ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

Multi-Class Support Vector Machine (MSVM) Enabled Indoor Scene Classification With SIFT Features

Japneet Kaur[1], Rimanpal Kaur[2]

Research Scholar, Department of Computer Science, CGC Technical Campus, Jhanjeri, Mohali, India [1]

Assistant Professor, Department of Computer Science, CGC Technical Campus, Jhanjeri, Mohali, India[2]

ABSTRACT: The indoor scene recognition methods require the inclusion of the various methods in the computer vision, image processing and feature recognition for the scene recognition by identifying the category of the input image by comparing it against the given training databases by the means of the feature descriptor (popularly based upon the colour or low level features) and the classification algorithm. The indoor scene classification algorithms require the number of the computations and feature transformations along with the normalization and automatic categorization. In this thesis, the multi-category dataset has been incorporated with the robust feature descriptor using the scale invariant feature transform (SIFT) along with the multi-category enabled support vector machine (mSVM). The multi-category support vector machine (mSVM) has been designed with the iterative phases to make it able to work with the multi-category dataset. The mSVM represents the training samples of main class as the primary class in every iterative phase and all other training samples are categorized as the secondary class for the support vector machine classification. The proposed model is made capable of working with the variations in the indoor scene image dataset, which are noticed in the form of the colour, texture, light, image orientation, occlusion and colour illuminations. Several experiments has been conducted over the proposed model for the performance evaluation of the indoor scene recognition system in the proposed model. The results of the proposed model have been gathered in the form of the various performance parameters of statistical errors, precision, recall, F1-measure and overall accuracy. The proposed model has clearly outperformed the existing models in the terms of the overall accuracy. The proposed model improvement has been recorded higher than ten percent for all of the evaluated parameters against the existing models based upon SURF, FREAK, etc.

KEYWORDS: Adaptive Multi-class Classification, Support Vector Machine, Feature Description, Indoor Scene Recognition.

I. INTRODUCTION

Scene classification is aimed at labeling an image into semantic categories (room, office, mountain etc.). It is an important task to classify [1][4], organize and understand thousands of images efficiently. From application point of view, scene classification is useful in-content based image retrieval. As accurate classification of an image [3], as better as it helps in better organization and browsing of the image data. Scene classification is highly valuable in remote navigation also.

Indoor scenes are cluttered with number of objects. So classification techniques simply based on color, texture and intensity are not very effective to classify indoor scenes [2]. Pioneering works used SIFT, SURF etc. in combination with supervised learning. But these techniques fail to distinguish many indoor scenes. One way to bridge semantic gap between image representation and image recognition is to make use of more and more sophisticated models[13][15], but good learning and inference is extremely difficult task for such models. Alternatively semantic gap between low-level features like color, intensity, texture etc. and high-level category label can be reduced by introducing object-based representation as intermediate representation[11]. As the performance of scene recognition is heavily dependent on feature representation, this object-based intermediate representation [14] proves to be useful in enhancing classification results. Recently objects-based techniques for indoor scene classification have proven to be showing promising performance over other state-of-art techniques.

DOI: 10.15680/IJIRCCE.2016. 0405307

In this work, we will review the recent and significant techniques that have been used for indoor scene classification. Besides we will identify the key approaches being used in indoor scene classification. The major contributions made by each significant work and the challenges posed to efficient classification will also be discussed.

Feature Descriptors

- **GIST-** A typical GIST is computer over a complete image for the scene classification task. It falls in the global image descriptor category.
- **SIFT-** The typical use of SIFT is to match the local regions in two images on the basis of their reconstruction, alignment or other similar. SIFT can be used for the purpose of identification of some specific objects by using BW (Bag of Words) model.
- **HOG-** Histogram of Oriented Gradient (HOG) [21] is used for object-recognition. It is based on computing edge- gradients. Typical HOG works in the sliding window fashion for object detection applications. HOG computes the complete image after dividing it into the smaller cells, called blocks. HOG can be used alongside SVM for feature detection using classification.
- **CENTRIST-** Census Transform Histogram (CENTRIST) is a novel visual descriptor, which is more robust to illumination changes, gamma variation etc. as compared to GIST [22] and SIFT [23]. CENTRIST is histogram of Census Transform (CT) values. CT compares intensity value of a pixel with its neighboring pixels and assigns value 1 or 0 to those pixels. After that the decimal number corresponding to this sequence of 8 neighboring binary digits is computed and used as CT value of center pixel. This descriptor retains the local as well as global structure of the scene. However, there are several limitations of this descriptor. It is not invariant to rotation and scale changes. It also does not consider color information. Further it cannot be used for precise shape description.

