Enhancement of Depth Value Approximation using Noise Filtering and Inverse Perspective Mapping Techniques for Image Based Modelling

Rahmita Wirza O.K Rahmat[1], Ng Seng Beng [2], Intan Syaherra Ramli[3*]

[1,2] *Multimedia Department, Faculty of Computer Science & Information Technology, Univeristi Putra Malaysia, Serdang, Malaysia*
[3] *Computing Science Studies, College of Computing, Informatics and Media,Universiti Teknologi MARA (UiTM) Cawangan Negeri Sembilan, Kuala Pilah, Malaysia*

*Corresponding author: * intan3885@uitm.edu.my*

HIGHLIGHTS

- The fast and accurate depth value approximation using trigonometry equation and image relation for complex object in 3D Modelling.
- Good feature to Track and Pyramidal Lucas Kanade Optical Flow is used for Feature Extraction and Matching.
- The noise filtering and Inverse Perspective Mapping Technique is used to enhance the depth value approximation.

ABSTRACT

This article proposes the methods to enhance the depth value approximation in 3D Image Based Modelling for complex object. Fundamentally, the fast and accurate depth value approximation is crucial as the 3D modelling used in virtual and augmented reality applications, reverse engineering, and the architecture. Therefore, the enhanced method must be robust against the challenges with noise, complexity, distortion and longer processing time. In this experiment, five small and complex objects were captured using a turntable, laptop, and a webcam. The feature points between images were tracked and matched using good features to tracks and Pyramidal Lucas Kanade's optical flow. Next, the depth value was approximated using trigonometry equation. To enhance the accuracy, the noise filtering, and Inverse Perspective Mapping (IPM) were introduced. The results show that the average error based on the approximated width and depth dimensions was 3.27% and 6.88% compared with the actual object. Furthermore, the processing speed was 1519 points per second. Therefore, this method enhanced the depth value approximation, which can be used to build the full texture 3D model in future.

Keywords: Depth value approximation, Optical Flow, Trigonometry

INTRODUCTION

As the demand for virtual and augmented reality applications has expanded in various research fields such as engineering and architecture, 3D reconstruction has emerged as an important topic in computer vision. 3D reconstruction is the process of replicating real-world objects from 2D images based on depth value

approximation using contact and non-contact-based techniques. The contact-based method utilized the specialised hardware such as Coordinate Measuring Machine (CMM) that physically contact the object. This approach has simplified the traditional method, which made use of software such as Maya, Blender and 3D Studio Max. Meanwhile, the non-contact-based method which consists of active and passive approach did not interfere the object. The active process uses Light Detection and Ranging (LiDAR) to measure the distance between the sensor and the object by emitting multiple laser beams simultaneously, but it is costly (Awange et al., 2019). Meanwhile, the Kinect sensor to measure the depth is inexpensive, but it has a limited operating range and sensitive to the environmental conditions. It is also having the challenges to reconstruct certain classes of objects such as dark and shiny (Sasadara et al., 2020). In contrast, the non-contact-based such as Image Based Modelling (IBM) is hardware-independent, widely used on mobile devices and suitable for beginner. Furthermore, the use of multi-view 3D tools such as PhotoScan, 123D Catch by Autodesk, Visual SfM, and Arc3D were easier but has visual quality and accuracy issues (Mousavi et al., 2018; Elkhrachy, 2022; Davide et al., 2022).

Image Based Modelling Approach

The most common methods to create the 3D models from 2D images are using the volumetric and feature-based approach (Shalma et al., 2021; Lin et al.,2019; Ding et al., 2021). The volumetric-based approaches, such as silhouette and space carving are relied on the extraction of the background from grayscale images. However, this approach suffers from complexity and longer processing time (Hwang et al., 2017; Agudo, 2022). This is due to multiple voxels with varying resolutions that involved in the process. Meanwhile, the feature-based methods involve feature extraction, matching, and merging of the feature points from wide range of complex objects (Tushar et al., 2017). However, feature based method using Structure from Motion (SfM) requires many related images to reconstruct the 3D points. Consequently, the time required to generate a complete 3D point was increased (Wang et al., 2019). Existing methods also have an issue with precision impacted by image uncertainty relations (Zhou et al., 2016; Chen et al., 2022). Figure 1 depicts the methods that will be discussed in more detail in this section.
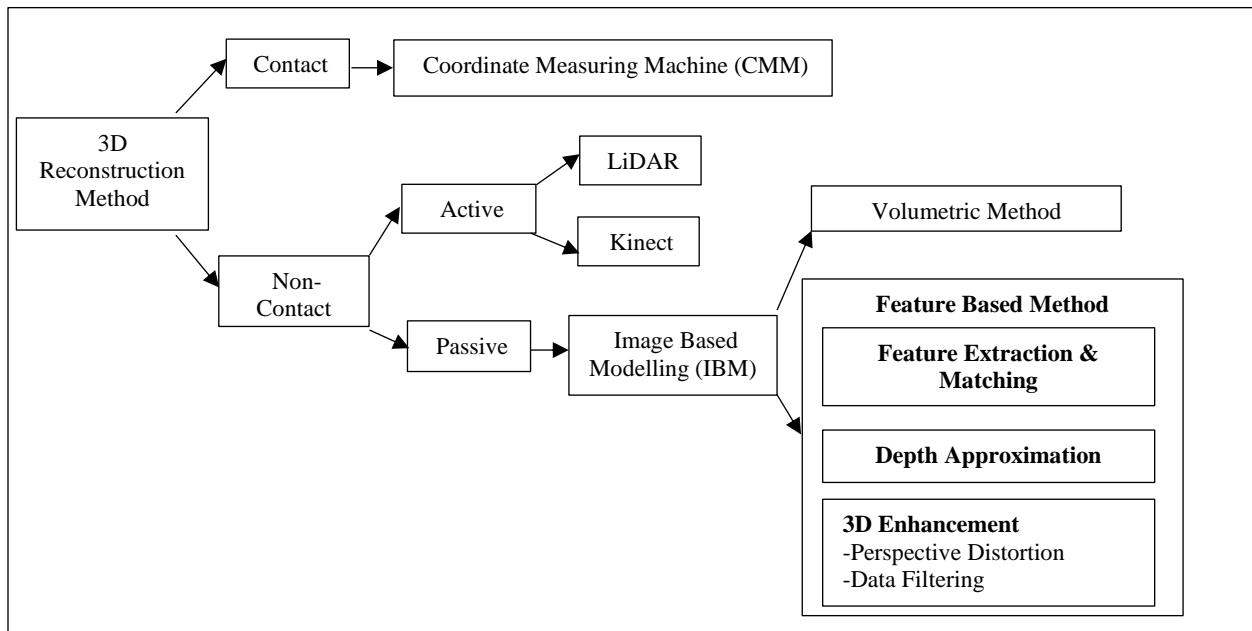

Figure 1: 3D Reconstruction Method

Feature Based Method

The feature-based methods involve feature extraction, matching, and merging of the feature points from wide range of complex objects (Tushar et al., 2017). The feature extraction process used the feature detectors such as Harris Corner Detector, Scale Invariant Feature Transform (SiFT), Good Features to Track, and Speeded Robust Features (SuRF) to select a set of points of interest from images (Mengxin et al., 2015; Tushar et al., 2017; Ze et al., 2022). The effectiveness of the methods is determined by the number of feature point extractions and the speed of the process, both of which influence the 3D point precision. Due to its resistance to changes in scale, translation, rotation, and lighting, the SiFT method has gained in popularity (Tareen et al., 2018; Escott et al.,2023). In addition, the SuRF method can be utilised to speed up the feature extraction. (Paul et al., 2019). Meanwhile, Good Feature to Track" made use of eigenvalues that are less likely to change if the image is translated or rotated to identify the features of interest (Shi & Tomasi., 1994)

