Optics and Laser Technology 119 (2019) 105590

Contents lists available at ScienceDirect

Optics and Laser Technology

journal homepage: www.elsevier.com/locate/optlastec

Full length article

Dynamic holographic imaging of real-life scene

Xiao-Ning Pang, Shao-Ji Jiang, Jian-Wen Dong*

School of Physics & State Key Laboratory of Optoelectronic Materials and Technologies, Sun Yat-sen University, Guangzhou 510275, China

HIGHLIGHTS

- A commercial device is used to acquired 3D information of real-life scenes.
- Triangulation of point-clouds and surface reconstruction are performed to enhance the realistic effect of the 3D model.
- Spatial domain Fraunhofer diffraction algorithm is used to accelerate the generation of the hologram.
- A universal image quality index is introduced to qualify the optical reconstruction.

ARTICLE INFO

Keywords:
Computer-generated holography
Real-time
Spatial-light modulator
Shape measurement

ABSTRACT

In this paper, a Kinect based three-dimensional real-life holographic imaging method is proposed. In this method, the three-dimensional information of a real-life scene is acquired by using a commercial Kinect device. The output depth map is converted into a point cloud using the pinhole camera model. Surface reconstruction of the point cloud is achieved by simply using directly uniform sampling method. Our proposed spatial domain Fraunhofer diffraction algorithm is used to generate dynamic hologram videos. Optical reconstruction is performed to test the validity of the proposed method. A universal image quality index is introduced to analyze the quality of holographic imaging. A depth map based triangular refinement method is used to increase the resolution of the 3D model. The final resolution of the reconstructed holographic image can be 3.0 times higher than the original one.

1. Introduction

Computer-generated holography (CGH) is one of the promising technologies of true three-dimensional (3D) display because of its ability to reconstruct complex wave field with both intensity and phase [1]. Unlike traditional holography, the hologram is mathematically calculated in CGH instead of optical recording processing. Lots of algorithms have been proposed to generate holograms, such as point source method [2] and polygon-based method [3–5]. Large screen area and small pixel pitch, implying a huge number of pixels and heavy computational burden, can ensure the large viewing angle so that CGH can represent a 3D optical field with full depth cues. Researchers have made great efforts to accelerate computational speed. Because the computational units of the polygon-based method are much less than those of the point-based method with the same vivid 3D effect, polygon-based method is used to save the computational times.

Another advantage of CGH compared with traditional holography is that virtual object can be encoded into a hologram and optically reconstructed. It widely extends the utilization range of CGH. However, the generation of a 3D holographic video using virtual models is labor-consuming. In order to ease the burden of creating 3D models, calculating holographic video using 3D information acquired from a real-life scene is a better choice. Three-dimensional information acquirement is also known as shape measurement. Many methods have already been proposed to measure the depth information of 3D objects, including stereo vision method, laser scanning method, and structured-light projection [6]. However, a sophisticated system is usually needed in these methods to generate 3D models with fast speed and high accuracy simultaneously.

As a commercial device, Kinect is a good alternative way to measure depth information of a dynamic 3D real-life scene with high rate and good accuracy for holographic display. The first-generation Kinect, released in November 2010, is a low-cost peripheral using with Microsoft Xbox 360 to detect the gesture of players [7]. Kinect has been widely used in many applications. John Stowers et al. used Hough transform and depth maps from Kinect to autonomously hover a quadrotor helicopter [8]. The Kinect device aided three-dimensional display system has also been reported. Md. Shariful Islam et al. presented a computer-

* Corresponding author.
E-mail address: dongjwen@mail.sysu.edu.cn (J.-W. Dong).

https://doi.org/10.1016/j.optlastec.2019.105590
Received 10 March 2019; Received in revised form 16 May 2019; Accepted 19 May 2019

generated integral imaging system which reconstructed 3D images of a real scene captured by Kinect [9]. Ji-Seong Jeong et al. captured a real object using Kinect and converted it into the polygon model, and then reconstructed the object in the integral imaging system [10,11]. Songlin Xie et al. proposed an augmented reality autostereoscopic display system with the aid of Kinect by tracking viewers' heads [12]. Yasuhiro Takaki et al. developed a holographic display module using a 4k2k-SLM cooperated with Kinect as a 3D image input device [13]. James Barabas et al. in MIT Media Lab used Kinect to capture real scene information and performed diffraction specific coherent panoramagram computation. The reconstructed image is displayed in their Mark II holographic display system with a frame rate of 15 frames per second [14,15], although the quality of the reconstructed image is needed to be improved.

In this paper, a Kinect based dynamic real-life scene 3D holographic imaging method is proposed. The rest of this paper is organized as follows. In Section 2, the principle of Kinect to capture 3D data of a real-life scene will be briefly introduced. In Section 3, our previously proposed spatial domain Fraunhofer diffraction algorithm will be recapitulated. In Section 4, the reconstruction of the point cloud and surface reconstruction are discussed. The optical experimental results are illustrated and discussed in Section 5. Finally, the conclusion is given in Section 6.

2. Kinect based dynamic 3D holographic imaging method

A Kinect based dynamic 3D real-life-scene optical holographic imaging method will be detailed described in this section. There are three key processes in the proposed method, i.e. 3D data acquisition, hologram encoding, and optical reconstruction, as shown in Fig. 1. In the process of 3D data acquisition, a series of depth maps presented in Fig. 1(b), which contains 3D information of a dynamic real-life scene, are acquired by using the Kinect as shown in Fig. 1(a). In order to calculate the hologram video of the dynamic real-life scene using our previous proposed spatial domain Fraunhofer diffraction method, the depth maps are then converted into triangle-based models. A detailed view of a triangle-based model is displayed in Fig. 1(c). Finally, illuminated by the coherent laser with a wavelength of 532 nm, optical reconstruction of the dynamic real-life scene is obtained using the spatial light modulator (SLM), where the hologram video is uploaded. The experimental setup can be found in Fig. 1(e). Note that the pixel pitch of the SLM (Holoeye Pluto VIS) used is 8 μm, and the total number of pixels of the SLM is 1920 * 1080.