Feature Representation Techniques

- **Bag of words-** Bag of words involves four steps:1) Detect point/region of interest 2) Compute local descriptor 3) Quantize local descriptors into words to form visual vocabulary 4) Find occurrences of these words in image for constructing BoW features[17]. In first step, image features are detected by interest point detectors. These interest point detectors are used to detect distinctive features of image. Amongst the most popular detectors are Harris Laplace Detector [24], which is used to detect corner like structures. Difference of Gaussian (DoG) is also a well-known detector to detect blob-like structures. It is not only faster but also compact. Hessian-Laplace has also been used in some of the previous works for scene classification. Another category of works are based on regular grids for feature detection. Image is partitioned into regular grids, then either dense or sparse features [29] are computed over these grids. The next step is to compute local descriptors over interest points or regular grids. SIFT (scale invariant feature transform) is popular local descriptor. It is invariant to scale, space and orientation. PCA (principal component analysis) is often used in combination with SIFT to reduce the dimensionality. DOG interest point detector plus SIFT feature descriptor [13] has proven to be good choice for scene classification techniques. The third step i.e. vector quantization is the most expensive step in this whole process of image representation. K-means clustering [16] is often used to quantize the feature vectors to form words.
- **Spatial pyramid matching-** Spatial pyramid matching is another well-known representation based on incorporating spatial lay-out of feature by first partitioning the image into increasingly fine fixed sub-regions and then computing histogram of local-features on these local sub-regions. Basically, it is extension of BoW representation [17]. The local features are computed over each fixed sub-region. The geometric correspondence between images is computed by using pyramid matching scheme [25]. More specifically, an image is partitioned into regular grids at resolutions 0,1…L such that there are 2^l cells in the grid at level l. Histogram intersection is used for finding number of matches between cells. Using clustering technique the feature vectors are quantized into M discrete types. The dimensionality of feature vector for L levels and M channels is $M\sum 4^l = M*1/3*(4^{l+1}-1)$.

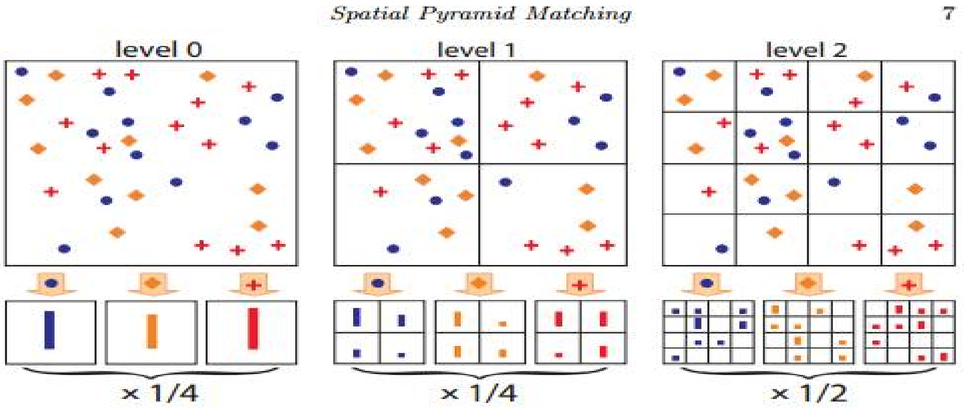


Fig 1.1: Spatial Pyramid Matching at Level 0, Level 1 & Level

- **Object bank representation-** Object bank consists of a number of object detectors. Due to availability of a large number of detectors, it seems reasonable to use objects representation for scene classification. In [26], the latent SVM object detectors and texture classifier have been used for blobby objects and texture-based objects respectively. For a given image several object detectors run on image. The image representation can be viewed as response of "generalized object convolution". The dimensionality of representation is a point to be considered. For O object detectors at S detection scales and L spatial pyramid levels, the dimensionality is O*S*L. In original work [26], 200 object detectors have been used at 12 detection scales and 3 spatial pyramids. But this representation also works well with even modest number of object-detectors. Object-bank representation is useful representation in scenes cluttered with many objects, where GIST [22] and SPM [27] fail to distinguish scenes.

