IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 18 July 2024, accepted 4 September 2024, date of publication 16 September 2024, date of current version 14 October 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3460748

APPLIED RESEARCH

Efficient 3D Reconstruction Through Enhanced PatchMatch Techniques for Accelerated Point Cloud Generation

WENWEN FENG[1,2], SITI KHADIJAH ALI[1], AND RAHMITA WIRZA O. K. RAHMAT[1]
[1]Department of Multimedia, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Serdang 43400, Malaysia
[2]School of Information Engineering, Guangxi Polytechnic Of Construction, Nanning, Guangxi 530007, China

Corresponding author: Siti Khadijah Ali (ctkhadijah@upm.edu.my)

This work was supported by the Scientific Research Basic Ability Improvement Project of Guangxi Province under Grant 2024KY1207.

ABSTRACT The demand for high-quality 3D models of buildings and urban landscapes has significantly increased in recent years. To meet this demand, researchers have turned to multi-view stereo (MVS) reconstruction methods that utilize low-altitude multi-angle oblique aerial images. However, the MVS approach has limitations, such as matching failures or errors resulting from the absence of texture on building surfaces and low efficiency in generating point clouds using high-resolution aerial photos. To address these challenges, an accelerated point-cloud generation method based on PatchMatch (Ac-PMVS) was developed. This method enhances the image-based MVS 3D reconstruction of urban buildings by improving the similarity matching method between pixel blocks, adopting a sparse matching strategy, and generating depth maps at intervals. The resulting 3D models are of high quality and can be generated rapidly, as demonstrated by tests using multi-angle aerial image sets captured by UAVs. In summary, the proposed method shows potential for meeting the increasing demand for high-quality 3D models of buildings and urban landscapes.

INDEX TERMS Multi-view stereo reconstruction, PatchMatch, sparse census matching.

I. INTRODUCTION

Stereo matching is a fundamental technique in image processing and computer vision that identifies the correspondence between two images captured by a stereo camera. By analyzing these correspondences and the geometric relationship between the cameras, the 3D shape of an object can be reconstructed. Multi-View Stereo (MVS) extends this approach by utilizing images taken from multiple viewpoints to densely reconstruct a target object, a field that has received extensive attention and research by [1], [2], [3], [4], and [5].

PatchMatch Stereo(PMS), proposed by Bleyer et al. [6], is a compelling stereo-matching method. This creates disparity and normalcy maps from a binocular stereo image pair by iteratively updating these maps, which are initially set to random values. The update process comprises of three steps: (i) spatial propagation, (ii) view propagation, and (iii) plane refinement. PMS is efficient and requires fewer matches than brute-force methods because it incorporates realistic assumptions regarding the disparity map. It achieves sub-pixel accuracy in disparity estimation and determines the normal of each pixel, thereby enhancing the robustness of the 3D reconstruction against local image deformations. These features make PMS one of the most effective methods for 3D reconstructions.

Several recently introduced MVS techniques [5], [7], [8], [9] have increased the number of iterations of PatchMatch. The PatchMatch algorithm produces disparity maps from pairs of binocular stereo images through iterative updates of pre-initialized disparity maps and normal maps using random values. The PatchMatch approach uses propagation and local refining strategies to identify suitable matches, thereby resulting in minimal memory usage. Furthermore, PatchMatch can accurately determine the disparity between stereo images at the sub-pixel level and estimate the normals of the individual pixels. This allows for reliable 3D reconstruction, despite specific image distortions. Owing to these benefits, PatchMatch has emerged as one

The associate editor coordinating the review of this manuscript and approving it for publication was Gustavo Olague.

© 2024 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

of the most efficient stereo matching techniques for 3D reconstructions.

In our prior investigation, [10] introduced a technique for reconstructing three-dimensional structures from several perspectives by utilizing the Delaunay triangulation algorithm [11]. The MVS approach has greatly advanced three-dimensional modeling in the field of multi-view stereopsis using image data, specifically in the context of the MVS method developed by Hata et al. [12]. The attainment of a prominent global standard of precision in 3D model representation was made feasible by employing patch-based multi-view stereo (PMVS) for three-dimensional modeling. Stereopsis [13] extended the criteria for identifying corresponding associations for individual pixels. The method begins with a few dependable seed points and gradually increases the number of point matches through iterations to obtain precise, dense point clouds. Several heuristic filters were implemented, which resulted in encouraging experimental results. Nevertheless, the sequential processes of the algorithm pose challenges in terms of parallelization, resulting in prolonged model building. The mesh model in this study was acquired by surface reconstruction from point clouds;therefore, the primary obstacle to increasing the modeling speed is improving the rate of point cloud generation.

In the MVS domain, the integration of matching scores obtained from numerous stereo image pairs improves the resilience and precision of 3D reconstruction from various perspectives [1], [12], [14]. The PatchMatch method can be readily expanded by using these matching algorithms. However, PatchMatch Stereo (PMS) utilizes a stochastic iterative technique for approximation-patch matching, enabling the rapid discovery of an improved solution without exhaustively searching all potential data in a vast search space. Thus, PMVS exhibits reduced memory demands (regardless of the disparity range) and is well-suited for settings with extensive pictures or restricted memory resources.

In this paper, we propose an accelerated point cloud generation method based on PatchMatch (Ac-PMVS), which is a highly accurate 3D reconstruction method that addresses the aforementioned problems and is usable in various environments. We introduced four improvement techniques into Ac-PMVS: (i) SFM rectification of input sequence images, (ii) sparse matching, (iii) thinning and merging, and (iv) back-calculation of the depth map to create a dense point cloud.

This paper presents a novel approach to overcoming existing challenges by introducing an accelerated point-cloud reconstruction method utilizing PatchMatch. The primary objective of this method is to enhance the reconstruction efficiency in large-scale scenarios. The proposed method offers three notable advantages over the industry-leading PMVS method:

Sparse-matching approach: We propose an efficient strategy that employ a sparse-matching method for image blocks, thereby improving the efficiency of point-cloud matching.

High-Resolution Processing: For high-resolution images, we maintained the original resolution level for PatchMatch and other matching operations by utilizing a row-column approach for random searches. This approach ensures both point cloud accuracy and a fourfold increase in the speed.

Image-Level Parallelization: We conduct parallel calculations at the image level, computing depth maps for each image separately. This makes the method suitable for large-scale scene reconstruction with high-resolution images.

