iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.81254

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Mathematical Foundations of Camera Models and Projection Techniques in Computer Graphics

Goldi Soni[1], Bhabani Sankar Biswal[2], Neha Biswal[3]
[1]Assistant Professor, [2, 3]B.Tech (CSE), Amity University Chhattisgarh

Abstract: Digital images play a very important role in computer graphics, computer vision, and real-world applications such as robotics, augmented reality, and 3D reconstruction. The accuracy of these applications mainly depends on how well camera models and projection techniques represent real-world scenes. Over the years, many methods have been developed to improve camera calibration, geometric modeling, and image transformation. This review paper presents a comprehensive study of mathematical foundations used in camera models and projection techniques. It focuses on important concepts such as pinhole camera models, homography, camera calibration, and geometric transformations. The paper also discusses recent advancements where traditional geometry is combined with machine learning and deep learning methods to improve accuracy and robustness. Various approaches including classical methods, optimization-based techniques, and modern deep learning frameworks are analyzed and compared. The study highlights the advantages of combining geometric principles with learning-based models for handling challenges like distortion, noise, and real-world complexity. The main objective of this review is to provide a clear understanding of existing techniques and identify future directions for improving camera modeling and projection in computer graphics.
Keywords: Camera Models, Projection Techniques, Homography, Camera Calibration, Computer Graphics, Deep Learning, 3D Reconstruction

I. INTRODUCTION

With the rapid growth of digital technology, computer graphics and computer vision have become essential in many applications such as gaming, virtual reality, medical imaging, and autonomous systems. At the core of these technologies lies the concept of how a 3D real-world scene is captured and represented as a 2D image. This process is governed by camera models and projection techniques. Camera models describe how light from a scene is projected onto an image plane, while projection techniques define the mathematical relationship between 3D world coordinates and 2D image coordinates. One of the most commonly used models is the pinhole camera model, which provides a simple yet effective representation of image formation. However, real cameras often introduce distortions due to lens imperfections, making accurate modeling a challenging task. To address these issues, researchers have developed various camera calibration techniques that estimate intrinsic and extrinsic parameters of a camera. Traditional methods rely on geometric constraints and predefined patterns, but they often struggle in complex environments. Recent advancements have introduced learning-based approaches that improve accuracy by combining deep learning with classical geometry.

Another important concept in computer graphics is homography, which describes the transformation between two images of the same planar surface. Homography plays a key role in applications such as image stitching, augmented reality, and motion estimation. While classical methods use feature matching and optimization, modern techniques apply neural networks and transformer models to achieve better performance in difficult conditions.

In addition, the integration of geometry and learning has opened new possibilities for improving 3D reconstruction, pose estimation, and image alignment. These hybrid approaches leverage the strengths of both mathematical modeling and data-driven learning to produce more reliable results.

This review paper aims to study and analyze various techniques related to camera models and projection methods. It provides a structured overview of traditional approaches, modern learning-based methods, and hybrid techniques, highlighting their strengths, limitations, and practical applications.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

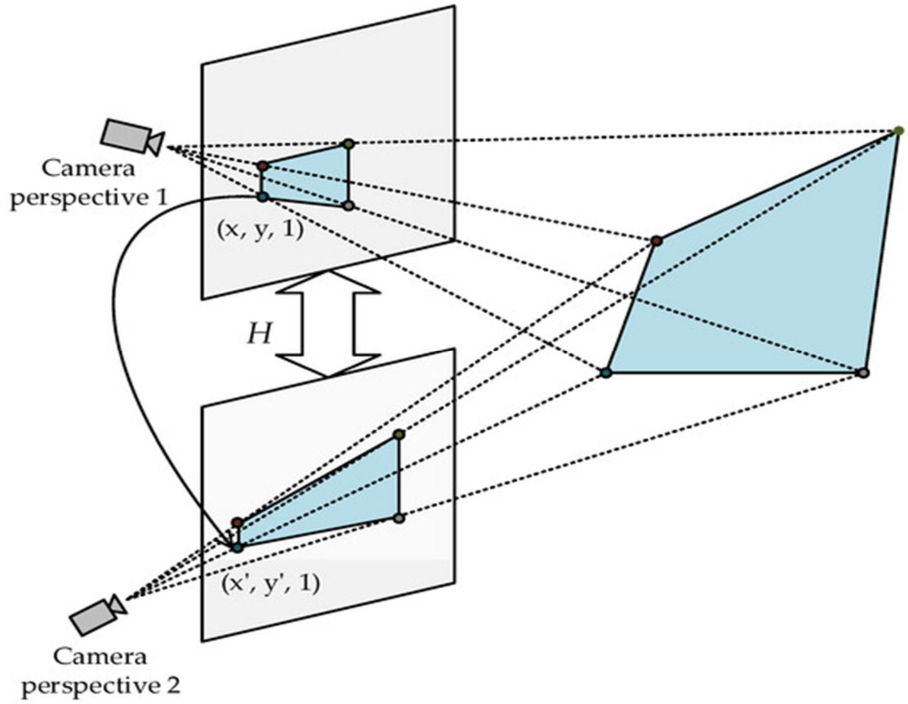


Figure 1. Mapping between two images using homography transformation.
This figure demonstrates how a planar surface viewed from different perspectives can be mapped using a homography transformation. It shows the relationship between corresponding points in two images.
(Source: https://www.mdpi.com/2079-9292/12/24/4977)

