

3D-Interactive-Depth Generation and Object Segmentation from Holoscopic Image

E. Alazawi, A. Aggoun, J. Cosmas, M.R. Swash, M. Abbod and O. Abdulfatah

Abstract— Holoscopic 3D Imaging is a technique for producing natural 3D objects that exist in our world without the need for wearing specific eyewear. An Auto-Feature-Edge (AFE) descriptor algorithm is used to simplify the edge detection of objects that uses a Multi-Quantize Adaptive Local Histogram Analysis (MQALHA) algorithm. This paper presents an exploitation of available depth estimation and Feature Edge (FE) segmentation techniques when generating a 3D-Interactive-Map (3DIM). The robustness and efficiency of the proposed method is successfully illustrated in the paper and compared with the current state-of-the-art techniques.

Index Terms—3D Holoscopic system, Edge detection, Auto-thresholding, 3D depth, Local histogram analysis, Segmentation

I. INTRODUCTION

FULL natural colour effects, which are almost like the real world, can be obtained from Holoscopic 3D Imaging (HI), also known as Integral Imaging (II), which offers true-3D imaging technique [1]. This technique uses a micro-lens array to produce a planar intensity distribution of the true volume of the object scene [2].

Current 3D display technology requires viewers to wear head-gear devices, and recently much work was undertaken to develop multi-view autostereoscopic 3D display technology where the glasses will not be required. Unfortunately, at the moment the resolution is limited and the viewing is unnatural, which causes problems for the viewer during extended viewing [3, 4]. Due to improvements in microlens manufacturing technologies and the greater availability of processing power it is now possible to use the holoscopic technique for 3D display technology. The basic principle of 3D HI is that a lens array captures the spatio-angular distribution of the light rays [5, 6]. Many micro-lenses form the lens array and from these the spatio-angular distribution of the light rays can be identified to form an Elemental Image (EI). The planar detector surface acts to maintain a 2D distribution of intensities that are the HI. The capture of the scene is through micro-lenses where each has a slightly different view from which it is possible to determine the 3D depth through the use of a decoder [7]. The principle concept for capturing and displaying the image as illustrated in Fig. 1.

Manuscript received December 9, 2013. This work was supported by the European Commission under FP7-ICT-2009-4 (3DVIVANT).

Eman Alazawi (Eman.Alazawi@brunel.ac.uk) is a member of Electronic and Computer Engineering, Brunel University London, UB8 3PH, UK., Amar Aggoun (Amar.Aggoun@beds.ac.uk) is a head of Computer Science and Technology, University of Bedfordshire, LU1 3JU, UK, John Cosmas (John.Cosmas@brunel.ac.uk), Rafiq Swash (Mohammad.Rafiq@brunel.ac.uk), Maysam Abbod (Maysam.Abbod@brunel.ac.uk), and Obaid Abdulfatah (Obaidullah.Abdulfatah@brunel.ac.uk) are members of Electronic and Computer Engineering, Brunel University London, UB8 3PH, UK.

This manuscript details the use of an HI system with a feature-based technique to prepare the strong geometric 3D structures. This is done by performing correspondence matching with the EI array and a small set of regularly spaced pixels.

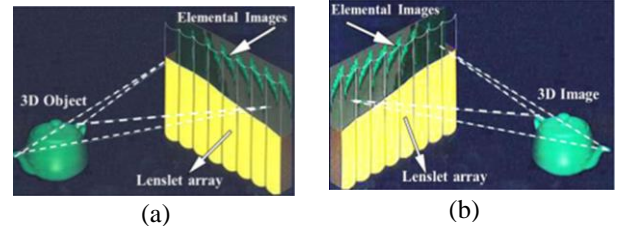


Fig.1 Unidirectional 3D-HI system using a cylindrical lens type lenticular sheet: a) capture of the 3D object process and b) display of the 3D image process.