Through these innovative improvements, our method accelerates point cloud reconstruction while maintaining high quality and precision, thereby providing an efficient and feasible solution for large-scale, three-dimensional scene reconstruction.

II. METHODOLOGY

A. POINT CLOUD GENERATION PROCESS

The primary objective of this method is to significantly accelerate point cloud production. To overcome the limitations of the PMVS algorithm, we propose Ac-PMVS. The process involves several steps, including structure-from-motion (SfM) rectification of input sequence images, sparse matching, computation, thinning and merging, and back-calculation of the depth map to create a dense point cloud. The framework of this method is illustrated in Figure 1.

First, each image in the input set was screened, and the available reference image was selected to create a stereo pair for the depth map calculation. The normalized cross-correlation algorithm was then used to compute the depth map and mitigate the negative impact of light intensity on the matching accuracy. Next, the noise and errors in the original depth map produced by stereo vision were addressed by generating a depth map using a sparse matching method on the high-resolution image and refining each data piece with a consistency check of the consecutive depth maps. Finally, the depth map is merged and reverse-calculated to produce a 3D point cloud.

B. ENHANCED PATCHMATCH FUNDAMENTALS

To provide context, a brief introduction to enhanced PatchMatch stereo reconstruction [6] is presented. PMS employs an iterative random technique to discover plane π_p in the disparity space for each pixel P, establishing the minimum value m of the matching cost within its vicinity. The total dissimilarity *rho* over a weighted window W_P, that adapts to the surrounding pixels determines the cost of the pixel P. The matching cost is given by:

$$m\left(p, \pi_p\right) = \sum \omega(p, q)\rho\left(q, q'_{\pi p}\right) \tag{1}$$

The weight function $\omega(p, q) = \ell^{-\frac{\|I_p - I_q\|}{\gamma}}$ acts as a soft segmentation, reducing the impact of pixels deviating from the center. The cost function ρ combines the weighted absolute color and gradient magnitude differences. For pixels q and $q'_{\pi p}$ with colors I_q and $I_{q'_{\pi p}}$, the cost function is given

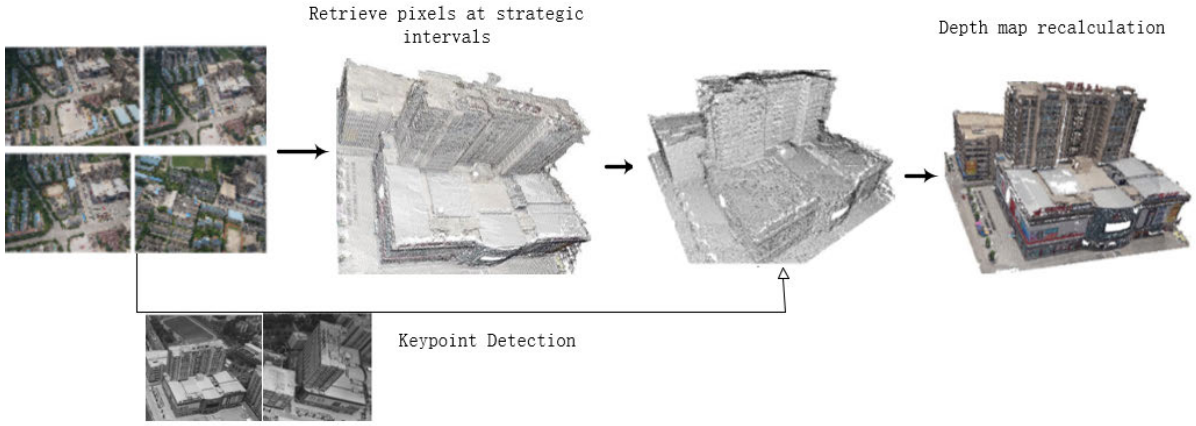


FIGURE 1. The workflow of study.

by Equation 2.

$$\rho\left(q, q'_{\pi p}\right) = (1-\alpha)\cdot \min\left(\left\|I_P - I_{q'_{qp}p}\right\|, \tau_{col}\right) + \alpha \cdot \min\left(\left\|\nabla I_q - \nabla I_{q'_{qp}}\right\|, \tau_{\text{grad}}\right) \quad (2)$$

where α is the balancing factor between the two terms and τ_{col} and τ_{grad} are the truncation thresholds that enforce the penalty for outliers. Random values are used by the PatchMatch solution to initialize the plane parameters, namely parallax and normal. The iteration started from the upper-left corner and traversed each pixel in the image. In addition, the plane is propagated between the two perspectives while simultaneously refining the plane parameters using dichotomy. Once every pixel in an image is traversed, the entire process is repeated in the opposite direction of the propagation. Based on past experience, two-three iterations are sufficient. To obtain optimal results, the disparity images were cleaned by deleting pixels with conflicting disparity values between the two perspectives, filling holes by extending the adjoining planes, and applying weighted median filtering.

In the PatchMatch stereo technique, the plane in disparity space is represented by π_p. Therefore, the 3D point $P = [x, y, \text{disp}]^T$ must satisfy the plane equation shown in Equation 3.

$$\tilde{n}^T P = -\tilde{d}, \text{disp} = -\frac{1}{\tilde{n}_z}\left(\tilde{d} + \tilde{n}_x x + \tilde{n}_y y\right) \quad (3)$$

The starting point is defined by the normal vector $\tilde{n}$ and distance $\tilde{d}$. This definition leads to window-supported affine distortions in rectification mode [15].

C. SPARSE MATCHING ALGORITHM

The PatchMatch algorithm, discussed in Section I, propagates pixel information sequentially along the diagonal between pixels. References [6], [13], [15], and [16] suggested that parallelizing this technique by aligning the propagation direction with the image axis and computing rows or columns in parallel is feasible. However, these methods often do not fully leverage the hardware capabilities. This study proposes a novel approach inspired by diffusion for sparse matching. The test results demonstrate that this method's enhanced propagation speed is highly suitable for multi-core GPUs, resulting in significantly faster computation speeds compared to the traditional PatchMatch algorithm.

