❐   2394

Computer vision based 3D reconstruction : A review

Hanry Ham, Julian Wesley, Hendra
Computer Science Department, School of Computer Science, Bina Nusantara University, Indonesia

Article Info	ABSTRACT
	3D reconstruction are used in many fields starts from the object reconstruction such as site, cultural artifacts in both ground and under the sea levels, medical imaging data, nuclear substantional. The scientist are beneficial for these task in order to learn, keep and better visual enhancement into 3D data. In this paper we differentiate the algorithm used depends on the input image: single still image, RGB-Depth image, multiperspective of 2D images, and video sequences. The prior works also explained how the 3D reconstruction perform in many fields and using various algorithms.

Corresponding Author:

Hanry Ham,
Computer Science Department,
School of Computer Science, Bina Nusantara University,
Jakarta, 11480 - Indonesia.
Email: hanry.ham@binus.edu

1.   INTRODUCTION

3D Reconstruction task is one of the interesting task that meet its maturity already. These can be seen from the commercial products such as product from Agisoft and Pix4D that are capable of produced high quality of large scale 3D models. Furthermore, the hardware such as the computer vision has been developed and improve since then. There are some setup camera introduced in the research such as stereo camera and Kinect.

In addition to the vision setup, kinect camera shows a great positive feedback from the researchers, proved by common vision setup that can be found in the literature review. Not only that, stereo camera setup can be found among the literature review. In addition to the stereo camera, custom stereo camera are quite popular among the researchers by combining two equals web camera that positioned by period of distance. The algorithm to perform 3D reconstruction between these camera are different due to the produced images are different as well. Kinect abilities allows RGB image and depth map produced, on the other hand Stereo camera has to perform another depth map acquisition algorithm by combining 2 RGB images.

Numerous numbers of 3D reconstruction task can be found in capturing the site, cultural artifacts both in ground and under the sea levels [1]. The extinction factor is the most prominent issue in these area. Moreover, 3D imaging data also could help improve the accuracy of the anatomical features in order to observe some areas before coming to the surgery action .Furthermore, in order to perform 3D reconstruction, there are multiple approaches found in the literature review such as from the broad ranges of vision setup, various types of inputted image to construct 3D reconstruction. Thus, In this paper will describe more on those approaches.

The great numbers of the researchers along with the hardware supports allows such algorithm to do high processing calculation in order to perform reconstruction task. There are some sections mentioned in part 2.. The benefits of reconstruction are to perform 3D recording, visualization, representation and reconstruction [2]. Moreover Tsiafaki and Michailidou explained that, there are 6 benefits in performing reconstruction and visualization: limiting the destructive nature of excavating, placing excavation data into the bigger picture, limiting fragmentation of archaeological remains, classifying archaeological finds, limiting subjectivity and publication delays, enriching and extending archaeological research.

Some algorithms found in the literature review introduced the usage of single and multiple images approaches to perform 3D reconstruction. There are some characteristics of the algorithms in the literature specifically built for single or multiple images, advantages and drawbacks explained in this paper.

In this paper will described the vision setup by 3 categories as follows:

1. Single Camera

    A single camera is simple to calibrate, computationally efficient more compact. However, they are lack of the depth information. It requires prior knowledge from other sensor to determine the depth scale [3].

2. Stereo Camera

    In stereo camera mechanism is that the images captured either using 2 equals web camera [4] or any cameras. They are set by a defined distance. In addition to 2 images captured, an algorithm is used to generate depth map. However, stereo matching have several issue when the scene contains weekly textured areas, repetitive patterns or occlusions occur in both indoor and outdoor environments [5] as shown in Figure 1.
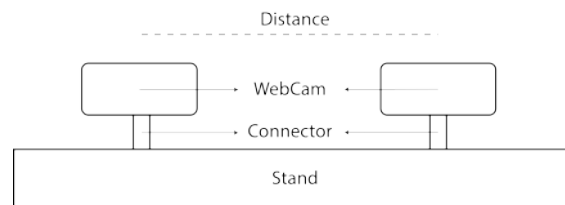

Figure 1. Stereo Camera

3. Kinect / Structured Light / Time of Flight

    Structured Light sensor is able to perform range detection, an accurate distance measurement is the output [6]. Kinect camera is a product from Microsoft that has an RGBD camera. The product comes with native SDK that allows user to call the API to perform some vision task such as skeleton detection.

