

Recognize Various Scenes using Classification methods in Million Images Dataset: A Review

Priya Singla
ME Student (ECE)
NITTTR Chandigarh, India,

Rajesh Mehra
Professor (ECE)
NITTTR Chandigarh, India.

Abstract: - The scene recognition provides the visual data from the level of the objects and the relation between the objects. The aim of the scene recognition is to reduce the semantic gap between the computer system and social life. In the form of the example, one can determine the contexts of an input image and classify them in to scenes such as forest, buildings seashores. Some of the issues of the scene recognition is forest, coast, tall buildings highways and street. The applications of the scene recognition are object detection , object recognition, text and video detection, matching human perception, robotics and graphics. This paper surveys on different techniques of scene recognition from single image to multi-million images. Scene recognition performs a task for context for the object recognition. The availability of the large datasets like Image Net leads to multi-million-item dataset initiatives. Scene Recognition is one of the successful methods of the recognition of the images. The variations in color, size of the image are often affected by complex backgrounds, different lighting conditions, shadows and reflections of the image. So solution to this problem is to enhance the accuracy of scene image recognition. This leads to a variety of applications such as understanding the scene, automatic localisation of the image.

Keywords: -ImageNet, multimillion images, Scene recognition.

I. INTRODUCTION

The object and texture are distinguished by scene and the distance between observer and the fixated zone as the discriminating factor. for the scene. An object is about 1 to 2 meters from observer but in case of scene it is more than 5 meter between the observer and the fixed point. A scene is mainly a place in which one can move. The information functions of the environment recognition scene using computational model of scene category. The bottom up approach is used to decide the category of the scene[20]. The segmentation of objects is done to Categorise the scene. The overview of the scene is analysed by object and the texture to accumulate information of the scene less than 200m. In form of example , the information for recognition of the image given by 4 to 8 cycles per image where the shape and density of the objects could not be recovered. In some research people are totally blind to changes of objects when there is part of the scene[11].

Some issues in Scene Recognition:-

- Highway
- Forest
- 3. Coast
- 4. Inside City
- 5. Street
- 6. Open Country
- 7. Mountain
- 8. Tall Buildings

The standard image class has been used extensively and Oliva Toralba dataset is used which is a subset of the Corel database. The procedure of manipulation and analysis of reputation in way to enhance the quality of image is called as digital image processing. The principle of image processing is to improve the difficult scene.

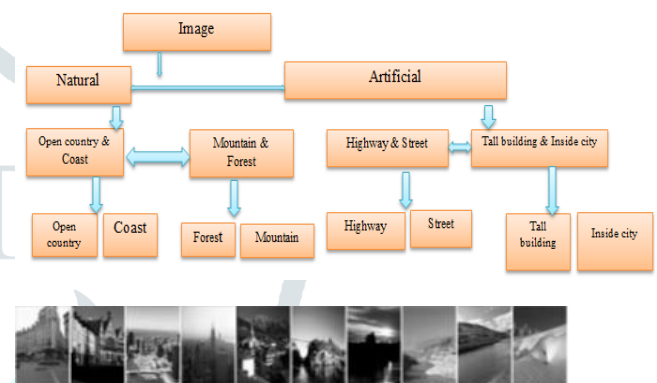


Fig. 1 Scene Recognition[12]

Applications of scene recognition are:-

- Object Detection
- Object Recognition
- Text and video detection
- Human Perception Match
- Environment understanding
- Robotics/ Video and Image
- Graphics (In-painting).

Challenges in Scene Recognition

The process of the scene recognition overcomes the all negative points of the image processing. The allowing of the context for determining the objects is the task task of computer vision in Scene Recognition. The large number of the datasets such as Image Net gives the formulation of the Convolutional Neural Networks[15].

The main challenges in scene detection and recognition can be categorized into three types :-

- i) Diversity of the scene images:- The document images in regular format , color, size and natural scenes taken in the different color scales, orientations and scenes.
- ii) Background Complexity:- The form in the complex set of the structure in videos and the natural scene images. So the confusions and errors are virtually scene such as elements are bricks , grasses and fences.
- iii) Intervention factors:- The main factors are noise, distortion, instance, blur, low resolution, illumination and partial occlusion leads to failures in scene detection and recognition.