II. CURRENT 3D OBJECT SEGMENTATION METHODS

The first work to identify depth extraction with HI systems used point spread functions [8]. In other work, for getting unidirectional 3D maps from HI, a practical feature-based matching approach was undertaken via disparity analysis and viewpoint image extraction [9]. From View-Point Images (VPI), the object's depth was determined via their displacements and a depth equation. There was a large amount of non-information and homogeneous regions produced from this. Another paper proposed a method to determine surface points from the intersections between pixel projection rays [10]. This method's output was a non-smooth and non-uniform sampling, which is a drawback. The same authors then used frameworks from [9] as well as graph cuts to estimate disparities on the surface of the extracted features between pairs of pixels. The authors were unsuccessful when trying to generate foreground masks directly from the HI system as it is hard to calculate depth on large non-informative and homogeneous regions [11].

The authors introduced an automatic masking procedure for use with depth map calculation results to ensure the removal of errors that were in the background to obtain better object depth by extending the work in [9] to get adequate object contour and noise reduction [12]. This masks the binary object onto the depth map to calculate where it is part of the background. This process is efficient and rather accurate when identifying areas for removal from the background.

Other work has enhanced the accuracy of the depth-through-disparity algorithm to overcome deficiencies in a local descriptor's special information by considering the depth and piecewise continuous space [13]. The method is better than the depth-through-disparity algorithm as it overcomes errors within the algorithm.

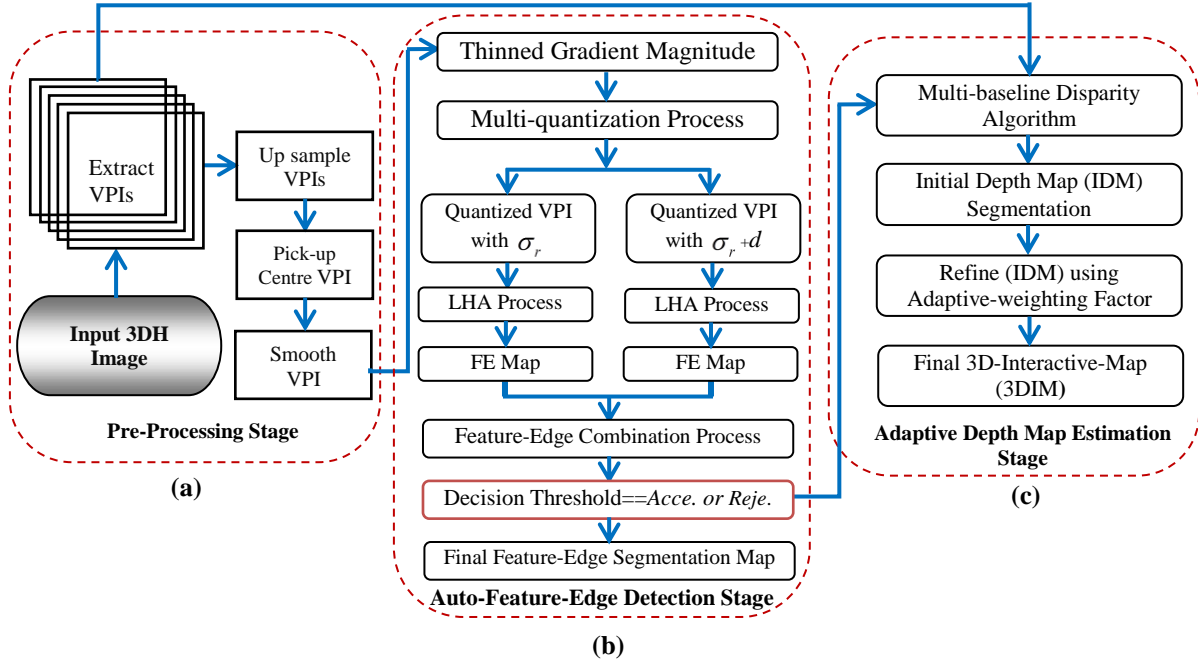


Fig. 2 Illustration of the three stages in the proposed methodology for the 3DHI technique, a) pre-processing process, b) auto-feature detection process and c) adaptive depth map estimation process.

A recent paper presented a new approach to estimate and extract from a 3D Omni-directional HI system a full parallax 3D depth map [14]. This approach is a great improvement on the current optimal 3D object estimation and segmentation techniques. A three-stage system is shown in Fig. 2 for the use of this approach that does not need human interaction and therefore does not suffer from human induced errors.

