This is a pre-copyedited, author-produced version of an article accepted for publication in Lecture Notes on Data Engineering and Communications Technologies following peer review. The final authenticated version is available online at https://doi.org/10.1007/978-3-030-70665-4_120

Disparity Estimation From Holoscopic Elemental Images

Bodor Almatrouk, Hongying Meng, and Mohammad Rafiq Swash

Brunel University London, London, UB8 3PH, UK
{Bodor.Almatrouk, Hongying.Meng, Rafiq.Swash}@brunel.ac.uk

Abstract. Abstract: Holoscopic 3D imaging system, also known as integral imaging, is an innovative 3D imaging principle that overcomes key limitation of traditional 2D imaging issues such as depth, scalability and multi-perspective with its simplistic form of 3D data acquisition and visualisation which provides robust and spatial information of the real world 3D scene. In this paper, an innovative 3D map generation technique is proposed which produces accurate 3D map of a scene from a single holoscopic 3D image based on angular information preserved in its elemental images. 3D depth map is generated from the elemental images based on the semi-global block-based matching algorithm. A weighted least squares filter is utilised to minimise the noise in the resulting disparity image. The evaluation result outperformed state of the art 3D depth generation techniques and the proposed technique enlarges the industrial application of 3D imaging applications such as AR/VR, inspection, robotics, security and entertainment.

Keywords: holoscopic 3D image; integral image; elemental image; viewpoint image; depth map; disparity; SGBM; holoscopic 3D camera

1 Introduction

3D depth estimation from a holoscopic 3D (H3D) image is a promising technique that has gained an interest in the past few years due to the advantage of calculating depth using a single aperture camera. The concept of H3D was first introduced in 1908 by Gabriel Lippmann [1] which he called integral photography. In today's H3D cameras such as Lytro [2], a micro-lenses array is placed behind the lens and in front of the camera sensor to capture the colour and intensity of light rays from every direction, through every point in space, breaking down the entire image into individual rays of light. H3D technology has the advantage of providing motion parallax, binocular disparity and convergence. Some of the most compelling feature that H3D cameras afford is the limitless refocus of the scene due to its infinite depth of field [3] and the wide depth of field.

The micro-lens array is the core of this technique as it is utilised to achieve the effect of an entire array of cameras but with only a single lens and a single sensor design. As a result, each elemental image, which is the micro-image is formed

behind each micro-lens, preserving different information about the direction, colour, and intensity of light in the scene.

In contrary to typical approaches to estimate depth map using the extracted and up-sampled viewpoint images from the H3D camera, in this paper, we estimate the depth map using elemental images. The reasoning behind this is that viewpoint images are orthographic image, where elemental images are perspective images that contain true angular information, thus, giving more accurate depth information.

In H3D image, light rays first pass through the primary lens into what is called a relay lens, which is used to focus the light rays to each micro-lens in the micro-lens array. A single lenslet among the full micro-lens array captures the light coming from the relay lens, meaning that the exit pupil is being re-imaged to the image sensor, giving the full amount of angular information. Accordingly, each lens is capturing a perspective of the scene recorded from that array location. Each micro-lens has a number of pixels behind it, creating a 2D format image that is referred to as the micro-lens image or elemental image. The H3D image captured is a 2D image of an array of elemental images.

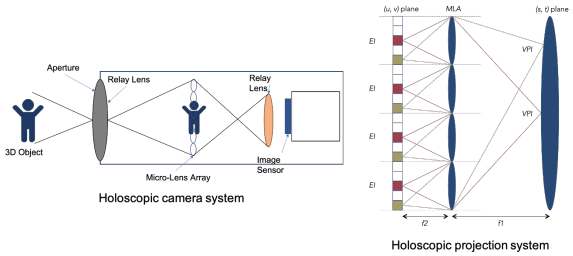

Fig. 1. H3D system projection

H3D plenoptic function can be represented by four dimensions $H3D(s,t,u,v)$. Each micro-lens corresponds to the position described in the image at *(s, t)*, where the orientation (u,v) resembles the image aperture, meaning that each pixel in the elemental image corresponds to a particular angle of light rays passing through the camera to a specific micro-lens. Consequently, H3D can be parameterised in two-plane schemes as seen in Fig. 1. The main lens can be considered as the first plane (s,t) and the micro-lens array as the second plane (u,v). The resolution of the micro-lens image is determined by the image sensor. Accordingly, the number of pixels of each elemental image is equal to the number of pixels of the entire H3D image divided by the number of elemental images. viewpoint images are a collection of images that can be extracted from a H3D image, each presenting a specific perspective.

