Low-delay single holoscopic 3D computer-generated image to multiview images

Taha Alfaqheri[1] · Akuha Solomon Aondoakaa[1] · Mohammad Rafiq Swash[1] · Abdul Hamid Sadka[1]

Abstract

Due to the nature of holoscopic 3D (H3D) imaging technology, H3D cameras can capture more angular information than their conventional 2D counterparts. This is mainly attributed to the macrolens array which captures the 3D scene with slightly different viewing angles and generates holoscopic elemental images based on fly's eyes imaging concept. However, this advantage comes at the cost of decreasing the spatial resolution in the reconstructed images. On the other hand, the consumer market is looking to find an efficient multiview capturing solution for the commercially available autostereoscopic displays. The autostereoscopic display provides multiple viewers with the ability to simultaneously enjoy a 3D viewing experience without the need for wearing 3D display glasses. This paper proposes a low-delay content adaptation framework for converting a single holoscopic 3D computer-generated image into multiple viewpoint images. Furthermore, it investigates the effects of varying interpolation step sizes on the converted multiview images using the nearest neighbour and bicubic sampling interpolation techniques. In addition, it evaluates the effects of changing the macrolens array size, using the proposed framework, on the perceived visual quality both objectively and subjectively. The experimental work is conducted on computer-generated H3D images with different macrolens sizes. The experimental results show that the proposed content adaptation framework can be used to capture multiple viewpoint images to be visualised on autostereoscopic displays.

Keywords Holoscopic 3D imaging · Macrolens array · Low delay · Image disparity · Viewpoint image · Multiview image · Autostereoscopic display

1 Introduction

In recent years, there has been an increased interest in exploring light field techniques in 3D visualisation. These techniques have been used extensively for solving the problems related to eye fatigue and discomfort viewing experience from wearing the 3D display glasses. Due to recent advances in theory and microlens manufacturing, holoscopic 3D (H3D) technology is becoming a practical, prospective 3D display technology and is thus attracting much interest in the 3D visualisation sector.

The main concept of H3D, also known as Integral imaging technology, is based on the fly's eye imaging system concept. The concept of this technology is based on the proposal of Gabriel Lippmann in [1]. He proposed an analogue multi-lens photo plate that is "reminiscent of the insect's compound eye, where a large number of crystalline lenses are arranged hexagonally.

At the capturing stage, the conventional 2D camera loses angular information of the incident light rays on the 2D camera sensor [2]. To acquire more angular information, the optical representation of the fly's eye is replaced by a 2D sheet containing large number of macrolenses with lenticular or hexagonal shapes [2]. A 3D holoscopic image is recorded using a lenslets arrays closely packed together with a recording device. Each lenslet views the scene at a slightly different angle than its neighbour, so a scene is captured from different viewpoints and parallax information is recorded [3].

Stereoscopy is a visualisation technique to provide the illusion of depth to the viewer's eyes with different image offsets. The offsets images are separately displayed in the left and right of the viewer's eyes.

This visualisation technique depends on channelling the light rays to the left and right of the viewer's eyes using specific 3D glasses.

✉ Taha Alfaqheri
   taha.alfaqheri2@brunel.ac.uk

1 Electronic and Computer Engineering, Brunel University London, London, UK

A major problem with this commonly used technique is that it causes headache and eye strain for the viewers in addition to intensive post-production work [4].

Besides, stereoscopic visual contents productions with dual cameras are complicated in terms of the encoding and decoding processing aspects and costly due to the required man work efforts in post-production. This may lead to make stereoscopic unsuitable for the current market requirements. Adding to these factors, the stereoscopic displays require physical 3D display head glasses to identify the perceived 3D effects by the viewers. Further research development on the lenticular sheet is made to provide 3D visualisation that is capable of ensuring a higher motion parallax continuity for the end-users. In the 2000s, Sharp was the first company that produced a 3D LCD screen implemented on a laptop device [5].

In 2007, Philips promoted lenticular-lens-based 3D displays that create more depth information. Such displays provide more 3D depth with higher texture details at the cost of a limited motion parallax continuity [6]. Contrary to conventional 2D displays or stereoscopic displays, which are only providing a monoscopic or a stereoscopic view, 3D displays provide various numbers of viewpoints capabilities, viewing distances, and display resolutions capabilities [7].

Autostereoscopic, also known as a glasses-free display type, is first developed by Reinhard Boerner [8]. This type of displays does not only provided higher 3D depth perception result compared to stereoscopic display type but also visualise full-motion parallax continuity. Therefore, it accommodates motion parallax and allows multiple viewers simultaneously without requiring viewers to wear glasses.

These 3D experiences can be achieved without the need to wear 3D glasses [9]. The main autostereoscopic principle is using a flat panel of many macrolenses. The first practical attempt of manufacturing a prototype of this display type was in 1985 by Boerner [8].

As autostereoscopic displays provide 3D viewing perception without the need for special glasses or any head device, for this reason, this technology may lead to incorporate escalation in larger variations of next-generation 3DTV services [10]. The autostereoscopic displays can be categorised into three categories: regular two-view display, head-tracked display, and multiview with three or more views display [11].

Nowadays, there is a great deal of research interest in developing content adaptation techniques for autostereoscopic displays [12–14] these research studies aiming to provide the display viewers with richer 3D viewing experiences.

To create a stereoscopic image, two images are projected onto the same screen through different polarising filters. The viewer wears 3D eyewear that also contains a pair of polarising filters. Each filter allows only light that is similarly polarised to pass through the viewer eye, ensuring each eye sees different image. This produces the 3D effect by projecting the same scene into both eyes but depicted from slightly different viewing perspectives. In contrast, autostereoscopic imaging systems use a large number of viewing regions that are recorded and presented on displays that do not require glasses for 3D viewing. A number of such systems are commercially available [15].

Manufacturing of autostereoscopic displays can be grouped into two types: head-tracked displays and multi-view displays.

In the first type, the virtual scene cameras provide different viewing points with a limited motion parallax compared to multiview displays [9]. This type of displays requires a precise tracking mechanism to switch between different viewpoints according to various moving head directions. In the second type, a multiview display is produced with a fixed number of viewpoint images which depend on the pitch size of the manufactured lenticular sheet. In contrast to the first type, a multiviewer with higher motion parallax continuity can be achieved at a time due to increase in the number of viewpoints of the display. The perceived viewing of depth information can be distinguished from slightly different viewing angles on each eye due to the disparity between the viewpoint images. As a result, higher motion parallax can also be perceived in this type of displays [9].

Although some progress has been made to improve 3D visualisation systems in terms of user experience, further research is needed.

The need arose due to issues with the current methods of commercially available 3D visualisation systems which are tend to cause eye strain, fatigue, and headaches after a short period of time. These visual discomfort issues are attributed to the fact that the display viewers should focus on the screen plane to accommodate their eyes on the 3D display screen and converge their eyes to a point in space in different screen plane (convergence), producing unnatural 3D viewing experience [16]. In this case, the viewer still has to use his/her brain attempts to interpret the 3D visual contents.