III. 3D OBJECT SEGMENTATION APPROACH

In the proposed approach, the efficiency is crucially dependent on the 3D object map segmentation and its explicit recovery of indirect inference of 3D objects and their robust estimation and connection. In this paper the 3DHI technique is used as the base for a system that extracts and segments the 3D object using depth estimation and segmentation techniques as explained in the following steps.

A. Pre-processing Stage

This stage resamples the collected data so it is as a 2D VPI. 2D images are recorded in which each image (an EI) has a slightly different view. These are then formed into a VPI where “viewpoint image segmentation” exploits the strong interval correlations between pixels displaced by one microlens. Previous work has detailed this process in detail [9, 13].

B. Auto-Feature-Edge Detection Stage

This stage requires a Modeling Auto-Feature-Edge (AFE) detector algorithm. In this, threshold detection uses both edge-based and region-based features to locate information in the 3D scene. The method presented here uses an automatic bi-level (one value) thresholding algorithm to sort images into regions based on the background. To ensure robust information about the feature’s edge is obtained, several techniques need to be

integrated to identify the Feature-Edge (FE) blocks. Fig. 2(b) shows the principle of the Multi-Quantize Local Histogram Analysis (MQLHA) algorithm as adopted in the AFE threshold algorithm.

The principle steps in the AFE detector algorithm (Fig. 3) are as follows:

1. Use a discrete Gaussian window=3 from the 2D Gaussian function to smooth the VPI to reduce noise: $G(x, y) = \frac{1}{2\pi\sigma^2} \exp[-(x^2 + y^2)/2\sigma^2]$, where, the parameter σ indicates the width of the Gaussian that defines the effective spread of the function.
2. Detection of the two boundary separating regions is via a gradient magnitude increase when there is a contrast between the FE on a smoothed VPI (\vec{x}, σ) and the background by computing the magnitude $m(\vec{x})$ and orientation $\theta(\vec{x})$ of the biggest alteration in intensity within a small region of each pixel [15] as the pixel differences:

$$m(\vec{x}) = \sqrt{[VPI(x+1, y) - VPI(x-1, y)]^2 + [VPI(x, y+1) - VPI(x, y-1)]^2}$$

$$\theta(\vec{x}) = \tan^{-1}[VPI(x, y+1) - VPI(x, y-1) / VPI(x+1, y) - VPI(x-1, y)]$$

The gradient of the smoothed VPI can therefore be written as:

$$VPI_{\sigma} = \nabla[G_{\sigma}(\vec{x}) * VPI(\vec{x})] = [\nabla G_{\sigma}(\vec{x}) * VPI(\vec{x})]$$

3. From this, only when the FE does not have a gradient magnitude that is the local maxima is the non-maxima suppression process performed. The gradient magnitude VPI_{σ} is set to zero apart from where it is the local maxima. Weak FE pixels are removed by using an average value from the surrounding 3x3 window.
4. This step uses the MQALHA algorithm to increase the FEs continuity, improve the speed of the multi-resolution aspect, and enable the use of the method with many noise levels (Fig. 3). The multi-resolution task is done by the MQALHA through the use of local histogram analysis from low and high

edge maps for two differently quantized thinned gradient magnitude VPIs. The algorithm is explained in the following two steps:

- a) Use the scheme described in [15] to automatically quantize the VPI into two 12-level intensities. Smooth the small peak fluctuations with a 1-D Gaussian filter with a standard deviation of 0.8. The first local maximum is in the grey level and is set as σ_r , and is used to estimate the thinned gradient magnitude VPI to give the first quantized VPI. Shift the σ_r value to the right by a distance (d) to produce the second VPI from the higher starting point. Where $d = 1/4 * f_{max}$ is the best distance for producing the second VPI from the higher starting point. The higher value ($\sigma_r + d$) was selected to match the Full-Width-at-Quarter-Maximum (FWQM) of the noise distribution peaks (Fig. 4 (a)). Depending on the noise level, the second quantize starting point as well as σ_r will both either increase or decrease.
- b) As in [16], divide the quantized VPIs into 4x4 non-overlapping blocks. A 1-D Gaussian kernel of width $W=3$ smooths the computed histogram of each block, and from this they are classified as either background or EF blocks. If the block is in the background, as determined from the histogram being unimodal, than all pixels are set to zero. Otherwise, the first valley in the differentiated smoothed histogram is the threshold for that block. Those pixels with quantized grey levels over the threshold are determined to be edge pixels and set to one [16]. On the VPI, with σ_r in the lower threshold FE map, save the threshold block, as with the VPI ($\sigma_r + d$) higher threshold FE map for further processing. For the “horseman” VPI the results are in Fig. 4 (b-e).