II. BACKGROUND

The lot of survey papers in scene recognition had done from past few years, a question arises that, 'Is it possible to recognize the context of a scene without recognizing the objects? A lot of reasons for scene recognition instead of using object recognition [21]The low level feature used for recognition of scene. The low-level features in the image used successfully to infer its semantic context. Along with that many features can be extracted from the image for the purpose of scene recognition, texture, orientation, and color have been used extensively in literature, implemented with different data sets and with different classifiers[17].

Instead of global features the traditional bottom up approach is used for scene recognition by segmentation of image. The concept of recognition of scenes on global geometric concept claims a high recognition rate [18].The image recognizes at different scales works on the classification of images. However the detailed explanation on power spectra of images differentiating images into 8 semantic categories. In this paper, the complexity of the same features of the image textures has been improved [19].

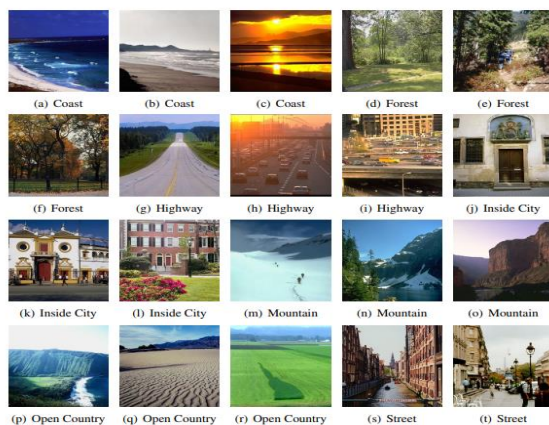


Fig. 2 Dataset Image in Scene

III. RELATED WORK

The recognition techniques build a semantic representation for reducing the semantic gap. The face recognition methods works on extracting the features from single resolution image. The redundant features of the single image resolution may reduce the accuracy of scene recognition.

Chen Cheng et al.,2018[1]proposed a research on military image scene image recognition by using CNN(Convolutional Neural Networks) algorithm. In this research there was less accuracy in military image scene recognition. The deep learning methods from CNN, s has common deep learning attributes of images from the image data. Convolutional Neural Network works best in field of image recognition. The military image training was limited so in order to improve the recognition effect of military images, a new image scene recognition algorithm based on CNN was proposed.**Teddy Mantoro et al.,2018[2]**proposed a research on face recognition process using hybrid process of Haar Cascades and Eigenface methods which can detect multiple faces at a time Face recognition had been an authentic process in various fields. . There was problem in face recognition using biometric identification with accurate results. In this research, different approaches were used to recognise the face image. Firstly, Haar Cascades method used to detect the input image face on the basis of feature extraction and then to detect the characteristics value Sobel edge detection was used.**Bolei**

Zhou et al.,2018[3] proposed a research on repository 10 million image scene using the Convolutional Neural Networks approaches. The visualisation of the CNN shows the object detectors for the classification of the scene image. In order to get the high coverage and diversity, the places .Place Database along with the Places-CNNs offer a resource to for the problems of scene recognition. The multimillion datasets have machine learning algorithm to identify the objects and scenes images.**Zhou Li et al.,2013[4]**proposed a research on scene recognition based on visual image model. The large sized codebook hashigh computational costs and memory consumption.So in order to overcome the weakness a new supervised approach was proposed .The approach based on multiple image resolution and extracting the structural and textural features of the image. In this research, weighted histogram of gradient descriptor was described to identify various scene categories.**Luis Herra and, Shuqiang Jiang et al., 2016[5]** studied on scene recognition in convolutional neural network. The different scale ranges used in same network in different scales. The objects in the scenes recognise specific kind of scales. By breaking down the reaction of ImageNet-CNNs and Spots CNNs at various scales .They allocate that both work in various scale ranges, so utilizing a similar system for every one of the scales actuates dataset predisposition bringing about restricted execution. In this manner, adjusting the component extractor to every specific scale (for example scale-explicit CNNs) is pivotal to improve acknowledgment, since the items in the scenes have their particular scope of scales**Atsumi, M.et al., 2012[6]** propose a research on the scene recognition based on learning and tree features.This paper proposes an item and scene arrangement strategy dependent on the probabilistic inert part tree with supported highlights. In this strategy, object classes are right off the bat acquired by grouping a lot of item portions extricated from scene pictures in every scene classification through the probabilistic idle part examination with the variable number of classes. At that point the probabilistic dormant segment tree with helped highlights at its branch hubs is created as an arrangement tree of all the item classes of all the scene classifications pursued by naming article classes. Ultimately, every scene classification is described by the structure of its named item classes. Item and scene acknowledgment is at the same time performed dependent on the probabilistic inactive part tree seek by utilizing composite helped highlights for the tree traversal. Through tests by utilizing pictures of plural classifications in a picture database, it was demonstrated that execution of item and scene acknowledgment is high and improved by utilizing composite helped. **Xiaodong Yu et al., 2012[7]** studied on active recognition based on vision and language.This paper displays a novel way to deal with using high level learning for the issue of scene acknowledgment in an dynamic vision system, which we call dynamic scene acknowledgment. In customary methodologies, abnormal state learning is utilized in the post-handling to consolidate the yields of the article locators to accomplish better grouping execution. Conversely, the proposed methodology utilizes high level information effectively by actualizing cooperation between a thinking module and a tangible module.**Xiaodong Yu et al., 2011[8]** proposed a novel technique for scene acknowledgment utilizing video pictures through the examination of human exercises. The perceiving three sorts of things, for example, human exercises, items and condition. In the past technique, areas and introductions of articles are assessed utilizing shape models, which are regularly professed to be reliant upon individual scene. Rather than shape models, they utilize applied information about capacity and additionally use of articles just as that about human activities. In thistechnique,