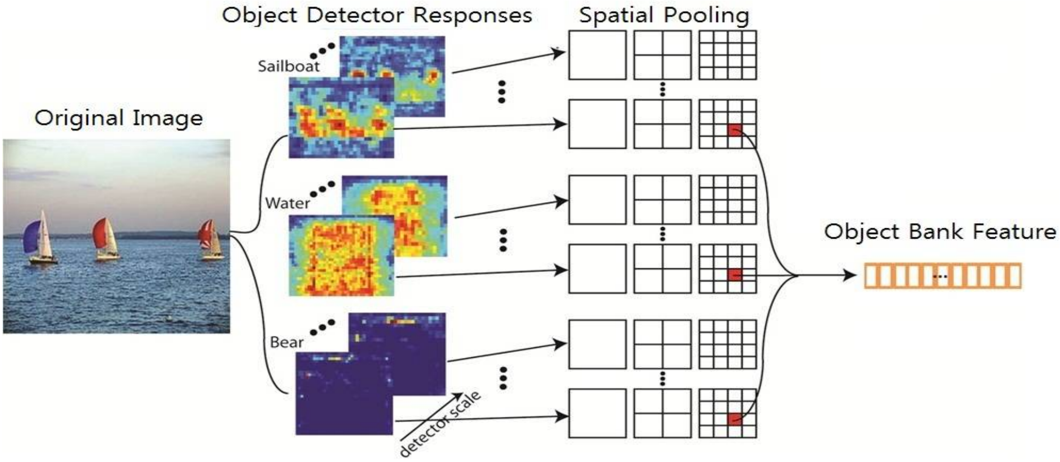


Fig 1.2: Object Detector responses

Classifiers

- **Support Vector Machine:** Support Vector machine (SVM) is a supervised learning basedClassifier. Support vector machines were proposed by Boser et al. in [12]. SVM is supervised machine learning approach specifically designed for pattern matching. SVMs construct a set of hyper-planes that separates the data points into two classes with maximal margin in high dimensional feature space. Mathematically, SVM learns a mapping $\chi \rightarrow \Upsilon$ where x € χ represents the feature vector and y € Υ represents scene category. The objective is to learn a function defined below, with function parameter α, given a set of training images (x1, y1), (x2, y2), (xn,yn)
  Y= f(x, a)

DOI: 10.15680/IJIRCCE.2016. 0405307

Once the mapping function is learnt, the classifier can render a category label for unobserved feature vector. According to the linear separability of data-points, SVM[20] can be linear or non-linear. Non-linear SVM is an extension of linear SVM. It maps the non-linearly separable points to higher dimensional plane in which these can be linearly separable. Linear SVM- In case of non-linear SVM, the hyper plane is defined as given below

$x_i.w+b\geq +1$, if $y_i=+1$

$x_i .w+b\leq+1$, if $y_i= -1$

Where x_i denotes observed data points, w is normal vector, b is offset of hyper plane w.r.t. to origin and $y_i$is target value. Supports vectors are those data points that lie exactly on hyper plane and satisfy following equations:

$x_i .w+b= +1$

$x_i .w+b= -1$

This linear SVM is used for binary classification. For multiclass classification multiple one-vs-all linear SVM can be used

II. RELATED WORK

In [1] authors have proposed an approach based on Object-Centric spatial Pooling (OCP). They have proved that if location of object of interest is known then foreground and background features can be pooled separately. This OCP representation has outperformed traditional Spatial Pyramid Matching (SPM). The only challenge is localization of objects. They have also remarked that good image understanding requires to figure out what objects are in image as well as where they are in image. In [2] authors have jointly worked upon to investigate the contribution of objects to scene classification. This paper discusses the usefulness of image representation based on objects in providing complementary information to the low-level features. In this work a large number of 'object filters' have been used. They have used object-bank of 200 pre-trained SVM object detector at 12 different detection scales and 3 spatial pyramid levels. Object bank representation yields promising results by capturing important information. However, limitation of this lies in computation speed and high dimensionality of object-bank. The training of a large number of object detectors is also difficult. They further enhanced performance by replacing SVM classifier with logistic regression classifier. They explore feature sparsity and object sparsity for dimensionality reduction. In [3] authors have discussed in detail the importance of each component like different detectors, spatial location, and scale classification models etc. for efficient scene recognition. They have provided guidelines for using high-level object representation for scene classification. In [4] the work is done on Indoor scene recognition by a mobile robot via adaptive object detection. In this paper, authors have used common objects, such as screen or furniture, as a key intermediate representation to recognize indoor scenes. Authors have used object category classifiers to match low level visual features to objects, and contextual relations to associate objects to scenes. The main contributions made by this paper were use of probabilistic generative model and 3D range sensor for incorporating geometric and structural information. For building scene detector they have utilized category level object detection and classifier confidence. Object detection has been done on the gray-scale feature, Gabor features and HOG-features. The sliding window approach has been used to avoid segmentation. Further they have applied adaptive scheme for guiding informative search of object.

