Floor Segmentation Approach Using FCM and CNN

Kavya Ravishankar, Puspha Devaraj, Sharath Kumar Yeliyur Hanumathaiah*

Department of Information Science & Engineering, Maharaja Institute of Technology Mysore, 571477 Mandya, India

* Correspondence: Sharath Kumar Yeliyu Hanumathaiah (hodise@mitmysore.in)

Abstract: Floor plans play an essential role in the architecture design and construction, which serves as an important communication tool between engineers, architects and clients. Automatic identification of various design elements in a floor plan image can improve work efficiency and accuracy. This paper proposed a method consists of two stages, Fuzzy C-Means (FCM) segmentation and Convolutional Neural Network (CNN) segmentation. In FCM stage, the given input image was partitioned into homogeneous regions based on similarity for merging. In CNN stage, the interactive information was introduced as markers of the object area and background area, which were input by the users to roughly indicate the position and main features of the object and background. The segmentation evaluation was measured using probabilistic rand index, variation of information, global consistency error, and boundary displacement error. Experiments were conducted on real dataset to evaluate performance of the proposed model. The experimental results revealed the proposed model was successful.

Keywords: U-net; Fuzzy; Probabilistic rand index; Variation of information; Global consistency error; Boundary displacement error

1. Introduction

Floor segmentation has a wide range of applications in engineering field and is considered as a complex task in image processing [1, 2]. Certain patterns in the set of image pixels need to be recognized to segment the floor. Texture and color are both important characteristics. However, due to illumination conditions, finding specular reflections and shadows in floors is commonplace, which is one of the main difficulties faced by most segmentation algorithms. Nowadays there are lot of mathematical calculations done by computer vision techniques, which can be used for floor recognition. The floor recognition process can be considered as the preliminary approach for autonomous robot navigation and several other applications. In industrial environments the use of autonomous robots have large impact and varied applications. Because working in toxic and other hazardous environments is very difficult for human being, the work is preferred to be done by unmanned vehicles and robots. Proper navigation scheme and floor maps are required to assure smooth working of the system. More intelligent and automatic devices are capable of doing several tasks. Navigation requires not only path planning but also current location tracking.

A lot of studies were conducted on navigation of living beings in neuroscience history showed that the information captured from environment was stored as special information in human brain, including location data. When a human being walks through an environment, the previously stored spatial information creates a proximity sense, which decides how much the person moves forward and backward in the environment. Taking car driving as another example. The person diving the car normally has an idea of car dimensions and road area. There are some blind spots which the driver cannot see while sitting inside the car. The side mirrors and the camera were activated when the car moves backward, which enables the driver to see the free empty space and helps to calculate the path to move in any directions easily [3-5]. The navigation of robots and autonomous vehicles uses visual information captured by camera. These approaches replaced traditional schemes because of their robustness and simplicity. The system developed in these scenario needs to detect static and moving objects to guide robots within

the given area, which may be indoor or outdoor [6, 7]. Compared with other sensor-based navigation system, the visual based system has more advantages, particularly, lesser costs and higher effectiveness. The information acquired from those images are higher than the sensor based systems.

Floor segmentation is a difficult task in real-time scenario. The pixel content recognized from the captured scene is used for recognition. Due to illumination in different scenarios like indoor and outdoor, some features important to detect edges and corners are missing in the image captured. In some cases the shadows cover some corners, whose segmentation is more complex. Reflections are one major issue affecting the segmentation. These issues cannot be avoided in real-world scenario but can be removed little bit by image pre-processing, such as brightness adjustment etc. This paper proposed a novel approach for floor segmentation, which overcomes all the limitations in previous research and mainly focuses on image pre-processing and segmentation in multiple steps.

The remainder of this paper was organised as follows: Section 2 began with a brief introduction of related research; Section 3 proposed floor region detection algorithm; Section 4 discussed the performance analysis of the proposed algorithm; Sections 5 and 6 showed the experimental results.

2. Literature Review

Incorporation of image processing framework into navigation opens up a new era for navigation technology, which is considered as one of the most important advancement in the field. Vision based navigation also includes moving robots, which have some features, such as segmenting floors, appliances, walls etc. Recent works have discussed different types of techniques used for floor segmentation and path planning. In this section, we discussed about those technologies and found there relevance. Adachi et al. [8] replaced costly three-dimensional image sensor with high definition webcam to acquire surrounding images of robot movement. They developed a road following robot with high accuracy and robustness and used semantic segmentation to identify boundaries of the movable area. The system initially detected the boundaries and estimated lines. Their study developed the target point through intersection between two lines and considered a centroid for target path. The centroid point was updated according to changes in boundary lines. They tested the system in an indoor lab, whose performance was good in obstacle detection and corner detection. But the system needed to be tested in a complex outdoor and indoor scenario with more corners and obstacles.

Barcel´o et al. [9] presented a single grey scale image sequence based indoor navigation, which combined previously detected vertical and horizontal lines together for segmentation. The system detected several types of indoor scenes and its performance was not affected by camera movement. The system paid more attention to moving features. Features of moving objects in the scene were discarded in order to estimate the moving features of segmented area. Orientation of the camera was kept relative to ground to improve the result. The system was tested in an indoor environment and its performance was around 60%. One important issue in mobile navigation is the detection of free space. The study of Geovani Rodr´ıguez-Telles et al. [10] showed that the robot navigation was possible using a low resolution image. The system focused on free-space segmentation in front of camera for robot navigation. For capturing the environment scenes, monocular camera was used and placed downward to face the floor. The system captured low resolution images from real-time environment, required neither camera calibration nor optical flow ground plane constraints, and used simple linear iterative clustering super pixel algorithm for segmentation. Free spaces and objects were easily segmented using the system. Then the system was tested in several indoor environments. In some cases, such as objects with shadow, smaller size objects, dark areas, etc. were not segmented well, which affected the system performance. The proposed system could be applied to indoor environments, but was not suitable for outdoor real-time navigation.

Ma et al. [11] introduced a fixed camera based floor segmentation for robot navigation. Shaking initially caused the problems of video capturing procedure. When camera was placed on a moving robot, the chances of getting noisy image were high, which may reduce the performance. In their proposed system, the camera was fixed on the ceiling to capture video of the floor and moving robot. The segmentation process was implemented in the detection system of floor area and moving robot position. The system had some major limitations. For example, the number of fixed cameras depended on the area covered; obstacles or objects with the same colour as the floor were not identified. Wang et al. [12] proposed the floor segmentation framework based on gravity vector estimation and developed semantic segmentation for indoor environment analysis. In this proposed system, RGB-D image sensor was used for surface segmentation. The system detected floor, wall, and small objects using surface normal clustering and captured geometry information of the object from the depth information. Then the information was used to objects and surface accordingly. The system was only capable to detect indoor environments and tested according to indoor scenarios. The effect of shadow also affected the system performance in real-time experiment.