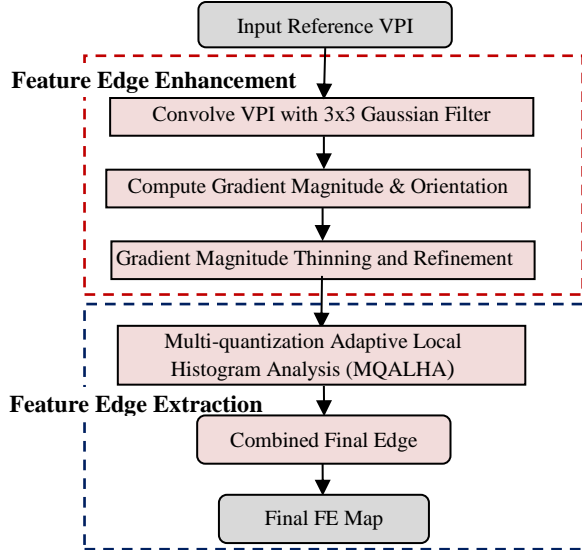


Fig. 3 Illustration of the steps of enhancement and extraction of the Feature Edge in AFE detector algorithm.

5. A “final edge map” is formed by combining the two EF maps. The higher threshold map is scanned to determine an edge segment and the lower threshold FE map is examined using a 3x3 local neighborhoods window to find possible connected successors. Then, when one of these successor pixels has an edge the algorithm will link it to the high

threshold FE map after extracting and linking it to the endpoint. This new pixel becomes the new endpoint.

6. The above process will only stop when all the connected edge points from the lower VPI have been determined. From this it is possible to transfer all the features from the higher VPI map to the FE map. The method to obtain the required information will vary from specific application to application as the task and object domain require.

C. Estimation of Adaptive Depth Map

This section details the stages in an approach to generate a 3DIM extracted from the VPI and disparity analysis using a previously presented method [13].

First: From stage B, interesting points (FE with local maxima more than the threshold) are extracted using the training center VPI to establish the description. Using sum of square difference (SSD), extract the disparity displacement (Dis) that is the distance between the pairs of pixels. For the initial disparity map estimation of the cost function $\forall Dis$ in x direction is:

$$C(bp, Dis) = SSD(d) = \sum_{x, y \in w} [\hat{I}_1(p) - \hat{I}_2(p + Dis)]^2 \quad (1)$$

where, $C(p, Dis)$ is the cost function, d is a horizontal displacement vector between pixel p of the window (block) w in the I_1 centre VPI and $(p + Dis)$ in the I_2 target VPI. Using all viewpoint image pairs and their different baselines, the initial matching decision is performed:

$$C(bp, Dis)_{initial} = SSSD(Dis) = \sum_{i=2}^P \sum_{p \in w} [\hat{I}_1(p) - \hat{I}_i(p + Dis)]^2 \quad (2)$$

where, SSSD is the sum of SSD that minimizes the depth map (D) and P is the pairs of the VPIs. The initial disparity for the central pixel $p(x_b, y_b)$ is determined and for all local pixels (bn) their scores are obtained from various values of Dis ($Dis=0, 1, \dots, Dis_{max}$). Then, so that pixel (p) has the minimal cost from all the computed disparity cost functions of bp , the value of \overline{Dis}_p^* is determined; where its minimum is given by:

$$\overline{Dis}_p^* = \arg\{\min_{d \in R} \{score(bp, \overline{Dis})\}\} \quad (3)$$

where, bp represents the window around pixel (p) and R is the research area.