III EXPERIMENTAL DESIGN

Final Indoor Scene Recognition Algorithm (FISRA)

The proposed model uses the scale invariant feature transform features (SIFT) points for the indoor scene detection in the forged images. The proposed model finds the SIFT points in the both of the images by using the traditional SIFT point algorithm. The SIFT points are the pointsdepictedas the strong visual points in the given image. The SIFT points are obtained in the form of point location, point width, point height and point descriptor vector. The point locations and point descriptors are evaluated for the comparative analysis by comparing the SIFT Points using the robust classifier. The proposed model has been designed with the Chi square distance based vector classification using the support vector machine (SVM) based classification between the SIFT points are gathered from the images in the processing. The proposed model uses the vectorized key point vectors and point locations to find the difference between the two points. The matching points are analysed in the first stage. Then the non-matching points in the first image (Original image) are shown as the points of removal and non-matching points in the indoor scene image data are analysis to compute the final decision on the indoor scene recognition.

Algorithm 1: SIFT Based indoor scene recognition

```
1.  Load the actual picture
2.  Get the size of original picture
3.  If actual picture depth is 3
    a.  Convert the picture to grayscale
4.  Apply SIFT over the actual picture
5.  Get the SIFT point data from the actual picture
6.  Load the input picture
7.  Get the size of input picture
8.  If suspected picture depth is 3
    aConvert the picture to grayscale
9.  Apply SIFT over the input picture
10. Get the SIFT point data from the input query image
11. Evaluate the point locations between the point location arrays obtained from actual and suspected image
12. Find the matching locations
13. Evaluate the point descriptors on the particular locations
14. Return the matching points
15. Return decision logic containing the detection scene category
```

SURF based Indoor Scene Recognition Algorithm

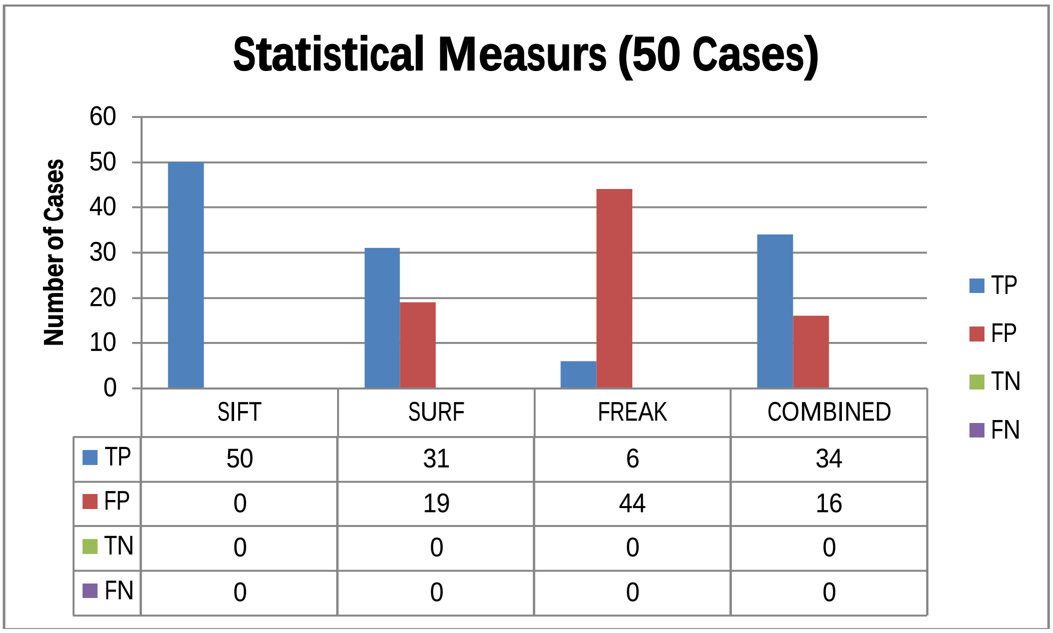
To mitigate the problems of the existing system, the new model has been proposed for the use of pixel based image region localization and neighbour pixel and neighbour region analysis to find the abnormalities in the pixel based pattern in order to detect the indoor scene along with speeded-up robust features (SURF) and SVM. The SVM algorithm can be used to match and find the matching pixel groups based upon the region localization. The following algorithm has been applied over the image data for the indoor scene recognition under the proposed model category:

Algorithm 2: SURF points based image forgery detection

```
1.  Load the actual image
2.  Get the size of actual image
3.  If original image depth is 3
        a.  Convert the image to grayscale
4.  Apply SURF over the original image
5.  Get the SURF data matrix from the SURF function
6.  Load the input image
7.  Get the size of input image
8.  If suspected image depth is 3
        a.  Convert the image to grayscale
9.  Apply SURF over the input image
10. Get the SURF data matrix from the SURF function
11. Evaluate the point locations between the point location arrays obtained from actual and suspected image using
    the SVM algorithm.
12. The Euclidean distance is calculated using the SVM algorithm
13. Evaluate the point descriptors on the location where it matches.
14. Return the matching points
15. Return the decision logic containing the detected category of indoor scenes
```

IV. RESULT ANALYSIS

The First experiment has been conducted over the 50 test samples, which has been randomly selected out of the given image set. The randomizer module generates the random index containing the fifty image ids, which are acquired from the given dataset. Such randomly chosen samples are further processed and analysed under the proposed model for the result evaluation.



