J. Integr. Sci. Technol. 2025, 13(5), 1120

Journal of Integrated
SCIENCE & TECHNOLOGY

A scalable structure-from-motion framework for efficient 2D-to-3D reconstruction of historical artifacts

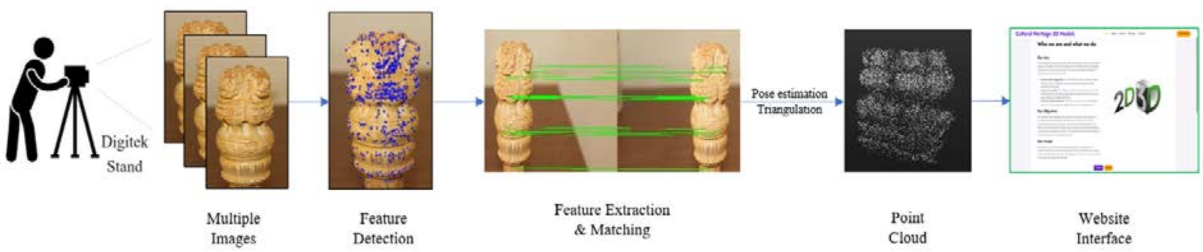
Swati Shilaskar,[1*] Shripad Bhatlawande,[1] Sourjadip Pramanik,[1] Anshul Surana,[1] Jyoti Madake,[1] Anjali Solanke[2]

[1]Vishwakarma Institute of Technology, Pune, India. [6]Marathwada Mitra Mandal College of Engineering, Pune, India.

Submitted on: 23-Nov-2024, Accepted on: 09-Feb-2025, and Published on: 28-Mar-2025

Article

ABSTRACT



Virtual museums are becoming increasingly popular, offering accessibility and engagement with historical artifacts for wider audiences. However, creating high-quality 3D models of artifacts can be time-consuming and expensive. Addressing these issues, we propose a comprehensive SFM based 2D-to-3D reconstruction method that enhances accessibility and scalability. Our approach integrates efficient camera calibration achieving up to 95% accuracy, robust feature detection and matching using algorithms such as SIFT, ORB, and AKAZE, and employs pose estimation, triangulation, and bundle adjustment to ensure high accuracy and detail. The method is lightweight, minimizing computational load, and is implemented on a user-friendly web-based platform. The solution demonstrated promising results, with reprojection errors as low as 15%, and effective 3D reconstructions of artifacts. Applications include virtual museums and the preservation and virtualization of artifacts, providing an interactive and immersive experience for users.

Keywords: 2D to 3D, Sfm Reconstruction, Camera Calibration, SIFT, Point Clouds, Triangulation, Web-based 3D Visualization

INTRODUCTION

Across many disciplines, including medical imaging, industrial design, virtual reality, and cultural heritage preservation, three-dimensional (3D) reconstruction has grown to be an increasingly useful tool.[1] 3D reconstruction fundamentally is the process of creating three-dimensional models from two-dimensional (2D) data, such as photographs or sensor inputs. Although this sounds basic, obtaining good reconstructions is difficult. The work requires advanced algorithms to understand depth, geometry, and spatial relationships from often lacking or confusing data. Fortunately, new advances in sensor technologies, machine learning, and computer vision have greatly enhanced the accuracy and efficiency of 3D reconstruction, hence increasing the availability of these techniques for use in the real world as well as in research.

One well-known instance is the work of Liritzis et al. (2021), who investigated the Sanctuary of Delphi using 3D reconstruction to show its capabilities.[2] By means of 3D scanning and modeling, they produced a digital replica of the location, therefore exposing the ability of technology to record precise geometric details and offer immersive visuals. In a similar study, Fadzli et al. (2023) carefully examined real-time 3D reconstruction methods with an eye toward their function in telepresence systems.[3] Their studies showed how these technologies might enable interactive presentations of difficult settings. Another study looked at using digital photogrammetry and laser scanning for exact 3D modeling,[4] therefore offering fresh opportunities for recording, analysis, and restoration methods particularly helpful when conventional preservation approaches are insufficient.

In 3D reconstruction, handling missing or unclear data remains a major difficulty, particularly in cases involving obscured areas of images. The creation of 3D-ReConstnet,[5] a neural network meant to create 3D object point clouds from single-view photos, offers one encouraging answer to this problem. This method highlights how

*Corresponding Author: Dr. Swati Shilaskar,
Vishwakarma Institute of Technology, Pune, India
Tel: +91 20 2420 2180; Email: swati.shilaskar@vit.edu

Cite as: *J. Integr. Sci. Technol., 2025, 13(5), 1120.*
URN:NBN:sciencein.jist.2025.v13.1120
DOI:10.62110/sciencein.jist.2025.v13.1120

deep learning may assist generate semantically rich 3D reconstructions from minimal data by modeling the uncertainty in obstructed areas using a Gaussian probability distribution. It also emphasizes how increasingly important machine learning is in overcoming constraints of conventional reconstruction methods.

New technologies include augmented reality, real-time processing, and deep learning are offering fascinating opportunities to increase 3D model correctness and interaction. Galabov (2015), for instance, presented a novel set of techniques based on motion parallax, depth cues, and shading effects for turning 2D photos into 3D models.[6] Particularly in situations when thorough 3D data is difficult to get, these methods enable the conversion of 2D data into 3D representations without the necessity of sophisticated motion analysis, therefore facilitating the creation of 3D models.

The possibility to interact with 3D models in real-time expands the possible uses of these technologies even more. Using their layer-by-layer scanning approach for additive manufacturing,[7] which combines augmented reality to offer real-time monitoring and flaw detection, Malik et al. (2019) investigated this capacity. This method shows how real-time 3D reconstruction may be used for dynamic, interactive applications, such virtual museums or instructional platforms, where users may interact with 3D models in real-time.

The following sections of this paper delve into the specifics of the proposed methodology, examining the technical foundations and practical applications of 3D reconstruction techniques. By leveraging advancements in computer vision, machine learning, and sensor technologies, this work aims to contribute to the ongoing evolution of 3D reconstruction methods, with a focus on their application in preserving and presenting cultural heritage. The integration of these technologies not only enhances the accuracy and efficiency of 3D modeling but also enables new ways of engaging with and understanding complex spatial data, paving the way for future innovations in the field.

LITERATURE SURVEY

In recent years, computer vision-based techniques for three-dimensional (3D) reconstruction have gained significant attention due to their wide-ranging applications[8] in various fields. A detailed overview of these methodologies, together including the several approaches and algorithms applied for 3D reconstruction, is given by Ham et al. (2019) based on the data collecting technologies used - single-camera setups, multiple-camera configurations, Time-of-Flight (ToF), Shredded Light, and Kinect-based systems.[9] The important processing phases, namely feature extraction, depth estimate, and surface reconstruction are also covered by the authors.