the area and utilization of items can be distinguished by watching collaboration of human. **Tran H.M. et al., 2017[9]** proposed a research on identification of different scene. In the traffic surveillance system (TSS), there are numerous variables influence the characteristics of the outcome. Through down to earth application, it is hard to figure out which scene changing amid the day time frame, from the light to evening time, the transformation of the bright and cloudy, wet and dry scene. Be that as it may, there have been no controlled examinations which represent the strategy to recognize condition scene, which is the one of six principle challenges in TSS. Accordingly, this paper shows the technique to identify and perceive the difference in scene amid throughout the day reconnaissance; Subsequently, TSS embraces the acknowledgment to decide the suitable strategy for every scene, for expanding execution. Our acknowledgment demonstrate depends on the mix of the CIE-

Lab shading space and the histogram of the area of-intrigue (return on initial capital investment) in each casing, which utilized for extricating the component for the Feed Forward Neural System to play out the recognition. In the investigation segment, our outcomes demonstrate that the advantages of our proposed strategy in reality traffic reconnaissance framework **K. E. A. van de Sande et al., 2010[10]** proposed research on shape descriptors of scene recognition images. In this paper, considers the invariance properties and the uniqueness of shading descriptors structured. The scientific invariance properties of shading descriptors are investigated, utilizing a scientific classification dependent on invariance properties as for photometric changes, and tried tentatively utilizing a dataset with realized brightening conditions. Moreover, the uniqueness of shading descriptors is evaluated tentatively utilizing two benchmarks, one from the picture space and one from the video area.

TABLE 1:- SURVEY OF THE VARIOUS TECHNIQUES AND LIMITATIONS

AUTHOR	RESEARCH DESCRIPTION	LIMITATION
Chen Cheng et al.,2018[1]	proposed a research on military image scene image recognition by using CNN(Convolutional Neural Networks) algorithm	less accuracy in military image scene recognition.
Teddy Mantoro et al.,2018[2]	a research on face recognition process using hybrid process of Haar Cascades and Eigenface method	There was problem in face recognition using biometric identification with accurate results. In this research,
Bolei Zhou et al.,2018[3]	proposed a research on repository 10 million image scene using the Convolutional Neural Networks approaches	Place Database along with the Places-CNNs offers a resource to for the problems of scene recognition.
Zhou Li et al.,2013[4]	It based on visual image model, a new supervised approach was proposed	The large sized codebook has high computational costs and memory consumption
Luis Herra et al., 2016[5]	The different scale ranges used in same network in different scales	adjusting the component extractor to every specific scale
Atsumi, M.et al., 2012[6]	an item and scene arrangement strategy dependent on the probabilistic inert part tree	probabilistic dormant segment tree with helped highlights at its branch hubs
Xiaodong Yu et al., 2011[7]	high level learning in scene recognition.	issue of scene acknowledgment in an dynamic vision
Xiaodong Yu et al., 2011[8]	scene acknowledgment utilizing video pictures through the examination of human exercises.	Less reliability.
Tran H.M. et al., 2017[9]	the technique to identify and perceive the difference in scene	hard to figure out which scene changing amid the day time frame
K. E. A. van de Sande et al., 2010[10]	visual information on the level of objects and scene types	More complex data.