Then, the extracted feature points were matched between the corresponding images. This step is very crucial as the accuracy of the 3D model is dependent on the correctness of the match. The feature points were matched using Fundamental Matrix (F-Matrix) which relies on the epipolar geometry and Zero-Mean Normalized Cross Relation (ZNCC) between corresponding points. Feature matching process becomes more challenging when it involves images captured by using non-parallel moving camera. Therefore, there is research utilising the optical flow method and brightness patterns to determine the correspondence between two corresponding features. The optical flow detects the displacement between the feature points in image sequence. The Lucas-Kanade (LK) method is the most popular and effective sparse method for optical flow estimation (Baker et al., 2011; Choi et al., 2022). Pyramidal LK is an enhanced version of LK that addresses the issue of large point displacements. The method begins tracking from the highest level of an image pyramid, which contains the least amount of detail, to the lowest level, which contains more detail. At the apex of the pyramid, the computation required significantly fewer unknowns and iterations. Moreover, the process of pyramid algorithms tends to be considerably faster and require less processing time (Wirza & Azmi, 2012; Dong, 2022). By using this approach, the same analogy applys in the real world, where the optical flow magnitude is greater when the object is closer to the observer than when the object is further away. Consequently, an optical flow feature match can be utilised to estimate the depth. In this perspective, the magnitude of the optical flow is inversely proportional to its distance.

Based on the feature match obtained from the corresponding images, the depth information can be approximated. However, the loss of depth information due to the projection of 3D world coordinates onto a 2D plane is the next challenge in 3D reconstruction from images. In general, the optimal triangulation method based on polynomial, linear, mid-point, L2, and L approaches is used to approximate the depth value (Seong & Civera, 2019; Yang et al., 2019; J. Chen, 2020). However, the previous methods always had issues with processing time and accuracy. Furthermore, complete 3D point reconstruction based on bundle adjustment require optimization, which still has accuracy and time consumption issues due to outliers (Alndiwee et al., 2022). Moreover, 3D improvement is a critical step in ensuring that the generated model is dependable and accurate enough to be used in a variety of applications. Therefore, elements such as perspective distortion and outliers are critical and must be considered because they affect the correctness of the resulting 3D model (Fan et al., 2018, Zhiliang et al., 2018).

Perspective Distortion and Noise Handling

Perspective distortion is a phenomenon of varying distances that happens when a three-dimensional subject is projected onto a flat plane, resulting in varying degrees of distortion. The difficulty lies in correcting

perspective distortion in real time, as digital cameras cannot extract depth information. In the real world, this phenomenon typically happens during 3D reconstruction of huge depth dimensions for surface faces, building facades, and scene reconstruction (Fan et al. 2018; Soycan, 2019; Yu-Fei,2020). Previous study employed projective transformation to reduce the perspective impact in images caused by central projection (Soycan, 2019). Moreover, Fan et al. (2018) improve the accuracy by transforming the perspective view of two images utilising the ground plane idea. Ondraovi et al. (2021) employed homography mapping to eliminate perspective distortion in photographs. Other researchers employ IPM to circumvent the issue of perspective distortion that could compromise precision (Just et al. 2020; Chungyup et al., 2019; Wongsaree et al., 2018).

Next, data filtering is necessary to eliminate outliers from false match feature points caused by variable optical flow magnitude, which impacts the precision of depth estimates. Previous research utilised RANSAC and ORSA to eliminate the erroneous match by utilising the epipolar geometry of the picture pair (Rodriguez-Gonzalvez et al., 2014; Zhiliang et al., 2018; Clément et al., 2022). Meanwhile, Mingwei et al. (2019) developed a geometry-constrain approach to eliminate the discovered outlier in the feature tracking method. Therefore, a new IBM approach for 3D reconstruction that is low in complexity, fast, and accurate is still in high demand for use in a variety of applications. Consequently, this study proposed an effective, fast, and accurate depth value approximation of 3D Image-Based Modelling (i3DM) based on Data Filtering Techniques and IPM. The proposed method is more accurate, reduces hardware dependence, has a low level of complexity, and is capable of 3D enhancement. The technique, results, discussion, and conclusion will all be covered in detail in the following section.

METHODOLOGY

The research framework consists of three main process including image acquisition, feature tracking, depth value approximation, and data enhancement as in Figure 2.
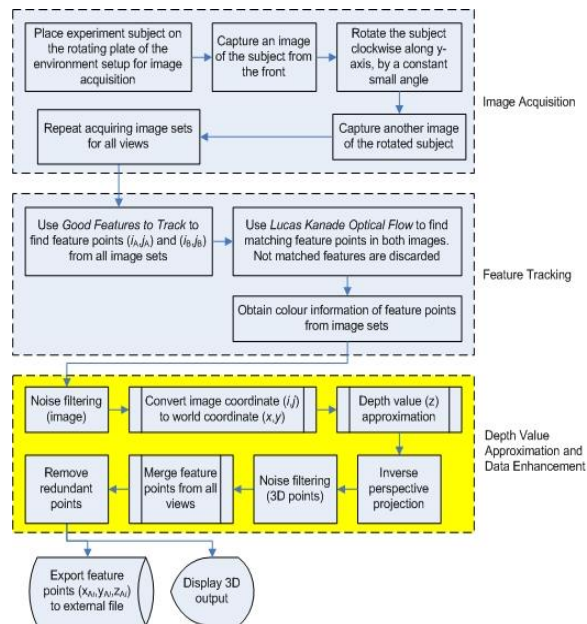

Figure 2: Research Framework

Image Acquisition

Firstly, the image acquisition setup is designed with a black background, black base, black rotating plate, and a webcam as in Figure 3. The black background was used to reduce the noises and simplify the pre-processing procedure. In this experiment, five different types of complex subjects which in irregular shape were captured using webcam as shown in Figure 4. The subject was place in the middle of the setup and rotated manually using the spinning plate. The black base marks with the object rotation angle. Each subject was captured three times to ensure the accuracy and consistency.
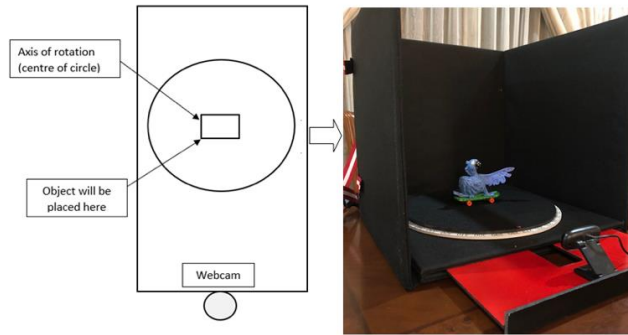

Figure 3: Sketch and the Experiment Setup

Subject	Image
1	
2	
3	
4	
5	

Figure 4: The subject used in the experiment

Feature Tracking

In this experiment, the first image captured without rotation and the second images captured after minor rotation angle. Next, the feature point from the first image A (i_A, j_A) and the feature point from the second image B (i_B, j_B), were detected using Good Features to Track. This detector is used to find the strongest corner in the images by using the Shi-Thomasi method. This method enhanced the existing Harris Corner Detection in which the corner region (R) was considered when λ_1 and λ_2 is above a minimum value as shown in Equation 1.

