International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

SceneGuide: An Indoor and Outdoor Scene Recognition Wearable Aid for Visually Impaired People

Jyoti Madake[1], Shripad Bhatlawande*[2], Anjali Solanke[3]

Submitted: 08/05/2023 **Revised**: 17/07/2023 **Accepted**: 06/08/2023

Abstract: This paper proposes a new SceneGuide wearable aid for providing information about the surrounding scene to the visually impaired people. Its main feature is its ability to understand the scene and offer simplified information in an intuitive way. SceneGuide aid is designed as a wearable jacket with low-power embedded processing unit, monocular camera, and Bluetooth headphones. It is a lightweight, low-cost, battery-operated blind assistive aid. The aid employs a novel, computationally efficient model, using multi-feature fusion and multi-level optimum feature selection approach. SceneGuide serves as a complementary assistive aid to the conventional white cane and helps reduce the cognitive information load and anxiety experienced by visually impaired people. The functional evaluation of the aid presented scene recognition accuracy of 95.25% on a custom dataset and 85.82% on the 15 Scene Standard Dataset. This aid was evaluated with 10 blindfolded volunteers. The volunteers expressed 77% acceptance towards usability to identify the scene with lower levels of confusion and anxiety. This highlights that the SceneGuide aid can enhance the understanding of visually impaired people about their surroundings.

Keywords: *Computer vision, machine learning, scene recognition, visually impaired, wearable aid*

1. Introduction

World health organization (WHO) reported approximately 2.2 billion people are affected with near or distance visual impairment [1]. There are a variety of circumstances that make it difficult for these people to move safely or recognize their surroundings. This excursion also presents numerous difficulties, including inadvertent falls, a feeling of being lost, and the incapacity to perform activities independently. Traditionally, long white canes are used to aid vision-impaired individuals. The white cane has been a widely used mobility aid due to its simplicity and portability. The blind user needs to heavily relay on can tapping for obstacle detection along the path, this increases their walking anxiety and cognitive load [2], also it has limitation in perceiving any visual information. Over the past 75 years, extensive research has been conducted in the field of assistive aids for individuals who are blind or visually impaired. Various electronic mobility aids, such as Venucane [3], MobiFree [4], SplitGrip Cane [5], and Tom Pounce III [6], have been developed to facilitate obstacle detection and avoidance. Wearable aids, including Ultrasonic spectacles and waist-belt [7], Array of Lidars and Vibrotactile Belt [8], Haptic Sensory Glove [9], Optical See-Through Glasses [10], and Mobility Shoes [11] have also been introduced to enhance the mobility and perception of visually impaired individuals. These aid with simultaneous localization and mapping, as well as recognition of traffic lights and crosswalks. Additionally, cloud-based smartphone aids like the Divya Dristi App [12], Uasis Aid [13], Tap-Tap-See App [14], and Be-My-Eyes App [15] have been developed to help with various tasks related to navigation and object recognition.

These aids have limited or no capability in providing detailed visual information about the surroundings. The smartphone-based aids [12, 13, 14, 15] heavily rely on cloud technologies, with stable internet connectivity, and are not suitable for areas with limited connectivity, creating dependency challenges. The usability of these assistive aids presents challenges due to slow learning curves and complex user interfaces. The complex operating processes further hinder the adoption and effectiveness of these aids.

Visual perception assistance has emerged as a transformative technology that offers significant benefits to visually impaired individuals by enhancing their understanding of the surrounding environment. It promotes environmental awareness by providing them with crucial information about objects, structures, and people in their surroundings. The use of computer vision and machine learning algorithms has facilitated the development of systems capable of recognizing various types of scenes and objects within them. Previous research has focused on scene recognition using computer vision and machine learning approaches [16, 21, 22, 30], as well as deep learning techniques [33 – 44]. However, scene recognition continues to pose challenges, primarily due to variations in internal scene details, changes in illumination, and occlusions.

In this paper, we propose a novel approach that leverages the fusion of local and global features to achieve more

[1,2] Vishwakarma Institute of Technology, Pune, Maharashtra – 411037, India ORCID ID : [1]0000-0002-5302-2508, [2]0000-0001-8405-9824
[3] Marathwada Mitra Mandal's College of Engineering, Pune, 411052, India, ORCID ID : 0000-0001-9602-3263
** Corresponding Author Email: shripad.bhatlawande@vit.edu*

meaningful scene perception for visually impaired individuals. By combining fine-grained local details with holistic global information, our approach aims to provide an understanding of scenes, enabling visually impaired individuals to perceive and navigate their environment more effectively. The proposed method uses a multi-feature fusion of ULBP (Uniform Local Binary Pattern), LBPHF (Local Binary Pattern Histogram Fourier), and SIFT (Scale Invariant Feature Transform) features to capture scene discriminative information with lower computational cost. It also employs a multi-level feature dimension optimization algorithm using PCA (Principal Component Analysis) and LDA (Local Discriminant Analysis) to obtain an optimal feature subset with high inter-class separation. The paper describes the implementation of a wearable SceneGuide assistive aid to help visually impaired individuals in real-time scene recognition. Functional evaluation on custom and 15 scene datasets demonstrates the superiority of the proposed method over other state-of-the-art methods. Real-time evaluation with 10 sighted volunteers confirms the system's suitability as a lightweight, portable assistive aid for scene recognition.

The paper is organized as follows: Section 1 provides a summary of the need assessment survey conducted for visually impaired people. Section 2 presents a review of related works. Section 4 introduces the proposed method of multi-feature fusion and multi-level feature optimization. Section 5 presents the experimental results. Finally, Section 6 concludes the paper.

1.1. Need Assessment Study with Visually Impaired People

This section of the paper details the findings of a survey conducted on a group of people living in the city area of the western region of India. The survey has been conducted on 60 blind or visually impaired people. The initial section of the survey asked participants to identify their age group, gender, and family status. This information was used to form groups based on their age groups to perform age-specific mobility assistive aid requirement assessment. The cities were not designed to cater to the mobility needs of blind people. The blind and visually impaired people face many challenges while traveling short distances on their own. The challenges faced by them vary as per the age group and type of activity they wish to carry. The objective of the survey is to understand the different types of hurdles to independent mobility experienced by blind and visually impaired people in indoor and outdoor situations, the shortcomings of the existing assistive aids., and the ergonomic considerations for the assistive aid.

The detailed questionnaire with 10 questions was prepared with the following questions: i) Challenges faced in identifying the indoor scene, ii) finding the correct bus stop, iii) identifying the auto rickshaw, iv) Shopping requirement for daily items, v) places visited frequently, vi) Any accidental falls, vii) Challenges faced while traveling, viii) challenges faced while shopping, ix) assistive aid design considerations, x) assistive aid cost requirement.

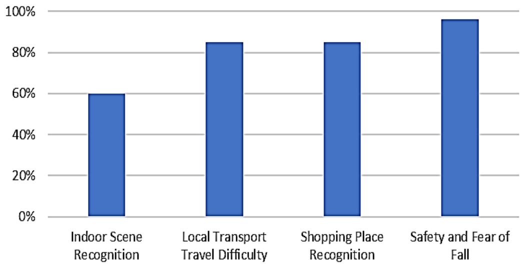


Fig.1. Type of Problems Reported by Blind and Visually Impaired People

The responses received from the blind and visually impaired people were analyzed and we found that there had been many challenges faced by these people during outdoor travel even for a short distance. 90% of the participants reported that they memorize the route. Also, the environmental noises and smell help them to recognize the place. 60% of them reported difficulty in the recognition of indoor scenes in known and unknown houses. 85% of respondents had difficulty traveling using local transport facilities such as autorickshaws or public buses. 75% of participants reported difficulty in identifying the shops and marketplaces for day-to-day shopping for groceries, vegetables, medicines, etc. Figure 1 details the distribution of the largest reported requirements by blind and visually impaired people.

1.2. Need Assessment Driven Findings

The last two questions in the survey were about the requirements from the aesthetics and functions needed in an assistive aid. 80% of them reported it should be wearable aid, so their hands can be free to hold guide-cane and other things. It should be battery operated, lightweight. The blind participants suggested the system should give them real-time responses either in voice or vibrations. Some suggested aid with a video calling facility to a known person. The figure 2 shows the requirements of the features in assistive aid vs the percentage respondents.

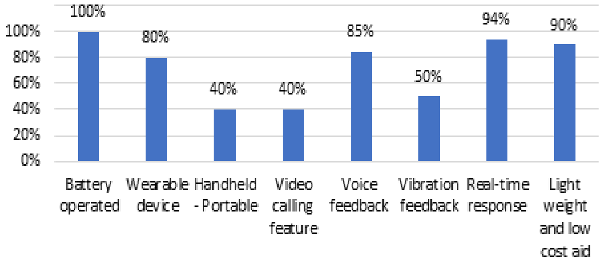


Fig. 2. The Assistive Aid Features vs Percentage of the Respondents

The findings of the survey helped to assess the feature requirements and the type of system to be designed for helping blind people. The data gathered from the survey helped to understand the budget limitation for assistive aid since almost 95% of them needed lightweight and low-cost aid. This paper proposes an assistive aid to address these difficulties reported by blind and visually impaired people and focuses on real-time, lightweight, wearable blind assistive aid.

2. Prior Art

The domain of scene recognition research encompasses the creation of computer vision algorithms and machine learning models that can automatically identify scenes. This task is particularly challenging, as scenes can vary greatly in complexity and diversity, and often display significant variability in visual appearance.

Scene recognition is an essential component of computer vision, which involves identifying the objects and natural surroundings depicted in a scene image. An effective scene recognition algorithm must account for the interrelationships between the various semantic partitions of the image. There exist a variety of scene recognition strategies, with early approaches such as the influential work of [16] considering scenes as a collection of objects with distinct shapes and structures. However, recognizing indoor images poses significant challenges due to the complexity of indoor environments, resulting in poor performance for this type of approach compared to the satisfactory results obtained for outdoor scenes.

The GIST [17] characteristics provide a statistical overview of the spatial arrangement of the scene, capturing its key perceptual features, such as naturalness, openness, roughness, expansion, and ruggedness. Oliva and Torrlba [17] suggested that images belonging to the same scene category have comparable spatial configurations that can be extracted without dividing the image into segments. These features are computed with a combination of multi-scale-oriented Gabor kernels. These were found to be more suitable for outdoor scenes than indoors. Hierarchical GIST [18] utilized the perceptual GIST layer and Kernel PCA layer.

Both the local and global descriptors can be utilized for scene description. The SIFT [19] algorithm extracts image features that remain unchanged despite variations in image scale, rotation, and illumination. A method to employ SIFT for scene categorization [20] involved extracting local features from dense patches, creating a dictionary through k-means algorithm from random local patches, and generating a feature vector for each input image. SIFT-based spatial pyramid matching technique (SPM) [21] and Sparse coding based spatial pyramid matching (ScSPM) [22] was used to represent local features. In contrast to the original SPM method that employs histograms, the ScSPM approach utilizes the max operator, which is more resistant to local spatial translations, resulting in improved robustness. The efficiency of the SIFT-FV method [23] surpasses previous techniques as it computes SIFT features on dense grids and creates global features using Fisher kernel coding.

LBP algorithm [24] for texture classification, was first used for scene categorization on SUN dataset [25]. LBP-HF [26], which combines uniform LBP and Fourier coefficients, was proposed and shown to have better rotation invariance than uniform LBP. Completed local binary pattern (CLBP) [27] for texture classification, which encodes both the signs and magnitudes of differences between center pixel and its neighbors, as well as the intensity of the center pixel. CLBP achieved superior texture classification accuracy com- pared to the original LBP algorithm. Pyramid Local Binary Pattern (PLBP) [28], utilizes a hierarchical spatial pyramid to extract texture resolution information by concatenating LBP features, more efficient than LBP. CENTRIST [29], a holistic visual descriptor representation of images' geometrical and structural properties. It employs the census transform (CT) to calculate feature maps, which is equivalent to LBP.

HOG [30] features are adept at capturing the distribution of edge directions and image gradients on a regular grid, making them useful for scene categorization. In a study on the SUN397 dataset [31], HOG features outperformed SIFT and GIST. The combination of SIFT and PCA in the PCA-SIFT approach [32] involves the computation of a projection matrix P using numerous image patches. This technique has been shown to yield superior results in feature selection and image matching compared to the use of SIFT features alone [33].

In recent years, the deep convolutional neural network (CNN) [36] first introduced in 1998, has gained popularity and achieved astounding performance on a variety of visual tasks, in addition to the traditional scene recognition approaches that employ handcrafted visual features, as mentioned above. The Multi-scale orderless pooling CNN (MOP-CNN) [37] is a method that involves using convolutional neural networks (CNNs) to extract activations from local patches of input images at different scales. These activations are then aggregated into vector descriptors using a locally aggregated descriptor (VLAD) technique, which is orderless pooling.

The DAG-CNN (Directed Acyclic Graph Convolutional Neural Network) [38], as a deep learning architecture, has achieved notable success by integrating local features extracted from lower layers and holistic features from top layers. This combination allows the model to learn and leverage both low-level and high-level features to improve its performance on image-related tasks. Several CNN architectures Inception [39], ResNet [40], GoogleNet [41] and DenseNet [42], have also demonstrated similar benefits. These architectures have shown that features extracted from intermediate and high layers, which correspond to parts and objects in images, are more informative for scene recognition tasks compared to low-level features such as edges and textures. This suggests that leveraging high-level features is crucial for achieving superior performance on image recognition tasks.

Recent deep learning-based scene recognition techniques propose use of discriminative information of the pre-trained deep learning architecture with metric learning [42, 43]. The fusion of deep learning features with extreme learning machine observed better scene recognition performance. Improved scene recognition proposed by [44] involves dense hog feature extraction, autoencoders and spatial pyramid pooling. Instead of using histogram of local visual descriptors over each region in image block, the modified spatial pyramid pooling with local normalization is used to produce representations of the various regions.

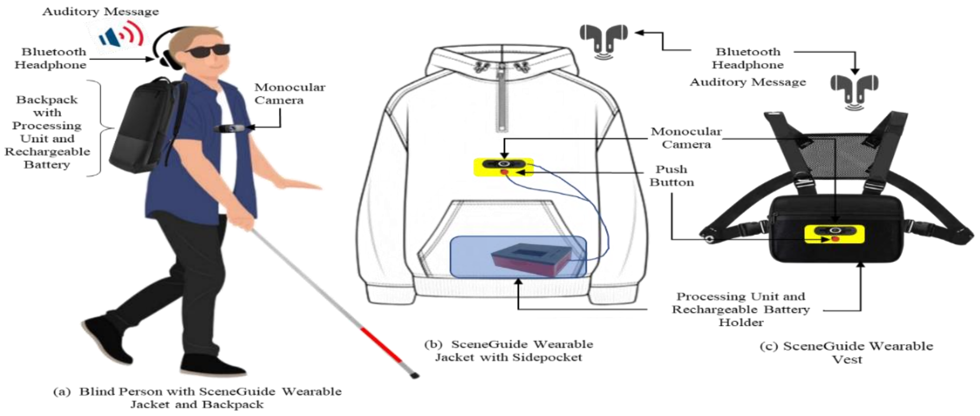
Recent advancements in deep learning and computer vision have led to significant progress in scene recognition research and have enabled the development of models that can recognize scenes with high accuracy and robustness. However, there are still many challenges and open problems in this field, such as dealing with large variability in scene appearance due to lighting, weather, and other factors, or recognizing scenes with little visual information or contextual cues. The large computational complexity, heavy models are not suitable for portable embedded development platform. This paper proposes computer vision and machine learning approach for scene recognition.

3. Design and Development of Sceneguide Wearable Aid

This research presents a real-time approach to recognizing indoor and outdoor scenes, namely indoor kitchen and living room scenes. The outdoor scenes include recognition of public bus station, autorickshaw stand, marketplace and shopping place. The responses to a survey of blind people with visual impairments were used to inform the selection of scene types. Fig 3 is a system-level diagram of the SceneGuide wearable aid. The blind or visually impaired people assistive wearable aid SceneGuide is implemented utilizing a low-power embedded microcontroller, the Jetson Nano board consists of a Quad-core ARM A57 CPU clocked at 1.43 GHz, a 4GB RAM, and a 128-core GPU, USB connected monocular camera, Bluetooth headphones. The figure 1 details the camera is mounted on the wear- able jacket worn by the blind user who is equipped with backpack with Jetson Nano processing unit and 22.5W rechargeable portable battery pack. The fig. 4 (a) and (b) details the developed wearable SceneGuide aid with the blind-folded volunteer. First, the camera's collected images are resized to 256X256 pixels and Gaussian blurred to eliminate high-frequency noise. The second step is to extract multiple features using SIFT, Uniform Local Binary Pattern, and Local Binary Pattern Histogram Fourier. In the third stage, clustering and feature fusion are carried out to generate an optimal feature subset. The high dimensionality of the feature space is optimized using PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis) is used to increase distinction between the classes. The last and final stage consists of indoor or outdoor scene recognition employing Random Forest, KNN, LGBM, and XGBoost classifiers, and scene recognition is successfully performed. The output of the classifier is then given to the text-to-speech conversion module. The Bluetooth-connected headphones are used to send an audible message alert regarding the scene category to the blind individual.

3.1. Dataset Collection and Pre-Processing

The two different datasets were used to evaluate the performance of the model. The custom dataset has a total of 4800 images of 6 scene categories. The images in the custom dataset are collected and labeled by the authors. The standard dataset of 15-Scene is also



(a) Blind Person with SceneGuide Wearable Jacket and Backpack

(b) SceneGuide Wearable Jacket with Sidepocket

(c) SceneGuide Wearable Vest

Fig. 3. System Level Diagram of SceneGuide Wearable Aid for Visually Impaired People

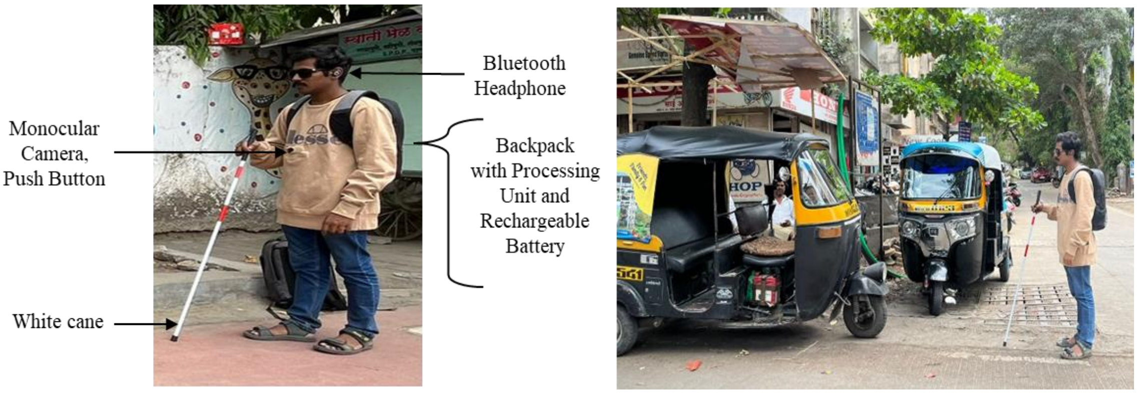

Bluetooth Headphone
Monocular Camera, Push Button
Backpack with Processing Unit and Rechargeable Battery
White cane

Fig. 4. (a) Blind-folded volunteer with SceneGuide Aid, (b) Blind-folded volunteer at Outdoor Rickshaw Stand

used. The dataset consists of 6 scene classes i) Indoor Kitchen, ii) Living Room, iii) Shopping Place, iv) Market Place, v) Bus stop, and vi) Rickshaw. Figure 5 shows examples of the images in the dataset. The diverse image database can easily accommodate additional image types in the training set without any changes to the existing algorithms. All the images have been resized to 256 x 256 dimensions. The images were augmented with rotation of 10 deg, 15% zoomed and translated horizontally and vertically by 20%. Also, the 15% shear and horizontal flip was introduced to build a model with translation, rotation, and view angle change invariant. The scene of interest at varied scales and positions, against an illumination change, improves model generalization, addresses class imbalance, and increases robustness to variations in the input data. All the images were converted to a grayscale. Then smoothing is applied using gaussian filter to remove extraneous noise.

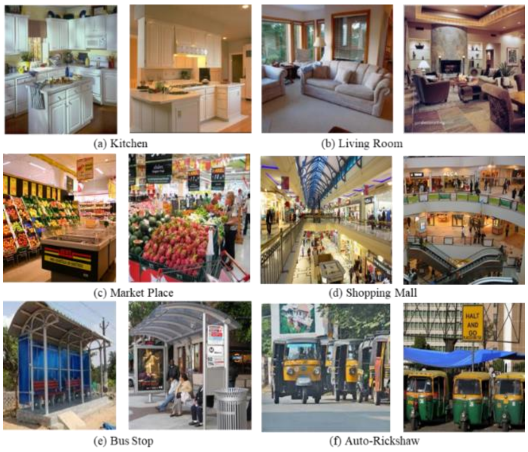

(a) Kitchen
(b) Living Room
(c) Market Place
(d) Shopping Mall
(e) Bus Stop
(f) Auto-Rickshaw

Fig. 5. Sample Images from the dataset with (a) Kitchen, (b) Living Room, (c) Market Place, (d) Shopping Mall, (e) Bus Stop, (f) Auto-Rickshaw

reduce high-frequency noise and preserving edges in the scene. The Gaussian filter produces the following estimate of the original scene, with kernel standard deviation sigma along horizontal and vertical direction, as shown in eq 1:

$$F(i,j) = \frac{1}{2\Pi\sigma^2} e^{-\left(\frac{i^2+j^2}{2\sigma^2}\right)} \qquad (1)$$

Here i and j are the pixels in the image. This helps in reducing the high-frequency components, resulting in features which are more robust to noise and other distortions.

In order to accurately capture the specific geometrical details, present in different scene categories such as kitchen, living room, shopping place, bus station, and rickshaw, it is important to compute the edge gradients and create image descriptors based on this information. This approach allows for a more precise description of the distinct structural details that are present in each category.

The Prewitt operator [35], developed by M. S. Prewitt is used to estimate the magnitude and orientation of the structural details present in the scene. The Prewitt operator with horizontal and vertical kernel is used to generate edge gradient images using eq. 2:

$$|\nabla F| = \sqrt{\nabla f_i^2 + \nabla f_j^2} \qquad (2)$$

This gradient image is used for computing SIFT features for the scene.

3.2 Multi-Feature Extraction and Feature-Fusion

The research introduces a novel scene recognition system that integrates three distinct feature extraction techniques: Scale Invariant Feature Transform (SIFT) [19], Uniform Local Binary Pattern (ULBP) features [24], and Local Binary Pattern Histogram Fourier (LBPHF) features [26]. SIFT is employed to extract scale-invariant and viewpoint-robust features from images, accurately capturing geometrical and structural variations present in each of the six scene categories. ULBP is utilized to capture fine-grained local image features, enhancing the system's ability to distinguish subtle texture patterns within the scenes. Additionally, LBPHF is used to capture both spatial and frequency information, providing a comprehensive representation of the scene's characteristics.

The scenes, including the kitchen, living room, shopping mall, marketplace, bus stop, and autorickshaw, exhibit unique features such as object arrangements, structural uniqueness, and lighting conditions. By combining these three feature extraction techniques, the scene recognition algorithm achieves a comprehensive representation, effectively recognizing the distinctive visual characteristics in the dataset's six different scenes. This feature fusion approach significantly improves scene recognition accuracy, outperforming existing algorithms and demonstrating its effectiveness in real-world scene recognition problems.

The unique features are generated by integrating image-based dense features and applying multi-level feature dimension reduction to obtain an optimized and highly discriminative feature subset for classification. This approach outperforms existing algorithms for scene recognition, demonstrating its effectiveness in solving real-world scene recognition problems.

Algorithm 1: Gradient image based key point detection and description

```
Input: Image dataset with N images, $I(i, j)$

Output: Feature vector $h(wf_{SIFTN}^{K})$

Steps:
1. for N images in the dataset:
2.     Resize images to 256 x 256
3.     Convolve Image, $I(i, j)$ with 3 x 3 Gaussian kernel:
           $F(i, j) = G * I(i, j)$
           Gaussian filter kernel equation:
           $F(i,j) = \frac{1}{2\Pi\sigma^2} e^{-\left(\frac{i^2+j^2}{2\sigma^2}\right)}$
4.     Find intensity gradient of filtered image $F(i, j)$
           Magnitude of image gradient, $|\nabla F| = \sqrt{\nabla f_i^2 + \nabla f_j^2}$
           Direction of image gradient, $\theta = t_{an}^{-1}\left(\frac{\nabla f_j}{\nabla f_i}\right)$
           Compute Edge Gradient image, $E(i, j)$
   end for
5. for each pixel in $E(i, j)$ compute
       SIFT Key points, $N_{SIFT}$
       Key point Descriptors, $DN_{SIFT} = N_{SIFT}$ x 128 d
6. Find minimum number of clusters needed for $N_{SIFT}$ key points
   for k = kmin → kmax do
       k, $N_{SIFT}$ → k-means++ and train
       for number of clusters Nc = k do
           Calculate distance of each NSIFT point from cluster center $N_{c1}, \dots N_{ck}$
           Compute sum of square distances for each cluster
           Find WCSS for all clusters
           Minimize $WCSS_k$ in range $[WCSS_{kmin}, \dots WCSS_{kmax}]$
           Choose optimum number of clusters $N_C$ at elbow point
               {the point where the WCSS starts to decrease more slowly}
       end for
   end for
7. Normalize the clustered feature descriptors, $f_{SIFTKN}$
8. Compute the histogram of features with bins = Nc
9. return $h(wf_{SIFTN}^{K})$
```

SIFT [4] is used to extract distinctive invariant key points, N_{SIFT} from images. SIFT detects all the key points which are invariant to scale and orientation. Each scene image has different number of key points. In this study, we have implemented a BOF framework to analyze structural elements and recognize scene. The descriptors extracted using SIFT, N_{SIFT} were grouped into a specified number of clusters to minimize the WCSS (within cluster sum of square differences) for the K clusters. The clustered descriptors were normalized N_{SIFTKN}. The Algorithm 1 details the SIFT feature points detection, description and choosing optimum number of clusters to generate final features f_{SIFTK}. These features are normalized to using min-max scaling to generate f_{SIFTKN}. The normalization improves the performance of the classification algorithm and helps to converge faster and more reliably. The optimization algorithm works better with the features are on similar scales, as it helps to avoid oscillations and divergences. Finally, the histogram $h(wf_{SIFTN}^{K})$ of the normalized features f_{SIFTKN} computed.

The SIFT feature extraction algorithm can struggle to accurately capture the characteristics of an object in an image when the background is complex or contains noise. In contrast, LBP features are effective at filtering out such noise when the image contains uniform patterns. Therefore, combining SIFT and LBP features can potentially yield better results for scene recognition tasks.

Local Binary Patterns (LBPs) are a type of feature descriptor that is computed from pixel intensities in a local neighborhood. This algorithm is simple yet effective for extracting local texture information and is also robust to changes in lighting and rotation. To compute the LBP value for a pixel, a 3X3 size circular neighborhood of pixels is defined around the central pixel. The LBP operator then compares the intensity values of the pixels in this neighborhood to the intensity value of the central pixel and assigns a binary label of 0 or 1 based on whether each pixel's intensity value is greater than or less than the central pixel's intensity value. These binary labels are then concatenated to form a binary number, which is converted into a decimal value to represent the LBP value for the central pixel.

The Uniform Local Binary Pattern [24], is an extension of the LBP algorithm that is designed to reduce the number of possible LBP patterns and improve the discriminative power of the LBP descriptor is used. Here all uniform patterns have their own separate bins, while the non-uniform patterns are collected into a single bin. Uniform patterns are those that have a maximum of two transitions between 0 and 1, and vice versa. In this study, the ULBP operator was utilized to convert pixel values to binary numbers, using the eight neighboring pixels around each pixel at a radius of r = 3. The image was divided into non-overlapping blocks of 9x9 and the histogram of each block was calculated to form a feature vector, f_{ULBP}.

The LBP-HF algorithm [26] is a feature descriptor that combines Uniform Local Binary Patterns (ULBP) and Fourier coefficients. Fourier coefficients, capture frequency information in an image. By combining these two methods, LBP-HF can achieve improved rotation invariance compared to ULBP alone. In the LBP-HF algorithm, ULBP is first computed on the image to generate a binary pattern. Then, Fourier transforms are performed on the binary patterns in each image block. The Fourier coefficients are computed from the transformed binary patterns and used to represent the texture information of the image. Finally, these Fourier coefficients are normalized to obtain the LBP-HF features, f_{LBPHF}. The combination of ULBP and LBPHF, more detailed, and robust to rotation and illumination change features are extracted.

The proposed method follows multi - feature fusion approach to integrate scene key point features and local texture features. The feature fusion forms a cumulative/joint histogram h_{scene}^{i} represented as:

$$h_{scene}^{i} = [\, h(wf_{SIFTN}^{K}), h(f_{ULBP}^{i}), h(f_{LBPHF}^{i})] \qquad (3)$$

The feature dimension after fusion was (12000 x 40 bins)

3.3 Multi-level Optimum Feature Subset Selection

The multiple-level feature subset is selected using two different dimension reduction techniques, PCA and LDA. The reason behind performing multi-level feature reduction is to obtain optimum features for classification, to improve the performance of scene classification.

3.3.1 Level 1 Optimization on N-dimensional feature set

PCA is a method in statistics that can change a group of correlated variables into a new group of uncorrelated variables, called principal components. It does this by using a special mathematical process called an orthogonal linear transformation. This technique can reveal the underlying structure of the data and help identify patterns or trends among the variables. Additionally, PCA can be used for dimensionality reduction, as it transforms the original variables into a lower-dimensional space while retaining most of the important information. This can be particularly useful for reducing the complexity of high-dimensional datasets and improving the performance of statistical models.

The extracted and fused multi-features are $h_{scene}^{i} = (h_{scene}^{1}, h_{scene}^{2}, \ldots, h_{scene}^{I})$ has dimension of I x N, where I is number of scene images with fused feature vector dimension n. The dimensions are reduced to (P << N) using PCA as described in algorithm 2. The input features are standardized to have a unit variance and zero mean. The covariance matrix's eigenvectors and eigenvalues are calculated. A feature transformation matrix is computed associated with largest eigenvalues. The projection of feature vectors into a p-dimensional subspace is done by including top p eigenvector of co-variance matrix. This is used as new basis for the further steps. This way the feature vectors h_{scene}^{i} with N dimensions have been reduced to P dimensional representation, as h_{scene}^{P} retaining maximum information about scene features.

This process transforms the d-dimensional features into k dimensional subspace (where k << d). This indicates that the first principal component has the largest variance and is orthogonal to the other principal components.

3.3.2 Level 2 Optimum feature subset selection

The h_{scene}^{P} features with P dimension were further reduced to obtain a lower dimensional space with higher inter class separation using Linear Discriminant Analysis [34]. This approach maximizes the separation of multiple scene classes, giving better recognition accuracy. The feature space h_{scene}^{P} is projected to lower subspace (L << C-1), where C is the number of scene categories. This approach performs dimension reduction without much loss of class discriminative features. The feature reduction is done. The feature optimization steps are mentioned in the algorithm 2.

```
Algorithm 2: Multilevel optimum feature subset selection (MOFSS)

Input: Cumulative histogram of fused features $h_{scene}^{i}$
Output: Optimum feature subset $h_{scene}^{P}$ and $h_{scene}^{L}$

Steps:
1.  Level 1 Optimization on N-Dimensional feature set
    $h_{scene}^{i} = \{h_{s1}, h_{s2}, \ldots h_{sN}\}$
2.  Standardize the features
    $h_{scene\ j}^{i} = \frac{h_{scene\ j}^{i} - \overline{h_{scene\,j}}^{i}}{\sigma_j} \quad \forall j$
3.  Compute mean: $\mu_s = \frac{1}{n}\sum_{i=1}^{n} h_{scene}{}^{i}$
4.  Compute covariance matrix
    $Cov(h_{scene}) = \frac{1}{N}\sum_{i=1}^{N}(h_{s_i} - \mu_s)(h_{s_i} - \mu_s)^T$
5.  Compute eigen vectors $V_i$ and eigen values $\lambda_i$ of $Cov(h_{scene})$
6.  $C_{vi} = \lambda_i V_i$    (I = 1, 2, 3, … N), N no. of features
7.  Estimating high valued eigen vectors
    $\lambda_1 > \lambda_2 > \lambda_3 > \ldots\ldots > \lambda_N$
8.  Transformation matrix: $T = [V_1, V_2, \ldots Vp]$, with p eigenvectors associated with largest eigenvalues
9.  Project the fused features into the P feature subspace:
    $h_{scene}^{P} = T\, h_{scene}^{i}$  for I = 1,2, …, N

10. Level 2 Optimization on P Dimensional feature set
    $h_{scene}^{P} \epsilon R^{d\times N}$, and $h_{scene}^{i} \epsilon R^{d\times 1}$ is i^th column of $h_{scene}$
    class $y_i \epsilon$ (1, 2,…,C),
    Nc=number of samples in each class C
11. for each class:
12.     Compute mean vector: $m_c = \frac{\sum_{i-1}^{Nc} h\,sec\,ne^{i}}{-Nc} \in R^{d\times 1}$
13.     Compute total mean vector: $m = \frac{\sum_{i=1}^{N} hsC\,ln\,l^{i}}{N} \in R^{d\times 1}$
14.     Compute within-class Scatter:
        $s_w = \sum_{C=1}^{C}\sum_{j=1}^{Nc}(h_{scene}{}^{j} - m_C)(h_{scene}{}^{j} - m_C)^T \epsilon$
15.     Compute between-class scatter:
        $S_b = \sum_{c=1}^{c} N_c \cdot (m_c - m)(m_c - m)^T \epsilon R^{d\times d}$
16.     Eigen-decomposition:
        $[V, D] = eig\,(Sw^{-1}\ Sb)$
17. end for
18.     Project the features into C subspace:
        $h_{scene}^{L}$ = Top – p eigenvectors corresponding to the largest eigenvalues
```

The mean vector, total mean vector for is class is used for computing within-class scatter and between class scatter. The eigenvectors and their corresponding eigenvalues are obtained. The eigenvectors are arranged in decreasing order and top p eigenvectors with highest eigenvalues are selected. The d * i eigenvector matrix is utilized to transform the input samples into a new subspace by multiplying the original data matrix h_{scene}^{P} by h_{scene}^{L}.

The paper follows the multi-level feature dimension reduction using two feature optimization techniques as detailed in fig. 6. The level 1 feature subspace selection maximizes the feature's variance, and level 2 maximizes the distance between different scene categories. Level 2 optimization achieves dimension reduction by identifying the directions of linear discriminants that maximize the ratio of the between-class variance to the within-class variance. The resulting linear discriminants provide a lower-dimensional representation of the data that maximizes class separability.

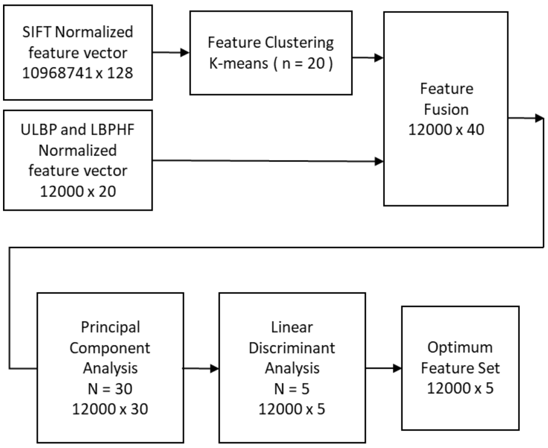

SIFT Normalized feature vector 10968741 x 128
Feature Clustering K-means (n = 20)
ULBP and LBPHF Normalized feature vector 12000 x 20
Feature Fusion 12000 x 40
Principal Component Analysis N = 30 12000 x 30
Linear Discriminant Analysis N = 5 12000 x 5
Optimum Feature Set 12000 x 5

Fig.6. Flow diagram for dimensionality reduction

The feature vectors obtained after fusion high dimensional feature of size h_{scene}^{i} 12000 x 40 is obtained. The dimension of this feature space was reduced using PCA, with number of principal components chosen as 30. Further to improve the inter class discrimination the feature vectors were optimized using LDA. The feature subspace obtained had dimensions of 12000 x 5. This resulted into significant reduction in the features with high discriminative power.

3.4 Detection and Recognition of Scene

After obtaining the optimal scene sub-features hsceneL by reducing dimensionality, the dataset consisting of six different scene categories was divided into a ratio of 80:20. The 80% of the features were utilized for training while the remaining 20% was allocated for testing. The scene classification performance of the model was evaluated using five classifiers: i) Decision Tree, ii) Random Forest, iii) KNN (K nearest neighbour), iv) LGBM (Light Gradient Boosting), and v) XGBoost (Extreme Gradient Boosting). The evaluation process utilized various metrics, including classification accuracy, precision, recall, F1-score, AUC, ROC Curve, and Confusion matrix. Accuracy measures the correct classification rate. Precision calculates the proportion of true positives among predicted positives. Recall measures the proportion of true positives among actual positives. F1 score balances precision and recall. The confusion matrix breaks down true positives, true negatives, false positives, and false negatives. The ROC curve displays sensitivity (true positive rate) versus 1-specificity (false positive rate) at various thresholds and helps assess the classifier's trade-off between precision and recall. AUC (Area Under the Curve) measures the classifier's overall performance across all possible thresholds and is useful for comparing classifiers.

4. Result & Discussions

The results of the experimentation are discussed in this section. The performance of the proposed system was carried out on three different datasets. Indoor and outdoor scene categories custom dataset, 15-Scene dataset. The evaluations of models were done to check how the individual features of scene images, and multi-feature fusion perform. Also, the analysis sis performed to compare the effectiveness of dimensionality reduction approaches using just PCA, and PCA-LDA combined. Table I details the classification results obtained on the custom dataset for different classifiers with multi-features combined (SIFT, ULBP, LBPHF) and dimensionality reduction performed using PCA and LDA together.

Table 1. Classifier Performance for Multi-Feature Fusion with PCA-LDA Dimensionality Reduction

Classifier	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Decision Tree	92.70%	92.70%	92.07%	92.07%
Random Forest	92.04%	94.37%	92.67%	93.51%
K-nearest neighbour	93.63%	93.63%	93.63%	93.63%
LGBM	**95.13%**	**95.13%**	**95.13%**	**95.13%**
XGBoost	93.958%	93.958%	93.958%	93.958%

The Decision Tree classifier achieved an accuracy of 92.70%, with corresponding precision, recall, and F1-Score also at 92.70%. The Random Forest classifier obtained an accuracy of 92.04%, with precision, recall, and F1-Score of 94.37%, 92.67%, and 93.51%, respectively. The K-nearest neighbor classifier showed an accuracy of 93.63%. The LGBM classifier demonstrated outstanding performance with an accuracy of 95.13%. The XGBoost classifier also performed commendably, achieving an accuracy of 93.958%.

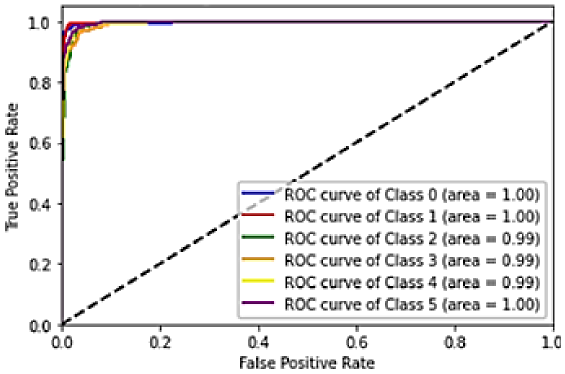


Fig. 7. ROC curve for LGBM classifiers with multi-feature fusion, PCA-LDA Dimension Reduction

LGBM and XGBoost are gradient-boosting algorithms that use a group of weak decision trees to make predictions. While both algorithms iteratively add decision trees to the model, they differ in the way they construct decision trees. LGBM uses a leaf-wise approach, while XGBoost uses a level-wise approach. In terms of both speed and accuracy, LGBM and XGBoost classifiers outperformed the Random Forest and KNN classifier. Specifically, the LGBM classifier showed fast training speed and achieved high accuracy of 95.13%.

Figure 7 shows the ROC curve for the LGBM classifier. An AUC value of 0.99 indicates excellent classification performance. LGBM implements GOSS (Gradient based one-side sampling) and EFB (Exclusive Feature Bundling) which optimizes the learning process. LGBM has achieved high accuracy with reduced computation time.

The table II details the performance of different classifiers on the feature subset generated after performing feature dimensionality reduction using PCA alone. The Decision Tree classifier achieved an accuracy of 80.16%, with corresponding precision, recall, and F1-Score also at 80.16%, indicating balanced performance. The Random Forest classifier obtained an accuracy of 80.83% and demonstrated higher precision of 95.50% and recall of 81.34%, resulting in an F1-Score of 87.80%. The K-nearest neighbor classifier showed an accuracy of 89.08%.

Table 2. Classifier performance for multi-feature fusion with PCA dimensionality reduction

Classifier	**Accuracy**	**Precision**	**Recall**	**F1-Scor**
Decision Tree	80.16%	80.16%	80.16%	80.16%
Random Fores	80.83%	95.50%	81.34%	87.80%
K-nearest neighbour	89.08%	89.08%	89.08%	89.08%
LGBM	93.25%	93.25%	93.25%	93.25%
XGBoost	**93.75%**	**93.75%**	**93.75%**	**93.75%**

The LGBM classifier achieved a accuracy of 93.25%, while the XGBoost classifier demonstrated the highest accuracy among all, with a commendable 93.75%. Both LGBM and XGBoost classifiers outperformed the others, achieving significantly higher accuracy and balanced performance across precision, recall, and F1-Score, minimizing both false positives and false negatives.

Figure 8 shows the ROC curve for the LGBM classifier with PCA alone. An AUC value of 0.99 and above indicates excellent classification performance. The AUC value close to 1 highlight that the LGBM classifier model has a high true positive rate with low false positive rate. This supports the model’s ability to effectively distinguish between classes.

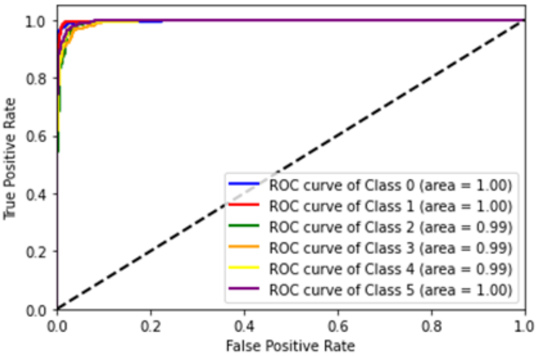


Fig. 8. ROC curve for LGBM classifiers with multi-feature fusion, PCA dimension reduction

Table 3 presents a comparative analysis of the LGBM classifier's performance on both a standard dataset and a custom dataset. The analysis evaluates the impact of different feature extraction techniques on the classifier's accuracy. When utilizing SIFT features, an accuracy of 74.25% was observed on the standard dataset. Incorporating ULBP + LBPHF feature fusion resulted in an improved accuracy of 78.85%. Further enhancement was achieved by applying PCA for feature reduction on SIFT, yielding a performance of 80.85%. The most significant performance boost was obtained through the fusion of three techniques, namely SIFT + ULBP + LBPHF + PCA, which achieved an accuracy of 93.25%. However, the SIFT + ULBP + LBPHF + PCA + LDA approach outperformed all other feature extraction and dimension reduction techniques, achieving an impressive recognition accuracy of 95.13%. The results demonstrate the efficiency of feature fusion and dimension reduction in significantly improving the LGBM classifier's performance on custom dataset.

The proposed model exhibited superior scene recognition accuracy when compared to previously published models on the 15-scene dataset. Table 4 provides a comprehensive overview of the performance comparison. The proposed model, which integrates multiple features from SIFT, ULBP, and LBPHF, outperformed other models developed using computer vision and traditional machine learning techniques. Specifically, the SIFT+ULBP+LBPHF+PCA configuration achieved an impressive precision of 84.25% and an AUC (Area Under the Curve) of 99%. Moreover, the other proposed method, SIFT+ULBP+LBPHF+PCA+LDA, achieved an even higher precision of 85.82% with the same AUC of 99%. These results highlight the effectiveness of the proposed feature fusion approach in improving scene recognition performance compared to existing models. The higher precision and AUC values indicate the model's ability to accurately identify and classify scenes from the 15-scene dataset, showcasing its potential for real-world applications in scene recognition tasks. Additionally, model is light-weight, with lower computational complexity making it more suitable for real-time deployment on embedded CPU for building portable vision-based projects. We have deployed this model on Jetson embedded development kit, equipped with monocular camera and Bluetooth headphones.

Table 3. Detailed Comparative Analysis of The Model On Custom Dataset

Method	Custom Dataset			
	Accuracy	**Precision**	**Recall**	**F1-Score**
SIFT	74.25	74.25	74.25	74.25
ULBP + LBPHF	78.85	78.85	78.85	78.85
SIFT + PCA	80.85	80.85	80.85	80.85
SIFT + ULBP + LBPHF + PCA	93.25%	93.25%	93.25%	93.25%
SIFT + LBP + LBP HF + PCA + LDA	**95.13%**	**95.13%**	**95.13%**	**95.13%**

Table 4. Comparative Analysis of The Proposed Model on 15-Scene Dataset

Method	Precision	Recall	F1-Score	AUC
LBP [26]	69.8	70.3	69.3	94.9
Uniform LBP [26]	55.6	54.7	52.6	92.8
LBP-HF [27]	63.8	63.9	62.5	92.8
PLBP [28]	52.7	52.4	51.5	92.1
SIFT_SPM [21]	62.3	57.8	57.0	94.4
SIFT_FV [23]	79.6	79.2	79.0	98.4
SIFT_ScSPM [22]	84.8	84.0	84.1	98.9
SIFT_ULBP_LBPHF_PCA (Proposed -1)	**84.25**	**84.25**	**84.25**	**99**
SIFT_ULBP_LBPHF_PCA_LD (Proposed – 2)	**85.82**	**85.82**	**85.82**	**99**

5. Test and Evaluation Of the Sceneguide Aid

The proposed scene recognition model was implemented on an embedded processing board called Jetson Nano, and the system utilized a monocular camera along with Bluetooth headphones or earplugs as input and output devices. The aid was designed in three different forms: a wearable jacket that housed the processing unit inside the backpack, a wearable jacket with the processing unit inside the side pockets of the jacket, and a wearable vest that contained both the camera and processing unit. All three forms of the aid were powered by rechargeable battery packs.

In order to capture scene details from a better front view, the scientists placed the monocular camera on the jacket and determined the placement dimensions as illustrated in Figure 3 (b). The distance between the shoulder joints (d (a, b)) and the distance between the shoulder and hip joint (d (a, e)) were measured, and the midpoint (d (a, b) / 2) was used to determine the horizontal position of the camera sensor center. For the vertical position, the center of the camera was fixed at a ratio of 2:3 from the shoulders (2/3 * d (a, e)). After calculating the placement point (c, d), the camera was secured inside a Velcro fixture to prevent movement during navigation. A push button was also fixed below the camera, with a separation of 1cm ± 0.5cm, to generate a request for the processing unit to perform the recognition operation.

During the testing and evaluation of the SceneGuide aid, ten blindfolded volunteers were recruited. The participants were comprised of six individuals aged between 18-30 years and four individuals aged between 30-45 years. Of the ten participants, seven were male and three were female. Within the age group of 18-30 years, there were five male and two female volunteers, while within the age group of 30-45 years, there were two male and one female volunteers.

The effectiveness of the proposed aid was compared in two cases: (i) volunteers equipped with only a white cane (180 trials), and (ii) volunteers equipped with a white cane along with a wearable SceneGuide Jacket or vest (180 trials). The volunteers were placed at a distance of 4 to 10 meters away from the scene event, and 360 scene recognition trials were conducted in a repeated set to ensure recognition accuracy in unknown scenes. During the evaluation, volunteers were trained for 10-12 hours across six different categories of scenes and instructed on how to use a white cane and rely on acoustic cues from the environment to make a final decision about the scene. Additionally, when equipped with the SceneGuide aid along with the white cane, volunteers were trained on how to generate requests for scene recognition. Each volunteer was presented with each category of scene three times, with and without the SceneGuide aid. The volunteers were provided with earphones to wear in one ear only, which allowed the other ear to capture ambient acoustic signals without obstruction.

Table 5. The Mean Response Time Observed for Scene Recognition by Blind-Folded Volunteers

Sr. No.	Scene Category	White cane and surrounding sensory cues Avg (s)	White Cane + SceneGuide Aid Avg (s)
1	Kitchen	22.42 ± 5.00	7.01 ± 1.40
2	Living Room	21.38 ± 6.47	7.38 ± 1.74
3	Market	15.84 ± 5.71	7.16 ± 1.77
4	Shopping Place	25.01 ± 5.37	8.46 ± 3.30
5	Bus Stop	21.59 ± 5.25	9.77 ± 3.27
6	Auto Rickshaw	20.71 ± 7.30	8.21 ± 2.69

Table V presents a comparative analysis of the time taken by volunteers with and without the SceneGuide aid as a complement to the white cane. The proposed aid requires 1.9 seconds to generate the scene type result after the volunteer presses the push button to request scene understanding. Additionally, it takes 1 second to produce an audio message such as "Shopping Place" or any other scene class.

The results of the evaluation indicate that the volunteers equipped with a white cane and SceneGuide aid had a significantly faster recognition time, with an average recognition time of 8.00 ± 2.36 seconds. In comparison, the volunteers equipped only with a white cane responded to the unfamiliar scene with an average response time of 21.16 ± 5.85 seconds. The difference between the two groups is quite significant, with the volunteers equipped with SceneGuide recognizing the scene 62.19% faster than those without it.

It was observed that 20% of volunteers between the ages of 30 and 45 were unable to recognize the scene category in their first attempt and required two or more responses from the SceneGuide aid to confirm the category. During indoor testing, some volunteers struggled to align themselves with the kitchen scene or living room scene and instead were misaligned with a wall, which led to incorrect responses from the aid. The effectiveness of the proposed aid in generating accurate scene responses is limited when the volunteer is not able to properly align with the scene. The recognition of public transport scenes such as bus stops and auto-rickshaws were more difficult due to the high level of ambient noise caused by vehicles and crowd. This made it challenging for volunteers to clearly hear the feedback from the aid, resulting in a longer recognition time. In 6.66% of cases, volunteers were unable to confirm the scene category even though the aid generated the correct response.

The Market scene and shopping place scene are typically characterized by crowded environments, which provide a variety of sensory cues for recognition. However, differentiating between the Market scene and shopping place scene can be challenging due to their similar auditory cues, making it difficult to distinguish between the two places. The evaluations demonstrate that in such cases, the vision-based recognition of the scene plays a crucial role. While a traditional white cane mainly relies on tactile feedback from the surrounding environment, the SceneGuide Aid provides a comprehensive and accurate understanding of the user's surroundings.

The volunteers experienced a reduction in anxiety and were able to relate the sensory channel feedback with the system's feedback. Additionally, the arching of the cane and perceptual efforts required to comprehend scene were reduced. As a result, volunteers experienced less confusion and a reduced cognitive load while navigating in the indoor and outdoor environment.

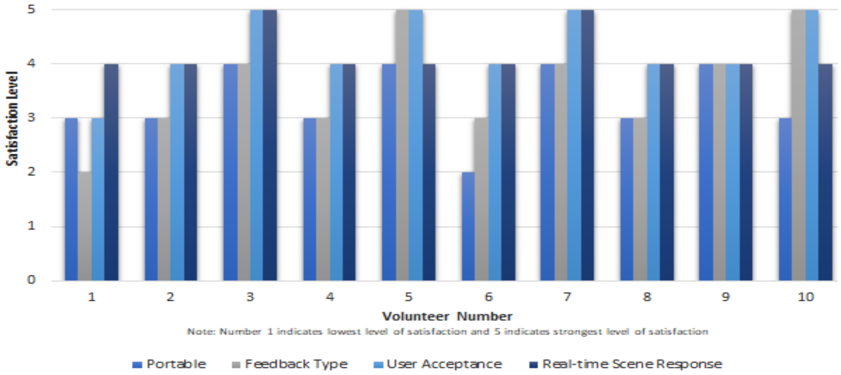
The proposed model is best suited for real-time performance with low memory requirement and faster recognition response. The SceneGuide Aid appears to be a promising solution for individuals with visual impairments who require assistance in navigating complex environments. Its real-time performance and ability to provide scene recognition make it a valuable tool for enhancing mobility and independence.

Fig. 9 presents insightful data on the usability evaluation of the SceneGuide aid based on feedback from the participating volunteers. The satisfaction levels for the aid's functionalities were rated on a scale of 1 to 5, with a higher score indicating a higher level of satisfaction. These evaluations were conducted to determine the aid's effectiveness in enhancing the navigation experience of individuals with visual impairments. The aid's design features were evaluated, including its portability, feedback type, real-time response, and user acceptance.

The aid's portability feature was rated moderately, with an average satisfaction score of 3.3 out of 5. This score indicates that the aid's lightweight, portable, and battery-operated design is suitable for easy carrying in a backpack or as a wearable vest or jacket. The portability of the aid is a critical aspect since it enables the user to carry it wherever they go, ensuring that they always have access to the aid's assistance. The auditory feedback system of the aid was found to be highly effective, with an average satisfaction score of 3.6 out of 5. This score indicates that the aid generates audio messages only when prompted by the user via a push button, which prevents unnecessary repetition of information. This feature allows the user to request multiple feedbacks in situations where confirmation was required, enabling them to utilize their hearing for other acoustic cues during navigation. The auditory feedback system provides optimum feedback without interfering with other sensory modalities, ensuring a seamless navigation experience for the user.

The aid's real-time response and user acceptance were highly rated, with an average satisfaction score of 4.2 and 4.3, respectively. These scores suggest that the aid's assistive functions were perceived as useful and appropriate by the users and could potentially improve their overall scene recognition experience. The real-time response of the aid is critical since it provides immediate feedback to the user, enabling them to navigate through their surroundings with confidence.

The usability evaluations of the SceneGuide aid highlight its potential to significantly enhance the scene recognition for individuals with visual impairments. The aid's design features, including its portability, feedback type, real-time response, and user acceptance, were rated favorably, indicating that the aid's assistive functions were perceived as useful and appropriate by the users.



Note: Number 1 indicates lowest level of satisfaction and 5 indicates strongest level of satisfaction

Fig. 9. Scene-Guide Real-time Evaluation Feedback from blind-folded volunteers

6. Conclusion

Real-time wearable SceneGuide Aid is a significant contribution to assistive technology for visually impaired individuals, as scene recognition is a challenging task for them. The proposed method's experimental results demonstrate its high accuracy compared to state-of-the-art methods. This can be attributed to the feature fusion of SIFT, ULBPH, and LBPHF used in the method, as well as feature dimension reduction using PCA and LDA resulted in optimum features for classification with strong recognition. The highest recognition accuracy of 95.24 % was observed using LGBM classifier. The evaluation process utilized various metrics, including classification accuracy, precision, recall, F1-score, AUC, ROC Curve, and Confusion matrix. The model was deployed on Jetson-Nano Development board, with monocular camera for capturing the scene. To validate the method a different standard dataset 15-scene was used. The real-time usability evaluation of the model was done with involvement of 10 blind-folded volunteers. The wearable, light weight characteristics of SceneGuide Aid make it accessible and convenient for blind users. The real-time nature of the system also ensures that users can receive immediate feedback on their surroundings, which can reduce their cognitive and perceptual efforts while indoor and outdoor navigation.

Future plans involve evaluating the aid with visually impaired individuals and further miniaturizing the system. The training curriculum for visually impaired users will also be assessed, and additional scene categories such as hospitals, pharmacies, ATMs, and temples will be incorporated into the next version. Through these efforts, the SceneGuide Aid has the potential to significantly enhance the quality of life for the visually impaired by promoting greater confidence and independence.

References:

[1] World Report on Vision. 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

[2] Seybold, Diana. "Investigating stress associated with mobility training through consumer discussion groups." Journal of Visual Impairment & Blindness 87, no. 4 (1993): 111-112.

[3] S. Bhatlawande, M. Mahadevappa, J. Mukherjee, M. Biswas, D. Das and S. Gupta, "Design, Development, and Clinical Evaluation of the Electronic Mobility Cane for Vision Rehabilitation," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 22, no. 6, pp. 1148-1159, Nov. 2014..

[4] Lopes, Sérgio I., José MN Vieira, Óscar FF Lopes, Pedro RM Rosa, and Nuno AS Dias. "MobiFree: a set of electronic mobility aids for the blind." Procedia Computer Science 14 (2012): 10-19.

[5] Balakrishnan, Meenakshi Rao, Parigi Vedanti Madhusudhan Valiyaveetil, Sashi Kumar Paul, Rohan Venkatesan, Arun Kumar Harikesavan, Karthikeyan Kolappan, Bhagavatheesh Chanana, Piyush Mehra, Dheeraj, "A Split Grip Cane Handle Unit With Tactile Feedback for Directed Ranging", Patent, WO/2015/121872, issued 20/08/2015

[6] Dernayka, Aya, Michel-Ange Amorim, Roger Leroux, Lucas Bogaert, and René Farcy. "Tom Pouce III, an electronic white cane for blind people: Ability to detect obstacles and mobility performances." Sensors 21, no. 20 (2021): 6854.

[7] S. S. Bhatlawande, J. Mukhopadhyay and M. Mahadevappa, "Ultrasonic spectacles and waist-belt for visually impaired and blind person," 2012 National Conference on Communications (NCC), Kharagpur, India, 2012, pp. 1-4.

[8] Katzschmann, Robert K., Brandon Araki, and Daniela Rus. "Safe local navigation for visually impaired users with a time-of-flight and haptic feedback device." IEEE Transactions on Neural Systems and Rehabilitation Engineering 26, no. 3 (2018): 583-593.

[9] Kilian, Jakob, Alexander Neugebauer, Lasse Scherffig, and Siegfried Wahl. "The unfolding space glove: A wearable spatio-visual to haptic sensory substitution device for blind people." Sensors 22, no. 5 (2022): 1859.

[10] J. Bai, S. Lian, Z. Liu, K. Wang and D. Liu, "Virtual-Blind-Road Following-Based Wearable Navigation Device for Blind People," in IEEE Transactions on Consumer Electronics, vol. 64, no. 1, pp. 136-143, Feb. 2018.

[11] Meshram, V. V., Patil, K., Meshram, V. A., & Shu, F. C. (2019). An astute assistive device for mobility and objectrecognition for visually impaired people. IEEE Transactions on Human-Machine Systems, 49(5), 449-460.

[12] Dutta, Senjuti, Mridul S. Barik, Chandreyee Chowdhury, and Deep Gupta. "Divya-Dristi: A smartphone-based campus navigation system for the visually impaired." In 2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT), pp. 1-3. IEEE, 2018.

[13] Garcia-Macias, J. Antonio, Alberto G. Ramos, Rogelio Hasimoto-Beltran, and Saul E. Pomares Hernandez. "Uasisi: A modular and adaptable wearable system to assist the visually impaired." Procedia Computer Science 151 (2019): 425-430.

[14] Tap Tap See App, 2012 [Mobile App]. Available: https://play.google.com/store/apps/details?id=com.msearcher.taptapsee.android&hl=en_US

[15] Be My Eyes App, 2015 [Mobile App]. Available: https://play.google.com/store/apps/details?id=com.bemyeyes.bemyeyes&hl=en&gl=US

[16] Szummer, Martin, and Rosalind W. Picard. "Indoor-outdoor image classification." In *Proceedings 1998 IEEE International Workshop on Content-Based Access of Image and Video Database*, pp. 42-51. IEEE, 1998.

[17] Oliva, Aude, and Antonio Torralba. "Modeling the shape of the scene: A holistic representation of the spatial envelope." *International journal of computer vision* 42 (2001): 145-175.

[18] Han, Yina, and Guizhong Liu. "A hierarchical GIST model embedding multiple biological feasibilities for scene classification." In *2010 20th International Conference on Pattern Recognition*, pp. 3109-3112. IEEE, 2010.

[19] Lowe, David G. "Distinctive image features from scale-invariant keypoints." *International journal of computer vision* 60 (2004): 91-110.

[20] Fei-Fei, Li, and Pietro Perona. "A bayesian hierarchical model for learning natural scene categories." In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, pp. 524-531. IEEE, 2005.

[21] Lazebnik, Svetlana, Cordelia Schmid, and Jean Ponce. "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories." In *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, vol. 2, pp. 2169-2178. IEEE, 2006.

[22] Yang, Jianchao, Kai Yu, Yihong Gong, and Thomas Huang. "Linear spatial pyramid matching using sparse coding for image classification." In *2009 IEEE Conference on computer vision and pattern recognition*, pp. 1794-1801. IEEE, 2009.

[23] Sánchez, Jorge, Florent Perronnin, Thomas Mensink, and Jakob Verbeek. "Image classification with the fisher vector: Theory and practice." *International journal of computer vision* 105 (2013): 222-245.

[24] Ojala, Timo, Matti Pietikäinen, and David Harwood. "A comparative study of texture measures with classification based on featured distributions." *Pattern recognition* 29, no. 1 (1996): 51-59.

[25] Xiao, Jianxiong, Krista A. Ehinger, James Hays, Antonio Torralba, and Aude Oliva. "Sun database: Exploring a large collection of scene categories." *International Journal of Computer Vision* 119 (2016): 3-22.

[26] Ahonen, Timo, Jiří Matas, Chu He, and Matti Pietikäinen. "Rotation invariant image description with local binary pattern histogram fourier features." In *Image Analysis: 16th Scandinavian Conference, SCIA 2009, Oslo, Norway, June 15-18, 2009. Proceedings 16*, pp. 61-70. Springer Berlin Heidelberg, 2009.

[27] Song, Kechen, and Yunhui Yan. "A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects." *Applied Surface Science* 285 (2013): 858-864.

[28] Qian, Xueming, Xian-Sheng Hua, Ping Chen, and Liangjun Ke. "PLBP: An effective local binary patterns texture descriptor with pyramid representation." *Pattern Recognition* 44, no. 10-11 (2011): 2502-2515.

[29] Wu, Jianxin, and Jim M. Rehg. "Centrist: A visual descriptor for scene categorization." *IEEE transactions on pattern analysis and machine intelligence* 33, no. 8 (2010): 1489-1501.

[30] Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 1, pp. 886-893. Ieee, 2005.

[31] Xiao, Jianxiong, Krista A. Ehinger, James Hays, Antonio Torralba, and Aude Oliva. "Sun database: Exploring a large collection of scene categories." *International Journal of Computer Vision* 119 (2016): 3-22.

[32] Ke, Yan, and Rahul Sukthankar. "PCA-SIFT: A more distinctive representation for local image descriptors." In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 2, pp. II-II. IEEE, 2004.

[33] Malhi, Arnaz, and Robert X. Gao. "PCA-based feature selection scheme for machine defect classification." *IEEE transactions on instrumentation and measurement* 53, no. 6 (2004): 1517-1525.

[34] Ye, Jieping, Ravi Janardan, and Qi Li. "Two-dimensional linear discriminant analysis." *Advances in neural information processing systems* 17 (2004).

[35] Prewitt, Judith MS. "Object enhancement and extraction." *Picture processing and Psychopictorics* 10, no. 1 (1970): 15-19.

[36] LeCun, Yann, Léon Bottou, Yoshua Bengio, and Patrick Haffner. "Gradient-based learning applied to document recognition." *Proceedings of the IEEE* 86, no. 11 (1998): 2278-2324.

[37] Gong, Yunchao, Liwei Wang, Ruiqi Guo, and Svetlana Lazebnik. "Multi-scale orderless pooling of deep convolutional activation features." In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VII 13*, pp. 392-407. Springer International Publishing, 2014.

[38] Yang, Songfan, and Deva Ramanan. "Multi-scale recognition with DAG-CNNs." In *Proceedings of the IEEE international conference on computer vision*, pp. 1215-1223. 2015.

[39] Chen, Liang-Chieh, George Papandreou, Florian Schroff, and Hartwig Adam. "Rethinking atrous convolution for semantic image segmentation." *arXiv preprint arXiv:1706.05587* (2017).

[40] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778. 2016.

[41] Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. "Going deeper with convolutions." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1-9. 2015.

[42] Huang, Gao, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. "Densely connected convolutional networks." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700-4708. 2017.

[43] Wang, Chen, Guohua Peng, and Bernard De Baets. "Deep feature fusion through adaptive discriminative metric learning for scene recognition." *Information Fusion* 63 (2020): 1-12.

[44] Wang, Chen, Guohua Peng, and Bernard De Baets. "Embedding metric learning into an extreme learning machine for scene recognition." *Expert Systems with Applications* 203 (2022): 117505.

[45] Parshapa, P. ., & Rani, P. I. . (2023). A Survey on an Effective Identification and Analysis for Brain Tumour Diagnosis using Machine Learning Technique. International Journal on Recent and Innovation Trends in Computing and Communication, 11(3), 68–78. https://doi.org/10.17762/ijritcc.v11i3.6203

[46] Muñoz, S., Hernandez, M., González, M., Thomas, P., & Anderson, C. Enhancing Engineering Education with Intelligent Tutoring Systems using Machine Learning. Kuwait Journal of Machine Learning, 1(2). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/116