IV. METHODS USED IN SCENE RECOGNITION

i) Scene recognition using object recognition (SR-OR): -This method is used for distinguishing very complex scenes which might be difficult in using low standard features in object recognition. To overcome the local features, high level image is presented called as object bank. The presence of multiple classes of objects within single image is a challenging set of features in objects.

ii) Scene recognition using low-level image features (SR-LLF):- There are many robust OR algorithms so using SR-OR can be problematic

V. CLASSIFICATION APPROACHES IN SCENE RECOGNITION

i) Decision based tree technique:-Decision Tree supervised learning consists of tree-like graph to predict the value based

because the object recognition algorithm is treated as a black box so object recognition propagate using scene recognition.

iii) Scene classification in spatial domain:-The different classes of scenes are represented by extracting the features of spatial domain. Scene recognition algorithm use superordinate level of descriptor, texture and edge feature[12].

iv) Scene classification in frequency domain:-The frequency domain effective source of information for scene image. The statistics of images in frequency domain reveals different spectral signatures for different image scenes[22].

on several input variables. A decision tree based on classification predicts the class of input in numeric variable.

The partition algorithm requires the set of variable. A decision Tree contains the tree like graph. The graph used for the prediction of the value of the target variable based on the different inputs. The decision tree contains the leaves and the branches. The prediction for the variable based on the decision tree to the leaf nodes. The classification tree outcomes class of the inputs nodes and the variables. The classification tree contains the class of the input and the output in the numeric form. The algorithms used to build decision tree for the recursive partitioning algorithm. Algorithm used to construct by splitting rule and stopping rule.

ii) Decision Tree Predictive model Technique:-The decision model predicts the image class .The bad weather and bright day night extracted from color obtain from classification.

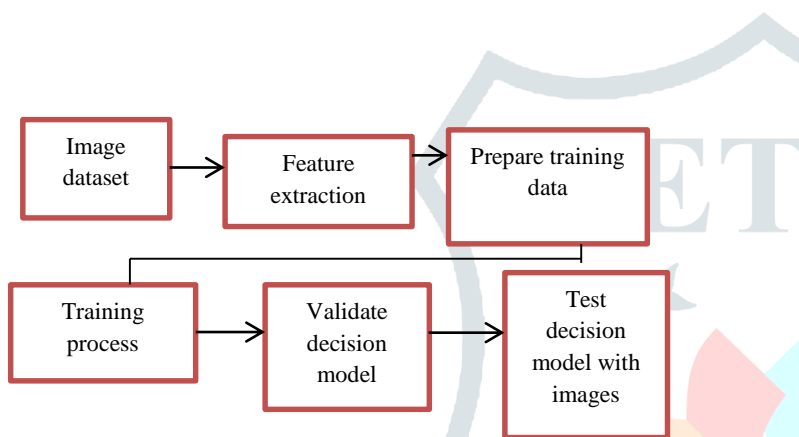


Fig. 3 Scene Recognition Structure

The above diagram depicts the steps for suitable decision tree. The decision tree model determines the training process for preparing database. The database has input variable to form desired output. The output has image class with five variables fixed in extraction stage [13]

iii) Independent Component Analysis Technique: - Independent Component Analysis (ICA) has main approach in image and signal processing. In ICA the independent signal is separated from source signal. To get IC image the binarized image obtained by extracting features[14]. The main idea of the ICA used to separate the source signal from the observed signal, as the connection of the source components. The mathematical model of the independent component analysis formulate by the independent processing components. The components have the mutual independence, stationary and zero. The observed signal regarded as the connection of the source components. The consideration of the model in the linear representations in noise free model described as the unknown full matrix model.

iv) Features of Haar Technique:-Haar is the fast image processing in which blocks of images divided into blocks sizes 64*64 and each block apply haar transform for 2 times to get LL component for block image. The set of feature vector is combined set of LL features. The designed weights for features are calculated for this approach. HAAR technique consists of an adjacent regions at specific location in the detection window, adding the intensities of the pixel in each region and computing the difference between the additional values. The difference used to categorise the subsections of the image. The example in HAAR technique is given in the form of the human face. It has been observed that all the faces region

around the eyes is darker than the cheeks. So, the position of the rectangle which lies between the eyes and the cheeks region. The location of the rectangle of the detection window acts like boundary box to target object. In The detection phase, a window of the target size is moved to the input image and HAAR like features is calculated. The difference is done by the separation of the objects from the non objects. The HAAR like features classifier like features are important the accuracy of the object. In the HAAR like features, organised in the way called classifier. The singularity of the matrix is obtained from the given figure [23].