$$R = \min (\lambda_1, \lambda_2) \tag{1}$$

Next, the corresponding features were matched using Lucas Kanade Optical Flow method based on vector and magnitude displacement. This method is used as the optical flow represent the 2D vector where the displacement shows the movement of feature point from the first frame to the second frame. It estimates the 2D translation of sparse feature points from two consecutive frames. This method assumes the neighbouring pixels have similar motions and constant brightness between the image intervals. Its normally work with small motion of images. Therefore, the pyramidal method improves the performance and robustness of the tracking process for larger translation. It calculates the optical flow at the top of images and integrate the result to the next level of pyramid until the original size of image was reached. Equation 2 shows the similarity checking of inverse the inverse matrix with the corner detector. In the equation, (u,v) is the displacement of x and y over time. Meanwhile, f_x and f_y are image gradients.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum_i f_{x_i}^2 & \sum_i f_{x_i} f_{y_i} \\ \sum_i f_{x_i} f_{y_i} & \sum_i f_{y_i}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i f_{x_i} f_{t_i} \\ -\sum_i f_{y_i} f_{t_i} \end{bmatrix} \tag{2}$$

At this phase, the output is the feature points matched between two consecutives imaged. The feature point match is used to obtain the accurate depth value estimation in the next process.

Depth Value Approximation

Next, the depth value was approximated based on the magnitude displacement and the trigonometry relationship between the feature points. In this experiment, the image acquired in a controlled environment where the extrinsic parameter such as distance and motion are predetermined in the previous step. Firstly, the feature point match obtained from the previous step must be converted into cartesian coordinates. The assumption is the camera being placed directly in front of the subject with no pitch angle. In this circumstance, the 2D Cartesian origin and camera lens origin are parallel. Consequently, the centre of the image's width becomes the origin of the vertical axis. Meanwhile, the origin of the horizontal axis is represented by the height of the camera lens. Consequently, the conversion of image coordinates (i,j) to Cartesian coordinates (x,y) using Equation 3 as follows:

$$(x, y) = (i - (\frac{imgwidth}{2}), (imgheight - j) - \min(j) - lensheight) \tag{3}$$

The formula translates the x-origin from the left side to the centre of the image using $i - (\frac{imgwidth}{2})$. Meanwhile, the difference between the cartesian coordinate and imaging system was determined using

(*imgheight – j*). Next, the subject's horizontal basis is referred to as min (*j*). Therefore, it is assumed that the camera lens is perpendicular to the horizontal plane. Besides, the lensheight is measured to ensure that the object's cartesian origin (0,0) is parallel to the origin of optical center.

Next, the experiment was configured with y-axis rotation of the second image. In this situation, the height of the subject remained consistent for all rotation degrees. Therefore, the same scaling factor is applied to both the width and height dimensions. This is to maintain the aspect ratio when using the actual and detected height, as shown in Equation (4). The world 2D coordinates were then determined using Equation (5), where $(x_p, y_p)_i$ is the image coordinate for point *i* and $(x_w, y_w)_i$ is the world coordinate for point *i*.

$$\text{Scaling Factor, } k = \frac{objheight}{detected\ height} \tag{4}$$

where *detectedheight* = max(*j*) - min(*j*)

$$(x_w, y_w)_i = (kx_p, ky_p)_i \tag{5}$$

Next, the process proceeds with the depth approximation value. z_w refers to the distance of the current feature point from the centre of rotation (COR) in the experimental setup. Meanwhile, $objdist - z_W$ refers to the z-distance between the feature point and the projection centre (COP). Therefore, the point (X_{w1}, Y_{w1}, Z_{w1}) and (X_{w2}, Y_{w2}, Z_{w2}) correspond in the world coordinate system. Meanwhile, the points (X_{p1}, Y_{p1}, Z_{p1}) and (X_{p2}, Y_{p2}, Z_{p2}) detected in the projection plane.
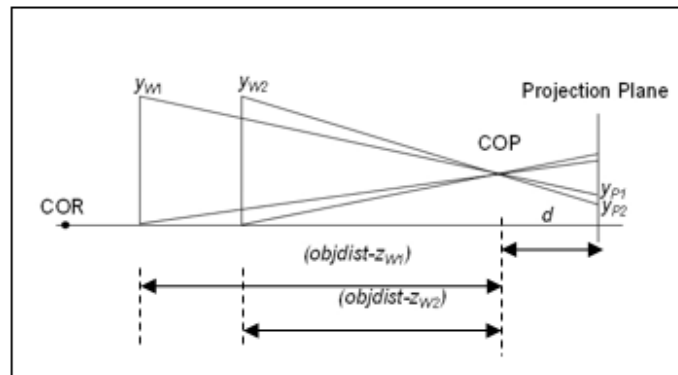

Figure 5: Drawing of the experiment's setup from the left.

Therefore, the depth approximation equation is obtained as in Equation (6) using various triangular ratios and the simultaneous equation.

$$z_{W1} = objdist \left(\frac{x_{P1} y_{P2} \cos\alpha - x_{P2} y_{P1}}{y_{P2} objdist(\sin\alpha) + x_{P1} y_{P2} \cos\alpha - x_{P2} y_{P1}} \right) \tag{6}$$

$$= \frac{objdist(x_{P1} y_{P2} \cos\alpha - x_{P2} y_{P1})}{y_{P2} objdist(\sin\alpha) + x_{P1} y_{P2} \cos\alpha - x_{P2} y_{P1}}$$

Next, the following step is to fully construct a 3D point using the merging technique for all images. The proposed method is simplified compared to previous method that consider unrelated image views. Meanwhile, in this study, the rotation angle between image views were known. Therefore the, the merging technique become simpler which used the inverse y-axis rotation matrix Ry(θ) as in Equation (7) to combine the 3D points between image sets. Figure 6 displays the fully merged 3D points for Subject 1.

$$R_y(-\vartheta) = \begin{pmatrix} \cos(-\vartheta) & 0 & \sin(-\vartheta) \\ 0 & 1 & 0 \\ -\sin(-\vartheta) & 0 & \cos(-\vartheta) \end{pmatrix} \tag{7}$$

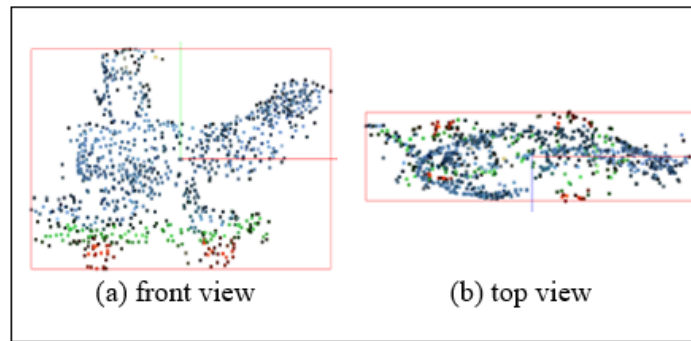where θ is the yaw orientation of the subject.