4. Fusion

    Some researchers also tried possibilities of using fusion approach where as combining depth map produced by Stereo and kinect camera to achieve higher accuracy in depth map precision. To such development allows to produce better 3D Reconstruction object, rich in features details. Range cameras are low cost and ease to use to construct 3D point clouds in real time. One issue arise is that the transparent and reflective surfaces [7]. on the other hand, 3D model produced by stereo vision are mostly incomplete in low texture regions. The possibilities of combining both approached could lead to better depth map quality. Fusion approach is shown in Figure 2.
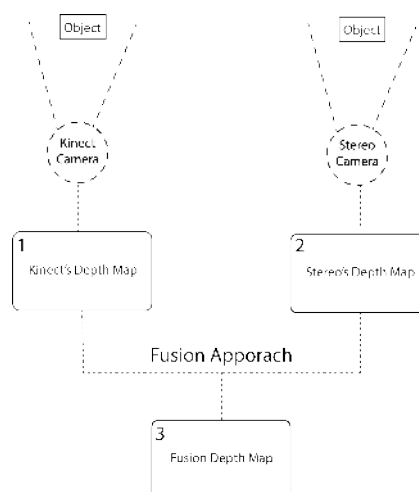

Figure 2. Fusion Approach

The algorithms vary due to the characteristics of the inputted image. Therefore, in this paper we described the inputted image into 2 categories : single and multiple images. Single image, the characteristic image can be described as:

1. Single Still Image
   Single still image here using an RGB image. This image can be taken by a regular camera.

2. RGB-Depth Image
   RGB image is taken with the setup camera that produced RGB-D format image. Mostly, the setup used is commercial camera such as Kinect, Intel real sense camera.

On the other hand, the multiple images can be described as:

1. Multiperspective of 2D images [8]
   The idea of this aprroach is to take some images differentiate in its perspective to the object. Thus the area of the object are covered properly using filter [9]. In addition to that, Xian-hua and Yuan-qing [10] said that in order to perform 3D reconstruction, an effective matching of a feature is the prominent factor in later stage. They implemented a feature matching error elimination method based on collision detection.

2. Video Sequences
   The using of the input video sequences as known as structure from motion. Sepehrinour and Kasaei explained that these methods are using the shared information of consecutive frames, in the form of tracking information of feature points in a sequence of images. The factors may impact to the developed methods: the knowledge or lack of knowledge of camera calibration parameters, having multiple cameras with different viewing angles or only one moving camera, and rigid or non-rigid shape reconstruction based on the incoming video stream.

2. TAXONOMY OF 3D RECONSTRUCTION

3D Reconstruction plays an important roles in several aspects such as medical imaging data, site and cultural artifact reconstruction.

(a) Medical Imaging Data
   Common surgery operation procedures uses X-Ray as a reference for the doctor to operate on specific section. However, some important features cannot be visualized well in 2D images [12]. In addition to 2D images, the accuracy may increase depends on several aspects such as: number of 2D Views, the image noise, and the image distortion. Magnetic resonance Images also holds an important method while considering the operation process. The given output of MRI are in 2D images, however there are some literature can be found in manipulating those images into 3D space. By implementing such method, they would like to prove the more features captures, the more accurate result is. a work from Hichem et al. introduced a geometric interpretation of the 3D model reconstruction of the blood vessel of the human retina. Sumijan et al. [14] in their work introduced a method to calculate volume Hemorrhage Brain on CT-Scan Image and 3D Reconstruction. The idea of this work is to calculate of the bleeding area in the brain on each image slide CT-scan. As it is said in the previous work[15], brain injury is one of the most causes that cause the death of human. In addition to the pipeline, the extraction the bleeding area of the brain using Otsu algorithm combining with the morphological features algorithm. Therefore by visualizing the brain volume aim at improving visual enhancement for the doctor to give the best medical treatment.

(b) Site and cultural artifacts Reconstruction
   The site reconstruction has been widely an issue to the archaeology in order to capture the social, culture through the building, they do the reconstruction. Regular camera can only allow to capture in 2D space format. Not all the details from the building can be captured and closely observed. Since then, by using stereo camera or Kinect make this task possible along with the algorithm developed in the current research. The archaeological sites are not only on the ground but also under the sea. The reconstruction which performed under the sea rises another issue to the images captured such as degradation quality

if underwater images, uneven illumination of light on the surface of objects. scattering and absorption effects [1].