2 Related Work

Many recent studies concerning H3D imaging system focus on depth estimation since it is one of the main features of a single aperture camera. Joen et al. [4] employed multiview viewpoint images, generated from a Lytro camera. They used traditional cost-volume-based stereo depth estimation approach [5], which assigns a hypothesised disparity to each pixel derived from a pixel-by-pixel matching. However, to overcome the narrow baseline problem, they have adjusted the cost-volume algorithm by incorporating sub-pixel displacement and Graph cuts algorithm has been utilised to refine the depth maps.

One of the most compelling features of the H3D cameras, is the limitless refocus of the scene due to its infinite depth of field. This feature has been used in [6] where depth is estimated from both defocus and correspondence depth cues. The depth algorithm was later refined using shading information to improve the shape estimation [7].

Occlusion-aware depth estimation algorithms were developed such as imposing the colour consistency in the viewpoint images in [8] and using multi-orientation epipolar plane images in [9].

Previous algorithms have achieved reasonably accurate depth maps. However, all previous works have focused on extracting depth maps from enhanced viewpoint images, generated mostly from Lytro camera. Here, we focus on obtaining depth maps using elemental images from the H3D images.

3 Methodology

A single H3D image contains both spatial and angular information of the scene. Therefore, the depth map of a scene can be estimated locally, i.e. from a single shot. The disparity can be generated from viewpoint images extracted from a calibrated H3D image by applying traditional stereo matching algorithms [10]. However, employing the current stereo matching algorithms directly to H3D's viewpoint images cannot offer a precise depth map due to the narrow baseline between them. In some approaches such as in [4], sub-pixel displacement algorithm is used. In our method, we have found that elemental images offer a more substantial baseline between two consecutive elemental images, particularly when the main lens is placed further away from the sensor.

3.1 Pre-processing

H3D image comprise of a large number of elemental images. To easily access and manipulate specific elemental images and viewpoint images, given a calibrated H3D image using [11], we produced an equation that loops through the entire H3D image pixels to generate a 4D parameterisation H3D function such as:

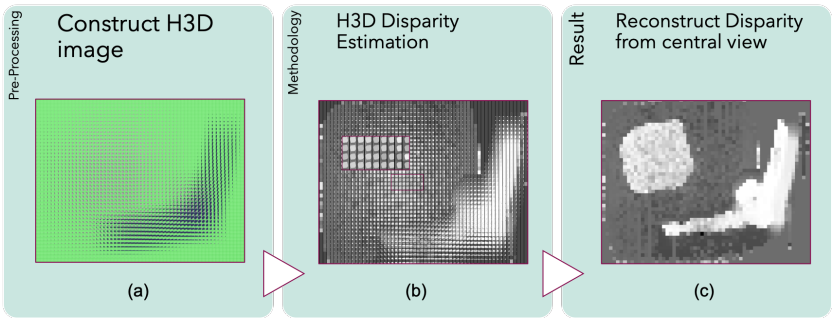

Fig. 2. The proposed method's framework and results for 3D disparity estimation froma H3D image

$$
\begin{aligned}
H3D &= (u, v, s_1 : s_n, t_1 : t_m) \\
s &= (u-1) * n + 1) \\
s_n &= (u * n) \\
t &= (v-1) * m + 1) \\
t_m &= (v * m)
\end{aligned} \tag{1}
$$

where u and v are the positions of elemental images and n and m are the number of pixels in each elemental image. An extra dimension has been added to the function, which consists of RGB colour channels generating a final 5D function.

Since elemental image are low-resolution images, detecting distinctive features is challenging. Thus, we used the up-sampling solution used in [12] to improve the quality of the image and detect unique features as seen in figure 3.

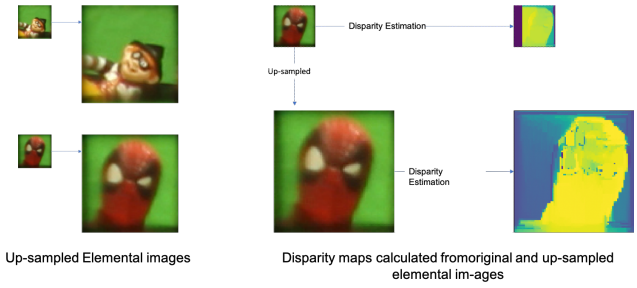

Fig. 3. Pre-processing H3D images

3.2 Disparity Estimation

For our stereo matching algorithm, we have used Semi-global Based Matching (SGBM) method [13]. SGBM has shown more significant results when used on images with shorter baseline. For each elemental image, block of pixels in the first elemental image are matched to their corresponding blocks of pixels in the next elemental image along the epipolar line, returning the corresponding vertical displacement, which is proportional to its depth. This problem can be expressed in a global cost function [14]:

$$E(D) = \sum_{d \in D} \left(C(d) + \sum_{d' \in N(d)} P_1 I\left[|d - d'| = 1\right] \; + \sum_{d'' \in N(d)} P_2 I\left[|d - d''| > 1\right] \right) \tag{2}$$

where I equals to 1 if its argument is true, and 0 if its argument false. $C(d)$ denotes the data term similarity metric of chosen disparity d, each similarity cost is stored in a 3D cost structure. This step is repeated for every block of pixels in the elemental images. The minimal costs in the stack of the 3D structure represent a potential disparity estimations. Finally, penalty terms P_1 and P_2, which rely on the difference to the neighbourhood disparities are introduced.

The produced minimal costs are not highly distinctive, which could result in incorrect disparity estimation. Cost aggregation within these 3D cost structures is applied to solve this matter. The final cost is derived by the accumulation of minimal costs along multiple paths in the image. In our case, we used 8 paths. Potential cost values are aggregated, and a weighted summation is performed on these cost possibilities. Accordingly, for each pixel, we analyse all its neighbouring pixels along the paths, the more significant the difference of the lateral parallax axis with the neighbouring pixels, the higher the penalty resulting in a high addition to the source value of the matching costs. This process is performed to force the strings in the path to be moderately continuous, which guarantees a smooth surface. We repeat this method on every path and every correspondence in the image to produce the final cost.

To minimise the noise in the resulting disparity image, we have applied weighted least squares (WLS) filter [15].

3.3 Disparity reconstruction

The resulting disparity image of the estimation process resembles the disparity maps of the multiple elemental images of the H3D image, as seen in figure 2 (b). We can extract multiple disparity maps from different viewpoints. Given a H3D disparity image coordinate, the first plane coordinates (s, t) can be used to represent the viewpoint disparity images and *(u, v)* representing the disparity elemental images. To extract the final depth, pixels from disparity elemental images are rearranged accordingly. From a particular (u, v) coordinate, the pixels sharing the same location of (s, t) from each elemental disparity image are

extracted and arranged according to each pixel's associated micro-lens position (u, v) to generate a single viewpoint disparity image. Thus, each elemental disparity image is considered as one pixel in the viewpoint disparity image such as in figure 4.

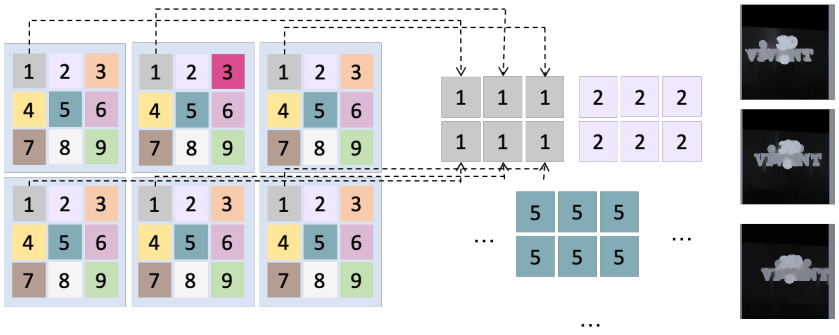

Fig. 4. Viewpoint disparity extraction based on re-arranging pixels, on the right different viewpoint disparity map can be viewed from different perspectives

4 Evaluation

The results of the proposed depth extraction algorithm were examined using synthetic H3D images created using POV-Ray software. Each image consists of around 97×53 elemental images; each is 100×100 pixels. Some of the elemental disparity images were not as accurate as of the surrounding elemental images. This is due to the fact that in our solution, we have employed fixed parameters throughout the whole elemental images. In the future, we wish to create a threshold to adjust the parameters according to the images. Our method was tested against Joen et al. [4] method, where we modified their algorithm to work with elemental images. Their algorithm has resulted in an incorrect disparity labelling as seen in 5 (a), This is mainly due to the fact that their algorithm uses the central images as a reference to compute disparity and since elemental images differ this approach is unreliable. However, we have used their algorithm on a slice of similar elemental images and we still did not get a robust result as seen in figure 5 (b). where our results provide a smoother, and more accurate disparity labelling. Even when we used viewpoint images with their method, we still got incorrect disparity due to the fact that viewpoint images are low in resolution.