The potential advantages of using H3D imaging over computer-generated 3D visualisation techniques are related to acquiring more viewpoints and parallax information at acquisition and visualisation stages.

The H3D data is recorded using a regularly spaced array of lenslets closely packed together with a recording device. Each lenslet views the scene at a slightly different angle than its neighbour, so a scene is captured from many viewpoints and parallax information is recorded. With H3D imaging, this process is more intuitive because the images are captured by recording more angular and natural light information without having to create a mental 3D image or use 3D eyewear [17].

The performance of such light field systems, however, has been limited by the camera sensor size (number of pixels) and of the characteristics of the microlens array [18, 19].

These two factors are calculated according to the sampling theorem to optimise the trade-off results between spatial and angular resolution of the recorded light field image [18].

The motion parallax continuity depends on how many the viewpoints can be generated at the same in the display. The 3D visual contents generated from multiview capturing system are captured using either multi-camera rigs set-up or single moving 2D imaging system. Several issues are arising from multiview capturing system requirements such as multiview camera installation set-up, colour mismatch and viewpoint image misalignment that are difficult to deal with in post-processing stage.

A single-aperture holoscopic camera can reduce the related calibration issues of multiview set-ups and implementation cost by eliminating the calibration issues of multiview set-ups.

The holoscopic 3D imaging system is designed depending on the fly's visual system; this is made possible by the introduction of the macrolens array placed right before the imaging sensor, making it the key component that differentiates an H3D imaging system from a 2D imaging system. This macrolens array gives the H3D imaging system the ability to record the angular and spatial information of any given scene. The MLA can be grouped into two categories: omnidirectional OH3D and unidirectional H3D types. Figure 1 shows a practical example using both MLA types. The type of the generated elemental images in H3D system depends on the used MLA characteristics at acquisition stage.

The acquisition with omnidirectional MLA provides more directional information (with recording full parallax of a scene) compared to unidirectional MLA. This type of MLA acquires more angular information at the cost of lower spatial information. The unidirectional MLA using vertical cylindrical MLA alignment can capture only horizontal parallax movement of the scene. Therefore, by using omnidirectional MLA, we can get more angular information in each

elemental image with different directions and full-motion parallax continuity. In omnidirectional H3D images, each part is represented by a group of replicated elemental images with a slightly different capturing angle.

One of the main holoscopic 3D applications is transforming raw H3D images into multiview images. The required calibration time at multiview capturing stage could be reduced significantly using H3D imaging system with higher reconstruction accuracy. Additionally, using H3D acquisition concept in autostereoscopic visualisation system improves the visualisation system performance by reducing required viewpoints extraction time at the post-processing stage.

To overcome the problem of generated low resolution viewpoint images from H3D images, we evaluated and optimised computer-generated images using H3D capturing concept with different MLA sizes. The evaluation work has been conducted at autostereoscopic visualisation stage. Furthermore, H3D content adaptation framework for autostereoscopic display presented and evaluated with two upsampling techniques (bicubic and nearest neighbour interpolation techniques) in terms of subjective and objective quality assessments.

The rest of this paper is organised as follows. Section 2 provides overview of the autostereoscopic display and H3D imaging system. In Sect. 3, a holoscopic adaptation framework for multiview visualisation is described. Section 4 presents experimental evaluation set-up for the proposed work and the achieved results. The section includes the characteristics of computer-generated H3D test images, the used evaluation methodologies in terms of quality image assessment, objective metric, and computational complexity. Section 5 concludes the proposed work with future work recommendations.

2 Hardware specifications and H3D imaging properties

This section is divided into two subsections. In the first subsection, construction of the autostereoscopic 3D display system is described. In the second subsection, an H3D capturing imaging system and its properties are presented, highlighting the process of generating holoscopic 3D computer-generated images.

2.1 Autostereoscopic display properties

The main feature of the autostereoscopic display is that it can visualise the generated 3D visual contents without the need of wear display glasses, or other devices at the viewer's end. This type of displays can be divided into four viewing scenarios: (1) electro-holographic where wave-front
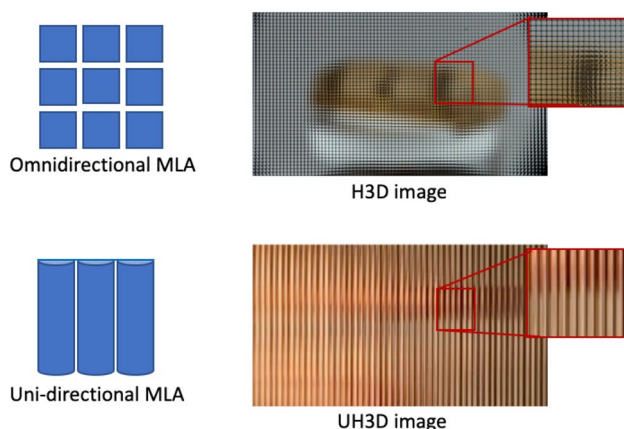

Fig. 1 Macrolens array types

reconstruction is used to visualise the image on the autostereoscopic display, (2) volumetric display where the image is formed within a volume of space without the use of light interference, (3) multiple images where the viewing field displays many 2D images at the same time, and (4) a light field display where the projector and macrolens arrays form the final 3D image [20]. A simple schematic diagram showing the main components of an autostereoscopic display is depicted in Fig. 2.

The viewing range of the autostereoscopic display depends on the maximum number of generated viewpoint images the display can provide simultaneously. This capability enables the viewers to perceive different levels of depth and motion parallax at a fixed viewing range.

2.2 Holoscopic 3D imaging system, 3D image properties and dataset

This section presents the components of the holoscopic 3D image acquisition process, its image properties and the synthetic data, in addition to the macrolens array specifications and their effects on the reconstructed autostereoscopic image quality. A prototype holoscopic 3D camera is shown in Fig. 3a.

The main components of the holoscopic 3D camera are macrolens array, relay lens and digital camera sensors. Its schematic diagram is featured in Fig. 3b. The macrolens array (rectangular white layout) is mounted at a predefined distance to the camera sensor as shown in Fig. 3b. The main lens image plane is placed in front of the macrolens array, which allows the macrolens array to capture the positions in the scene from different perspectives. The combination of lens layouts integrated with the Sony camera sensor model alpha mark ii is shown in Fig. 3a. The total weight of the prototype H3D camera is 1.57 kg making it lighter than other multiview capturing systems and more flexible in mobility than most stereo imaging rigs [21].
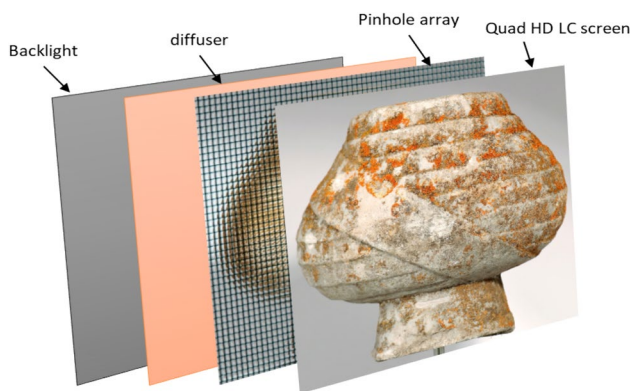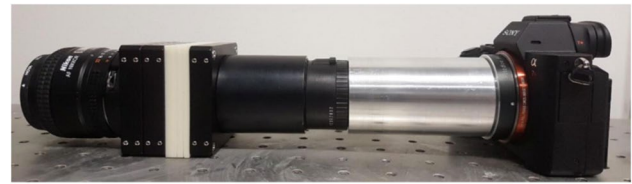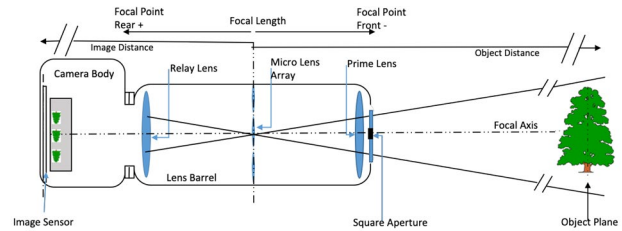