The edges is rough idea of the scene. The many long vertical edges of tall buildings correspond to horizontal edges is an example of edge. Simultaneously, forest would show uniform distribution of dominant long vertical edges. On other hand, mountains show the diagonal edges. Streets and highways have characteristics edges.

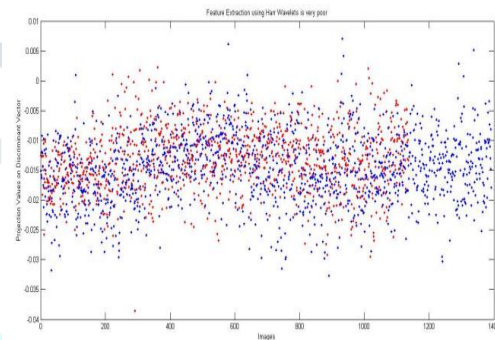


Fig. 4 HAAR Features

vi) Binarisation Process:- In this process the independent component analysis based on the probability of objective function and selected optimization scheme. The fixed set of the ICA algorithm separate the scene images from the background images. The binarisation process described as :-

a) Separate Model:-The connection of the pixels from the three sources and noiseless connection mixture. A set of single image on the observed signals such as R,G and B channels. So the intensity of each signal observed from other three signals with regular intensity of that pixel.

b) Threshold value:-The algorithm recognise the global threshold from an image by decreasing the class variance if the other classes of the front and the front and the back pixels. That can be done by increasing the class variance for the value of the threshold. The mean of each class determine the potential threshold value.

VI. MULTI-MILLION DATASETS

i) Data sets:-The convolutional neural network benefits for the visual recognition and recursive neural network for language processing. The training dataset has the coverage of classes and various classes. The space for the data set depends on the tasks and performance in domain of artificial intelligence.

The increasing availability of the data sets ranges to 30 million items in convolutional neural networks matches to human level recognition ranges from. In visual recognition object classes 1 million to 2.5 million scene images. Human would experience data natural system and right recognition of data with coverage of spaces of classes.



Fig. 5 Dataset Categories (Indoor, Nature and Urban)

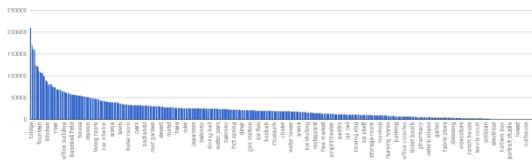


Fig. 6 Sorted Image Categories

ii) Single scale:- In this the different scales are determined on the basis of the CNN. The architecture consists of the different images at different scales. The two scales are concatenated and taken at SVM. The final output shows that combination of the different images at different scales are hybrid to the global sets. The CNN features are extracted from the local scales with image NET. The final determination will boost the performance of the accuracy and reliability of the desired form of the set. The accuracy based on the final set of the values.

iii) Multiscales:- In the complex complex set of the networks, there is the combination of the scales. The features of the sets of the networks are concatenated for each specific scale of the networks. In this type of the complex networks the PCA method is used to reduce the dimension of each combined dimension of the network. In this method the scales using Image Net and scales using CNN are corresponds to the network [26].

The information seen during the training and tested at the different sets of the data are seen in the convolutional neural networks. The ImageNets in the convolutional neural networks are used in the scene recognition during the test bias and the data about the objects recognise the scenes. The characteristics of datasets plays a main role in explaining the causes of the data sets. The scene and the sets of the objects datasets helps in recognising the main properties of the target data sets. The places of the data sets with 200 categories consist of large data with the low data sets bias. The main focus on the Image Net which contains the object data. In this way the statistics are collected and the distribution are compared in accordance to the main data.