Figure 6: Complete 3D point after merging

Data Enhancement with Noise Filtering and Inverse Perspective (IPM) Method

The noise and perspective distortion may influence the accuracy of the depth measurement. Therefore, the noise filtering and inverse perspective distortion technique are required as shown in Figure 7. Noise filtering is the process of removing outliers from the false matches and after depth approximation process. Meanwhile, inverse perspective distortion is used to correct the perspective distortion effect when the depth approximation does not precisely portray the real item.
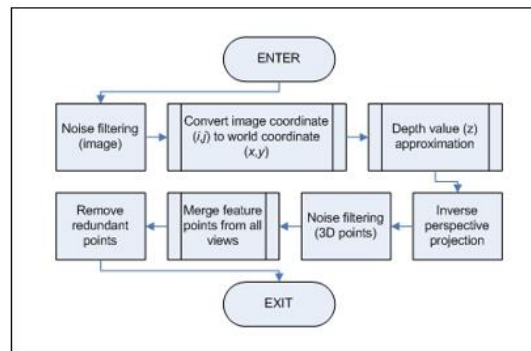

Figure 7: Processing Steps in Depth Approximation

Feature false matches happened when the second image during the y-axis rotation is hidden and cause the feature points have no pair. Based on Figure 8, the entire front region facing the camera was considered as the maximum viewing area. Meanwhile, the wave pattern represents the largest hemisphere viewable region prior to the rotation of α° in y clockwise. In addition, the stripes pattern is represented by the largest hemisphere that can be seen after rotation. The area with the wave pattern exclusively in the first hemisphere, nevertheless, does not cross into the second hemisphere. Thus, the feature points found in this area should be filtered out as noise.
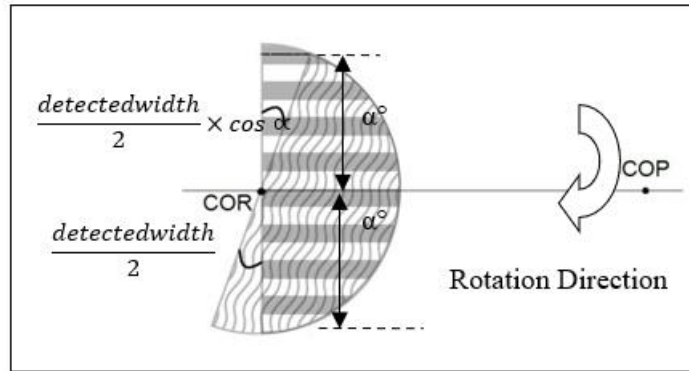

Figure 8: The viewable region before and after the y-axis rotation

Next, the process of matching features began with the extraction of feature points from the first image, followed by the search for matches in the second image. Therefore, the filtering procedure relies on the detected feature points in the initial image. Figure 9 depicts the zone of noise that must be filtered out. Based on Figure 9, the second image with wave pattern only sector has no feature match. The x- value of feature points was chosen for the filtering process which represents the viewable width. Equation (8) shows the maximum viewable width with possible feature points matched in the first image with $width_A$.
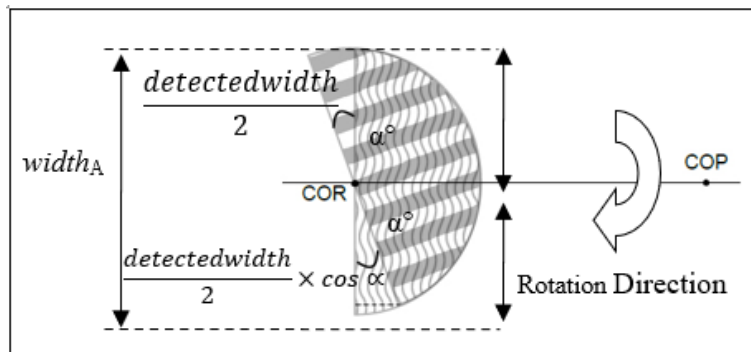

Figure 9: Noise Sensitive Region to be Filtered

$$width_A = \frac{detectedwidth}{2} + \left(\frac{detectedwidth}{2} \times cos\,\alpha\right)$$

$$= \frac{detectedwidth}{2}(1 + cos\,\alpha) \tag{8}$$

Therefore, the noise caused by a false match that could affect the accuracy of the constructed 3D model could be removed. Figure 10 depicts the subject's detected 2D feature points before and after the implementation of the filter.
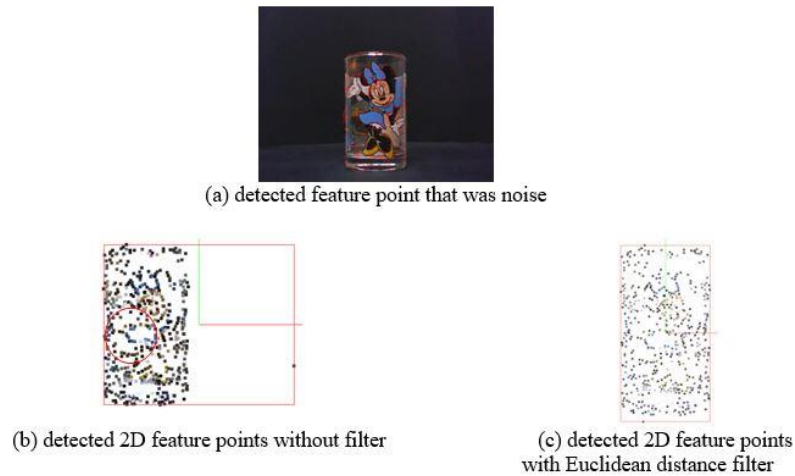

(a) detected feature point that was noise

(b) detected 2D feature points without filter

(c) detected 2D feature points with Euclidean distance filter

Figure 10: The identified 2D feature points of the subject prior to and during application of the filter

The previous phase of filtering aims to eliminate the feature points that lack a pair in the second image. However, the LK Optical flow can still generate the false matches that may impact the approximation of the depth value due to redundant pattern on the objects or parallel displacement direction. Therefore, the depth value approximation was examined which may contain the extreme values deemed noise. Based on the likelihood of noise distribution in this experiment, the sliding window filter technique was introduced as shown in Figure 11. Firstly, the image was divided into two parts. The blue rectangle represents the subject and the points detected in the centre of the image had a greater priority for representing the subject because the subject was positioned in the centre of the camera field of view. Therefore, the variable distance was considered for three successive points and labelled "case 1," "case 2," and "case 3."

In this section, we discussed the left side of the feature points. Initially, the detected feature points were sorted from smallest to largest, and feature points with Euclidean distances greater than the threshold value were identified. The threshold value was established based on the picture dimensions with 1% value. For instance, the x-axis threshold value is 6.4 pixels, which is generated from 640 x 0.01. In the meanwhile, the y or z axis consists of 4.8 pixels, which is generated from 480 x 0.001 pixels. The detection was based on a sliding window scan from the image's left edge to its centre.
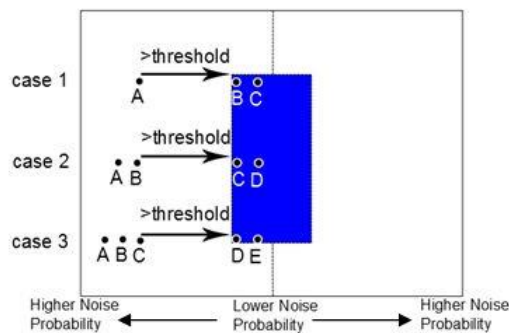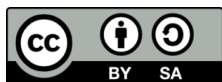