(a) Prototype Holoscopic 3D camera at Brunel University

(b) Schematic diagram of Lightray distribution and registration process

Fig. 3 Holoscopic 3D capturing

The synthetic holoscopic 3D data presented in this paper is recorded with a virtual holoscopic 3D imaging system with the same system design demonstrated in Fig. 3, with the only exception of different macrolens array sizes.

The acquisition process in holoscopic 3D camera is done by recording the camera sensor light rays with different angular information due to diffraction or splitting of viewpoint light arrays. The diffracted light arrays of the prime lens are then passed through the macrolens arrays, making their way to the relay lens and then finally recorded on the imaging sensor. The recorded light ray from the same viewpoint is captured on the imaging sensor with equal step size intervals between viewpoints. The final holoscopic 3D image is represented by replicating the group of macroimages captured with slightly different angles (Fig. 5). The H3D capturing process concept is illustrated in Fig. 4.
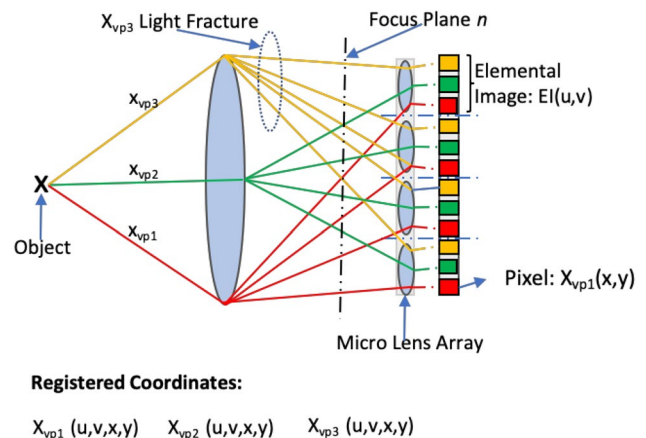

Fig. 2 Main components of an autostereoscopic display
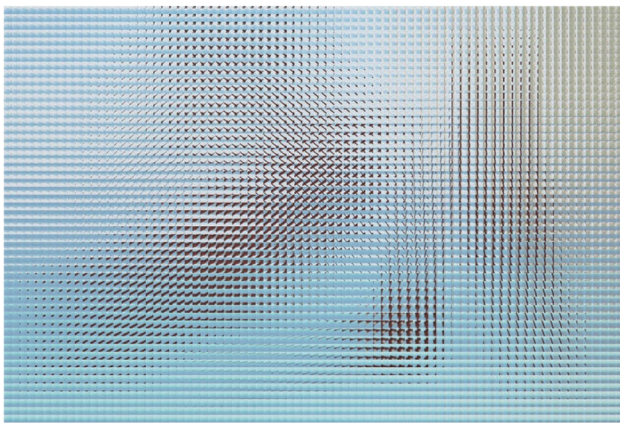

Fig. 4 Holoscopic 3D image registration concept

Fig. 5 Holoscopic 3D synthetic image data sample

Thus, the synthetic holoscopic 3D data presented in this paper is recorded with a virtual holoscopic 3D imaging system with the same specifications mentioned and as demonstrated in Fig. 4. The main components of the holoscopic 3D imaging system are the macrolens array, camera sensor, square aperture and relay lens array. The square aperture controls the amount of light that is allowed into the camera and helps to avoid severe vignetting the generated image corners. The final generated image, as depicted in Fig. 5, shows a sample of the final computer-generated image with image resolution (7952×4472 p.).

In the experimental work, the properties of the generated synthetic test images are reported in Table 1. The synthetic images are generated from the original image size 7952×4472 p.

The extracted viewpoint images are extracted from elemental images with resolution ranging from 10×10 pixels to 100×100 pixels. The implemented camera sensor size was 35.9×24 mm with a dot pixel pitch of 0.0041 mm.

The macrolens array records multiple viewpoint images; however, the initial H3D data is recorded with inverse depth before being corrected by the relay lens for data registration at the camera sensor. Based on the implemented macrolens size, the elemental holoscopic images are recorded with different resolutions. In another way, the elemental image size is the same as the total number of viewpoint images recorded in that single image.

Figure 6 presents the synthetic holoscopic 3D extracted viewpoint results achieved with different MLA sizes. As can be seen from the figure, the resolution of the extracted viewpoint images is increased with increase in the MLA size as indicated in Table 1.

In the experimental work, the standard disparity size between viewpoint images is 0.02 cm, which is equivalent to 5 pixels recorded by the H3D imaging sensor used for this paper. The viewing position in front of the multiview
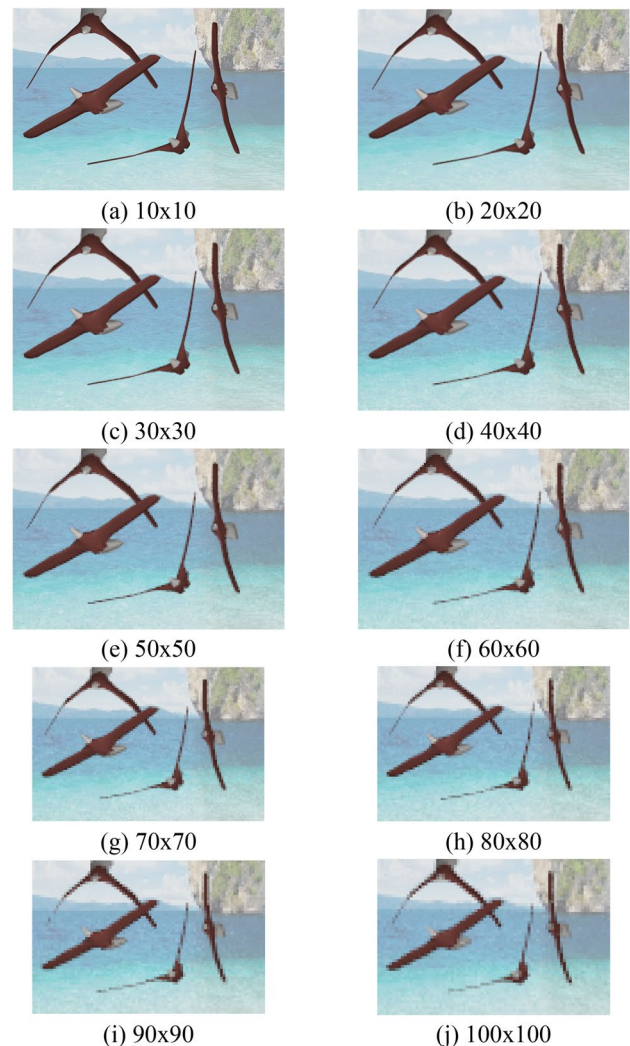
Table 1 Synthetic H3D image resolution vs different MLA sizes