VII. COMPARISON BETWEEN SIFT AND HOG FEATURE EXTRACTION ALGORITHM

(i) Scale-invariant feature transform (or SIFT)

The detection of the local features in images and the object of images can be extracted to describe the features of the image is called as Scale-invariant feature transform (or SIFT). The SIFT is an algorithm in computer visualisation. The first research of the SIFT algorithm was done by David Lowe in

1991 and now reviewed. The feature of the extraction the images and the properties matching different objects and scenes used is recognised by SIFT algorithm. The features of the images are particularly invariant to view point images and the rotation of the image scaling. The spatial and frequency domains, decrease the chances of the disruption by occlusion, clutter, or noise. The complex set of the features with typical images can be extracted using SIFT algorithms. The relative location between the original image scene cannot be changed from one point to another image [27]. The different corners of the square box used as the features position and regardless of the other point positions. The points in the frame used already in this the square box may be opened or closed. The features of the flexible objects change the internal geometry between the different images. The feature of the SIFT detects the features and reduces the contribution of the errors by variations in the local images.

The features are extracted in the SIFT features from the set of reference images stored in the database. The new image is compared to the last images features based on the Euclidean distance of their feature vectors. The keypoint descriptors discovers the match possibility in a large database of features. The background features may not match the database, so in order to find the correct database from the subsets of the keypoints of objects, local scale and orientation in the new image.

i). **Detection of extrema of space**:- The primary phase of calculation looks over all scales and picture areas. It is executed effectively by utilizing a distinction of Gaussian capacity to recognize potential intrigue indicates that are invariant scale and orientation.

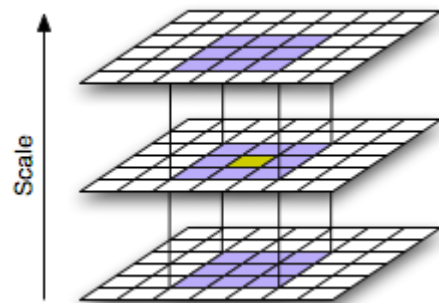


Fig. 7 scale space extrema detection [24]

ii. **Key point Localisation** :- At every applicant area, a definite model is fit to decide area and scale. Key focuses are chosen dependent on proportions of their steadiness.

ii. **Orientation assignment**:- In this type, at least one introduction are assigned to each key point area dependent on nearby gradient direction. Every future task are performed on image information that has been changed with respect to the allocated orientation, scale, and area for each component, accordingly giving invariance to these changes.

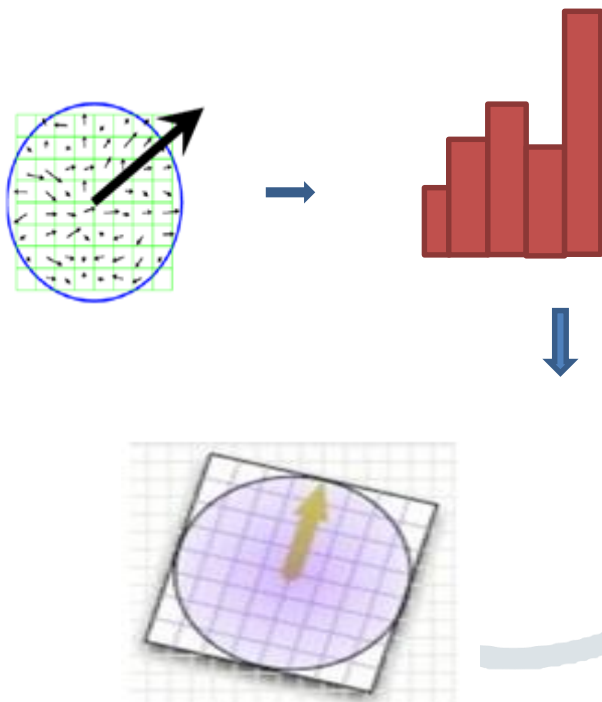


Fig. 8 Orientation assignment

iv.)Keypoint descriptor:-The local picture gradients are measured at the designated scale in the region around each key points which are then converted into a representation that allows for significant levels of local shape distortion and change in illumination.