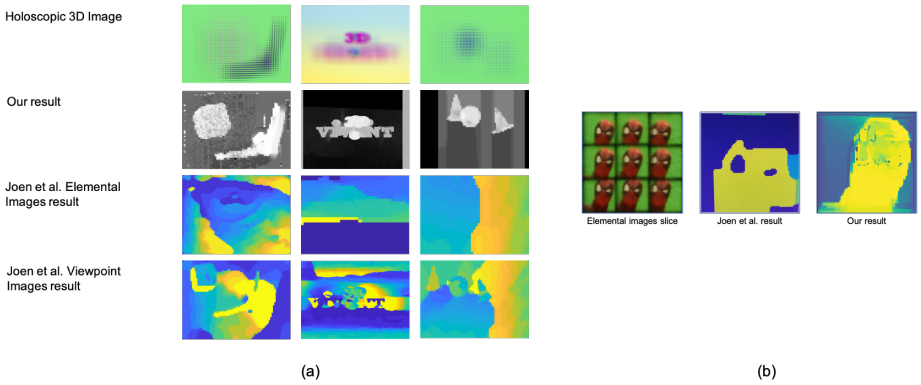

Fig. 5. Results against Joen et al. [4] algorithm. (a) fist row shows the original H3D image, second row shows our result from elemental images, third row is the result of [4] on elemental images and forth row is the result of [4] on viewpoint images. (b) result from a slice of similar elemental images

5 Conclusion

An innovative technique for 3D depth generation is proposed and evaluated by comparing to the state of the art techniques. 3D disparity maps are obtained from a single H3D image by constructing a 5D matrix based on holoscopic elemental images. The disparity map is measured from elemental images using a semi-global block-based matching algorithm and Weighted least squares Filter is applied to minimise the noise in the resulting disparity image. In contrary to typical approaches to estimate depth generation using the extracted and up-sampled viewpoint images from a H3D camera, we have estimated the depth map based the elemental images rather than viewpoint images as they contain true angular information, giving more accurate depth. The evaluation result demonstrated a reasonable good outcome compared to the state of the art techniques.

References

[1] Gabriel Lippmann. "Epreuves reversibles Photographies integrals". In: *Comptes-Rendus Academie des Sciences* 146 (1908), pp. 446–451.

[2] Todor Georgiev et al. "Lytro camera technology: theory, algorithms, performance analysis". In: *Multimedia Content and Mobile Devices*. Vol. 8667. International Society for Optics and Photonics. 2013, 86671J.

[3] Ren Ng et al. "Light field photography with a hand-held plenoptic camera". In: *Computer Science Technical Report CSTR* 2.11 (2005), pp. 1–11.

[4] Hae-Gon Jeon et al. "Accurate depth map estimation from a lenslet light field camera". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 1547–1555.

[5] Asmaa Hosni et al. "Fast cost-volume filtering for visual correspondence and beyond". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35.2 (2012), pp. 504–511.

[6] Michael W Tao et al. "Depth from combining defocus and correspondence using light-field cameras". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2013, pp. 673–680.

[7] Michael W Tao et al. "Depth from shading, defocus, and correspondence using light-field angular coherence". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015, pp. 1940–1948.

[8] Ting-Chun Wang, Alexei A Efros, and Ravi Ramamoorthi. "Occlusion-aware depth estimation using light-field cameras". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2015, pp. 3487–3495.

[9] Hao Sheng et al. "Occlusion-aware depth estimation for light field using multi-orientation EPIs". In: *Pattern Recognition* 74 (2018), pp. 587–599.

[10] Kuk-Jin Yoon and In So Kweon. "Adaptive support-weight approach for correspondence search". In: *IEEE transactions on pattern analysis & machine intelligence* 4 (2006), pp. 650–656.

[11] Donald G Dansereau, Oscar Pizarro, and Stefan B Williams. "Decoding, calibration and rectification for lenselet-based plenoptic cameras". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2013, pp. 1027–1034.

[12] Loıc Baboulaz. *Feature extraction for image super-resolution using finite rate of innovation principles*. Tech. rep. Imperial College London, 2008.

[13] Heiko Hirschmuller. "Stereo processing by semiglobal matching and mutual information". In: *IEEE Transactions on pattern analysis and machine intelligence* 30.2 (2007), pp. 328–341.

[14] Raden Arief Setyawan et al. "Implementation of Stereo Vision Semi-Global Block Matching Methods for Distance Measurement". In: *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)* 12.2 (2018), pp. 585–591.

[15] Wei Liu et al. "Semi-global weighted least squares in image filtering". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2017, pp. 5861–5869.