Elemental image size (pixel)	Macrolens array (MLA) size	Viewpoint image size (pixel)
100×100	79×53	79×53
90×90	88×58	88×58
80×80	99×66	99×66
70×70	113×75	113×75
60×60	132×88	132×88
50×50	159×106	159×106
40×40	198×132	198×132
30×30	265×178	265×178
20×20	397×265	397×265
10×10	795×533	795×533

(a) 10x10   (b) 20x20
(c) 30x30   (d) 40x40
(e) 50x50   (f) 60x60
(g) 70x70   (h) 80x80
(i) 90x90   (j) 100x100

Fig. 6 H3D viewpoints images resolution without interpolation

display plays an important factor in experiencing 3D depth information and motion parallax continuity. Furthermore, the pitch size of the implemented lenticular sheet on the multiview display will affect the perceived scaling depth from horizontal disparities.

3 Proposed holoscopic 3D content adaptation framework

The holoscopic 3D content adaptation workflow for multiview displays is divided into three major steps, namely (1) holoscopic 3D content recording, (2) selective H3D viewpoint image extraction, and (3) viewpoint image interlacing, as shown in Fig. 7. This section describes the processing pipeline for H3D contents adaptation on multiview display.

3.1 Holoscopic 3D content capturing

The first step in the content adaptation workflow is the capturing of good quality holoscopic data.

At this stage, the generated nonlinear distortion from the optical design of the holoscopic camera lens (MLA alignment with camera sensor) is corrected. The correction process includes H3D camera calibration process presented in [21]. A white background is used in the calibration process, and a grid of circular dots is generated from the implemented MLA [21].

In order for a holoscopic 3D dataset to be classified of sufficiently good quality, the following conditions must be met:

- Resolution should be at least 1080 p and above.
- The square aperture must be properly aligned to produce a fine square grid of elemental images.
- There should be no overlapping or ghosting between elemental images.
- Image noise should be kept to the minimum value.
- The image should be in full-frame focus.

Once all the following requirements are met, the possibility of recording/capturing high-quality 3D images is greatly improved. Figure 8 shows an extracted single frame from
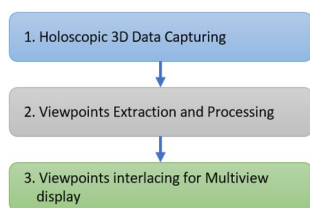

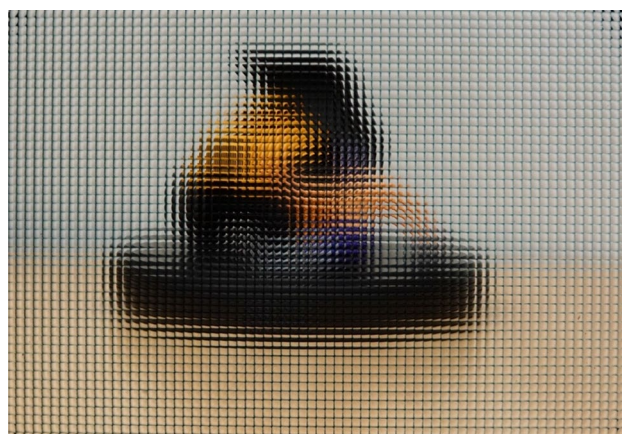**Fig. 8** Extracted frame from holoscopic 3D video data

holoscopic 3D video captured in normal room lighting conditions. A total number of 90×64 active MLA (fully framed MLA) is used in the capturing process.

There are two capturing methods: linear and angular. In the linear method, a recording of targeted objects is mostly applied when the dataset is image-based or static. This allows viewers to mainly see the depth information of the scene without being able to experience multiview scene when the user changes his/her position or changes viewing angles.

In regard to the angular method of recording, the aim here is to record a slightly different viewing angle of the scene, resulting in the display viewers being able to experience motion parallax and depth. This results in the extraction of multiview point images that are closely packed together, providing higher motion parallax continuity when moving slightly with different viewing images, is very important. Figure 9 is an illustration of the two capturing methods used in recording 3D data for the autostereoscopic 3D display.

When the holoscopic 3D camera is being moved linearly while recording, only autographic information of a scene is captured, resulting in the visualisation of the same viewpoint at different locations. However, when
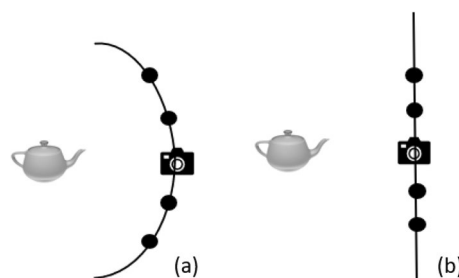

Fig. 7 Holoscopic 3D content adaptation framework

Fig. 9 Autostereoscopic capturing approaches. **a** Angular capturing. **b** Linear capturing
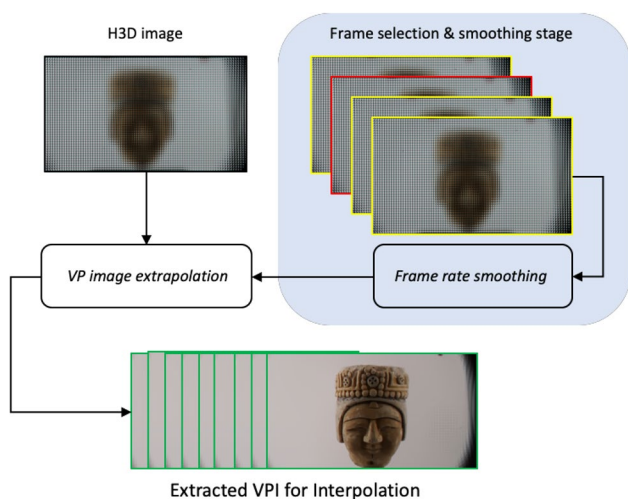
Fig. 10 Viewpoint extraction workflow for holoscopic 3D data

Fig. 11 Basic viewpoint (VP) image extraction from an OH3DI consisting of nine elemental images (EI)

angular motion is applied, different viewpoints can be seen along with depth information depending on the viewer's location.

3.2 Selective H3D viewpoint extrapolation

Figure 10 demonstrates the implemented process workflow for viewpoint extraction of raw holoscopic 3D data. The best pixel selection for viewpoint extraction process depends on achieving the best trade-off between the perceived visual quality and perceived 3D depth.