In the proposed holographic imaging method, the Kinect is used to perform shape measurement. It contains an RGB camera, an infrared (IR) camera and projector. The wavelength of structured-light emitted from the IR projector is 830 nm. An 830 nm bandpass filter is mounted before the IR camera so that only the structured-light pattern is captured [16]. Therefore, the Kinect can only work indoor in case of interfering with the IR spectrum of sunlight. To capture 3D data, a known structured-light pattern which varies as depth increases is projected into a 3D scene from the IR projector [17]. Structured-light patterns at different depth planes are saved in the local memory of the device as reference patterns in advance. Then, the structured-light pattern mapping into the scene, known as a measured pattern, is captured by the IR camera. Comparing with the reference patterns, the depth clue of each pixel in the measured pattern can be obtained, which is known as the depth map. The output raw depth map is quantized to be 11 bits [8], which is not the real distance between the 3D scene and Kinect. The real depth can be obtained using pre-calibration information of an object with known distance [18,19]. At the same time, the RGB camera captures color images as the texture of 3D scenes. The IR camera and RGB camera can output depth maps and color images at the frame rate of 30 Hz with a total number of pixels 640 × 480 simultaneously [8].

Note that due to the limitation of the IR projector's power, the measurement range of Kinect is limited to a maximum of about 3 m. Moreover, the measured depth resolution varies with depth. At about 2 m away from Kinect, the lateral resolution, in *xy* plane, is about 3 mm while the longitudinal resolution, at the same distance along the *z* direction, is about 10 mm [8]. To capture 3D data with fine resolution, the best working distance is about 1 m away from Kinect.

To generate hologram videos, our previously proposed spatial domain Fraunhofer diffraction algorithm is applied [4,5]. The holograms are generated from triangle-based models using spatial domain Fraunhofer diffraction algorithm. In order to match the input of the algorithm, the depth maps are then converted to triangle-based models based on pinhole camera models, which is known as surface reconstruction. An objective file that contains the information of the 3D scene, such as vertices, texture maps and normal vectors of each triangle, is then generated.

3. Recapitulation of spatial domain Fraunhofer diffraction algorithm

In this section, a brief recapitulation of our previously proposed method, spatial domain Fraunhofer diffraction algorithm, is described. The triangle-based model is used in this algorithm. As shown in Fig. 2,

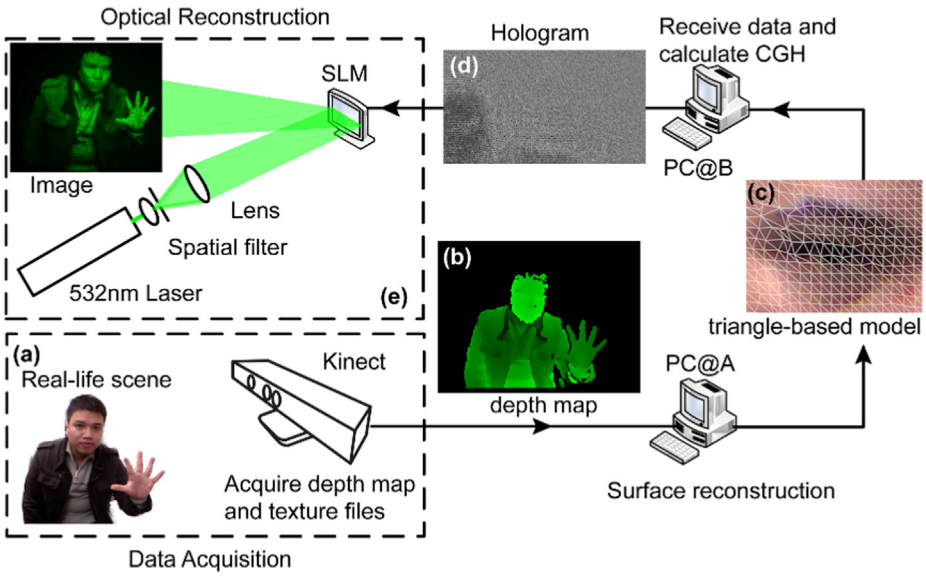

Fig. 1. Schematic flow chart of Kinect based dynamic 3D real-life scene optical holographic imaging method. (a) The Kinect is used to acquire the depth information of a real-life scene. (b) Two-color-channel depth map exported from Kinect. (c) Triangle-based model reconstructed from depth map. Only the detailed view of eyes is shown in (c). (d) Hologram generated using spatial domain Fraunhofer diffraction algorithm. (e) Optical setup is shown. A DPSS laser with a wavelength of 532 nm is expanded and collimated and then illuminated on the spatial light modulator uploaded with a hologram. Image of a real-life scene is then optically reconstructed.