Second: Determine if the disparity is dominant for the block centered on pixel p (bp), and if it is not, further refinement needs to be performed. However, some noise is still part of the disparity map so it is suboptimal. The method from [13] is used to stop many neighboring edges from being considered as interesting and to assist large untextured areas as well as to ensure the extrapolation of unreliable pixels from local high confidence pixels. There are three confidence terms form the adaptive weighting factor: $conf(w_p) = \exp(-\|I_{bp} - \overline{I}_{bn}\|/\sigma^2)$ representing the high variance of the center block bp and neighboring blocks bn , where, $I_{bp}, \overline{I}_{bn}$ represent the center pixel value bp and average value of the pixels within the block bn , respectively. With the use of a small ω it is possible to reduce the variance of each block. Use the spatial Euclidean distance from the coordinates for which the variance is high and close to the center block bp to calculate the distance term; $dist(w_p) = \sqrt{(i_{bp} - i_{bn})^2 + (j_{bp} - j_{bn})^2}$.

Use CIELab to calculate the colour space of a pixel bp ($C_{bp} = [L_{bp}, a_{bp}, b_{bp}]$) and pixel bn ($C_{bn} = [L_{bn}, a_{bn}, b_{bn}]$), donated as:

$$C_{diff}(w_p) = \sqrt{(L_{bp} - L_{bn})^2 + (a_{bp} - a_{bn})^2 + (b_{bp} - b_{bn})^2}.$$

The aggregated cost is computed as a *weighted* factor of the bp for different neighbouring blocks as:

$$W(bn, bp) = conf(w_p) \cdot dist(w_p) \cdot C_{diff}(w_p) \quad (4)$$

In the voting scheme, reliable neighbours gave each pixel p a vote: $Vote_p(Dis) = \{n | n \in U(p) \wedge Dis_n = Dis\}$, where, n is a consistent pixel from a reliable neighbour in the support aggregation window of the centre pixel p and the disparity Dis_n contributes one vote, which is accumulated in the set $Vote_p(Dis)$. The maximum majority weighted vote decides the final disparity of the bp as: $(Dis_p^* = arg\ msx | Vote_p(Dis) |)$. From Eqs. (2) and (4) with a bp block, the filtered cost match function $C(bp, Dis)_{out}$ becomes two terms: $C_{initial}$ is the visual cost match function of the center block bp and C_{smooth} indicates the smooth cost match function of the nearest neighbor block nb and is defined as:

$$C(bp, Dis)_{out} = C(bp, Dis)_{initial} + C_{smooth} \quad (5)$$

Rewrite Eqs. (2), (4) and (5) using a smoothing term as:

$$C(bp, Dis)_{out} = C(bp, Dis)_{initial} + \sum_{bn \in n(bp)} W(bn, bp) \times \min\{C(bn, Dis)\} \quad (6)$$

where, $\sum_{bn \in n(bp)} W(bn, bp) \times \min\{C(bn, Dis)\}$ is the smooth match function for the nb nearest neighbour block. This makes the disparity map more accurate.

Third: Derive the depth equation $D = \frac{Dis \cdot \psi \cdot F}{\Delta}$ from geometrical analysis of the recording process, where, D is the corresponding depth to be calculated, ψ is the pitch size of the microlens, F is the focal length of the microlens, within the two extracted VPIs the disparity of the object point is Dis and the sampling distance is Δ . Calculate the 3DIM from Eq. 5.

Fourth: Apply a median filter to further smooth the 3DIM. Calculate a median window size filter of 5x5 by systematizing all the neighboring pixels to a numerical order and substitute the middle ordered pixel for the considered pixel.

IV. DISCUSSION / RESULTS

The main contributions of this method are as follows:

- A.** This method's FE map is more distinctive and robust when dealing with camera VPI changes. It can provide robust local feature information that can provide the geometric constraints of the features. The real-world "Horseman" Fig. 4 shows the results from the AFE descriptor; while, Fig. 5 shows the results from the proposed algorithm for unidirectional real and synthetic 3DHs. The objects are correctly identified.
- B.** This approach also separates the objects from the background in the extracted VPI. The FE detection guides the region identification through the generation of a binary mask (see Fig. 6). Detected points are classified as either 1 or 0, and connected FEs are converted into connected segments (contour). The detected edge pixels are used to automatically extract the seeds required for growing a region through pixel aggregation of the interior pixel area of each object [13, 17]. This is beneficial as the segmentation is determined without

any prior knowledge and it does not require a homogeneity criterion threshold. The proposed method's results were compared with those from previous [13] work as shown in Fig 6 (a, b).