Deep learning techniques have revolutionized the field of computer vision, offering enhanced capabilities for complex tasks such as 3D reconstruction. Liu et al. (2020) presented a novel method using deep learning to rebuild 3D structures from image sequences.[10] Along with feature fusion, their approach combines unsupervised and supervised learning methods to produce accurate depth estimate and surface modeling. Using artificial neural networks (ANN), Graph Convolutional Networks (GCN), and MarrNet, they produced striking 3D scene reconstruction findings. Similarly, Sakai et al. (2020) concentrated on deep learning 3D shape reconstruction from a single image.[11] Their approach utilizes Convolutional Neural Networks (CNN) to automatically learn valuable image features and reconstruct 3D shapes through two methods: normal vector estimation and direct 3D reconstruction. Experimental results on human face images demonstrate higher accuracy compared to previous methods, highlighting the potential of deep learning in 3D reconstruction.

Shalma et al. (2023) investigate several techniques applied in reconstruction of 3D images of both particular and generic objects in another thorough review.[12] The authors address the difficulties reconstructing increasingly complicated objects in addition to covering a broad spectrum of methods including 3D shape representations, depth estimate, and multi-view representations. Susheel Kumar et al. (2011) also provide a thorough study of 3D reconstruction algorithms, with specific attention on volumetric methods including voxel coloring.[13] Their work investigates the difficulties of these techniques, including the variation in performance when working with various image numbers and resolutions, and emphasizes the growing inclination for volumetric methods due of their lower complexity and the increasing capabilities of modern computing systems.

These methods also find interesting uses for the preservation of cultural legacy. Bent et al. (2022) explain how historical landmarks including Florence and Orsanmichele might be preserved via 3D reconstruction.[14] Their method generates very detailed 3D models by integrating photogrammetry, laser scanning, and sophisticated software processing, therefore helping to save and record these significant cultural sites. Likewise, Zhu et al. (2021) offer a novel technique to generate thermal point clouds by combining Thermal Infrared (TIR) images with Mobile Laser Scanning (MLS) point clouds.[15] Their approach involves key-points extraction using line-intersection, semi-automatic and automatic correspondence determination using restricted RANSAC for 6DOF pose estimation, and a non-local mean strategy for data fusion. Keypoint extraction and posture estimation combined in this approach produces comprehensive geometric and thermal data suitable for uses like structural analysis and energy efficiency evaluations.

Especially in relation to dynamic situations, real-time 3D reconstruction has also witnessed notable developments. Mehta et al. (2017) presented a single RGB camera based real-time 3D skeleton posture capture technique.[16] Stable and temporally consistent 3D reconstructions are generated by integrating a CNN-based pose regressor with kinematic skeleton fitting. Even in outdoor environments or with low-quality cameras, this method provides possibilities for use in several real-time applications including 3D character control and virtual reality. Figure 1 shows a synopsis of these approaches and tools.

With an eye toward virtual museums especially, the current effort aims to advance 3D reconstruction methods. Even if a lot of research has been done on these techniques, some difficulties still exist including problems with camera calibration accuracy, limited feature identification and matching, and the great computational needs of deep learning algorithms. Moreover, there is a clear

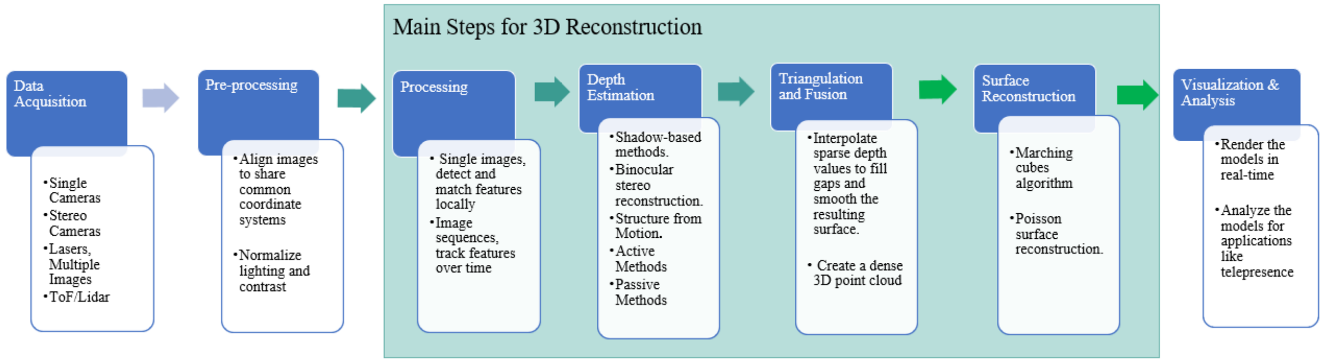


Figure 1. A comparison of various steps used by various researches

discrepancy between theoretical models and actual implementations; issues about the usefulness and accessibility of the produced 3D models abound. This work presents a scalable and strong 2D-to-3D reconstruction technique to address these difficulties. Along with a mix of feature recognition and matching techniques for dependable keypoint identification, it uses an effective camera calibration procedure to reach great accuracy. To guarantee great authenticity in the 3D models, also used are methods such as triangulation, bundle adjustment, and pose estimation. This method is lightweight and efficiency-oriented unlike many deep learning-based techniques, which are sometimes computationally costly. Moreover, the system is meant to be reachable via a web-based platform, which enables the development of virtual museums helping to preserve cultural legacy and educate about it.

METHODOLOGY

High-quality 3D models are becoming more and more important as virtual museums grow more well-liked due to their accessibility and ability to interact with artefacts. However, the process of making these models may be costly and time-consuming, and it frequently fails to adequately capture textures and small details.

This research proposes a straightforward yet effective method to improve 2D images of artefacts into 2D-to-3D reconstruction techniques for the development of a virtual museum. This improvement will lead to a more robust digital preservation of cultural and historical legacy by improving the visual appeal and immersiveness of virtual museum exhibits. The basic idea in 3D image reconstruction is that given a set of images with a different viewpoint, the goal is to use these images to reconstruct a three-dimensional representation of the object. More specifically, the motion of the cameras is found with respect to a world coordinate frame F. In simpler terms, the camera's movement in the 3D world needs to be tracked and translated to the 2D image it captures. This translation is done through what's known as camera projection matrices, representing the motion of the cameras, denoted as W. Using this set of camera projections, different algorithms are used to recover the 3D structures of the scene. This will provide a light-weight architecture which is integrated to a website to make it accessible to users. The various steps of the architecture used for 2d-3d modeling in this setup are shown in Figure 2.

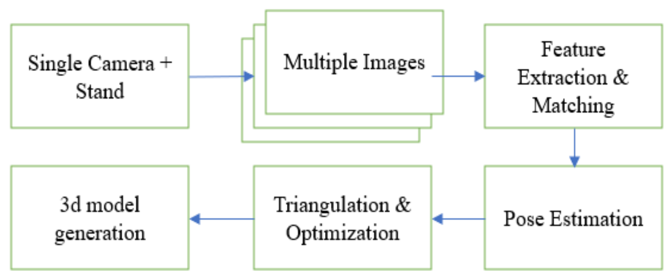


Figure 2. Overview of the steps followed for 3d reconstruction.