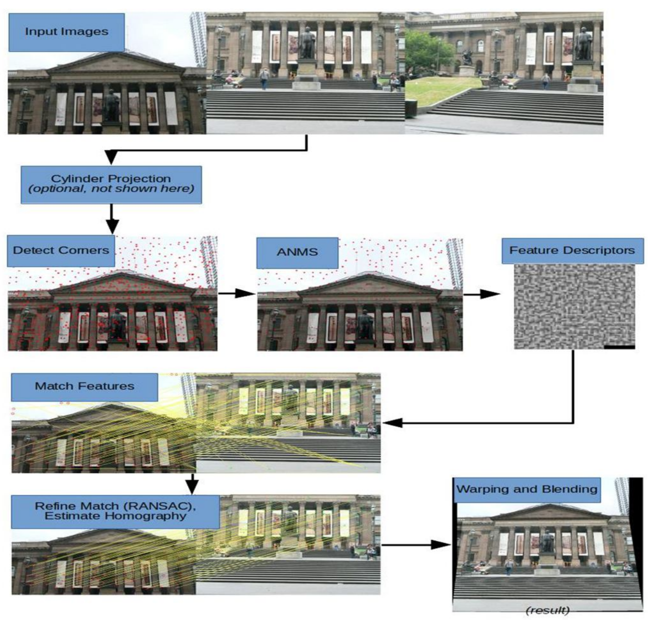


Figure 2: Image stitching pipeline using homography.
This figure shows the process of image stitching. Features are detected and matched between images, and homography is used to align them. Finally, images are warped and blended to create a single panorama.
(Source: Online image source)

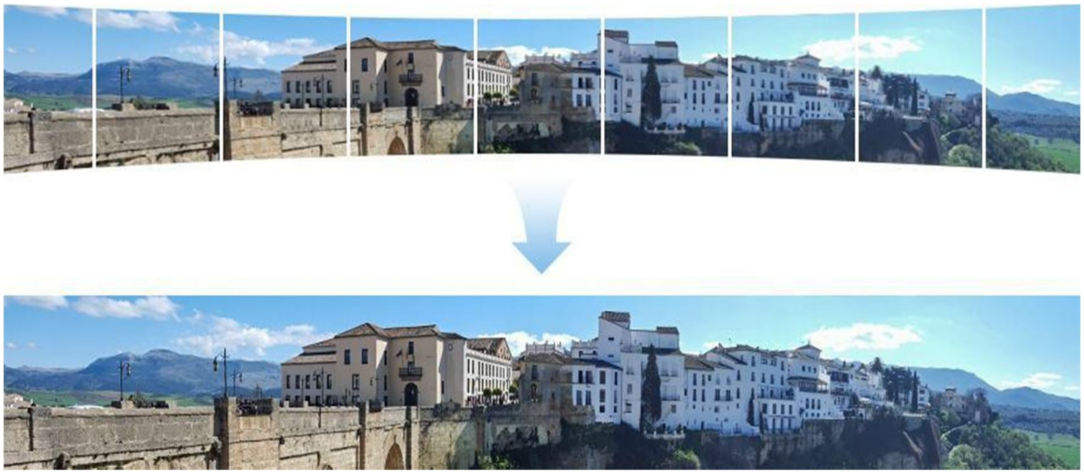


Figure 3: Panorama generated using image stitching based on homography.
This figure shows how multiple overlapping images are combined to form a single panoramic image. Homography aligns the images, and blending techniques ensure a smooth and continuous output.
(Source: Online image source)

II. OBJECTIVES

The main objective of this review paper is to study and analyze different camera models and projection techniques used in computer graphics and computer vision. The paper focuses on understanding how mathematical foundations help in accurate image formation and transformation. Another objective is to examine various approaches for camera calibration, homography estimation, and 3D reconstruction. The study includes both traditional geometric methods and modern learning-based techniques to understand their performance in different scenarios. The paper also aims to analyze how deep learning methods improve accuracy and robustness compared to classical approaches, especially in challenging environments such as noise, distortion, and lighting variations. In addition, the objective is to compare different research works based on their methodology, efficiency, and real-world applicability. This helps in identifying the strengths and limitations of existing techniques. Finally, the paper aims to highlight future research directions by exploring hybrid approaches that combine geometry and learning for more accurate and efficient computer vision systems.

III. LITERATURE REVIEW

The authors focused on improving camera calibration by converting real cameras into an ideal pinhole model. This is important because many computer vision tasks such as SLAM, 3D reconstruction, and pose estimation depend on accurate geometric measurements. Traditional calibration methods, like Zhang’s method, rely on predefined distortion parameters and iterative optimization, which often fail when images have strong nonlinear distortions.To solve this problem, the authors proposed a preprocessing method based on Gaussian Processes (GP). This approach learns the relationship between distorted image points and a perfect square grid using only a single checkerboard image. The GP model is capable of capturing complex distortions such as lens defects, mirror effects, and sensor noise. As a result, it generates a corrected image that behaves like a perfect pinhole camera.After this correction step, the calibration process becomes simpler because it does not require estimating distortion parameters or repeated optimization. Experimental results on both synthetic and real datasets show that this method reduces reprojection error and improves straight-line accuracy. The study concludes that this approach provides a more flexible, accurate, and reliable calibration method for different types of cameras.[1]

The authors proposed a fully automatic method for calibrating LiDAR and camera systems without using any calibration targets or manual steps. This is important because traditional calibration methods require special patterns, controlled setups, and multiple data captures, which are difficult to use in real-world applications like robotics and autonomous driving.The proposed method performs calibration using only a single pair of image and LiDAR point-cloud data. First, dense point clouds are created by combining LiDAR scans, and a virtual camera is used to generate LiDAR intensity images.

Then, correspondences between LiDAR and camera images are found using a deep learning-based matching technique.To improve accuracy, the method applies RANSAC to remove incorrect matches and minimizes reprojection error to estimate the initial alignment. Further refinement is done using a registration process that considers similarity between LiDAR and image data while removing occluded points.Experimental results show that the method achieves lower rotation and translation errors compared to traditional approaches. The study concludes that this single-shot and target-less calibration framework is efficient, accurate, and suitable for real-world sensor alignment tasks.[2]

This paper focuses on improving the accuracy of stereo camera systems used in close-range photogrammetry. In many practical situations, camera calibration parameters may change due to vibrations or slight movements, which leads to errors in 3D reconstruction. Such changes increase triangulation error and reduce measurement reliability. To address this issue, the authors proposed an optimization-based recalibration method that minimizes reprojection error to recover the correct camera parameters. The approach considers both simulation and real-world experiments. In the simulation stage, camera position and orientation are intentionally disturbed and then corrected using the Nelder–Mead optimization algorithm. The method estimates six parameters, including three translations and three rotations, to achieve accurate recalibration. In physical experiments, a stereo camera setup with adjustable positions is used along with an image-based measurement technique to validate the results. The experimental results show a significant reduction in reprojection error, restoring accuracy close to the original calibration even after displacement. The study concludes that this optimization-based approach effectively improves measurement precision and is useful in dynamic environments where camera positions may change.[3]

