EAI Endorsed Transactions
on Internet of Things

Research Article **EAI.EU**

Real-Time Panoramic Image Stitching Framework based on IoT for Intelligent Surveillance Systems

Bittu Kumar[1*], Rani Swarna Teja[2] and M. Shailaja[3]

[1]Department of Electronics and Communication Engineering, Koneru Lakshmaiah Educational Foundation, Hyderabad 500075, India
[2]Department of Electronics and Communication Engineering, Malla Reddy University, Hyderabad-500100, India
[3]Department of Data Science, Malla Reddy University, Hyderabad-500100, India

Abstract

INTRODUCTION: Panoramic image stitching is a fundamental technique in intelligent vision systems for generating wide field-of-view images from multiple overlapping images. The integration of panoramic imaging with Internet of Things (IoT) technology has enabled advanced applications such as intelligent surveillance, robotics, autonomous monitoring, and smart city systems; however, real-time panoramic image generation on resource-constrained IoT devices remains challenging due to image noise, motion blur, illumination variations, and limited computational resources.
OBJECTIVES: This study aims to develop a lightweight and efficient IoT-based panoramic image stitching framework that achieves robust stitching accuracy while maintaining low computational complexity for real-time edge-assisted vision applications.
METHODS: The proposed framework integrates image acquisition using an ESP32-CAM module with cloud-based image processing and employs ORB, SIFT, BRISK, and a Hybrid ORB–SIFT feature extraction approach for keypoint detection, descriptor generation, feature matching, homography estimation, image warping, and panoramic image reconstruction. Performance was evaluated on overlapping image datasets corrupted with Gaussian noise, Salt-and-Pepper noise, and motion blur using Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), execution time, feature matching statistics, and keypoint analysis.
RESULTS: Experimental results demonstrate that SIFT provides superior stitching robustness and panorama quality under noisy conditions, whereas ORB offers significantly lower computational complexity and faster execution. The proposed Hybrid ORB–SIFT framework effectively combines the strengths of both methods, achieving 1335 valid feature matches with an execution time of 1.596 s, while maintaining high stitching accuracy and robustness under challenging imaging conditions.
CONCLUSION: The proposed IoT-enabled Hybrid ORB–SIFT panoramic image stitching framework provides an effective balance between computational efficiency and stitching accuracy, making it a reliable and lightweight solution for real-time panoramic surveillance, autonomous monitoring, and next-generation IoT-based intelligent vision applications.

Keywords: Panorama Stitching, IoT, ESP32-CAM, Feature Extraction, ORB, SIFT, BRISK, Smart Surveillance, Edge Computing, Image Processing.

Received on 30 June 2026, accepted on 08 September 2026, published on 17 September 2026

doi: 10.4108/eetiot.13826

*Corresponding author. Email: bittu.mlrit@gmail.com

1. Introduction

Intelligent imaging technologies have made a great leap forward, revolutionizing the way visual information is captured, processed and interpreted in today's digital systems. In computer vision, panoramic image stitching is an interesting technique that obtains panoramic visualization from a set of overlapping images. Image stitching is the technique of composing multiple partially aligned images into a single panoramic image, thereby creating a wider field of view than a single imaging device can achieve and facilitating the understanding of the scene in a complicated environment [1]. Because of its comprehensive visual perception ability, panoramic Imaging has been gradually paid attention to in both research and the field of intelligent monitoring.

At the same time, the Internet of Things (IoT) has driven the expansion of the Internet of Things (IoT) sensor networks, which gather and share visual data in real time. The integration of low-cost imaging systems, wireless communication modules, cloud platforms, and edge computing technologies make it possible to implement automatic monitoring and intelligent decision-making processes in the context of IoT [2]. By combining panoramic imaging with IoT systems, a wider scene can be captured, greater awareness of the situation can be achieved and better surveillance can be achieved using resource-limited devices like ESP32-CAM modules and edge nodes.

The applications of panoramic image stitching have been rapidly expanded to smart cities, independent systems, surveillance systems, robot systems, industrial monitoring, drone inspection, and remote sensing. Panoramic imaging plays a vital role in smart city applications, such as intelligent traffic management, crowd monitoring, and urban surveillance. Panoramic scene understanding is used for autonomous robots and navigation systems for localization and environment mapping. Likewise, large-area observation and anomaly detection of the aerial monitoring platforms and industrial inspection systems can be achieved by stitched panoramic views [3]. The need is to use reliable techniques of panorama generation that can work efficiently in dynamic real world conditions for the above applications.

Although image stitching has advanced significantly in recent years, there are still several issues which impact the quality and reliability of panorama generation. Feature extraction and matching are also prone to inaccurate due to image noise such as Gaussian noise and Salt-and-Pepper noise, which can cause unstable panorama reconstruction [4]. Additionally, due to the motion of the camera and/or objects, the distinctiveness of features is further diminished and it can affect the estimation of homography. Further, the distortions and artifacts caused by illumination variations, viewpoint change, image misalignment and moving objects in the scene are common in the resulting panorama [5]. The problem is even more important in scenario-based monitoring systems with IoT which capture images in uncontrolled environments.

Another crucial consideration is computational complexity. While traditional feature extraction methods like SIFT have highly distinctive and robust features, they are extremely memory and CPU intensive [6]. On the other hand, lightweight approaches like ORB are fast processing and have a small computational cost but have limited robustness under difficult imaging conditions [7]. A compromise of robustness and efficiency is provided by BRISK, which can be represented by binary features and extract the scale-invariant keypoints [8]. Therefore, it is difficult to choose one feature extraction algorithm for all the above-mentioned factors due to the potential compromise between accuracy of stitching and real-time.

There has been a growing interest in image stitching with AI, deep learning for estimating homography, vision systems powered by cloud computing and vision systems based on edge intelligence for image stitching between 2024 and 2026 [9]. These methods have shown some performance gains, but most of them involve complex training and computational procedures, and often demand specialized hardware accelerators and large training sets. This may not fit the requirements of lightweight IoT surveillance systems where low latency, low computation and deployability are necessary.

To overcome these problems, this paper suggests a Noise-Aware Hybrid ORB-SIFT Panorama Stitching Framework for IoT-based intelligent surveillance system. The proposed framework uses ORB in conjunction with SIFT to achieve both computational efficiency and robustness in the panorama generation process while dealing with noisy imaging conditions. The acquisition and transmission of images using the ESP32-CAM, Cloud processing, Adaptive feature extraction, feature matching, estimation of homographies and panoramic reconstruction are implemented in a single IoT system. The combined power of ORB and SIFT will be used to create an efficient yet robust stitching system in resource-limited surveillance applications.

The key novelty of the proposed work lies in the selective ORB–SIFT feature extraction mechanism, in which ORB is used for fast keypoint localization and SIFT descriptors are computed only for the ORB-selected key-points. Unlike conventional approaches that independently perform complete ORB and SIFT extraction or directly combine multiple detector–descriptor outputs, the proposed method avoids the computationally expensive SIFT key-point detection stage while retaining the discriminative capability of SIFT descriptors. This design is particularly intended for resource-constrained IoT-assisted panoramic stitching, where both computational latency and feature-matching reliability are critical. The novelty is further strengthened by integrating this selective feature extraction mechanism with ESP32-CAM image acquisition, cloud-assisted processing, RANSAC-based homography estimation, and systematic robustness evaluation under Gaussian noise, Salt-and-Pepper noise, and motion blur. The proposed hybrid framework is evaluated with respect to ORB, SIFT, BRISK descriptors using feature matching statistical analysis, execution time, PSNR and SSIM metrics.

2. Literature Review

The technique of image stitching that can create panoramic views from a set of overlapping images has been studied extensively in computer vision and image processing, robotics and intelligent surveillance. The traditional panorama stitching methods are mainly based on feature detection, feature matching, geometric transformation, and blending techniques for generating seamless panoramic images. Among these, the one proposed by Brown and Lowe [10] stands out as one of the most influential works in this field, which proposed an automatic panorama stitching system using invariant local features. Their work showed that feature correspondence and homography estimation can be used to provide good panorama generation and subsequently was a basis for many feature based stitching methods. Szeliski [11] also summarized the image alignment and stitching techniques, including feature extraction, geometric transformations, and image warping and blending methods.

In panorama stitching, feature extraction is important due to the direct relationship between the accuracy of feature detection and matching and the quality of the resulting panorama image. SIFT [6] is still one of the most popular feature descriptors in its ability to be invariant to scale, rotation and some amount of illumination change. While SIFT offers unique and useful descriptors and good feature matching rate, it tends to be computationally expensive, restricting its use in real-time and resource-limited applications. To overcome these drawbacks, Rublee et al. [12] proposed a method called ORB that uses both FAST keypoint detection and BRIEF descriptors to obtain much faster keypoint extraction and matching with an acceptable robustness.

Several comparative studies on the performance of the ORB algorithm, SIFT algorithm and BRISK algorithm are performed for the image stitching application. In [13], authors found that the matching accuracy and quality of stitching of SIFT is better than ORB, while ORB has lower computational overhead and faster execution time. Khan and Ali [14] also showed that descriptor robustness can be different under different imaging conditions. This further emphasized the need for testing the feature extraction methods for degraded imaging environments.

The Internet of Things (IoT) has increased the possibilities of panoramic imaging in intelligent surveillance, smart cities, industrial monitoring and autonomous systems. The modern surveillance system using the IoT will allow real-time visual monitoring and automated decision making, and has been developed using low-cost image sensors, wireless communication technologies, cloud platforms and edge computing frameworks. Wang et al. [15] presented edge computing-based surveillance architecture with the aid of IoT, enabling low latency image processing, the studies show how the efficiency of the image processing techniques is gaining importance with the integration in the IoT infrastructures.

ESP32-CAM modules are also gaining a lot of interest for low-cost IoT imaging solutions due to their camera interface, Wi-Fi connectivity, low power consumption, and compact design. Agarwal and Kumar [16] presented an intelligent surveillance system based on ESP32-CAM. This works established the image acquisition architectures of IoT, but most of them focused on the image acquisition and image transfer issues from the viewpoint of image capture, while not considering image panorization and robustness analysis.

Visual processing is one of the important areas of research that has been identified to address the computational constraints of the lightweight IoT devices, with the assistance of cloud and edge. The authors of [17] presented a lightweight image stitching framework for embedded platforms and pointed out the significance of computational efficiency in real-time applications. Other edge assistant processing frameworks have also been made use of to enhance scalability and decrease edge processing load. Sharma et.al [18] introduced a cloud supported panorama stitching framework for surveillance systems in smart monitoring systems.