Figure 3 shows the block diagram of the proposed solution and includes the following steps: i) Camera preparation, ii) Data Acquisition & Preprocessing, iii) 2d-3d reconstruction algorithm, iv) Optimization techniques and v) platform building.

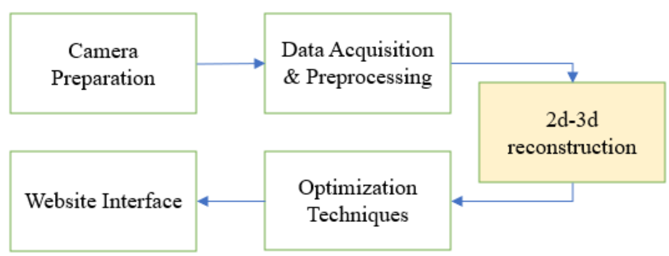


Figure 3. Flowchart for the proposed solution

The subsequent sections delve into the specifics of each of these blocks, beginning with the preparation of the camera for capturing the artifact images.

Camera Preparation

Camera calibration plays a pivotal role in ensuring accurate and reliable results.[17] Calibration serves to rectify distortions inherent in camera lenses and to establish the intrinsic parameters necessary for accurate geometric reconstruction. In this project, the checkerboard method was employed for camera calibration. This method involves capturing multiple images of a checkerboard pattern with known dimensions from different viewpoints. The geometric structure of the checkerboard allows for the extraction of corresponding image points, which are then used to estimate the camera's intrinsic parameters, such as focal length and lens distortion coefficients, through a calibration process. To facilitate the calibration process, a series of images capturing the checkerboard from various orientations and distances were acquired as shown in Figure 4.

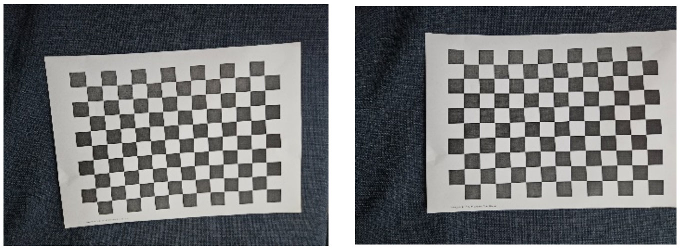


Figure 4. Images for Camera Calibration

The known parameters included the size of each box taken as 15mm x 15mm and the number of corners present in the checkerboard image (width, height) as (9,13). These values are used for calculating the parameters needed to estimate the camera matrix, as given in Equation (1):

$$x = K[R|t]X \quad (1)$$

where, respectively, x and X stand for the coordinates in the image plane and the three-dimensional world space. The camera matrix with the intrinsic parameters is denoted by K. The camera matrix with the extrinsic parameters translation and rotation is denoted as [R|t]. This formula illustrates how to use the camera matrix K and a rigid-body transformation [R|t] to project 3D points X to the 2D picture plane x. Finding the ideal K and [R|t] to minimise the re-projection error—that is, the discrepancy between the projected and observed picture points—is the process of camera calibration. This is usually achieved via a least-squares optimization routine. The calibration results, comprising the camera matrix, distortion parameters, rotation vectors, and translation vectors were stored, facilitating the undistortion of images captured by the calibrated cameras for further processing.

Data Acquisition & Preprocessing

A number of systematic procedures are involved in the 3D reconstruction process from 2D pictures, especially for statues and monuments. For further geometric computations, the intrinsic parameters—which comprise elements like focal length, primary points, and lens distortion coefficients—are essential. The camera was attached to a sturdy stand after calibration. The first step of this procedure calls for meticulous setup and calibration of the camera gear, which includes a suitable camera that can capture 4000x3000 pixel images and a solid position for the Digitek® DTR 550 LW 67-inch Foldable Tripod Stand. It is essential that the camera settings stay the same for every shot; manual exposure, white balance, and focus adjustments. The choice of the object itself is crucial to achieving the best 3D reconstruction; sculptures with more texture and no obstacles produce better results. The camera setup for this project is seen in Figure 5.

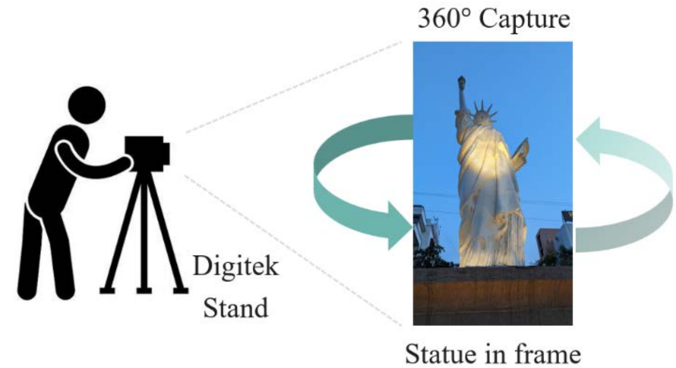


Figure 5. Camera setup for data acquisition

An pivotal factor in taking pictures is the actual distance between the statue and the camera. A constant distance was kept while taking pictures of the monument from various perspectives in order to ensure the most realistic 3D reconstruction. The angles were chosen so that there was at least 60% overlap between each image and its predecessor. During the 3D reconstruction phase, this overlap makes image alignment easier.

The images are arranged chronologically, creating a logical flow for further examination, to facilitate methodical processing. A custom function preprocesses these images with Gaussian blurring, normalisation, and histogram equalisation are included in this function. Image noise and detail can be decreased with the aid of Gaussian blurring. Through the normalisation process, the intensity of the image is scaled to fall within a predetermined range, often 0 to 255. By applying histogram equalisation to the luma (Y), or brightness, blue projection (U), and red projection (V) colour spaces, the image's contrast is improved, making details easier to see. Downsampling is then utilised to lower each image's resolution while maintaining important visual details. The goal of this meticulous downsampling procedure is to balance image quality and computational effectiveness. This balance is critical for ensuring optimal performance during subsequent processing stages.

The Structure from Motion (SfM) algorithm[18] was applied for the reconstruction step after the image capture process was finished. Key-point detection, key-point matching, bundle modification, and dense point cloud production are some of the phases involved in this process.

2D-3D Reconstruction Algorithm

A two-step procedure starts as soon as the images are prepared. Initially, 2D photos are transformed into 3D models using SfM. Because of its scalable and user-friendly design, it is ideal for online platforms such as virtual museums. In the second section, known as data association, pictures are compared to determine how similar they are. This aids in producing a precise 3D model that may be improved upon.

Feature Detection and Matching

Finding and matching features across several images is the initial step in the SfM process, which is accomplished by employing feature descriptor methods. These consist of Scale-Invariant Feature Transform (SIFT),[19] Oriented FAST and Rotated BRIEF (ORB),[20] and Accelerated-KAZE (AKAZE).[21] These algorithms identify distinct points in the pictures, which are subsequently compared to produce three-dimensional points. SIFT identifies local characteristics in pictures. It is based on the object's appearance and is unaffected by small changes in viewpoint, noise, illumination, rotation, or image scale. SIFT identifies possible key spots in an image, applies different filters to it, determines their stability, assigns an orientation, and then computes a descriptor for these key points. Conversely, ORB is a feature detector that is both quick and robust.