	SIFT	SURF	FREAK	COMBINED
🟦 TP	50	31	6	34
🟥 FP	0	19	44	16
🟩 TN	0	0	0	0
🟪 FN	0	0	0	0

Figure 4.1: Type 1 and type 2 errors for 50 test cases

The figure 4.1 shows the statistical errors of type 1 and type 2 from the fifty random samples obtained from the dataset. The variation can be clearly observed from the above bar graph in the figure 4.1. The proposed model has been found with the maximum true positive cases, which shows the robustness of the proposed model. The figure 4.1 clearly defines the samples classified in the four major classes of true positive, true negative, false positive and false negative errors, which are primarily classified as the statistical errors. These errors are the definitive parameters to calculate the overall accuracy, precision, recall and f1-measure to understand the performance of the proposed model from various aspects.

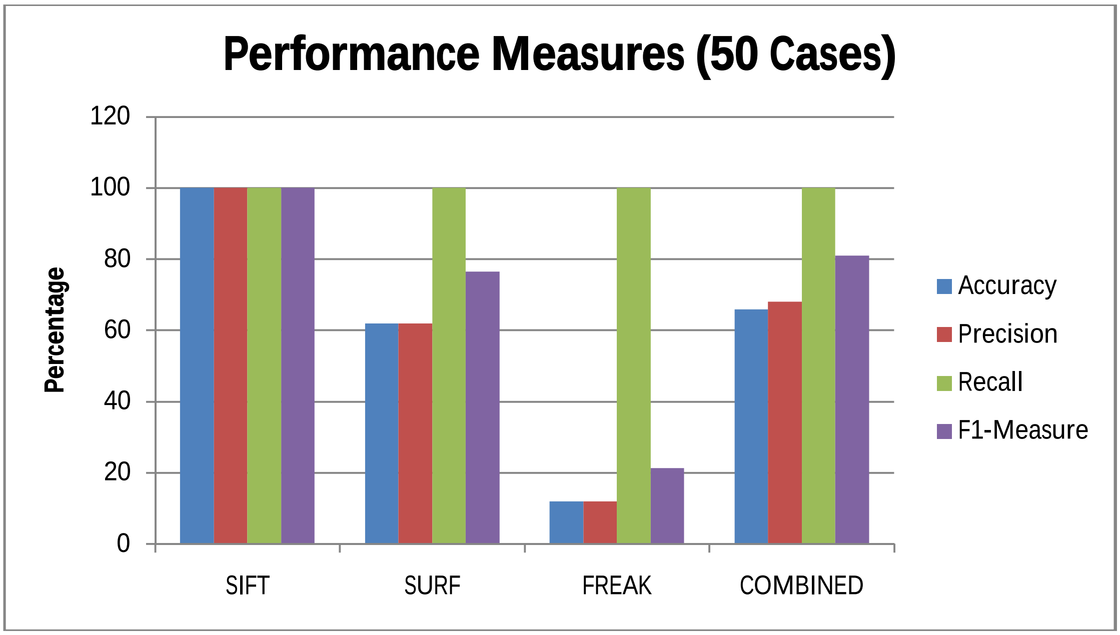
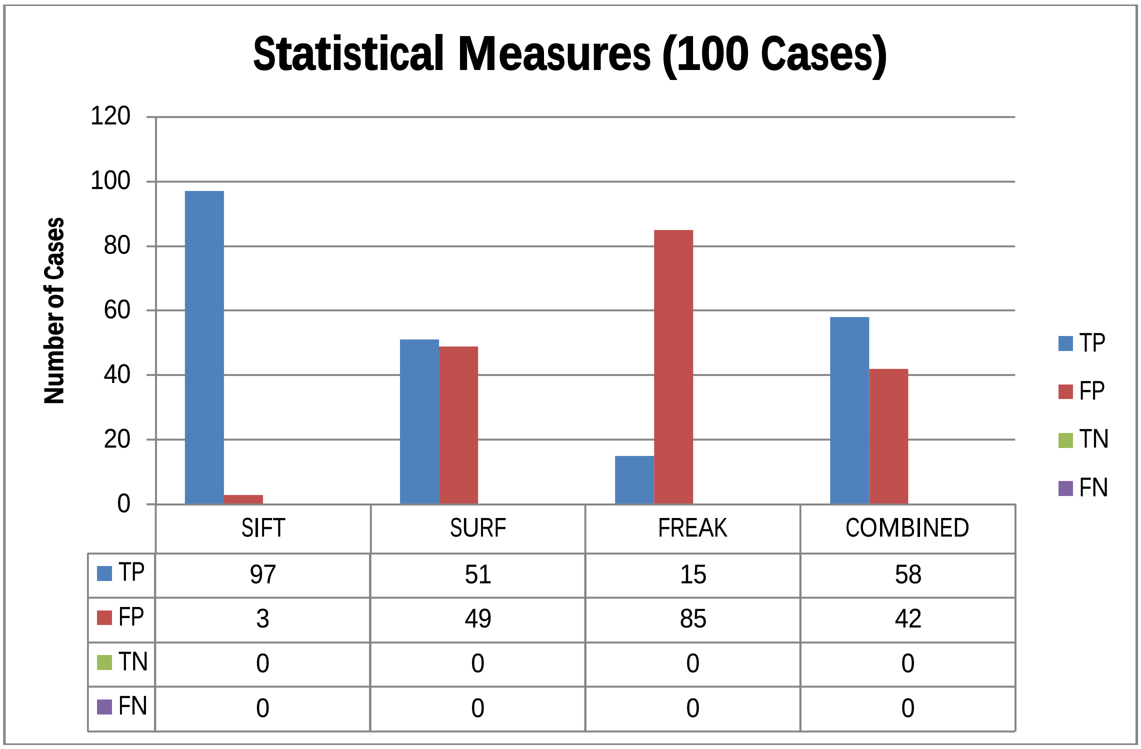