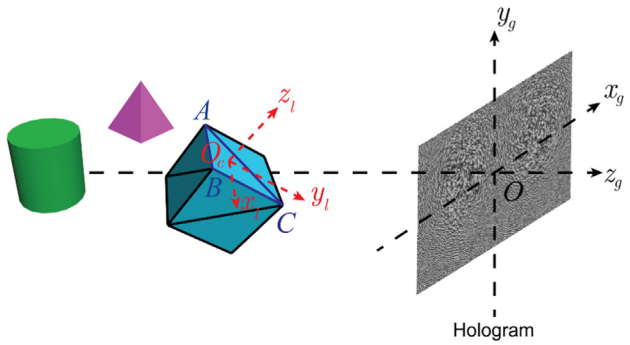

Fig. 2. Schematic diagram of a 3D scene and its diffraction field. A scene consists of three triangle-based models is shown. Δ_{ABC} is one of triangles of the scene. O_c is the centroid of Δ_{ABC}. $Ox_gy_gz_g$ is the global coordinate. The hologram locates on the plane of $z_g = 0$. $O_cx_ly_lz_l$ is the local coordinate with its origin at the centroid of Δ_{ABC} and the normal vector direction of the triangle is parallel to z_l.

there are three triangle-based objects in the 3D scene, i.e. a cube, a cylinder and a pyramid. Only the mesh wire of the cube is shown. Δ_{ABC} is one of the triangles make up the cube. $Ox_gy_gz_g$ is the global coordinate and $O_cx_ly_lz_l$ denotes the local coordinate of current diffraction apertures, i.e. Δ_{ABC}. $O_c(x_c, y_c, z_c)$ is the centroid of Δ_{ABC} in the global coordinate. The hologram is located at the plane $z = 0$. Then the total diffraction field of the 3D scene O_{total} is the linear superposition of that of all triangles. The diffraction field of an arbitrary triangle, such as the blue triangle Δ_{ABC} in Fig. 2, can be derived from Rayleigh-Sommerfeld diffraction integral as follows [4,5],

$$O_H(x_H, y_H) = \frac{\exp[ik(z_c + r_0)]}{i\lambda r_0}\mathcal{F}\{O_\Delta(x_l, y_l)\} \tag{1}$$

where λ is the wavelength of illumination light source and $k = 2\pi/\lambda$ is the wavenumber. r_0 is the distance between the centroid of the diffraction aperture and any hologram pixel (x_H, y_H). $\mathcal{F}\{\cdot\}$ is the operator of Fourier transform (FT).

By applying affine transformation in the local coordinate, the triangle O_Δ is transformed into a simple right triangle O_0 with vertices (0, 0), (0, 1) and (1, 1) in the local coordinate. The FT of O_Δ can be obtained from that of O_0 with the relation as follows,

$$\begin{aligned}&\mathcal{F}\{O_\Delta(x_l, y_l)\}\\ &= (a_{22}a_{11} - a_{12}a_{21})\exp\left(-i2\pi\frac{a_{13}x'_H + a_{23}y'_H}{\lambda r_0}\right)F_0\\ &\left(\frac{a_{11}x'_H + a_{21}y'_H}{\lambda r_0}, \frac{a_{12}x'_H + a_{22}y'_H}{\lambda r_0}\right)\end{aligned} \tag{2}$$

where $x'_H = x^l_H - r_{13}r_0$, $y'_H = y^l_H - r_{23}r_0$, and (x^l_H, y^l_H) is the hologram pixel in the local coordinate, $r_{ij}(i, j = 1, 2, 3)$ are elements of the rotation matrix between the global coordinate and the local coordinate, a_{ij} $(i = 1, 2, j = 1, 2, 3)$ are elements of the affine transform matrix, and F_0 is the FT of triangle O_0.

Consider that the complex transmittance of triangle O_Δ is uniform, as well as that of triangle O_0, the FT of triangle O_0 can be analytically expressed. The total diffraction field of uniform intensity triangle-based object, O_{total}, can also be calculated analytically, i.e. $O_{total} = \sum_{i=1}^{N} C_i\mathcal{F}\{O_{\Delta_i}\} = \sum_{i=1}^{N} C_iA_ig_i\mathcal{F}\{O_0\}$, where N is the total number of triangles of the 3D scene, O_{Δ_i} is the transmittance of the ith triangle patch, C_i is the coefficient outside FT in Eq. (1), and A_i is the coefficient outside FT in Eq. (2), g_i is the gray level of the uniform texture map of the ith triangle. Note that since only the FT of the given triangle O_0 is calculated here, the computational time of the total diffraction field can be dramatically reduced.

If the objects of the 3D scene are mapped with complicated grayscale texture, the complex transmittance of diffraction aperture $O_{\Delta_i} = a_i(x_l, y_l)e^{i\phi_i(x_l,y_l)}$ is not uniform, where the distribution of amplitude $a_i(x_l, y_l)$ and phase $\phi_i(x_l, y_l)$ of transmittance are spatial variable. In this case, the FT of the triangle cannot be calculated analytically. However, the triangle can be subdivided into numbers of triangles which are small enough that amplitude and phase distribution are considered to be uniform. Then the FT of the grayscale-texture-mapped triangle can be approximately obtained by adding up that of all sub-triangles, i.e. $\mathcal{F}\{O_\Delta\} = A\mathcal{F}\{O_0\} = A\sum_{j=1}^{M} l_j\mathcal{F}\{O_j\}$, where M is the total number of sub-triangles in one slanted triangle, O_j is the jth sub-triangle, and l_j is the complex transmittance of O_j. So even if the 3D scene with complicated grayscale texture, the diffraction field on the hologram plane can be computed using Eqs. (1) and (2).

Note that unlike the angular spectrum method, in which the diffraction field relates to the propagation of angular spectrum in Fourier space, the diffraction field in this paper directly relates to the spatial distribution of object wave. That is why the term *spatial domain Fraunhofer diffraction method* is used.