This paper aims to improve the accuracy and robustness of camera calibration by solving common problems such as distortion, inaccurate feature detection, and unstable parameter estimation. Traditional calibration methods often fail in complex conditions where images contain noise or lighting variations. To overcome these issues, the authors proposed a hybrid framework that combines deep learning with classical geometric methods. First, a convolutional neural network is used to correct radial distortion by transforming distorted images into rectified ones. This step helps in improving the quality of input data for calibration. Next, a heatmap-based corner detection method is applied, where each corner is represented as a Gaussian distribution. This allows precise detection of feature points, which are further refined using sub-pixel accuracy techniques and outlier removal. Finally, a simplified calibration process based on RANSAC is used to select reliable data and estimate stable parameters. Experimental results show that the method achieves lower reprojection error and better stability compared to traditional approaches. The study concludes that combining learning-based correction with geometric optimization leads to a highly accurate and reliable calibration system.[4]

This paper focuses on estimating both intrinsic and extrinsic camera parameters using a learning-based approach instead of traditional calibration methods. Conventional techniques depend on geometric constraints and calibration patterns, which may not work well in complex real-world conditions. To address this, the authors proposed a multi-task learning framework that integrates camera model equations directly into a neural network. The model is trained using a special loss function called Camera Projection Loss, which helps in improving parameter estimation accuracy. The system uses a deep learning architecture with a feature extractor to analyze input images and predict important camera parameters such as focal length, principal point, and camera position. These parameters are estimated by projecting 2D image points into 3D space and minimizing reconstruction error. The method is tested using both synthetic data generated from simulation and real-world datasets. Experimental results show that the approach achieves better accuracy and lower error compared to traditional and other learning-based methods. The study concludes that combining deep learning with camera geometry provides a more flexible and reliable calibration solution for practical applications.[5]

This work focuses on providing a strong geometric foundation for understanding concepts used in computer vision and computer graphics. Many modern techniques depend on mathematical models, but they are often applied without proper understanding of the underlying geometry. The author aims to connect classical Euclidean geometry with modern projective geometry to improve practical applications. The study explains important concepts such as coordinate transformations, perspective projection, and camera models in a structured manner. It also covers key topics like camera calibration, homography, epipolar geometry, and two-view reconstruction. These concepts help in understanding how 3D scenes are represented and reconstructed from 2D images. Instead of proposing a new algorithm, the work focuses on explaining mathematical relationships that describe image formation and spatial understanding. Examples and derivations are used to show how geometric constraints improve accuracy. The study concludes that a strong understanding of geometry is essential for building reliable and accurate computer vision systems. It highlights that combining theory with practical applications improves performance and interpretability.[6]

This paper focuses on estimating the homography matrix using only a single image of a planar surface. Normally, homography estimation requires multiple images, but this study addresses the challenge of working with limited data. The authors proposed an optimization-based approach that begins by detecting edges in the image using an edge detection technique.

Then, dominant parallel lines are extracted and their intersections are used to form an initial geometric structure. These points are used to compute an initial homography using a linear transformation method. After that, the result is improved using a nonlinear optimization technique to increase accuracy. The method is tested on a dataset of textured surfaces and evaluated using standard performance metrics. The results show that the approach produces accurate and visually correct rectified images even from a single input image. The study concludes that reliable homography estimation is possible without multiple views, making the method useful for applications such as image correction, augmented reality, and image stitching.[7]

This work focuses on improving the performance of 3D computer vision systems by combining geometric methods with deep learning techniques. The goal is to achieve better accuracy, robustness, and efficiency in tasks such as reconstruction and motion analysis. The study mainly addresses three important components, which include camera calibration, rotation estimation, and motion segmentation. It introduces a generalized calibration method that can handle different types of camera distortions effectively. For rotation estimation, an advanced mathematical approach is used to improve accuracy even in the presence of noise and uncertainty. The method also includes an efficient optimization technique that reduces computation time while maintaining reliable results. In addition, a learning-based approach is applied for motion segmentation by grouping similar motion patterns using feature representations. This helps in simplifying complex motion analysis tasks. Experimental results show that the proposed hybrid approach performs better than traditional and purely learning-based methods. The study concludes that combining geometry with learning techniques leads to more accurate, stable, and efficient 3D vision systems.[8]

This paper focuses on improving the accuracy and reliability of camera pose estimation using two images. In many real-world applications, incorrect feature matches and planar scenes create difficulties for traditional methods. The authors proposed a new approach called Linear Relative Pose (LiRP), which is based on pose-only imaging geometry. Instead of relying on complex models, the method uses a linear relationship between translation and depth to simplify calculations. The approach processes multiple feature points at the same time and uses advanced techniques to reduce the effect of incorrect matches. It combines weighted estimation methods and robust algorithms to identify and remove outliers effectively. This helps the system maintain stability even when a large number of incorrect correspondences are present. The method performs well in situations where traditional approaches usually fail, such as pure rotational motion or planar scenes. Experimental results show that the proposed method maintains high accuracy even with a high percentage of outliers. The study concludes that pose-based geometry combined with robust estimation techniques improves performance in challenging vision tasks.[9]

This work focuses on developing efficient methods for understanding 3D environments by combining geometric principles with deep learning techniques. The aim is to improve performance in applications such as autonomous driving, virtual reality, and large-scale scene reconstruction. The authors proposed a framework that integrates multi-view geometry with learning-based models. First, an improved reconstruction pipeline is introduced to generate accurate 3D structures by combining information from multiple views. The method also enhances neural radiance field training by selecting important data points and using geometric information as guidance. This helps in improving both speed and accuracy of the learning process.In addition, a transformer-based approach is used to improve point-cloud representation and enable flexible scaling. The framework also includes model compression techniques to reduce computational cost and make the system more efficient. Experimental results show improved reconstruction quality, faster processing, and better performance on devices with limited resources. The study concludes that combining geometry with learning techniques leads to more scalable and practical 3D perception systems.[10]