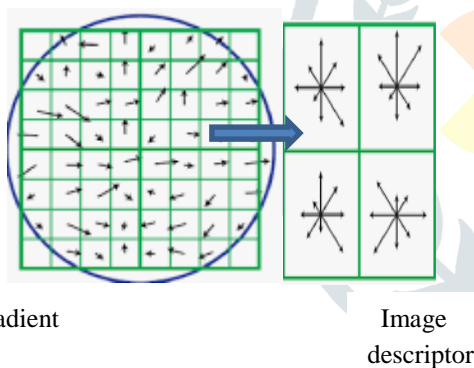


Fig. 9 Key Point Descriptor

(ii) Histogram of Oriented Gradients(HOG)

Histogram of Oriented Gradients is used for the detection of the image in image processing and feature extraction. HOG is feature descriptor technique computes the gradient orientation in limited portion of the image. This technique is same as the edge detection and scale invariant feature featute transform technique[25].In the overview method , a set of input image is taken then the color is detected is detected. The next procedure ia the calculation of the gradients. After that, the differentiation of the image window in to small spatial regions is called as cells. Then, the celss divided in to small spatial cells of the gradient over the pixels of the cells.The cells of the spatiall blocks, overlapps the spatial blocks. In next procedure,The collection of the Histogram oriented gradients over the detection window.Then the support vector is put in the linear form. In the final step the detection and classification of the image. HOG used in evaluating the local histograms of image in desired in image. The shape of an object characterised distribution of local gradients or the edge directions.

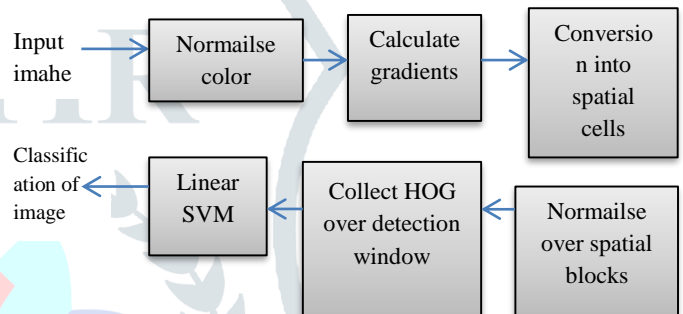


Fig. 9 Overview method of HOG

In the overview method , a set of input image is taken then the color is detected is detected. The next procedure ia the calculation of the gradients. After that, the differentiation of the image window in to small spatial regions is called as cells. Then, the cells divided in to small spatial cells of the gradient over the pixels of the cells.The cells of the spatiall blocks, overlapps the spatial blocks. In next procedure,The collection of the Histogram oriented gradients over the detection window.Then the support vector is put in the linear form. In the final step the detection and classification of the image.

TABLE 2:- COMPARISON TABLE BETWEEN SIFT AND HOG

	SIFT	HOG
defination	Transform an image in to large collection of feature vectors.	Transform an image in to blocks, then construct histogram gradient.
use	SIFT is used for matching local regions.	HOG is used in sliding window.
Identification	SIFT is used for identification of specific objects and for classification	HOG is used to classify patches using classifiers such as SVM's.
Features	Cannot helps in detection of cells.	False markup rate is low in dection of cells then other algorithm.

TABLE 3:- COMPARATIVE ANALYSIS IN FEATURE EXTRACTION IN (SIFT and HOG)

SIFT	HOG
Evaluated around scale invariant feature points obtained using the difference of gaussian (DoG) key point detector.	Dense evaluated all over the image
It does not learn the representation by itself, it is hard coded.	No learning algorithms for HOG features.
First order gradients	Based on first order image gradients pooled in orientation bins.
Compute descriptor and histogram of orientations	Make histogram of orientations

VIII.CONCLUSION AND FUTURE SCOPE

It is concluded that a millions of image recognition is done by scene recognition. In this survey, scene recognition using objects and image features whereas multiple classes of objects within single image is a challenging set of features in objects. The frequency domain becomes an effective source of

information in scene recognition. The process of facial recognition with the Haar Cascade and Eigenface method recognizes the face both during the day and night. In scene recognition, Tiny Image dataset to ImageNet leads to multi-million-item dataset initiatives. In this survey paper, different techniques and approaches are determined for scene recognition from single image to multimillion images. The comparison between the Scale Invariant Feature Transform and Histogram of Oriented Gradients(HOG) is described in this paper. In the scale Invariant Feature Transform the image is formed by the collection of the feature vectors and the Histogram of Oriented Gradients(HOG) there is the blocks of the images. The increasing availability of the datasets increases the performance of the images.