(c) Nuclear Substantial Reconstruction

Monterial et al. [16] used 3D image reconstruction of neutron sources that emit correlated gammas. This aim at preventing nuclear threat search, safeguards and non-proliferation. This research is prominent and under supervision of legal division. In addition to that, nuclear had been used as source of energy, yet some controversies arise about the impact of harmful substantial.

2.1. Single still image approach

The first part will describe the algorithms found in the literature review using single still image. Compared to the multiple images, single image occurs tend to have more challenges. Saxena et al. explained that one of the issued is to create a depth map due to the local features are insufficient to estimate depth at a point. In addition, single still image approach is relatively less studied in the literature.

Saxena et al. [17] introduced a 3D depth reconstruction using a single still image. A supervised learning approach was proceeded by taking a training set including the unstructured indoor and outdoor environments and their corresponding ground-truth depthmaps. Their proposed algorithms aware of the global structure of the image, based on modeling depths and relationships between depths using proposed multiple spatial scales using a hierarchical, multiscale Markov Random Field. Ground truth were taken using 3D scanner.

Yan et al. [8] proposed a system called Perspective Transformer nets. The model was built by ignoring the color and texture factors. In addition to that, the experiments shows that excellent performance of the proposed model in reconstructing the object without ground-truth 3D volume as supervision. The input used were provided by Chang et al. [18] works. The images input proposed is a single view 3D volume reconstruction [19] with perspective transformation [20] run through defined encoder-decoder network that consists of a 2D convolutional encoder, a 3D up-convolutional decoder and a perspective transformer networks.

Fan et al. [21] applied a region-based growing algorithm for 3D reconstruction by using brain MRI images. There are 3 steps in their proposed pipeline : First, the seed element is the initial state of the segmentation. Second, start the growing process from the seed element. There are 4 areas of the growth area. However there are some defined threshold value to meet the pattern of growth. Third, use the points which satisfy the growing requirement as seed element, and continue to grow. in addition to the result, their proposed method could achieve 90.52% compared to Nadu [22] works.

2.2. RGB-depth image approach

Zhang et al. [23] developed a feature-based RGBD camera pose optimization for real-time 3D reconstruction. Their proposed work are ignoring corner-based feature detectors such as BRIEF and FAST due to acquired images contains huge noise around object contours. Subsequently, SURF detector was chosen due to the fact that its robustness, stability, scaleable and rotation invariant [24]. In addition to that, SURF can be computed in parallel on the GPU [25]. The miss-matched pairs in feature matching can be removed using RANSAC algorithm. The consistency of the global positions of matched features are tracked by proposed feature correspondence list and camera pose optimization both in the spatial and temporal dimension. Subsequently, in order to evaluate the method, voxel-hashing was used for each camera poses compared to the proposed method. It is proved that their proposed optimized camera poses outperforms the structure of the reconstruct model for the real scene data captured by a fast moving camera.

Group et al. [26] explained that a fully convolutional 3D denoising autoencoder neural network. They experimented using RGBD dataset and it is proved that the network could reconstruct a full scene from a single depth image by filling holes and hidden element. The network is capable of learn the object shape by inferring similarities in geometry. A real-word dataset of table top scenes [27] was used using KinectFusion. Their steps can be mentioned as follows: acquisition RGBD image using Kinect, denoising and hole filling depth channel using [28] algorithm, projection of the pixel into 3D space using preset equations, retrieve sensor pose from accelerometer and align point cloud data, voxelize the point cloud, and A predefined CNN layer was trained. In addition to that, the network is not constrained to a fixed 3D shape and it is capable successfully reconstructing arbitrary scenes.

Jaiswal et al. [29] used Kinect to assess 3D object modelling. The proposed pipeline are as follows: first, 3D point cloud, a green surface was placed behind and under the object to do the histogram-based seg-

mentation out the object from the RGB images. Afterwards, RANSAC algorithm is used to perform a coarse alignment. Second, the registration using SIFT based [30] to overcome the lack structural features or undergo significant changes in camera view. Third, global alignment is used ti eliminate inaccuracy at each registration that could lead to significant misalignment between the first and last frame. Fourth, 3D point cloud denoising is performed to refine the 3D object model, in this case Moving Least Square (MLS) 3D model denoising method [31]. Fifth, surface reconstruction using Delaunay triangulation method [32] to convert 3D point clouds into meshed. Afterwards, coloring task is performed to each vertex and simply interpolate the color in each triangle faces.