It performs well in real time and is accurate and efficient. ORB makes rotation invariant by identifying important points using a pyramid structure and determining their orientation. A modified version of the BRIEF descriptor is then used to further process these important points. The feature matching algorithm AKAZE is precise, consistent, and effective. It employs the Additive Operator Splitting (AOS) scheme, a novel technique. Instead of smoothing images, it diffuses them using a nonlinear scale space. In this scale space, the approach finds important spots and calculates their descriptors. The end product is a collection of unique and robust key points that work well with a variety of noise and transformations.

Each of these approaches has advantages and disadvantages of its own. Despite being resilient to changes in scale and rotation, SIFT can be computationally demanding. Conversely, ORB is a quicker approach that is resistant to rotational but not scale changes. An algorithm called AKAZE offers a decent balance between accuracy, robustness, and efficiency. A FlannBased Matcher is developed to match the critical spots when the SIFT detector is utilised. A collection of algorithms designed for quick nearest neighbour searches in big datasets and for high dimensional features may be found in the Fast Library for Approximate Nearest Neighbours (FLANN) library. It performs better than conventional algorithms, offering a quicker and more efficient way to look for patterns in data. Here, the two nearest neighbours (k=2) for each descriptor are found using the knn Match technique and the FlannBased Matcher, as indicated in Equation (2).

Let D1 and D2 be descriptors from the first and second images, respectively. For each descriptor d1 D1, FLANN finds the two nearest neighbors d2_1 and d2_2 in D2 :

$$Match\ is\ good = \frac{distance(d1, d2_1)}{distance(d2, d2_2)} < ratio\ threshold \tag{2}$$

When the ORB or AKAZE detectors are used, Brute-Force Matcher (BFMatcher) is created to match key points. As shown in Equation (2), the BFMatcher matches each descriptor in the first image with the descriptor in the second image that it is most like, according to the specified distance measurement method. Here, Hamming distance is used as the norm type when 'ORB' is used, and Hamming2 is used for AKAZE. Regardless of the type of matcher used, the Lowe's ratio test is then applied to filter out good matches. When ORB or AKAZE is used, all matches are considered good matches as they are already cross-checked.

Pose Estimation and Triangulation

Post establishment of correspondences between images, the geometric constraints and the Perspective-n-Point (PnP) method are used to estimate the relative position and orientation of pairs of images. The PnP method is used to determine the camera's position and orientation, or 'pose', in 3D space. This is done by using a set of known points in 3D space and their corresponding 2D projections in the image. Accurate 3D models require an understanding of the camera's pose at the time of each shot. The method is implemented by reducing the reprojection error given in Equation (3):

$$\frac{min}{R,t} \sum_i ||m_i - \pi(RM_i + t)||^2 \tag{3}$$

where is the projection function, R is the rotation matrix, t is the translation vector, and M_i, m_i are the 3D and corresponding 2D points. The RANdom SAmple Consensus (RANSAC) technique is utilised to address outliers in the point correspondences and strengthen the pose estimate.

Following the establishment of the camera's pose for each images, the 3D coordinates of the scene's points are estimated using the triangulation technique. This entails determining a point's location in three dimensions by comparing its positions in two or more images. The triangulation technique serves as the basis for creating an extensive 3D model by reconstructing the scene's spatial arrangement using fundamental geometric concepts. Equation (4) illustrates the triangulation process. In order to determine the 3D point X, two projection matrices P1 and P2 and the associated points x1 and x2 in the images are required.

Triangulation is used to estimate the 3D coordinates of points in the scene once the camera's posture has been determined for each picture. This involves determining the location of a specific point in 3D space based on its position in two or more images. By using basic geometric principles, the triangulation process reconstructs the spatial layout of the scene, forming the foundation for generating a comprehensive 3D model.

The triangulation process can be described as shown in Equation (4). Given two projection matrices P1 and P2 and the corresponding points x1 and x2 in the images, the 3D point X can be found.

$$AX = 0 \tag{4}$$

Where A is constructed as:

$$A = \begin{bmatrix} x1P1_3 - P1_1 \\ y1P1_3 - P1_2 \\ x2P2_3 - P2_1 \\ y2P2_3 - P2_2 \end{bmatrix}$$

Here, $P1_i$ and $P2_i$ denote the i^{th} row of the projection matrices $P1$ and $P2$, and $(x1, y1)$ and $(x2, y2)$ are the coordinates of the points in the images. The entire process followed for model building is represented through Figure 6.

The triangulation process takes as input matched points from two images and the respective projection matrices of the cameras that captured these images. Figure 7 shows the triangulation method.

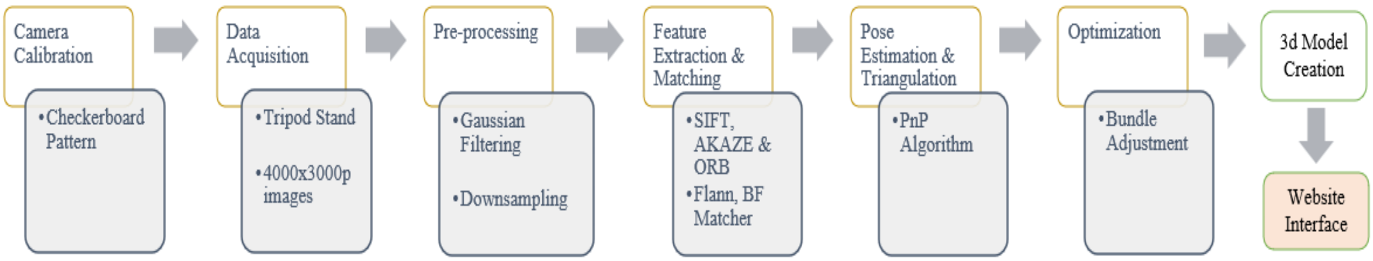


Figure 6. Complete steps of the 2d to 3d pipeline followed for 3D model building

The point in 3D space where these lines best intersect is calculated using linear algebra technique like the Direct Linear Transform (DLT). This intersection point in 3D space is the estimated real-world position of the matched point. A collection of 3D points that constitute a sparse reconstruction of the scene is produced by repeating this process for each matching point.