The difference between the two data types used for viewpoint extrapolation is that holoscopic video data will include two extra steps before viewpoint image extraction, while the H3D image data does not. The first step of viewpoint extraction in H3D video data includes frame separation and selection; this is where keyframes are selected when there is a significant activity change from the last selected frame. In the second step, the H3D video data goes through a frame rate smoothing process to ensure no skipping from frame to frame. The selection and smoothing stages are introduced into the viewpoint extrapolation workflow to reduce computational expense when dealing with H3D video data. Depending on the video time duration and frame rate, one could end up with H3D image frames up to tens of thousands, leading to a high amount of computational power required to complete the task in near real-time. The following step is the extraction of viewpoint images.

To extract VPIs from the H3DI, the computational reconstruction algorithm generates VPIs independently by superimposing the pixels from all EIs, as shown in Fig. 11.

The following viewpoint extraction process can be defined as in Eq. (1):
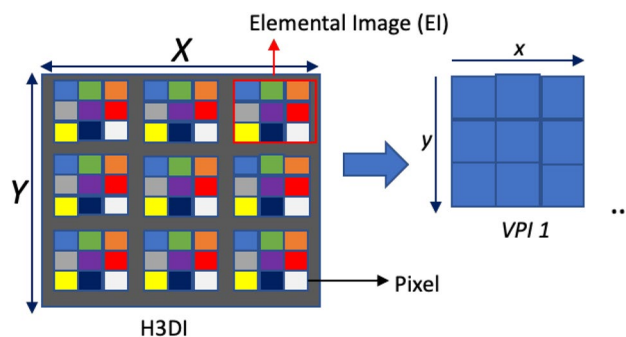
$$\text{VPI}(k, l) = \sum_{k=1}^{K-1} \sum_{l=1}^{L-1} \text{EI}_{k,l}(x - kS, y - lS), \tag{1}$$

where VP(k, l) is the extracted viewpoint image, EI(k, l) is the coordinates of the elemental images making up the holoscopic image, and (S) is the shift step size used in extracting all corresponding pixels that make up a particular image.

As discussed earlier in the previous sections, the corresponding viewpoint image size is always the same as the number of elemental images, as illustrated in Fig. 11. As for this paper, the viewpoint extraction process presented in Eq. (1) is further extended to accommodate a patch-based interpolation technique that reconstructs high-definition viewpoint images (HDVPI) by superimposing neighbouring viewpoint images with the shift and integrate Eq. (2) [22].

$$\text{HDVPI}_{ij} = \sum_{k=1}^{N} \sum_{p=1}^{N} V_{i\pm\Delta(1\mp k),k,j\pm\Delta(1\mp p),p}, \tag{2}$$

where HDVPI$_{ij}$ is the result of the reconstructed high-definition viewpoint image and coordinates i, j; k, p are the indexed number of VP ranging from 1 to N.

On the other hand, the image quality of the extracted viewpoint images can be improved without the introduction of noise by using different MLA characteristics. By increasing the number of macrolenses in MLA, the default spatial resolution of the extracted image is automatically increased. Accordingly, the patch size or the number of neighbouring viewpoint images which are required for interpolation is reduced, decreasing the amount of noise in the final constructed multiview image.

However, this increase in spatial resolution comes at the cost of the disparity range of multiviewpoint images. Therefore, finding the trade-off between image quality and image perceived depth is important and must be considered in autostereoscopic visualisation.

3.3 Multiviewpoint interlacing

Multiviewpoint interpolation consists of remapping and interlacing the extracted viewpoint images together to be visualised on the autostereoscopic display. Figure 12 illustrates the principle behind the viewpoint interlacing technique for autostereoscopic displays.

The extracted viewpoint images (*n*) are remapped in such a way that each lenticular strip refracts pixels of similar viewpoint images to a specific direction and while the viewer moves, the chance of visualising other viewpoint images becomes possible.

The overall process of the proposed adaptation framework for H3D content is demonstrated in Fig. 12.

In the following sections, we present the evaluation work results and supportive narrative analysing the obtained viewpoint extracted images at different MLA sizes, computational complexity, and interpolation techniques (Fig. 13).

4 Experimental results

The holoscopic depth vs viewpoint image resolution trade-off analysis is presented in this section. Moreover, a brief introduction to the evaluation platform and its metrics is presented. This is done to facilitate a better understanding of the nature of errors that occur during the viewpoint extrapolation process.

An experimental study is conducted to point out how much-unwanted artefact is introduced in a viewpoint image. The objective image quality evaluation uses the conventional PSNR metric. The section concludes with the best trade-off specifications.

4.1 Evaluation methodology

In this section, an evaluation platform to evaluate different interpolation techniques on extracted H3D images is demonstrated in Fig. 14. The extracted 2D image is compared
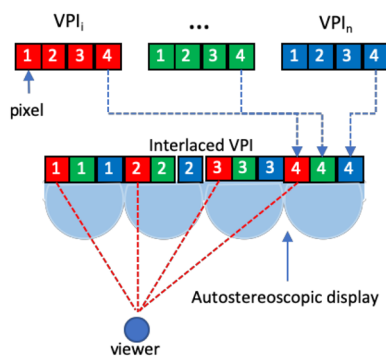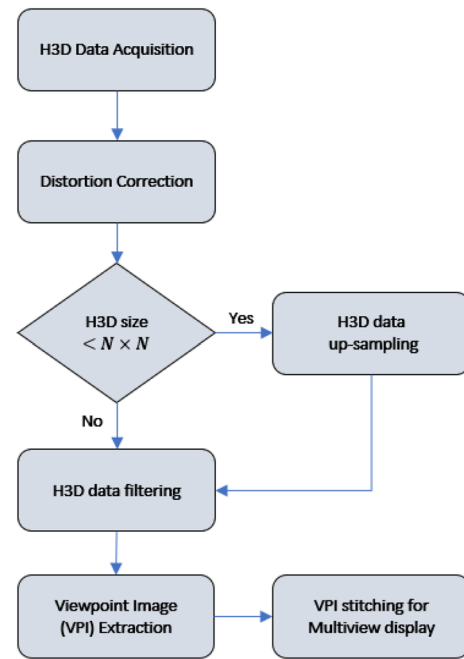

Fig. 13 Holoscopic 3D to multiview conversion process

to different image interpolation methods. An evaluation is applied to all the extracted viewpoints. The number of viewpoints depends on how many viewpoints the multiview display can display at one time. The objective evaluation is implemented on the extracted viewpoints images $(N-1)$, where N is the maximum number of viewpoints for the implemented multiview display.