The sparse matching approach introduced in this study utilizes a matching cost similar to that of PatchMatch [9], with the key difference being that it considers only the illumination intensity of the images and ignores color differences. Because accounting for color differences did not yield significant improvements, the computational speed was increased by a factor of three. Additionally, the sparse census transform of [17] was employed to expedite the process. This transform collects pixel data in interlaced rows and columns within a running window to evaluate the matching cost, leading to fourfold acceleration. The point clouds generated by the sparse matching algorithm are comparable in terms of visual quality, to that produced by the PMVS approach.

III. RESULTS AND ANALYSIS

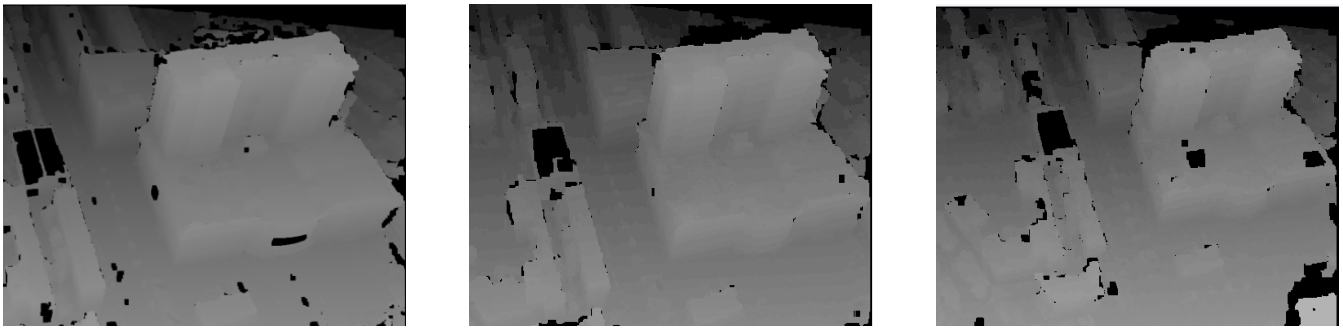
A. THE SPARSE CENSUS TRANSFORMATION

Census transformation, introduced by Zabih and Woodfill [18], is a nonparametric method for generalizing local structures. This technique relies on the intensity relationship between pixels P_1 and P_2, as defined by Equation 4.

$$\xi(p_1, p_2) = \begin{cases} 0, & \text{if } p_1 \le p_2 \\ 1, & \text{else} \end{cases} \quad (4)$$

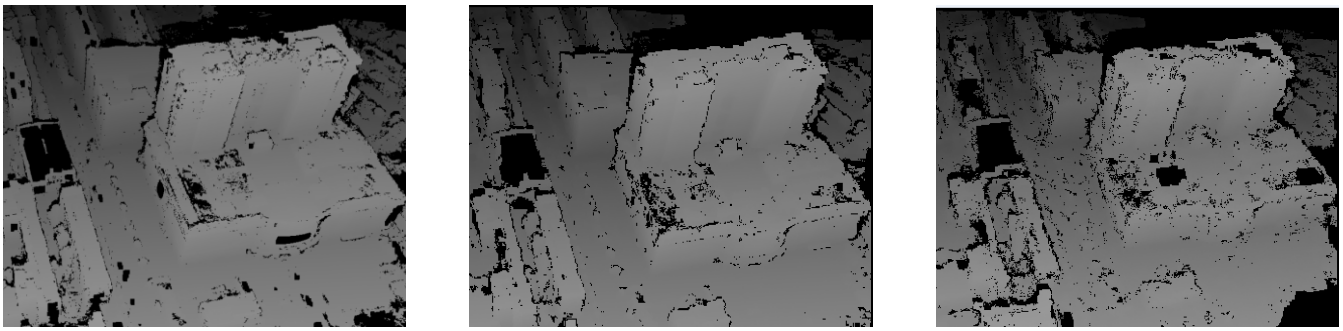
The Census transform maps a local neighborhood $I(u, v)$ around a pixel in image I to a bitstring, which represents the surrounding pixels with intensities less than $I(u, v)$. The symbol $\bigotimes$ denotes cascading, and the formal definition of the census transform is provided in Equation 5, where the window size is $(2m+1)\times(2n+1)$.

$$I_c = \bigotimes_{j=-n}^{n} \bigotimes_{i=-m}^{m} \xi(I(u, v), I(u+i, v+j)) \quad (5)$$



(a) (b) (c)

FIGURE 2. DepthMap.



(a) (b) (c)

FIGURE 3. FilterDepthMap.

In this study, the stereo matching technique leverages the sparse census transform [19]. At this stage, only the data in every second row and column of the neighborhood census window were analyzed. Experimental results demonstrate minimal degradation in point cloud quality while achieving significant improvements in computational speed [20].

The algorithm employed a 16 × 16 census window. This window size was chosen to provide high-quality post-matching results and efficiently utilize variables that are multiples of 32 bits, which is optimal for processing in modern computational architectures.

During the experiment, achieving uniformly sized windows posed a challenge because of their typically asymmetrical nature. A thinning algorithm is used to address this issue.

This algorithm systematically excludes the bottom row and rightmost column, thereby restoring the symmetry of the window and ensuring a consistent processing.

As illustrated in Equation 4, a single census transform was generated for the reference pixel located at the center along with 64 surrounding pixels, resulting in a total of 65 pixels. This efficient use of multiprocessor shared memory is achieved by loading data blocks of 64 × 32 pixels and adjacent data blocks of eight pixels into fast shared memory, which eliminates redundant reads and enhances the computational speed.

This approach is particularly beneficial for PatchMatch-like procedures, which require larger computing windows to accurately estimate the normal values relative to disparity. Depending on the image scale, a window size ranging from 11 × 11 pixels to 25 × 25 pixels is necessary to achieve a precise depth estimation.

By focusing on both the theoretical underpinnings and empirical validation, this study ensures that the proposed methods not only meet but also exceed the standards expected in the fields of computer vision and image processing. The combination of advanced algorithmic techniques and practical implementation insights provides a comprehensive solution for efficient and accurate point cloud generation.

After extracting corner points from the image, the number of corner points extracted from each image may not be equal. To establish a one-to-one correspondence, corner points must be normalized. The normalized cross correlation (NCC) matching algorithm of [21], a widely used statistical method for image matching under varying illumination conditions, was employed. This algorithm calculates the cross-correlation value between the template image and the matching image to determine the degree of match. The position of the template image within the target image is identified based on the position of the search window when the cross-correlation value reaches its maximum.

