# Image-based Indoor Positioning: Current Status and Challenges

Melek ŞENTÜRK[1], Emrullah DEMİRAL[1], Hacer Kübra SEVİNÇ[2], İsmail Rakıp KARAŞ[1] and Uznir Ujang[3]

[1] Faculty of Computing and Information Science, Karabuk University, Karabuk, Turkey
meleksenturk@karabuk.edu.tr, emrullahdemiral@karabuk.edu.tr, ismail.karas@karabuk.edu.tr
[2] Vocational School of Information Technology, Karabuk University, Karabuk, Turkey
kubrasevinc@karabuk.edu.tr
[3] Universiti Teknologi Malaysia, Malaysia, mduznir@utm.my

**Keywords:** Image-based indoor positioning, camera-based indoor positioning, vision-based indoor positioning

## Abstract

The loss of GNSS (Global Navigation Satellite Systems) signals in enclosed spaces has revealed the need for potential positioning solutions in indoor environments that require high accuracy. Although the solutions developed so far have not yet provided a universally accepted system in terms of metrics such as cost, feasibility, and performance, image-based positioning systems have recently become the focus of researchers' interest in response to the current search. The literature review conducted revealed a gap in the literature, as no study presenting the current developments in this field was found. This study aims to fill this gap and provide researchers with a reference source. Accordingly, studies between 2020 and 2025 were collected using the keywords "vision-based," "image-based," "camera-based," "visual," "indoor localization," "indoor positioning," "indoor navigation," "indoor tracking," "visual SLAM," "VIO", "simultaneous localization and mapping", "feature matching", "image matching", "feature-based") were collected. Since 2025 has not yet been completed, a prediction was made using linear regression with data from previous years. The collected publications have been disaggregated using DOI numbers, and studies not containing the keywords ("visual", "image", "camera", "SLAM", 'feature', "indoor positioning", "indoor navigation", "indoor tracking", "indoor localization") have been excluded. Existing studies were examined as traditional methods and deep learning-based approaches, and deep learning-based approaches were found to be superior to traditional methods in terms of speed, reliability, and accuracy metrics. However, surfaces with low texture, moving environments, and variable lighting conditions remain limitations for both methods. Future studies should test the developed systems in different areas and scenarios.

## 1. Introduction

From the past to the present, location information required in many different fields and scenarios, such as military applications [1], smart cities [2], personal navigation applications [3], and industrial automation systems [4], can be provided outdoors via GNSS. However, the use of these systems in indoor environments is limited due to various obstacles [9], and high-accuracy positioning poses a problem in enclosed spaces. In this context, this problem is the focus of researchers, and the search for alternative solutions is actively ongoing.

GNSS satellites orbiting the Earth transmit signals containing time and orbital information. A GNSS receiver on the ground can determine its own position using signals received from these four satellites. Among these systems, the most widely used are GPS (US), GLONASS (Russia), Galileo (EU), and Beidou (China), which are satellite navigation systems that provide global positioning information [5][6]. While these systems work efficiently in open areas [7], they encounter various limitations in indoor environments [8] that require high accuracy and precision. The limitations mentioned are presented in Table 1.

| Obstacle | Effect and result |
|---|---|
| Multi-Path Propagation | Objects in enclosed spaces cause problems such as scattering and reflection of electromagnetic waves, resulting in signals being transmitted via different paths [9]. |
| Signal Loss | The building materials used cause signal loss indoors [9] [10]. |
| Transition Zones | Transitions between outdoor and indoor areas cause sudden signal loss [9][10]. |
| Line of Sight Obstacle | Insufficient number of satellites in the field of view causes errors [11][12]. |
| Signal Interference | Electronic devices in enclosed environments cause signal distortion [9] [13]. |

**Table 1.** Challenges, Impact, and Outcomes in Indoor Positioning

To overcome the challenges encountered in indoor positioning, existing literature has proposed high-precision GNSS receivers and ground-based alternative solutions [14]. However, the high cost of these solutions limits their applicability [15]. Although various technologies such as Wi-Fi, Bluetooth, and VLC (Visible Light Communication) have offered solutions for different scenarios such as navigation and inventory tracking in indoor environments [16], there is still no universally accepted solution in this field [18]. **Figure 1** provides a classification of indoor positioning technologies.

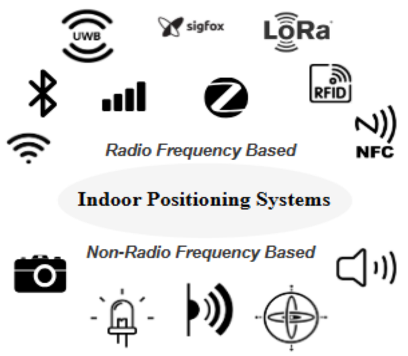

**Figure 1:** Indoor Positioning Technologies

Although RF (Radio Frequency)-based solutions, which are widely used among the technologies in question, offer advantages such as wide coverage area and real-time tracking, they also bring problems such as multipath propagation and signal interference in closed environments [16]. **Table 2** presents the advantages and disadvantages of RF-based solutions.

| Technology | Advantage | Disadvantage |
|---|---|---|
| Wi-Fi [19][20] | Extensive infrastructure area, high accuracy in advanced methods. | Multi-path propagation effect, sensitivity to environmental changes |
| UWB [16][20] | High accuracy, suitable for precise tracking, resistant to multipath propagation effects. | High hardware costs limit its widespread use. |
| RFID [21][22] | Low cost (with a single reader), high accuracy can be achieved when used with advanced models | The need for multiple readers to achieve high positioning accuracy increases the cost. |
| Hybrid [23][24] | Using the strengths of different technologies together increases accuracy and robustness | Increased system complexity increases application and maintenance costs |

**Table 2.** Indoor Positioning Technologies Advantages and Disadvantages

The fundamental constraints in indoor positioning solutions are stated as multipath propagation effects, limitations in large-scale applications, and maintaining system stability against changing environmental conditions [19]. To overcome these existing constraints, researchers are focusing on machine learning algorithms, the use of heterogeneous data sources, and hybrid models [25]. Additionally, non-RF-based solutions are noted to offer numerous advantages over RF-based systems due to their privacy and security, resilience to signal interference, and centimeter-level accuracy [26].