2.3. Multiperspective of 2D images approach

Kowalski et al. [33] created an open source system for live, 3D data acquisition using multiple kinect v2 Sensors. To overcome the ability of the native Kinect V2 SDK, they made this flexible framework. There are 3 coordinates system of a markers: Kinect v2 sensor, coordinate system of a marker which is located at a center on a given marker and the world coordinate. The proposed pipeline as follows: first calibrations were done by calibrating 2 types of defined markers. Subsequently the Iterative Closest Points (ICP) algorithm [34] was used to refine the initial estimation.

Evangelidis et al. [5] combined low-resolution depth data with high resolution stereo data to overcome the construction of high-resolution depth maps for the range-stereo fusion problem. The input used stereo images (high resolution) and depth data (low resolution) from the range camera. The low resolution depth data are projected into the color data and refined a high resolution sparse disparity map. Subsequently, the depth up-sampling algorithms were perform such as triangulation-based interpolation and join bilateral filter. then a region growing fusion were performed and final denser High resolution map as the result.

Burns [35] introduced a texture super resolution (TSR) method for 3D multi-view reconstruction. In addition, their work used video sequence as the input. Moreover to the proposed pipeline, a Photoscan from Agisoft is used to do multi-view stereo reconstruction and 3D mesh model. Then, optical flow algorithm is integrated in order to register each pixel of neighboring to the closest key-frame using KLT feature tracker [36]. Afterwards, to support robustness to outliers the fundamental matrix filtering of the tracked 2D points and RANSAC filtering of the 2D/3D correspondences. Due to the piece-wise affine surface approximation constructed in 3D mesh, this may lead to pixels registration error. To overcome that issue, to locate the displacements, an optical flow estimation is used [37]. The object used is 2mx1m desk that has many textured objects on it as gray-scale images along with the subsampling applied to it. It is acquired using a camera with 5.5mm focal length at f/2.8 mounted on a Byaer 1/18" e2v detector. There are 3 experiments conducted and it shows that the proposed methods outperforms compared to the registration with mesh and camera poses only, registration with optical flow only.

Tulsiani et al. [38] studied multi-view supervision for single-view reconstruction and a differentiable ray consistency (DRC) term was introduced which allows computing gradients of the 3D shape given an observation from an arbitraty view. The dataset used is called ShapeNet dataset. The following steps to perform their methods are: *formulation*, view consistency loss function is introduced aim at measuring the inconsistency between a predicted 3D share and a corresponding observation image.*shape representation*, The assumption made was it is possible to trace trays accross the voxel grid and compute intersection with cell boundaries. The 3D shape representation is parametrized in a discretized 3D voxel grid. Observation, This aim at achieving the shape to be consistent with some available observation such as depth image, object foreground mask. Also CNN model was used as a simple encoder-decoder which predicts occupancies in a voxel grid from the input RGB image. The result outperformed all the algorithms found in the literature review.

Martin-Brualla et al. [39] extended 3D time-lapse reconstruction where a virtual camera moves continuously in time and space using internet photos. Previous work assumed a static camera, the addition of camera motion during the time-lapse produces a very compelling impression of parallax. The first step is a pre-processing step, computing 3D pose of the inputted image using structure from motion algorithm. Subsequently, the desired path has to be specified through the reconstructed scene. Then, the algorithm compute time-varying, temporally consistent depthmaps for all output frames in the sequences. Proposed 3D time-lapse reconstruction computes time varying, regularized color profiles for 3D tracks in the scene. output video frames are reconstructed from the projected color profiles.

2.4. Video sequences

Sepehrinour and Kasaei [11] introduced a novel algorithm for perspective projection reconstruction using single view videos of non-rigid surfaces. The system input is a single view video that taken in a totally natural environment. In addition to that, the features extracted: projective depth coefficients of all points in each of the input frames, projection matrix components (camera calibration, rotation matrix, and transmission vector).