Figure 11: Probability Distribution of Noise

Based Figure 11, feature point A has a Euclidean distance with the next point that exceeds the threshold value and is therefore considered noise in Case 1. In Case 2, feature points A and B were judged to be part of the same cluster because they both surpassed the threshold value. Even though the three locations were near together, case 3 was subject to the same circumstance. To detect noise in the right side of the image, the same sliding technique was utilised. Figure 12 depicts a top view of the 3D point features with and without the filter.
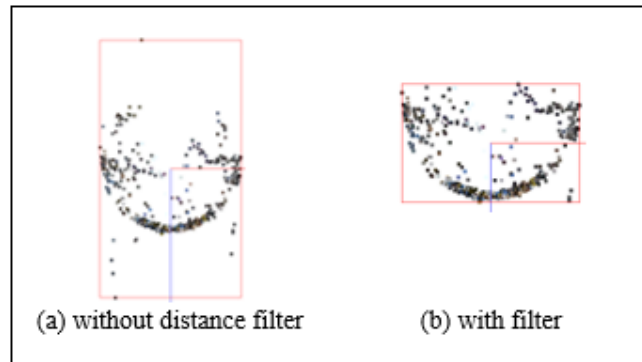

(a) without distance filter   (b) with filter

Figure 12: Comparison of the filter's effect on 3D feature points

As discussed in the previous section, perspective distortion can have an impact on the precision of depth value approximation, and this effect is especially noticeable for subjects with large depth dimensions as in Subject 1. Figure 13 depicts the result of the projective projection. The relationship between subject 1's head and left wing is indicated by the upper red line. Meanwhile, the bottom red line depicts the subject's foundation.

(a) 0°   (b) 60°   (e) 180°   (f) 240°

(c) 90°   (d) 120°   (g): 270°   (h): 300°
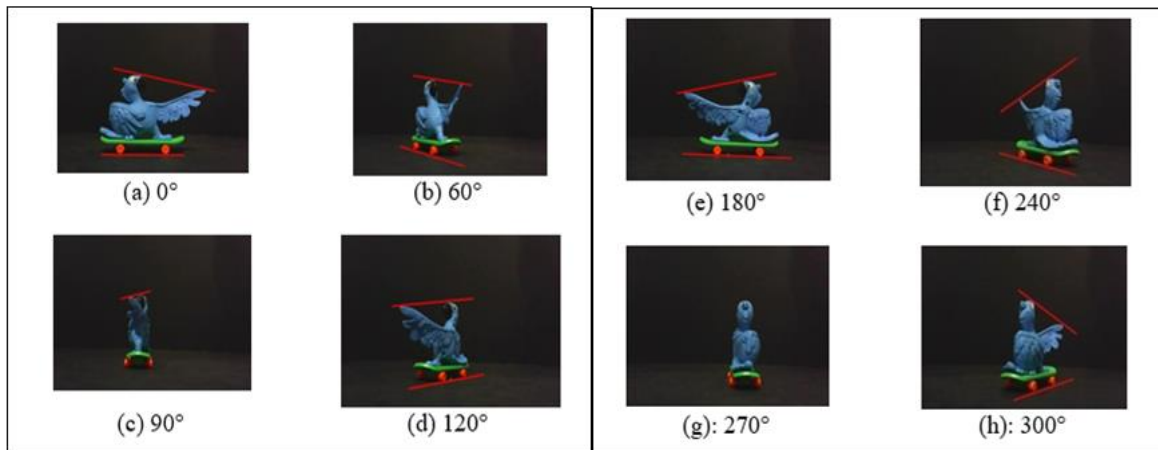
Figure 13: Effect of Perspective Projection on Subject 1

Based on Figure 13, subject 1's head was much higher than the wing at 0° and 180° perspectives. The bottom line is currently horizontal. However, when the subject was rotated from 0° to 90° anticlockwise, the effect of perspective distortion was visible. This indicates when the top and bottom red lines of the subject become nearly parallel. Because the slope of the red lines was inconsistent, the impact was also visible in photos taken at 120°, 240°, and 300°. Figure 14 contrasts the effect of a 3D point obtained after

rotating the image by 120° with the effect of a 3D point generated when the image was not affected by perspective distortion at 0°.

(a) initial 0° 3D points   (b) rotated 120° 3D points

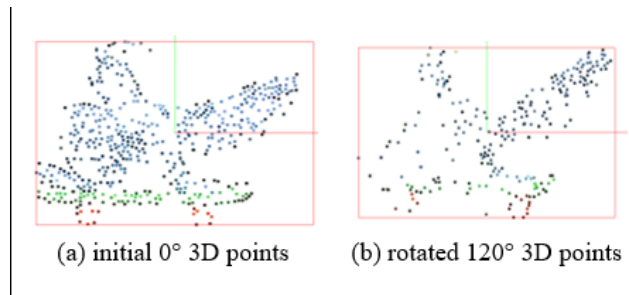**Figure 14:** Initial 0^0 and Rotated 120^0 Approximated 3D Points

As a result, the IPM technique is critical for correcting the projected image coordinate to the actual world coordinate. This information defines the distance between the point and the centre of rotation (COR), as shown in Figure 15. The trigonometry equation is used to establish a relationship between the projected image and the actual world. The projected image distance was converted to the same magnitude as the subject's distance from COP. As a result, the projected image on the left side overlapped the subject.
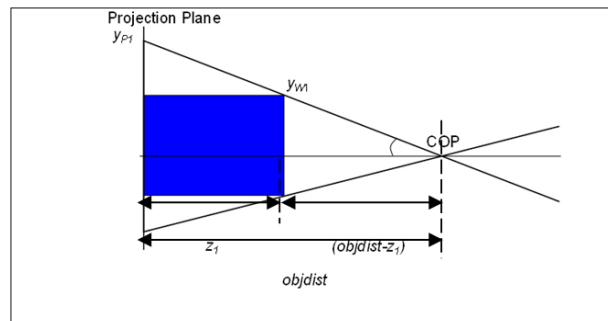
Figure 15: Inverse Perspective Mapping

Therefore, Let y_{W1} be the real height and the height projected onto the projection plane. Following that, z_1 expressed as the distance between the feature point and the projection plane, and β, the pitch angle, were calculated. This information was used to determine the IPM triangulation ratio by:

$$tan\beta = \frac{y_{W1}}{(objdist - z_1)} = \frac{y_{P1}}{objdist}$$

$$y_{W1} = \frac{y_{P1}(objdist - z_1)}{objdist}$$

$$= y_{P1}\left(1 - \frac{z_1}{objdist}\right) \qquad (7)$$

Next, the x-axis IPM can be generated

$$x_{W1} = x_{P1}\left(1 - \frac{z_1}{objdist}\right) \qquad (8)$$

Therefore, perspective distortion for image coordinates (x_{P1}, y_{P1}) can be adjusted using the following formula:

$$(x_{W1}, y_{W1}) = (x_{P1} \left(1 - \frac{z_1}{objdist}\right), y_{P1} \left(1 - \frac{z_1}{objdist}\right)) \tag{9}$$

Table 1 compares the approximate wingtip height of Subject 1 before and after IPM implementation. After implementing IPM, the difference between the detected and real measurement has decreased in average of 4.83mm.