De Jesús Osuna-Coutiño and Martinez-Carranza [13] proposed a combined approach for floor segmentation made two types of analysis. The first one was floor connection, used for increasing recognition robustness. The second one was the recognition framework avoiding misrecognition. Experiments showed that the system performed better than previous work. Three different datasets were used for analysis using binary features. The system ran without Graphic Processing Unit (GPU), which made its application in low budget. Chen et al. [14]

proposed a deep learning approach for image segmentation. They applied atrous convolution for feature extraction and used the upsampling filters for increasing the features. The system provided accurate segmentation with proper boundaries for objects. The work was done through experiments with several complex datasets, which achieved sufficient result. One of the major limitation was the efficiency of the system. Some objects, such as bicycles, chairs etc. were not detected properly and their boundaries were not accurate. It was suggested to improve encoder and decoder with high resolution features.

Skoryk et al. [15] proposed computer vision floor segmentation methods, which showed multiple algorithms and segmentation quality. Considering classical computer vision approach and deep learning approach in a fusion format, they used two different datasets for analysis, NYUDv2 and SUN-RGB-D. They initially analysed using classical approach based on super pixels and then used deep learning techniques like CNN and Fast CNN. They finally fused two approaches for better performance because both approaches had limitations. But the system was effected due to lack of segmentation quality. Kida et al. [16] proposed an algorithm to detect objects and surface in real-time environment, which focused on detecting human bodies and properties. But surfaces and objects were also segmented using sampling methods. The free space was detected using the position vectors distributed in the area. When being tested in an indoor environment with one human standing and one moving, the system segmented human body and surface properly.

Du and Arslan [17] proposed a magnetic field based indoor positioning scheme. The segmentation based K nearest neighbour (KNN) algorithm was implemented in the proposed system for estimating location. This approach decreased the computational complexity by partitioning and selecting desired area for calculation. This partitioned target area was fed to KNN which increased the accuracy of the entire system. The experiment results showed 9.24% performance improvement than the standard KNN model. Bormann et al. [18] conducted a survey on multiple room segmentation techniques to determine the most effective one. The comparison showed the advantages and disadvantages of each algorithm. Four algorithms were selected for implementation with help of locally available ROC package. Their study helped new developers to find the best segmentation method. Honto et al. [19] developed a floor map segmentation system for indoor navigation, which captured the map of indoor area and detected the passage ways. A semi-automatic way was proposed to segment floor from the map. Information retrieved from the map was used for simple operation for users. Both grabcut method and snake method were used to segment passageway and other surface.

Ambruş et al. [20] introduced 3D point information based room segmentation approach, which reconstructed room information from raw points and created a floor plan with minimized energy consumption. The system initially detected primitive points from the 3D cloud data and then detected ceiling and wall area. The study detected openings using 2D projection and calculated last view point for reconstruction. Lu et al. [21] proposed floor planning approach for rural housing, which was a new method for room segmentation. Deep learning neural network algorithm was utilized to decode information for floor plan. 1D loop was used to detect the wall, window and door openings, which was considered as room. Then the room region was identified using text information. The system converted the floor plan to room layout which had all room attributes. The major limitation of the system was lack of semantic information of rooms. Fleer [22] created a room cleaning robot with panoramic camera and laser diodes. The idea was to develop a robot which considered the room as human being. The robot compared the image input and sensor data for floor detection and path planning for navigation. The surrounding video was captured by the panoramic camera and distance between objects was detected using lasers. The system was tested in real and simulated environments, which segmented the room environments. But the system failed in some room border detection, which effected the performance.

The above recent segmentation research was based on deep neural network and other models. However, due to storage limitations, it was not workable to implement segmentation neural network architecture on an embedded platform, such as a field-programmable gate array (FPGA). Therefore, apart from machine learning technology, this paper proposed floor region segmentation algorithm, which combined boundary identifying methods with classification topology to differentiate the floor region from non-floor region. The proposed algorithm first identified the boundaries of the object using FCM method. Then CNN classification topology distinguished the floor region from non-floor region. By far most of the vision-based floor regions detection algorithms are still dependent on the depth sensors or complex models with relatively more power consumption. This paper proposed a new floor estimating algorithm based on image segmentation using a single image, which combined extracting surface texture characteristics with a specific geometric area to discover object boundaries and distinguish between the floor and non-floor regions using CNN classification.

3. Proposed Method

The proposed method consisted of two stages: FCM and CNN segmentation. In FCM segmentation, the given input image was partitioned into homogeneous regions based on similarity for merging. In CNN stage, the interactive information was introduced as markers of the object area and background area, which were input by users to roughly indicate the position and main features of the object and background. The markers were in the form of simple strokes. The proposed method calculated the similarity between different regions and merged them

based on the proposed maximal similarity rule with the help of these markers. Then the objects were extracted from the background when the merging process ended. Figure 1 showed the proposed block diagram.
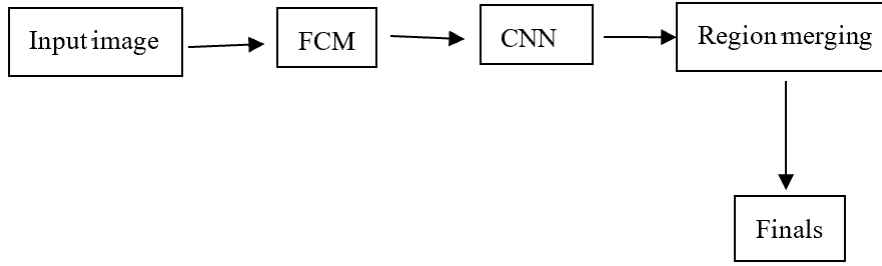

Figure 1. Block diagram of the proposed method

3.1 FCM

Clustering is the method of forming homogenous data by separating the data and considering the object relationship, which allocates v feature vectors into N clusters. Every n-th cluster has Cn as its center. Fuzzy clustering was employed in numerous areas, such as pattern recognition and fuzzy detection. FCM was extensively used among various kinds of fuzzy clustering processes. FCM utilized reciprocal distance to determine fuzzy weights. The input of this process was known amount of clusters N. The mean spot of every member of a cluster was identified. The output was segregating N clusters in the object class. The FCM clustering was performed to reduce the total weighted mean square error (MSE).

The FCM allowed every feature vector to match several clusters with different fuzzy membership values. The final segmentation was based on the feature vector's optimum weight in all clusters. The steps involved in FCM algorithm were given below.

Algorithm-1: FCM process
By implementing the clustering process FCM, the lung of the input chest computed tomography (CT) image was segmented into left and right parts.
Input: Feature vectors (image voxels) $v = \{v1, v2, \ldots, vn\}$, N = count of clusters.
Output: N, number of cluster groups with less sum of distance error.
Steps:
1. Fuzzy weighting was used to set random weight for every pixel, with positive weights $\{W^v_n\}$ ranging from 0 to 1.
2. The starting weights for each v-th voxel on all N clusters were normalized.
3. The weights on n= 1,…,N for each v were normalized to get W^v_n.
4. Estimated new centroids Cn, n = 1,…,N
5. Updated the weights $\{W^v n\}$
6. If the input was altered, repeat from step 3. Or the process should be stopped.
7. Every pixel was matched its corresponding cluster. The clustering process FCM segmented the image into left and right parts. Figure 2 shows the initial segmentation using FCM.

3.2 CNN

Among the available CNN, U-Net had its U-shaped encoder-decoder structure. It segmented images by down-sampling and up-sampling using the original images and took out the feature map similar to its actual type of image. Composed of contracting and expanding paths, U-Net was used in delineations needed for radiation treatment planning because it integrated local features with general location information of the object. Even with a large number of samples, Unet still produced relatively acceptable results. Both 2D and 3D format was used to construct it, with respective benefits and drawbacks. In the 2D approach, the U-net planning was practiced with input-output pair of part of 2D; a fractional architecture was fed by any 3D sub volume by replacing all 2D operations with equivalent 2D operations for 3D approach.

An accurate 3D U-net extension was accomplished by U-net only. The 3D U-net authorized to proceed in 3D sub volume and received output of each graphical information of 3D space in the volume specifying tumour expectation. 2D convolutions produced a single image by applying the same weights throughout the entire depth of the frame stack (many channels). In order to preserve temporal information of the frame stack, 3D convolutions employed 3D filters and resulted in a 3D volume. 2D U-net reduced the 3D direction of information because each image was treated separately. However, this system was capable of learning from several samples. The 3D

direction of information was enriched while the number of samples was reduced in 3D U-net, thus increasing the quantity of information per sample. Semantic segmentation involved labeling each pixel in an image or voxel in a 3D volume. Because 3D U-net offered the best segmentation for brain tumour sub-regions in Magnetic Resonance Imaging (MRI) modalities, 3D U-net was used for semantic segmentation of MRI image. The 3D U-net, made of a contractive (encoder path) and expanding (decoder path), used convolution and pooling to build a bottleneck in the middle of the path (Figure 3).
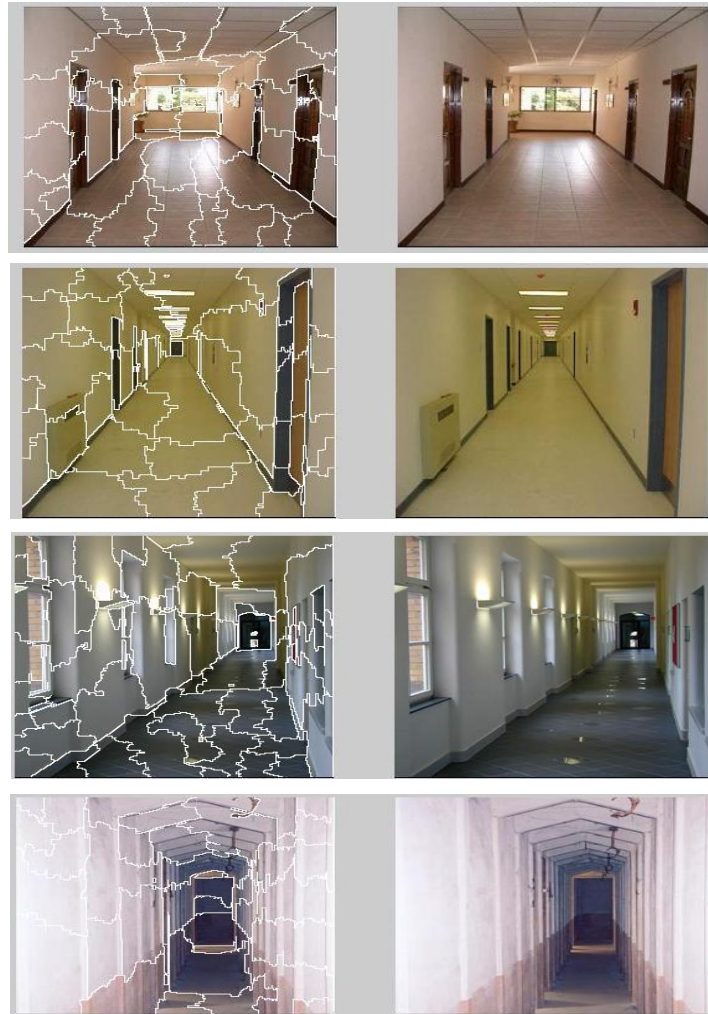

Figure 2. Set of floor images with homogeneous regions using FCM segmentation
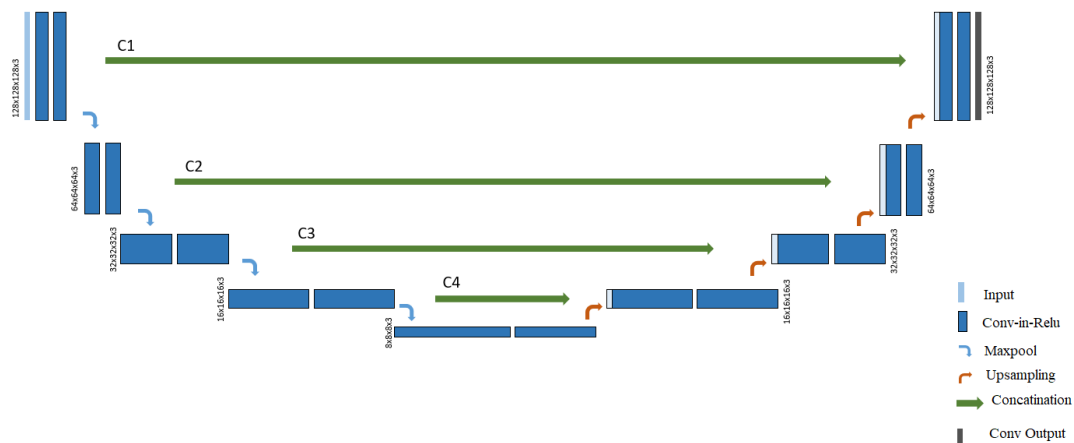

Figure 3. 3D U-net architecture

Convolutions and up-sampling were used to recreate the image after this bottleneck. As part of the deep convolution neural network, the 3D U-net segmentation presented a network and test approach based on application of data enlargement to given images, which was more effective. A contracting path was used to capture circumstances, and a balancing growing path was used for exact localization. A CNN-based architecture was frequently utilized to categories labels. But the goal in medical imaging should be more than just classification. Apart from localization, the goal was set to foretell the class label of each pixel using the context of its immediate surroundings as input. The image context, captured via the encoder path, was merely a stack of maximum pooling and convolutional layers. Depth of the context increased as the image size was gradually decreased by the encoder. This basically meant the network forgot to learn the "WHAT" information while learning the "WHERE" information in the image. The encoder network learned a conceptual representation of the input image using a sequence of encoder blocks, thus performing the function of a feature extractor.

Set of two 3x3 convolutions formed one block, which followed a rectified linear unit (ReLU) activation function. ReLU added nonlinearity to the network, assisting in the generalization of training data. The associated decoder block was the skip connection for the ReLU output. The ReLU activation function was followed by the dropout function. By deleting (ignoring) a few randomly selected neurons, the network was compelled to learn a new representation. Neurons became independent with the help of network. In turn, this promoted generalisation and prevented overfitting in the network. Then a 2x2 max-pooling was used to cut the structural dimensions (height and width) of the feature maps in half. The computational cost was decreased by lowering the number of trainable parameters. The bridge completed the information flow by connecting the encoder and decoder networks. The flow consisted of two 33-convolutional layers, with a ReLU activation function placed after each. The size steadily grew while the depth gradually lowered, with the decoder path providing precise localisation. A semantic segmentation mask was produced from the abstract representation using the decoder network. Starting with a 2x2 transpose convolution, the decoder block was activated. Then the block joined the relevant skip connection feature map for the encoder block. These skip connections offered functionality that occasionally lost owing to network depth. Then two 3x3 convolutions were utilised, with each followed by a ReLU activation function. The final decoded output was subjected to 1x1 convolutions with sigmoid activation. The pixel wise classification was represented by masked segmentation which was created by sigmoid activation function.

As a result, this network was entirely convolutional from beginning to end. To retrieve the "WHERE" information, the decoder gradually used up-sampling (precise localization). Output of the transposed convolution layers was concatenated with the feature maps from these encoders in order to leverage skip connections at each stage of the decoder to achieve more precise positioning. These skip connections gave the decoder additional data, enabling it to produce more precise semantic characteristics.

In addition, the skip connections acted as a shortcut connection, enabling gradients to pass to lower layers not in a degraded way. Skip connections, to put it simply, increased the gradient flow during backpropagation, enabling the network to learn better representation. After each concatenation, two consecutive regular convolutions were applied so that the model learned to put together a more precise result. Due to the size, complexity, and memory requirements, a full image could not be offered for training. As a result, the data was standardized, which generated random sub-volumes. The image we utilized had the size 128x128x128x3, but the dimensions of the input image were 256x256x256x3. As a result, the size varied from that of the original at various spots, but the essential element stayed the same.

4. Evaluation of Segmentation Approaches

Evaluation results of different evaluators varied significantly, because each evaluator may have distinct standards for measuring the segmentation quality. S1 and S2 were two images.

Rand index [23]

$$R\left(S_1, S_2\right) = \frac{1}{2^N} \sum_{\substack{i,j \\ i \neq j}} \left[I\left(l_i = l_j \wedge l_i^{'} = l_j^{'}\right) + I\left(l_i \neq l_j \wedge l_i^{'} \neq l_j^{'}\right) \right] \qquad (1)$$

where, I is the identity function, and the denominator is the number of possible unique pairs among N data points. This gave a measure of similarity ranging from 0 to 1.

Variation of information [24]

$$VI\left(S_{test}, S_K\right) = H\left(S_{test} \mid S_K\right) + H\left(S_K \mid S_{test}\right) \qquad (2)$$

The first term in the above equation measured the amount of information about S_{test} that we lost, while the

second term measured the amount of information about S_K that we had to gain, when going from segmentation S_{test} to ground truth S_K. Where, H (.|.) is the conditional entropy.

Global consistency error [25]

Global consistency error measured the extent to which the regions in one segmentation were subsets of the regions in second segmentation (i.e., the refinement). Let $R(S, p_i)$ be the set of pixels in segmentation S containing pixel P_i, then the local refinement error was defined as:

$$E\left(S_1, S_2, p_i\right) = \frac{\left|R\left(S_1, p_i\right) \setminus R\left(S_2, p_i\right)\right|}{R\left(S_1, p_i\right)} \tag{3}$$

Boundary displacement error (BDE) [25]

$BDE = min(\ |x - y_i|)$, where x∈B1 and y_i∈B2,i=1,2,...,n, n is the number of boundary points in B2.
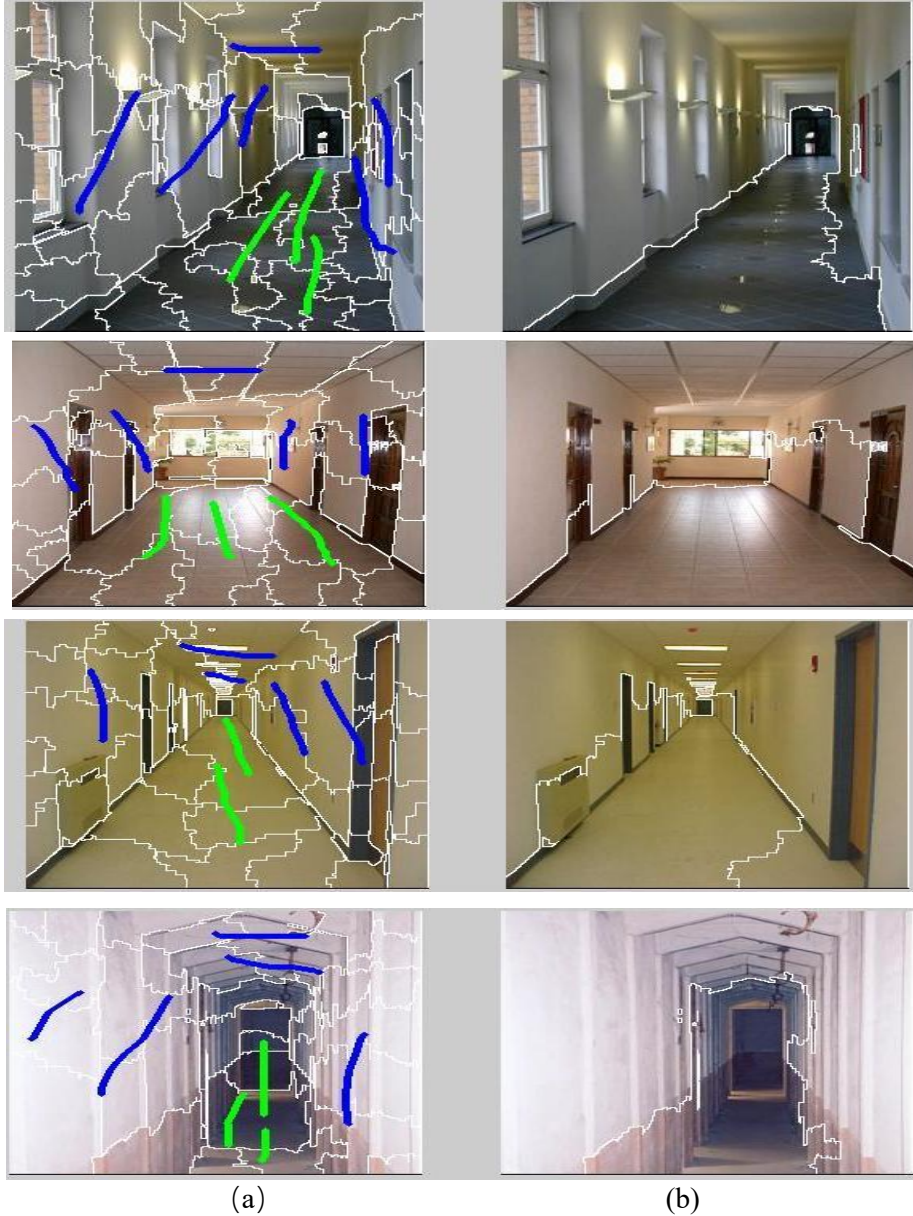
5. Experiment

Figure 4. (a) Floor images with foreground and background markers; (b) Extracted floor regions

To evaluate the segmentation results produced by different algorithms, a database was compiled, which contained 300 corridor images along with ground truth segmentations. The corridor images were taken in twenty different buildings, exhibiting a wide variety of different visual characteristics. Segmentation method was evaluated by assessing its consistency with the ground truth segmentation given by the human expert. Any evaluation metric desired should take into account the effects of over-segmentation and under segmentation. The former refered to the reference region represented by two or more regions in the examined segmentation. The latter refered to two or more reference regions represented by a single region in the examined segmentation. In accurate boundary localization, the ground truth was usually produced by humans that segmented at different granularities. Finally, when there were different numbers of segments, two segmentations needed to be compared. Table 1 shows the parameter values of different segmentation methods. An image should have higher PRI value and lower VOI, GCE and BDE values. Each parameter was described by ground truth and proposed method. Each row was represented by average of each class of totally about 100 images. According to the Table 1, the proposed method achieved 0.9725 PRI, 2.23 VI, 2.14 GCE and 1.24 BDE. Y-axis represented number of samples. The proposed method achieved good results. From this evaluation, it was found that region merging segmentation was well suited for the corridor images. Figure 4 showed examples of segmentation results obtained by the proposed floor segmentation method. Further, to evaluate the performance of the proposed method, we selected images from MIT Scene dataset [26] and implemented the proposed method (Figure 5). In addition, this paper analyzed the measures of PRI, VI, GCE and BDE in Table 1 in graphical representation for floor segmentation shown in Figures 6, 7, 8 and 9. Finally, our method was compared with other well-known methods, such as SegNet [27], Deep Lab and FCN [28], shown in Figure 10.

Figure 5. Images with floor segmentation for MIT Scene Dataset
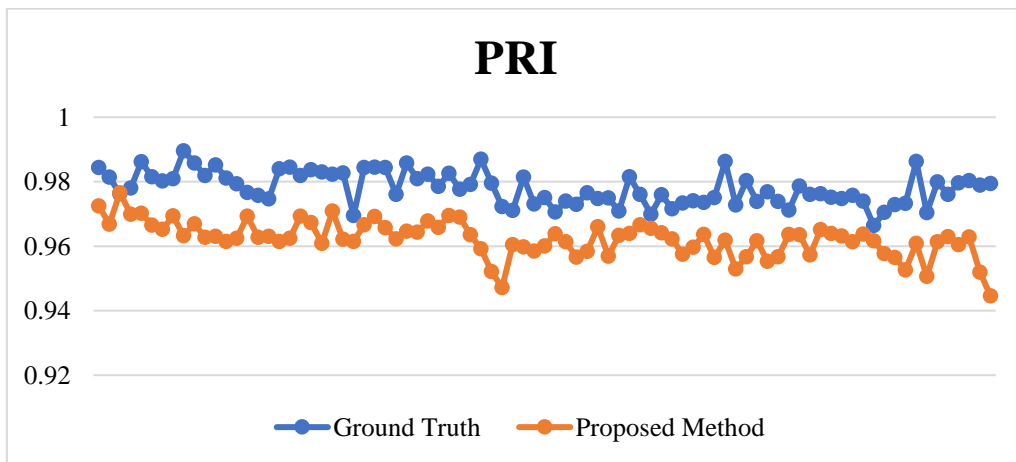

Figure 6. PRI measures for the proposed method
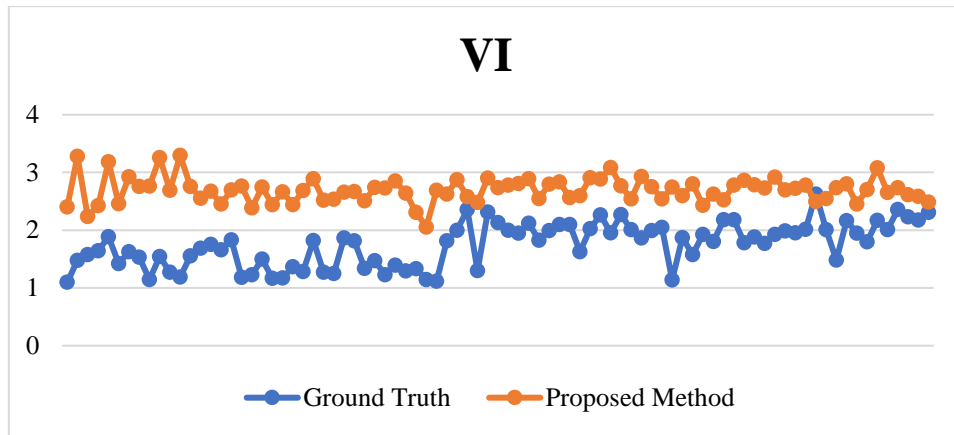
40

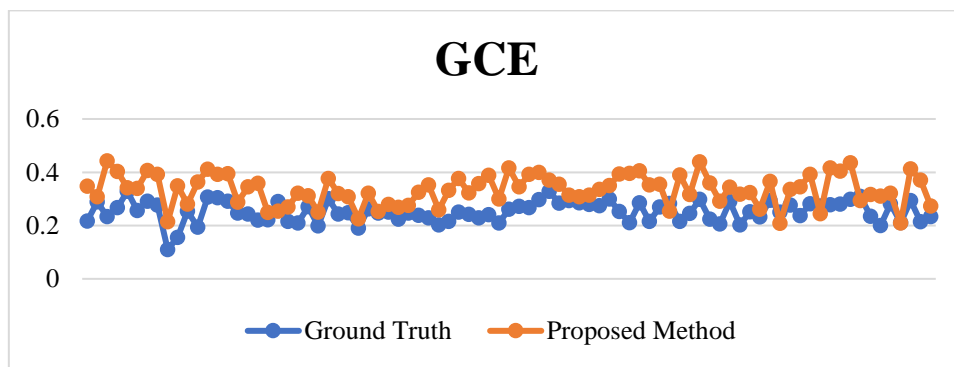**Figure 7.** VI measures for the proposed method

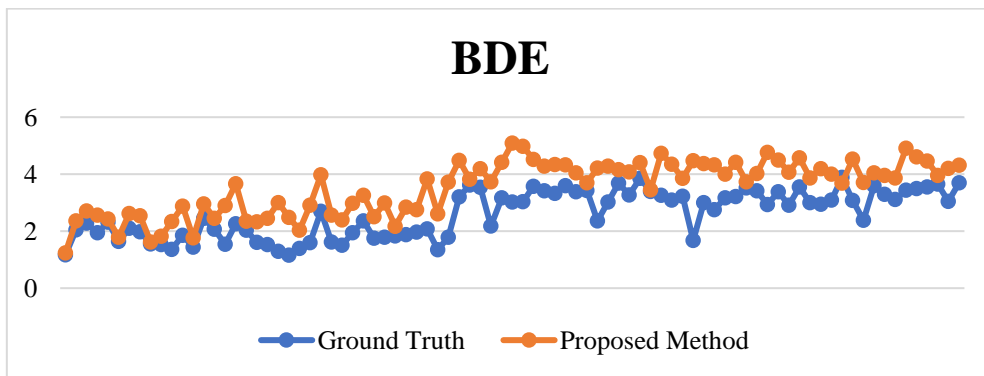**Figure 8.** GCE measures for the proposed method
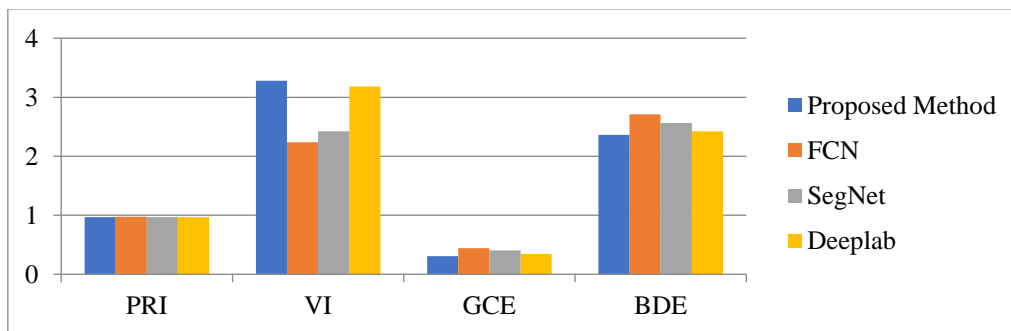

Figure 9. BDE measures for the proposed method

Figure 10. The proposed method with other well-known methods

Table 1. Segmentation evaluation results

PRI		VI		GCE		BDE	
Ground Truth	Proposed Method	Ground Truth	Proposed Method	Ground Truth	Proposed Method	Ground Truth	Proposed Method
0.9844	**0.9725**	1.0986	2.4053	0.2163	0.348	1.1726	**1.2421**
0.9814	0.9668	1.4768	3.2799	0.2886	0.3087	2.0427	2.3632
0.9763	0.9766	1.5755	**2.239**	0.2333	0.4429	2.2872	2.7106
0.978	0.9699	1.6447	2.423	0.2678	0.403	1.9413	2.5653
0.9862	0.9702	1.8866	3.1828	0.3283	0.3421	2.3115	2.4248
0.9815	0.9665	1.4235	2.4569	0.2568	0.3388	1.6375	1.7912
0.9802	0.9652	1.6247	2.9266	0.2914	0.4066	2.1042	2.6198
0.9809	0.9694	1.5332	2.7568	0.2774	0.3917	1.9846	2.5463
0.9896	0.9633	1.1464	2.7651	0.1104	**0.2144**	1.554	1.6153
0.9858	0.9669	1.5422	3.2584	0.1555	0.3495	1.5256	1.8215
0.9819	0.9628	1.2741	2.6944	0.2544	0.2803	1.3667	2.3399
0.9852	0.9631	1.1895	3.2977	0.194	0.3639	1.8585	2.883
0.9811	0.9615	1.5554	2.7562	0.3065	0.4117	1.4353	1.7689
0.9793	0.9625	1.6896	2.5537	0.3046	0.3925	2.4477	2.96
0.9767	0.9693	1.7558	2.6725	0.2913	0.3946	2.0737	2.4518
0.9758	0.9628	1.6596	2.4516	0.2473	0.2882	1.539	2.9012
0.9747	0.9631	1.8337	2.6995	0.2438	0.3454	2.2588	3.6632
0.984	0.9615	1.1838	2.7629	0.2207	0.3583	2.0361	2.3465
0.9845	0.9625	1.2308	2.3854	0.2221	0.2488	1.6088	2.324
0.9819	0.9693	1.4975	2.7491	0.2898	0.2541	1.5319	2.4548
0.9837	0.9673	1.1659	2.4426	0.2152	0.2701	1.2912	2.997
0.983	0.9609	1.1743	2.6636	0.2101	0.321	1.1602	2.4829
0.9823	0.9709	1.3659	2.4417	0.2714	0.3113	1.4004	2.0409
0.9827	0.9622	1.2817	2.687	0.1985	0.2504	1.5968	2.9105
0.9695	0.9615	1.8211	2.8936	0.3014	0.3766	2.6958	3.9803
0.9844	0.9666	1.2739	2.5193	0.2436	0.3199	1.6184	2.5612
0.9845	0.9692	1.2528	2.5369	0.2476	0.3081	1.5076	2.3913
0.9844	0.9657	1.8664	2.6567	0.1903	0.2248	1.9409	2.9759
0.9761	0.9623	1.8152	2.6675	0.262	0.3217	2.3612	3.2545
0.9858	0.9646	1.3397	2.5086	0.2467	0.2525	1.7528	2.5038
0.9809	0.9643	1.4724	2.7404	0.2499	0.2795	1.7894	2.9891
0.9823	0.9678	1.2281	2.7297	0.2243	0.2687	1.8376	2.1728
0.9785	0.9658	1.3923	2.8518	0.2467	0.2761	1.8729	2.8445
0.9826	0.9695	1.2941	2.6399	0.2382	0.325	1.966	2.7535
0.9776	0.969	1.3325	2.3117	0.2294	0.3528	2.0811	3.8312
0.9791	0.9635	1.1472	2.0563	0.2026	0.2592	1.3498	2.61
0.987	0.9592	1.1198	2.6928	0.2156	0.3321	1.7931	3.7153
0.9795	0.9521	1.8158	2.6255	0.25	0.3769	3.2065	4.4892
0.9723	0.9472	1.9978	2.8758	0.2418	0.3227	3.6194	3.822
0.9711	0.9605	2.3591	2.5796	0.2291	0.3574	3.5443	4.197
0.9814	0.9598	1.2983	2.4789	0.2413	0.3884	2.1826	3.736
0.9731	0.9585	2.3162	2.9	0.2093	0.2999	3.1678	4.4181
0.9751	0.9601	2.134	2.7343	0.2617	0.4162	3.0275	5.0911
0.9706	0.9638	1.9994	2.7816	0.2727	0.3457	3.0365	4.9766
0.974	0.9614	1.9515	2.8076	0.2677	0.3917	3.5692	4.5186
0.973	0.9566	2.118	2.8908	0.2979	0.3991	3.4184	4.2814
0.9766	0.9584	1.825	2.5481	0.3288	0.3707	3.3257	4.3362
0.9748	0.966	1.995	2.7952	0.2842	0.3546	3.5953	4.3257
0.975	0.9569	2.0982	2.8364	0.2936	0.3144	3.3842	4.0463
0.9709	0.9634	2.0977	2.5619	0.2852	0.3082	3.43	3.7023
0.9815	0.9639	1.6292	2.5962	0.2795	0.3145	2.3586	4.2116
0.9761	0.9666	2.0269	2.9081	0.2747	0.336	3.0208	4.2804
0.9699	0.9655	2.2658	2.8876	0.3003	0.35	3.6914	4.1619
0.976	0.9641	1.9568	3.0829	0.253	0.393	3.2725	4.0808
0.9716	0.9623	2.2723	2.7695	0.2112	0.396	3.8544	4.4018
0.9734	0.9575	2.0116	2.542	0.2849	0.4054	3.3971	3.448
0.9741	0.9597	1.8673	2.9311	0.2151	0.3529	3.2655	4.7314
0.9736	0.9636	1.9959	2.7529	0.2693	0.3545	3.0959	4.3545
0.9751	0.9565	2.0472	2.5417	0.2834	0.2542	3.2285	3.8546
0.9863	0.9619	1.1389	2.7464	0.2158	0.3901	1.6811	4.4786
0.9728	0.9529	1.8693	2.5948	0.2459	0.3165	3.0038	4.3767
0.9803	0.9567	1.5757	2.8028	0.2988	0.4396	2.7547	4.3225

0.9739	0.9617	1.9276	2.4321	0.2236	0.3598	3.1714	4.0037
0.9769	0.9553	1.803	2.6269	0.2066	0.2916	3.2129	4.416
0.9739	0.9567	2.1842	2.5269	0.2896	0.3445	3.5158	3.7313
0.9712	0.9636	2.1839	2.7779	0.2024	0.3181	3.4122	4.0188
0.9786	0.9635	1.7804	2.8643	0.2521	0.3242	2.9318	4.7692
0.9761	0.9573	1.8827	2.786	0.2319	0.2611	3.3862	4.4987
0.9763	0.9651	1.7724	2.7303	0.2952	0.3659	2.9138	4.0633
0.9752	0.9639	1.9275	2.9188	0.2507	0.2086	3.5568	4.5692
0.9748	0.9632	1.9895	2.696	0.2773	0.3359	2.997	3.8722
0.9758	0.9614	1.9564	2.7258	0.2378	0.3454	2.944	4.1932
0.974	0.9637	2.0156	2.7797	0.2821	0.3922	3.087	4.0052
0.9664	0.9617	2.6242	2.4953	0.2733	0.2447	3.8839	3.6849
0.9705	0.9577	2.0104	2.5479	0.278	0.4168	3.0765	4.5349
0.9729	0.9565	1.4845	2.7384	0.281	0.4046	2.3783	3.7116
0.9733	0.9526	2.1643	2.8014	0.2982	0.4352	3.5918	4.0439
0.9863	0.9609	1.9464	2.4546	0.3096	0.2932	3.2987	3.9619
0.9704	0.9506	1.7974	2.7018	0.2354	0.3161	3.1184	3.8707
0.9799	0.9614	2.1686	3.079	0.2005	0.3101	3.444	4.9152
0.9761	0.963	2.0117	2.6509	0.2817	0.3215	3.4953	4.6039
0.9796	0.9605	2.3599	2.7359	0.2103	0.2101	3.5545	4.4675
0.9803	0.9629	2.2335	2.6158	0.2938	0.413	3.6486	3.9593
0.9789	0.9519	2.1769	2.5879	0.2147	0.3704	3.0478	4.2046
0.9794	0.9446	2.3085	2.4883	0.234	0.2728	3.6977	4.3199

6. Conclusion

This paper proposed an approach for floor segmentation using FCM and CNN. In FCM segmentation, the given input image was partitioned into homogeneous regions based on similarity for merging. In CNN stage, the interactive information was introduced as markers of the object area and background area, which were input by the users to roughly indicate the position and main features of the object and background. The segmentation performance was measured using probabilistic rand index, variation of information, global consistency error, and boundary displacement error. Experiments were conducted on a relatively large database of floors. It was observed that the proposed floor segmentation achieved relatively good segmentation results, compared with other existing well-known methods. Further, we worked on large datasets for mobile navigation of robots.

Data Availability

The data used to support the research findings are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Z. C. Chen and S. T. Birchfield, "Visual detection of lintel-occluded doors from a single image," In 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Anchorage, AK, USA June 23-28, 2008, pp. 1-8, 2008. https://doi.org/10.1109/CVPRW.2008.4563142.

[2] D. C. Lee, M. Hebert, and T. Kanade, "Geometric reasoning for single image structure recovery," In 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, June 20-25, 2009, pp. 2136-2143, 2009. https://doi.org/10.1109/CVPR.2009.5206872.

[3] R. G. Golledge and G. Zannaras, "Cognitive approaches to the analysis of human spatial behaviour," In *Environment and Cognition*, W. Ittelson (Ed.), New York, USA: Seminar Press, pp. 59-94, 1973.

[4] M. J. Farah, *The Handbook of Neuropsychology, Disorders of Visual Behavior,* Amsterdam, Netherlands: Elsevier, pp. 395-413, 1989.

[5] M. Riddoch and G. Humphreys, *Neuropsychology of Visual Perception*, Hillsdale: Lawrence Erlbaum Associates, pp. 79-103, 1989.

[6] D. Conrad and G. DeSouza, "Homography-based ground plane detection for mobile robot navigation using a modified EM algorithm," In 2010 IEEE International Conference on Robotics and Automation, Anchorage, AK, USA, May 03-07, 2010, IEEE, pp. 910-915. https://doi.org/10.1109/ROBOT.2010.5509457.

[7] G. Panahandeh, N. Mohammadiha, and M. Jansson, "Ground plane feature detection in mobile vision-aided inertial navigation," In 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems,

Vilamoura-Algarve, Portugal, October 07-12, 2012, IEEE, pp. 3607-3611. https://doi.org/10.1109/IROS.2012.6385503.

[8] M. Adachi, S. Shatari, and R. Miyamoto, "Visual navigation using a webcam based on semantic segmentation for indoor robots," In 2019 15th International Conference on Signal-Image Technology & Internet-Based Systems, (SITIS), Sorrento, Italy, November 26-29, 2019, IEEE, pp. 15-21. https://doi.org/10.1109/SITIS.2019.00015.

[9] G. C. Barcelʹo, G. Panahandeh, and M. Jansson, "Image-based floor segmentation in visual inertial navigation," In 2013 IEEE International Instrumentation and Measurement Technology Conference, (I2MTC 2013), Minneapolis, MN, USA, May 06-09, 2013, IEEE, pp. 1402-1407. https://doi.org/10.1109/I2MTC.2013.6555644.

[10] F. Geovani Rodrʹıguez-Telles, L. Abril Torres-Mʹendez, and E. A. Martʹınez-Garcʹıa, "A fast floor segmentation algorithm for visual-based robot navigation," In 2013 International Conference on Computer and Robot Vision, Regina, SK, Canada, May 28-31, 2013, IEEE, pp. 167-173. https://doi.org/10.1109/CRV.2013.40.

[11] L. Ma, J. M. Wang, B. Zhang, and S. B. Wang, "Automatic floor segmentation for indoor robot navigation," In 2010 2nd International Conference on Signal Processing Systems, Dalian, China, July 05-07, 2010, IEEE, pp. 684-689. https://doi.org/10.1109/ICSPS.2010.5555399.

[12] S. Wang, X. X. Zuo, W. W. Yu, R. X. Wang, and K. Madani, "Towards robotic semantic segmentation of supporting surfaces," In 2015 IEEE International Conference on Computational Intelligence & communication technology, Ghaziabad, India, February 13-14, 2015, IEEE, pp. 775-779. https://doi.org/10.1109/CICT.2015.89.

[13] J. A. de Jesús Osuna-Coutiño and J. Martinez-Carranza, "Binary-patterns based floor recognition suitable for urban scenes," In 6th International Conference on Control, Decision and Information Technologies, April 23-26, 2019, IEEE, pp. 1574-1579. https://doi.org/10.1109/CoDIT.2019.8820296.

[14] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution and Fully Connected CRFs," IEEE Trans. Comp., vol. 2016, 2016. https://doi.org/10.48550/arXiv.1606.00915.

[15] A. Skoryk, Y. Chyrka, I. Gorovyi, O. Grechnyev, and P. Vyplavin, "Comparative analysis of classic computer vision methods and deep convolutional neural networks for floor segmentation," In 2020 IEEE Third International Conference on Data Stream Mining & Processing, (DSMP 2020), Lviv, Ukraine, August 21-25, 2020, IEEE, pp. 217-221. https://doi.org/10.1109/DSMP47368.2020.9204339.

[16] Y. Kida, S. Kagami, T. Nakata, M. Kouchi, and Hiroshi Mizoguch, "Human finding and body property estimation by using floor segmentation and 3D labelling," In 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583), Hague, Netherlands, October 10-13, 2004, IEEE, pp. 2924-2929. https://doi.org/10.1109/ICSMC.2004.1400777.

[17] Y. C. Du and T. Arslan, "A segmentation-based matching algorithm for magnetic field indoor positioning," In 2017 International Conference on Localization and GNSS, Nottingham, UK, June 27-29, 2017, IEEE, pp. 1-5. https://doi.org/10.1109/ICL-GNSS.2017.8376237.

[18] R. Bormann, F. Jordan, W. Z. Li, J. Hampp, and M. Hägele, "Room segmentation: Survey, implementation, and analysis," In 2016 IEEE International Conference on Robotics and Automation, (ICRA), Stockholm, Sweden, May 16-21, 2016, IEEE, pp. 1019-1026. https://doi.org/10.1109/ICRA.2016.7487234.

[19] T. Honto, Y. Sugaya, T. Miyazaki, and S. Omachi, "Analysis of floor map image in information board for indoor navigation," In 2017 International Conference on Indoor Positioning and Indoor Navigation, (IPIN 2017), Sapporo, Japan, September 18-21, 2017, IEEE, pp. 1-7. https://doi.org/10.1109/IPIN.2017.8115896.

[20] R. Ambruş, S. Claici, and A. Wendt, "Automatic room segmentation from unstructured 3D data of indoor environments," IEEE Robot. Autom. Lett., vol. 2, no. 2, pp. 749-756, 2016. https://doi.org/10.1109/LRA.2017.2651939.

[21] Z. D. Lu, W. G Teng, J. W. Guo, W. L. Meng, J. Xiao, W. Zhang, and X. P. Zhang, "Data-driven floor plan understanding in rural residential buildings via deep recognition," Inf. Sci., vol. 567, pp. 58-74, 2021. https://doi.org/10.1016/j.ins.2021.03.032.

[22] D. Fleer, "Human-Like Room Segmentation for Domestic Cleaning Robots," Robotics, vol. 6, no. 4, 2017. https://doi.org/10.3390/robotics6040035.

[23] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating algorithms and measuring ecological statistics," In Proceedings Eighth IEEE International Conference on Computer Vision, (ICCV 2001), Vancouver, BC, Canada, July 07-14, 2001, IEEE, pp. 416-423. https://doi.org/10.1109/ICCV.2001.937655.

[24] R. Unnikrishanan and M. Hebert, "Measures of similarity," In 2005 Seventh IEEE Workshops on Applications of Computer Vision, (WACV), Breckenridge, CO, USA, January 05-07, 2005, IEEE, pp. 394-394. https://doi.org/10.1109/ACVMOT.2005.71.

[25] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "A measure of objective evaluation of image segmentation

algorithms," In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Diego, CA, USA, September 21-23, 2005, IEEE, pp. 34-34. https://doi.org/10.1109/CVPR.2005.390.

[26] "MIT places database for scence recognition," Places Database, 2015, http://places.csail.mit.edu/.

[27] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 39, no. 12, pp. 2481-2495, 2017. https://doi.org/10.1109/TPAMI.2016.2644615.

[28] J. D. Chen, Y. C. Wu, Y. Yang, S. P. Wen, K. B. Shi, A. Bermak, and T. W. Huang, "An efficient memristor-based circuit implementation of squeeze-and-excitation fully convolutional neural networks," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 33, no. 4, pp. 1779-1790, 2021. https://doi.org/10.1109/TNNLS.2020.3044047.