A literature review revealed no studies presenting current developments in this field. In this context, our aim is to classify current studies on Image-Based Positioning Systems, which stand out among indoor positioning technologies, under two headings: traditional methods and deep learning-based approaches, and to compare them based on accuracy, contribution, and limitations. Furthermore, the study aims to systematically analyze publications in this field over recent years to present the current trends in the research area.

## 2. Methodology

Within the scope of the study, Scopus, IEEE Xplore, Web Of Science and Springer Link databases were filtered according to the keywords (“vision-based”, “image-based”, “camera-based”, “visual”, “indoor localization”, “indoor positioning”, “indoor navigation”, “indoor tracking”, “visual SLAM”, “VIO”, “simultaneous localization and mapping”, “feature matching”, “image matching”, “feature-based”) and publications between the years 2020-2025 were collected and deduplicated according to their DOI numbers and publications in refereed journals and conference papers were included. Publications were filtered again appropriately according to the keywords (“visual”, “image”, “camera”, “SLAM”, “feature”, “indoor positioning”, “indoor navigation”, “indoor tracking”, “indoor localization”) and the exclusion process was performed. Additionally, because 2025 has not yet been completed, a linear regression model was used to estimate data from 2020 to 2024. **Figure 2** presents a flowchart regarding the acquisition of publications. **Figures 3 and 4** graphically display the number of publications by year.

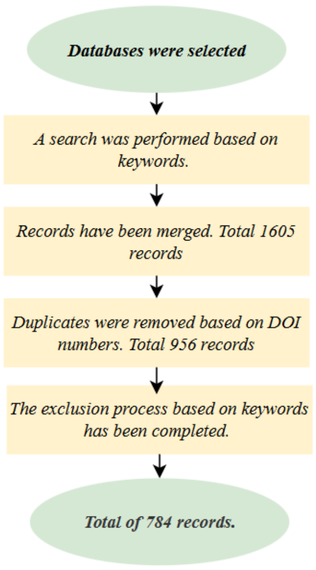
Databases were selected
A search was performed based on keywords.
Records have been merged. Total 1605 records
Duplicates were removed based on DOI numbers. Total 956 records
The exclusion process based on keywords has been completed.
Total of 784 records.

**Figure 2:** Collection of publications: Flowchart

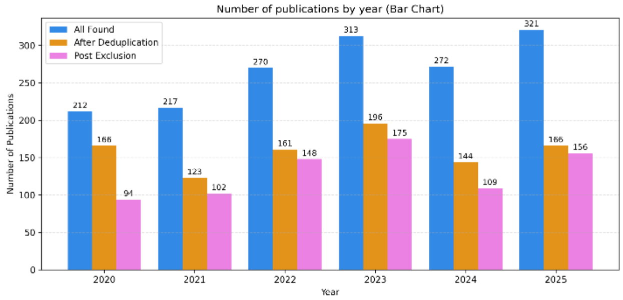

**Figure 3:** Number of publications by year bar chart

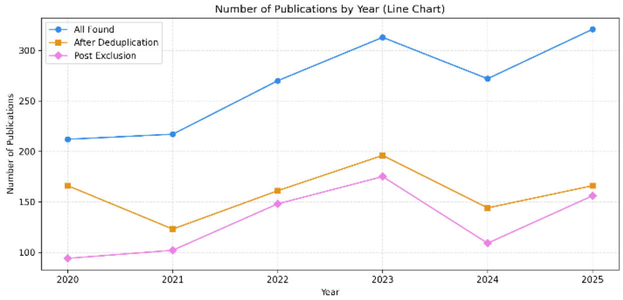

**Figure 4:** Line graph showing the number of publications by year

The studies obtained will be examined under two headings: “traditional-based” and “deep learning-based,” according to the methods they use. In addition, the research area in question will be analyzed according to trends over the years, publication titles, and keyword repetitions.

## 3. Literature Review

### 3.1. Image-Based Positioning in Indoor Environments

Indoor image-based positioning systems fundamentally perform the positioning process by comparing images with an existing reference database or using artificial markers. However, deep learning has recently introduced innovative methods in this field. This system utilizes the best matches or the geometric relationships between detected features and other known points to determine the user's location [27][28]. In this context, the advantages of these systems are that the methods used do not require additional infrastructure and that their application costs are relatively lower compared to other methods [29].

### 3.1.1. Traditional Feature Extraction-Based Methods

Feature extraction in systems developed using traditional methods involves the process of identifying distinctive points (such as color and texture) that represent the semantic and geometric properties of an existing image. This process reduces the size of the image data, creating a semantic and processable representation of the data for computer vision or machine learning algorithms. Feature extraction is a preprocessing step that significantly affects the performance of operations such as classification and object recognition [30] [31].

Feature-based image processing;

- Detection of distinctive and recurring key points in an image,
- Expression of these points as mathematical descriptors (such as SIFT, ORB) representing the distinctive features of the visual content,
- Matching of images using similarity measures based on these descriptors,

includes the following steps [32].