Table 1: The effect of the approximate wingtip height of subject 1 before and after IPM implementation

Experiment	Points	Approximated (mm)	Actual (mm)	Difference (mm)
1	553	61.57	55.38	6.19
2	542	59.26	55.38	3.88
3	441	59.80	55.38	4.42
Average		**60.21**	**55.38**	**4.83**

FINDINGS AND DISCUSSIONS

In this research, the noise filtering for 2D and 3D feature points, as well as IPM, are used to improve the depth approximation method. As a result, a prototype system for automatically generating the 3D model was built to evaluate the intended procedure. The evaluation's goal is to determine the effectiveness of the proposed method based on processing time, bounding box, and the difference between approximated and actual dimensions. Therefore, the complexity of the algorithm was determined by measuring the processing time. The simpler an algorithm is, the less processing time it requires. The experiment used six sets of images with a 3° rotational angle. The computer system automatically measured the processing time in seconds. Figure 16 depicts the desktop computer specifications used to run the system.

Type:	Desktop Computer
Operating system:	Microsoft Windows 7 Professional
Processor:	Intel ® Core ™ i5-3470 CPU @3.20GHz 3.20GHz
Memory (RAM):	8GB
System type:	64-bit Operating System
Graphic Card:	AMD Radeon HD7470
Solution's Platform:	Microsoft Visual C++, OpenCV, OpenGL
Webcam:	Logitech C170, 5MP

Figure 16: Specification of desktop computer

The processing time in this experiment was associated to the number of 3D feature points produced. As shown in Table 2, each subject was repeated three times to determine the average time and processing capacity in terms of the number of 3D points constructed. This is to ensure the consistency of the average processing capacity.

Table 2 : Processing Time

Sub	Exp. #	Points	Ave. Points	Time (s)	Ave. Time (s)	Processing Capacity (points/s)	Average Processing Capacity (points/s)
1	1	2092	2002	1.31	1.257	1596.95	1592.68
	2	1969		1.27		1592.68	
	3	1945		1.19		1634.45	
2	1	1337	1386	0.83	0.837	1572.94	1655.91
	2	1400		0.83		1686.75	
	3	1421		0.85		1671.76	
3	1	544	499	0.43	0.407	1265.12	1226.04
	2	519		0.42		1235.71	
	3	433		0.37		1139.47	
4	1	1128	1177	0.75	0.777	1464.94	1514.80
	2	1235		0.80		1543.75	
	3	1168		0.78		1497.44	
5	1	3090	3032	2.00	1.920	1545.00	1579.17
	2	2926		1.83		1598.91	
	3	3080		1.93		1595.85	
Average							**1513.72**

The result indicates that it took one second for every 1514 feature points to complete the process. Meanwhile, Subject 5 required 1.92 seconds to analyse six images which consider the longest processing time. This average processing time is corelated with the highest number of feature points generated for Subject 5. As a result, each view took approximately 0.32 seconds to process. The graph in Figure 17 depicts the relationship between the number of actual and average feature points and the processing time in seconds (s). In overall, the graphs show that the processing time increased linearly with the number of feature points.
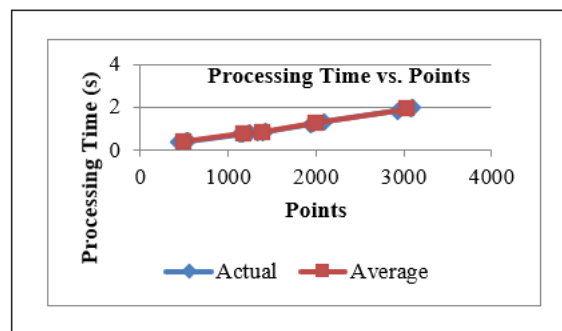

Figure 17: Processing Time Based on Detected Feature Points

Next, the correctness of the constructed 3D models was then assessed using a bounding box and a representation of the actual object's shape. The bounding box compares the size of the constructed 3D model to the actual dimensions of the original object. The difference value was then calculated by subtracting the

actual measurement from the detected dimension. The accuracy increases as the disparity decreases. The approximation of height, width, and depth obtained using the proposed methods is shown in Table 3.

Table 3: Result of height, width, and depth approximation

Sub	Height (mm)			Width (mm)			Depth (mm)		
	App. Avg	Actual	Diff	App. Avg	Actual	Diff	Approx. Avg	Actual	Diff
1	94.90	94.90	0.0	138.23	134.03	4.20	37.05	35.12	1.93
2	49.10	49.10	0.0	102.20	95.41	6.79	46.90	40.70	6.20
3	109.19	109.19	0.0	65.89	62.80	3.09	41.08	55.38	-14.30
4	72.94	72.94	0.0	73.80	68.63	5.17	63.26	65.04	-1.78
5	136.39	136.39	0.0	66.31	66.62	-0.31	73.32	66.33	6.99

According to Table 3, there was no difference between the real and detected height values because the scaling factor in this proposed method used the actual height value. Subjects 2 and 4 have a wider variation in width, measuring roughly 6.79 mm and 5.17 mm, respectively. This is because the subjects' structure, which includes a lot of smooth, simple surfaces with little texture, caused the false matched error and affected the dimension approximation. Furthermore, despite being entirely blue, Subject 1 had a width error of approximately 4.2mm when compared to earlier subjects with larger errors. This was due to the texture and a variety of properties that could be tracked from the subject. As a result, the feature points were correctly matched.

Subject 5's width estimation, on the other hand, was extremely accurate. This is due to the subject contains unique and colourful elements that can be tracked. The width dimension difference for Subject 3 was approximately 3.09mm, while the depth dimension difference was 14.30mm. This happen due to the fact that some regions of the object had a low number of detected feature points and were parallel to the viewing direction. As a result, the features had been mismatched. Therefore, the depth dimension filter has been disabled in this case, as shown in Table 4. The outcome shows a positive change, indicating that Subject 3's area was not cut off. As a result, the average inaccuracy dropped from -14.30mm to 4.83mm.

Table 4: Difference of Approximated Depth with the Actual Measurement of Subject 3

Experiment	Points	Approximated (mm)	Actual (mm)	Difference (mm)
1	553	61.57	55.38	6.19
2	542	59.26	55.38	3.88
3	441	59.80	55.38	4.42
Average		**60.21**	**55.38**	**4.83**