This paper focuses on improving image registration accuracy in real-world 3D scenes where depth variations cause alignment errors. Traditional homography methods assume that the scene is flat and apply a single transformation, which leads to incorrect results when objects are at different distances. To solve this issue, the authors proposed a depth-aware framework called DPH-Net. This model uses a deep learning approach to estimate transformations more accurately by considering depth information. The system first predicts camera motion using a neural network structure and then uses a depth map to apply different transformations to each pixel. This allows the model to handle complex scenes more effectively. An additional optimization step is included to refine the alignment by comparing structural similarity between images. This helps in improving the final output quality. Experimental results show that the proposed method achieves better accuracy and lower error compared to traditional and existing deep learning approaches. The study concludes that using depth-aware homography significantly improves performance in non-planar scenes and is useful for applications like image stitching and motion analysis.[11]

This paper focuses on solving the problem of limited datasets available for evaluating homography estimation methods. Many existing approaches are tested on small or unrealistic datasets, which makes it difficult to compare their performance accurately.

To address this issue, the authors introduced a large-scale dataset called Pi3D, which contains real 3D planar surfaces obtained from large reconstruction systems. They also developed a Homography Estimation Benchmark that includes a large number of image pairs with known transformations. The dataset includes variations in viewpoint, lighting, and scene conditions, making it more suitable for real-world evaluation. The authors tested both traditional methods, such as RANSAC-based approaches, and modern deep learning techniques using this benchmark. The evaluation follows a standard training and testing process to ensure fair comparison between different methods. The results show that combining advanced sampling techniques with deep feature matching improves both accuracy and robustness. The study concludes that this benchmark provides a reliable way to compare homography methods and helps in understanding their strengths and limitations for real-world applications.[12]

This paper focuses on estimating homography in situations where traditional feature matching methods do not work well. In applications like sports videos, images often contain occlusions, zoom changes, and varying viewpoints, making feature detection difficult. To solve this problem, the authors proposed a modified algorithm called H-RANSAC. Unlike traditional RANSAC, this method does not depend on matched feature points. Instead, it works with unpaired and featureless points, which makes it more flexible. The method introduces geometric constraints to remove incorrect configurations during the estimation process. It also includes a validation step to eliminate unrealistic transformations after computation. Instead of using descriptors, the approach forms correspondences based on reprojection distance, which improves efficiency and reduces the number of iterations required. Experimental results show that the method achieves lower error and better alignment compared to traditional and advanced feature-based methods. The study concludes that this approach is highly effective in challenging scenarios with large viewpoint changes and limited feature information.[13]

This paper focuses on improving the accuracy and number of feature matches between image pairs. Traditional feature matching methods often fail in complex scenes where there are repeated patterns, depth changes, or occlusions. To overcome these limitations, the authors proposed a multi-stage feature matching approach based on homographic decomposition. Instead of using a single transformation, the method estimates multiple homographies using clustering and iterative refinement. In the first stage, an initial set of matches is refined by repeatedly estimating transformations and grouping similar regions. Then, Delaunay triangulation is applied to remove incorrect matches and improve correspondence accuracy. The method also performs localized matching within smaller regions to detect additional reliable feature points. An extrapolation step is used to identify matches in difficult areas such as occluded regions.

Experimental results show that the approach achieves higher accuracy, better feature density, and lower error compared to traditional and learning-based methods. The study concludes that using multiple homographies significantly improves feature matching performance in complex visual scenes.[14]

This paper focuses on analyzing how different feature detection and matching algorithms affect the accuracy of homography estimation. These methods are important for applications such as image stitching, alignment, and 3D reconstruction. The study compares several commonly used algorithms, including SIFT, SURF, KAZE, ORB, and FAST. Each method is evaluated based on its ability to detect feature points and match them correctly between images. The process involves extracting features, matching them using distance-based techniques, and estimating homography using a robust optimization method. The performance of each algorithm is tested under different conditions and settings. The results show that SIFT provides the most accurate and stable matching results, especially for high-precision tasks. KAZE performs well in preserving image details, while ORB and FAST are faster but less reliable in complex scenes. The study concludes that the choice of algorithm depends on the application requirements. Accurate methods are suitable for detailed analysis, while faster methods are useful for real-time systems.[15]

This paper focuses on improving augmented reality (AR) tracking for users who remain mostly stationary. Traditional SLAM-based methods rely on movement, so they often fail when there is very little motion. To solve this problem, the authors proposed a global tracking framework that combines spherical motion-based SLAM with global localization techniques. This allows the system to track position even when the user is not moving much.The method uses feature extraction techniques to detect important points in images and track them over time. It also estimates camera pose using geometric methods and improves accuracy through optimization techniques. A local map is first created and then aligned with a global model using precomputed data. This helps in maintaining consistency and improving robustness in dynamic environments. Experimental results show that the method provides better tracking accuracy and stability compared to existing systems. The study concludes that combining local tracking with global alignment improves AR performance, especially in scenarios with limited movement.[16]

This paper focuses on improving pose estimation for rolling-shutter cameras, which capture images line by line instead of all at once. Traditional methods assume simple motion patterns, which often fail in real-world situations with complex movement. To address this issue, the authors introduced a new concept called scanline homography.

This method models the transformation for each image row separately, allowing more flexible handling of camera motion. The approach establishes a relationship between scanline homography and traditional plane homography. It uses smooth mathematical functions to model image warping and estimate camera motion accurately. An optimization process is applied to refine the results, starting from an initial approximation. This helps in improving the accuracy of pose estimation and image correction. Experimental results show that the method performs well in both simulated and real-world conditions. It provides accurate pose estimation and effectively corrects rolling-shutter distortions. The study concludes that this approach is reliable and suitable for handling complex camera motion in practical applications.[17]

This paper focuses on developing an automated system to extract traffic information using existing CCTV cameras. The goal is to obtain accurate data such as vehicle count, speed, and movement patterns without installing additional sensors. The authors proposed a three-stage framework for this purpose. First, vehicles are detected using a deep learning model that identifies objects in video frames. Then, homography transformation is applied to convert the camera view into a top-down perspective. This transformation helps in reducing occlusion and provides a clearer view of vehicle movement. After that, tracking algorithms are used to follow vehicles across frames and reconstruct their trajectories. The system also uses a calibration method to convert image measurements into real-world distances. This allows accurate estimation of speed and position. Experimental results show that the method achieves high accuracy with minimal error in traffic measurements. The study concludes that combining homography with deep learning and tracking provides an efficient and cost-effective solution for traffic monitoring.[18]