In a study conducted by Patruno C. et al., a high-accuracy, low-cost system using a monocular camera was presented to enable the positioning of autonomous vehicles capable of moving freely in closed industrial environments. The system utilizes the SURF (Speeded-Up Robust Features) algorithm for keypoint detection and feature extraction, and Euclidean distance and direction histograms are used to eliminate erroneous values. The study achieved an error rate of 0.21% (at 17.2 m) on wooden floors and 0.94% (at 80 m) in industrial areas, with an average relative positioning error of approximately 3%. Real-time operation was achieved at a processing frequency of 35-70 Hz. However, the system is limited by the difficulty of feature extraction on weak-textured and reflective surfaces and the accumulation of deviation errors over long distances [33].

In a study conducted by Zhang, T., et al., point and line feature extraction were integrated to improve the weaknesses of traditional SLAM (Simultaneous Localization and Mapping) systems and ensure real-time operability. In the study, erroneous features were reduced using the Bilateral filter. SURF was used for the extraction of point-based features, while the FLANN library was used for their matching. For the processing of line features, the LSD algorithm was applied together with geometric constraints.The study achieved an approximate 66.6% increase in line matching speed and, using the proposed method, produced 51.3% fewer errors compared to VINS-Mono and 58% fewer errors compared to PL-VIO in real-world tests. The method in question has been limited in areas with excessive darkness and repetitive textures, along with the computational load in low-power systems [34]. With the same objective, Sun, X., et al. proposed a SLAM algorithm that combines point and line features. In the proposed algorithm, ORB (Oriented FAST and Rotated BRIEF) was used for extracting point features and LSD for extracting line features. The results of the study show that the proposed method achieves lower error rates of up to 92.8% and 75% compared to methods such as ORB-SLAM2 and LSD-SLAM, while also successfully demonstrating real-time operation (30 FPS) and semi-dense point cloud mapping. However, extremely dark environments and high-speed movements are noted as limitations of the system [35]. Similarly, a smartphone-based, infrastructure-free method has been proposed to overcome the shortcomings of SLAM technologies in environments with weak and repetitive textures. The proposed method involves extracting features from images captured by the phone camera using the ORB algorithm and filtering mismatches using the HMM (Hidden Markov Model) model. The system's limitations include its high processing power requirements, its failure to account for moving environmental conditions, and its difficulty in environments with repetitive structures [36].

A study conducted by Zhang, X., et al. the goal is to achieve continuous and highly accurate positioning indoors. The study was based on matching video frames captured by a smartphone camera with geotagged images. The method used involved extracting feature points using the SIFT (Scale-Invariant Feature Transform) algorithm and minimizing mismatches using the RANSAC (RANdom SAmple Consensus) algorithm. Dividing the closed environment into subspaces and modeling it with a graph structure increased the efficiency of the method. The system, tested using 484 geotagged images collected in a closed environment, resulted in an error of ~0.4 m in image matching, ~0.7 m in offline positioning, and ~0.9 m in online positioning. However, the system was limited by the requirement for the camera to be continuously on and poor performance in low-texture areas [37]. In another study using both SIFT and RANSAC algorithms, the aim was to overcome the limitations of the classical RANSAC algorithm. To this end, features were extracted using the SIFT algorithm, and high accuracy was achieved using methods such as fundamental matrix calculation and slope estimation using epipolar geometry. The developed method reduced positional error by 40% compared to the classical RANSAC algorithm. Furthermore, successful results were achieved even in low-resolution images with 30-40% fewer iterations. The study resulted in a positioning error below 75 cm and a computation time below 3 seconds. The limitations of the developed system were noted as having few feature points, being sensitive to dynamic environmental conditions, and having high computational costs [38].

In a study conducted by Liu, X., et al., a system was developed using RGB-D camera images. In the developed system, the ORB algorithm was used for feature point extraction, the BoVW (Bag of Visual Words) model was used for image indexing, and the PROSAC (PROgressive SAmple Consensus) algorithm was used for matching. The positioning error obtained as a result of the study was 13.9 cm (room) and 12.9 cm (corridor). However, the developed method was limited in repetitive and low-texture environments [39]. In another study conducted with a similar objective, an SLAM system using a depth camera was designed to determine the position of mobile robots in indoor environments. This system uses the ORB algorithm to detect feature points. To improve the distribution and make it homogeneous, these points were processed using a Quadtree structure. The matching accuracy was increased by integrating the Hamming distance and cosine similarity for feature matching, and the RANSAC algorithm was used to minimize mismatches in the system. The developed method showed a 33% better matching performance compared to the original ORB, and the average positioning error was measured as 0.027 m [40].

The method proposed by Zhou et al. utilizes ORB in the feature extraction stage and the LSD algorithm in the matching stage of these extracted features. The research was tested using 136 labeled images collected from a shopping mall environment. The ORB algorithm performed feature extraction in 0.06 seconds, while LSD performed line processing in 0.09 seconds. Experimental results have shown that the system offers significant advantages over traditional methods in terms of both speed and accuracy, and that it is suitable for practical applications with a positioning error of less than 5 meters. The proposed solution offers positive features such as low computational power requirements and high processing speed, but it also has disadvantages such as limited robustness against small-scale training data and variable environmental conditions [41].