A comparison was made with state-of-the-art techniques [9-13]. This showed that the system presented in this paper was superior to the other tested systems as it used learnt parameters (given in Table 1). Table 2 shows this new system outperforming the comparison systems on the "Horseman" image. The proposed system is robust to camera VPI changes and it can provide information related to 3D depth estimation and 2D foreground object segmentation. However, there were some ambiguities that caused some problems. These were due to the thinning edge term (non-maximum suppression) that prevented multiple responses. Also, not many details are visible as the image resolution is low, due to illumination and focusing as well as other factors. The VPI resolution constrains the feature detection and extraction.

V. CONCLUSION

In this work, an automatic 3D-Interactive-Map generator was presented that was implemented in an Holoscopic 3D Imaging system. The proposed method is based on the edges as the main features and was developed so it could handle both real and synthetic HIs. The 3DIM was generated automatically from a combination of semi-local disparity map estimation and FE detection techniques. The basis of the success was the use of the feature edge in both directions, i.e. estimation of 3D map and object segmentation. The proposed method outperformed current state-of-the-art techniques, but the depth estimation still needs improving and extending with the Omni-directional 3DHI being viewed from more orthographic VPIs.

ACKNOWLEDGMENT

This work was supported by the 3D VIVANT (Live Immerse Video-Audio Interactive Multimedia) project funded by the EU. The authors acknowledge the support of the European Commission under the Seventh Framework Programme (FP7).

REFERENCES

- [1] J. Hong, Y. Kim, H.-J. Choi, J. Hahn, J.-H. Park, H. Kim, S.-W. Min, N. Chen, and B. Lee, "Three-Dimensional Display Technologies of Recent Interest: Principles, Status, and Issues," *Applied optics*, Vol. 50, No. 34, pp. H87-H115, Dec. 2011.
- [2] Y. Zhu, T. Zhen, "3D Multi-View Autostereoscopic Display and Its Key Technologies," *Proc. APCIP Conference on Image Processing*, Vol. 2, Shenzhen, China, pp. 31-35, 2009.
- [3] L. Onural, "Television in 3-D: What are the Prospects," *Proc. IEEE*, Vol. 95, No. 6, 2007.
- [4] G. Lawton, "3D Displays Without Glasses: Coming to a Screen Near You Computer," *Computer*, Vol. 44, No. 1, 2011, pp. 17-19.
- [5] J.-H. Park, K. Hong, B. Lee, "Recent Progresses in Three-Dimensional Information Processing Based on Integral Imaging," *Appl. Opt.* Vol. 48, No. 34, pp. H77-H94, 2009.
- [6] J.-H. Park, G. Baasantseren, N. Kim, G. Park, J.-M. Kang, B. Lee, "View Image Generation in Perspective and Orthographic Projection Geometry Based on Integral Imaging," *Opt. Express*, Vol. 16, No. 12, pp. 8800-8813, 2008.
- [7] A. Aggoun, "3D Holoscopic Imaging Technology for Real-Time Volume Processing and Display," *High-Quality Visual Experience Signals and Communication Technology*, IV, pp. 411-428, 2010.

