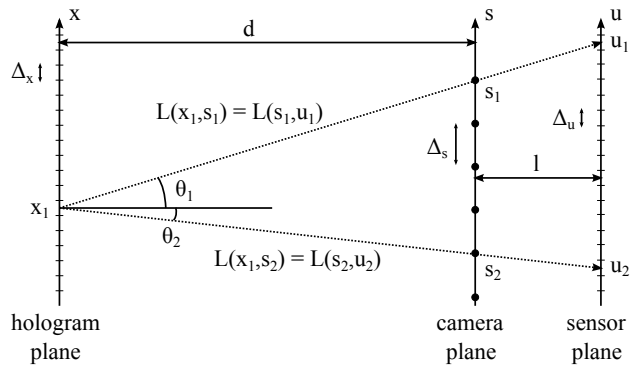

Fig. 1. Parametrization of light field capture for holographic stereograms.

Let us denote the hologram, capture and camera sensor planes by x, s and u, respectively. We define the LF parametrized by the x and s planes as $L_1(x, s)$, similarly another LF is defined using the s and u planes as $L_2(s, u)$. Denoting the distance between the camera and hologram planes as d, and the distance between the sensor and the center of projection of the camera as l, the relation between L_1 and L_2 is given by

$$L_1(x, s) = L_2(s, u_x), \tag{1}$$

where $u_x = s + l(s - x)/d$. The hologram, capture and sensor planes are discretized by the sampling steps Δ_x, Δ_s, Δ_u, which represent the holographic element (hogel) size, distance between adjacent cameras and the pixel size of the camera sensor, respectively. If the magnification equation given by $\Delta_x = \Delta_u d/l$ is satisfied and Δ_s is chosen to provide integer pixels of disparity D for adjacent views, i.e. $D = \Delta_s l/d \in \mathbb{Z}$, for those points on the hologram plane, then there is a one-to-one correspondence between the discrete LFs $L_1[m, i]$ and $L_2[i, k]$. Thus, the LF $L_1[m, i]$ can be directly read from the captured images.

HSs encode the LF information in the form of holographic fringes. The object field can be expressed as a superposition of windowed plane waves emitted from different hogels to different directions, and the amplitudes of the plane waves are specified by the corresponding LF (intensity) samples. Thus,

$$O_{HS}(x) = \sum_m \mathrm{rect}\left(\frac{x - m\Delta_x}{\Delta_x}\right) \times \sum_i \sqrt{L_1[m, i]}\exp(j2\pi f_x^{mi} x), \tag{2}$$

where f_x^{mi} are the spatial frequencies of the plane waves on the hologram plane. They are determined by the propagation directions of the corresponding rays as $f_x^{mi} = (1/\lambda)\sin\theta_x^{mi}$, where λ is the wavelength of the monochromatic light and θ_x^{mi} represent the incidence angles of the rays along the $x-$axis. The inner sum in Eq. 2 produces the spatial pattern to be written inside each hogel and it can be found via applying an inverse Fourier transform operation to the reordered images according to Eq. 1. The interference pattern of a reference wave and the object wave found by Eq. 2 creates the intensity fringe patterns of the HS which then reconstruct the object wave when illuminated with the same reference wave. In this paper, we consider the complex object wave as the HS to avoid the reconstruction noise that would be introduced by the conjugate object wave so as to evaluate our approach in a more reliable way, as will be demonstrated in Sec. 4.

The diffractive properties of HSs are mainly determined by the hogel size. The hogel size of the HS is usually chosen based on the properties of HVS and an average intended observation distance. Assuming that HVS is a diffraction-limited imaging system, the minimum distance between two points at distance d that can be resolved by the HVS is given by the Rayleigh criterion as [15]

$$\Delta_x^{HVS} = \frac{1.22\lambda d}{T}, \tag{3}$$

where T is the pupil size of the human eye (which is typically in $2mm$-$8mm$). Thus, having a hogel size smaller than or equal to this minimum resolvable distance will ensure maximized perceived image resolution. On the other hand the pupil size sets an upper limit for the angular resolution of HS which needs to be satisfied for smooth experience of view-dependent image properties such as motion parallax [16]. These two criteria usually impose a dense LF sampling

for the capture setup which complicates the capture process. In the following section, we propose a method to relieve the sampling requirements regarding the capture setup.

3. LIGHT FIELD RECONSTRUCTION

One widely used way of LF analysis is to utilize the epipolar-plane image (EPI) representation. An epipolar-plane image can be formed by taking the slices of the LF, i.e. for a 4D LF $L(s,t;u,v)$ it is obtained as $E(s,u)$ or $E(t,v)$, depending on the direction for which the analysis will be carried out. The problem of densely sampled LF reconstruction can be formulated as reconstruction of each densely sampled EPI slice from only a sparse (decimated) set of samples as illustrated in Fig. 2(a) and Fig. 2(b). Here, we refer to the LF sampling case as the densely sampled LF, when the disparity range of the scene between adjacent views is within $-1:1$ pixels with respect to the disparity of points on the focused scene plane. It has been shown that the continuous LF function can be obtained from such a densely sampled LF using linear interpolation [17].