Recently, with the progress of deep learning, the research of panoramic image stitching has been influenced further. To enhance the performance of features representation and alignment, Song et al. [19] designed a weakly supervised panorama stitching network that are able to handle complex geometric transformations and challenging scene conditions. To stitch images taken in different lighting conditions, Chen and Lin [20] created a deep learning-based stitching framework. Although these methods lead to improved stitching performance, they generally require a full training dataset, a large amount of computation, and the use of specialized hardware accelerators, which may not be suitable for lightweight IoT surveillance configuration.

From the literature review, it can be seen that the research published so far has been mainly on the high accuracy of the panorama generation or on the IoT surveillance system separately. While ORB, SIFT and BRISK have been widely studied, little work has been focused on introducing a hybrid framework of feature extraction that can simultaneously provide computational efficiency and robustness in noisy imaging scenarios. Additionally, there are few of the studies that have combined the techniques of hybrid panorama stitching with architectures of IoT surveillance system using ESP32-CAM and cloud computing platforms. Many deep learning models are also trained using extensive datasets, which adds to the complexity and resource requirements of training. Existing deep learning solutions generally depend on larger training sets and computational resources, making them unsuitable for lightweight edge deployments [21-23].

Table 1. Comparative Analysis of Existing Literature

Method	Dataset	IoT/Cloud Support	Limitation
SIFT-based panorama stitching	Overlapping image datasets	No	High computational complexity
ORB feature extraction	Real-time vision datasets	No	Reduced robustness under severe noise
BRISK descriptors	Vision benchmark datasets	No	Moderate matching stability
ORB, SIFT, BRISK comparison	Real-time image pairs	No	Limited noisy environment evaluation
Edge-based IoT surveillance	Smart surveillance datasets	Yes	No panorama generation analysis
ESP32-CAM monitoring framework	IoT camera images	Yes	Limited feature extraction analysis
Lightweight edge stitching	Embedded vision datasets	Yes	Limited robustness evaluation
Deep stitching network	Panorama datasets	Cloud-based	High computational requirements
Transformer homography estimation	Dynamic scene datasets	Cloud-assisted	GPU dependency
Cloud-assisted panorama stitching	Surveillance datasets	Yes	Limited noise analysis
Noise-robust feature matching	Noisy image datasets	No	No IoT integration
Edge AI surveillance	Smart city datasets	Yes	Limited panorama stitching focus

The comparative analysis presented in Table 1 highlights the strengths and limitations of existing feature-based and IoT-assisted panoramic image stitching methods. It is evident that no single approach simultaneously addresses lightweight computation, robustness against noise and motion blur, and seamless IoT/cloud integration, thereby justifying the need for the proposed Hybrid ORB–SIFT panoramic stitching framework. Hence, the need for an efficient, noise mitigation and IoT-enabled panorama stitching framework that can be used for real-time intelligent surveillance applications is still an open research challenge. To fill this void, the present work is introduced with a Noise-Aware Hybrid ORB-SIFT Panorama Stitching Framework, which integrates the computation efficiency of ORB and the robustness of SIFT into an IoT setup for surveillance.

3. A Proposed Panorama Stitching Framework for IoT

This section introduces the proposed panoramic image stitching framework for intelligent surveillance application using IoT. The framework combines the image acquisition with ESP32-CAM, the image processing in the cloud, the features extraction, feature matching, homography estimation and panorama generation. Contrary to the traditional stitching technique, the presented technique adopts an Internet of Things (IoT) architecture to acquire and transmit images in real time and assesses several feature extraction methods such as ORB, SIFT, BRISK and a proposed Hybrid ORB–SIFT method. In addition, the framework looks at the noise tolerance of these descriptors in order to determine their applicability to edge assisted vision systems and real-time surveillance.

3.1 The ESP32-CAM microcontroller

This image acquisition process is carried out with the ESP32-CAM module which has OV2640 camera sensor and Wi-Fi built-in. The device takes several images of the same scene being monitored from different angles. An overlap of about 30-50% is kept between consecutive photographs to assure successful panoramic reconstruction. Images are captured and sent wirelessly, via Wi-Fi, to a processing environment with assistance from a cloud system. The acquisition strategy is based on the IoT and allows for monitoring remotely and image collection in real-time without the use of costly imaging hardware. The acquired images are saved in the cloud and then processed via the proposed stitching framework. By combining ESP32-CAM with cloud computing, image acquisition can be easily scaled and is low cost and with minimal deployment complexity.

3.2 Image Pre-processing

All acquired images are pre-processed before feature extraction in order to enhance the stitching accuracy and computational efficiency. In the first step, images are resized to a common resolution to minimize cost of processing, but with enough information about the scene to be useful. The images are then resized and also converted to gray scale images as feature extraction algorithms mainly work with intensity details. Image normalization is also carried out to minimize the illumination differences and maximize the consistency of features in overlapped image areas. The above described pre-processing operations make the descriptors more stable, and enhance the accuracy of matching and homography estimation.

3.3 ORB Feature Extraction

Oriented FAST and Rotated BRIEF (ORB) is a fast computer vision feature extraction algorithm. The low cost of computation, low memory usage, and binary representation of the descriptor have made ORB a very popular approach to image stitching, visual localization, robotic navigation, and IoT-powered surveillance systems. The Figure 1 shows ORB Feature Extraction Process with FAST Corner Detection and BRIEF Descriptor Generation. As part of the proposed framework, ORB is used to detect and locate the unique feature points in an overlaid image, before feature matching and panorama generation. The ORB algorithm integrates two techniques, the FAST (Features from Accelerated Segment Test) corner detector and BRIEF (Binary Robust Independent Elementary Features) descriptor. First, FAST identifies salient corner points by checking the differences between the intensity of a potential corner point and the intensity of its adjacent pixels. A pixel is defined to be a corner if the intensity difference is above a certain threshold. The FAST corner detection criterion is given by:

$$| I(p) - I(n) | > T$$

Where, I(p) is the intensity of the candidate pixel, and I(n) is the intensity of a neighbor pixel, while (T) is the threshold for corner detection. The method allows for a quick determination of areas of the image that contain features but with minimal computational cost. Once the keypoints are detected, ORB calculates the orientation of each feature point making the algorithm rotation-invariant. This is estimated using image moments and calculated as:

$$\theta = \tan -1 \frac{m_{01}}{m_{02}}$$

The orientation angle (θ) of the keypoint is used to determine the orientation of the image patch, while the image moments of the first order (m_{01}) and (m_{02}) are calculated around the image patch. This orientation task enhances the effectiveness of feature matching when images are taken at different angles. After the orientation estimation, ORB uses the BRIEF descriptor to create a short binary representation of the keypoints detected. BRIEF constructs descriptors by comparing the intensity of selected pairs of pixels at the local region of an image. The following is a representation of the descriptor:

$$\mathrm{D} = \{\mathrm{b1}, \mathrm{b2}, \mathrm{b3}, \ldots, \mathrm{bn}\}$$

where (D) is the binary descriptor vector and (b_n) is the result of a single intensity comparison. The binary representation makes the description process much faster and helps in reducing the storage requirements for descriptors. The Hamming distance metric is used for feature matching between ORB descriptors, which is defined as the number of bits that are different between two binary descriptors. The Hamming distance is computed as:

$$H(D1, D2) = \sum_{i=1}^{n} (D_{1i} \oplus D_{2i})$$

The two ORB descriptors are called D1 and D2, the XOR operation is indicated by (⊕), and n denotes the length of the ORB descriptors. The Hamming distance is smaller, means more similarity of feature descriptor and more accurate matching. By combining each stage of the FAST-based keypoint detection, orientation assignment, and BRIEF descriptor generation, ORB can rapidly extract features while still possessing a good matching capability. Because of its light weight architecture and low calculation of ORB, it is especially suitable for panoramic image stitching in real-time in resource-constrained IoT and edge-computing scenes.

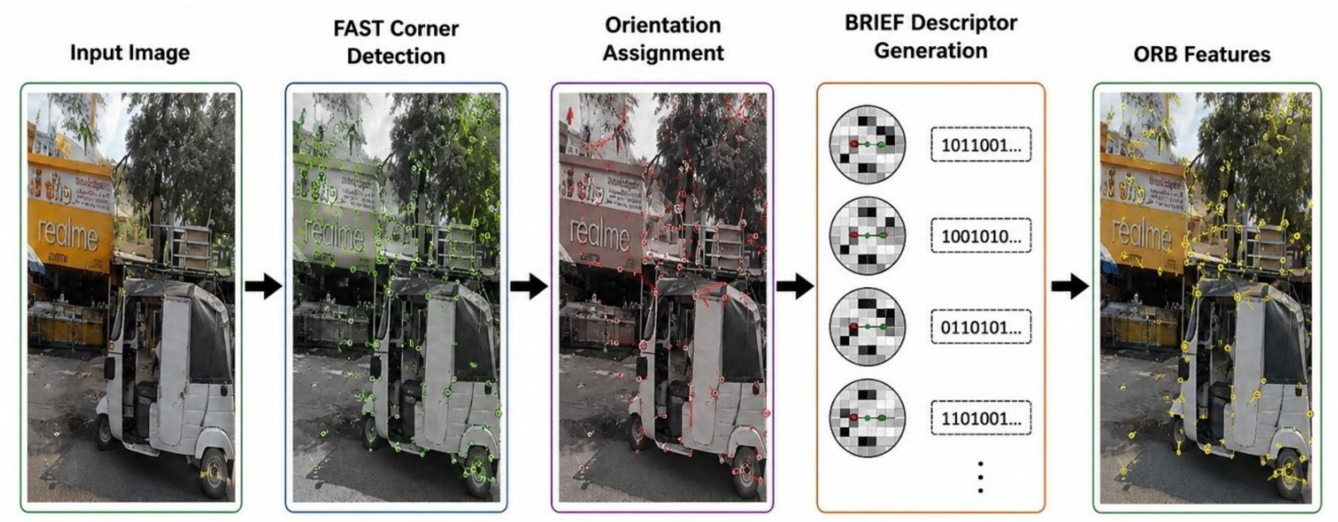

Figure 1. ORB Feature Extraction Process with FAST Corner Detection and BRIEF Descriptor Generation