4. Triangulation of the point cloud and surface reconstruction

As described in the previous section, a triangle-based model is needed in spatial domain Fraunhofer diffraction algorithm. The depth maps obtained from Kinect should be converted into triangle-based models. A simple pinhole camera model is applied to reconstruct the 3D global coordinates of the point cloud from two-dimensional depth maps [20]. In a pinhole camera model, homogeneous coordinate is used to represent points in space. Homogeneous coordinates use $n + 1$ elements representing a point in space instead of n in Cartesian coordinate. The relation between these two coordinates is given by $[x_1, x_2, x_3, w]^T \leftrightarrow (x_1/w, x_2/w, x_3/w)^T$, where the square brackets $[\cdot]$ denote the vectors in homogeneous coordinates, and the parentheses $(\cdot)$ are interpreted as a set of Cartesian coordinates. The symbol $\leftrightarrow$ means that vectors in the two coordinates represent the same point in the space. As shown in Fig. 3, $Oxyz$ and $O_cx_cy_cz_c$ are respectively the global and local Cartesian coordinate systems of the pinhole camera in the real world. The pinhole is located at the origin of the local coordinate system. $O'uv$ is the image plane and $O'(u_0, v_0)$ is the camera center. Note that the image plane always parallels to the xy plane of the camera local coordinate. The distance between the pinhole and the image plane is considered to be the focal length of the camera f. Point $P(x, y, z)$ is an arbitrary point in a 3D point cloud (i.e. the yellow duck in Fig. 3). Point

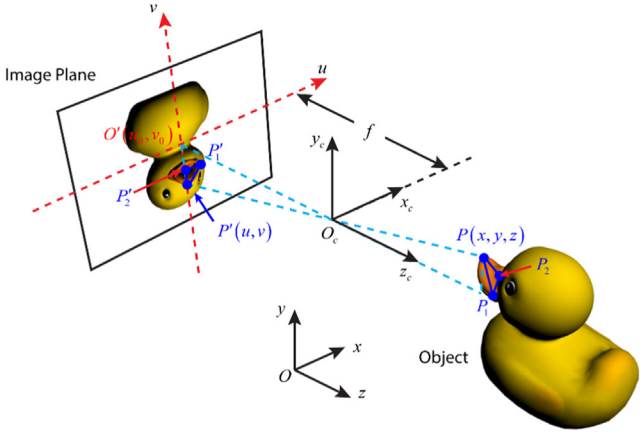

Fig. 3. Schematic diagram of the pinhole camera model. $Oxyz$ is the global coordinate of real world. The pinhole locates at the origin of the local camera coordinate $O_cx_cy_cz_c$. $O'uv$ is the image coordinate. $O'(u_0, v_0)$ is the center of the image plane. The distance between the pinhole and the image plane of the camera is f. P, P_1 and P_2 are any points of a point cloud, e.g. the yellow duck in space. And P', P'_1, and P'_2 are the corresponding projection in the image plane. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$P'(u, v)$ is the perspective projection of point P on the image plane. The perspective projection between the 3D point cloud and the two-dimensional image plane can be represented as matrix multiplication between positions in homogeneous coordinate and a projection matrix shown as follows [20],

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & s & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{21} & r_{31} & t_1 \\ r_{12} & r_{22} & r_{32} & t_2 \\ r_{13} & r_{23} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = [K \quad 0] \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$
$$= K[R \quad T] \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{3}$$

where K is the intrinsic parameters matrix of the infrared camera. f_x and f_y are the effective focal length of the infrared camera in the x and y directions, respectively. For non-square camera pixels, $f_x \neq f_y$, while $f_x = f_y = f$, for square pixels. The parameter s is referred as the skew parameter. R with elements $r_{ij}(i, j = 1, 2, 3)$, and T with elements $t_i(i = 1, 2, 3)$, are respectively the rotation matrix and translation matrix between global Cartesian coordinate and camera local Cartesian coordinate. These two matrices, known as extrinsic parameters, are important when multiple cameras are used to capture the scene. According to Eq. (3), the 3D point cloud in the global coordinate system can be converted from the depth map. Furthermore, a triangle-based model can be reconstructed from the 3D point cloud. This process is known as surface reconstruction.

Surface reconstruction from the point cloud has been widely investigated. Many methods, which can be categorized into explicit surface and implicit surface methods, have been proposed to reconstruct the surface of a model [10,21]. The explicit surface methods, such as the parametric surfaces method and the triangulated surfaces method, denote the exact location of a surface. In triangulated surfaces methods, a set of triangles, especially the vertices of triangles, are generated using the sample points directly. However, these methods are difficult to deal with noise and nonuniform data problems. On the other hand, implicit surfaces methods, such as the Poisson surface reconstruction method [22,23], represent the surface as a particular isocontour of a function, considering as surface fitting. One of the advantages of implicit surface methods is that they can fill up the surface holes automatically.

A directly uniform sampling surface reconstruction method is used in this paper. Note that depth maps captured by Kinect are limited by the narrow viewing angle because only a single IR camera is used to capture the structured-light pattern projected onto the 3D scene. The reconstructed point cloud is approximately spatially uniform distributed due to the uniform camera sensor matrix. The surface reconstruction from the point cloud is achieved directly by connecting neighboring points both in vertical and horizontal directions. As shown in Fig. 3, $P'(u, v)$, $P_1'(u_1, v_1)$ and $P_2'(u_2, v_2)$ are points on the image plane of the camera, and the corresponding reconstructed points in the space are $P(x, y, z)$, $P_1(x_1, y_1, z_1)$ and $P_2(x_2, y_2, z_2)$ respectively. Point P' and P_2' are the adjacent points in the vertical direction with the sampling distance $\delta v = |v_2 - v|$. Similarly, point P_1' and P_2' are the adjacent points in the horizontal direction with the sampling distance $\delta u = |u_1 - u_2|$. Applying the uniform sampling, all triangles of the model can be reconstructed from the depth map. The normal vector of each triangle can be calculated using the vertices of each triangle. The area that encloses with the corresponding vertices in the color image captured by the RGB camera is used to map as texture.