Xu et al. [40] developed underwater 3D object reconstruction with multiple views in video stream via structure from motion (SFM). They are trying to capture the inherent geometrical variation of 3D objects at multiple visual angles using a myring streamline AUV system with CCD camera with resolution of 480 TVL/PH and the minimum scene illumination 0.28 lux on board. The proposed pipeline : continuous videos stream combining SFM with object tracking strategies. An object tracking so called particle filter has been introduced in image sequence with multiple views to focus on the motion trajectories of underwater 3D objects all the time. a process of triangulation, iterative process, and other parameter adjustment is set for SFM algorithm to recover and estimate the position of the camera calibration and the geometry of underwater scene with sparse 3D point cloud.

Lapandic et al. [41] introduced a framework for automated reconstruction of 3D model from multiple 2D Aerial images using Unmanned Aerial Vehicle (UAV). The objective of this work is to achieve near real-time performance with reliable accuracy and execution time. The proposed pipeline as follows: feature detection and extraction using FAST algorithm and Lucas-Kanade method respectively, 2D point correspondence, point cloud filtering, camera pose estimation, points triangulation and point cloud calculation.

3.   DISCUSSION

The oldest paper cited in this paper is 1981 and the research about 3D reconstruction is still going on. This proved that the maturity of the research in this area is achieved. There are numerous algorithms is described in solving numerous of problems. In addition, the commercial software such as Microsoft, Agisoft, intel real sense, asus and many others companies develop software and hardware to perform such calculation. The general pipelines found in the literature reviews are: first, image acquisition. There are some datasets available that can be used in order to evaluate the performance of the proposed algorithms. Moreover, chances to create own object using vision setup mentioned earlier in section **??**. Second, Pre-processing step by allowing some filters applied to get the best images to construct. Third, 3D cloud points. The alignment algorithm plays an important role to get decent accuracy. Along with the refinement method in mismatched 3D cloud registration. Fourth, 3D reconstruction is where the texturing and meshed are applied as the final result.

4.   CONCLUSION

In this paper explains several current 3D reconstruction methods from literature review. There are various algorithm in order to perform each step of general algorithm of 3D reconstruction. Each object constructed required special algorithms depends on the vision setup, the texture and size of the observed object. The improvement of the sensor could lead to the higher accuracy of creating 3D reconstruction in the future besides the efficient algorithms. Modeling using neural network shows a great advantages [26], [8]. The defined network will try to learn the shapes and will fill the occlusion region automatically.

REFERENCES

[1]   A. Anwer, S. S. A. Ali, and F. Meriaudeau, "Underwater online 3D mapping and scene reconstruction using low cost kinect RGB-D sensor," *2016 6th International Conference on Intelligent and Advanced Systems (ICIAS)*, pp. 1–6, 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7824132/

[2]   D. Tsiafaki and N. Michailidou, "Benefits and Problems Through the Application of 3D Technologies in Archaeology: Recording, Visualisation, Representation and Reconstruction," *SCIENTIFIC CULTURE Tsiafaki & Michailidou SCIENTIFIC CULTURE*, vol. 1, no. 3, pp. 37–45, 2015.

[3] F. Santoso, M. Garratt, M. Pickering, and M. Asikuzzaman, "3D-Mapping for Visualisation of Rigid Structures: A Review and Comparative Study," *IEEE Sensors Journal*, vol. PP, no. 99, pp. 1–1, 2015. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7322186

[4] A. Harjoko, R. M. Hujja, and L. Awaludin, "Low-cost 3D surface reconstruction using Stereo camera for small object," *2017 International Conference on Signals and Systems (ICSigSys)*, pp. 285–289, 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7967057/

[5] G. D. Evangelidis, M. Hansard, and R. Horaud, "Fusion of Range and Stereo Data for High-Resolution Scene-Modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 11, pp. 2178–2192, 2015.

[6] G.-v. J. M and M.-v. J. C, "Simple and low cost scanner 3D system based on a Time-of-Flight ranging sensor," pp. 3–7, 2017.

[7] R. Ravanelli, A. Nascetti, and M. Crespi, "Kinect V2 and Rgb Stereo Cameras Integration for Depth Map Enhancement," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLI-B5, no. July, pp. 699–702, 2016. [Online]. Available: http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLI-B5/699/2016/isprs-archives-XLI-B5-699-2016.pdf

[8] X. Yan, J. Yang, E. Yumer, Y. Guo, and H. Lee, "Perspective Transformer Nets: Learning Single-View 3D Object Reconstruction without 3D Supervision."