Figure 4.2: Performance measures for 50 test cases

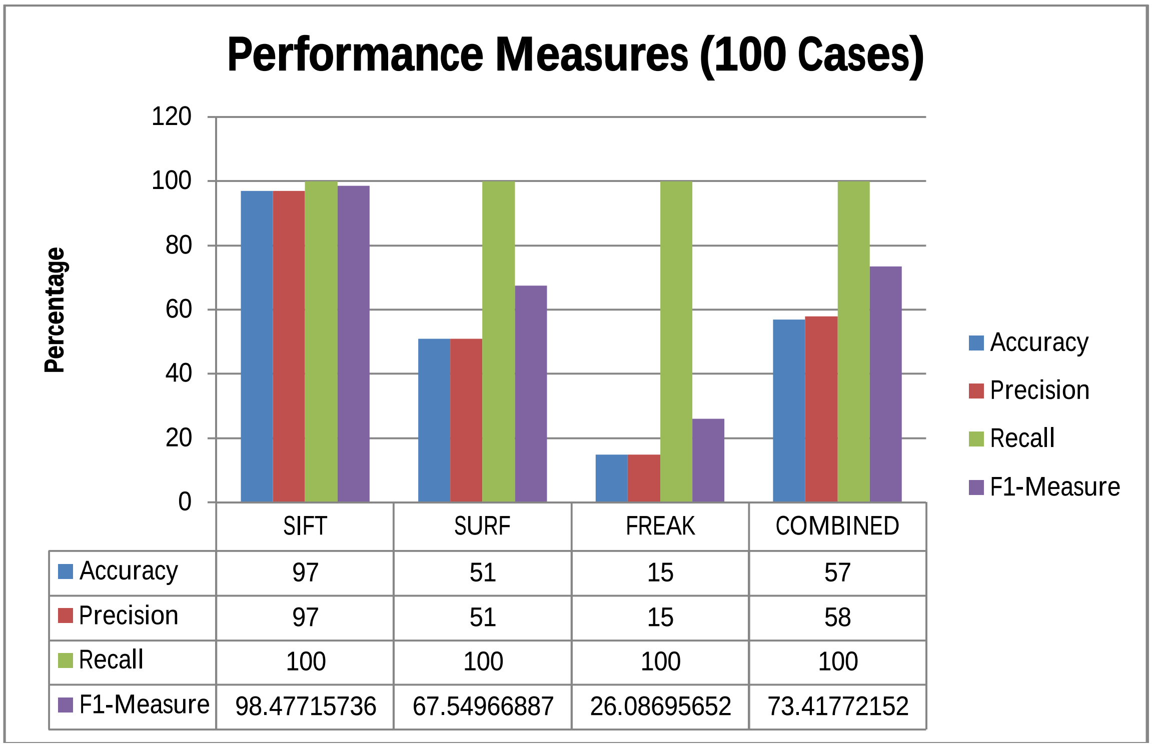
The figure 4.2 shows the performance measures calculated over the obtained 50 samples. The proposed model and other models based on SIFT, SURF and FREAK has been recorded with the variable measured values for all of the performance measures. The first experiment has been conducted over the 100 test samples, which has been randomly selected out of the given image set. The randomizer module generates the random index containing the hundred image ids, which are acquired from the given dataset. Such randomly selected samples are further processed and analyzed under the proposed model for the result evaluation.