This paper focuses on improving supervised homography estimation by addressing the problem of unrealistic training datasets. Many existing models are trained on synthetic data, which does not represent real-world conditions accurately. To solve this issue, the authors proposed an iterative deep learning framework that improves both dataset quality and model performance. The approach consists of two main stages: data generation and model training. In the first stage, new training samples are created using real image pairs along with estimated transformations. This helps in generating more realistic data that reflects actual scene conditions. In the second stage, the model is trained using this improved dataset, while filtering out low-quality samples to enhance learning. The process is repeated multiple times, allowing continuous improvement in both data and model accuracy. Experimental results show that the method achieves better performance and lower error compared to traditional and existing learning-based approaches. The study concludes that using realistic datasets and iterative training improves the accuracy and reliability of homography estimation in real-world scenarios.[19]

This paper focuses on improving homography estimation in situations where training data does not have ground-truth labels. Traditional unsupervised methods often struggle in scenes with multiple planes due to parallax effects. To address this issue, the authors proposed a framework called HomoGAN, which combines deep learning with geometric constraints. The method uses a transformer-based model to estimate homography in a step-by-step manner, improving accuracy gradually. A generative adversarial network is also used to identify important regions in the image and focus on the main planar surface. This helps in reducing errors caused by multiple planes in the scene. Additional loss functions are introduced to maintain consistency and stabilize the training process without requiring labeled data. These techniques improve the reliability of the model. Experimental results show that the proposed method performs better than traditional and existing deep learning approaches, especially in challenging conditions such as low texture and poor lighting. The study concludes that combining geometry with adversarial learning improves homography estimation accuracy and robustness.[20]

This paper focuses on improving the accuracy and efficiency of homography estimation between image pairs. This task is important for applications such as image stitching, SLAM, and augmented reality. Traditional methods either rely on feature matching or use deep learning with fixed optimization steps, which limits performance. To overcome this, the authors proposed a fully trainable model called Iterative Homography Network (IHN).The model uses a neural network to extract features from images and calculates relationships between them. It then updates the homography parameters step by step in an iterative manner to improve accuracy. A special version of the model is designed to handle scenes with moving objects by identifying reliable regions. A multi-scale approach is also used to capture both global and local information. Experimental results show that the method achieves lower error and faster processing compared to traditional and existing deep learning methods. The study concludes that iterative refinement improves both accuracy and robustness in homography estimation.[21]

This paper focuses on improving visual place recognition by replacing traditional geometric verification methods with a deep learning approach. Conventional techniques rely on RANSAC, which can be slow and less efficient for large-scale applications. To address this, the authors proposed a transformer-based Deep Homography Estimation (DHE) network. This model analyzes feature maps from images and predicts the transformation needed to align them correctly.

The method uses a special loss function based on reprojection error, allowing the network to learn without requiring labeled homography data. This makes the training process more flexible and efficient. The system follows a two-step process where candidate images are first selected using global features, and then refined using the deep homography model. This improves matching accuracy. Experimental results show that the proposed method achieves higher accuracy and better robustness under different conditions such as lighting and viewpoint changes. It also reduces computation time significantly. The study concludes that deep homography estimation provides a faster and more reliable alternative for place recognition tasks.[22]

This paper focuses on improving homography estimation in difficult scenarios such as low-texture regions, limited overlap, and depth variations. Traditional methods often fail in such cases because they depend heavily on feature matching and cannot capture complex relationships. To overcome these challenges, the authors proposed a deep learning-based multi-grid homography model. Instead of using a single transformation, the method estimates multiple transformations for different regions of the image. The model introduces a special layer to capture relationships between different parts of the image more effectively. This helps in improving both accuracy and computational efficiency. To handle depth variations, the method uses a depth-aware approach that maintains consistency only in regions with similar depth. This reduces errors caused by parallax. The model is trained without labeled data using alignment and structural constraints. Experimental results show improved accuracy, better image quality, and higher robustness compared to traditional and existing deep learning methods. The study concludes that this approach is highly effective for complex real-world image alignment tasks.[23]

This paper focuses on improving unsupervised homography estimation by handling instability and motion-related errors in real-world scenes. Traditional methods often fail when there are moving objects or complex depth variations. To solve this problem, the authors introduced a new representation called homography flow. Instead of directly estimating transformation parameters, the method models motion as a combination of predefined motion patterns. The approach also uses a subspace projection technique to reduce noise and remove unwanted motion effects. This helps in focusing on the main planar motion while ignoring dynamic objects. Additionally, a consistency-based loss function is used to ensure that features remain stable before and after transformation. This improves the reliability of the model during training. The system is implemented using a deep learning architecture and tested on challenging datasets. Experimental results show better accuracy and robustness compared to traditional and previous learning-based methods. The study concludes that combining motion representation with subspace techniques improves homography estimation in complex environments.[24]

This paper focuses on improving homography estimation in difficult environmental conditions such as rain, fog, and low-light scenes. In such cases, traditional feature-based methods fail because image details become unclear and feature detection becomes unreliable. To address this problem, the authors proposed a deep learning framework called SeFENet. The method combines semantic information with structural image features to improve matching accuracy. The model uses a multi-scale feature extraction process to capture both global and local information from images. It also includes a semantic module that identifies important regions and objects in the scene. A special fusion mechanism is used to combine semantic and structural features, which helps in reducing incorrect matches. The system is trained using an adaptive learning strategy to improve performance in degraded conditions. Experimental results show that the method achieves lower error and better robustness compared to existing approaches. The study concludes that integrating semantic information significantly improves homography estimation in challenging real-world environments.[25]

This paper focuses on improving both accuracy and efficiency of homography estimation for applications such as image stitching, SLAM, and video stabilization. Traditional deep learning methods often rely on single-scale refinement, which limits performance and increases computation time. To overcome these issues, the authors proposed a multiscale neural network called MCNet. The model processes image features at different resolutions and refines the homography estimation step by step. The approach uses correlation-based matching to improve feature correspondence between images. It also introduces a special optimization strategy that increases learning efficiency and improves precision during training.