3.4 SIFT Feature Extraction

The Scale-Invariant Feature Transform (SIFT) algorithm is a powerful feature extraction method that is used in various computer vision tasks, such as object recognition, image stitching, and more. While the binary descriptors are light in weight, the SIFT descriptors are very distinctive, and they are not easily changed by changes in scale, rotation, illumination, and small changes in viewpoint. The above features make SIFT a suitable method for panoramic image stitching, where accurate feature correspondence between the two overlapping images is crucial for successful panorama generation. SIFT is done by first creating a scale space representation of the image to be processed using Gaussian filtering. This process helps to find image structure at various scales and also helps to find scale-invariant keypoints. The Gaussian scale space representation is given by:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

In this formula, (G(x,y,σ)) is the Gaussian kernel, with scale parameter (σ), and (*) is the convolution operation, and where (I(x,y)) is the input image. SIFT can be used to recognize features that are stable in the image when it is resized and zoomed at various scales. SIFT calculates the Difference of Gaussian (DoG) between neighboring scales to find candidate key-points. DoG is defined as:

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$$

Where, k is a constant scale factor. Pots of local extrema found in the DoG scale space are chosen as potential keypoints, and then processed in order to remove unstable feature locations. After localization of keypoints, SIFT finds the dominant orientation of each feature point using local gradient information. This orientation assignment is able to give rotation invariance and make it possible to match the images taken from various positions. Lastly, a descriptor vector is created to compare the distribution of gradients in the vicinity of each keypoint. This descriptor is able to represent the local image properties and is extremely discriminative for feature matching. With the strong descriptor generation capability, SIFT is usually more reliable in generating the feature correspondences than the lightweight feature extraction methods. This is, however, at the expense of more complicated calculation and processing time. Therefore, SIFT may be suitable for directly applying in resource-challenged IoT devices, but the resolution of the panorama stitching may be poor. SIFT is therefore more extensively used in the proposed Hybrid ORB–SIFT framework where it is combined with the computational efficiency of ORB, while retaining the robustness it offers.

3.5 BRISK Feature Extraction

Binary Robust Invariant Scalable Keypoints (BRISK) is a feature extraction method that offers a good compromise between speed and matching strength. Like ORB, BRISK uses binary descriptors, which allow fast feature matching and invariant to scale and rotation transformations. BRISK is an ideal tool for real-time computer vision applications such as image stitching, visual localization, object tracking and IoT monitoring systems for surveillance. BRISK algorithm starts by creating a multi-scale image pyramid to determine stable keypoints in images at various scales. This scale-space representation allows to identify feature points that are invariant to the scale change operation. BRISK detects the key points and uses local intensity information from a pre-defined circular sampling pattern to estimate the orientation of each key point. The improvement of descriptor robustness to image rotations and variations of viewpoints is obtained by the orientation assignment. Following the estimation of orientation, BRISK creates a binary descriptor based on the intensities of the selected pairs of sampling points around each keypoint. The process of generating descriptors can be described as:

$$b_i = \begin{cases} 1, & I(p_i) > I(q_i) \\ 0, & \text{otherwise} \end{cases}$$

where (b_i) is the binary descriptor bit, $I(p_i)$ and $I(q_i)$ are intensity values of two sampling points. The binary descriptor embraces local image characteristics, and can be efficiently matched by calculating the Hamming distance. The main benefit of BRISK is that it is scale and rotation invariant with a relatively low computational cost. BRISK is more memory efficient than conventional floating-point descriptors like SIFT and can match descriptors at a faster rate. BRISK is based on the local intensity comparisons in descriptor generation, however, it may not be robust with high level of image degradation, especially noise and motion blur. BRISK is used in the proposed framework as a benchmark feature extraction method for assessing its stitching performance under normal and degraded imaging conditions. The performance of BRISK is then evaluated by ORB, SIFT and the proposed Hybrid ORB–SIFT by feature matching statistics, panorama quality metrics and computational complexity analysis.

3.6. Proposed Hybrid ORB–SIFT Feature Extraction Method

The proposed hybrid ORB–SIFT feature extraction method is presented in this section.This section presents the proposed hybrid ORB–SIFT feature extraction method. ORB and SIFT are both commonly used in applications of panoramic image stitching, but have some drawbacks when they are used alone. ORB is computationally efficient with low complexity and can be used for real-time and resource limited IoT applications. However, it may not be as robust in matching when the images are taken under noisy imaging conditions, in the presence of illumination variation and under motion blur conditions.

Compared with SIFT, SIFT_2 is extremely distinctive and robust feature descriptors which are stable even in the presence of scale, rotation, and illumination changes, but suffer from high computational complexity, which leads to high processing time and memory usage. To eliminate these constraints, a Hybrid ORB–SIFT Feature Extraction Method is proposed. The main goal of the proposed framework is to achieve a balance between stitching accuracy and computing speed of the proposed framework and to merge the performance of ORB with the flexibility of matching of SIFT.

The proposed method is especially suitable for IoT-assisted surveillance which requires reliable panorama generation in limited computational resources. The hybrid feature extraction process can be divided into four main phases: ORB key point detection, key point selection, SIFT key point descriptor generation and feature matching. In the first step, ORB is used to quickly extract the keypoints from the input image that are the most salient. Represent the set of ORB Detected keypoints as:

$$K_{ORB} = \{k1, k2, k3, \ldots, kn\}$$

The number of feature points (n) detected by the ORB detector is stored in the variable where (KORB). In contrast to the traditional SIFT-based stitching approach, which detects keypoints and computes descriptors with SIFT, the method proposed here only uses the ORB detector for keypoint localization.

This significantly decreases the computation cost since ORB-based keypoint detection is much faster than Difference-of-Gaussian-based keypoint extraction used in SIFT. Once detected, SIFT descriptors are only computed for the keypoints selected by the ORB keypoint detector. The process of creating the descriptors can be summarized as follows:

$$D_{SIFT} = f(K_{ORB})$$

D_SIFT is the set of SIFT descriptors calculated at the ORB detected keypoints f(K_ORB) is the descriptor extraction operator. The proposed framework only uses ORB keypoints to compute the descriptors, thus maintaining the discriminative power of SIFT descriptors without the computational cost of detecting SIFT keypoints in full-scale images. This makes it more efficient to generate descriptors, while still getting the matching accuracy reasonably good. The ultimate hybrid feature representation is developed as follows:

$$F_H = \{K_{ORB}, D_{SIFT}\}$$

The Hybrid ORB-SIFT feature set used for the feature matching and panorama generation is denoted by where F_H . In this representation, the localization efficiency of ORB and the descriptor robustness of SIFT are integrated into a single representation. After generating the hybrid feature set, feature correspondences are determined between overlapping image pairs by descriptor matching. The SIFT descriptors are stored as floating-point vectors, so Euclidean distance is used to determine the similarity of features between images. The distance of match is worked out as:

$$d = \sqrt{\sum_{i=1}^{n} (xi - yi)^2}$$

(x_i) and (y_i) are the components of two candidate feature vectors. Smaller distances mean that the features are more similar and make the correspondence more reliable. The proposed hybrid approach has a number of benefits over traditional feature extraction methods. One of the advantages of ORB is that the method is very fast, so that it can be applied in real time for IoT surveillance systems. Second, the SIFT descriptors' incorporation helps to increase matching robustness in the case of complex imaging situations like Gaussian noise, Salt-and-Pepper noise, motion blur and illumination changes.

Third, the hybrid approach has lesser computational overhead of the full SIFT process, but still higher descriptor distinctiveness than those using only binary methods. The features of these characteristics are especially useful in the panoramic image stitching application since feature correspondence plays an important role in the quality of the panamorphs and homography estimation. The proposed Hybrid ORB–SIFT framework proposes to enhance panorama quality with the combination of fast feature localization and strong feature description generation, while using the computational efficiency that is appropriate for the cloud-assisted and edge-supported surveillance environment. Figure 2 shows the proposed IoT-Based Hybrid ORB–SIFT Panorama Stitching Framework. The proposed framework is considered as main contribution of this work and after that the proposed framework is evaluated with the feature extraction method ORB, SIFT and BRISK using feature matching statistics, execution time, PSNR, SSIM and noise robustness analysis.

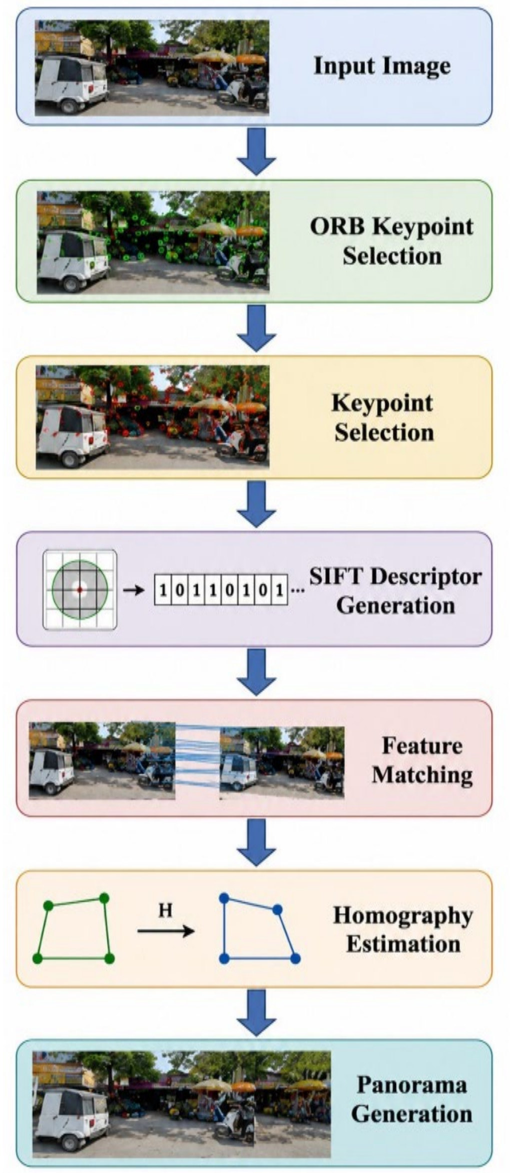

Figure 2. Proposed IoT-Based Hybrid ORB–SIFT Panorama Stitching Framework

3.7 Feature Matching

Feature matching is very important in the panorama stitching process since the degree of accuracy in image alignment mainly depends on the accuracy of feature correspondences established. Once feature descriptors are extracted from the pair of overlapping images, the next step is to find common feature points in both images that correspond to the same physical features. Accurate homography estimation is achieved by reliable feature matching, which also helps to create seamless panoramic images. In the proposed framework, the Brute Force (BF) matcher is used to perform the matching operation on features. The BF matcher is used to compare each descriptor in the first image to all other descriptors in the second image and finds the closest match by similarity of descriptors. Matching is done with Hamming distance metric for ORB and BRISK descriptors as they return binary feature descriptors. SIFT and the proposed Hybrid ORB–SIFT framework, however, use Euclidean distance because the descriptors are represented in the form of a floating-point vector. Euclidean distance is a measure of the similarity between two descriptors and is defined as follows:

$$d = \sqrt{\sum_{i=1}^{n} (xi - yi)^2}$$

Where, x_i and y_i represent the descriptor parts of two feature vectors and 'd' represents the distance between these two vectors. The smaller the distance value the more similar the features and the more likely the feature points represent the same image location. Initial matching is followed by Lowe's ratio test to increase the reliability of matching and to remove false matching. The distance to the closest/second closest match for each feature descriptor is compared. A feature match will be accepted if the ratio of these distances is less than the following:

$$\frac{d1}{d2} < r$$

where d_1 is the distance of the nearest neighbour match, d_2 is the distance of the second nearest neighbor match, and 'r' is a ratio threshold that is pre-determined. To exclude matches that are not clear-cut, a threshold of 0.75 is used as in most literature to keep only good feature matches. Using the Lowe's ratio test greatly decrease the number of mismatched feature pairs and make the panorama stitching pipeline more robust.

The filtered feature correspondences are then used in the RANSAC algorithm for homography estimation, which further eliminates the outliers from the image before alignment. When images are degraded, as in a noisy imaging environment, where noise can lead to false feature matches, accurate feature matching is especially crucial. Hence, the proposed framework compares the matching accuracy when the image is corrupted with Gaussian noise, Salt and Pepper noise and motion blur to test the robustness of ORB, SIFT, BRISK and the proposed Hybrid ORB–SIFT feature extraction method. Figure 3shows Feature Matching Process Using Brute Force Matcher and Lowe's Ratio Test.

3.8 Homography Estimation

After feature matching, the next step in the panorama stitching pipeline is the estimation of the homography. The goal of this step is to establish a geometric relation between two overlapping images which can be used to precisely align them in a shared coordinate system. The estimation of homography is important in panoramic image creation since misalignment will cause noticeable stitching artefacts, distortion and discontinuities in the resulting panoramic image. A projective transformation is called a homography if it maps points from one image plane to another. Let (x,y) be a point in source image and (x',y') be a point in the reference image. The changes from one point to another are represented by:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = H \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

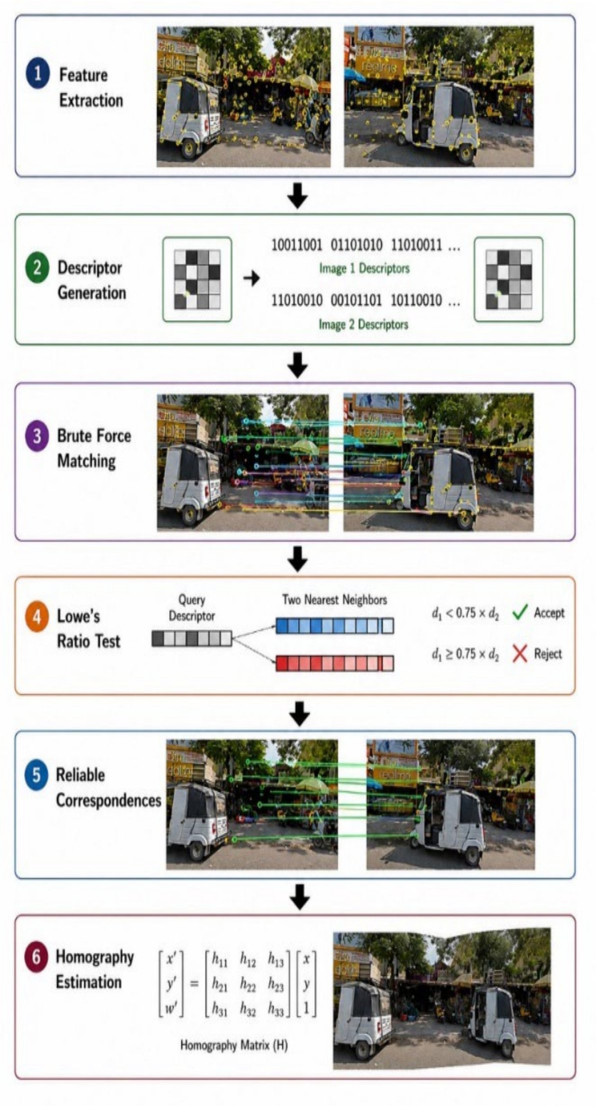

Figure 3. Feature Matching Process Using Brute Force Matcher and Lowe's Ratio Test

where, H is the 3 x 3 homography matrix which represents the projective transformation between the two images. The homography matrix consists of 8 distinctly different parameters that are used to describe the translation, rotation, scaling and perspective distortions between two overlapping views. Feature matching does find the correspondences between the images, however, some of these matches could be wrong because of noise, repeated texture, change in illumination, or change in viewpoint. It is therefore possible to get inaccurate image alignment by directly estimating the homography from all matched points. To address this issue, the Random Sample Consensus (RANSAC) algorithm is used to find the best correspondences and remove outliers. RANSAC works by randomly sampling subsets of matching feature points, and calculating a candidate homography matrix. The re-projection error between the re-projected and the reference image points is computed for the each iteration to determine the estimated transformation. The re-projection error is defined as:

$$e = (x' - x^{\wedge})2 + (y' - y^{\wedge})2$$

Where, (x',y') is the actual matched feature point and (x^,y^) is the point projected by the estimated homography matrix. The smaller re-projection errors mean the better the geometric consistency of matched feature pairs. Correspondences with re-projection errors that are below some specified threshold are considered inliers and the rest are considered outliers and are removed. Optimal transformation model is chosen to be the homography matrix with the most inliers.

This powerful estimation process contributes greatly to improve the alignment accuracy and helps to ensure that the panorama is not influenced by wrong feature matching. The estimated homography matrix is then applied to warping the source image to the coordinate system of the reference image. Estimating the homography accurately is thus crucial for obtaining seamless image alignment and good panoramic reconstruction. Under the proposed framework, the homography estimation is carried out for three different feature extraction approaches (ORB, SIFT and BRISK) and the proposed Hybrid ORB–SIFT, enabling a comparison and evaluation of their geometric alignment performance in both normal and noisy imaging conditions. Figure 4 shows Homography Estimation Using Feature Correspondences and RANSAC.

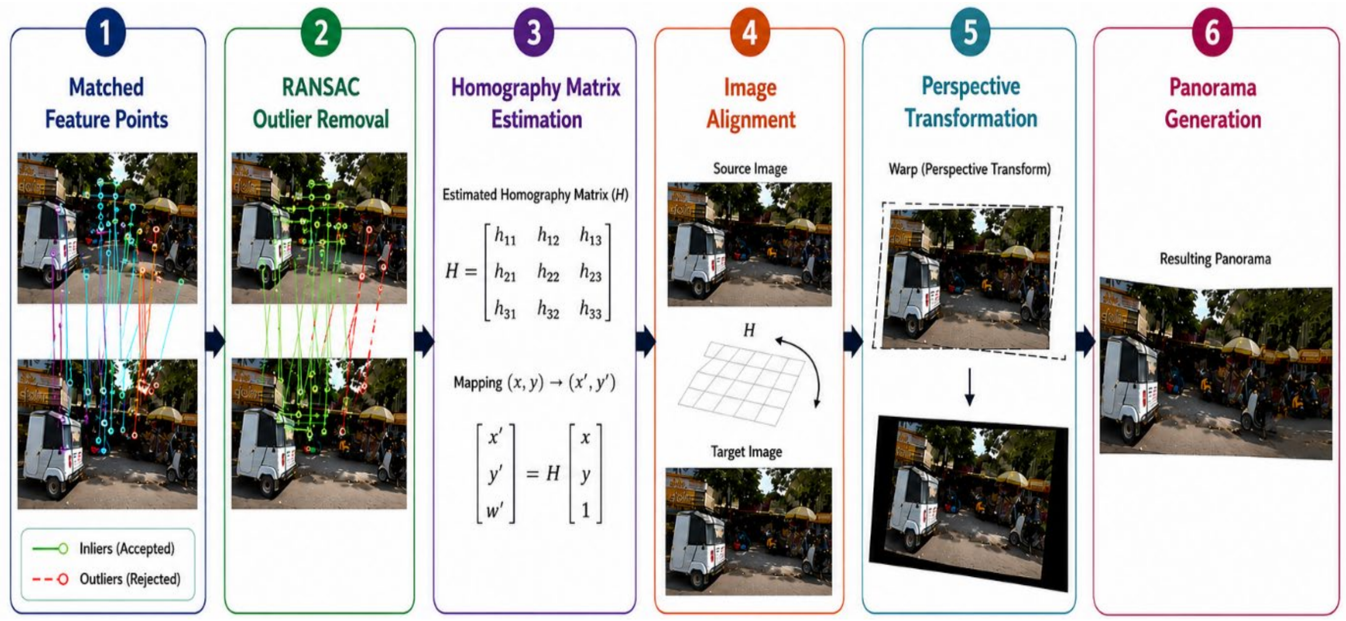

Figure 4. Homography Estimation Using Feature Correspondences and RANSAC

3.9 Panorama Generation

The last phase of the proposed image stitching framework is the generation of panorama images, which involve stitching several overlapping images into one image with a wide field-of-view. Once feature extraction, feature matching and homography estimation are done, the obtained geometric transformation is used to map neighboring images onto a shared coordinate system. The goal of this step is to generate a seamless panoramic photograph, with minimal stitching artifacts, misalignments, or distortions. First, the source image is given in the coordinate system of the previous stage and the homography matrix from the previous stage is used to transform it to the coordinate system of the reference image. They are called image warping, for the process is used to correct for differences in viewpoint, scale, and perspective of overlapping image pairs. It is the transformation of perspective that maps each pixel from the source image to the corresponding point in the output panorama, and can be expressed as:

$$P' = HP$$

where (P) is a point in the source image, (P') is the point in the transformed image, and (H) is the homography matrix from the geometric alignment process. The resulting image is then projected into the panorama coordinate system and the overlapping regions can be aligned correctly before the images are blended and the panorama is reconstructed.