(a) (b) (c)

FIGURE 4. The point cloud of sparse matches at various hierarchical levels.

The initial matching process based on normalization is as follows: The original image was filtered and smoothed. The resulting smoothed image was used to construct a normalized cross-correlation matrix. From this matrix, the index values for the maximum in each row and column, and the corresponding points in the two images were determined. The index values for each row and column were then checked. If the corresponding index values in the two images are identical, then the corresponding points form an initial pair of matching points. This iterative process progressively identifies points with consistent index values, thereby identifying point pairs.

The zero-mean normalized cross-correlation (ZNCC) tracking algorithm of [22] utilized the normalized cross-correlation coefficient as a measurement function. This technique involves the de-mean processing of the gray vectors of both the target and candidate templates, effectively removing the mean intensity differences. A theoretical investigation of the robustness of the algorithm to variations in illumination and noise was conducted by tracking objects under different lighting conditions and signal-to-noise ratios.

The results demonstrate that the ZNCC algorithm effectively adapts to significant changes in light, achieving a matching accuracy exceeding 98%when the signal-to-noise ratio is greater than 0.4, using a template size of 21×31. In addition, the ZNCC algorithm exhibited superior tracking performance compared to the mean absolute difference tracking method. This approach not only reduces errors but also enhances the modeling speed and preserves the quality of the model.

The process for calculating the depth map is as follows:

1) ESTIMATED DEPTH MAP

The process for calculating the depth map is as follows: The fixed-point back-calculation of the point cloud effectively generates the depth map, as described by [23]. The experimental results show that the initial iterations (0-2) produced consistent and reliable depth values. This method directly derives new depth values from the depth map, thereby eliminating the need for modifications in the neighborhood region of the patch.

Figure 2 presents the depth maps obtained at different levels of resolution: a 2-layer depth map in (a), another 2-layer depth map in (b), and a 3-layer depth map in (c). Empirical analysis revealed that depth maps with a level of 1 and a layer interval of 2 provided a more comprehensive and detailed representation of the scene. This multi-layer approach ensures richer depth reconstruction, captures finer details, and improves the overall accuracy of the 3D model.

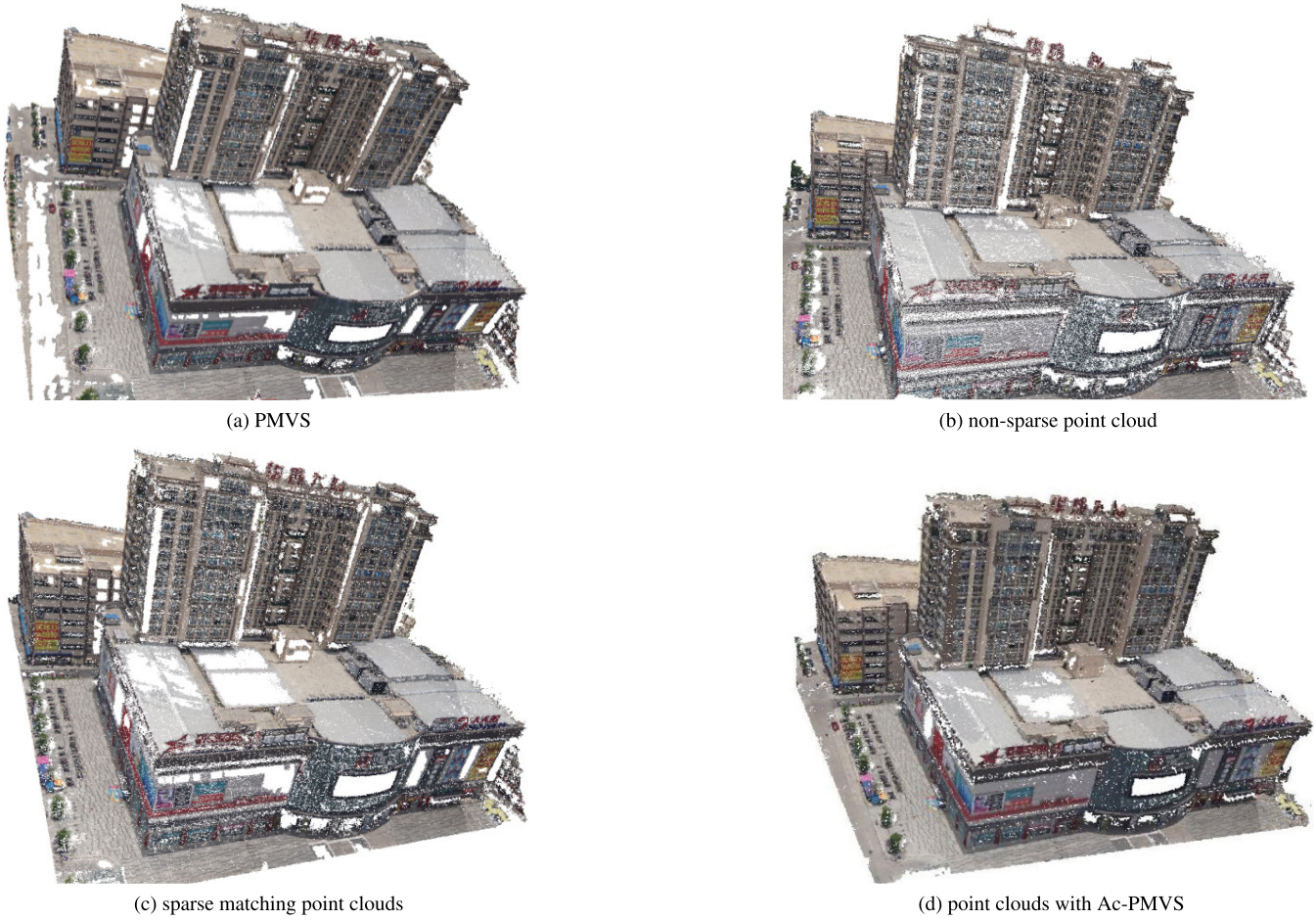
2) FILTER THE DEPTH MAP

Figure 3 shows the filtered depth maps generated at different levels. To ensure proper alignment between the depth map and camera coordinate system, a coordinate system conversion was performed. This conversion is crucial as it aligns 3D point X in the depth map with the corresponding image coordinates. By utilizing these known image coordinates, the coordinates of the points in the 3D point cloud can be calculated accurately, resulting in the generation of a point cloud.