In a study conducted by He, Y., et al., they presented a stepwise feature matching algorithm to solve problems encountered by UAVs in indoor environments, such as insufficient feature matching, high rates of matching errors, and loss of accuracy in low light conditions. The proposed algorithm is an enhanced FAPP algorithm that incorporates adaptive FAST (Features from Accelerated Segment Test) thresholding (by calculating dynamic threshold values based on local brightness, texture complexity, and standard deviation), feature thinning using DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clustering, geometric clustering, and adaptive threshold optimization. The developed system, together with IMU integration, has provided more robust and accurate positioning. Test results show that the developed method outperforms ORB with a 96.5% correct matching rate, 13.2 ms/frame processing time, and 1.2 pixel RMSE value. Indoor flight tests yielded a deviation of 0.03 m on the X/Y axis and an error close to zero on the Z axis [42].

| Method | Accuracy and Performance | Contributions and Limitations |
|---|---|---|
| Monocular Camera + SURF + Histogram Filtering + CLL (Conditional Log-Likelihood) [33] | 17.2 m at 21% error<br>80 m at 94% error<br>Average relative positioning error: ~3% | It has been tested on different surfaces and proven to be applicable. Performance may decrease on low-textured and reflective surfaces. |
| EuRoC Dataset + SURF + LSD + FLANN (point) + Geometric Constraints and RANSAC (line) + Bilateral Filtering [34] | According to VINS-Mono/PL-VIO, 22% higher positioning accuracy, 51.3/58% fewer errors, and 66% faster line matching. | Real-time SLAM RGB-D-based high-precision positioning with point-line fusion. Extremely dark and repetitive environments are the limitations. |
| TUM RGB-D, ICL-NUIM Feature sets + ORB + Octree Filtering + LSD + LBD (Line Binary Descriptor) + Depth Integration [35] | UPL-SLAM (recommended method) has a significantly lower ATE (0.019 m) compared to ORB-SLAM2 (26.9%) and LSD-SLAM (92.8%). | Superior accuracy has been achieved through feature fusion and semi-dense mapping. Performance degradation and computational complexity have been observed under challenging conditions. |
| Smartphone camera images + ORB + DBoW2+ HMM + PnP [36] | HMM, Direct, and Hloc, with RMSE (0.48 m) that is 70.9% and 91.1% lower, respectively | HMM does not require infrastructure and offers high accuracy. Database dependency is one of the limitations of feature extraction in low-resolution environments. |
| Geo-Tagged Database and Smartphone Sensors + SIFT + Symmetry and Ratio Constraint + RANSAC + SFM-based PDR [37] | Image matching: ~0.4 m, offline/online positioning: ~0.7 m/~0.9 m. Computation time: 1.8 seconds per successful match, 0.59 seconds on average per position estimate. | The system, which operates without infrastructure and switches to PDR in the event of visual data loss, relies on continuous camera use and requires an initial position and labeled data set. |
| ETH3D Image Library + SIFT + Adaptive Ransac + Slope Similarity [38] | Average positioning error <75cm<br>Calculation time < 3 seconds | A system that is 30-40% less iterative and noise-resistant than classic RANSAC. It is limited by feature scarcity and database update difficulties. |
| Kinect V2 RGB-D Kamera + ORB + BoVW + Hamming Mesafesi + PROSAC + EPnP (Efficient Perspective-n-Point) [39] | EPnP can achieve an accuracy level of 20 cm with less positional error than epipolar constraint. | BoVW simplifies the creation of 3D databases, but it requires pre-calibration, performs poorly in low-texture/repetitive scenes, and requires database updates. |
| Oxford Optical Image and TUM dataset + ORB + Quadtree + Hamming Distance and Cosine Similarity + ICP (Iterative Closest Point) [40] | Oxford has a 53% better feature distribution and 33/6.5 higher accuracy compared to others. An average error of 0.027 m was confirmed with TUM. | This low-cost and easy-to-set-up method exhibits limited performance in low-texture environments and during fast camera movements, despite its improved feature distribution. |
| Data set consisting of 136 images + ORB + LSH [41] | Image matching success has been demonstrated with a processing time faster than 0.6 seconds and an error margin of 5 meters. | Despite its advantages of real-time positioning and low cost, the method is limited in dynamic environments and should be tested on large datasets. |
| CVPR and Oxford Affine Dataset + ORB + DBSCAN + FAPP [42] | The test results have been verified with a 96.5% matching rate and 0.02 m accuracy. | Despite its performance in low light and complex textures, the method is sensitive to dynamic obstacles, is not optimized for millimeter accuracy/computational speed, and has not been validated with real-time tests. |

**Table 3.** Traditional-Based Method

### 3.1.2. Deep Learning-Based Methods

Deep learning-based methods used in image-based indoor positioning can extract feature points from visual or synthetic images. While these methods offer alternative and efficient solutions for precise positioning indoors, they can be applied to various data types such as depth and RGB [43].

In a study conducted by Uygur et al., a method utilizing sparse semantic information to enable indoor positioning was proposed. Unlike traditional methods, the proposed method was developed using semantic information from sparsely distributed objects such as doors, tables, and windows, rather than point clouds or dense data structures. The developed system used a spherical camera with a 360° field of view, an IMU (Inertial Measurement Unit), and a Hokuyo range finder. This study performed the object recognition and categorization task using the Tiny YOLOv2 deep learning model, while the positioning process was supported by relative angle calculation and the MCL (Monte Carlo Localization) technique. The mathematical modeling of the relationships between objects was established using the PoE (Products of Experts) method, while IMU (inertial measurement unit) data underwent advanced processing to minimize measurement noise. Experimental results show that an average position error of 0.37 ± 0.15 meters and an angle error of 0.21 ± 0.25 radians were obtained in the scenario where all object classes were used. On the other hand, object classification difficulties arising from similar environmental features and cumulative errors in IMU data constitute the main limitations of the study [44]. This study adopts a semantic information-based approach, utilizing the Region-Based Fully Convolutional Network (R-FCN) for object recognition and classification. The method first semantically labels objects and then employs the SURF algorithm for feature extraction within these regions. and performed the matching phase by using epipolar geometry and semantic relationships as constraints only through semantically relevant features. The results of the study show high classification accuracy on an object basis (99.21% for doors and 90.91% for trash cans), a significant 80% reduction in the number of feature points to be processed compared to traditional methods, and impressive performance with maximum positioning errors measured at 46 cm, 59 cm, and 68 cm in three different test scenarios [45].