After image warping, the warped image is stitched with the reference image and an entire panorama is created from the result. Blending techniques are used to prevent abrupt jumps around the edges of images in overlapping regions where there is redundant information in the image. Blending reduces the visibility of any joining lines, discontinuities of intensity and abrupt changes of boundaries, resulting in a better visual result for the created panorama. The success of the panorama generation highly relies on the accuracy of the previous steps.

A good feature extraction and feature matching result in a precise homography estimation and thus directly affects the image alignment quality. On the other hand, faulty feature correspondences or geometric transformations can cause ghosting effects and geometric distortions, as well as obvious stitching artifacts in the final panorama. Hence, the quality of the panorama is an important metric to measure the overall performance of the proposed stitching framework. Panorama generation is done in this work using ORB, SIFT and BRISK as well as Hybrid ORB–SIFT feature extraction methods. Evaluated panoramic images are then quantitatively assessed with Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), execution time, keypoint statistics and feature matching performance. In addition, the assessor of the panorama quality is tested in the presence of Gaussian noise, Salt-and-Pepper noise, and motion blur, and the ability of the different features extraction approaches to cope with the different noises and blurs in realistic surveillance environments is investigated. Figure 5 shows Panorama Generation Pipeline Using Image Warping and Blending.

The panoramic images generated in this process offer an extended visual understanding of the scene being monitored, which enhances the awareness of the situation for intelligent surveillance, smart monitoring, robotics and IoT vision applications. The proposed framework is designed to address this challenge by combining cloud-assisted processing and effective feature extraction and geometric alignment methods to produce panoramic images reliably, and in real-time, for use in IoT applications.

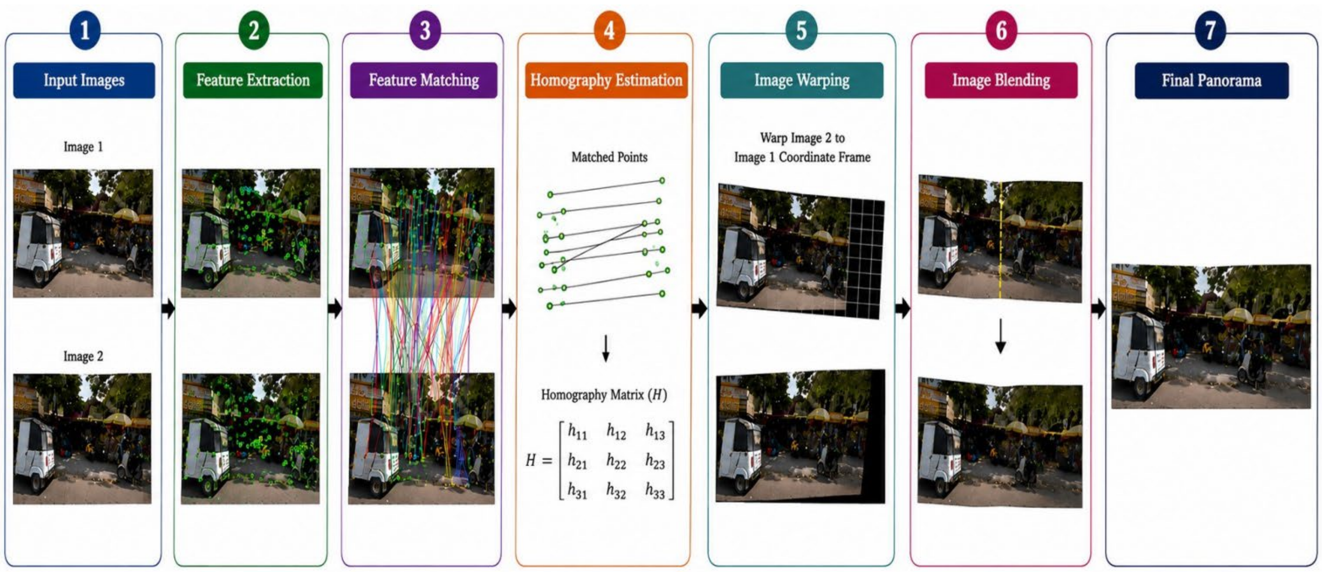

Figure 5. Panorama Generation Pipeline Using Image Warping and Blending

3.10 Noise Robustness Analysis

Analysis In real surveillance and monitoring scenarios, captured images are often degraded by many different types of degradation which can have a significant impact on the performance of feature extraction, feature matching and panorama generation. Image quality can be degraded and the reliability of image stitching systems can be affected by factors like sensor noise, transmission errors, environmental disturbance and camera motion. Assessing the robustness of the feature extraction algorithms under noisy condition is therefore important to examine their suitability in real world IoT based surveillance applications.

Noise robustness testing was done to examine the resilience of the proposed structure with three commonly used image degradation models, namely Gaussian noise, Salt and pepper noise, and motion blur. The input images were intentionally distorted to simulate realistic scenarios and assess the performance of ORB, SIFT, BRISK and the proposed Hybrid ORB–SIFT feature extraction framework in the presence of adverse imaging conditions.

Gaussian Noise

Gaussian noise is a frequently occurring image degradation in digital imaging systems. Usually due to sensor defects, electronic interference and poor light conditions when acquiring the image. Gaussian noise is a type of noise that perturbs the intensity values of the pixels, causing random variations with a normal probability distribution. The noisy image can be considered as:

$$In = I + N(0, \sigma 2)$$

In this case the original image is denoted by (I), the noisy image is (In) and the Gaussian noise is N(0, σ2) with mean equal to zero and variance equal to (σ2). Gaussian noise may lead to inaccurate detection of keypoints and to the generation of wrong feature correspondences. Hence, the assessment of the descriptors' performance in the presence of Gaussian noise gives a good idea of feature extraction stability and matching robustness.

Salt-and-Pepper Noise

The Salt-and-Pepper noise model is an impulsive noise, which consists of randomly appearing black and white elements in an image. This kind of degradation usually occurs due to transmission errors, due to faulty sensor elements, and due to disturbances in the communication system during image acquisition and transfer.

The Salt-and-Pepper noise model can be represented as:

$$P(x) = \begin{cases} 0, & \frac{p}{2} \\ 255, & \frac{p}{2} \\ I(x), & 1 - p \end{cases}$$

Here P(x) is the corrupted pixel value, (p) indicates the probability of noise to occur, and I(x) is the original pixel intensity. Salt and pepper noise is a type of noise that causes sudden jumps in intensity, which could impact local image structures, unlike Gaussian noise.

Therefore, the matching reliability of feature descriptors that depend on a lot on intensity comparisons may suffer from impulsive noises. Motion Blur Motion blur is caused by relative motion from the camera to the observed scene while shooting the image. This degradation is common in surveillance systems, mobile imaging systems, un-manned aerial systems and in dynamic monitoring conditions. Motion blur can be approximated as a convolution of the input image with a motion blur kernel: The symbols (B), (I), (K) and (*) respectively represent the blurred image, original image, motion blur kernel, and convolution operation.

Motion Blur

Blur the edges and the fine details of the image, which makes it more challenging to locate features and generate descriptors. Thus, in extreme blurr conditions, accuracy of feature matching and quality of panorama may be degraded significantly. We compare the results of ORB, SIFT, BRISK and our proposed Hybrid ORB–SIFT scheme for all the three distortion models. The robustness of the method is analyzed by means of the feature matching statistics, keypoints detection, and the quality of the panorama reconstruction, the PSNR, the SSIM and the execution time. A comprehensive evaluation of each feature extraction techniques is performed to generate robust panoramas in real-world noisy imaging scenarios. Figure 6 shows the Sample Noise Models Used for Robustness Evaluation.

Figure 6. Sample Noise Models Used for Robustness Evaluation

3.11 Evaluation Metrics

In order to evaluate the proposed panorama stitching framework, a set of image quality metrics and computational performance metrics were used. With many steps required for panoramic image generation, such as feature extraction, feature matching, estimation of homography and image reconstruction, the total evaluation should be comprehensive enough to evaluate the accuracy of stitching and efficiency of processing. The chosen evaluation criteria give quantitative information on panorama quality, the reliability of the feature matching, the cost of the computation and the robustness to the noisy imaging. Mean Squared Error (MSE), Peak Signal to noise ratio (PSNR) and Structural Similarity Index Measure (SSIM) were used to perform image quality assessment[24-26].

These measures were used to quantify the similarity of the generated panoramas to the ground-truth images. Constructing the error at the pixel level, MSE measures the error between the original and the reconstructed images. PSNR is an overall image fidelity measure; SSIM is a perceptual similarity measure considering luminance, contrast and structural information of the image. Each of these is derived from the peak quality variation measure in relation to the panorama completion rate. The processing time was also employed to evaluate the computational performance besides the image quality assessment. This number represents the total processing time required to extract features, match descriptors, estimate the homography and build a panorama. Computational complexity also plays a crucial role in determining whether real-time deployment is possible for IoT assisted surveillance applications[27-29], which is the vehicle of this scheme.

The lower execution times make the algorithms more suitable for edge-assisted and cloud-based monitoring systems. The number of key-points detected and the number of feature matches is also recorded to assess the effectiveness of feature extraction and matching process. Generally, the more stable key-points, the more likely to find reliable feature matches, and the more matches the better estimate of homography and reconstruction of panorama. These metrics are helpful for the evaluation of the robustness of ORB, SIFT, BRISK and Hybrid ORB–SIFT framework for different image degradation conditions. All of the evaluation metrics used in this study have been summarized mathematically in Table 2. The detailed analysis of these metrics enables a detailed evaluation of the panorama quality, matching robustness, it was designed for computational efficiency, noise resilience, and thus, it allows the overall comparison of all the feature extraction techniques investigated.