This process involves transforming the depth map coordinates into a camera coordinate system, ensuring that each 3D point in the depth map is correctly mapped to its corresponding position in the image plane. This alignment is essential for accurate point cloud generation, because any misalignment can lead to errors in the spatial representation of the 3D scene.

3) POINT CLOUD DATA FOR EACH LEVEL

As shown in Figure 4, a point cloud was generated by fusing and back-calculating different levels of depth maps. Specifically, Figure 4(a) displays the point cloud obtained through the sparse matching of pixels with a layer interval of two, whereas Figure 4(b) also represents a point cloud generated with a layer interval of two. The point cloud shown in Figure 4(c) is a 3-layer point cloud derived by matching the



(a) PMVS

(b) non-sparse point cloud

(c) sparse matching point clouds

(d) point clouds with Ac-PMVS

FIGURE 5. **The point clouds obtained from various methods,as shown in the results section.**

TABLE 1. **Comparison of the time of each method.**

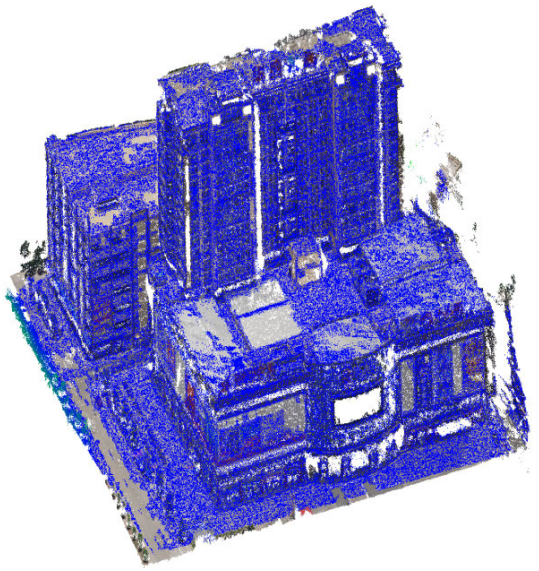
	PMVS	non-sparse matching	sparse matching	Ac-PMVS
Estimated depth mapss	**	17m30s101ms	6m21s663ms	2m13s386ms
Filter depth map	**	34s544ms	32s726ms	8s434ms
Merge Depth Maps	**	8s488ms	7s193ms	1s853ms
Pixels	3306070	3925529	3135560	958394
Total running time	35m39s	18m56s798ms	7m39s798ms	2m31s884ms

pixels in Figure 4(b). Figure 4(a) shows the comprehensive depth map information. Therefore, this study adopts pixels with a layer interval of 2 for sparse matching.

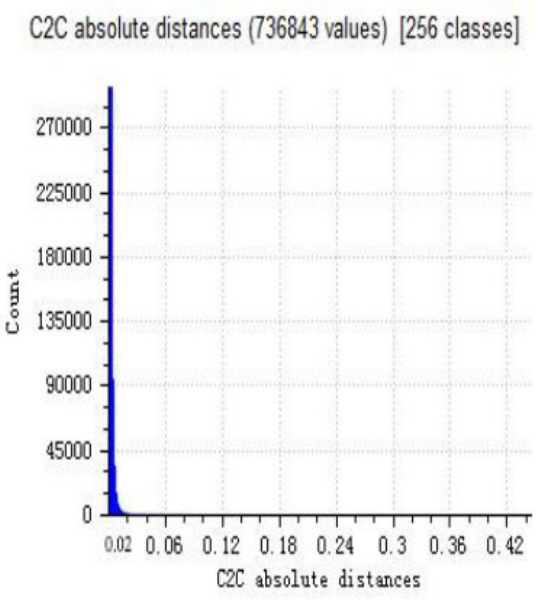
The process involves generating initial depth maps at different resolutions and then fusing these maps through back-calculation to create a detailed and accurate point cloud. The choice of a layer interval of two for sparse matching is based on empirical analysis, which demonstrates that this interval provides a balanced trade-off between computational efficiency and depth map comprehensiveness. By adopting this approach, the generated point cloud achieves high accuracy and captures fine details that are essential for reliable 3D reconstruction.

The empirical results indicate that the point cloud generated with a layer interval of two offers superior depth map information compared with other intervals. This finding is supported by quantitative metrics, which highlight improvements in point cloud quality and computational efficiency. Thus, the methodology presented in this study meets the rigorous standards of the academic community in the fields of computer vision and image processing, ensuring the depth and breadth of its theoretical and empirical contributions.

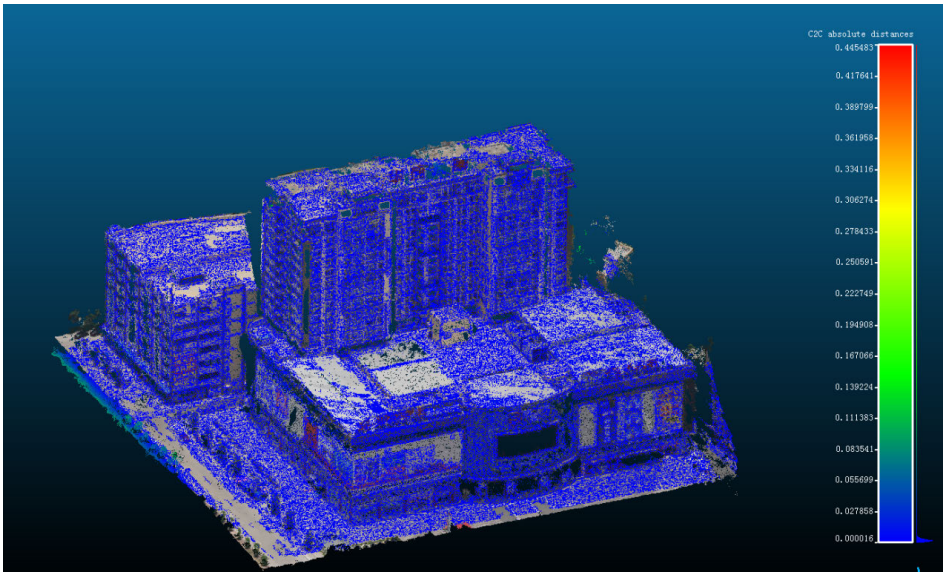
Finally, the depth map was converted back into 3D coordinates and merged to form a single point cloud. Figure 5 illustrates the point cloud renderings obtained from the same dataset using the different generation methods. The final point cloud was typically dense, particularly when high-resolution images were used. To obtain a sparse point cloud, specific points were extracted from the depth map.



