RESEARCH PAPER

Appearance-based indoor place recognition for localization of the visually impaired person

Dlnya S. Salih, Abbas M. Ali
Department of Software and Informatics, College of Engineering, Salahaddin University-Erbil, Kurdistan Region, Iraq.

ABSTRACT:
Indoor localization and mapping is an important issue in computer vision. Many approaches have been proposed and used to give an accurate process of localization, most of them have limitations and cannot precisely recognize places since this challenge involves many issues like a random representation of features not based on spatial domains let mismatch of finding the corresponding image in an accurate way. In addition, some other minor problems are related to the way of features extraction like octave, Haar,…etc. Hence, it still is regarded as an open problem. The proposed system uses and compares different machine learning techniques for feature extraction like BOW, HOG, and EOH for visual place recognition in a way that improves accuracy and robustness of indoor localization for the visually impaired person. Here we combined several powerful approaches, then applied them to two international datasets (COLD and IDOL) and found more accurate results as compared to using each method separately.

KEY WORDS: Visual place recognition BOW, HOG, EOH, indoor localization, visually impaired
DOI: http://dx.doi.org/10.21271/ZJPAS.31.4.8.
ZJPAS (2019), 31(4);70-81.

INTRODUCTION:
Indoor localization and Mapping system are very significant for assisting visually impaired person in order to localize themselves and also navigating indoor environments that are not familiar for them in their normal daily lives. Based on the information from the World Health Organization, there are 285 million people with disabilities in their vision where 39 million among them are blind (Hu et al., 2016).

Currently, so many buildings are generally built for sighted people, thus localization and navigation tasks in the environment that normal people do it without any problem could be a great issue for the visually impaired person. Even various systems have designed for the purpose of assisting visually impaired person for robotic navigation, efficient solutions are not found yet (Hu et al., 2016). Here it is proposed to use and compare different types of approaches to accomplish the task; handling some issue which involves enhancing the accuracy and robustness of appearance-based recognition of locations this will aim to implement a visual localization system for the visually impaired person to demonstrate improvements in robustness and accuracy over existing methods.

The image in front of the visually impaired person will be compared with other images which are already stored in the computer database. This will lead us to think of transforming the images into another form of information and store it in the
database. Different types of algorithms will be used to complete this task and will be examined and compared (Ugave, 2014).

Here machine-learning techniques are combined for speeding up localization or visual appearance-based place recognition for the visually impaired person in a robust enough with fewer constraints in real time. This paper aims To design an efficient algorithm in machine learning for appearance-based place recognition using common datasets in this field could be applied to a sequence of live images and then be implemented in real-time modes.

2. LITERATURE REVIEW AND THEORETICAL BACKGROUND

2.1 Place Recognition and Simultaneous Localization and Mapping

A map is required for a robotic localization system, for recognizing a place where it is already visited. Simultaneous Localization and Mapping (SLAM) are the processes of creating a map about the environments and localization within the map. The process of simultaneously building a map about the surroundings and localizing itself within the map is called Simultaneous Localization and Mapping (SLAM), and it has been a very active field in the past four decades. To support visually impaired person for the purpose of possessing safe and quality life, researchers have developed aiding ways and machines.

SLAM is applied in several applications like VSLAM. Several vital features like strength, scalability in detailed mapping, semantic mapping representations and metric explained during a review regarding SLAM. This review utilizes novel sensors and concepts like deep learning to handle active SLAM and its new boundaries exploration (Cadena et al., 2016). Different sorts of VSLAM implementations that uses no filter has been reviewed so far that describes some methods to find out ingredients of SLAM (Younes et al., 2016), for estimating offline sub-maps and aligning them to international map in efficient way an indirect monocular SLAM that uses no filter is proposed (Bourmaud and Mégret, 2015).

Though numerous inquiries were completed utilizing monocular and stereo vision priing harily based algorithms with applications, for instance, obstacle avoidance, localization, and following robots, only some of them concentrate on visually impaired person’s navigation. Using a wired camera for sence enlargement and 3D map building a combination of mono SLAM and object recognition were utilized (Castle, Klein and Murray, 2010). Location of the object is discovered by the homography decomposition between the pair of current and former image. Device with hand-pushed mobility ability has been created utilizing dual cameras – one among them for recognizing the scence utilizing GIST features and estimating pose whereas the other is for enhancing Fast-Appearance-Based map (FAB-MAP) via marker detection.

Scene recognition includes a training procedure offline before the online approach is performed in a precise online direction (Nguyen et al., 2017). 6DOF posture estimation technique concentrates on using a 3D camera that is firmed on the white cane with an active Rolling Tip (ART) at its termination that can lead itself in the desirable way of a tour. It utilizes Gaussian measurement Model (GMM) based on pattern recognition approach presenting good outcomes for stairway and obstacle detection (Hong, Qian and Wu, 2016). For interaction of human and device an interface for speech that uses a Bluetooth headphone and keyboard is planned to be used. In order to get the final pose utilizing VO and VSLAM, visual depth measurements are needed. With utilizing a static monocular camera the depth from an object can be measured values of object and height of the camera are known (Diamantas, Astaras, and Pnevmatikakis, 2016).

Sparse-depth data depth gained from a RGBD camera 3D lidar with monocular camera for estimating posture and recovery utilized by Depth improved VO (Zhang, Kaess and Singh, 2014). An intelligent system was developed for localization associate deegred navigation for visually impaired utilizing a monocular camera in interior surroundings. The proposed system is tested for 2 datasets that are standard Karlsruhe and indoor environment recorded datasets (Ramesh et al., 2018).

A wireless system for indoor way finding developed to help visually impaired person so as to navigate the unknown environment with steady and dynamic obstacles. The most effective rout can be found by the system for connecting a starting and a goal point in an indoor framework whereas giving hints to the visually impaired person regarding moving on the surface by passing obstacles (Milici, et al 2018).
2.2 Speeded Up Robust Features (SURF)

SIFT feature descriptor which is most widely accepted shows the big cost of computations. H.Bay et al. presented a speeded up variant of SIFT named as SURF (Speeded Up Robust Features). For each key point, SIFT calculates a 128-dimensional feature descriptor of neighbors based on local gradients histogram while SURF produces a 64-dimensional feature descriptor relying on sums of Haar wavelet components. SURF is approximated. By using Laplacian of Gaussian with a box filter instead of the difference of Gaussian. The calculation time has decreased highly in SURF because of utilizing integral images which performs box type filter calculation very fast. Based on the Hessian matrix determinant the interest points are detected. The selection of interest point at each location in the image over different scales is computed based on local maxima of Hessian matrix determinant, to permit invariance of rotation orientation is associated for every key point.

By calculating the Haar wavelet response of each key point in a circular neighborhood the orientation is computed. With the strength of horizontal and vertical responses, the wavelet responses are plotted as points on the two axes. Every response within the sliding window is computed and the leading orientation is chosen as the key point orientation. By considering a square box filter of size 20 around each key point oriented along its perspective orientation the SURF descriptors are computed. The area around each key point is split into 4×4 subregions then for each block, Haar wavelet response in the horizontal and vertical direction is calculated. Then each subregion is represented by using a four-dimensional descriptor that contains the wavelet responses summation ($d_x$ and $d_y$) and the summation of absolute values of $d_x$ and $d_y$ as illustrated by equation (1) (Abedin, Dhar and Deb, 2017):

$$v = (\Sigma d_x, \Sigma d_y, \Sigma |d_x|, \Sigma |d_y|)$$  \hspace{1cm} (1)

2.3 Bag of Words (BOW)

for analyzing text document bag-of-words (BoW) technique was initially presented in the problem of text retrieval domain, it was also acclimatized for applications of computer vision. In the BoW model, a visual analogue of a word is utilized for analyzing the image, which depends on the process of quantizing the vector though clustering low-level visual features of local areas or points, like color, structure and so on. For extracting BoW features from images consist of the following four steps:

1. Interest area or point detection in an automatic way,
2. Local descriptor calculation over detected areas/points,
3. Creating visual vocabulary by quantizing the descriptors to words,
4. BoW feature construction by finding the appearance inside the image for every certain word in the visual vocabulary.

The BoW approach can be explained as the following. Having a dataset S including m images expressed by $S = s_1, s_2, ..., s_m$, where $s_i$ represents visual features, a particular algorithm for unsupervised learning, like k-means, is utilized for grouping $S$ depending on a constant number of visual words $O$ expressed by $O = o_1, o_2, ..., o_C$, when $C$ represent the cluster number. Then, the data can be summarized in a C×M concurrence table of counts $M_{ij} = m(o_i, s_j)$, where $n(o_i, s_j)$ indicates how many times the image $s_i$ contains the word $o_i$ (Tsai, 2012).

2.4 Edge Oriented Histogram (EOH)

Edge orientation histograms (EOH) is utilized in some tasks such as classification and detection as a feature descriptors. The local orientation or the directions of the edge are described by those descriptors. The histograms are constructed by starting with calculating the image edge. Here, after converting the image to greyscale the edge is calculated through filtering the image, the following kernels are utilized for the filtering process: $[-1,0,1]$ and $[-1,0,1]^T$. The obtained filtered images are indicated by $d_x$ and $d_y$ respectively. The direction ($\alpha$) and magnitude ($M$) of the edge for the pixels within the image are calculated by Equation (2) and Equation (3) respectively. The producing $\alpha$ will be perpendicular to real direction of the edge in the image.

$$\alpha = \arctan\left(\frac{d_y}{d_x}\right)$$  \hspace{1cm} (2)

$$M = \sqrt{d_x^2 + d_y^2}$$  \hspace{1cm} (3)

The next phase is separating the image into small areas which are not overlapping, named cell. Here rectangular cells of $c_x \times c_y$ pixels is used by
this step $n_x \times n_y$ cells are generated. If the size of the image is $x \times y$, then $n_x = \lfloor x/c_x \rfloor$ and $n_y = \lfloor y/c_y \rfloor$. Single orientation histogram will describe each cell. The bins of edge orientation histogram are equally spread through direction of the edge $0^\circ – 180^\circ$ where the gradient sign is neglected. If $d^\circ$ is the histogram bin size, then number of histogram bins $n_d$ is $180^\circ/d^\circ$. Based on the magnitude of edge, the histogram is calculated depending on the weighted vote (Timotius and Setyawan, 2015).

**Histogram of Gradient (HOG):** HOG features give a summarized but robust representation of images for classification of the image in general (Peter A. Torrione et al, 2014). The fundamental concept beyond HOG features is that the allocation of gradients’ intensity or the direction of edge describes the appearance of local object and shape in the image. HOG splits the image into cells which are small areas connected to each other, and for the pixels inside each cell, a histogram of gradient directions is compiled by HOG. Based on the founded value in the computation of gradient, every pixel inside the cell casts for the channel of orientation-based histogram a weighted vote. Along 0 to 180 degrees the histogram channels are equally spread, then the feature vector is represented by the collection of those cell histograms. It will be contrast normalized via calculation of an intensity measure through the greater image region, named block, afterward this value is utilized for normalizing every cell inside the block. More preferable invariant in lightening or shadowing results from this normalization (Ren, and Li 2014). Figure 1 illustrates the HOG extraction process.

![HOG extraction process](image)

**Figure 1:** HOG feature

### 2.4 Support Vector Machine (SVM)

SVMs are originally classifiers with two class which is presented to be an appealing and more formal technique for linear or non-linear decision boundaries learning. Having a group of points, that to either of two classes belong, the hyperplane is found by SVM and it leaves the biggest possible number of fraction points on the same side of the same class while increasing the space from hyperplane of either class. This is equal to minimizing structural risk for obtaining generalization with good results from two classes supposing 1 example

$$(x_1,y_1), (x_2,y_2) \ldots (x_i,y_i), x_i \in \mathbb{R}^N, y_i \in \{-1,+1\}$$

(4)

Quadratic programming is utilized to find the most preferable hyperplane implies fixing a constrained improvement issue. The margin width between classes is the criteria for optimization. The distinguished hyperplane is illustrated by Equation (5).

$$f(x) = \sum_{i=1}^{l} y_i a_i k(x,x_i) + b$$

(5)

Where kernel function is $k(x,x_i)$ and the membership of $x$ is indicated by the $f(x)$ sign. The optimal hyperplane construction is equal to finding every $a_i$ that are nonzero. The optimal hyperplane support vector is every $x_i$ data point that corresponds to nonzero $a_i$ (Sun, Bebis and Miller, 2002).

### 3. METHODOLOGY AND DESIGN

#### 3.1 Proposed Method

In general the proposed method consists of the following steps

---

73 ZANCO Journal of Pure and Applied Sciences 2019
1. Input stage: Initially the two international datasets that have been used for place recognition purpose in the literature are obtained from a reliable source of information such as machine learning websites and used here for enhancing the accuracy percentage which is COLD and IDOL datasets. Here all the images from datasets are going to be ready for the next step which is feature extraction using a powerful descriptor, this phase also includes data collection and other preprocessing activities such as removing noisy data and redundant information and image filtering such as converting a colored image to grayscale is performed in this stage as well.

2. Feature extraction: This step is the core phase of the system where local features are extracted using speeded up robust feature (SURF). There are two types of feature detection global and local feature here interest points are extracted using local features. Local feature extraction consists of two steps which are detection and description. In this thesis, a robust descriptor which is SURF is utilized for selecting the 100 strongest local features within each image of the whole three databases (COLD and IDOL). SURF size of 128 is used and the threshold of 600 is used as well. HOG descriptor which is an excellent descriptor for extracting local features is also used. In order to enhance recognition accuracy, another technique is used for feature extraction called EOH. Here some of these techniques are combined as well, for obtaining better results as shown in Figure 2.

![Gray Scale image (from COLD dataset)](image)

Figure 2: Features are first extracted from each image within the dataset then concatenated together to create a new feature vector for every image.

3. Localization process: This step also can be referred to as a recognition process. For recognition task clustering techniques and machine learning, classification methods are used. For clustering purpose, unsupervised learning technique which is known as k-mean is used for grouping similar features together. Training the dataset using a different number of centroids within the range [50-600] and figuring out which one obtains the best localization accuracy. BOW technique has been used for the result of every used centroid of the images inside the datasets. One row of BOW feature is obtained and saved together for all the images in an excel sheet then the result is saved in a readable format by Weka application in...
order to be classified using SVM, random forest and naïve Bayes or any other classifier.
To specify the type of the image feature for example, for idol dataset the type field contains five labels (one-person office, two-person office, printer area, corridor, and kitchen), a tag is assigned for every feature vector that represents each image in the dataset as shown in Figure 3.

**Figure 3**: graphical representation of the training and classification process

4. Reading next frame from the live images then returning back to step two to calculate features for new frames. The overall process is shown in Figure 4.

**Figure 4**: The general process
The following subsections give the explanation for both datasets that have been used in this paper.

### 3.2 COsy Localization Database (COLD)

The COLD is the abbreviation of COsy (Cognitive System for Cognitive Assistant) Localization Database. For assessing the localization system which is primarily vision-based for working on mobile platforms in a rational setting, this database is representing an attempt for providing an environment on a large scale and has flexibility in testing. This database has three types (COLD-Freiburg, COLD-Ljubljana, COLD-Saarbrücken) that acquired inside three labs in three different cities in Europe: in Germany at Freiburg University inside the laboratory of autonomous intelligent System; Slovenia, at Ljubljana University inside the laboratory of Visual Cognitive Systems; and the laboratory in Saarbrücken, Germany at the Artificial Intelligence research center.

Regular and omni-directional cameras with laser range scans and odometry data are utilized for capturing the images of the dataset. During several days under different weather and illumination conditions (cloudy, night and sunny weather) utilizing triple different mobile platforms and the same camera setup, the data were recorded. Within every three laboratories, the image capturing was carried out inside some different function rooms. Thus, for evaluating the strength of localization and recognition algorithms with respect to variations caused by human activities and changes in illumination conditions, COLD dataset considered as a perfect testing database for this purpose (Pronobis and Caputo, 2009). Figure 7 shows some examples of camera images within the database for each of the three laboratories.

**Figure 7:** samples of COLD dataset images presenting some rooms for each of three types (Pronobis and Caputo, 2009).

### 3.3 Image database for rObot Localization (IDOL)

The word IDOL is the abbreviation of Image Database for robot Localization. The dataset contains 24 sequences of images accompanied by laser scans and odometry data obtained utilizing two mobile robot platforms. In the duration of 6 months, The data was acquired inside an indoor lab circumstance including five places with diverse working (one-person office, two-people office, corridor, kitchen, and printer area) under several illumination variations (in cloudy, night and sunny weather). Consequently, natural diversity is captured that exist in the real world surroundings caused by both human activity and illumination conditions. The KTH-IDOL1 database has an extension which is KTH-IDOL2 composed of 12 image sequences brought out from KTH-IDOL1 dataset and another 12 sequences captured after 6 months (Luo et al., 2006). Figure 8 shows sample of images of the dataset.
4. RESULTS AND DISCUSSION

Here we have used several approaches to calculate Recognition accuracy which are Bag Of Word (BOW) combined with speeded up robust features (SURF) descriptor, Histogram Of gradient (HOG), Edge Oriented Histogram (EOH) and SVM classifier. Then we have combined HOG once with EOH and also combined it with the soft assignment of BOW using SVD in order to achieve more accurate results for two international datasets (COLD and IDOL). Table 1 includes accuracy percentage of COLD dataset when features of the images are extracted using SURF descriptor then K-mean clustering with a deferent number of centroids is used within the range [50-600], then BOW is calculated for each cluster the final result is achieved after SVM classifier is used. It can be noticeable that for cold sunny dataset it reaches a maximum accuracy of 68.710% when cluster 150 is used. The highest accuracy achieved is 81.369 for cold cloudy at a cluster of 100. The last column of the table 1 includes accuracies of cold night dataset when it gains best results 76.037% with a cluster of 500. Here SMO classifier is used which is an improved application of SVM classifier. Table 2 contains correctly classified instances of IDOL dataset when the combination of SURF and BOW model is used for finding features and clustering then SVM classifier is applied to the result to gain overall accuracy. It can be seen that for IDOL dataset results are reduced. Both IDOL sunny and IDOL night reach the maximum accuracy of 65.995% and 65.388% respectively at a cluster of 150. Meanwhile, IDOL cloudy gains best result 64.885% accuracy at a cluster of 100. Soft assignment feature of BOW technique using SVD is calculated also in our paper for both COLD and IDOL dataset as shown in Table 3 and Table 4 respectively. The second way that we have used for accomplishing localization task is EOH which is a powerful machine learning technique. We applied this method to the same datasets that we have used previously COLD and IDOL. As shown in Figure 5(a) cold sunny, cold cloudy and cold night reach their peak accuracy of 93.617 %, 94.109%, and 93.983% respectively when the number of regions is 4 using random forest classifier. Figure 5(b) presents the accuracy results of IDOL dataset using random forest classifier. IDOL sunny, IDOL cloudy and IDOL night gain the peak accuracy of 90.380%, 89.640% and 88.601% at r=4. Figure 6(a) represents EOH accuracies for cold dataset using SVM classifier when cold sunny, cold cloudy and cold night reach their peak accuracy of 97.246 %, 96.164%, and 93.983 % respectively when number of regions is 8. As shown in Figure 6(b) IDOL Sunny reaches the top value 87.695% at r=4, while IDOL cloudy and IDOL night have the best result 88.222% and 86.321 % respectively at r=8. The third method applied to COLD and IDOL datasets to find recognition accuracy is HOG. Table 5 contains percentage accuracy for both datasets when SVM classifier is used. Both EOH and HOG can obtain good results but we can still optimize the accuracy by using a combination of this two method (EOH+HOG) this combination obtains best results as shown in Table 6. We also combined HOG and soft assignment of BOW using SVD a notable improvement can be seen with this combination as shown in Table 7. It is worth noting that in all the obtained results no filter is used.
Table 1: Localization accuracy (in percentage) of COLD dataset when BOW and SVM classifier is used

| No. of Cluster | 50   | 100  | 150  | 200  | 250  | 300  | 350  | 400  | 450  | 500  | 550  | 600  |
|----------------|------|------|------|------|------|------|------|------|------|------|------|------|
| COLD sunny     | 66.708 | 68.335 | 68.710 | 65.832 | 64.330 | 60.45 | 67.459 | 68.335 | 64.330 | 64.455 | 59.949 | 60.700 |
| COLD cloudy    | 79.315 | 81.369 | 78.082 | 78.356 | 80.137 | 77.123 | 76.301 | 73.698 | 74.246 | 76.986 | 76.027 | 71.232 |
| COLD night     | 68.049 | 73.755 | 74.481 | 70.435 | 70.746 | 70.954 | 71.991 | 68.879 | 72.199 | 76.037 | 74.792 | 73.029 |

Table 2: Localization accuracy (in percentage) of IDOL dataset when BOW and SVM classifier is used

| No. of Cluster | 50   | 100  | 150  | 200  | 250  | 300  | 350  | 400  | 450  | 500  | 550  | 600  |
|----------------|------|------|------|------|------|------|------|------|------|------|------|------|
| IDOL sunny     | 63.758 | 63.758 | 65.995 | 60.962 | 59.172 | 57.941 | 54.474 | 57.382 | 58.053 | 55.704 | 53.803 | 53.355 |
| IDOL cloudy    | 60.414 | 64.885 | 61.395 | 57.033 | 59.978 | 59.869 | 59.760 | 56.488 | 55.725 | 56.597 | 55.943 | 53.435 |
| IDOL night     | 58.963 | 64.145 | 65.388 | 61.036 | 59.067 | 61.761 | 63.626 | 57.409 | 61.036 | 57.927 | 58.238 | 62.590 |

Table 3: Localization accuracy (in percentage) of COLD dataset for soft assignment of BOW using SVD when SVM classifier is utilized

| No. of Cluster | 50   | 100  | 150  | 200  | 250  | 300  | 350  | 400  | 450  | 500  | 550  | 600  |
|----------------|------|------|------|------|------|------|------|------|------|------|------|------|
| COLD sunny     | 58.698 | 59.199 | 58.573 | 58.573 | 58.322 | 58.322 | 59.073 | 58.197 | 58.573 | 58.573 | 58.448 | 58.322 |
| COLD cloudy    | 75   | 78.630 | 80.411 | 80.137 | 80.958 | 80.958 | 81.780 | 81.643 | 82.328 | 81.917 | 82.191 | 82.602 |
| COLD night     | 64.211 | 63.070 | 61.825 | 62.551 | 62.033 | 62.759 | 62.240 | 61.929 | 62.344 | 62.240 | 62.551 | 62.136 |
**Table 4:** Localization accuracy (in percentage) of IDOL dataset for soft assignment of BOW using SVD when SVM classifier is utilized

| No. of Cluster | 50   | 100  | 150  | 200  | 250  | 300  | 350  | 400  | 450  | 500  | 550  | 600  |
|----------------|------|------|------|------|------|------|------|------|------|------|------|------|
| IDOL sunny     | 60.962 | 60.962 | 62.080 | 60.850 | 62.639 | 59.172 | 60.962 | 61.073 | 61.745 | 61.745 | 59.955 |
| IDOL cloudy    | 55.943 | 55.507 | 54.198 | 53.762 | 53.435 | 54.198 | 54.198 | 54.307 | 52.780 | 52.780 | 51.690 | 53.653 |
| IDOL night     | 55.440 | 56.994 | 56.476 | 56.580 | 58.238 | 56.580 | 57.202 | 57.437 | 58.134 | 59.585 | 58.445 | 56.373 |

**Figure 5:** EOH classification accuracy for COLD and IDOL datasets using random forest classifier

**Figure 6:** EOH classification accuracy for COLD and IDOL datasets using SVM classifier
Table 5: HOG classification result for COLD and IDOL datasets using SVM classifier

| Dataset     | Accuracy (percentage) |
|-------------|-----------------------|
| COLD sunny  | 88.235%               |
| COLD cloudy | 92.054%               |
| COLD night  | 90.871%               |
| IDOL sunny  | 86.800%               |
| IDOL cloudy | 85.278%               |
| IDOL night  | 85.595%               |

Table 6: accuracy results of COLD and IDOL datasets for combination of EOH and HOG (EOH+HOG) using SVM classifier

| Dataset     | Accuracy (percentage) |
|-------------|-----------------------|
| COLD sunny  | 97.496%               |
| COLD cloudy | 98.082%               |
| COLD night  | 98.962%               |
| IDOL sunny  | 99.105%               |
| IDOL cloudy | 99.018%               |
| IDOL night  | 98.238%               |

Table 7: accuracy results of COLD and IDOL datasets for the combination of soft assignment of BOW using SVD and HOG (SVD+HOG) classified by SVM.

| Dataset     | Accuracy (percentage) |
|-------------|-----------------------|
| COLD sunny  | 93.241%               |
| COLD cloudy | 96.712%               |
| COLD night  | 93.879%               |
| IDOL sunny  | 96.085%               |

5. CONCLUSION

Three different methods have been used in this paper to find an accuracy percentage for two international datasets (COLD and IDOL). The first method is using a combination of SURF and BOW model where the obtained results by this method where not so accurate for COLD dataset maximum reached accuracy is 81.369%, while for IDOL it achieves its best result of 65.995%. The second used method is EOH technique where COLD dataset reaches a peak of 97.246% accuracy which is a very good result and IDOL reaches top value 88.222% accuracy. A third technique that we have used is HOG for the COLD dataset best-obtained result is 92.054% while IDOL achieves the highest accuracy of 86.800%. Although EOH obtains good result we can still improve localization accuracy by using a combination of EOH and HOG. This combination gains most accurate results in all cases where it obtains results in the range of 97.496% and 99.105%. HOG is also combined to soft assignment feature of BOW using SVD where the results are more accurate than applying each of the two methods separately. This work can be improved using Deep learning for the purpose of localization.

REFERENCES

Abedin, Z., Dhar, P. and Deb, K. (2017) ‘Traffic sign recognition using SURF: Speeded up robust feature descriptor and artificial neural network classifier’, Proceedings of 9th International Conference on Electrical and Computer Engineering, ICECE 2016, pp. 198–201. doi: 10.1109/ICECE.2016.7853890.

Bourmaud, G. and Mégret, R. (2015) ‘Robust large scale monocular visual SLAM’, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 07–12 June, pp. 1638–1647. doi: 10.1109/CVPR.2015.7298772.

Cadena, C. et al. (2016) ‘Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age’, IEEE Transactions on Robotics, 32(6), pp. 1309–1332. doi: 10.1109/TRO.2016.2624754.

Castle, R. O., Klein, G. and Murray, D. W. (2010) ‘Combining monoSLAM with object recognition for scene augmentation using a wearable camera’,
Salih, D. and Ali A /ZJPAS: 2019, 31 (4): 70-81

Image and Vision Computing. Elsevier B.V., 28(11), pp. 1548–1556. doi: 10.1016/j.imavis.2010.03.009.

Diamantas, S., Astaras, S. and Pnevmatikakis, A. (2016) ‘Depth estimation in still images and videos using a motionless monocular camera’, IST 2016 - 2016 IEEE International Conference on Imaging Systems and Techniques, Proceedings, (October), pp. 129–134. doi: 10.1109/IST.2016.7738210.

Hong, S., Qian, X. and Wu, W. (2016) ‘A Co-Robotic Cane’, (August), pp. 33–42.

Hu, F. et al. (2016) ‘Indoor Localization for the Visually Impaired Using a 3D Sensor’, pp. 192–204.

Luo, J. et al. (2006) ‘The KTH-IDOL2 Database’, Environment, pp. 1–12.

Milici, Stefano; Esposito, Ambra; Staderini, E. M. (2018) ‘A Wireless Indoor Navigation Aid for Visually Impaired People using UWB Localization Infrastructure and an Original Wayfinding Algorithm Stefano’. doi: 10.3929/ethz-b-000251651.

Nguyen, Q. H. et al. (2017) ‘Developing a way-finding system on mobile robot assisting visually impaired people in an indoor environment’, Multimedia Tools and Applications, Multimedia Tools and Applications, 76(2), pp. 2645–2669. doi: 10.1007/s11042-015-3204-2.

Pronobis, A. and Caputo, B. (2009) ‘COLD: The CoSy localization database’, International Journal of Robotics Research, 28(5), pp. 588–594. doi: 10.1117/027836490903912.

Ramesh, K. et al. (2018) ‘Real-time localization and navigation in an indoor environment using monocular camera for visually impaired’, 2018 5th International Conference on Industrial Engineering and Applications, ICIEA 2018. IEEE, pp. 122–128. doi: 10.1109/IEA.2018.8387082.

Ren, H. (2014) ‘OBJECT DETECTION USING EDGE HISTOGRAM OF ORIENTED GRADIENT Haoyu Ren Ze-Nian Li Vision and Media Lab School of Computing Science Simon Fraser University’, International Conference on Image Processing(ICIP), pp. 4057–4061.

Sun, Z., Bebis, G. and Miller, R. (2002) ‘On-road vehicle detection using Gabor filters and support vector machines’, International Conference on Digital Signal Processing, DSP, 2, pp. 1019–1022. doi: 10.1109/ICDSP.2002.1028263.

Timotius, I. K. and Setyawan, I. (2015) ‘Evaluation of Edge Orientation Histograms in smile detection’, Proceedings - 2014 6th International Conference on Information Technology and Electrical Engineering: Leveraging Research and Technology Through University-Industry Collaboration, ICITEE 2014, pp. 2–6. doi: 10.1109/ICITEEED.2014.7007905.

Tsai, C.-F. (2012) ‘Bag-of-Words Representation in Image Annotation: A Review’, ISRN Artificial Intelligence. Hindawi, 2012, pp. 1–19. doi: 10.5402/2012/376804.

Ugave, V. A. (2014) ‘SMART INDOOR LOCALIZATION USING MACHINE LEARNING TECHNIQUES’, pp. 1–149.

Younes, G. et al. (2016) ‘Keyframe-based monocular SLAM: design, survey, and future directions’, Robotics and Autonomous Systems, 98(July), pp. 67–88. doi: 10.1016/j.robot.2017.09.010.

Zhang, J., Kaess, M. and Singh, S. (2014) ‘Real-time depth enhanced monocular odometry’, IEEE International Conference on Intelligent Robots and Systems, (Iros), pp. 4973–4980. doi: 10.1109/IROS.2014.6943269.