An OBJ file is always generated to record all information of the reconstructed triangle-based models, which contains 3D coordinates of vertices, normal vector of vertices, triangle faces, and the relative coordinate of texture maps. This OBJ file, i.e. the triangle-based model, is then encoded into hologram using spatial domain Fraunhofer diffraction method mentioned above. It is convenient to transfer the OBJ file from one computer to another through Internet, as shown in Fig. 1. However, holographic communication, such as the 3D holographic conference, is limited by the computational speed of hologram and the large quantity of data to transfer. In our proposed method, the computational speed is accelerated by using GPU-based CUDA programming language, and the speed of CUDA algorithm is about 500 times faster than that of the serial algorithm based on CPU [4].

Although the uniform sampling method is efficient in converting the point cloud into a triangle-based model, there are still problems in the output models, such as noise in depth and mesh holes. Fatih Calakli et al. had proposed a smooth signed distance surface reconstruction method to address the problem of noisy data and reconstruct high-quality polygon meshes with smooth color maps [24]. Shahram Izadi et al. developed a system called KinectFusion to generate 3D models with high quality and geometrically accuracy [25]. As mentioned above, the viewing angle of model captured by a single Kinect is very narrow. To capture a model with a viewing angle of 360 degrees, either multiple Kinects are used or time-series models are captured by single Kinect [26]. On the other hand, there are missing data in the captured depth maps, resulting in holes in the reconstructed triangle-based model, due to the inaccessible regions, such as the hair and eyeballs. To make up these defects of the model, the implicit surface reconstruction method is always used for its capacity of automatic hole filling [27].

5. Experiment results

To test the validity of the proposed Kinect based real-life holographic imaging method, two dynamic real-life scenes are captured and optically reconstructed. In Fig. 4, a dynamic real-life scene is shown, in which a man walks from about 3.0 m to about 1.0 m away from the Kinect sensor and waves his hand. The experimental setup shown in Fig. 1(e) is used to obtain the optical reconstruction. Two selected

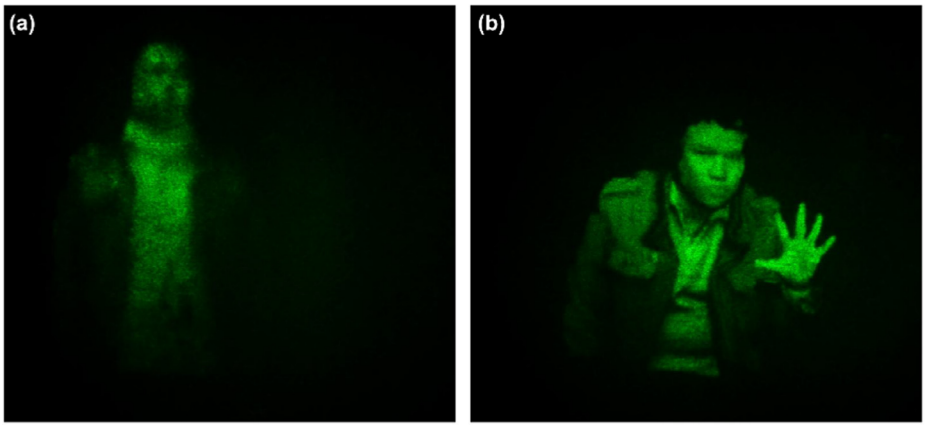

Fig. 4. Optical reconstruction of a dynamic scene in which a man walks and waves hands. Note that the camera always focuses on the man. (a) The man is far away from the Kinect, (b) the man is near to the camera. The optical reconstruction of the dynamic real-life scene can be found in the uploaded video (Video 1, MP4, 2.9 MB).

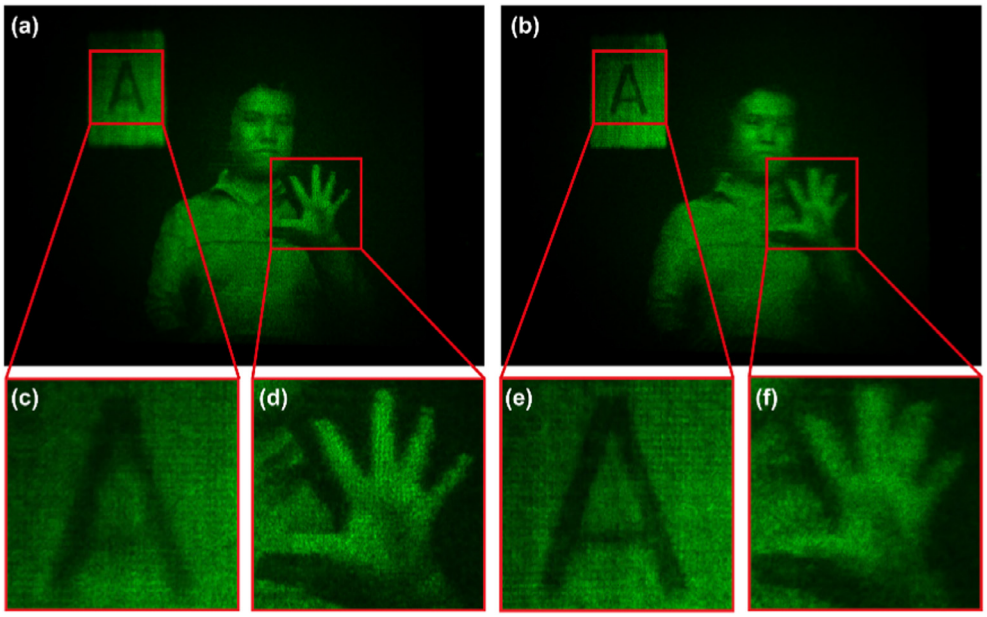

Fig. 5. Optical reconstruction of a scene in which a man sits near the camera. And an object with character "A" is behind the man as a reference. (a) The optical reconstruction captured when the camera focus on the hand. The hand is clearly reconstructed, while the character is out of focus. The detailed view is shown in (c) and (d). (b) The same reconstructed optical field captured by a camera when the character is in focus. The character is clear while the hand is defocused. More details are shown in (e) and (f). The optical reconstruction of the dynamic real-life scene with the waving hand in focus can be found in the uploaded video (Video 2, MP4, 3.1 MB).