The network is designed to extract features once and then refine results across multiple scales, reducing computational cost. It can also handle dynamic objects without requiring additional processing steps. Experimental results show that the method achieves lower error and faster performance compared to existing approaches. The study concludes that multiscale learning combined with efficient optimization provides accurate and real-time homography estimation for practical applications.[26]

This paper focuses on understanding how deep learning models for homography estimation perform when applied to new datasets that are different from the training data. In many computer vision tasks, models fail under such conditions, but homography models often show good generalization.To study this, the authors designed a lightweight homography estimation network and tested it on multiple datasets with different characteristics.

They also created a special dataset to analyze how the model behaves under various conditions. The experiments included cross-dataset testing and visualization of feature attention to understand what the model learns. The results show that the network mainly depends on local features such as edges and corners rather than high-level semantic information. Further analysis was done by modifying texture properties, which confirmed that the model's performance strongly depends on local geometric details. The study concludes that deep homography models are naturally robust to domain changes because they rely on consistent texture features. This makes them behave similarly to classical geometric methods while maintaining the advantages of deep learning.[27]

This paper focuses on improving homography and motion estimation in challenging environments such as rain, fog, and low-light conditions. Traditional methods often fail in such situations because they rely heavily on image brightness and texture information. To address this problem, the authors proposed an unsupervised deep learning framework that combines gyroscope data with image features. The gyroscope provides reliable motion information independent of visual quality. The method converts gyroscope readings into a motion representation, which helps in capturing global camera movement. This information is then combined with optical flow obtained from images to improve motion estimation accuracy.A fusion module is used to integrate both sources of information, and a decoder estimates the final homography transformation. The model is trained without labeled data using consistency-based learning. Experimental results show that the approach achieves better accuracy and robustness compared to traditional and existing methods, especially in difficult conditions. The study concludes that combining sensor data with deep learning improves performance in real-world vision tasks.[28]

This paper focuses on reconstructing 3D planar surfaces from a single image in real-world environments. Many existing methods work only for specific datasets, such as indoor or outdoor scenes, and do not generalize well across different conditions. To overcome this limitation, the authors proposed a transformer-based model that can learn plane geometry from diverse datasets. The model is trained using a large collection of images from multiple sources to improve generalization. The approach uses a two-step strategy where it first identifies plane orientation and then estimates its position. This helps in achieving stable and accurate reconstruction results. In addition, the model incorporates depth and surface information to better understand the structure of the scene. This improves its ability to detect and reconstruct planes correctly. Experimental results show that the method performs better than existing approaches, especially in unseen environments. The study concludes that combining large-scale training with geometry-aware learning improves the accuracy and reliability of 3D plane reconstruction.[29]

This paper focuses on improving feature detection and matching between images captured using different camera types, such as fisheye and perspective cameras. Traditional feature detection methods often fail in such cases due to strong distortions present in fisheye images. To address this problem, the authors proposed a self-supervised deep learning framework inspired by existing feature detection models. The method uses a shared neural network to extract features and separate modules to detect keypoints and generate descriptors. A hybrid homography model is introduced to relate fisheye and perspective images, allowing the system to learn correspondences between different camera geometries. The model is trained using both synthetic and real datasets to improve generalization. A contrastive learning approach is applied to ensure that corresponding points from different camera types produce similar feature representations. This improves matching accuracy and consistency. Experimental results show that the method performs better than traditional and existing deep learning approaches, especially in highly distorted images. The study concludes that this approach provides a robust solution for feature matching in multi-camera systems used in applications like SLAM and 3D reconstruction.[30]

IV. COMPARATIVE STUDY OF FIVE PUBLISHED RESEARCH PAPERS

The study analyzed five significant research papers related to camera models, calibration techniques, and homography estimation in computer graphics and computer vision. The table highlights the proposed solutions, methodologies, and outcomes of each paper. This comparison helps in identifying which techniques provide better accuracy, robustness, and efficiency in real-world applications. It also gives a clear understanding of the strengths and limitations of different methods.

Table 1: Comparison of Selected Five Research Papers

S. No	Title of Papers	Year	Proposed Solution	Methodology	Conclusions
1	Gaussian Process Camera Calibration	2023	Convert real camera into ideal pinhole model	Uses Gaussian Process to correct distortion and map coordinates	Improves calibration accuracy and reduces distortion errors

2	LiDAR-Camera Calibration Toolbox	2023	Automatic target-less calibration	Uses feature matching and optimization between LiDAR and camera data	Provides accurate and robust sensor alignment
3	Multi-Task Learning Calibration	2022	Estimate camera parameters using deep learning	Neural network with camera projection loss and feature extraction	Improves accuracy and generalization
4	Depth-Aware Homography Estimation	2023	Pixel-wise transformation using depth information	Deep learning with depth maps and image alignment optimization	Handles non-planar scenes effectively
5	MCNet Homography Estimation	2024	Efficient multi-scale homography estimation	Uses correlation learning and multi-scale refinement	Provides fast and accurate real-time results

V. CONCLUSION

This review paper presents a detailed study of camera models and projection techniques used in computer graphics and computer vision. It focuses on important concepts such as camera calibration, homography estimation, and 3D reconstruction. From the analysis of different research papers, it is observed that traditional geometric methods provide a strong theoretical foundation for understanding image formation and transformation. However, these methods often face challenges in complex real-world conditions such as noise, distortion, and varying lighting. Modern approaches based on machine learning and deep learning have shown significant improvements in accuracy and robustness. These methods are capable of handling complex scenarios and large datasets more effectively than classical techniques. The comparison of selected research papers shows that hybrid approaches, which combine geometric modeling with learning-based techniques, provide better performance in terms of accuracy, efficiency, and adaptability. In conclusion, the integration of mathematical foundations with modern learning methods plays a crucial role in developing reliable and efficient computer vision systems. These advancements will continue to improve applications such as augmented reality, robotics, and 3D scene understanding.