	SIFT	SURF	FREAK	COMBINED
🟦 TP	97	51	15	58
🟥 FP	3	49	85	42
🟩 TN	0	0	0	0
🟪 FN	0	0	0	0

Figure 4.3: Type 1 and type 2 errors for 100 test cases

The type 1 and type 2 errors has been evaluated from the testing of the input test samples. The graph obtained from the values of the statistical type 1 and 2 errors has been presented in the figure 4.3. The figure 4.3 has been obtained from the hundred testing samples and all of the samples show the equal statistical errors from the first evaluation.



	SIFT	SURF	FREAK	COMBINED
🟦 Accuracy	97	51	15	57
🟥 Precision	97	51	15	58
🟩 Recall	100	100	100	100
🟪 F1-Measure	98.47715736	67.54966887	26.08695652	73.41772152

Figure 4.4: Performance measures for 100 test cases

The figure 4.4 shows the performance measures calculated over the obtained 100 samples. The proposed model and other models based on SIFT, SURF and FREAK has been recorded with the variable performance measure over the given statistical errors.

V. Conclusion and Future Work

The proposed model has been designed by using the scale invariant feature transform (SIFT) with the support vector machine (SVM) for the multi-class classification in the adaptive manner. The proposed multi-class enabled support vector machine (mSVM) has been developed with the adaptive classification in the iterative manner. One primary class is compared against all other classes signified the secondary class for the realization of the support vector machine (mSVM). In this thesis, the scale invariant feature transform (SIFT) along with the multi-class support vector machine (mSVM) has been proposed for the indoor scene recognition. The probabilistic classification with the multi-class support vector machine has been utilized for the robustness in the classification. The proposed model has undergone several experiments and has been found better than the previous option in the terms of precision, recall, f1-measure and overall accuracy and SIFT based model has outperformed the other descriptors with mSVM classification.

Future Scope

In the future the proposed model can be improved by using the hybrid low level feature extracted along with the efficient color illumination to find the dual-mode attacks over the images to determine the indoor scene in the image data. Also the swarm intelligent algorithm can be utilized for the indoor scene recognition in the digital image dataset.

References

1. Alberti, Marina, John Folkesson, and PatricJensfelt. "Relational approaches for joint object classification and scene similarity measurement in indoor environments." In *AAAI 2014 Spring Symposia: Qualitative Representations for Robots*. 2014.
2. Zhang, Lei, Xiantong Zhen, and Ling Shao. "Learning object-to-class kernels for scene classification." *Image Processing, IEEE Transactions on* 23, no. 8 (2014): 3241-3253.