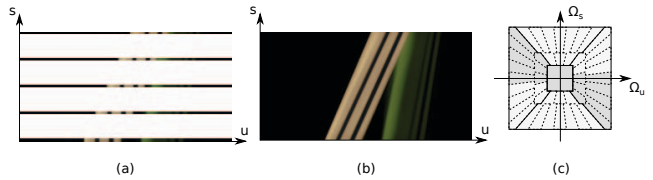

Fig. 2. Reconstruction of the densely sampled EPI. (a) Decimated EPI. (b) Corresponding reconstruction result. (c) Tiling of the frequency domain of EPI by the shearlet atoms.

The LF reconstruction problem can be efficiently solved using regularization in the shearlet domain, since the LFs exhibit sparse representations in this domain [1]. Fig. 2(c) illustrates how the frequency domain of EPI is tiled by the shearlet atoms. The shearlet atoms are distributed such that each disparity value in EPI is revealed as direction in tiling. The reconstruction of unknown samples in EPI can be modeled as estimation of $\boldsymbol{a}$ given that $\boldsymbol{b} = \boldsymbol{H}\boldsymbol{a}$, where $\boldsymbol{a}$ and $\boldsymbol{b}$ are the vectorized versions of the densely sampled and decimated EPI, respectively, $\boldsymbol{H}$ is the masking matrix representing the known samples positions. Reconstruction is obtained by iterative hard thresholding procedure with decreasing threshold [1]. That is,

$$\boldsymbol{a}_{n+1} = \boldsymbol{S}^* \left\{ \boldsymbol{T}_{\lambda_n} \left\{ \boldsymbol{S} \left[\boldsymbol{a}_n + \alpha(\boldsymbol{b} - \boldsymbol{H}\boldsymbol{a}_n) \right] \right\} \right\},$$

where $\boldsymbol{S}, \boldsymbol{S}^*$ are the shearlet analysis and synthesis transform matrices, respectively, α is acceleration coefficient and $\boldsymbol{T}_{\lambda_n}$ is the hard thresholding operator with threshold λ_n. After sufficient number of iteration, $\boldsymbol{a}_n$ is obtained as the solution with the corresponding sparse representation of $\boldsymbol{S}\boldsymbol{a}_n$.

In this paper, we consider full parallax viewing of HSs. Therefore, the full parallax LF reconstruction is achieved by consecutively reconstructing each horizontal parallax set and then repeating the same procedure for each vertical parallax set. For a more detailed discussion of the LF reconstruction algorithm, we refer the reader to [1].

4. SIMULATION RESULTS

We consider the simulation setup illustrated in Fig. 3, where we use the 3D modeling software Blender [18] for designing the scene and rendering the perspective images. The holographic stereogram is intended to be viewed by an observer at distance $d = 200mm$. The pupil size of the observer is assumed to be $2mm$. The hologram parameters are then set (according to Sec. 2) as: pixel size $X_x = 2\mu m$, hogel size $\Delta_x = 64\mu m$ and total number of pixels $N = 8092$.

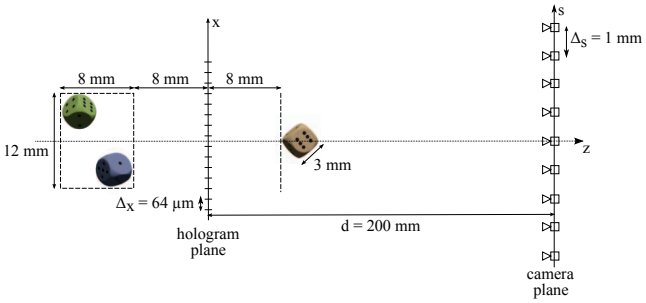

Fig. 3. The simulated scene.