[9] Q. Hao, R. Cai, Z. Li, L. Zhang, Y. Pang, F. Wu, and Y. Rui, "Efficient 2D-to-3D correspondence filtering for scalable 3D object recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, no. 1, pp. 899–906, 2013.

[10] J. Xian-hua and Z. Yuan-qing, "Error Elimination Algorithm in 3D Image Reconstruction," vol. 12, no. 4, pp. 2690–2696, 2014.

[11] M. Sepehrinour and S. Kasaei, "Perspective reconstruction of non-rigid surfaces from single-view videos," *2017 25th Iranian Conference on Electrical Engineering, ICEE 2017*, no. Icee20 17, pp. 1452–1458, 2017.

[12] J. Yao and R. Taylor, "Assessing accuracy factors in deformable 2D/3D medical image registration using a statistical pelvis model," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2, no. Iccv, pp. 1329–1334, 2003. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-0344983014&partnerID=tZOtx3y1

[13] G. Hichem, F. Chouchene, and H. Belmabrouk, "3D model reconstruction of blood vessels in the retina with tubular structure," *International Journal on Electrical Engineering and Informatics*, vol. 7, no. 4, pp. 724–734, 2015.

[14] S. Sumijan, S. Madenda, J. Harlan, and E. P. Wibowo, "Hybrids Otsu method, Feature region and Mathematical Morphology for Calculating Volume Hemorrhage Brain on CT-Scan Image and 3D Reconstruction," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 15, no. 1, p. 283, 2017. [Online]. Available: http://journal.uad.ac.id/index.php/TELKOMNIKA/article/view/3146

[15] F. Caregiver, A. Introduction, D. Traumatic, M. Tbi, M. Tbi, S. Tbis, A. Tbi, T. B. I. Penetration, F. Violence, C. Changes, and P. Changes, "Fact Sheet Traumatic Brain Injury," pp. 1–6, 2018.

[16] M. Monterial, P. Marleau, and S. A. Pozzi, "Single-View 3-D Reconstruction of Correlated Gamma-Neutron Sources," *IEEE Transactions on Nuclear Science*, vol. 64, no. 7, pp. 1840–1845, 2017.

[17] A. Saxena, S. H.Chung, and A. Y. Ng, "Depth reconstruction from a single still image." *Ijcv*, vol. 74, no. 1, 2007.

[18] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu, "ShapeNet: An Information-Rich 3D Model Repository," 2015. [Online]. Available: http://arxiv.org/abs/1512.03012

[19] D. J. Rezende, S. M. A. Eslami, S. Mohamed, P. Battaglia, M. Jaderberg, and N. Heess, "Unsupervised Learning of 3D Structure from Images," 2016. [Online]. Available: http://arxiv.org/abs/1607.00662

[20] J. Wu, T. Xue, J. J. Lim, Y. Tian, J. B. Tenenbaum, A. Torralba, and W. T. Freeman, "Single image 3D interpreter network," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9910 LNCS, pp. 365–382, 2016.

[21] B. Fan, Y. Rao, W. Liu, and Q. Wang, "Region-Based Growing Algorithm for 3D Reconstruction from MRI Images," pp. 521–525, 2017.

[22] T. Nadu, "Brain Tumor Segmentation of MRI Brain Images through FCM clustering and Seeded Region Growing Technique," vol. 10, no. 76, pp. 427–432, 2015.

[23] M. Zhang, Z. Zhang, and W. Li, "3D Model Reconstruction based on Plantar Image ' s Feature Segmentation," pp. 1–5, 2017.

[24] L. Juan and O. Gwun, "A comparison of sift, pca-sift and surf," *International Journal of Image Processing (IJIP)*, vol. 3, no. 4, pp. 143–152, 2009.

[25] W. Yan, X. Shi, X. Yan, and L. Wang, "Computing OpenSURF on OpenCL and general purpose GPU," *International Journal of Advanced Robotic Systems*, vol. 10, pp. 1–12, 2013.

[26] M. L. Group, M. Intel, D. Ireland, A. Palla, D. Moloney, and L. Fanucci, "Fully Convolutional Denoising Autoencoder for 3D Scene Reconstruction from a single depth image," no. Icsai, pp. 566–575, 2017.