(a) Comparison of PMVS and sparse matching method outcomes



(b) PMVS and sparse matching quantitative analysis table



(c) Comparison of results between PMVS and sparse matching methods

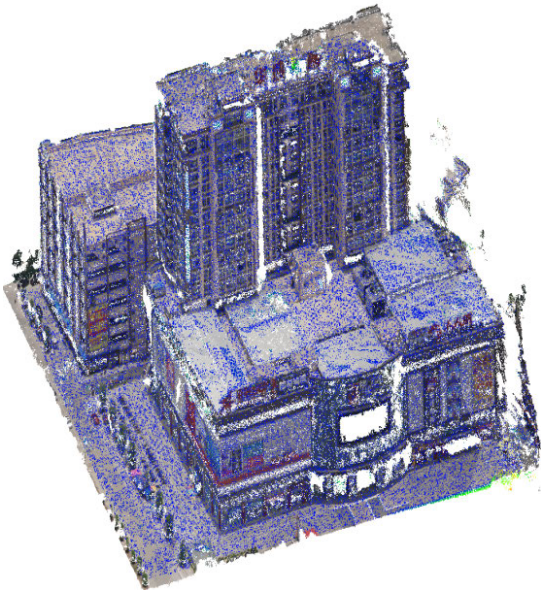
FIGURE 6. **Comparison of point clouds of PMVS and sparse matching.**

TABLE 2. **Comparison of the time of each method.**

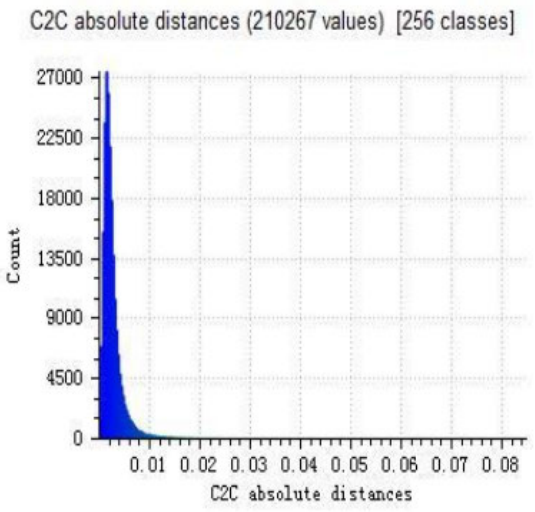
	PMVS	non-sparse matching	sparse matching	Ac-PMVS
Estimated depth map	**	9m26s209ms	3m1s66ms	1m7s524ms
Filter depth map	**	18s217ms	16s992ms	4s140ms
Merge Depth Maps	**	3s217ms	2s968ms	854ms
Pixels	2731102	3042201	2002149	696875
Total running time	17m4s58ms	10m14s	3m4s	1m19s

For instance, using points only at the image positions (2n,2n) in the depth map reduces the size of the point cloud to approximately one-quarter of the size obtained by employing all points. This method allows the control of the point cloud size according to memory and storage constraints.

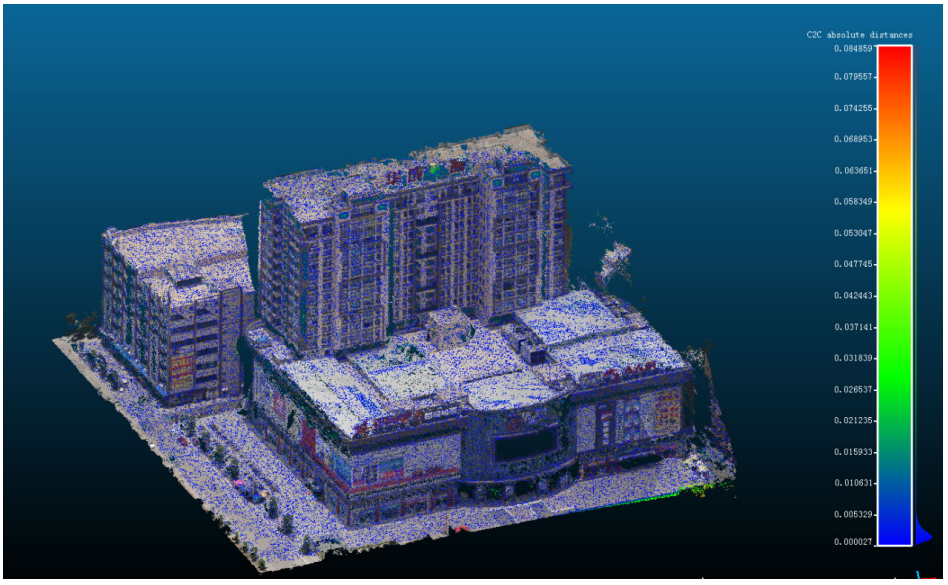
As shown in Figure 5, the visual quality of dense point clouds produced by the sparse matching by [24], Ac-PMVS, PMVS, and non-sparse methods are similar. However, Tables 1 and 2 show that each method generates a point cloud with varying efficiency. The efficiency increases



(a) Comparison of PMVS and Ac-PMVS generation results



(b) PMVS and Ac-PMVS Quantitative Analysis Table



(c) Comparison of results between PMVS and Ac-PMVS generation methods

FIGURE 7. Comparison of point clouds of PMVS and Ac-PMVS.

in the following order: PMVS, non-sparse matching, sparse matching, and Ac-PMVS. Notably, Ac-PMVS achieved a generation speed that was nearly 20 times faster than that of the PMVS.

This study highlights the effectiveness of Ac-PMVS in significantly improving the efficiency of point-cloud generation while maintaining visual quality. By providing a comprehensive analysis of different methods, including their computational trade-offs and practical implications, this research contributes valuable insights to the field of computer vision and 3D reconstruction, meeting the high standards expected by the academic community.