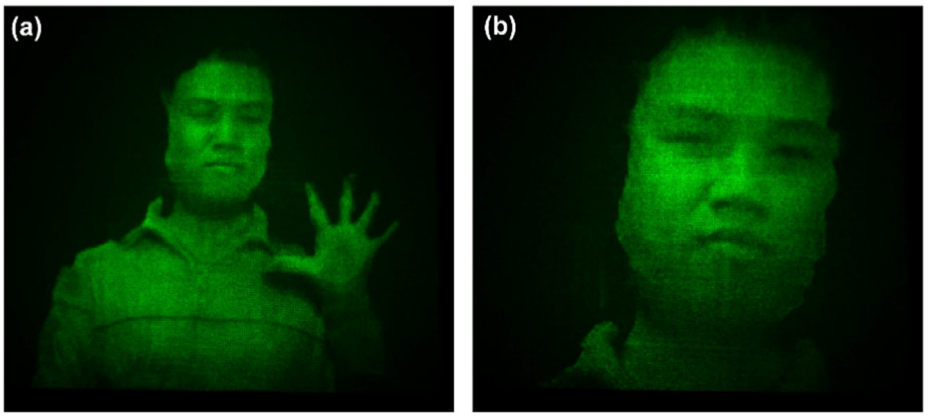

Fig. 6. Reconstructed holographic image of the same scene of Fig. 5 which is digitally magnified using depth map based triangular refinement method. (a) The digital magnification is 1.5 times. (b) The digital magnification is 3.0 times. The detailed information, such as eyes and nose, can be clearly recognized.

frames of optical reconstruction of the dynamic scene are illustrated in Fig. 4. The optical reconstruction of the man far away from the sensor is illustrated in Fig. 4(a), while in Fig. 4(b), the man waving the hand near the camera is shown. The reconstructed image near the Kinect is evidently clear than that far away from the sensor. On one hand, the digital camera always focuses on the plane near the Kinect. It means that Fig. 4(b) is in focus while Fig. 4(a) is out of focus. On the other hand, as mentioned above, the lateral resolution of model captured by Kinect is decreasing as the measurement distance increases. When the man is far away from Kinect, only the profile of the optical reconstruction is recognized. More details, such as the face and the decoration in clothes, are missed due to the sparse sampling points. When the man walks close to the sensor, the lateral resolution becomes higher, and detailed information, e.g. the eyes and nose can be clearly recognized. In November 2013, the second-generation Kinect with Xbox One, also known as Kinect V2, was released. Generally speaking, the Kinect V2 performs better than the Kinect V1. The main differences between these two devices related to our holographic display system are listed as follows. First, the resolution of the depth camera of Kinect V2 is increased. Then a finer triangle-based model can be reconstrued by using the Kinect V2. Second, the Kinect V2 has a larger field of view than Kinect V1 so that a larger scene can be captured. Third, the methods used to capture the depth maps are different. The Kinect V1 acquires the 3D information of a real-life scene based on the structured light coding method while the time of flight method is used in the Kinect V2. However, both devices have the same kind of outputs, i.e. series of depth maps of the dynamic 3D scene, which are used to generate 3D models in our proposed system. Then, the workflow of the whole system remains the same no matter which generation of Kinect is used. It means that our proposed method is applicable to both Kinect V1 and V2. Therefore, by replacing the Kinect V1 with Kinect V2, the output resolution of holographic reconstruction in our proposed method will increase.

Furthermore, objective image quality assessments, such as the peak signal-to-noise ratio (PSNR), root mean squared error (RMSE), and speckle noise, are always used to evaluate the holographic image quality [28]. In this paper, a universal image quality index (UIQI) is introduced to qualify the optical reconstruction using the proposed 3D real-life holographic imaging method [29,30]. The UIQI, one of full-reference image quality assessments, considers the relation between the test image and the reference image, which can be defined as follows,

$$UIQI = l(x, y)\cdot c(x, y)\cdot s(x, y) = \frac{2\mu_x\mu_y}{\mu_x^2 + \mu_y^2}\cdot\frac{2\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2}\cdot\frac{\sigma_{xy}}{\sigma_x\sigma_y}, \quad (4)$$

where x and y are two nonnegative image signals, i.e. the test image and reference image. μ_i and $\sigma_i\,(i = x, y)$ are the mean intensity and the standard deviation of the two images, respectively. And σ_{xy} is the

covariance between the test image and reference images. As shown in Eq. (4), there are three independent contributing components to UIQI, i.e. luminance comparison $l(x, y)$, contrast comparison $c(x, y)$ and structure comparison $s(x, y)$. Note that the value of UIQI is in $[-1, 1]$. The larger UIQI indicates a higher correlation between the test image and the reference image. The UIQI equals to 1 if and only if these two images are identical. As reported by Zhou Wang et al., the performance of UIQI is always consistent with the subjective method [30]. To calculate the UIQI of the reconstructed image of our proposed real-life holographic imaging method, an orthographic projection of the corresponding 3D triangle-based model is used as the reference image. After simple arithmetic, the UIQI of Fig. 4(b) is 0.58 while that of Fig. 4(a) is 0.22. The result clearly shows that a camera captured the reconstructed image in Fig. 4(b) is much more similar to its corresponding model than that of Fig. 4(a).