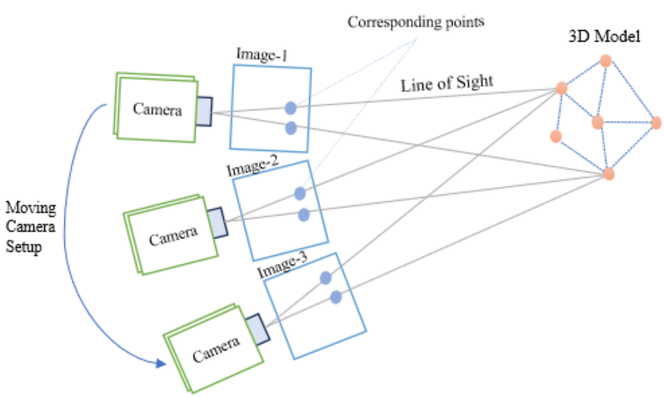


Figure 7. Illustration of the Triangulation Method[22]

Optimization & Model building

The triangulation approach yields a dense cloud of points for 3D reconstruction. The bundle adjustment algorithm is used to fine-tune these 3D points and camera pose estimates obtained from the PnP approach. This step helps align the images accurately and prepares them for reconstructing the scene's shape. Bundle Adjustment (BA), a computer vision optimisation technique, maximises the accuracy of the 3D structure and camera settings. Using the initially calculated 3D points and camera poses, it minimises the discrepancy between the observed and predicted positions of points. The nonlinear least squares method is used in the optimisation process. Using the initial estimations, the error is calculated, the parameters are changed to minimise the error, and the process is repeated until either the maximum number of iterations is achieved or the error can no longer be considerably decreased. The reconstructed scene's shape is then represented by a PLY file for visualisation and analysis, which is created from the reconstructed 3D points and their colours for simple rendering and manipulation in 3D modelling tools.

All of the procedures needed to perform Structure from Motion (SfM) and produce a 3D reconstruction from 2D images are described in Algorithm 1.

Algorithm 1: 2d to 3d reconstruction method

```
Input: 2d images
Output: 3d model file for visualization
Initialization
IL ← Image_loader
FF ← FindFeatures
EM ← ComputeEssentialMatrix
BA ← BundleAdjust
for img in IL(dir, scale) do:
        pts1, pts2 ← FF(img_0, img_1)
        matches ← MatchFeatures(pts1, pts2)
        P1, P2, E ← EM(matches)
        R, T ← RecoverPose(E)
        p_cloud ← TriangN(pts1, pts2, P1, P2)
        opt_params ← BA(K, R, T, p_cloud, re)
        pts, colors ← FormatColors(p_cloud)
        verts ← PointsColors(pts_out, colors)
        ply file ←GeneratePLY(len(verts), verts)
return 3D model file
End
```

The process of turning 2D photos into a 3D model is represented by the algorithm. Images are loaded, their features are recognised and matched, and their projection matrices are calculated. When triangulating points to create a 3D point cloud, this information is used to calculate the necessary matrix and retrieve posture data. The point cloud is normalized and optimized to reduce reprojection error. After this optimization, the data is formatted, post processed, and combined with color information. Finally, the output is written into a .ply file, which represents the 3D reconstruction of the original 2D images. This file can be visualized on different platforms. Here, cloudcompare, an open-source software is used to visualize the 3d model.

Platform building

The generated 3D models are integrated on a website. This website serves as a platform for virtual meuseum and showcasing historical artifacts. This makes it possible for users to view exhibitions from any device with internet access. To ensure the best possible viewing and interaction on various screen sizes, the responsive design complements the user-friendly interface. A comprehensive service area for generating 3D models from 2D pictures, navigational tools, and a carousel showcase are among the interactive elements. The website incorporates technologies such as Google Web Fonts and Icon Libraries, Animate.css, WOW.js, Owl Carousel, and Lightbox to improve user experience and picture presentation. It has fast loading speeds, accessible design, and cross-browser compatibility to optimise usage.

The primary function of the website is the 3D model viewer, which enables in-depth engagement with rebuilt artefacts. WebGL technology is used in the viewer's implementation to provide responsive and fluid model manipulation. By rotating, zooming, and panning the models, users may view the artefacts from all directions. The complexities and historical relevance of each artefact can be better understood by users thanks to this access.

Figure 6 details the entire process for the suggested 2D to 3D reconstruction model, including all preprocessing and postprocessing steps.

RESULTS AND DISCUSSION

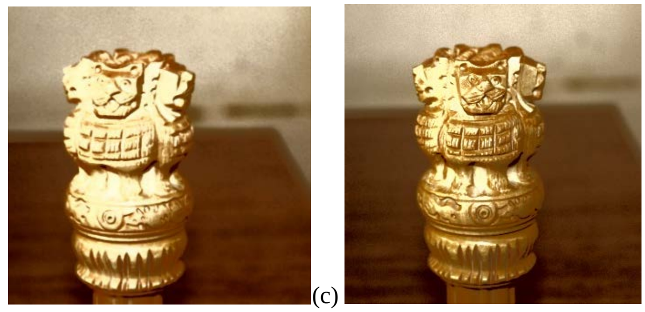
The implementation of the proposed 2D-to-3D reconstruction method yielded significant and promising results. The suggested 2D-to-3D reconstruction method's use produced noteworthy and encouraging outcomes. The results from several steps of the procedure are shown in this part, including camera calibration, data collection and preprocessing, feature matching and detection, pose estimation and triangulation, and the finished 3D reconstruction.

Performance Evaluation of the 2d to 3d reconstruction

Using the chequerboard method, the camera calibration procedure was successfully finished. A number of pictures were taken of the chequerboard from various angles and distances. Accurate estimations were made of the intrinsic camera parameters, such as the distortion coefficients and camera matrix. The camera was calibrated up to 95%, according to the calibration results, which included the camera matrix, distortion parameters, rotation vectors, and translation vectors. Additionally, an inaccuracy of roughly 0.07% was demonstrated when a distorted image was undistorted using the camera settings.This calibration made sure that later photos were distortion-free and recorded the scene's geometric characteristics.

The images were preprocessed using methods such histogram equalisation, normalisation, and Gaussian blurring to improve there quality. In order to balance image quality and computing efficiency, down sampling was also used. The images were consistently prepared thanks to the preprocessing stages, which made 2D-to-3D reconstruction precise and effective. Figure 8 displays the various preprocessing methods used on one image.

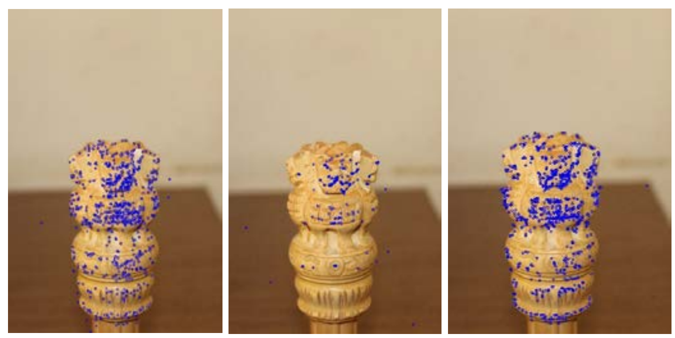
(a) (b)



(c) (d)

Figure 8. Preprocessing techniques applied on the images (a) Original Image (b) Gaussian Blur (c) Down Sampling (d) Histogram Equalization

The SIFT, ORB, and AKAZE algorithms were used in the feature detection and matching procedure. Every technique found distinctive keypoints in the images, which were compared to create correlations. While ORB and AKAZE gave computational efficiency, SIFT offered robust feature recognition and matching. A comparison of keypoints captured in various approaches, which served as the basis for the 3D reconstruction, namely, SIFT, ORB, and AKAZE, respectively are depicted in Figure 9.