[27] M. Firman, O. M. Aodha, S. Julier, and G. J. Brostow, "Structured Prediction of Unobserved Voxels from a Single Depth Image," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5431–5440, 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7780955/

[28] S. Liu, C. Chen, and N. Kehtarnavaz, "A computationally efficient denoising and hole-filling method for depth image enhancement," vol. 9897, p. 98970V, 2016. [Online]. Available: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2230495

[29] M. Jaiswal, J. Xie, and M. T. Sun, "3D object modeling with a Kinect camera," *2014 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA 2014*, 2014.

[30] J. Xie, Y. Hsu, R. Feris, and M. Sun, "Fine registration of 3D point clouds with iterative closest point using an RGB-D camera," *Circuits and Systems (ISCAS . . . ,* pp. 1–4, 2013. [Online]. Available: http://staff.washington.edu/junx/publication/Fine Registration_ISCAS13.pdf%5Cnhttp://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6572486

[31] H. Avron, A. Sharf, C. Greif, and D. Cohen-Or, " $<$sub$>$1$<$/sub$>$ -Sparse reconstruction of sharp point set surfaces," *ACM Transactions on Graphics*, vol. 29, no. 5, pp. 1–12, 2010. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1857907.1857911

[32] M. Isenburg, Y. Liu, J. Shewchuk, and J. Snoeyink, "Streaming computation of Delaunay triangulations," *ACM Transactions on Graphics*, vol. 25, no. 3, p. 1049, 2006. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1141911.1141992

[33] M. Kowalski, J. Naruniec, and M. Daniluk, "Live Scan3D: A Fast and Inexpensive 3D Data Acquisition System for Multiple Kinect v2 Sensors," *Proceedings - 2015 International Conference on 3D Vision, 3DV 2015*, pp. 318–325, 2015.

[34] P. Besl and N. McKay, "A Method for Registration of 3-D Shapes," pp. 239–256, 1992.

[35] C. Burns, "Texture Super-Resolution for 3D Reconstruction," pp. 4–7, 2017.

[36] J.-y. Bouguet, V. Tarasenko, B. D. Lucas, and T. Kanade, "Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm," *Imaging*, vol. 130, no. x, pp. 1–9, 1981.

[37] A. Plyer, G. Le Besnerais, and F. Champagnat, "Massively parallel Lucas Kanade optical flow for real-time video processing applications," *Journal of Real-Time Image Processing*, vol. 11, no. 4, pp. 713–730, 2016.

[38] S. Tulsiani, T. Zhou, A. A. Efros, and J. Malik, "Multi-view supervision for single-view reconstruction via differentiable ray consistency," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 209–217, 2017.

[39] R. Martin-Brualla, D. Gallup, and S. M. Seitz, "3D Time-Lapse Reconstruction from Internet Photos," *International Journal of Computer Vision*, vol. 125, no. 1-3, pp. 52–64, 2017.

[40] X. Xu, R. Che, R. Nian, and B. He, "Underwater 3D Object Reconstruction with Multiple Views in Video Stream via Structure from Motion," pp. 0–4, 2016.

[41] D. Lapandic, J. Velagic, and H. Balta, "Framework for automated reconstruction of 3D model from multiple 2D aerial images," *Proceedings Elmar - International Symposium Electronics in Marine*, vol. 2017-Septe, no. September, pp. 18–20, 2017.

BIOGRAPHY OF AUTHORS

Hanry Ham is a lecturer and research assistant at Bina Nusantara University with Master of Engineering from The Sirindhorn International Thai-German Graduate School of Engineering in Thailand and German (2016). He obtained Bachelor Degree in Computer Science from Bina Nusantara University (Indonesia) in 2014. His researches are in fields of image processing, computer vision and computer graphics. He is affiliated with IEEE as student member. Besides, he is also involved in student associations, and committee of several competitions such as BNPCHS and ACM-ICPC Regional Asia Site.

Julian Wesley is a lecturer at Bina Nusantara University with Master of Computer Science (M.TI.) major from Bina Nusantara University in 2016. His researches are in fields of image processing, computer vision, and virtual reality. Besides, he is also work as a technology consultant who focused on IT financial industries. He is leading a R&D team in Emerio Indonesia and guide intern students from multiple universities in Indonesia.

Hendra is a lecturer at Bina Nusantara University. He was born in Tanjungpandan, 18 July 1992. He completed his bachelor degree in Bina Nusantara University on 2010. Subsequently he obtained his master degree on 2018. Both degree are in Information Technology. Now, he is working as a Software Engineer at a start-up company in Indonesia.