The camera plane is also chosen to be $d = 200mm$ away from the hologram plane. We first implement the dense LF sampling case for which the LF samples required for HS calculation can be simply obtained from the captured LF data by linear interpolation, as pointed out in Sec. 3. For the scene shown in Fig. 3, it is required that the camera spacing should be at most $1mm$ to be able to capture the dense LF. Therefore, we put 49×49 cameras at $s,t \in \{-24\Delta_s, -23\Delta_s, ..., 24\Delta_s\}$ with spacing of $\Delta_s = 1mm$ on the camera plane. In the second setup, we employ the sparse set of 7×7 cameras at $s,t \in \{-24\Delta_s, -16\Delta_s, ..., 24\Delta_s\}$ to demonstrate how our LF reconstruction algorithm relieves the LF capture stage of HSs. Given this sparse set of cameras we estimate all 49×49 dense set of views using two different approaches: one is our shearlet-based LF reconstruction method discussed in Sec. 3 and the other one is depth-based light field reconstruction which utilizes ground truth depth maps (provided by Blender) for each sparse viewpoints. In the depth-based approach, following the procedure in [13], we separate the depth in three dominant layers (corresponding to depths of three dice) that are consequently used in the rendering of the dense light field. Then, we calculate three holographic stereograms by using both the original and the estimated dense sets of views for the above-mentioned two approaches. In all HS calculations,

the scene is assumed to be illuminated by a monochromatic light with wavelength of $534nm$ and correspondingly only the green channels of the images are utilized. The HSs are calculated hogel by hogel using IFFT, as suggested by Eq. 2, where the discrete set of spatial frequencies are obtained by resampling the continuous set of spatial frequencies via bi-linear interpolation.

We compare the qualities of obtained HSs by simulating the viewing process. In particular, we employ wave field modeling and find the image perceived by the observer using the Fresnel diffraction model as

$$I(u,v) = |\mathcal{F}_l \{T(s,t)\mathcal{F}_d \{O_{HS}(x,y)\}\}|^2, \qquad (4)$$

where $\mathcal{F}_z$ is the Fresnel propagation operation by distance z, and $T(s,t)$ represents the lens transfer function of the human eye [15]. The human eye is modeled as a camera with a thin lens having a circular aperture of diameter $2mm$. The distance between the pupil and retina, l, is assumed to be $25mm$ and the eye is focused on the hologram plane. The pixel size on the retina is set to $\Delta_x l/d = 8\mu m$, i.e. it corresponds to one hogel size according to lens magnification. In order to reduce the speckle noise in the reconstructed images, we do the calculation given by Eq. 4 in a multiplexed manner. The HSs are expressed as a sum of their sub-sampled versions (sub-sampling is done over hogels), where the sub-sampling factor is 6×6. The corresponding 36 reconstructed images are then (incoherently) summed up to find the final reconstructed image. By this way, we eliminate the interference between a given hogel and the hogels within its 11×11 neighborhood.

In Fig. 4, we show reconstructed images from three HSs, corresponding to different LFs, at different observer positions on the camera plane. The HS reconstructions corresponding to original dense set of images, estimated sets of images via the depth-based reconstruction, and estimated sets of images via our LF reconstruction algorithm are given in columns (b), (c), and (d), respectively. At each observation point shown in different rows, we take the green channels of *ideal* images rendered in Blender as reference (shown in column (a)) and calculate PSNRs for the image regions within the bounding boxes of dice. From column (b) to column (d), the corresponding PSNRs are found as $22.83dB$, $22.63dB$, $22.55dB$ for the top row; $22.90dB$, $22.89dB$, $22.87dB$ for the middle row; and $23.13dB$, $22.37dB$, $23.12dB$ for the bottom row.

Regarding the comparison between the reconstructed and corresponding ideal images, the significant parts of degradations are caused by the common factors that are the errors introduced in HS generation (e.g. due to plane wave assumption) and the speckle noise inherent to subsequent holographic reconstruction process. Otherwise we can provide acceptable quality reconstructed (perceived) images which are similar to those obtained from the original dense set of views, and in doing this we relieve the view sampling requirement by a factor of 8×8 compared to the dense LF capture case. Furthermore, our LF reconstruction and the depth-based method also result in similar reconstructed images from HSs. This makes our method more preferable, since it does not require depth estimation which is usually susceptible to artifacts such as misregistration. There are significant implications of these achievements. For example, it would be possible to capture the required LF information for the setup considered in this section by using a 7×7 array of cameras which is placed at $800mm$ away from the hologram plane, where each camera has $50mm$ focal length and $4K \times 4K$ sensor with pixel size of $\sim 4.3\mu m$. The distance between the adjacent cameras of this multi-camera setup would be $8mm \times (800/200) = 32mm$ which is a feasible value for camera spacing (it should have been $1mm \times (800/200) = 4mm$ for direct capture of dense LF).