The most widely used objective quality metric is in terms of peak signal to noise ratio (PSNR), which can emulate the perceived video quality as observed by the human visual system it has been used widely in image and video processing measurements due to it being relatively simple [23]. Furthermore, the PSNR quality metric is considered to be one of the most reliable indicators of visual quality variations in image/video processing algorithms in both industry and academia, and a reference benchmark for video quality evaluations in the context of developing video coding standards [24].
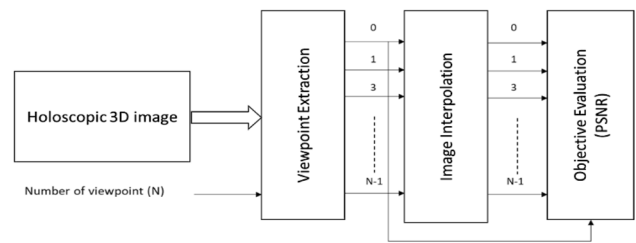

Fig. 12 Viewpoint image interpolation process of a multiview display

Fig. 14 Holoscopic 3D evaluation platform

A referenced image is generated from Pov-Ray software. The viewpoint images are represented as reference pictures and without any form of image interpolation. Then, the processed images are upsampled with the nearest neighbour and bicubic, respectively.

The PSNR calculation is based on finding a ratio between the signal energy and noise energy. For each frame, a pixel in the Luminance Y component of the reference frame (signal energy) is compared with the corresponding pixel from the processed frame pixel (noise energy). PSNR can be calculated on a logarithmic scale as shown in Eq. (3) [23]:

$$\text{PSNR} = 20\log_{10}\left[\frac{(2^n - 1)}{\sqrt{\frac{1}{x}\sum_i\sum_j\left(Y_{\text{ref}}(i,j) - Y_{\text{prc}}(i,j)\right)^2}}\right], \quad (3)$$

where $(2^n - 1)$ is the square of the peak signal value, (n) represents the number of bits per pixel in the luminance component, $Y_{\text{ref}}(i,j)$ represents the pixel values of the referenced image, $Y_{\text{prc}}(i,j)$ represents the pixel values of the processed frame, and X is the total number of pixels in the frame [23].

4.2 Evaluation of interpolation-based upsampling techniques

The holoscopic 3D imaging technology has the ability to record the full parallax of any given scene; the work associated with this paper takes advantage of that unique ability to aid the creation of multiple viewpoints for multiview displays since this serves as a cost-effective mechanism for capturing multiview content. However, the H3D inability to produce high-resolution viewpoint images poses a serious challenge and results in a fragile multiview camera set-up that is prone to acquisition errors like image misalignment and colour mismatch that is extremely difficult to fix in post-production.

The analysis of the trade-off between holoscopic depth and viewpoint image resolution is presented in this section. A brief introduction to the interpolation-based image upsampling technique is presented. This aims at providing the readers with an appreciation of the nature of errors that occur during the viewpoint extrapolation process.

An experimental study is conducted to point out how much-unwanted artefacts are introduced in the generated viewpoints. The image quality is evaluated using the objective PSNR metric. The experiments are conducted using three preselected synthetic images, namely Cone, Dice and Bird (Fig. 15).

The nearest neighbour, bilinear and bicubic techniques are the most common upsampling techniques used in today's image processing applications. The nearest neighbour upsampling technique upsamples images by
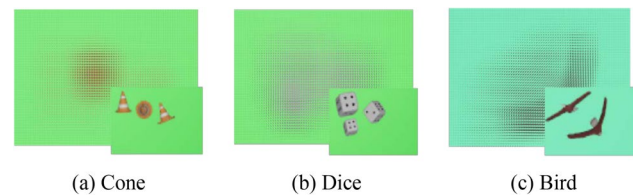

(a) Cone (b) Dice (c) Bird

Fig. 15 Synthetic test images

Table 2 Effects of varying interpolation step sizes on (PSNR) levels for (cone image)

Image resolution	Macrolens array size	Average PSNR	
		Nearest neighbour interpolation	Bicubic interpolation
795×533	10×10	36.9702	40.3415
395×265	20×20	32.9798	35.4198
265×178	30×30	30.8665	32.5521
198×132	40×40	29.574	30.8291
159×106	50×50	28.5545	29.7666
132×88	60×60	27.9159	29.0576
113×75	70×70	27.2896	28.4607
99×66	80×80	26.9702	28.0415
88×58	90×90	26.5798	26.4198
79×53	100×100	26.2745	26.1674
Average		29.3975	30.7056

predicting the value of pixels based on their neighbours; this can lead to pixelated images particularly when the upscaling factor is large. The bilinear upsampling technique upscales images by considering the weights of two neighbouring pixels resulting in a more consistent transition during image upscaling; however, this interpolation technique tends to wash out feature edges [25]. The bicubic being the most accurate of the three techniques takes in a complex sixteen weight matrix to derive pixel values that are introduced during image upsampling [26]. However, depending on the step size, these techniques could still result in producing unpleasant results. Tables 2, 3 and 4 report the obtained objective quality results for the nearest neighbour and bicubic interpolation techniques using different image resolutions. The overall PSNR levels of the three processed images are shown in Fig. 16.

From the achieved PSNR results, it is clear that bicubic interpolation preserves far more details compared with nearest neighbour upscaling interpolation. However, both interpolation techniques result in increased image artefacts in the obtained images. In view of the overall PSNR results featuring in Fig. 17, the quality of the extracted viewpoint images increases as the step size reduces.

Table 3 Effects of varying interpolation step sizes on (PSNR) levels for (Dice image)

Image resolution	Macrolens array size	Average PSNR	
		Nearest neighbour interpolation	Bicubic interpolation
795×533	10×10	36.1699	39.1291
395×265	20×20	32.508	34.6637
265×178	30×30	30.4311	32.3013
198×132	40×40	28.9616	30.7435
159×106	50×50	28.1377	29.7876
132×88	60×60	27.2164	28.8643
113×75	70×70	26.6171	28.0446
99×66	80×80	25.9999	27.3845
88×58	90×90	25.5582	26.8559
79×53	100×100	25.0435	26.2966
Average		28.6643	30.4071

Table 4 Effects of varying interpolation step sizes with PSNR values for (bird image)

Image resolution	Macrolens array size	Average PSNR	
		Nearest neighbour interpolation	Bicubic interpolation
795×533	10×10	33.1343	36.4722
395×265	20×20	29.2368	31.2757
265×178	30×30	27.2622	28.9457
198×132	40×40	25.9195	27.4454
159×106	50×50	24.9655	26.3631
132×88	60×60	24.4068	25.6595
113×75	70×70	23.6355	24.9141
99×66	80×80	23.1678	24.4908
88×58	90×90	22.7137	23.9981
79×53	100×100	22.2401	23.5911
Average		25.6682	27.3155

Fig. 16 Objective quality results using nearest neighbour and bicubic interpolation techniques with different MLA sizes