A second dynamic real-life scene, in which the man is sitting and waving the hand at the position of 1.5 m, with a character "A" behind as reference, is used to demonstrate the depth of field of the reconstructed 3D images. The optical reconstruction of this 3D scene is shown in Fig. 5. When the camera is focused on the man's hand, as shown in Fig. 5(a), the hand is clearly reconstructed, and meanwhile, the reference object, i.e. the character "A", is out of focus. The detail view is shown in Fig. 5(c) and (d), respectively. On the other hand, when the camera is focused on the reference object, the character "A" is in focus and the hand is out of focus, as shown in Fig. 5(b), (e), and (f). That is to say that the reconstructed image is a 3D scene with a large depth of field, and can satisfy observers with depth cues, such as the monocular accommodation effect. Note that the hologram is always uploaded onto SLM to obtain the optical reconstruction. With the reconstructed distance of 800 mm, the length of the maximum image area in *xy* plane is about 53 mm, since the pixel pitch of the SLM used in the experiment is 8 μm. The 3D model of real-life scene is resized to fit the image area of the SLM. So that the distance between the man and the reference object "A" is resized to be about 120 mm accordingly. The UIQI is also used to analyze the quality of reconstruction in this scene. The UIQI of the man on focus in Fig. 5(a) is 0.26, while that of the man out of focus in Fig. 5(b) is 0.23.

However, it's difficult to distinguish the detailed information on the man's face clearly, as shown in Fig. 5(a) and (b). In order to recognize more details of the reconstructed holographic image, a triangle-based model with fine mesh grid should be obtained. Unfortunately, the resolution of mesh grid of the reconstructed 3D model is determined by that of the output depth map from Kinect. The 3D model is refined into smaller triangles in order to represent the holographic image with high quality. A depth map based triangular mesh refinement method is proposed to refine the 3D model. First, the original depth map $D(x, y)$ with a total number of pixels $N_x \times N_y$ is resized with a given magnification $\alpha(\alpha > 1)$ using bicubic interpolation method. The output interpolated depth map $D(x/\alpha, y/\alpha)$ is then used to reconstruct a new triangle-based model with fine mesh wire using the directly uniform sampling surface reconstruction method as mentioned in Section 4. The computational time to encode the hologram will increase by α times due to more triangles in the new model. Consider that observers can only focus on a small area at one time, a part of the interpolated depth map with size $N_x \times N_y$ is chosen. So that the computational consuming is the same as that of the original model. The same scene in Fig. 5 is used to verify the depth map based triangular mesh refinement method. To illustrate the details of the man, the face is digital zoomed in. The corresponding holographic reconstructed image of the refined triangle-based model is shown in Fig. 6. By using the digital magnification, the resolution of the model increases by 1.5 and 3.0 times with respect to the original one, as shown in Fig. 6(a) and (b), respectively. The refined structure in Fig. 6(a) and (b), such as eyes, nose and so on, can be seen clearly.

6. Conclusions

In this paper, a Kinect based 3D dynamic real-life scene holographic imaging method is proposed, which extends the objects to encode in computer-generated hologram from virtual 3D models to real-life scenes. In this method, a Kinect sensor is used to acquire the depth information of the 3D scene and record it as a depth map. A simple pinhole model is introduced to reconstruct 3D point cloud from the depth map. Surface reconstruction is accomplished by using the directly uniform sampling method and information of the triangle-based model is exported as an OBJ file. Then the holograms are generated using spatial domain Fraunhofer diffraction algorithm. The final optical reconstructed images are then captured by a camera under coherent light illumination. A universal image quality index is introduced to evaluate the quality of the holographic reconstruction. When the reconstructed image is in focus, the UIQI index can be more than 0.5, which means a good agreement with the 3D model. Finally, a depth map based triangular refinement method is used to increase the resolution of the holographic image. And the resolution can be about 3.0 times higher than the original one.

Acknowledgments

The authors thank the technological help of Kinect by Dr. Chong Wang and Prof. Shing-Chow Chan in the University of Hong Kong. This work is supported by National Natural Science Foundation of China (Grant No. 61805288, 61775243, 11761161002), Natural Science Foundation of Guangdong Province (Grant No. 2017A030310510, 2018B030308005), Guangzhou Science, Technology and Innovation Commission (Grant No. 201804020029), and the Fundamental Research Funds for the Central Universities (Grant No. 181gpy72).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.optlastec.2019.105590.