In a study conducted by Rostkowska, M., et al., the KNN (K-Nearest Neighbors) algorithm was used for feature matching, and an efficient, real-time system was presented for low-cost service robots. In the proposed system, lightweight CNN (Convolutional Neural Network) architectures such as EfficientNet and MobileNet were used to extract feature points, and global descriptors were extracted from images. The study achieved up to 98% accuracy and 0.20-0.27 m positional error with EfficientNet V2L, and it was reported to be implementable at 3.13 FPS on Jetson TX2. The developed system consumed fewer resources while providing similar accuracy compared to complex models such as NetVLAD. However, metric errors may increase in sparse maps, and additional training data may be required for adaptation to different environments [46].

In the system presented by Zhou, T., et al., face images collected specifically for the Pascal VOC dataset were used. In the current study, an improved CNN (MSE-CNN) architecture was used for feature extraction, and the PnP algorithm was used for feature matching. The developed model achieved 98.1% recognition accuracy in a test using an indoor dataset with markers. Furthermore, MSE-CNN significantly outperformed the traditional Faster-RCNN model in terms of FPS [47]. The proposed method, which aims to achieve high-accuracy camera localization from a single RGB image in repetitive or sparse indoor environments, utilizes the 7-Scenes and 12-Scenes datasets. In the study, a ResNet-based feature extractor, the CoordConv scheme to reduce ambiguity between similar image patches in repetitive and sparse texture areas, the DFAM module to combine feature maps at different levels, and uncertainty modeling approaches were used. The study achieved high accuracy of up to 99.6% and low error rates (0.014m (12-Scenes)) in areas with repetitive textures, providing real-time performance. The proposed method outperformed methods such as DSAC++ and HSCNet [48].

Zhou, B., et al., proposed a deep learning-based system (Darkloc+) that uses thermal images to provide high-accuracy indoor positioning in low-light environments, overcoming the limitations of RGB-based methods. Supported by a Siamese network architecture, non-local attention mechanisms, and multi-task learning (absolute/relative pose estimation), the method achieved an average position error of 0.490 m and an orientation error of 9.201' in a corridor environment, in room environments (0.469 m position error and 25.542' orientation error), and in the mobile robot dataset with an average error of 0.314 m and orientation error of 4.004'. It achieved a 53.4% improvement in position accuracy compared to PoseNet and a 47.2% improvement compared to HourglassNet. The use of the combined loss function (G+C+R) outperformed GlobalLoss [49], achieving an improvement of 0.092m in position error and 0.752° in orientation error. In a parallel study, indoor positioning performance was improved using an LSTM-based RNN and a position-focused loss function, achieving an error range of 0.15-0.18m on the Microsoft 7-Scenes dataset and outperforming competing methods (with an average error of 0.16m). However, high computational cost, sensitivity to light changes, and generalization issues indicate that multi-camera systems and Transformer-based approaches should be evaluated in future studies [50].

Wang, C. et al. aimed to overcome the high cost and complexity issues of indoor positioning systems with the system they developed. The two-stage system, which utilizes technologies such as SuperPoint, MobileNet V3-Small, VLAD, and SuperGlue, has achieved high accuracy with an average error margin of 0.12 meters and 0.15 meters in 90% of cases, with the ability to operate in real time on mobile devices. Despite advantages such as hardware independence and low resource consumption, the system has limitations such as overly similar environments, low light conditions, and generalization difficulties [51]. In a similar study, a lightweight system was developed for real-time indoor fire detection and localization for embedded systems (Jetson Nano). Based on the FCOS architecture and using CSPNet (Cross Stage Partial Network), PAN (Path Aggregation Network), and GFL (Generalized Focal Loss), the system demonstrated localization performance with an average error margin of 0.44 m using stereo image processing. The model, trained with a specialized dataset, achieved accuracy comparable to YOLOv3 (95.4% AP50) while being 10 times smaller in size (7.7-26.69 MB) and processing at 14 FPS. The study has limitations such as a limited data set, the inability to distinguish some fires, and the need for specialized hardware [52]. In a study, the Selective Optimal Network (SoN) method was presented, which reduces the computational cost for real-time indoor positioning using a smartphone. The proposed method improves both speed and accuracy by selecting a shallow or deep network structure (such as MobileNet, ResNet) based on the complexity of the input image. Trained with semi-supervised learning and a complexity indicator, the system achieves up to a

78.59% increase in speed and a 1.5% improvement in accuracy, while being able to work in conjunction with smartphone cameras and depth sensors. The study has limitations such as environmental similarity, lighting conditions, and generalization [53].