This is due to the fact that the starting resolution of the extracted viewpoint image is higher than the current standard viewpoint images. However, depending on the pitch size of the lenticular sheet used in the autostereoscopic display, this value can change. As for the autostereoscopic display used in this work, the disparity size is estimated to be a distance of 5 pixels, meaning that if all the first pixels of the array elemental images are used to reconstruct the first viewpoint image, then the next five are ignored, and the sixth pixels are used to reconstruct the second viewpoint image, and so on. With this in mind, the disparity range needed for creating content suitable for autostereoscopic display falls within the range of 30×30 pixels to 60×60 pixels per elemental image. With
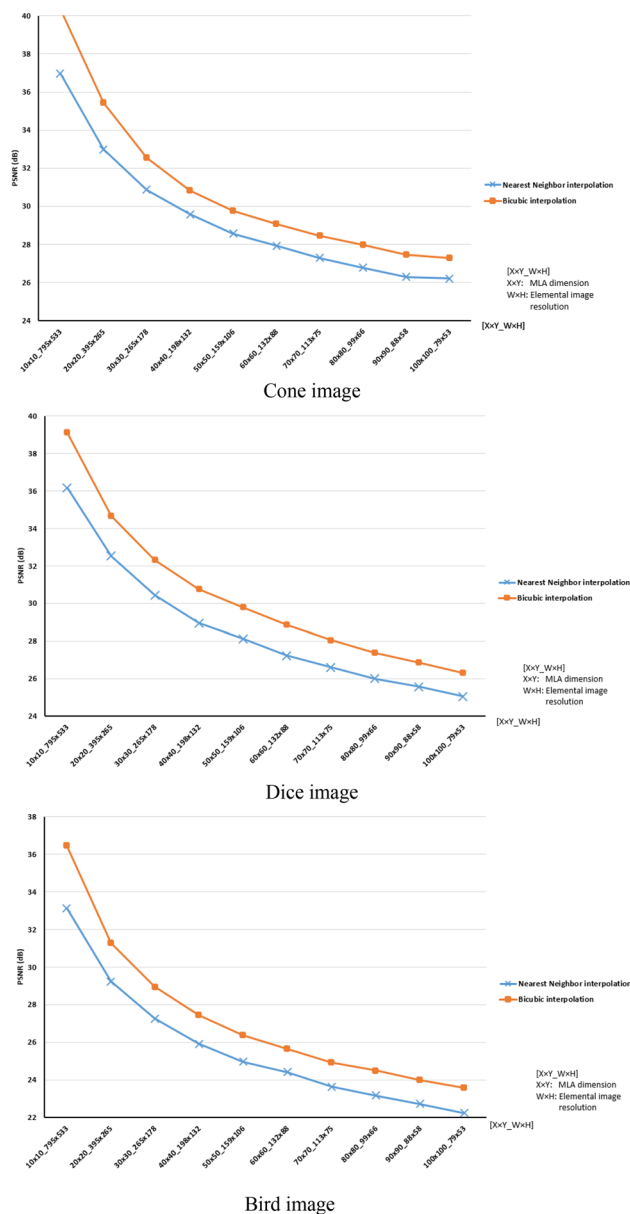
fewer pixels to be interpolated, the possible introduction of unwanted artefacts is reduced, leaving content creators with the option of using the upsampling image technique that is most suitable to their needs and requirements.

4.3 Computational complexity

The computational complexity of the holoscopic content adaptation framework is measured in accordance with the consumed processing time required to extract the viewpoint images, since it is known that there is a direct correlation between step size and the generated undesired image
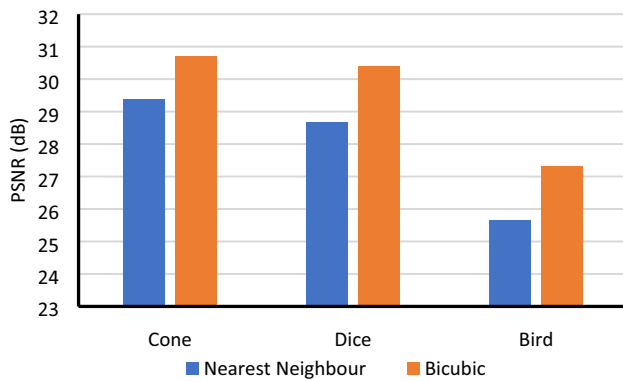
Fig. 17 Average PSNR of different extracted viewpoints

Table 5 Execution time for the viewpoint extraction process

Patch size	Processing time (s)
9	5.20
13	9.34
15	12.28
17	14.91
19	18.21
21	21.88

artefacts. This section highlights the effects of increasing the extracted viewpoint image resolution on the framework computation complexity.

Table 5 reports the processing time spent on different patch sizes. The experimental work is tested on PC with an Intel(R) Core (TM) i7-4790 @ 3.60 GHz CPU.

The viewpoint images are all extracted from the selected H3D images with an elemental image size of 90×90 pixels. As a result, the default resolution is 88×58 pixels, and larger patch size is used to upsample the extracted viewpoint image to high-definition quality. By using a larger patch size, the computational complexity increases along with the exposure to noise. However, by increasing the MLA number, the step size need can be reduced from 21×21 pixels to 9×9 pixels for the best quality and computational complexity results.

5 Conclusion

The unique feature of the holoscopic 3D imaging technique is that it can acquire more angular information than in 2D imaging. However, this advantage comes at the cost of a lower spatial resolution.

In this paper, we presented holoscopic 3D content adaptation framework that is optimised for autostereoscopic displays with low-delay constraints. The proposed adaptation framework is validated with computer-generated holoscopic

3D images. During the design stage, the best trade-off between image resolution and disparity range at the lowest computational complexity is taken into consideration based on the activity differences between elemental images.

Moreover, the nearest neighbour and bicubic interpolation techniques are selected to improve the spatial resolution of the extracted viewpoints. This pre-processing stage is essential to reduce the generated errors during the viewpoint extraction process. The evaluation work includes testing the performance of the viewpoint extraction process in terms of the perceived spatial resolution, interpolation step size, macrolens size, and the required processing time.

From the obtained results, we conclude that the proposed adaptation framework works in an efficient way with holoscopic 3D computer-generated contents. The spatial resolution of the extracted multiview images is increased with increase in step size of the viewpoint extraction process.

Moreover, the spatial resolution of the extracted viewpoint images increases, while the number of pixels in each macrolens decreases.

When employing the bicubic interpolation with the adaptation framework, the gain in objective quality increased by 1.56 dB compared with the nearest neighbour interpolation. For future work, the adaptation framework will be further optimised and evaluated for real captured holoscopic 3D content using both image and video datasets. It is recommended to implement the generative adversarial network (GAN) for image super-resolution (SR) with the proposed holoscopic 3D content adaptation framework in order to improve the spatial resolution of the extracted viewpoint images.