3. Antanas, Laura, M. Hoffmann, Paolo Frasconi, TinneTuytelaars, and Luc De Raedt. "A relational kernel-based approach to scene classification." InApplications of Computer Vision (WACV), 2013 IEEE Workshop on, pp. 133-139. IEEE, 2013.
4. Antanas, Laura, Marco Hoffmann, Paolo Frasconi, TinneTuytelaars, and Luc De Raedt. "A relational kernel-based approach to scene classification." In*Applications of Computer Vision (WACV), 2013 IEEE Workshop on*, pp. 133-139. IEEE, 2013.
5. Espinace, Pablo, Thomas Kollar, Nicholas Roy, and Alvaro Soto. "Indoor scene recognition by a mobile robot through adaptive object detection." Robotics and Autonomous Systems 61, no. 9 (2013): 932-947.
6. Giannoulis, Dimitrios, Dan Stowell, EmmanouilBenetos, Mathias Rossignol, Mathieu Lagrange, and Mark D. Plumbley. "A database and challenge for acoustic scene classification and event detection." In Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European, pp. 1-5. IEEE, 2013.
7. Gupta, Saurabh, Pablo Arbelaez, and Jitendra Malik. "Perceptual organization and recognition of indoor scenes from rgb-d images." In Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, pp. 564-571. IEEE, 2013.
8. Juneja, Mayank, Andrea Vedaldi, C. V. Jawahar, and Andrew Zisserman. "Blocks that shout: Distinctive parts for scene classification." In Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, pp. 923-930. IEEE, 2013.
9. Mesnil, Grégoire, Salah Rifai, Antoine Bordes, Xavier Glorot, YoshuaBengio, and Pascal Vincent. "Unsupervised and Transfer Learning under Uncertainty-From Object Detections to Scene Categorization." In *ICPRAM*, pp. 345-354. 2013.
10. Li, Li-Jia, Hao Su, Yongwhan Lim, and Li Fei-Fei. "Objects as attributes for scene classification." In *Trends and Topics in Computer Vision*, pp. 57-69. Springer Berlin Heidelberg, 2012.
11. Russakovsky, Olga, Yuanqing Lin, Kai Yu, and Li Fei-Fei. "Object-centric spatial pooling for image classification." In *Computer Vision–ECCV 2012*, pp. 1-15. Springer Berlin Heidelberg, 2012.
12. A. Pronobis, Semantic mapping with mobile robots, Ph.D. Thesis, Royal Institute of Technology, Stockholm, Sweden, 2011.
13. A. Pronobis, O. Mozos, B. Caputo, P. Jens-felt, Multi-modal semantic place classification, The International Journal of Robotics Research 29 (2–3) (2010) 298–320.
14. Li, Li-Jia, Hao Su, Li Fei-Fei, and Eric P. Xing. "Object bank: A high-level image representation for scene classification & semantic feature sparsification." In *Advances in neural information processing systems*, pp. 1378-1386. 2010.
15. L.-J. Li, H. Su, E. Xing, L. Fei-Fei, Object bank: a high-level image representation for scene classification and semantic feature sparsification, in: Neural Information Processing Systems, 2010.
16. P. Espinace, T. Kollar, A. Soto, N. Roy, Indoor scene recognition through object detection, in: IEEE International Conference on Robotics and Automation, 2010.
17. P. Viswanathan, T. Southey, J. Little, A. Mackworth, Automated place classification using object detection, in: Canadian Conference on Computer and Robot Vision, 2010.
18. P. Felzenszwalb, R. Girshick, D. McAllester, Cascade object detection with deformable part models, in: IEEE International Conference on Computer Vision and Pattern Recognition, 2010.
19. S. Helmer, D. Lowe, Using stereo for object recognition, in: IEEE International Conference on Robotics and Automation, 2010.
20. I. Posner, M. Cummins, P. Newman, A generative framework for fast urban labeling using spatial and temporal context, Autonomous Robots 26 (2–3) (2009) 153–170.
21. P. Dollar, Z. Tu, P. Perona, S. Belongie, Integral channel features, in: British Machine Vision Conference, 2009.
22. T. Kollar, N. Roy, Utilizing object–object and object–scene context when planning to find things, in: International Conference on Robotics and Automation, 2009.
23. Quattoni, Ariadna, and Antonio Torralba. "Recognizing indoor scenes." In*Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 413-420. IEEE, 2009.
24. B. Douillard, D. Fox, F. Ramos, Laser and vision based outdoor object mapping, in: Robotics: Science and Systems, 2008.
25. B. Russell, A. Torralba, K. Murphy, K. Freeman, Labelme: a database and webbased tool for image annotation, International Journal of Computer Vision 77 (1–3) (2008) 157–173.
26. M. Cummins, P. Newman, FAB-MAP: probabilistic localization and mapping in the space of appearance, International Journal of Robotics Research 27 (6) (2008) 647–665.
27. A. Bosch, X. Muñoz, R. Martí, A review: which is the best way to organize/ classify images by content? Image and Vision Computing 25 (2007) 778–791.
28. A. Bosch, A. Zisserman, X. Muñoz, Image classification using random forests and ferns, in: IEEE International Conference on Computer Vision, 2007.
29. C. Siagian, L. Itti, Rapid biologically-inspired scene classification using features shared with visual attention, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (2) (2007) 300–312.
30. E. Brunskill, T. Kollar, N. Roy, Topological mapping using spectral clustering and classification, in: Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, 2007, pp. 3491–3496.

BIOGRAPHY

Japneet Kaur is pursuing M.Tech in Computer Science from CGC Technical Campus, Jhanjeri, Mohali (affiliated to PTU)**. S**hereceived the B.Tech degree from Punjab Technical University, Punjab, India in 2014.

Ms. Rimanpal Kaur is Assistant Professor at Department of Computer Science, CGC Technical Campus, Jhanjeri, Mohali. She received her B.Tech degree from bbsbec Fategarh Sahib and M.Tech from PEC University of Technology