| Method | Accuracy and Performance | Contributions and Limitations |
|---|---|---|
| Spherical Camera and IMU + 2D Annotated Map + Tiny YOLOv2 + MCL + PoE [44] | Object-based positioning has achieved high accuracy with an error of 0.37 m (position) and 0.21 rad (angle). | The method offering visual obstacle resistance with semantic 2D mapping and a 360° camera is limited due to IMU errors and object recognition difficulties in dynamic environments |
| Data Collected with RGB-D Camera (1000) + R-FCN + SURF + Epipolar Geometry and Graphics-Based Ranking [45] | The system, supported by door (99.21%) and trash can (90.91%) semantic classification accuracy, achieved positioning accuracy of 46-68 cm in three scenarios. | Semantics-focused optimization reduced attributes by 80% and increased accuracy, but limitations exist in low object recognition accuracy, computational complexity, and performance in dynamic environments. |
| Data collected with Labbot Robot and Professional Camera and COLD Freiburg dataset + EfficientNet and MobileNet + KNN [46] | EfficientNetV2L achieved 3.13 FPS performance on Jetson TX2 with 98% accuracy and 0.20-0.27 m error. | This work has directly processed catadioptric images by presenting an energy-efficient CNN for low-resource devices and has increased generalization with balanced data. However, performance degradation occurs in different environments, and FPS optimization is required. |
| Pascal VOC and manually collected face dataset + CNN + PnP [47] | It has been verified with a 98.1% recognition accuracy, exceeding the traditional model by 1.6%. | The method that outperforms traditional Faster R-CNN is limited due to the lack of real-world testing and data dependency. |
| 7-Scenes and 12-Scenes Datasets + CoordConv Scheme + ResNet18 + DFAM + RANSAC-based PnP [48] | A remarkable success was achieved in the 12-Scenes dataset with 99.6% accuracy and an error of 0.014 m. | Lightweight architecture that provides superior localization on repetitive textures and sparse textures has limitations in generalization and real-time deployment due to specific data dependency, blur sensitivity, and RGB constraints. |
| TI-SLAM and Mobile Robot Dataset + ResNet34-Based Encoder Decoder Network + Multi-Task Learning + Geometric Loss Functions + Siamese Network Architecture [49] | In three different scenarios (corridor, room, mobile robot), consistent performance was demonstrated with error values of 0.490m/9.201', 0.469m/25.542', and 0.314m/4.004', respectively. | The method, whose accuracy has been proven with thermal semantic features, is limited due to tissue deficiency, thermal data limitations, and the difficulty of collecting large-scale data. |
| Microsoft 7-Scene Dataset + ResNet34 + RNN(LSTM) + Euclidean Loss Function. [50] | The system, which outperforms existing methods with an average error of 0.16 m, has optimized position estimation with an accuracy of 0.15-0.18 m. | Optimization reaching an error of 0.16 m, although suitable for indoor applications, makes practical use difficult due to high GPU requirements, limited data sets, and a single camera-focused structure. |
| Manually collected data + Superpoint + VLAD + MobileNet V3-Small + SuperGlue + Homography Disassembly [51] | An average error of 0.12 m and a 90% error distribution below 0.15 m demonstrate that the system provides both accuracy and reliability. | The system, which offers mobile compatibility and reduced data load with lightweight CNN, has disadvantages such as loss of accuracy in similar environments, lack of generalization, and inconsistency on low-performance devices. |
| Custom-built fire dataset +FCOS + PAN + GFL+Stereo Camera Calibration [52] | 0.44 m error, equivalent accuracy to YOLOv3 (95.4% AP50), and a balanced accuracy-speed ratio with 14 FPS performance on Jetson Nano. | High-accuracy and lightweight fire localization systems are limited by challenges such as data constraints, uncontrolled fires, and small flame detection. |
| College Dataset Complex Mall Dataset InLoc Dataset ImageNet Scene Classification Dataset + t-SNE and DBSCAN +SoN+ Adaptive Learning [53] | SoN outperformed traditional models with a 78.59% increase in speed and a 1.50% improvement in accuracy, achieving an accuracy rate of 0.914 on both datasets. | The SoN framework offers improvements in the speed-accuracy tradeoff and mobile compatibility, but it is limited by data constraints, dynamic environment incompatibility, and performance degradation with large images. |

**Table 4.** Deep Learning-Based Approaches

## 4. Results and Discussion

Due to the loss of GNSS effectiveness indoors, solutions have been sought for positioning in enclosed spaces. Although various technologies, algorithms, and methods have been used in the search for solutions, a system that is acceptable in terms of low cost, high performance, and applicability has not yet been developed. Among the technologies used, interest in image-based positioning systems has increased in recent years as an alternative to RF-based systems, which face various obstacles. While these systems offer advantages such as signal independence and easy installation, deep learning-based approaches have become the focus of researchers in this field in recent years.

As a result of the trend analysis conducted,

- The studies conducted between 2020 and 2021 generally focused on application-based indoor positioning systems, low-cost alternative solutions, and visual odometry,

- Research conducted between 2022 and 2023 highlights self-learning-based methods, stereo/monocular camera comparisons, wearable solutions, semantic information, and reinforcement-supported SLAM,
- In 2024-2025, the focus will be on solutions such as deep learning-based object recognition, RGB-D SLAM, multimodal fusion, and algorithms compatible with low-light environments.

The literature review reveals that deep learning-based approaches are superior to traditional methods in terms of accuracy, reliability, and speed. However, both methods still face challenges such as low-texture surfaces, moving environments, and variable lighting conditions, which are common difficulties for these systems. Furthermore, it is evident that the developed systems need to be tested in different fields and scenarios.

In conclusion, while image-based positioning systems offer solutions to the limitations of RF-based solutions, there are gaps within them that need to be addressed. In particular, the success of deep learning-based approaches is directly dependent on the diversity, sufficient size, and ability of the data sets used to represent real-world conditions. In this field, it is also essential to test systems developed under different environmental conditions and scenarios, develop models that work efficiently on real-time systems and mobile devices, and use sensor fusion methods to improve system stability.

## References

[1] Lee, H., Tak, J., & Choi, J. (2017). Wearable antenna integrated into military berets for indoor/outdoor positioning system. *IEEE Antennas and Wireless Propagation Letters*, *16*, 1919-1922.

[2] Huang, X. (2020). Multi-node topology location model of smart city based on Internet of Things. *Computer Communications*, *152*, 282-295.

[3] Potortì, F., Palumbo, F., & Crivello, A. (2020). Sensors and sensing technologies for indoor positioning and indoor navigation. *Sensors*, *20*(20), 5924.