- [8] S. Manolache, S.Y. Kung, M. McCormick, A. Aggoun, “3D-Object Space Reconstruction From Planar Recorded Data of 3D-Integral Images,” *Journal of VLSI Signal Processing Systems*, vol.5, pp. 5-18, 2003.
- [9] Wu, C., et. al., “Depth Map from Unidirectional Integral Images using a Hybrid Disparity Analysis Algorithm,” *IEEE Journal of Display Technology*, Vol. 4, No. 1, pp. 101-108, 2008.
- [10] D. Zarpalas, et. al., “Depth Estimation in Integral Images by Anchoring Optimization Technique,” *IEEE International Conference on Multimedia & Expo (ICME)*, 2011.
- [11] D. Zarpalas et. al., “Anchoring-Graph-Cuts Towards Accurate Depth Estimation in Integral Images,” *IEEE J. Display Technology*, Vol. 8, No. 7, pp. 405-417, 2012.
- [12] O. Fatah, A. Aggoun, M. Nawaz, J. Cosmas, E. Tsekles, M. Swash, E. Alazawi, “Depth Mapping of Integral Images Using Hybrid Disparity Analysis Algorithm,” *IEEE International Symposium Broadband Multimedia Systems Broadcasting*, pp. 1-4, South Korea, 2012.
- [13] E. Alazawi, A. Aggoun, O. Fatah, M. Abbod, R. M. Swash, “Adaptive Depth Map Estimation from 3D Integral Image,” *IEEE International Symposium Broadband Multimedia Systems Broadcasting*, pp. 1-6, London, UK, 2013. Best Student Paper Award.
- [14] E. Alazawi, A. Aggoun, M. Abbod, R. M. Swash, O. Fatah, “Scene Depth Extraction from Holographic Imaging Technology,” *IEEE 3DTV-CON: Vision Beyond Depth*, AECC, Aberdeen, Scotland, 7-8 October 2013.
- [15] A. Aggoun and M Khallil, “Multi-Resolution Local Histogram Analysis for Edge Detection,” *IEEE, ICIP*, pp. 45-48, 2007.
- [16] M Khallil and A. Aggoun, “Edge Detection Using Adaptive Local Histogram Analysis,” *IEEE, ICASSP*, pp. 757-760, 2006.
- [17] M. M. Abdelsames, “An Automatic Seeded Region Growing for 2D Biomedical Image Segmentation,” *International Conference on Environment and BioScience*, IACSIT Press, pp.1-5, Singapore, 2011.
- [18] http://ftp.itl.gr/pub/Holoscopy/Integral_Database/Database/Data/.

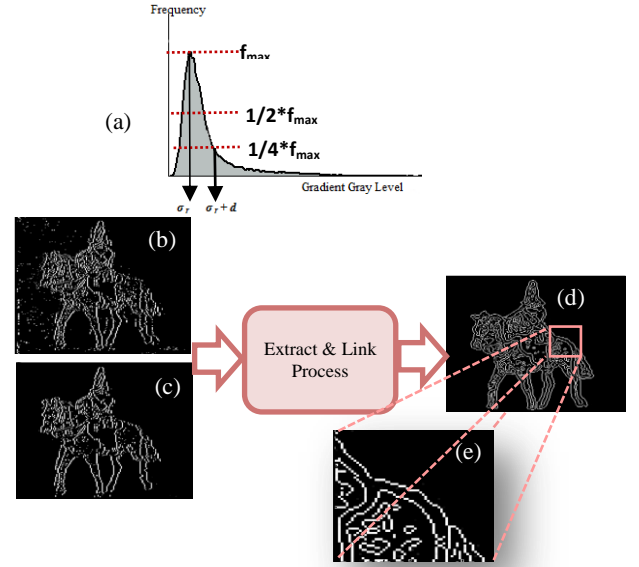


Fig. 4 AFE detection stage, a) first starting point (σ_r) (low threshold) setting strong edges and second starting point (σ_r+d) (high threshold) setting weak edges, b) low thresholding result on “Horseman” real-world VPI (noisy map), c) high thresholding result on “Horseman” (without noise) d) applying extraction and link process to set final EF map, and e) magnified section of a final FE map.