Next, the generated 3D point cloud's visualisation appearance is then compared to that of the actual model. In this experiment, only half of the 3D feature points were superimposed on the front view (0°) image to reduce the confusion caused by 3D feature points projected on 2D images. The result, as shown in Figure 18, shows an approximated 3D feature point in red that can represent the actual object. However, it also demonstrates that Subject 3 had very few feature points discovered. The feature point on Subject 3's fists has been removed because of the noise filtering approach. Furthermore, this area was perpendicular to the viewing direction, which could make determining the depth value difficult. Furthermore, the contrast between the upper and lower portions of the arm was quite low, contributing to the low number of visible feature points.
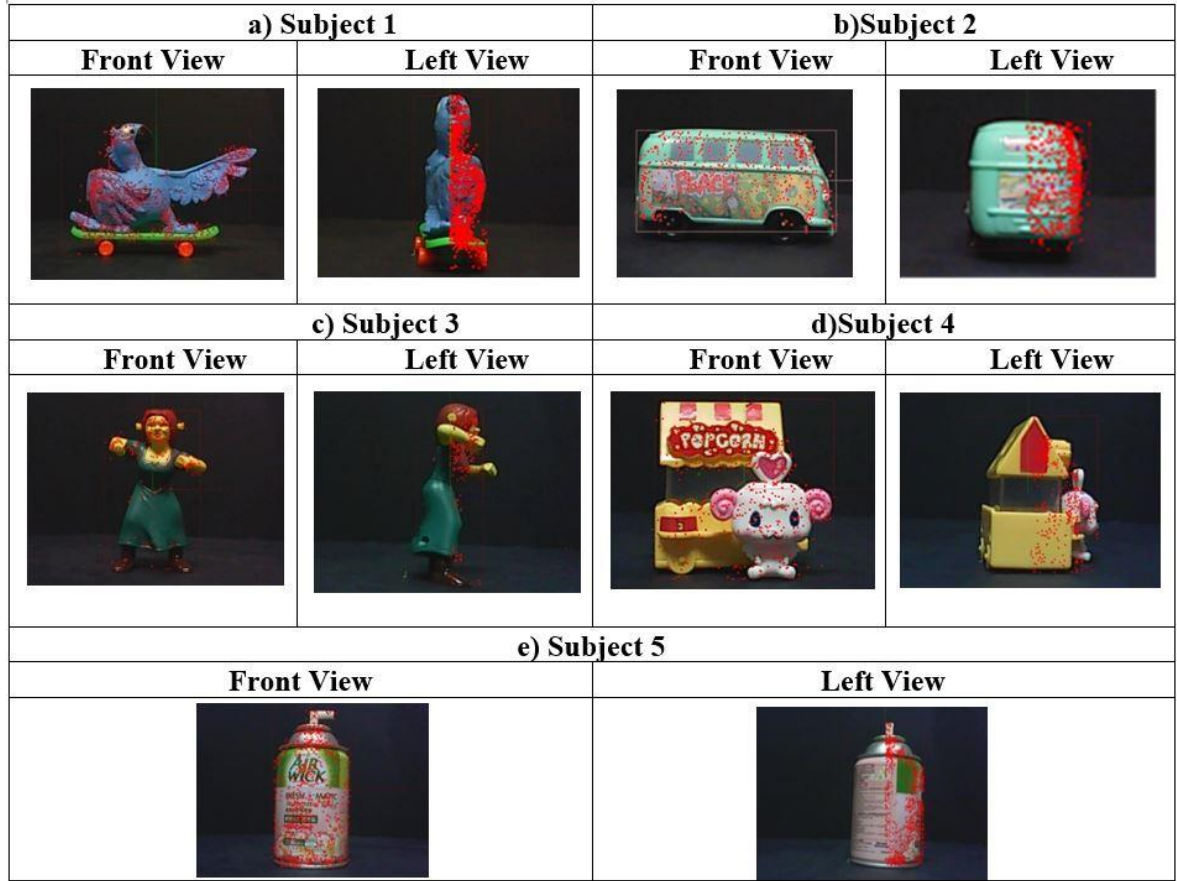
Figure 18: Detected Feature Points of five subjects being Superimposed on the Image

CONCLUSION AND RECOMMENDATIONS

In conclusion, the proposed method manages to enhance the depth approximation using noise filtering and IPM for complex subject. This approach also able to construct the 3D points using only four photos which contribute to improve the average of processing capacity to around 1,519 per second. Meanwhile, the longest processing time required to generate the 3D points for the subject is around 1.92 seconds. The IPM and data filtering procedures are important for enhancing the precision of 3D reconstruction points. In this study, the accuracy has been calculated based on the actual and detected dimension of width and depth of the object. As a result, the average width dimension error was around 3.27 percent of the actual width. The average error in depth dimension was 6.88 percent of the actual depth. In future, the accuracy of the proposed method will be compared with the existing methods to ensure the accuracy. Overall, this method also can simulate transparent objects like glass compared to other laser scanner that is sensitive to light reflections and refractions. This approach is also inexpensive and work in bright light compared to Kinect. Consequently, this method employs a simple algorithm to build complex-shaped tabletop items. It also reduces processing time and yields accurate results. However, there are still challenges to overcome, as the image acquisition setup utilises a rotating table, which resulted in a relative increase in human error during manual rotations. Furthermore, the bigger rotation degrees can result in ambiguity while monitoring the relevant feature point. In addition, this method has difficulties with textureless objects that lack distinguishing characteristics. Therefore, future study must improve the image acquisition, calibration technique and deal with the subject with less texture.

REFERENCES

Adikari, S. B., Ganegoda, N. C., Meegama, R. G., & Wanniarachchi, I. L. (2020). Applicability of a single depth sensor in real-time 3D clothes simulation: augmented reality virtual dressing room using Kinect sensor. *Advances in Human-Computer Interaction*. https://doi.org/10.1155/2020/1314598

Agudo, A. (2022). Safari from Visual Signals: Recovering Volumetric 3d Shapes, ICASSP 2022 IEEE International Conference on Acoustics, *Speech and Signal Processing (ICASSP)*, pp. 2495-2499.

Awange, J., & Kiema, J. (2019) Light Detection and Ranging (LiDAR). *In: Environmental Geoinformatics. Environmental Science and Engineering. Springer, Cham*. https://doi-org.ezaccess.library.uitm.edu.my/10.1007/978-3-030-03017-9_21

Baker, S., Scharstein, D., Lewis, J., Roth, S., Black, M., & Szeliski, R. (2011). A database and evaluation methodology for optical flow. *A Journal of Computer Vision*, 92(1), 1-31. https://doi:10.1007/s11263-010-0390-2

Cao, M., Jia, W., Lv, Z., Li, Y., Xie, W., Zheng, L., & Liu, X. (2019). Fast and robust feature tracking for 3D reconstruction. *Optics & Laser Technology*, 110, 120-128. https://doi.org/10.1016/j.optlastec.2018.05.036

Chen, S., Liang, L., & Ouyang, J. (2022). Accurate structure from motion using consistent cluster merging. *Multimed Tools Appl 81*, 24913–24935. https://doi.org/10.1007/s11042-022-12202.

Cho, J., Jung, Y., Kim, D.-S., Lee, S., & Jung, Y. (2019). Moving Object Detection Based on Optical Flow Estimation and a Gaussian Mixture Model for Advanced Driver Assistance Systems. *Sensors*, *19*(14), 3217. https://dx.doi.org/10.3390/s19143217

Choi, H., Kang, B., & Kim, D.(2022). Moving Object Tracking Based on Sparse Optical Flow with Moving Window and Target Estimator. *Sensors*.22(8):2878. https://doi.org/10.3390/s22082878

Chungyup, L., Soohyeon, C., Jung, Y., & Kwanghee, W.(2019). Instance segmentation in urban scenes using inverse perspective mapping of 3D point clouds and 2D images.*RACS '19: Proceedings of the Conference on Research in Adaptive and Convergent Systems*. pp.147–152 https://doi.org/10.1145/3338840.3355677

Clément, R., Vincent, N., & Pascal M. (2022). Automatic RANSAC by Likelihood Maximization, *Image Processing On Line*, (12),pp. 27-49.
https://doi.org/10.5201/ipol.2022.357

Crisnapati, P. N., Setiawan, M., Wikranta Arsa, I. G. N., Devi Novayanti, P., Wibawa, M. S., & Oka Ciptahadi, K. G.(2019). Real-Time Hand Palm Detection and Tracking Augmented Reality Game Using Lucas Kanade Optical Flow Combined with Color Blob Detection. *2019 1st International Conference on Cybernetics and Intelligent System (ICORIS)*, 263-268, https://doi:%2010.1109/ICORIS.2019.8874892

Davide, M., Simone, B., & Gianluigi, C. (2022). SfM Flow: A comprehensive toolset for the evaluation of 3D reconstruction pipelines, SoftwareX, (17),100931, ISSN 2352-7110. https://doi.org/10.1016/j.softx.2021.100931