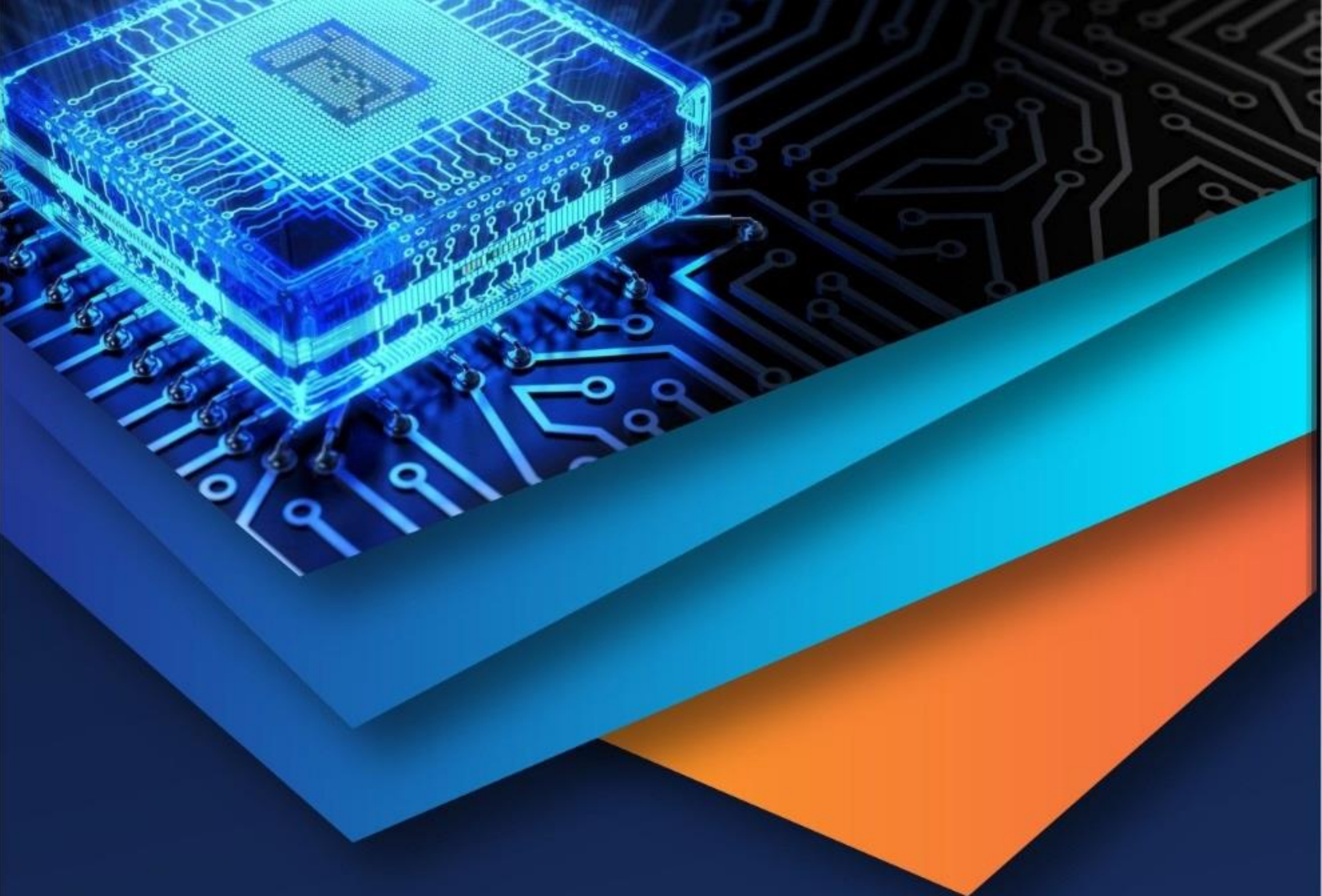
VI. FUTURE SCOPE

Future research in camera models and projection techniques can focus on developing more efficient and real-time solutions for practical applications. As computer vision systems are increasingly used in dynamic environments, there is a need for methods that can handle complex conditions such as noise, occlusion, and lighting variations more effectively. One important direction is the integration of advanced deep learning models such as transformers and hybrid networks with traditional geometric approaches. This combination can further improve accuracy and adaptability in tasks like 3D reconstruction, image alignment, and camera calibration. Another area of improvement is multi-sensor fusion, where data from cameras, LiDAR, and inertial sensors can be combined to achieve better performance in applications such as autonomous driving and robotics. Additionally, future work can focus on developing lightweight and computationally efficient models that can be implemented on mobile and embedded devices without compromising performance. Overall, the future of this field lies in creating scalable, robust, and intelligent systems that can operate effectively in real-world scenarios across various domains.

REFERENCES

[1] De Boi, I., Pathak, S., Oliveira, M., & Penne, R. (2023). How to turn your camera into a perfect pinhole model. arXiv preprint. https://arxiv.org/abs/2309.11326
[2] Koide, K., Oishi, S., Yokozuka, M., & Banno, A. (2023). General, single-shot, target-less, and automatic LiDAR-camera extrinsic calibration toolbox. arXiv preprint.
[3] https://arxiv.org/abs/ 2302.05094
[4] Pinchukov, V., Shmatko, E., Poroykov, A., & Bogachev, A. (2021). Computer and physical modeling for verification of the stereo system calibration parameters optimization results. CEUR Workshop Proceedings, 3027. https://ceur-ws.org/Vol-3027/paper41.pdf