[4] Barbieri, L., Brambilla, M., Trabattoni, A., Mervic, S., & Nicoli, M. (2021). UWB localization in a smart factory: Augmentation methods and experimental assessment. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1-18.

[5] Bhatta, B. (2021). *Global navigation satellite systems: new technologies and applications*. CRC Press.

[6] Paziewski, J. (2020). Recent advances and perspectives for positioning and applications with smartphone GNSS observations. *Measurement Science and Technology*, *31*(9), 091001.

[7] Egea-Roca, D., Arizabaleta-Diez, M., Pany, T., Antreich, F., Lopez-Salcedo, J. A., Paonni, M., & Seco-Granados, G. (2022). GNSS user technology: State-of-the-art and future trends. *IEEE Access*, *10*, 39939-39968.

[8] Fei, H., Xiao, F., Sheng, B., Huang, H., & Sun, L. (2019). Motion path reconstruction in indoor environment using commodity Wi-Fi. *IEEE Transactions on Vehicular Technology*, *68*(8), 7668-7678.

[9] Jiang, W., Cao, Z., Cai, B., Li, B., & Wang, J. (2021). Indoor and outdoor seamless positioning method using UWB enhanced multi-sensor tightly-coupled integration. *IEEE Transactions on Vehicular Technology*, *70*(10), 10633-10645.

[10] Liu, T., Li, B., Chen, G. E., Yang, L., Qiao, J., & Chen, W. (2023). Tightly coupled integration of GNSS/UWB/VIO for reliable and seamless positioning. *IEEE Transactions on Intelligent Transportation Systems*, *25*(2), 2116-2128.

[11] Cao, S., Lu, X., & Shen, S. (2022). GVINS: Tightly coupled GNSS–visual–inertial fusion for smooth and consistent state estimation. *IEEE Transactions on Robotics*, *38*(4), 2004-2021.

[12] Wang, Y., Pan, H., Qiu, L., Zhong, L., Liu, J., Ma, R., ... & Ren, J. (2024, December). Gpms: Enabling indoor gnss positioning using passive metasurfaces. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking* (pp. 1424-1438).

[13] Guo, Y., Zheng, J., Di, S., Xiang, G., & Guo, F. (2022). A Beacons Selection Method under Random Interference for Indoor Positioning. *Remote Sensing*, *14*(17), 4323.

[14] Zandbergen, P. A., & Barbeau, S. J. (2011). Positional accuracy of assisted GPS data from high-sensitivity GPS-enabled mobile phones. *The Journal of Navigation*, *64*(3), 381-399.

[15] Kunhoth, J., Karkar, A., Al-Maadeed, S., & Al-Ali, A. (2020). Indoor positioning and wayfinding systems: a survey. *Human-centric Computing and Information Sciences*, *10*(1), 1-41.

[16] Kim Geok, T., Zar Aung, K., Sandar Aung, M., Thu Soe, M., Abdaziz, A., Pao Liew, C., ... & Yong, W. H. (2020). Review of indoor positioning: Radio wave technology. *Applied Sciences*, *11*(1), 279.

[17] Mendoza-Silva, G. M., Torres-Sospedra, J., & Huerta, J. (2019). A meta-review of indoor positioning systems. *Sensors*, *19*(20), 4507.

[18] Quezada-Gaibor, D., Torres-Sospedra, J., Nurmi, J., Koucheryavy, Y., & Huerta, J. (2021). Cloud platforms for context-adaptive positioning and localisation in GNSS-denied scenarios—A systematic review. *Sensors*, *22*(1), 110.

[19] Subedi, S., & Pyun, J. Y. (2020). A survey of smartphone-based indoor positioning system using RF-based wireless technologies. *Sensors*, *20*(24), 7230.

[20] Yang, Y., Chen, M., Blankenship, Y., Lee, J., Ghassemlooy, Z., Cheng, J., & Mao, S. (2024). Positioning using wireless networks: Applications, recent progress and future challenges. *IEEE Journal on Selected Areas in Communications*.

[21] Wu, Y., Lin, J., Chen, H., Lan, H., & Yang, L. (2025). A transformer-based double-order RFID indoor positioning system. *Expert Systems with Applications*, *271*, 126530.

[22] Demiral, E., Karaş, İ. R., & Turan, M. K. (2013). Rfid sistemleri ile konum belirleme uygulamaları. *TMMOB Harita ve Kadastro Mühendisleri Odası*, *14*, 14-17.

[23] Guo, X., Ansari, N., Hu, F., Shao, Y., Elikplim, N. R., & Li, L. (2019). A survey on fusion-based indoor positioning. *IEEE Communications Surveys & Tutorials*, *22*(1), 566-594.

[24] Albraheem, L., & Alawad, S. (2023). A hybrid indoor positioning system based on visible light communication and bluetooth RSS trilateration. *Sensors*, *23*(16), 7199.

[25] Wu, Y., Lin, J., Chen, H., Lan, H., & Yang, L. (2025). A transformer-based double-order RFID indoor positioning system. *Expert Systems with Applications*, *271*, 126530.

[26] Alam, F., Faulkner, N., & Parr, B. (2020). Device-free localization: A review of non-RF techniques for unobtrusive indoor positioning. *IEEE Internet of Things Journal*, *8*(6), 4228-4249.

[27] Yang, S., Ma, L., Jia, S., & Qin, D. (2020). An improved vision-based indoor positioning method. *IEEE Access*, *8*, 26941-26949.

[28] Chen, C., Chen, Y., Zhu, J., Jiang, C., Jia, J., Bo, Y., ... & Hyyppä, J. (2024). An up-view visual-based indoor positioning method via deep learning. *Remote Sensing*, *16*(6), 1024.