Dhal, K., Karmokar, P., & Chakravarthy, A. et al. (2022). Vision-Based Guidance for Tracking Multiple Dynamic Objects. *J Intell Robot Syst 105*, 66 https://doi.org/10.1007/s10846-022-01657-6

Ding, T., Yuan, J., Lin, X., Zhang, N., Zhang, Y., & Gao, X. (2021). Three Dimensions Reconstruction of Single-spectrum Multi-X-ray Views of Contraband Based on Space Carving Method. *In Journal of Physics: Conference Series* 1986(1), p.p 012129. https://doi.org/10.1088/1742-6596/1986/1/012129

Domen, G., Janko S., Aleš B.,& Miha B.(2021). Still-camera multiview Spectral Optical Flow Imaging for 3D operating-deflection-shape identification. *Mechanical Systems and Signal Processing*,152,107456, ISSN 0888-3270. https://doi.org/10.1016/j.ymssp.2020.107456

Dong,Y. (2022). Faint Moving Small Target Detection based on Optical Flow Method, 7th *International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an*, pp. 391-395.

Elkhrachy, I. (2022). 3D Structure from 2D Dimensional Images Using Structure from Motion Algorithms. *Sustainability*,14(9):5399. https://doi.org/10.3390/su14095399

Escott, K., Shaw, R., & Lensen, A. (2023). Feature-based Image Matching for Identifying Individual K\=ak\=a. *ArXiv*. https://doi.org/10.48550/arXiv.2301.06678

Fan, R., Ai, X., & Dahnoun, N. (2018). Road surface 3D reconstruction based on dense subpixel disparity map estimation. *IEEE Transactions on Image Processing*, 27(6), 3025-3035. EC Accession Number: 17665641. https://doi.org.10.1109/TIP.2018.2808770

Justs, D. J., Novickis, R., Ozols, K., & Greitāns, M. (2020). Bird's-eye view image acquisition from simulated scenes using geometric inverse perspective mapping. *In 2020 17th Biennial Baltic Electronics Conference (BEC)* (pp. 1-6). https://doi.org/10.1109/BEC49624.2020.9277042

Kang, Z., Yang, J., & Yang, Z., & Cheng, S.(2020). A Review of Techniques for 3D Reconstruction of Indoor Environments. *ISPRS International Journal of Geo-Information*, 9(5),330. https://doi.org/10.3390/ijgi9050330

Li, M., Zheng, D., Zhang, R., Yin, J., & Tian, X. (2015). Overview of 3d reconstruction methods based on multi-view. *In 2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics* (2), pp. 145-148. https://doi.org/10.1109/IHMSC.2015.117

Menna, F., Nocerino, E., Morabito, D., Farella, E. M., Perini, M., & Remondino, F. (2017). An open-source low-cost automatic system for image-based 3D digitization. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 155. https://doi.org/10.5194/isprs-archives-XLII-2-W8-155-2017

Miller, A., Miller, B., Popov, A., & Stepanyan, K. (2019). UAV landing based on the optical flow video navigation. Sensors, 19(6), 1351.http://dx.doi.org/10.3390/s19061351

Nadour, M., Boumehraz, M., Cherroun, L., & Puig, V. (2019). Mobile robot visual navigation based on fuzzy logic and optical flow approach. *International Journal of System Assurance Engineering and Management*, 10(6), 1654-1667. https://doi.org/10.1007/s13198-019-00918-2

Ondrašovič, M., & Tarábek, P. (2021). Homography Ranking Based on Multiple Groups of Point Correspondences. Sensors, 21(17), 5752. https://dx.doi.org/10.3390/s21175752

Paul, M., Karsh, R. K., & Talukdar, F. A. (2019). Image hashing based on shape context and speeded up robust features (SURF).*International Conference on Automation, Computational and Technology Management (ICACTM)* (pp. 464-468). https://doi.org.10.1109/ICACTM.2019.8776713

Rodriguez-Gonzalvez, P., Gonzalez-Aguilera, D., Lopez-Jimenez, G., & Picon-Cabrera, I. (2014). Image-based modelling of the built environment from an unmanned aerial system. *Automation in Construction*, 48, 44-52. https://doi.org/10.1016/j.autcon.2014.08.010

Shalma, H., & Selvaraj, P. (2021). A review on 3D image reconstruction on specific and generic objects. *Materials Today: Proceedings*. https://doi.org/10.1016/j.matpr.2021.06.371

Shi, J., & Tomasi, C. (1994). Good features to track. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. 593-600). https://doi.org/10.1109/CVPR.1994.323794.

Stathopoulou, E. K., Welponer, M., & Remondino, F. (2019). Open-source image-based 3D reconstruction pipelines: Review, comparison and evaluation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLII-2/W17*, 331-338.

Tan, P. (2020) Image-Based Modeling. In: Ikeuchi K. (eds) Computer Vision. Springer, Cham. https://doi.org/10.1007/978-3-030-03243-2_11-1

Tareen, S. A. K., & Saleem Z., (2018). A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK, *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, pp. 1-10, doi: 10.1109/ICOMET.2018.8346440.

Tushar J., Kulbir S., Aditya A. (2017) A review and comparison of multi-view 3D reconstruction methods. *Journal of Engineering Research, Vol(5)(3).* https://kuwaitjournals.org/jer/index.php/JER/article/view/2307

Wang,X., Rottensteiner,F., & Heipke, C. (2019). Structure from motion for ordered and unordered image sets based on random k-d forests and global pose estimation, *ISPRS Journal of Photogrammetry and Remote Sensing, Volume 147*, Pages 19-41, ISSN 0924-2716, https://doi.org/10.1016/j.isprsjprs.2018.11.009.

Wirza, R., & Azmi, S. (2012). Depth Value Deduction for 3D Reconstruction using Optical Flow to Reverse Engineered a Geometrical Shape. *In Workshop On Advanced Information Technology (WIT-A2012)* (p.6).

Wongsaree, P., Sinchai, S., Wardkein P., & Koseeyaporn, J., (2018). Distance Detection Technique Using Enhancing Inverse Perspective Mapping, *3rd International Conference on Computer and Communication Systems (ICCCS)*, pp. 217-221, https://doi.org/10.1109/CCOMS.2018.8463318.

Liu, Y-F., Nie, X., Fan, J-S., Liu, X-G.(2020). Image-based crack assessment of bridge piers using unmanned aerial vehicles and three-dimensional scene reconstruction. Comput Aided Civ Inf.; 35: 511– 529. https://doi.org/10.1111/mice.12501

Ze, L.,Yong, Q., Hui,W., Xiaoli, Z., Gehao, S., & Xiuchen, J.(2022). A novel image-orientation feature extraction method for partial discharges.*IET Generation, Transmission & Distribution*,16(6),1139-1150.https://doi.org/10.1049/gtd2.12356

Zhiliang, M., & Shilong L., (2018). A review of 3D reconstruction techniques in civil engineering and their applications, *Advanced Engineering Informatics*,37,163-174, ISSN 1474-0346. https://doi.org/10.1016/j.aei.2018.05.005.

Zhou, Z.X. Gong, J. Guo, M.Y. (2016) Image-based 3D reconstruction for post-hurricane residential building damage assessment, *J. Comput. Civil Eng*. 30 (2). https://doi.org/10.1061/(ASCE)CP.1943-5487.0000480