[5] Zhang, Y., Zhao, X., & Qian, D. (2023). Learning-based framework for camera calibration with distortion correction and high precision feature detection. arXiv preprint. https://arxiv.org/abs/2202.00158
[6] Butt, T. H., & Taj, M. (2022). Multi-task learning for camera calibration. arXiv preprint.
[7] https://arxiv.org/abs/2211.12432
[8] Pajdla, T. (2025). Elements of geometry for computer vision and computer graphics [Lecture notes]. Czech Technical University in Prague. https://cw.fel.cvut.cz/b242/_media/courses/gvg/pajdla-gvg-lecture-2025.pdf
[9] Kumari, P., & Das, S. (2023). Single view homography estimation for an inclined textured planar surface: Overcoming the inverse and ill-posed challenge. In Proceedings of the ACM Conference. https://doi.org/10.1145/3627631.3627633
[10] Lochman, Y. (2025). Geometry and learning in 3D computer vision [Master's thesis, Chalmers University of Technology]. https://research.chalmers.se/publication/549335/file/549335_Fulltext.pdf
[11] Cai, Q., Li, X., & Wu, Y. (2024). Linear relative pose estimation founded on pose-only imaging geometry. arXiv preprint. https://arxiv.org/pdf /2401.13357
[12] Orsingher, M. (2023). Geometry and learning for efficient 3D perception [Master's thesis, Università degli Studi di Parma]. https://www.repository.unipr.it/bitstream/1889/5572/4/tesi.pdf
[13] Huang, C., Pan, X., Cheng, J., & Song, J. (2023). Deep image registration with depth-aware homography estimation. IEEE Signal Processing Letters. SCI_Q2_2023_Huang_SPL_DeepImageRegistration.pdf - Google Drive
[14] Barath, D., Mishkin, D., Polic, M., Förstner, W., & Matas, J. (2023). A large-scale homography benchmark. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). https://openaccess.thecvf.com/content/CVPR2023/papers/Barath_A_Large-Scale_Homography_Benchmark_CVPR_2023_paper.pdf
[15] Nousias, G., Delibasis, K. K., & Maglogiannis, I. G. (2025). Intelligent sampling consensus for homography estimation in football videos using featureless unpaired points. IEEE Access. https://doi.org/10.1109/ACCESS.2025.3627538
[16] Seibt, S., Lipinski, B. V. R., & Latoschik, M. E. (2022). Dense feature matching based on homographic decomposition. IEEE Access. https://doi.org/10.1109/ACCESS.2022.3152539
[17] Qiu, X. (2024). Comparison and application of implementing image homographs in computer vision. In Proceedings of the International Conference on Image, Algorithms and Applications. Atlantis Press.https://www.atlantis-press.com/article/126004121.pdf
[18] Baker, L., Ventura, J., Langlotz, T., Gull, S., Mills, S., & Zollmann, S. (2024). Localization and tracking of stationary users for augmented reality. The Visual Computer, 40, 227–244. https://doi.org/10.1007/s00371-023-02777-2
[19] Bai, F., Sengupta, A., & Bartoli, A. (2022). Scanline homographies for rolling-shutter plane absolute pose. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 12345–12354). https://openaccess.thecvf.com/content/CVPR2022/html/Bai_Scanline_Homographies_for_Rolling-Shutter_Plane_Absolute_Pose_CVPR_2022_paper.html
[20] Zhang, L., Yu, X., Daud, A., Mussah, A. R., & Adu-Gyamfi, Y. (2024). Application of 2D homography for high resolution traffic data collection using CCTV cameras. arXiv preprint. https://arxiv.org/abs/2401.07220
[21] Jiang, H., Li, H., Han, S., Fan, H., Zeng, B., & Liu, S. (2023). Supervised homography learning with realistic dataset generation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) https://openaccess.thecvf.com/content/ICCV2023/html/Jiang_Supervised_Homography_Learning_with_Realistic_Dataset_Generation_ICCV_2023_paper.html
[22] Hong, M., Lu, Y., Ye, N., Lin, C., Zhao, Q., & Liu, S. (2022). Unsupervised homography estimation with coplanarity-aware GAN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). https://openaccess.thecvf.com/content/CVPR2022/html/Hong_Unsupervised_Homography_Estimation_With_Coplanarity-Aware_GAN_CVPR_2022_paper.html
[23] Cao, S.-Y., Hu, J., Sheng, Z., & Shen, H.-L. (2022). Iterative deep homography estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). https://openaccess.thecvf.com/content/CVPR2022/html/Cao_Iterative_Deep_Homography_Estimation_CVPR_2022_paper.html
[24] Lu, F., Dong, S., Zhang, L., Liu, B., Lan, X., Jiang, D., & Yuan, C. (2024). Deep homography estimation for visual place recognition. In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI-24). https://ojs.aaai.org/index.php/AAAI/article/view/28901
[25] Nie, L., Lin, C., Liao, K., Liu, S., & Zhao, Y. (2021). Depth-aware multi-grid deep homography estimation with contextual correlation. arXiv. https://arxiv.org/abs/2107.02524
[26] Ye, N., Wang, C., Fan, H., & Liu, S. (2021). Motion basis learning for unsupervised deep homography estimation with subspace projection. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).
[27] Motion Basis Learning for Unsupervised Deep Homography Estimation With Subspace Projection
[28] Shi, Z., Zhang, Z., Cui, K., An, R., Liu, J., & Jiang, Z. (2025). SeFENet: Robust deep homography estimation via semantic-driven feature enhancement. arXiv.https://arxiv.org/abs/2412.06352
[29] Zhu, H., Cao, S.-Y., Hu, J., Zuo, S., Yu, B., Ying, J., Li, J., & Shen, H.-L. (2024). MCNet: Rethinking the core ingredients for accurate and efficient homography estimation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).MCNet: Rethinking the Core Ingredients for Accurate and Efficient Homography Estimation
[30] Shao, M., Tasdizen, T., & Joshi, S. (2024). Analyzing the domain shift immunity of deep homography estimation. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV).Analyzing the Domain Shift Immunity of Deep Homography Estimation
[31] Li, H., Luo, K., Zeng, B., & Liu, S. (2023). GyroFlow+: Gyroscope-guided unsupervised deep homography and optical flow learning. arXiv. https://arxiv.org/abs/2301.10018
[32] Liu, J., Yu, R., Chen, S., Huang, S. X., & Guo, H. (2025). Towards in-the-wild 3D plane reconstruction from a single image. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).Towards In-the-wild 3D Plane Reconstruction from a Single Image
[33] Mera-Trujillo, M., Patel, S., Gu, Y., & Doretto, G. (2023). Self-supervised interest point detection and description for fisheye and perspective images. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).Available at: Self-Supervised Interest Point Detection and Description for Fisheye and Perspective Images



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY