Vision-based Position estimation and Indoor scene recognition algorithm for Quadrotor Navigation

B Anbarasu1* and G Anitha2
1School of Aeronautical Sciences, Hindustan Institute of Technology & Science, 603103, Chennai,
2Department of Aerospace Engineering, Madras Institute of Technology, Anna University, 600044, Chennai
*Corresponding author’s e-mail: avianbu@gmail.com

Abstract. In this paper, an effective and simple Grid based vanishing point detection position estimation algorithm and Enhanced GIST descriptors based indoor scene recognition algorithm for navigation of MAV in indoor corridor environment is described. Two different classifiers, k-nearest neighbour classifier and support vector machine is employed for the categorization of indoor scenes into corridor, staircase or room. Indoor scene classification was performed on Dataset-1. In the training phase of the indoor scene recognition algorithm, GIST, HODMG and Enhanced-GIST feature vectors are extracted for all the indoor training images in the Dataset-1 and indoor scene classifiers are trained for the extracted image feature vectors and assigned image labels of the indoor scenes (corridor-1, staircase-2 and room-3). In the testing phase of the indoor scene recognition algorithm, for each unknown test image frame GIST, HODMG and Enhanced-GIST feature vectors are extracted and the indoor scene classification is performed using a trained scene recognition model. The proposed indoor scene recognition algorithm using SVM with Enhanced GIST descriptors produced high recognition rates of 99.33% compared to the KNN classifiers. After recognizing the indoor scene as corridor, the MAV has to estimate its position based on the detection of vanishing point in the indoor corridor image frames. Experimental results show that the proposed method is suitable for real time operations.

1. Introduction
Indoor Unmanned Aerial systems with embedded vision system have been developed [1]. Scene recognition is the important perceptual ability required for the indoor navigation of the MAV. An integrated vision based position estimation and scene recognition algorithm have been proposed in this work for indoor navigation of the MAV. Indoor Scene recognition is indispensable for the navigation of the Micro Aerial Vehicle in GPS-denied indoor environment. After recognizing the indoor scenes, MAV has to follow a suitable indoor navigation strategy. Indoor scene recognition is a highly complex task due to variations in the intra-class indoor scenes and similarities in the inter-class indoor scenes. MAV has to identify the surrounding environment by recognizing diverse indoor scenes and this perceptual capability is very important for safe navigation through indoor environment. Vision-based navigation system has been used in this work for the navigation of Micro Aerial Vehicles using computation of the positions of vehicles in corridor environment and categorizing the indoor scenes (corridor, staircase or room).
2. Related Works

New vision method for extraction of vanishing point was introduced based on the Hough transform algorithm and K-Means clustering method [2]. This method completely eliminates the necessity to find the intersection of lines to extract the vanishing point. Position estimation method of UAV is proposed using feature points identified by image matching using SIFT algorithm [3]. This approach can be used for indoor localization of UAVs in GPS-denied environments. Indoor unmanned aerial vehicle pose has been computed using vanishing geometry [4]. Micro scale Unmanned Aircraft Systems was developed with hovering capability in indoor environments [5]. Robust vision-based navigation method is used to navigate through doorways, hallways using visual cues [6]. Robust attitude estimation method was proposed which uses combination of inertial gyro data and visual straight-line segments extracted from the man-made structures [7]. In this method classified line segments are used to estimate the current attitude on a Gaussian sphere. Vision based quadrotor micro-UAV was developed for autonomous navigation [8]. Real time pose estimation method is used for unmanned aerial vehicles navigation in an urban environment using vanishing points [9]. Vanishing point was detected in three mutually orthogonal directions of a man-made environment using dominant clusters of line segments. Vanishing points are detected in architectural environments [10]. Single image perspective cues are used for autonomous indoor navigation of MAV [11]. These cues can be used to autonomously navigate a MAV in a variety of corridors and staircases. Autonomous indoor navigation is considered with a Computationally Constrained MAV [12]. Omnidirectional and perspective vision sensors are used for robust attitude estimation and UAV stabilization [13]. The main goal of scene classification is to classify an unknown test image into one of the scene image categories, e.g. airport inside, corridor, staircase, etc.

3. Proposed Methodology

This paper has proposed an efficient vision based real time indoor scene recognition method and vanishing point detection method for MAV indoor navigation as shown in Figure. 1. Parrot AR Drone version2 Quadrotor is shown in Figure 2. Forward and bottom cameras, namely, 720p and QVGA sensors are mounted on the Parrot AR Drone version2 Quadrotor. Using this Forward and bottom cameras, images can be captured with an image resolution of 1280×720 and 320×240 pixels respectively. Using a WIFI adhoc network, the real time corridor videos are transmitted by 720p forward camera sensor at 30 frames per second. Received video is converted into a sequence of image frames and image frames are resized into a resolution of 240×320 pixels.

In the preprocessing stage, resized RGB color image frames are finally converted into a grayscale image. Videos acquired by Micro Aerial Vehicle and make the MAV to recognize the indoor environment such as corridor, staircase and room. Real time indoor scene recognition model contains Training phase and Testing phase. In the training phase, global image features and low-level features are extracted. Classifiers are learned using these extracted image feature vectors and trained image labels. In the Testing Stage of the indoor scene recognition model, for each image frame acquired by the MAV, global and local image features are extracted and fed to the classifiers for indoor scene classification.
Figure 1: Scene recognition and vanishing point detection method

Figure 2. Parrot AR drone2 quadrotor

4. Position estimation and indoor scene recognition
The proposed technique for vision-based position estimation and scene recognition for indoor navigation of MAV is described in the following sub-sections.
4.1. Canny edge detection

In the canny edge detection algorithm [22], Edges in the grayscale indoor corridor image frames are detected using the following four steps. To remove noise, Gaussian filter is applied to the input image. High spatial derivative image regions are extracted by computing the Gradient-magnitude and Gradient-orientations in the image frames. Finally, weak and strong edge pixels are extracted by using the high and low threshold values in the non-maximum suppression method and suppressing the weak edge pixels.

4.2. Hough transform algorithm for line detection

Straight lines are detected in the corridor image by applying the Standard Hough transform method [23]. In the Hough transform algorithm, straight line can be computed as follows

\[ \rho = x \cdot \cos \Theta + y \cdot \sin \Theta \]  

(1)

Where \( \rho \) denotes the perpendicular distance is denoted by \( \rho \) from the origin to the straight line, and \( \Theta \) the angle made by the perpendicular to the x-axis is denoted by \( \Theta \). For each edge pixels extracted, \( \rho \) and \( \Theta \) values are calculated and increment the accumulator array \( P \). local maxima value in the accumulator space is considered to detect the straight lines in the corridor image frames.

4.3. Grid based approach for Vanishing point detection

One of the main contributions of this work is the development of grid and straight lines based vanishing point detection method for MAV position estimation in corridor environment. Angle of straight lines with an angle greater than 80˚ and angle of straight lines with an angle less than 10˚ is removed for the vanishing point detection. Threshold values are used to remove the near vertical, vertical and near horizontal straight lines, since these lines cannot converge to a vanishing point. Each Input corridor image frames (240×320 pixels) are divided into 10×10 grids. Total number of straight line intersecting inside the grid element \( G_{p,q} \) can be obtained as follows:

\[ G_{p,q} = \sum_{i=1}^{I} p \leq \frac{w}{n_x} < p + 1,q \leq \frac{h}{n_y} < q + 1 \]  

(2)

Where \( I, w \) and \( h \) denote the total straight-line intersections, width and the height of the image frame respectively.

Maximum number of line intersections falling inside the grid cell can be computed as follows

\[ (p^*,q^*) = \arg \max_{p,q} G_{p,q} \]  

(3)

Then the initial estimate of the vanishing point can be computed as follows

\[ (a^*,b^*) = \left( \frac{w}{n}(p^* + 0.3) , \frac{h}{n}(q^* + 0.3) \right) \]  

(4)

Width and height of the input corridor image frames are denoted by \( w \) and \( h \) respectively.
To detect the vanishing point using grid method, divide the input corridor image frame into a maximum of 10×10 grid cells. Maximum number of intersections of straight lines detected inside the grid cells are indicated by p* and q* respectively. Next, compute the weighted average of all indices lying near (a*, b*) to estimate the actual and accurate vanishing point location.

Indices set (S) lying close to (a*, b*) can be computed as follows:

$$S = \{ s \in [0, I) : P_2(a_s, b_s) - (a^*, b^*) \leq \alpha \}$$  \hspace{1cm} (5)

Where α indicates the threshold distance.

Finally, actual and accurate vanishing point location can be estimated as follows:

$$\left( \overline{a}, \overline{b} \right) = \frac{1}{|S|} \sum_{s \in S} \left( a_s, b_s \right)$$  \hspace{1cm} (6)

Where \(\left( \overline{a}, \overline{b} \right)\) denotes the actual and accurate vanishing point location computed in the corridor image frames. Using the detected actual and accurate vanishing point location, the MAV position (Centre, left or right) in the indoor corridor environment can be estimated.

### 4.4. Image feature extraction

Proposed Enhanced GIST visual descriptors and GIST and HODMG state of the art visual descriptors are presented in the following subsections in detail.

#### 4.4.1. GIST

The GIST visual state of the art descriptor was proposed by Oliva & Torralba [20]. Input indoor RGB image frame (1280×720 pixels) is converted into a grayscale indoor image frame (256×256 pixels). Filtering the grayscale image to produce 32 feature maps by applying 32 Gabor filters at 8 orientations and 4 scales. 4×4 grids were obtained by dividing the 32 feature maps into 16 regions. Extracted GIST descriptor is a 512 (16×32) dimensional GIST descriptor as shown in Figure 3.

#### 4.4.2. HODMG

Line structural element is used to compute the Directional morphological gradient features [24] in horizontal and vertical directions as shown in Figure. 3. These Directional morphological gradient features exhibit sensitivity to different directions in indoor images and provide edge orientation information while suppressing detailed textural information.

Directional morphological vertical and horizontal gradients \(gL_\alpha(f)\) can be computed for a given direction \(\alpha\) as follows:

$$gL_\alpha(f) = \delta L_\alpha(f) - \varepsilon L_\alpha(f)$$  \hspace{1cm} (7)

Where, \(\delta L_\alpha(f)\), \(\varepsilon L_\alpha(f)\) and \(L\) denotes the dilated image, eroded image and structuring element respectively. Horizontal and vertical directional gradient features of the three indoor scenes, namely, corridor, staircase and room are extracted from the input videos as illustrated in Figure. 3. (c) & (d). Dimensionality of Histogram of Directional Morphological Gradient (HODMG) is a 512-dimensional descriptor.
4.4.3. Enhanced gist descriptors

Proposed Enhanced-Gist descriptor is extracted by applying a 32 Gabor filters at 4 scales and 8 orientations on the grayscale image frame (256×256 pixels). Filtered 32 feature maps outputs are averaged within each 16 regions to produce a 16×32=512-dimensional Gist descriptor. Enhanced-Gist visual descriptors (1024-dimensional visual descriptors) are extracted by combing the extracted Gist and HODMG visual descriptors.

4.5. Indoor Datasets

Dataset-1 have a total of 450 images of 3 indoor scene classes. Specifically, 150 images per class (corridor, staircase and room) are considered; in    Dataset-1 100 and 50 images are used for training and testing the classifier respectively. The 100 training images per class are collected using the internet. For testing images out of 50 images per class, the first 20 images are collected using the internet and the remaining 30 images are the image frames extracted from the transmitted real time video from the Micro Aerial Vehicle.

4.6. SVM classifier

One-Against-All (OAA) SVM [25] method is used in this study. In OAA method, k SVM models will be constructed to compare each indoor class with all other indoor classes of scenes for a k indoor classes. A training dataset, \( \{ x_i, y_i \}_{i=1}^l \) Where \( x_i \in R^n, i = 1, ......, l \) and \( y_i \in \{1, ...., k\} \) represent the class of \( x_i \), then the mth SVM parameters can be determined by solving the following equation:

\[
\min_{w^m, b^m, \xi^m} \frac{1}{2} (w^m)^T w^m + C \sum_{i=1}^l \xi^m \quad (w^m)^T \phi(x_i) + b^m \geq 1 - \xi^m \text{ if } y_i = m, \\
(w^m)^T \phi(x_i) + b^m \leq -1 + \xi^m \text{ if } y_i \neq m, \quad \xi^m \geq 0, i = 1, ......, l
\]

(8)

Where \( \xi^m \) and C denotes the positive slack variable and regularization parameter (penalty value) respectively. By solving Eqn (8), k decision functions are obtained as follows: \( (w^1)^T \phi(x) + b^1, ... \)
Finally, indoor scene class ‘x’ can be determined based on the maximum score of the decision function as follows
\[ x = \arg \max_{m=1,\ldots,k} (w^m)^T \phi(x) + b^m \]  
\[ (9) \]

4.7. KNN classifier
For the kNN classifier [26], the Euclidean distance was selected as the distance metric expressed as follows:
\[ d(r,a) = \sqrt{(r_1-a_1)^2 + (r_2-a_2)^2 + \ldots + (r_n-a_n)^2} \]  
\[ (10) \]
Where \( r \) (r1, r2, ..., rn) and a(a1, a2, ..., an) are n-dimensional vectors.

In the kNN classifier training stage, extracted feature vectors of the training samples and the corresponding class labels were trained. In the classification phase the testing set feature vectors are classified based on the value of k between 1 and 25 and the distance metric function.

5. Experimental Results and Discussions
A robust vanishing point detection algorithm has been developed to estimate the MAV position and navigate in indoor environment. Based on the position of vanishing point the MAV can turn to the left or right to align with the corridor centre. The video stream is obtained from the onboard camera and received in the ground station and the video stream is further processed using MATLAB. The MAV position in the corridor centre has been estimated using the detected vanishing point coordinates in the sequence of real time image frames as shown in Figure 4. Vanishing point is detected in the centre of the corridor image for different image resolutions 480×640, 256×256, 240×480 & 512×512 pixels. Edge detection, line detection splitting of grid into 10×10 grid cells and vanishing point detection are difficult at low image resolutions (30×40, 15×20 pixels). At low image resolutions, the detection of straight lines and vanishing point is more difficult in the corridor image. Proposed Enhanced GIST descriptors is evaluated on Dataset-1. Graphical User Interface Output of real time vanishing point detection based on grid based method is shown in Figure 5. A total of three real time videos (2.4 GHZ, 1.5 to 3m) of indoor scenes (corridor, staircase and room) obtained from parrot AR drone MAV are considered and the image frames of the indoor videos (1280×720 pixels) are included for analysis in Dataset-1. Enhanced GIST descriptors are mainly proposed to recognize three indoor scenes namely, corridor, staircase or room. A ‘linear’ kernel function is used in the SVM classifier. Euclidean distance was selected as the distance metric in the kNN classifier. The classification results for GIST and HODMG and the proposed Enhanced GIST descriptors using SVM and k-NN classifiers for Dataset-1 are presented in Table 1. From Table 1, it can be observed that our proposed enhanced GIST descriptors using SVM classifier gives better classification accuracy compared to the state of the art visual descriptors and KNN classifier. Computational cost of the proposed Enhanced GIST descriptors using SVM and KNN classifiers are 1.91 and 1.18 seconds respectively. Proposed Enhanced GIST visual descriptors are robust and suitable for real time indoor navigation of Micro Aerial Vehicle.
Figure 4. MAV position in corridor (a). Corridor Left (b). Corridor Centre and (c). Corridor Right

Figure 5. Real time Graphical User Interface Output

| Table 1. Classification results of the SVM and KNN classifiers for Dataset-1 |
|-------------------------------|-------------------|-------------------|
| Classifier                | Types of Features | Dataset-1 Accuracy (%) |
| SVM Classifier            | Enhanced GIST     | 99.33              |
|                           | GIST              | 72.66              |
|                           | HODMG             | 45.33              |
| KNN classifier            | Enhanced GIST ( K = 20,21) | 56.00             |
|                           | GIST ( K = 1.2)   | 74.66              |
|                           | HODMG ( K = 20,21) | 56.00             |
6. Conclusions
Grid based vanishing point detection method for the estimation of MAV position in the corridor environment and Enhanced GIST descriptors based indoor scene recognition algorithm for the categorization of indoor scenes (corridor, staircase or room) have been discussed. We have evaluated the three image descriptors, namely GIST descriptors and Histogram of Directional Morphological Gradient features and Enhanced-GIST descriptors on Dataset-1. Enhanced-GIST descriptors produced highest recognition rate for classifying the corridor, staircase and room type of indoor scenes, with a recognition rates of 99.33% for dataset-1 compared to kNN classifiers. Enhanced-GIST descriptors gives better scene recognition results. Experimental results demonstrate the ability of the MAV to navigate in the indoor environment by classifying the indoor scenes into corridor, staircase or room using the proposed Enhanced-GIST descriptors. Proposed grid based vanishing point detection method and Enhanced GIST descriptor is robust for the estimation of the real time MAV position and scene categorization in GPS denied corridor indoor environments.

References
[1] Wang, F, Phang, SK, Ong, JJ, Chen, BM & Lee, TH 2012, ‘Design and Construction Methodology of an Indoor UAV system with Embedded Vision’, Journal of Control and Intelligent Systems, 40(1): 22-32.
[2] Ebrahimpour, R.; Rasoolinezhad, R.; Hajiabolhasani, Z. & Ebrahimi, M. Vanishing point detection in corridors: using Hough transform and K-means clustering. J. Comp. Vis, IET., 2012, 6(1): 40-51.
[3] Lee, J.O.; Kang, T.; Lee, K.H.; Im, S.K. & Park, J. Vision-Based Indoor Localization for Unmanned Aerial Vehicles. J. Aerosp. Eng., 2011, 24(3): 373-377.
[4] Yuxiang, W. An efficient algorithm for UAV indoor pose estimation using vanishing geometry. In Proceedings of the IAPR Conference on Machine Vision Applications, Nara, Japan, June 2011.
[5] Rudol, P. & Wzorek, W. & Doherty, P.Vision-based Pose Estimation for Autonomous Indoor Navigation of micro-scale Unmanned Aircraft Systems. In Proceedings of the IEEE International Conference on Robotics and Automation, Alaska, USA, May 2010.
[6] Zheng, C. & Green, R. Vision-based Autonomous Navigation in Indoor Environments. In Proceedings of the 25th International Conference of Image and Vision Computing, Queenstown, New Zealand, Nov 2010.
[7] Hwangbo, M. & Kanade, T. Visual-Inertial UAV Attitude Estimation Using Urban Scene Regularities. In Proceedings of the IEEE International Conference on Robotics and Automation, Shanghai, China, May 2011.
[8] Fowers, S.; Lee, D.J.; Tippetts, B.J.; Lillywhite, K.D.; Dennis, A.W.; & Archibald, J.K. Vision aided stabilization and the development of a quad-rotor micro UAV. In proceeding of the IEEE international symposium on computational intelligence in robotics and automation, Florida, USA, June 2007.
[9] Kim, Z. Geometry of Vanishing points and its Applications to External Calibration and Real time Pose Estimation. Institute of Transportation Studies, University of California, Berkeley, 2006.12p.
[10] Rother, C. A new approach for Vanishing Point detection in architectural environments. In Proceedings of the 11th British Machine Vision Conference, Bristol, UK, September 2000.
[11] Bills, C.; Chen, J. & Saxena, A. Autonomous MAV flight in Indoor Environments Using Single Image Perspective Cues. In Proceedings of the IEEE International Conference on Robotics and Automation, Shanghai, May2011.
[12] Shen, S.; Michael, N. & Kumar, V. Autonomous Multi-Floor Indoor Navigation with a Computationally Constrained MAV. In Proceedings of the IEEE International Conference on Robotics and Automation, Shanghai, China, May 2011.
[13] Shabayek, A.E.R.; Demonceaux, C.; Morel, O. & Fofi, D. Vision based UAV Attitude Estimation: Progress and Insights. J. Intell. robot. syst., 2012, 65(1-4), 295-308.
[14] Quattoni, A, & Torralba, A 2009, ‘Recognizing indoor scenes’, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1-8.
[15] Szummer, M, & Picard, RW 1998, ‘Indoor-outdoor image classification’, Proceedings of the IEEE International Workshop on Content-Based Access of Image and Video Database, 42-51.
[16] Zhou, B, Lapedriza, A, Xiao, J, Torralba, A & Oliva, A 2014, ‘Learning deep features for scene recognition using Places database’, Proceedings of the twenty-seventh International Conference on Neural Information Processing Systems, 487–495.
[17] Krizhevsky, A, Sutskever, I & Hinton GE 2012, ‘Imagenet classification with deep convolutional neural networks’, Proceedings of the twenty-fifth International Conference on Neural Information Processing Systems, 1106–1114.
[18] Khan, SH, Hayat, M, Bennamoun, M, Togneri, R & Sohel, FA 2016, ‘A Discriminative representation of convolutional features for indoor scene recognition’, IEEE Transactions on Image Processing, 25(7): 3372-3383.
[19] Oliva, A & Torralba, A 2001, ‘Modeling the shape of the scene: a holistic representation of the spatial envelope’, International Journal of Computer Vision, 42(3): 145-175.
[20] Oliva, A & Torralba, A 2006, ‘Building the gist of a scene: the role of global image features in recognition’, Progress in Brain Research, 55: 23-36.
[21] Wu, J & Rehg, JM 2011, ‘CENTRIST: a visual descriptor for scene categorization’, IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(8): 1489-1501.
[22] J. F. Canny, “A computational approach to edge detection,” IEEE Trans. Pattern. Analy. Mach. Intell., 8: 679-698, Nov. 1986.
[23] Duda, RO & Hart, PE 1972, ‘Use of the hough transformation to detect lines and curves in pictures’, Communications of the ACM, 15(1): 11-15.
[24] Söille, P. Morphological Image Analysis: Principles and Applications. Springer-Verlag, 2003.
[25] Cortes, C. & Vladimir, V. Support-Vector Networks. Mach Learn, (1995), 20(3): 273-297.
[26] Cover, T.M. & Hart, P.E. Nearest Neighbor Pattern Classification. IEEE Trans Inform Theory, (1967), 13(1): 21-27.