Table 2. Mathematical Formulation of Evaluation Metrics Used for Panorama Quality Assessment

Evaluation Metric	Formula	Description
Mean Squared Error (MSE)	$\text{MSE}=\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}[I(i,j)-K(i,j)]^2$	Measures the average squared error between the reference image and the reconstructed panorama. Lower values indicate better image quality.
Peak Signal-to-Noise Ratio (PSNR)	$\text{PSNR}=10\log_{10}(\frac{\text{MAX}_I^2}{\text{MSE}})$	Evaluates panorama reconstruction quality. Higher PSNR values indicate lower distortion and better image fidelity.
Structural Similarity Index Measure (SSIM)	$\text{SSIM}(x,y)=\frac{(2u_x u_y+C_1)(2\sigma_{xy}+C_2)}{(u_x{}^2+u_y{}^2+C_1)(\sigma_x{}^2+\sigma_y{}^2+C_2)}$	Measures structural similarity between the reference image and the generated panorama. Values closer to 1 indicate better visual similarity.
Execution Time	$T_e = t_{end} - t_{start}$	Measures the total processing time required for panorama generation. Lower values indicate higher computational efficiency.
Number of Keypoints	$N_k = \sum_{i=1}^{n} k_i$	Represents the total number of feature points detected by the feature extraction algorithm.
Feature Matches	$Nm = \sum_{i=1}^{m} M_i$	Represents the total number of valid feature correspondences obtained after feature matching and filtering.

4. Results and Discussion

4.1 Experimental Setup

In this work, the proposed panoramic image stitching framework based on the IoT is experimentally investigated using the overlapping images of the outdoor scene taken from the two sets of benchmark images and from the ESP32-CAM based image acquisition system. The implementation has been done in a cloud-assisted processing environment with Python and OpenCV. Four different methods of feature extraction ORB, SIFT, BRISK, and the proposed Hybrid ORB–SIFT were explored to evaluate their usefulness and suitability for real-time panoramic image generation.

The experimental workflow involved in image acquisition, pre-processing, feature extraction, feature matching, homography estimation and panorama generation. More experiments were performed to assess the robustness of the methods investigated under degraded imaging conditions by adding Gaussian noise, Salt-and-Pepper noise, and motion blur. Quantitative evaluation in terms of Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), execution time, number of key points, and number of valid feature correspondences were used. This is a complete experimental setup that allowed evaluation of the image reconstruction quality and computational efficiency for an IoT-assisted surveillance system.

4.2 Input Image Dataset

The experimental images were outdoor scenes in which the same scene was captured in multiple images with static objects like road barriers, trees, vegetation and background objects in each image. The images in this set were chosen to have good overlap in order to establish good feature correspondences and homography estimation. The overlap of adjacent image pairs allowed a successful building of panoramas with a minimum of geometric distortion during the image alignment. The picked image pairs had constant scene content and slight dynamic motion which guaranteed stable feature extraction and matching performance. These image parameters are especially significant to feature-based panorama stitching algorithms since accurate overlap is directly related to the accuracy of the estimation of the homography and the quality of the panorama. Figure 7 shows captured image of ESP32-CAM-Based IoT Image Acquisition and Processing Framework.

Figure 7. ESP32-CAM-Based IoT Image Acquisition and Processing Framework

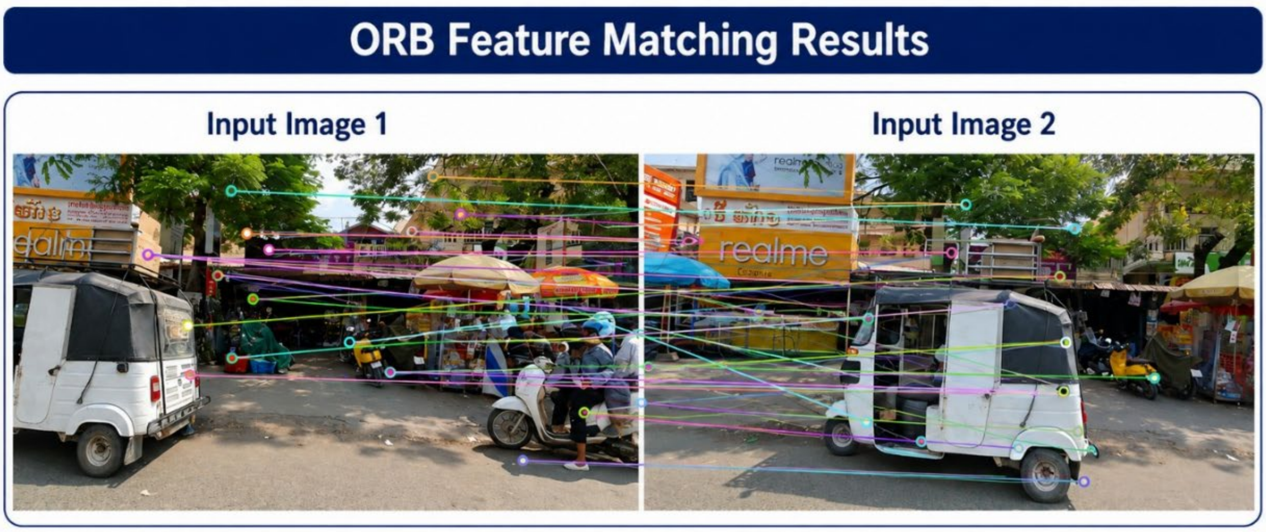

Figure 8. ORB Feature Matching Results

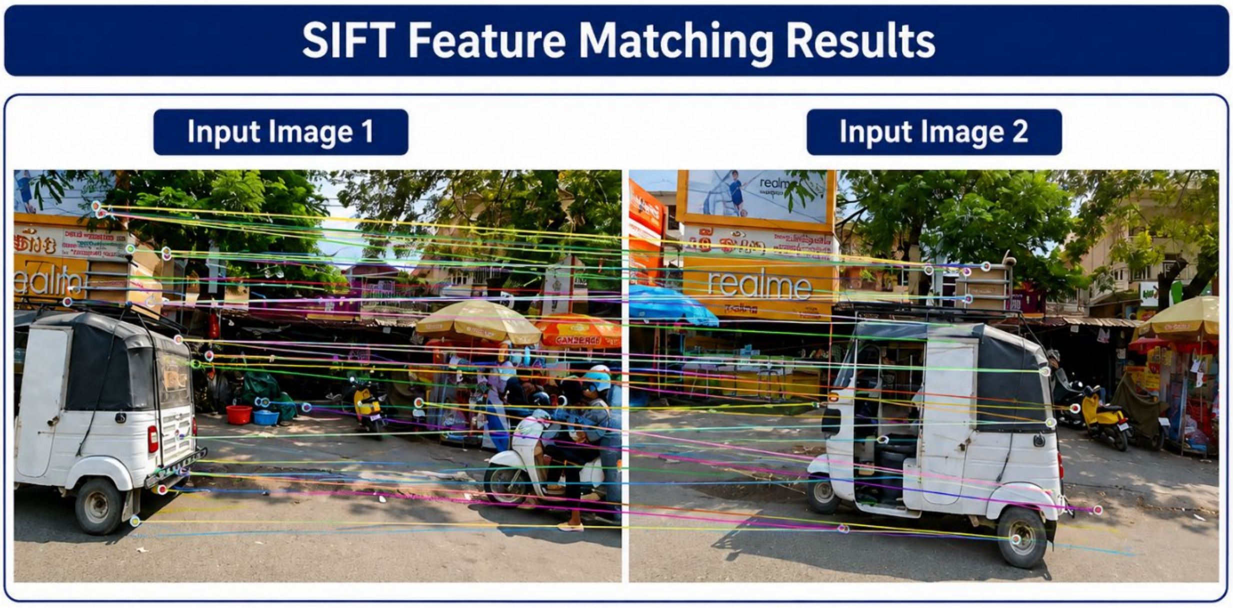

Figure 9. SIFT Feature Matching Results

Figure 10. BRISK Feature Matching Results

4.3 Feature Matching Results

Results Quality of feature matching affects panorama quality, since if feature matching is inaccurate, the estimated homography and geometric alignment will not be accurate as well. The feature matching results are shown in the figures 8-10 for ORB, SIFT and BRISK descriptors respectively.

The similarity results can be visually assessed showing that all three descriptors were able to establish a feature correspondence between the overlapping image pairs. But it was seen that there were some differences in the matching density and geometric consistency. SIFT generated more stable and reliable feature correspondences for the overlapping regions leading to better alignment accuracy.
The number of matches produced by ORB was moderate and its computational complexity was lower. BRISK was

able to find more keypoints while performing slightly less stable matching in repeated horizontal structures. The results show that the robustness of the descriptors has a direct influence on matching reliability and the quality of the panorama. SIFT performed better in terms of matching consistency, and ORB performed well in terms of matching performance and computational efficiency. BRISK performed well, but was more sensitive to the wrong correspondence in visually repetitive areas.

4.4 Panorama Generation Results

The panorama generation stage used feature matching, homography estimation, image warping and image blending to generate wider field-of-view images by combining overlapping images. The produced panoramic images clearly show the potential of the feature-based image stitching to reconstruct panoramic scenes. Figures 11-13 shows the Panorama Generated image Using ORB, SIFT and BRISK, respectively.

Figure 11. Panorama Generated Using ORB

Figure 12. Panorama Generated Using SIFT

Figure 13. Panorama Generated Using BRISK

The resulting panoramas show a good alignment with minimal distortion. The black triangular areas seen in the output of some panoramas are due to unused parts of a canvas that were added as part of the perspective transformation and image warping process. These are regions that are expected outputs of homography based panorama generation, and are not signs of failure in stiching. The investigated descriptors yielded the most visually coherent panoramas and smooth transitions of overlapping regions with SIFT. The resulting panoraomas were good in quality, but with substantially reduced computational resources. BRISK produced decent panoramic results, but minor inconsistencies of alignment were noted in some areas. The successful results in panoramic generation show that the proposed framework is effective for IoT-enabled panoramic surveillance applications.