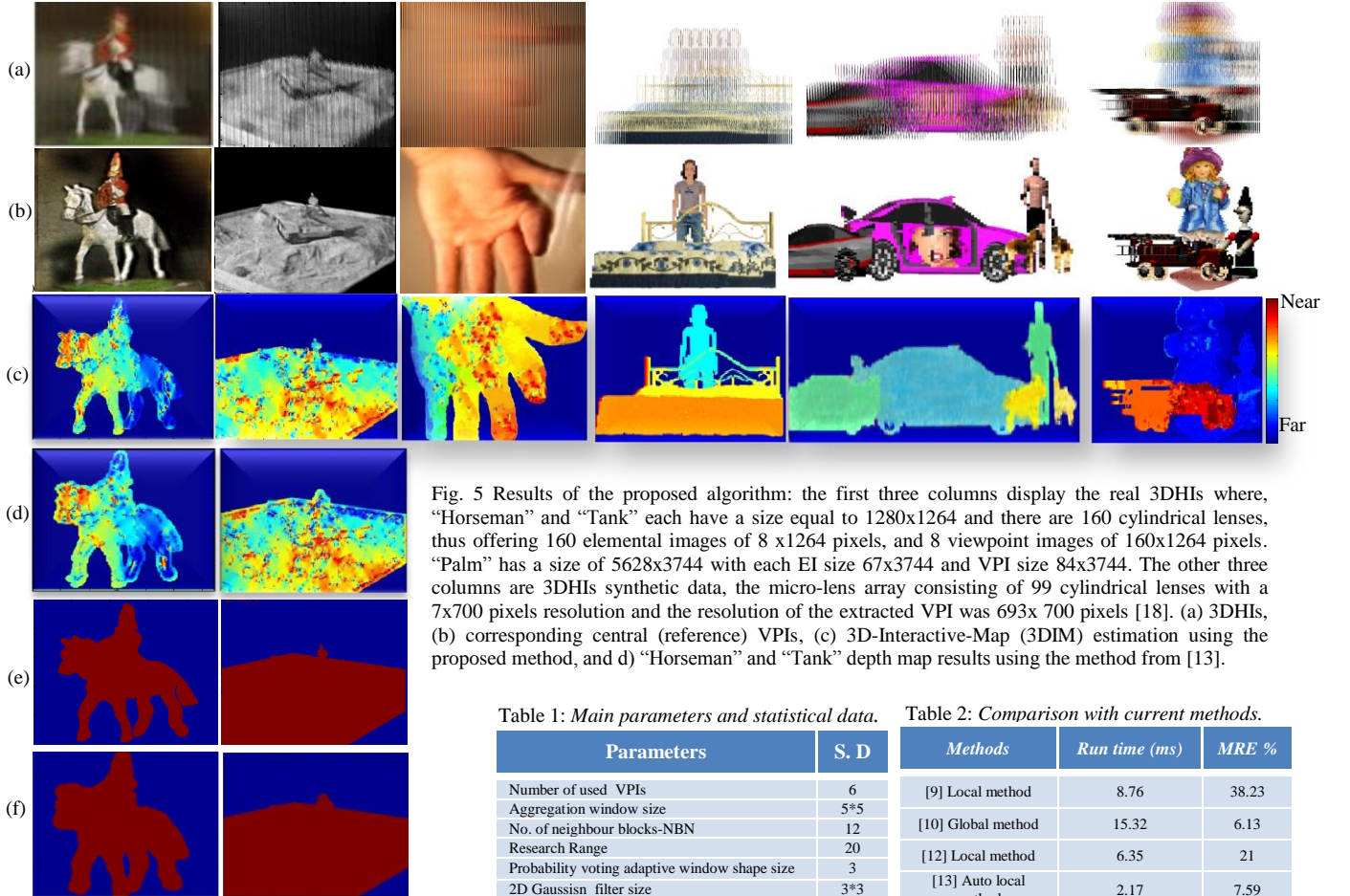


Fig. 5 Results of the proposed algorithm: the first three columns display the real 3DHIs where, “Horseman” and “Tank” each have a size equal to 1280x1264 and there are 160 cylindrical lenses, thus offering 160 elemental images of 8 x1264 pixels, and 8 viewpoint images of 160x1264 pixels. “Palm” has a size of 5628x3744 with each EI size 67x3744 and VPI size 84x3744. The other three columns are 3DHIs synthetic data, the micro-lens array consisting of 99 cylindrical lenses with a 7x700 pixels resolution and the resolution of the extracted VPI was 693x 700 pixels [18]. (a) 3DHIs, (b) corresponding central (reference) VPIs, (c) 3D-Interactive-Map (3DIM) estimation using the proposed method, and d) “Horseman” and “Tank” depth map results using the method from [13].

Table 1: Main parameters and statistical data.

Parameters	S. D
Number of used VPIs	6
Aggregation window size	5*5
No. of neighbour blocks-NBN	12
Research Range	20
Probability voting adaptive window shape size	3
2D Gaussisn filter size	3*3
Colour threshold pixel neighbor	2
Gaussian standard deviation starting σ_r	0.5
Similarity measurement method	SSD

Table 2: Comparison with current methods.

Methods	Run time (ms)	MRE %
[9] Local method	8.76	38.23
[10] Global method	15.32	6.13
[12] Local method	6.35	21
[13] Auto local method	2.17	7.59
Proposed approach / auto local method	1.68	7.73

Fig. 6 Foreground masks obtained from the real holographic images “Horseman” and “Tank” using, e) proposed algorithm and f) previous algorithm [13].