(a) (b) (c)

Figure 9. Feature matching from different applied algorithms (a) SIFT (b) ORB (c) Akaze

SIFT yields 6821 key points and 128-dimensional descriptors, providing a rich structure for 3D modeling, despite being computationally intensive. ORB, on the other hand, is a faster and more efficient alternative to SIFT, designed for real-time applications. However, it detects fewer features (500 key points with 32-dimensional descriptors) which are not evenly distributed. The A-KAZE algorithm identifies the greatest number of features (6985 key points with 61-dimensional descriptors) fairly evenly distributed across the image. Depending on the specific use case, while A-KAZE was valuable for computational efficiency and dense feature matching, SIFT excelled in providing a better structure for the 3D model due to its robustness and distinctive descriptors.

The table (1) below summarizes the performance of SIFT, ORB, and AKAZE algorithms in terms of the number of key points detected, the dimensions of their descriptors, and their key features. The values were observed over a set of images, and the average number of keypoints detected by each algorithm is presented.

Table 1. Comparison of Algorithms & their performances

Algorithm	No. of Key points detected	Descriptor Dimensions	Key Features
SIFT	6821	128	Robust feature detection, computationally intensive
ORB	500	32	Fast, efficient, fewer and unevenly distributed features
AKAZE	6985	61	High number of features, evenly distributed, efficient

Upon acquiring the features and their descriptions, they were feature matched. The application of Lowe's ratio test[23] filtered out good matches, ensuring the reliability of the correspondences. The feature matching image is shown in Figure 10.

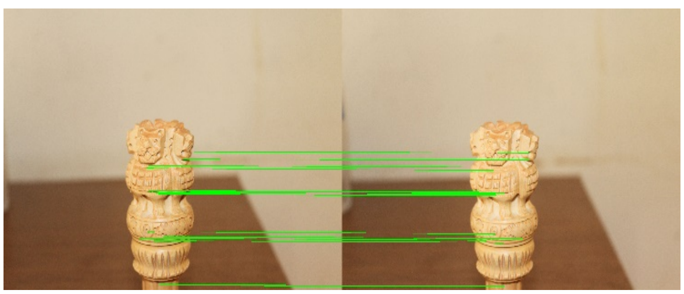


Figure 10. Feature matching on the keypoint detected algorithm

Post completion of feature matching and pose estimation using the Perspective-n-Point (PnP) algorithm, triangulation was then used to estimate the 3D coordinates of scene points based on their 2D image correspondences and camera projection matrices. The initial 3D point cloud, obtained from triangulation, was refined using an optimization technique known as bundle adjustment. This technique minimizes the reprojection error by iteratively fine-tuning the camera parameters and the positions of the 3D points. The result of this fine-tuning is a coherent and consistent 3D reconstruction. An optimized 3D model exhibits reduced reprojection error, which is indicative of a high-fidelity reconstruction of the scene.

The term reprojection error refers to the alignment degree between a 3D point, estimated via triangulation, and its corresponding 2D projection in the image. Ideally, for accurate 3D reconstruction, this error should be minimized. This error is measured by projecting the 3D points back onto the camera image plane and comparing the projected points with the original 2D image points. The distance between these projected and original points is then computed, typically in pixels. To illustrate the variation of the reprojection error across the iterations, Matplotlib was used to create a scatter plot. This plot is shown in Figure 11(a), maps the reprojection error against its corresponding image, and serves as an important tool for visually tracking the behavior of the error as more images are processed. It indicates the worst-case error scenario obtained from the various models reconstructed. As observed, except for a few spikes reaching 15%, reprojection error either reduces or at least maintains stability as the process progresses. This indicates an improvement in the accuracy of the 3D reconstruction. A significant increase in the error for the 15th image is due to less features in it affecting feature matching, complications with the essential matrix and pose recovery. With the error plotted on the website, users can visually follow its trend, making it easier to identify and resolve potential problems by updating their images. Figure 11(b) shows the plot for the reconstruction error obtained from the algorithms (SIFT, AKAZE and ORB) for the artifact taken in study for the demonstration of all the processes earlier. As observed, the reconstruction error is typically much less than 15% for SIFT and 35% for AKAZE, which are the worst-case values obtained during the reconstruction of various models during our study.

Figure 11 (a). Plotting the reprojection error matrix against images using SIFT for the worst-case reprojection

Finally, the 3D points, along with their associated colors, were converted into a PLY file format. This file was used for visualization and analysis, providing a detailed and accurate representation of the reconstructed artifact. The reconstructed 3D models were then integrated into a virtual museum platform, accessible via a dedicated website. There are various ways of modeling and visualizing these models.[24] The 3D model viewer, powered by WebGL technology, was used due to its lightweight faster integration with the website and its effective representation of point clouds. This enabled users to rotate, zoom, and pan the models, allowing them to examine the artifacts from various angles. Figure 12 showcases the user interface for the website.

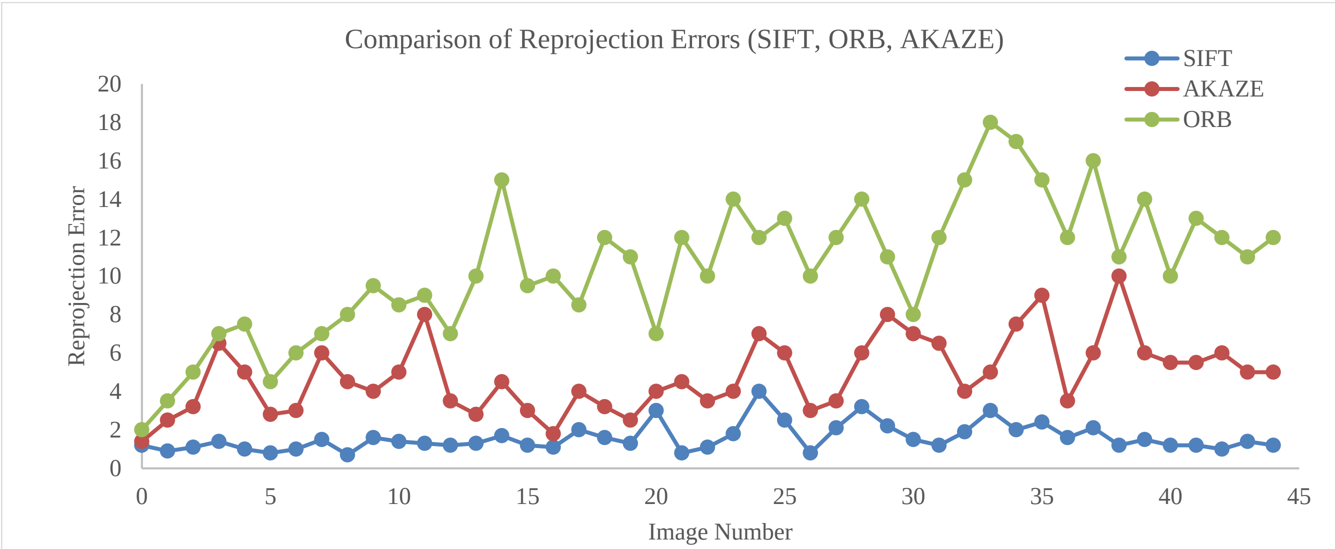
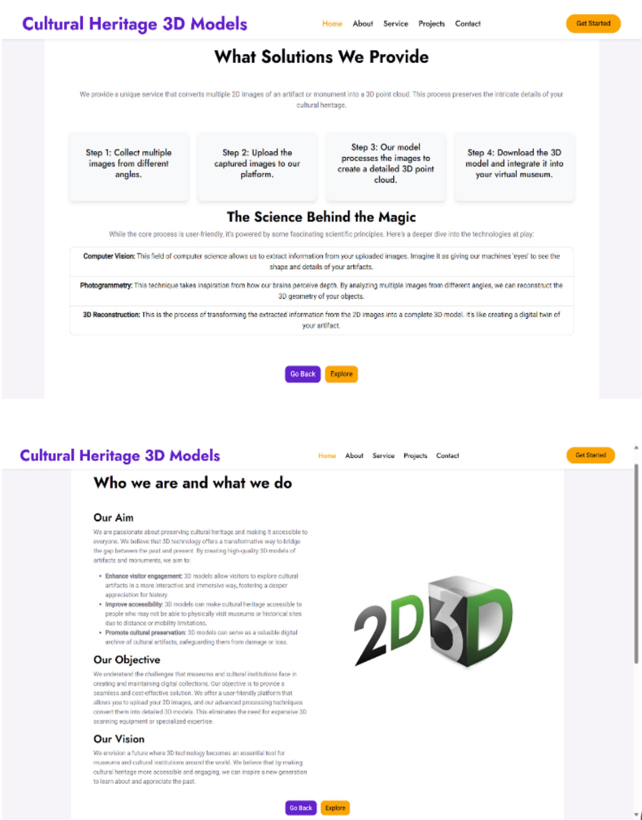


Figure 11 (b). Plotting the reprojection error matrix against images using all algorithms for the reconstructed artifact.



(a)

(b)

Figure 12. Website Interface for 3d reconstruction (a) About Page and (b) Services Page

Functional/Usability evaluation

The evaluation of the developed model was based on its application in virtual museums, with focus placed on two primary cases: small figurines and other statues or heritage sites. The Structure from Motion (SfM) technique, utilized by the model, is used for low-cost 3D reconstruction that creates a sparse representation of the scene, resulting in sparse point clouds.[25] Despite the sparsity, substantial detail was retained, and an accurate representation of the original form was provided by the generated 3D models. An example of point clouds for such small artifacts is illustrated in Figure 13, using the same artifact referenced in the preprocessing and feature matching results.

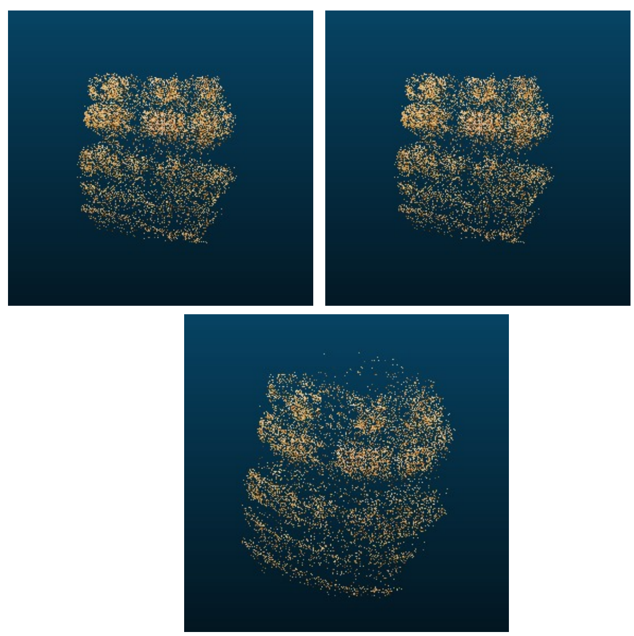


Figure 13. Sparse Representation of Point Clouds for smaller artifacts

Another example of a reconstructed artifact is shown in Figure 14. In the case of other artifacts and heritage sites, the reprojection error was found to be higher, averaging around 30%. This increased error could be attributed to the larger scale and complexity of these structures. However, valuable insights into their 3D structure and spatial characteristics were still provided by the reconstructed models of these sites.

(a) (b)

Figure 14. (a) Original Image (b) Sparse Representation of Point Cloud

Apart from objects, reconstructions were also carried out for historical monuments. Figure 15 showcases the reconstruction derived from smaller representations of historical sites, as traveling to these sites to capture images was not feasible. The details of the original artifact were preserved in the virtual representations (15 (b), 15 (d)), allowing for a more comprehensive visualization. The observed results demonstrated the model's potential for reproducing similar results in a more realistic setting.

(a) (b)

(c) (d)

Figure 15. Historical structures visualization (a) Original image of Petra (b) Point cloud visualization of Petra and (c) Original Image of Taj Mahal (d) Point cloud visualization of Taj Mahal

A thorough understanding of the parameters influencing 3D reconstruction was found to be essential as they directly impacted the success and accuracy of the reconstruction process. These parameters, including depth information, number of images, changes in baseline, camera calibration, image resolution, and feature matching techniques, among others, played significant roles in the 3D reconstruction pipeline. The level of detail and completeness of the reconstruction primarily differed when reconstructing a 3D model using 40 images compared to 60. A denser and more complete reconstruction was generally yielded by more images, but this also increased the computational complexity and time required for processing. Instances of overfitting with many random points in the 3D model were also observed, obscuring the determination of shape and structure. Therefore, the optimal number of images were dependent on the specific features and size of the artifact, balancing between reconstruction quality and computational efficiency. The baseline in 3D reconstruction refers to the distance between two camera positions and was found to be crucial in determining the accuracy and depth resolution of the 3D reconstruction. The distance between the camera setup and the monument or artifact played a crucial role in the quality of the reconstruction. This was because a smaller distance will lead to better image resolution and hence, better depth information. However, this also increases the error margin as in small distances, little fluctuations can cause huge changes in the information obtained. This error margin can be reduced by increasing the distance. However, if the distance increases too much, the depth information might be lost. Therefore, it is crucial to keep a balance between the two.

Conclusion

Rebuilding 3D models from 2D images was the main focus of this project, which has applications in digital preservation and artefact visualisation in a virtual museum context. In order to guarantee precise image acquisition, the procedure started with camera calibration. Data collection and high-resolution image preprocessing came next. Using feature detection and matching methods like SIFT, ORB, and AKAZE, the 3D reconstruction's accuracy was attained. The FLANN algorithm was then used to match them. After estimating the pose using the Perspective-n-Point technique, 3D coordinates were determined using triangulation. Bundle adjustment was used to optimise the original 3D point cloud, resulting in a precise and comprehensive 3D model. This model was then added to a website, providing viewers with an immersive experience. By transforming 2D images into 3D models, the platform helps in visualizing the virtual museum.

The virtual meuseum and visualization of historical artifacts are not the only use for three-dimensional reconstruction. Reconstructed 3D meshes are widely used in gaming and virtual reality (VR) to produce realistic and immersive environments.[26] Moreover, 3D reconstruction is crucial in the industrial and medical fields.[26] By enhancing the completeness and quality of 3D models, sophisticated reconstruction techniques can get around the limitations of traditional approaches. This is important for applications like endoscopic inspection and non-destructive testing in industrial settings.[27]

Looking further, more development and improvement are yet possible. The process of turning 2D images into 3D models can be refined even further. The virtual museum platform could also be expanded to include more artifacts and interactive features. Deep learning, an advanced technology, can be used to improve the process even more.[28]

The project achieved its goal of providing a simple yet efficient approach to enhance 2D images of artifacts with depth information, contributing to the digitalization of meuseums. The outcomes highlight the potential for further development and application of these techniques in various fields, including education, research, and tourism.

Conflict of Interest Statement

The authors declare that there is no conflict of interest regarding the publication of this work.

References and notes

1. L. Zhou, G. Wu, Y. Zuo, X. Chen, H. Hu. A Comprehensive Review of Vision-Based 3D Reconstruction Methods. Sensors 2024, Vol. 24, Page 2314 2024, 24 (7), 2314.
2. I. Liritzis, P. Volonakis, S. Vosinakis. 3D Reconstruction of Cultural Heritage Sites as an Educational Approach. The Sanctuary of Delphi. Applied Sciences 2021, Vol. 11, Page 3635 2021, 11 (8), 3635.
3. F.E. Fadzli, A.W. Ismail, S. Abd Karim Ishigaki. A systematic literature review: Real-time 3D reconstruction method for telepresence system. PLoS One 2023, 18 (11), e0287155.
4. E. Agosto, L. Bornaz. 3D Models in Cultural Heritage: Approaches for Their Creation and Use. International Journal of Computational Methods in Heritage Science (IJCMHS) 2017, 1 (1).
5. B. Li, Y. Zhang, B. Zhao, H. Shao. 3D-ReConstnet: A Single-View 3D-Object Point Cloud Reconstruction Network. IEEE Access 2020, 8, 83782–83790.
6. M. Galabov. A Real Time 2D to 3D Image Conversion Techniques. International Journal of Engineering Science and Innovative Technology (IJESIT) 2015, 4 (1), 297–304.
7. A. Malik, H. Lhachemi, J. Ploennigs, A. Ba, R. Shorten. An Application of 3D Model Reconstruction and Augmented Reality for Real-Time Monitoring of Additive Manufacturing. Procedia CIRP 2019, 81, 346–351.
8. W. Jeong, J. Kim, H. Kweon, K.J. Yoon. Multi-View 3D Scene Abstraction from Drone-Captured RGB Images. IEEE Access 2025.
9. H. Ham, J. Wesley, Hendra. Computer Vision Based 3D Reconstruction : A Review. International Journal of Electrical and Computer Engineering (IJECE) 2019, 9 (4), 2394–2402.
10. Y.P. Liu, X.P. Yan, N. Wang, X. Zhang, Z. Li. A 3D reconstruction method of image sequence based on deep learning. J Phys Conf Ser 2020, 1550 (3), 032051.
11. K. Sakai, Y. Yasumura. Three-Dimensional Shape Reconstruction from a Single Image by Deep Learning. International Journal of Advanced Computer Science and Applications 2020, 11 (2), 347–351.
12. H. Shalma, P. Selvaraj. A review on 3D image reconstruction on specific and generic objects. Mater Today Proc 2023, 80, 2400–2405.
13. K. SusheelKumar, V. Bhaskar, S. Prasad, R.C. Tripathi. Generating 3D Model Using 2D Images of an Object. International Journal of Engineering Science and Technology (IJEST) 2011, 3 (1), 406–415.
14. G.R. Bent, D. Pfaff, M. Brooks, R. Radpour, J. Delaney. A practical workflow for the 3D reconstruction of complex historic sites and their decorative interiors: Florence As It Was and the church of Orsanmichele. Heritage Science 2022, 10 (1), 1–17.
15. J. Zhu, Y. Xu, Z. Ye, L. Hoegner, U. Stilla. Fusion of urban 3D point clouds with thermal attributes using MLS data and TIR image sequences. Infrared Phys Technol 2021, 113, 103622.
16. D. Mehta, S. Sridhar, O. Sotnychenko, et al. VNect: Real-time 3D human pose estimation with a single RGB camera. ACM Trans Graph 2017, 36 (4).
17. O. Semeniuta. Analysis of Camera Calibration with Respect to Measurement Accuracy. Procedia CIRP 2016, 41, 765–770.
18. S.K. Karmacharya, N. Ruther, U. Shrestha, M.B. Bishwakarma. Evaluating the Structure from Motion Technique for Measurement of Bed Morphology in Physical Model Studies. Water 2021, Vol. 13, Page 998 2021, 13 (7), 998.
19. S. Castillo-Carrión, J.E. Guerrero-Ginel. SIFT optimization and automation for matching images from multiple temporal sources. International Journal of Applied Earth Observation and Geoinformation 2017, 57, 113–122.
20. A. Vinay, C.A. Kumar, G.R. Shenoy, K.N.B. Murthy, S. Natarajan. ORB-PCA Based Feature Extraction Technique for Face Recognition. Procedia Comput Sci 2015, 58, 614–621.
21. Y. Liu, C. Lan, C. Li, F. Mo, H. Wang. S-AKAZE: An effective point-based method for image matching. Optik (Stuttg) 2016, 127 (14), 5670–5681.
22. M. Rahman, H. Liu, M. Masri, I. Durazo-Cardenas, A. Starr. A railway track reconstruction method using robotic vision on a mobile manipulator: A proposed strategy. Comput Ind 2023, 148, 103900.
23. D.G. Lowe. Distinctive image features from scale-invariant keypoints. Int J Comput Vis 2004, 60 (2), 91–110.
24. M.N. Koeva. 3D modelling and interactive web-based visualization of cultural heritage objects. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 2016, XLI-B5, 297–303.
25. A. Eltner, G. Sofia. Structure from motion photogrammetric technique. Developments in Earth Surface Processes 2020, 23, 1–24.
26. M. Sarmah, A. Neelima, H.R. Singh. Survey of methods and principles in three-dimensional reconstruction from two-dimensional medical images. Visual Computing for Industry, Biomedicine, and Art 2023 6:1 2023, 6 (1), 1–19.
27. Y. Gu, J. Lv, J. Bo, et al. Joint Dense 3D Reconstruction Method for Endoscopic Images of Weak Texture Scenes. IEEE Access 2021, 9, 138254–138266.
28. A.K. Ingale, D.U. J. Real-time 3D reconstruction techniques applied in dynamic scenes: A systematic literature review. Comput Sci Rev 2021, 39, 100338.