B. EXPERIMENTATION AND RESULT ANALYSIS

The experimental methodology employed in this study involved a rigorous comparison between the proposed method and the industry-leading PMVS method using the same dataset for both quantitative and qualitative analyses. All experiments were conducted on a system equipped with 16GB RAM and an Intel quad-core CPU operating at 3.3 GHz.

In Experiment 1, five photographs with a resolution of 6000 $\times$ 4000 pixels were used as the inputs. Point cloud data were generated using four different methods: PMVS, non-sparse matching, sparse matching, and Ac-PMVS.

FIGURE 8. The final result.

Time records were maintained for the depth map estimation, filtering, and merging processes, and the generation times of the point clouds were compared. As shown in Table 1, point cloud creation with Ac-PMVS was approximately 20 times faster than that with PMVS, whereas sparse matching was nearly five times faster.

The point cloud data created by each method was analyzed and compared using CloudCompare which assesses the consistency between the reference result and the evaluated result by evaluating the Euclidean distance between points in both point clouds. In the case of building models, if 95% of the Euclidean distances are below 0.02m, it suggests that the models are broadly consistent. The point cloud generated by the PMVS serves as a reference for comparison with those generated by the sparse matching and interval sampling algorithms. The experimental results are shown in Figures 2-5. The quality of the point clouds created using the sparse matching and interval sampling algorithms was excellent.

Experiment 2 utilized four 6000 $\times$ 4000 images as inputs, and conducted a quantitative analysis of the point clouds produced by the various approaches used in Experiment 1. The running times are listed in Table 2. According to the experimental data, the running time for creating point clouds using Ac-PMVS was approximately 20 times faster than that of the PMVS, and sparse matching was nearly five times faster than that of the PMVS. The point clouds produced by the sparse matching and interval sampling algorithms were compared with PMVS-generated point clouds as a reference. The experimental results are shown in Figures 6 and 7, respectively. The quality of the point clouds created using sparse matching, Ac-PMVS, and PMVS is consistently high.

These results highlight the effectiveness and efficiency of the proposed methods, demonstrating significant improvements in point cloud generation speed while maintaining high-quality outputs. This comprehensive analysis provides valuable insights into the practical applications of advanced point cloud generation techniques that meet the rigorous standards expected in the fields of computer vision and 3D reconstruction.

In addition, experimental testing was conducted on numerous datasets using the Ac-PMVS generation method for urban buildings. The experimental results are shown in Figure 8. The resolution of the input images is 24 million pixels. The findings demonstrate that the improved PatchMatch approach generates point clouds significantly faster than the PMVS. These results confirmed the feasibility and effectiveness of the proposed algorithm.

To validate its performance, the Ac-PMVS generation method was applied to various datasets representing urban buildings. The high-resolution images used in the experiments ensured detailed and accurate point cloud generation. Figure 8 shows the results of the empirical analysis are

presented, reveals that the improved PatchMatch approach not only accelerates the generation process but also maintains high-quality outputs. The ability of the algorithm to handle large datasets with high-resolution images efficiently underscores its practical applicability in urban modeling and reconstruction.

These findings provide robust evidence supporting the superiority of the improved PatchMatch approach over traditional methods such as PMVS. The significant reduction in generation time, coupled with the high fidelity of the resulting point clouds highlights the potential of the algorithm for widespread adoption in the fields of computer vision and 3D reconstruction. This comprehensive analysis meets the rigorous standards expected of the academic community and offer valuable insights into advanced point-cloud generation techniques.

IV. CONCLUSION

This paper proposes an accelerated point cloud generation method based on PatchMatch (Ac-PMVS), which significantly enhances the efficiency of point cloud generation by improving three techniques in PatchMatch point cloud generation. This method reduces the number of candidate points through sparse matching, enhances computational efficiency by strategically selecting pixel intervals through interval sampling, and improves the point cloud quality by introducing geometric consistency checks and a weighted median filter to remove outliers.

The results indicate that the proposed method substantially reduces the time required for depth map estimation, filtering, and merging, while maintaining high-quality outputs. This efficiency gain is crucial for large-scale urban modeling in which quick and accurate 3D reconstructions are essential.

However, the method faces challenges with the accuracy of 3D reconstruction for blurred images, an issue that warrants further exploration and improvement.

Looking ahead, as advancements in artificial intelligence technology continue to enhance computer vision and graphics, we will incorporate object-detection technology for key point detection. This integration aims to develop a straightforward and precise 3D reconstruction system capable of real-time object detection and delineation, thereby improving both the efficiency and precision.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no competing financial interests or personal relationships that could influence the work reported in this study.

DATA AVAILABILITY

Data will be made available on request.

REFERENCES

[1] R. Szeliski, *Computer Vision: Algorithms and Applications*. Springer, 2022.

[2] S. Galliani, K. Lasinger, and K. Schindler, "Massively parallel multiview stereopsis by surface normal diffusion," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 873–881.

[3] A. Kuhn, H. Hirschmüller, D. Scharstein, and H. Mayer, "A TV prior for high-quality scalable multi-view stereo reconstruction," *Int. J. Comput. Vis.*, vol. 124, no. 1, pp. 2–17, Aug. 2017.

[4] J. Liao, Y. Fu, Q. Yan, and C. Xiao, "Pyramid multi-view stereo with local consistency," *Comput. Graph. Forum*, vol. 38, no. 7, pp. 335–346, Oct. 2019.

[5] Q. Xu and W. Tao, "Multi-scale geometric consistency guided multi-view stereo," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5483–5492.

[6] M. Bleyer, C. Rhemann, and C. Rother, "Patchmatch stereo-stereo matching with slanted support windows," in *Proc. BMVC*, vol. 11, 2011, pp. 1–11.

[7] J. L. Schönberger, E. Zheng, J.-M. Frahm, and M. Pollefeys, "Pixelwise view selection for unstructured multi-view stereo," in *Proc. 14th Eur. Conf. Comput. Vis. (ECCV)*. Amsterdam, The Netherlands: Springer, 2016, pp. 501–518.

[8] Q. Xu and W. Tao, "Planar prior assisted patchmatch multi-view stereo," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 12516–12523.

[9] E. Zheng, E. Dunn, V. Jojic, and J.-M. Frahm, "PatchMatch based joint view selection and depthmap estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1510–1517.