References

1. Lippmann, G.: Épreuves réversibles Photographies intégrales. Comptes Rendus de l'Académie des Sciences **146**(9), 446–451

(1908). **(Reprinted in Benton 'Selected Papers on Three-Dimensional Displays)**

2. Pesch, M., Velisavljevic, V., Aggoun, A., Hahne, C., Fiebig, S.: Refocusing distance of a standard plenoptic camera. Opt. Express **24**(19), 21521 (2016)
3. Aggoun, A. et al.: Immersive 3D holoscopic video system. IEEE Multimed. **20**(1), 28–37 (2013)
4. Fatah, O.A.: Post-production of holoscopic 3D image. Brunel University London, Uxbridge (2015)
5. Adrian, J. et al.: 2D/3D Switchable Displays. Shapu Giho/Sharp Tech. J. 15–18 (2003)
6. Redert, A., Berretty, R.P., Varekamp, C., Willemsen, O., Swillens, J., Driessen, H.: Philips 3D solutions: From content creation to visualization. In: Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06), pp. 429–431 (2006)
7. Blinder, D., et al.: Signal processing challenges for digital holographic video display systems. Signal Process. Image Commun. **70**(2018), 114–130 (2019)
8. Boerner, R.: 3D-Bildprojektion in Linsenrasterschirmen. Fernseh- und Kino-Technik **39**, 431–435 (1985)
9. Barré, R., Hopf, K., Jurk, S., Leiner, U.: 34.1: invited paper: transformers—autostereoscopic displays running in different 3D operating modes. SID Symp. Dig. Tech. Pap. **42**(1), 452–455 (2011)
10. Zhang, Y., Ji, Q., Zhang, W.: "Multi-view autostereoscopic 3D display", OPEE 2010–2010. Int. Conf. Opt. Photonics Energy Eng. **1**, 58–61 (2010)
11. Dodgson, N.A.: Autostereoscopic 3D displays. IEEE Comput. **38**(8), 31–36 (2005)
12. Plasencia, D.M., Karnik, A., Muñoz, J.M., Subramanian, S.: Portallax: Bringing 3D displays capabilities to handhelds. In: Human-Computer Interaction with Mobile Devices and Services, pp. 145–154 (2014)
13. de Boer, C.N., Verleur, R., Heuvelman, A., Heynderickx, I.: Added value of an autostereoscopic multiview 3-D display for advertising in a public environment. Displays **31**(1), 1–8 (2010)
14. Swash, M.R., Abdulfatah, O., Alazawi, E., Kalganova, T., Cosmas, J.: Adopting multiview pixel mapping for enhancing quality of holoscopic 3D scene in parallax barriers based holoscopic 3D displays. In: International Symposium on Broadband Multimedia Systems and Broadcasting, pp. 1–4 (2014)
15. Makanjuola, J.K., Aggoun, A., Swash, M., Grange, P.C.R., Challacombe, B., Dasgupta, P.: 3D-holoscopic imaging: a new dimension to enhance imaging in minimally invasive therapy in urologic oncology. J. Endourol. **27**(5), 535–539 (2013)
16. Shibata, T., Kim, J., Hoffman, D.M., Banks, M.S.: The zone of comfort: predicting visual discomfort with stereo displays. J. Vis. **11**(8), 11 (2011)
17. Aggoun, A., et al.: Immersive 3D holoscopic video system. IEEE Multimed. **20**(1), 28–37 (2013)
18. Bishop, T.E., Zanetti, S., Favaro, P.: Light field superresolution. In: 2009 IEEE International Conference on Computational Photography (ICCP). IEEE, San Francisco, CA, USA, 16-17 April 2009
19. Belhi, A., Bouras, A., Alfaqheri, T., Aondoakaa, A.S., Sadka, A.H.: Investigating 3D holoscopic visual content upsampling using super-resolution for cultural heritage digitization. Signal Process. Image Commun. **75**, 188–198 (2019)
20. Favalora, G.E.: Volumetric 3D displays and application infrastructure. Computer **38**(8), 37–44 (2005). https://doi.org/10.1109/MC.2005.276
21. Swash, M.R.: Holoscopic 3D Imaging and Display Technology : Camera / Processing / Display. Brunel University London (2013)
22. Fatah, O.A. et al.: Three-dimensional integral image reconstruction based on viewpoint interpolation. In: International Symposium on Broadband Multimedia Systems and Broadcasting, pp. 1–4 (2013)
23. Richardson, I.E.G.: Video Coding Design Developing Image and Video Compression Systems, 3rd edn. The Robert Gordon University, Wiley, London (2011)
24. Huynh-Thu, Q., Ghanbari, M.: Scope of validity of PSNR in image/video quality assessment. Electron. Lett. **44**(13), 800–801 (2008). https://doi.org/10.1049/el:20080522
25. Patel, V., Mistree, K.: A review on different image interpolation techniques for image enhancement. Ijetae **3**(12), 129–133 (2013)
26. Dengwen, Z.: An edge-directed bicubic interpolation algorithm. In: Congress Image and Signal Processing, vol. 3, pp. 1186–1189 (2010)

Taha Alfaqheri received his B.Sc. and M.Sc. degrees from Electrical and Electronic Engineering and Communications Engineering department/University of Technology, Baghdad, in 2007 and 2009, respectively. He received his Ph.D. degree in Electronic & electrical engineering from Brunel University London, London, in 2020. He is currently researcher at Electronics and Computer Engineering Department/Brunel University London, London. His research interests are 2D/3D visual signal processing, Machine learning, and Visual Communications.

Akuha Solomon Aondoakaa was born in Benue, Nigeria, in 1992. He is a final year Ph.D. candidate in Electrical engineering and research assistant at Brunel University London. He received a bachelor's degree in Computer science and a master's degree in Advanced multimedia and 3D technologies from Brunel University London. His current research focuses on estimating depth information from 3D images, Image sampling using deep-learning and big 3D image data analysis.

Mohammad Rafiq Swash is a seasoned professional with a decade of experiences in international technology-driven industries, academia and world-class research centres. In particular, he has years of professional working experience in 3D imaging and display technologies, expert systems, autonomous vehicles, IOT, advanced imaging and visualisation, information search & retrieval, knowledge representation with behavioural and pattern modelling. Dr Swash was a research scientist and contributed in multi-millions funded innovation research projects. In the past, he has worked as a senior software engineer, implementation manager, technical lead, and senior technology architect in tech-related industries. Dr Swash is currently the MSc programme director for Digital Design and Branding at Brunel University London, and his teaching includes multidisciplinary areas such as MSc Advanced Multimedia Design and 3D Technologies, MSc Digital Design and Branding, BSc Digital Design and BEng Computer Systems Engineering in the College of Engineering, Design and Physical Sciences, Brunel University London. He has a Ph.D. in Electrical and Electronic Engineering and BEng in Computer Systems Engineering. He is a recipient of many outstanding awards such as Prestigious VC's doctoral research prize, Best paper award of IEEE BTS 2013, Best poster prize 2009, Best final year project prize, Graham Hawkes prize, and the university Medal for overall excellent achievement. He has over 50 published white papers in international journals and conferences.

Abdul Hamid Sadka Professor Sadka is an internationally recognised figure in Visual Media Processing and Communications. He is currently the Director of the Institute for Digital Futures at Brunel University London. He is the former Head of the Department of Electronic and Computer Engineering at Brunel (Oct 06–Jul 12) and the Founding

Director of the Centre for Media Communications Research (CMCR). He and his team have made numerous innovations in video processing and his book "Compressed Video Communications" is regarded as a seminal textbook in this area. He frequently serves on influential advisory boards, evaluation panels and international committees, and provides consultancy services to international Law firms in the area of Video Compression and 2D/3D Visual Media Systems. He served as elected member on the Steering Board of New European Media (NEM) European technology platform (2011–14). He is widely supported by industry and has several entrepreneurial business engagements. He is a chartered engineer (CEng), Fellow of HEA, Fellow of IET and Fellow of BCS.