[29] Park, H., & Mu Lee, K. (2017). Joint estimation of camera pose, depth, deblurring, and super-resolution from a blurred image sequence. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 4613-4621).

[30] Kumar, G., & Bhatia, P. K. (2014, February). A detailed review of feature extraction in image processing systems. In *2014 Fourth international conference on advanced computing & communication technologies* (pp. 5-12). IEEE.

[31] Mutlag, W. K., Ali, S. K., Aydam, Z. M., & Taher, B. H. (2020, July). Feature extraction methods: a review. In *Journal of Physics: Conference Series* (Vol. 1591, No. 1, p. 012028). IOP Publishing.

[32] Murat, I. S. I. K. (2024). Comprehensive empirical evaluation of feature extractors in computer vision. *PeerJ Computer Science*, *10*, e2415.

[33] Patruno, C., Colella, R., Nitti, M., Reno, V., Mosca, N., & Stella, E. (2020). A vision-based odometer for localization of omnidirectional indoor robots. *Sensors*, *20*(3), 875.

[34] Zhang, T., Liu, C., Li, J., Pang, M., & Wang, M. (2022). A new visual inertial simultaneous localization and mapping (SLAM) algorithm based on point and line features. *Drones*, *6*(1), 23.

[35] Sun, X., Zhao, Y., Wang, Y., Li, Z., He, Z., & Wang, X. (2024). UPL-SLAM: Unconstrained RGB-D SLAM With Accurate Point-Line Features for Visual Perception. *IEEE Access*.

[36] Zhou, C., Kuang, Y., Hu, T., & Zhang, X. (2025). Research on scene matching positioning technology for large-scale indoor scenes. *Urban Lifeline*, *3*(1), 9.

[37] Zhang, X., Lin, J., Li, Q., Liu, T., & Fang, Z. (2020). Continuous indoor visual localization using a spatial model and constraint. *IEEE Access*, *8*, 69800-69815.

[38] Bai, J., Qin, D., Ma, L., & Teklu, M. B. (2021). An improved ransac algorithm based on adaptive threshold for indoor positioning. *Mobile Information Systems*, *2021*(1), 2952977.

[39] Liu, X., Huang, H., & Hu, B. (2022). Indoor Visual Positioning Method Based on Image Features. *Sensors & Materials*, *34*.

[40] Yin, Z., Wen, H., Nie, W., & Zhou, M. (2023). Localization of mobile robots based on depth camera. *Remote Sensing*, *15*(16), 4016.

[41] Zhou, X., Meng, B., Dong, Y., Huang, X., & Zhang, K. (2021, October). An efficient image-based indoor positioning approach using ORB and LSH. In *2021 China Automation Congress (CAC)* (pp. 2985-2989). IEEE.

[42] He, Y., & He, X. (2025). Research on an Improved Stepwise Feature Matching Algorithm for UAV Indoor Localization. *IEEE Access*.

[43] Castillo-Cara, M., Martínez-Gómez, J., Ballesteros-Jerez, J., García-Varea, I., García-Castro, R., & Orozco-Barbosa, L. (2025). MIMO-Based Indoor Localisation with Hybrid Neural Networks: Leveraging Synthetic Images from Tidy Data for Enhanced Deep Learning. *IEEE Journal of Selected Topics in Signal Processing*.

[44] Uygur, I., Miyagusuku, R., Pathak, S., Moro, A., Yamashita, A., & Asama, H. (2020). Robust and efficient indoor localization using sparse semantic information from a spherical camera. *Sensors*, *20*(15), 4128.

[45] Jia, S., Ma, L., Yang, S., & Qin, D. (2021). Semantic and context based image retrieval method using a single image sensor for visual indoor positioning. *IEEE Sensors Journal*, *21*(16), 18020-18032.

[46] Rostkowska, M., & Skrzypczyński, P. (2023). Optimizing appearance-based localization with catadioptric cameras: small-footprint models for real-time inference on edge devices. *Sensors*, *23*(14), 6485.

[47] Zhou, T., Ku, J., Lian, B., & Zhang, Y. (2022). Indoor positioning algorithm based on improved convolutional neural network. *Neural Computing and Applications*, *34*(9), 6787-6798.

[48] Xie, T., Dai, K., Wang, K., Li, R., Wang, J., Tang, X., & Zhao, L. (2022). A deep feature aggregation network for accurate indoor camera localization. *IEEE Robotics and Automation Letters*, *7*(2), 3687-3694.

[49] Zhou, B., Xiao, Y., Li, Q., Sun, C., Wang, B., Pan, L., ... & Li, Q. (2024). Darkloc+: Thermal image-based indoor localization for dark environments with relative geometry constraints. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1-12.

[50] Alam, S., Mohamed, F., & Hossain, B. (2025). Integrating Recurrent Neural Networks and Loss Function Optimization for Efficient Indoor Camera Positioning. *Computer Science*, *33*(1), 97.

[51] Wang, C., Bi, K., Zhao, B., Li, M., Chen, Y., Tao, S., & Yang, J. (2024). Lightweight Indoor Positioning System Based on Multiple Self-Learning Features and Key Frame Classification. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *10*, 373-379.

[52] Li, Y., Shang, J., Yan, M., Ding, B., & Zhong, J. (2023). Real-time early indoor fire detection and localization on embedded platforms with fully convolutional one-stage object detection. *Sustainability*, *15*(3), 1794.

[53] Lee, K., Lee, H., & Hwang, J. Y. (2025). SoN: Selective Optimal Network for smartphone-based indoor localization in real-time. *Expert Systems with Applications*, *272*, 126639.