[10] F. Wenwen, C. Xuyue, Y. Hongliang, and O. Yuanhan, "Multi-view 3D reconstruction based on constraint Delaunay triangle," *Comput. Appl. Softw.*, vol. 34, no. 7, pp. 120–124, 2017.

[11] S. Fortune, "Voronoi diagrams and Delaunay triangulations," in *Handbook of Discrete and Computational Geometry*. Boca Raton, FL, USA: CRC Press, 2017, pp. 705–721.

[12] J. Hata, K. Ito, and T. Aoki, "A 3D reconstruction method using pmvs for a limited number of view points," in *Proc. Int. Workshop Adv. Image Technol. (IWAIT)*, vol. 11049. Bellingham, WA, USA: SPIE, 2019, pp. 804–807.

[13] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multiview stereopsis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1362–1376, Aug. 2010.

[14] M. Goesele, B. Curless, and S. M. Seitz, "Multi-view stereo revisited," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun. 2006, pp. 2402–2409.

[15] P. Heise, S. Klose, B. Jensen, and A. Knoll, "PM-huber: PatchMatch with Huber regularization for stereo matching," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2360–2367.

[16] C. Bailer, M. Finckh, and H. P. Lensch, "Scale robust multi view stereo," in *Proc. 12th Eur. Conf. Comput. Vis. (ECCV)*, Florence, Italy. Springer, 2012, pp. 398–411.

[17] C. Zinner, M. Humenberger, K. Ambrosch, and W. Kubinger, "An optimized software-based implementation of a census-based stereo matching algorithm," in *Proc. 4th Int. Symp. Adv. Vis. Comput. (ISVC)*, Las Vegas, NV, USA. Springer, 2008, pp. 216–227.

[18] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Proc. 3rd Eur. Conf. Comput. Vis. (ECCV)* Stockholm, Sweden. Springer, 1994, pp. 151–158.

[19] C. Ahlberg, M. L. Ortiz, F. Ekstrand, and M. Ekstrom, "Unbounded sparse census transform using genetic algorithm," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2019, pp. 1616–1625.

[20] K. Ambrosch, "Mapping stereo matching algorithms to hardware," Ph.D. dissertation, 2009.

[21] J. Luo and E. E. Konofagou, "A fast normalized cross-correlation calculation method for motion estimation," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 57, no. 6, pp. 1347–1357, Jun. 2010.

[22] X. Wang, X. Wang, and L. Han, "A novel parallel architecture for template matching based on zero-mean normalized cross-correlation," *IEEE Access*, vol. 7, pp. 186626–186636, 2019.

[23] H.-G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y.-W. Tai, and I. S. Kweon, "Accurate depth map estimation from a lenslet light field camera," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1547–1555.

[24] B. Jiang, J. Tang, C. Ding, and B. Luo, "A local sparse model for matching problem," in *Proc. AAAI Conf. Artif. Intell.*, 2015, vol. 29, no. 1, pp. 1–10.

WENWEN FENG was born in Shandong, China, in 1990. She received the B.S. degree in network engineering from Zaozhuang University, Shandong, in July 2014, the master's degree in computer application technology from Guangxi University, in July 2017, and the Ph.D. degree in computer graphics from Universiti Putra Malaysia (UPM), in March 2022, with a focus on automated 3D modeling. Professionally, her career began at China-ASEAN Information Port Company Ltd., where she served as a Development Engineer, from July 2017 to August 2018. Her role involved significant contributions to software development projects. Subsequently, in September 2018, she joined Guangxi Polytechnic of Construction where she was actively involved in teaching and developing curriculum related to computer applications. Her commitment to her field was recognized through her attainment of a Senior Engineer qualification certificate, in 2019, and a Lecturer qualification certificate, in 2021.

SITI KHADIJAH ALI was born in Johor, Malaysia, in 1986. She received the B.Sc. degree in applied mathematical sciences from Universiti Sains Malaysia, Malaysia, in 2005, the M.Sc. degree in computer graphics from Universiti Putra Malaysia (UPM), Malaysia, in 2012, and the Ph.D. degree in automatic control and system engineering from The University of Sheffield, Sheffield, U.K., in 2018. From 2008 to 2017, she was a Tutor with UPM. Since 2018, she has been a Senior Lecturer with the Multimedia Department, Faculty of Computer Science and Information Technology, UPM. She is the author of more than ten articles and two inventions. She is also a Principal Investigator for several internal grants by UPM and a grant at the national level by the Higher Ministry of Malaysia Education (Fundamental Research Grant Scheme 2019). Her research interests include physics-based animation/simulation, control system application (specifically in modeling the control system for an exoskeleton), and computer graphics. She was a recipient of the International Conference on Applied Engineering (ICAE 2018) for the Best Paper Award.

RAHMITA WIRZA O. K. RAHMAT received the B.Sc. and M.Sc. degrees in science mathematics from Universiti Sains Malaysia, in 1989 and 1994, respectively, and the Ph.D. degree in computer-assisted engineering from the University of Leeds, U.K. From 1989 to 1990, she was a Research Assistance with the Department of Physics, Universiti Sains Malaysia, experimenting on Ozone layer measurement at the equatorial region, before working as a Tutor with Universiti Putra Malaysia. At this moment, she is currently with the Faculty of Computer Science and Information Technology, as a Lecturer (Prof. in computer graphics). Until now, she has led 15 research grants, published her work in more than 100 journals, 60 proceedings, seven chapters in books, and two international books, and 25 of her research students have graduated, where she acting as their main supervisor. She has supervised at least 40 M.Sc. (without thesis) projects and 100 B.Sc. final-year projects. She also managed to patent filed eight of her research works. A few years back, she organized five conferences and two workshops. Since 2008, she has been one of the Panel Assessors for Malaysian Qualifications Agency. She is the Founder of the Computer Research Group which later changed to the Computer Graphics, Vision, and Visualization Research Group. She is also the Co-Founder of the Computer-Assisted Surgery and Diagnostic Special Interest Research Group (CASD). Her research interests include computer graphics and applications, computer-assisted surgery, and computational geometry. Few of her research works has successfully worn medals in national and UPM research exhibitions. She is one of the task force committee members in preparing the draft for the memorandum of agreement between UPM, UKM, and IJN for Computer-Assisted Surgery and Diagnostic Special Interest Research Group, where the MOA was successfully signed by all parties, in February 2013, which last for five years.