References

[1] J.W. Goodman, Introduction to Fourier Optics, McGraw-Hill, New York, 1996.
[2] T. Nishitsuji, T. Shimobaba, T. Kakue, T. Ito, Review of fast calculation techniques for computer-generated holograms with the point light source-based model, IEEE Trans. Ind. Inform. (2017) 1.
[3] K. Matsushima, H. Nishi, S. Nakahara, Simple wave-field rendering for photo-realistic reconstruction in polygon-based high-definition computer holography, J Electr. Imag. 21 (2012) 023002-023001-023002-023008.
[4] Y.-Z. Liu, J.-W. Dong, Y.-Y. Pu, B.-C. Chen, H.-X. He, H.-Z. Wang, High-speed full analytical holographic computations for true-life scenes, Opt. Express 18 (2010) 3345–3351.
[5] Y.-Z. Liu, J.-W. Dong, Y.-Y. Pu, H.-X. He, B.-C. Chen, H.-Z. Wang, H. Zheng, Y. Yu, Fraunhofer computer-generated hologram for diffused 3D scene in Fresnel region, Opt. Lett. 36 (2011) 2128–2130.
[6] J. Geng, Structured-light 3D surface imaging: a tutorial, Adv. Opt. Photon. 3 (2011) 128–160.
[7] Wikipedia, "Kinect," https://en.wikipedia.org/wiki/Kinect, Accessed 20 July 2017.
[8] J. Stowers, M. Hayes, A. Bainbridge-Smith, Quadrotor Helicopter Flight Control Using Hough Transform and Depth Map from a Microsoft Kinect Sensor, Proceedings of the IAPR Conference on Machine Vision Applications, 2011, pp. 352–356.
[9] M.S. Islam, M. Tariquzzaman, Computer generated integral imaging (II) system using depth-camera, Int. J. Comput. Sci. Inform. Technol. Res. 3 (2015) 256–263.
[10] J.-S. Jeong, M.-U. Erdenebat, K.-C. Kwon, B.-M. Lim, H.-W. Jang, N. Kim, K.-H. Yoo, Real object-based integral imaging system using a depth camera and a polygon model, Opt. Eng. 56 (2017) 013110.
[11] J.-S. Jeong, K.-C. Kwon, M.-U. Erdenebat, Y. Piao, N. Kim, K.-H. Yoo, Development of a real-time integral imaging display system based on graphics processing unit parallel processing using a depth camera, Opt. Eng. 53 (2014) 015103.
[12] S. Xie, P. Wang, X. Sang, C. Li, Augmented reality three-dimensional display with light field fusion, Opt. Express 24 (2016) 11483–11494.
[13] Y. Takaki, J. Nakamura, Development of a Holographic Display Module Using a 4k2k-SLM Based on the Resolution Redistribution Technique, Biomedical Optics and 3-D Imaging(Optical Society of America, Miami, Florida, 2012, p. DM2C.5.
[14] V.M. Bove, Live Holographic TV: From Misconceptions to Engineering, SMPTE 2nd Annual International Conference on Stereoscopic 3D for Media and Entertainment,

2011, pp. 1–8.
[15] J. Barabas, S. Jolly, D.E. Smalley, J.V.M. Bove, Diffraction specific coherent panoramagrams of real scenes, Proc. SPIE 7957 (2011) 795702.
[16] J. Kramer, M. Parker, D. Herrera C., N. Burrus, and F. Echtler, Hacking the Kinect (Apress, 2012).
[17] B. Freedman, A. Shpunt, M. Machline, Y. Arieli, Depth Mapping Using Projected Patterns, PRIME SENSE LTD (Tel Aviv, IL), United States, (2010).
[18] J. Smisek, M. Jancosek, T. Pajdla, 3D with Kinect, Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference, 2011, pp. 1154–1160.
[19] K. Khoshelham, Accuracy Analysis of Kinect Depth Data, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XXXVIII-5/W12, 2012, pp. 133–138.
[20] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2004.
[21] S.P. Lim, H. Haron, Surface reconstruction techniques: a review, Artif. Intell. Rev. 42 (2014) 59–78.
[22] M. Kazhdan, M. Bolitho, H. Hoppe, Poisson surface reconstruction, Proceedings of the fourth Eurographics symposium on Geometry processing (Eurographics Association, Cagliari, Sardinia, Italy, 2006, pp. 61–70.
[23] M. Bolitho, M. Kazhdan, R. Burns, H. Hoppe, Parallel Poisson Surface Reconstruction, Proceedings of the 5th International Symposium on Advances in Visual Computing: Part I (Springer-Verlag, Las Vegas, Nevada), 2009, pp. 678–689.
[24] F. Calakli, G. Taubin, SSD: smooth signed distance surface reconstruction, Comput. Graphics Forum 30 (2011) 1993–2002.
[25] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, A. Fitzgibbon, KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera, Proceedings of the 24th annual ACM symposium on User interface software and technology ACM, Santa Barbara, California, USA, 2011, pp. 559–568.
[26] K. Berger, K. Ruhl, M. Albers, Y. Schröder, A. Scholz, J. Kokemüller, S. Guthe, M. Magnor, The capturing of turbulent gas flows using multiple Kinects, 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), 2011, pp. 1108–1113.
[27] H. Xie, K.T. McDonnell, H. Qin, Surface Reconstruction of Noisy and Defective Data Sets, Proceedings of the conference on Visualization '04 IEEE Computer Society, 2004, pp. 259–266.
[28] Y. Qi, C. Chang, J. Xia, Speckleless holographic display by complex modulation based on double-phase method, Opt. Express 24 (2016) 30368–30378.
[29] W. Zhou, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (2004) 600–612.
[30] W. Zhou, A.C. Bovik, A universal image quality index, IEEE Signal Process Lett. 9 (2002) 81–84.

Xiao-Ning Pang received his BS and PhD degrees from Sun Yat-sen University, China. He is currently an associate research fellow at the School of Physics, Sun Yat-sen University, China. His research interests include holographic three-dimensional display and nanophotonics.

Shao-Ji Jiang graduated from Sun Yat-sen University in 2004 with a PhD degree in optical engineering. He is currently a full professor at the State Key Laboratory of Optoelectronic Materials and Technologies, Sun Yat-sen University, China. His research interests include optical thin-film component and optical fiber sensors.

Jian-Wen Dong is a full professor at the School of Physics and State Key Laboratory of Optoelectronic Materials and Technologies, Sun Yat-sen University, China. His research interests include holographic three-dimensional display and nanophotonics.