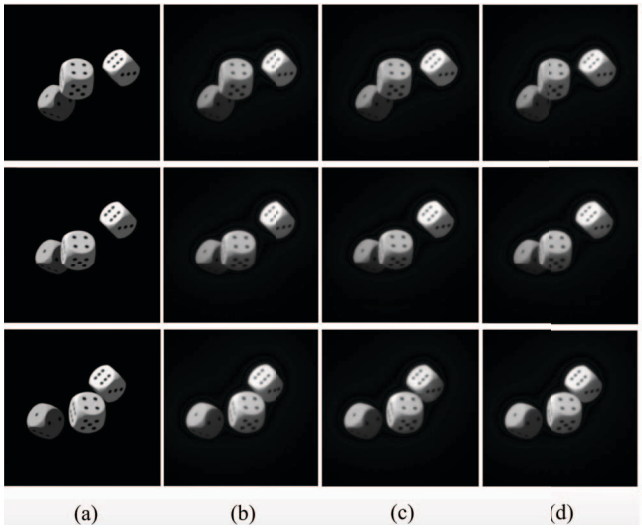

Fig. 4. HS reconstructions for views given in (a). HS is calculated from: (b) original dense set of images, (c) estimated dense set of images via depth-based approach (ground truth depth is utilized), (d) estimated dense set of images via our light field reconstruction algorithm. The observer is at: top row: $(6\Delta_s, 12\Delta_s)$, middle row: $(-12\Delta_s, 8\Delta_s)$, bottom row: $(-10\Delta_s, -10\Delta_s)$.

5. CONCLUSIONS

We have presented a way to relieve the light field sampling required by the holographic stereograms. In particular, we have utilized the sparse representation of light fields in the shearlet domain. The ability of acceptable quality dense light field reconstruction from its highly under-sampled versions (e.g. by a factor of 8×8) have led us to capture a much sparser sets of multi-view images than that is originally required for holographic stereogram calculation. As an important implication of this, we have demonstrated that the usually employed scanning camera setups can be replaced with the more convenient multi-camera arrangements.

6. REFERENCES

[1] S. Vagharshakyan, R. Bregovic, and A. Gotchev, "Image based rendering technique via sparse representation in shearlet domain," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1379–1383.

[2] B. R. Brown and A. W. Lohmann, "Complex spatial filtering with binary masks," *Appl. Opt.*, vol. 5, no. 6, pp. 967–969, 1966.

[3] B. P. Waters, "Holographic image synthesis utilizing theoretical method," *Appl. Phys. Lett.*, vol. 9, pp. 405–407, 1966.

[4] M. Yamaguchi, H. Hoshino, T. Honda, and N. Ohyama, "Phase-added stereogram: calculation of hologram using computer graphics technique," *Proc. SPIE*, vol. 1914, pp. 25–31, 1993.

[5] T. Yatagai, "Stereoscopic approach to 3-d display using computer-generated holograms," *Appl. Opt.*, vol. 15, no. 11, pp. 2722–2729, 1976.

[6] H. Kang, E. Stoykova, J. Park, S. Hong, and Y. Kim, "Holographic printing of white-light viewable holograms and stereograms," in *Holography - Basic Principles and Contemporary Applications*, E. Mihaylova, Ed. Intech, 2013.

[7] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*. 1996, SIGGRAPH '96, pp. 31–42, ACM.

[8] S. J. Hart, "Methods and apparatus for making holograms," 1998, US Patent 5,796,500.

[9] X. Cao, X. Sang, Z. Chen, Y. Zhang, J. Leng, N. Guo, B. Yan, J. Yuan, K. Wang, and C. Yu, "Fresnel hologram reconstruction of complex three-dimensional object based on compressive sensing," *Chin. Opt. Lett.*, vol. 12, no. 8, pp. 080901, 2014.

[10] Y. Rivenson, A. Stern, and J. Rosen, "Compressive multiple view projection incoherent holography," *Opt. Express*, vol. 19, no. 7, pp. 6109–6118, 2011.

[11] B. Katz, N. T. Shaked, and J. Rosen, "Synthesizing computer generated holograms with reduced number of perspective projections," *Opt. Express*, vol. 15, no. 20, pp. 13250–13255, 2007.

[12] Y. Ohsawa, K. Yamaguchi, T. Ichikawa, and Y. Sakamoto, "Computer-generated holograms using multiview images captured by a small number of sparsely arranged cameras," *Appl. Opt.*, vol. 52, no. 1, pp. A167–A176, 2013.

[13] J. Jurik, T. Burnett, M. Klug, and P. E. Debevec, "Geometry-corrected light field rendering for creating a holographic stereogram," in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Providence, RI, USA, June 16-21, 2012*, 2012, pp. 9–13.

[14] X. Cao, Z. Geng, and T. Li, "Dictionary-based light field acquisition using sparse camera array," *Opt. Express*, vol. 22, no. 20, pp. 24081–24095, 2014.

[15] J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, 2nd edition, 1996.

[16] M. Lucente, *Diffraction-Specific Fringe Computation for Electro-Holography*, Ph.D. thesis, Massachusetts Institute of Technology, USA, 1994.

[17] Z. Lin and H.-Y. Shum, "A geometric analysis of light field rendering," *Int'l J. of Computer Vision*, vol. 58, no. 2, pp. 121–138, 2004.

[18] "Blender," http://www.blender.org/about/.