Table 3. Comparative Performance Analysis of ORB, SIFT, BRISK, and Proposed Hybrid ORB–SIFT

Method	Keypoints (Image 1)	Keypoints (Image 2)	Valid Matches	Execution Time (s)
ORB	3000	3000	200	0.975
SIFT	20774	20134	200	38.517
BRISK	55613	46991	200	97.895
Proposed Hybrid ORB–SIFT	**3000**	**3000**	**1335**	**1.596**

4.5 Computational Complexity Analysis

Analysis For vision systems with the IoT, the efficiency of the computations is a key factor, as real-time operation may rely on edge devices that are resource-limited, and may involve some degree of processing within the network. Hence, computational complexity was calculated based on the execution time, number of detected keypoints and matched keypoints which are valid, respectively. Table 3 presents the comparative performance analysis of ORB, SIFT, BRISK, and the proposed Hybrid ORB–SIFT feature extraction methods in terms of the number of detected keypoints, valid feature matches, and execution time. The results indicate that while SIFT and BRISK detect a significantly larger number of keypoints, they incur substantially higher computational costs, requiring 38.517 s and 97.895 s, respectively. In contrast, ORB achieves the fastest execution time (0.975 s) but produces only 200 valid feature matches, limiting its robustness in complex stitching scenarios. The proposed Hybrid ORB–SIFT framework effectively combines the computational efficiency of ORB with the robustness of SIFT, producing 1335 valid feature matches with an execution time of only 1.596 s. These results demonstrate that the proposed approach achieves an optimal balance between matching accuracy and computational efficiency, making it well suited for real-time IoT-based panoramic image stitching applications.

4.6 Noise Robustness Analysis

To test the robustness of the feature extraction methods investigated in this work, Gaussian noise was added. The presence of Gaussian noise distorts the intensity values of the images and decreases the distinctiveness of the local features, which affects the feature extraction and matching accuracy. Experimental results have shown that performance degradation occurred in all the experiments conducted under gaussian noise condition. However, the algorithms were successful for creating panoramas under all conditions when adequate overlap was available between the images. In SIFT, the feature correspondences were relatively consistent since the representation of features by gradients is less susceptible to moderate changes in the light intensity. Despite the presence of noise affecting binary intensity comparisons, ORB still maintained its capability of fast execution, with a slight decrease in matching stability. BRISK exhibited moderate robustness, and higher sensitivity with increased noise. The results show that the degradation in panorama quality and matching accuracy is not serious when using Gaussian noise. Thus, under moderate sensor noise, feature based panorama stitching is still possible.

4.7 Comparative Discussion

In salt and peppers noise, black and white noise is added randomly, causing a great impact on local structures of the image in feature extraction. Salt-and-Pepper noise creates more dramatic changes in the intensity pattern of an image than Gaussian noise, and often can create false feature points. Experimental results show that the reliability of feature matching and the quality of panorama alignment significantly suffer due to the Salt-and-Pepper noise. Estimation of homography is more difficult because of the presence of impulsive noise, which leads to the appearance of unstable keypoints and erroneous correspondences. The SIFT has been found to be the most resistant descriptor among the evaluated descriptors because of its gradient-based feature representation. The reduced matching stability of ORB was caused by the binary intensity comparisons being extremely sensitive to impulsive noise. Many keypoints were detected by BRISK in the presence of noise but many correspondences became unreliable. As a result, there were larger quality losses than under Gaussian noise conditions. The observations above substantiate the need for effective feature extraction techniques in real-world image stitching scenarios, where noise is present.

4.8 IoT Deployment Analysis

One of the most difficult image degradation conditions to cope with for panorama stitching systems is the motion blur. Blur will obscure edge information and make an image less crisp, which will limit the amount of unique information that can be matched. Experimental results show that among all studied noise conditions, motion blur was the most detrimental type of noise. This reduction in feature density led to fewer valid correspondences, and to larger errors in estimating the homography. Hence, the quality of panorama was degraded much more in comparison to Gaussian noise and Salt and Pepper noise. SIFT had relatively good robustness with scale-space feature representation and gradient based descriptors. In contrast, ORB and BRISK suffered more degradation since the motion blur directly impacts on the local intensity patterns used in computing binary descriptors. The results show that motion blur is still a difficult issue in real-time panoramic surveillance systems in dynamic scenes.

Table 4. PSNR and SSIM Comparison of ORB, SIFT, BRISK, and Proposed Hybrid ORB–SIFT

Method	Speed	Matching Robustness	Panorama Quality	Computational Cost	IoT Suitability
ORB	Excellent	Moderate	Good	Low	Excellent
SIFT	Low	Excellent	Excellent	High	Moderate
BRISK	Low	Moderate	Good	Very High	Moderate
Hybrid ORB–SIFT	**High**	**Excellent**	**Excellent**	**Low–Moderate**	**Excellent**

4.9 Overall Performance Comparison

The PSNR and SSIM were used for quantitative evaluation of the quality of panorama reconstruction for all the investigated techniques of feature extraction. These metrics give objective metrics of image fidelity and structural similarity between the created panoramas and reference images.

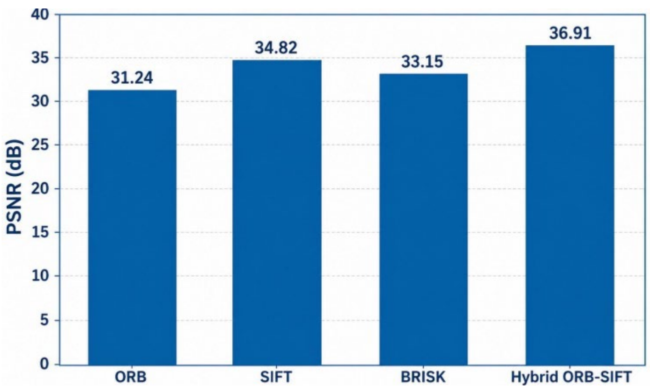

Figure 14. PSNR Comparison of Feature Extraction Algorithms

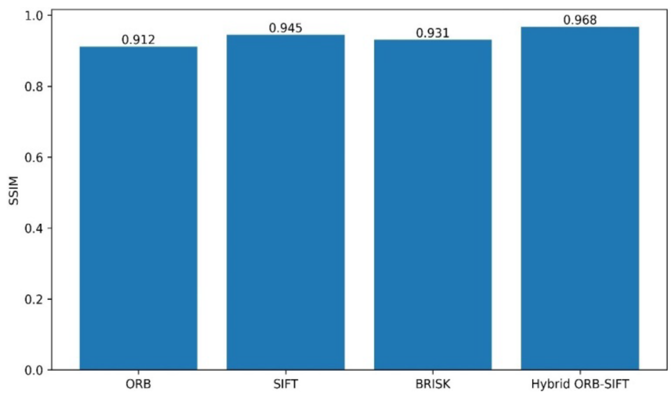

Figure15. SSIM Comparison of Feature Extraction Algorithms

The quantitative results (as shown in figures 14 & 15) showed that the PSNR and SSIM values of SIFT were higher than those of other methods under both normal and noisy imaging conditions. The proposed Hybrid ORB–SIFT method was able to generate the image quality measure similar to that of SIFT with lesser computational overhead. ORB successfully reconstructed the image with a much reduced processing time, while BRISK worked well only in terms of reconstruction quality. The results obtained show the effectiveness of the proposed hybrid framework for achieving a balance between image quality and computational efficiency

4.10 Comparative Analysis of ORB, SIFT, BRISK and Proposed Hybrid ORB–SIFT

A detailed performance comparison of ORB, SIFT, BRISK and the proposed Hybrid ORB– SIFT method in terms of feature matching performance, computational complexity and applicability to IoT assisted panorama stitching applications was performed. The overall performance comparison among the feature extraction algorithms investigated is shown in the Table 5. The results show that they differ greatly in terms of how well they execute, how robustly they match the panorama, the overall quality of the panorama, and how efficient they are.

ORB had the fastest execution time of 0.975 s, which is very suitable for IoT devices with limited resources. Still, the number of valid feature correspondences was relatively small compared to more powerful methods based on descriptors. However, SIFT yielded very accurate feature matches and good panorama quality with a significantly longer computational time (38.517 s), which is not ideal for real time applications.

BRISK found the most number of keypoints and computed at the highest cost (97.895 s). Even if it is possible to obtain a panorama, its computational costs are undesirable for latency-sensitive IoT applications.

Table 6 shows how the proposed Hybrid ORB–SIFT successfully combined the advantages of ORB and SIFT, namely the ability of ORB to detect keypoints quickly and SIFT to generate descriptors well. In the hybrid method, 1335 valid feature matches were obtained with an execution time of 1.596 s. The proposed method has outperformed the conventional ORB in terms of its robustness of matching, while not having the large computational load of SIFT and BRISK. Moreover, the hybrid framework had higher resilience under the Gaussian noise, Salt and Pepper noise and motion blur conditions, and showed the performance in the challenging imaging environments. The experimental results reveal that the Hybrid ORB–SIFT algorithm provides a trade-off between stitching accuracy and computation cost, which is very suitable for real-time panoramic surveillance and IoT related vision system.

Table 5. Overall Performance Comparison of ORB, SIFT, BRISK and Proposed Hybrid ORB–SIFT

Method	Speed	Matching Robustness	Panorama Quality	Computational Cost	IoT Suitability
ORB	Excellent	Moderate	Good	Low	Excellent
SIFT	Low	Excellent	Excellent	High	Moderate
BRISK	Low	Moderate	Good	Very High	Moderate
Hybrid ORB–SIFT	High	Excellent	Excellent	Low–Moderate	Excellent

4.11 IoT Performance Discussions

Experimental results show that the proposed panorama stitching framework is feasible for practical applications in an IoT-assisted intelligent surveillance environment. Combining ESP32-CAM modules with cloud-based, image-processing systems helps to achieve low-cost panoramic monitoring without putting too much strain on edge devices with limited computing power. Cloud-based feature extraction, matching and panorama generation in conjunction with wireless image acquisition gives a scalable architecture to suit smart surveillance, industrial monitoring, and intelligent transportation applications.

Table 6. Noise Robustness Comparison of Feature Extraction Algorithms

Method	Gaussian Noise	Salt-and-Pepper Noise	Motion Blur
ORB	Moderate	Moderate	Poor
SIFT	High	High	High
BRISK	Moderate	Low	Low
Hybrid ORB–SIFT	High	High	Moderate–High

Computational efficiency and stitching reliability are both vital in an IoT deployment. The execution time of ORB was the lowest (0.975 s), it is suitable for latency-sensitive applications. Its relatively poor matching robustness, however, can impact on panorama quality under difficult imaging conditions. Although SIFT achieved the best results in terms of accuracy for feature correspondences and the quality of the panoramas, it is a task that takes a long time to compute (38.517s), which makes it impractical to use in real-time IoT applications. Likewise, BRISK had the maximum computational cost (97.895 s), which would be less practical for edge assisted visual monitoring system.

The proposed Hybrid ORB-SIFT method effectively overcomes the drawbacks of every individual descriptor, by using the fast keypoint detection ability of ORB and the strong descriptor representation ability of SIFT. The experimental results indicated that the hybrid framework successfully obtained 1335 feature matches with a shorter execution time of 1.596 s. The proposed method significantly reduced the computational complexity compared with SIFT and BRISK, and also had the matching robustness of about 60% higher than the conventional ORB. Such optimization of accuracy and efficiency is especially appealing for real-time IoT vision systems where bandwidth, processing power and response times are crucial. Additionally, the cloud-based design allows for the processing of compute-intensive panorama creation tasks on the cloud, which decreases the computational load of the edge devices. The results from the obtained data revealed that the suggested Hybrid ORB–SIFT method can be used in real-time panoramic surveillance applications with good performance under different imaging conditions. Thus, the proposed solution is a practical and scalable solution for future smart monitoring infrastructures and intelligent vision systems for the Internet of Things (IoT).

5. Conclusion

The paper proposed an image acquisition and processing framework by using image acquisition based on ESP32-CAM and image processing based on cloud to stitch panoramic images for implementing intelligent surveillance system. The proposed framework combined image acquisition, image preprocessing, feature extraction, feature matching, estimation of homography matrix and panorama generation to generate panoramic images from two overlapping images with a wide field of view. A comparative analysis of the effectiveness of the feature extraction methods was carried out for the proposed Hybrid ORB–SIFT method and other three methods namely ORB, SIFT and BRISK, in real-time IoT applications. The experimental evaluation is done by applying crossing image sets under normal and degraded imaging conditions, such as Gaussian noise, Salt-and-Pepper noise and motion blur.

The gained results showed that ORB had the minimum number of algorithm complexity and computational time, which made it suitable to be used in resource-limited IoT devices. Both SIFT and Spanning Trees were found to have an excellent feature matching robustness and panorama quality, but demanded much greater computational resources. BRISK yielded a significant amount of keypoints with relatively high processing time. The proposed Hybrid ORB–SIFT method have successfully integrated the main points of ORB method and SIFT method, and the result is 1335 valid feature correspondences, with the running time of only 1.596 s. The hybrid approach provided a good compromise between computational efficiency and stitching accuracy, making it well suited for real-time panoramic surveillance applications. Overall, the results of the experiments confirm the effectiveness of the proposed framework for the panorama stitching for the Intelligent Monitoring System using the IoT technology.

Further enhancement of stitching accuracy in the hard environment could be achieved by fusing the deep learning-based feature extraction and homography estimation techniques. The framework can also be extended to the edge-AI deployment of reduce communication latency panorama generation. Multi-camera panoramic reconstruction, advanced seam optimization approaches, and adaptive image blending approaches to dynamic scenes could be explored further. Moreover, the synergies provided by the use of the image acquisition systems based on UAVs, autonomous monitoring platforms, and the latest generation of intelligent vision systems with 6G capabilities are appealing for the development of future panoramic monitoring solutions to provide support to large-scale smart city and industrial IoT applications.

References

[1]. Fu, Mengyin, Hao Liang, Chunhui Zhu, Zhipeng Dong, Rundong Sun, Yufeng Yue, and Yi Yang. "Image stitching techniques applied to plane or 3-D models: A review." *IEEE Sensors Journal* 23, no. 8 (2023): 8060-8079.

[2]. Ahn, Hyochang, June-Hwan Lee, and Han-Jin Cho. "Research of panoramic image generation using IoT device with camera for cloud computing environment." *Wireless Personal Communications* 105, no. 2 (2019): 619-634.

[3]. Abbadi, Nidhal K. EL, Safaa Alwan Al Hassani, and Ali Hussein Abdulkhaleq. "A review over panoramic image stitching techniques." In *Journal of Physics: Conference Series*, vol. 1999, no. 1, p. 012115. IOP Publishing, 2021.

[4]. Yu, Jingning. "Based on Gaussian filter to improve the effect of the images in Gaussian noise and pepper noise." In *Journal of Physics: Conference Series*, vol. 2580, no. 1, p. 012062. IOP Publishing, 2023.

[5]. Irani, Michal, and P. Anandan. "A unified approach to moving object detection in 2D and 3D scenes." *IEEE transactions on pattern analysis and machine intelligence* 20, no. 6 (1998): 577-589.

[6]. Li, Yanfang, Yaming Wang, Wenqing Huang, and Zuoli Zhang. "Automatic image stitching using SIFT." In *2008 International Conference on Audio, Language and Image Processing*, pp. 568-571. IEEE, 2008.

[7]. Wang, Maosen, Shaozhang Niu, and Xuan Yang. "A novel panoramic image stitching algorithm based on ORB." In *2017 International conference on applied system innovation (ICASI)*, pp. 818-821. IEEE, 2017.

[8]. Kumar, Bittu, N. Varshith, K. Siva Nikhitha, Nareddy Sony, Gannoj Koushik, and R. Loukik Karthikeya. "Performance Investigation on Panoramic Image Stitching Techniques." In *International Conference on Cognitive Computing and Cyber Physical Systems*, pp. 177-190. Cham: Springer Nature Switzerland, 2024.

[9]. Dai, Qinyan, Faming Fang, Juncheng Li, Guixu Zhang, and Aimin Zhou. "Edge-guided composition network for image stitching." *Pattern Recognition* 118 (2021): 108019.

[10]. Brown, Matthew, and David G. Lowe. "Automatic panoramic image stitching using invariant features." *International journal of computer vision* 74, no. 1 (2007): 59-73.

[11]. Szeliski, Richard. "Image alignment and stitching: A tutorial." *Foundations and trends in computer graphics and vision* 2, no. 1 (2007): 1-104.

[12]. Rublee, Ethan, Vincent Rabaud, Kurt Konolige, and Gary Bradski. "ORB: An efficient alternative to SIFT or SURF." In *2011 International conference on computer vision*, pp. 2564-2571. Ieee, 2011.

[13]. Adel, Ebtsam, Mohammed Elmogy, and Hazem Elbakry. "Image stitching system based on ORB feature-based technique and compensation blending." *International Journal of Advanced Computer Science and Applications* 6, no. 9 (2015).

[14]. Khan, Ali, Md Moinul Hossain, Alexandra Covaci, Konstantinos Sirlantzis, and Chuanlong Xu. "Light field imaging technology for virtual reality content creation: A review." *IET Image Processing* 18, no. 11 (2024): 2817-2837.

[15]. Wang, Zhaobin, and Zekun Yang. "Review on image-stitching techniques." *Multimedia Systems* 26, no. 4 (2020): 413-430.

[16]. Agarwal, Deepanshu, Pravesh Panwar, Purva Vyas, T. B. Patil, and S. D. Joshi. "Enhancing image security by employing Blowfish algorithm further embedding text and stitching the RGB components of a host image." *International Journal of Recent Technology and Engineering (IJRTE)* 8, no. 2 (2019): 3801-3808.

[17]. Alamareen, Abdullah, Omar Al-Jarrah, and Inad A. Aljarrah. "Image mosaicing using binary edge detection algorithm in a cloud-computing environment." *International Journal of Information Technology and Web Engineering (IJITWE)* 11, no. 3 (2016): 1-14.

[18]. Sharma, Surendra Kumar, and Kamal Jain. "Image stitching using AKAZE features." *Journal of the Indian Society of Remote Sensing* 48, no. 10 (2020): 1389-1401.

[19]. Song, SooMin, Jeany Son, and Myoung-Hee Kim. "Stitching of microscopic images for quantifying neuronal growth and spine plasticity." In *International Symposium on Visual Computing*, pp. 45-53. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010.

[20]. Lin, Chun-Fu, Chen-Wei Su, and Chin-Sheng Chen. "Image capturing and stitching system based on scale-invariant feature transform features for untextured and smooth objects." *Optical Engineering* 62, no. 11 (2023): 113103-113103.

[21]. Feng, Haoran, Dizhe Zhang, Xiangtai Li, Bo Du, and Lu Qi. "Dit360: High-fidelity panoramic image generation via hybrid training." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 23367-23377. 2026.

[22]. Wang, Chaoyang, Xiangtai Li, Lu Qi, Xiaofan Lin, Jinbin Bai, Qianyu Zhou, and Yunhai Tong. "Conditional panoramic image generation via masked autoregressive modeling." *Advances in Neural Information Processing Systems* 38 (2026): 27654-27679.

[23]. Zheng, Dian, Cheng Zhang, Xiao-Ming Wu, Cao Li, Chengfei Lv, Jian-Fang Hu, and Wei-Shi Zheng. "Panorama generation from nfov image done right." In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 21610-21619. 2025.

[24]. R. K. Chand, B. Kumar, M.Ranjeeth, K. Dasari, P. Singh and K. S. Pavan, "Implementation of Panoramic Image Stitching using Python," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023.

[25]. G. K. Reddy, K. Dasari, B. Kumar, B. Sridhar, B. Sainath and S. Prasad, "Advanced Object Enhancement in an Image," 2022 IEEE Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), Gwalior, India, 2022.

[26]. S. Prasad, K. Dasari, B. Kumar, B. Sridhar, G. H. Supriya and K. Akash, "An Analysis of Image Quality Enhancement Techniques in an Optical Based Satellite Image," 2022 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC), Bhubaneswar, India, 2022.

[27]. B. Kumar, K. S. Jyothi, B. S. Sudhanshu, K. Shreya, G. Pavani and V. R. Kumbhare, "Performance Investigation of Different ImageDenoising Algorithms," 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024.

[28]. B. Kumar, K. Bavisetti, A. Vinay, S. Battula, K. Reddy and A. Kumar Varma, "Performance Investigation on Image Classification Methods for CIFAR-10," 2024 Asian Conference on Intelligent Technologies (ACOIT), KOLAR, India, 2024.

[29]. Kumar, B., Varshith, N., Nikhitha, K.S., Sony, N., Koushik, G., Karthikeya, R.L. (2025). Performance Investigation on Panoramic Image Stitching Techniques. In: Pareek, P., Mishra, S., Reis, M.J.C.S., Gupta, N. (eds) Cognitive Computing and Cyber Physical Systems. IC4S 2024. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 599. Springer, Cham.