The future scope will collect more data and train more scenario classes and research on improving the accuracy of scene recognition by improving the accuracy of target image detection should be done.

REFERENCES

- [1] Chen, C., Huang, J., Pan, C., & Yuan, X. (2018, September). Military Image Scene Recognition Based on CNN and Semantic Information. In *2018 3rd International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*(pp. 573-577). IEEE.
- [2] Mantoro, T., & Ayu, M. A. (2018, May). Multi-Faces Recognition Process Using Haar Cascades and Eigenface Methods. In *2018 6th International Conference on Multimedia Computing and Systems (ICMCS)* (pp. 1-5). IEEE.
- [3] Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2018). Places: A 10 million image database for scene recognition. *IEEE transactions on pattern analysis and machine intelligence*, 40(6), 1452-1464.
- [4] Zhou, L., Hu, D., & Zhou, Z. (2013). Scene recognition combining structural and textural features. *Science China Information Sciences*, 56(7), 1-14.
- [5] Herranz, L., Jiang, S., & Li, X. (2016). Scene recognition with CNNs: objects, scales and dataset bias. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 571-579).
- [6] Atsumi, M. (2012). Object and Scene Recognition Based on Learning Probabilistic Latent Component Tree with Boosted Features. *International Journal of Machine Learning and Computing*, 2(6), 762.
- [7] Yu, X., Fermüller, C., Teo, C. L., Yang, Y., & Aloimonos, Y. (2011, November). Active scene recognition with vision and language. In *2011 International Conference on Computer Vision* (pp. 810-817). IEEE.
- [8] Higuchi, M., Aoki, S., Kojima, A., & Fukunaga, K. (2004, August). Scene recognition based on relationship between human actions and objects. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.* (Vol. 3, pp. 73-78). IEEE.
- [9] TRAN . H. M (2017, july). Scene recognition based on relationship between human actions and objects. ICSSE.IEEE
- [10] K. E. A. van de Sande. Evaluating Color Descriptors for Object and Scene Recognition. In *IEEE Transactions on Pattern .IEEE*
- [11] Candamo, J., Shreve, M., Goldgof, D. B., Sapper, D. B., & Kasturi, R. (2010). Understanding transit scenes: A survey on human behavior-recognition algorithms. *IEEE transactions on intelligent transportation systems*, 11(1), 206-224.
- [12] Farinella, G. M., & Battiato, S. (2010). Representation Models and Machine Learning Techniques for Scene Classification. *Pattern recognition and machine vision*, 6, 199.
- [13] Rouabeh, H., Abdelmoula, C., & Masmoudi, M. (2014, November). Performance evaluation of Decision Tree and neural network techniques for road scene image classification task. In *Image Processing, Applications and Systems Conference (IPAS), 2014 First International* (pp. 1-6). IEEE.
- [14] Kherada, S., & Nambodiri, A. M. (2013, November). An ICA based Approach for Complex Color Scene Text Binarization. In *Pattern Recognition (ACPR), 2013 2nd IAPR Asian Conference on* (pp. 125-129). IEEE.
- [15] Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive psychology*, 41(2), 176-210.
- [16] Oliva, A., Torralba, A., Guérin-Dugué, A., & Héroult, J. (1999, February). Global semantic classification of scenes using power spectrum templates. In *Proceedings of the 1999 international conference on Challenge of Image Retrieval* (pp. 9-9). BCS Learning & Development Ltd.
- [17] Torralba, A. B., & Oliva, A. (1999, September). Semantic organization of scenes using discriminant structural templates. In *iccv* (p. 1253). IEEE.
- [18] Vogel, J., & Schiele, B. (2004, August). A semantic typicality measure for natural scene categorization. In *Joint Pattern Recognition Symposium* (pp. 195-203). Springer, Berlin, Heidelberg.
- [19] Fei-Fei, L., & Perona, P. (2005, June). A bayesian hierarchical model for learning natural scene categories. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* (Vol. 2, pp. 524-531). IEEE.
- [20] Lazebnik, S., Schmid, C., & Ponce, J. (2006, June). Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *null* (pp. 2169-2178). IEEE.
- [21] Bolc, L., Wojciechowski, K., Tadeusiewicz, R., & Chmielewski, L. J. (Eds.). (2012). *Computer Vision and Graphics: International Conference, ICCVG 2012, Warsaw, Poland, September 24-26, 2012, Proceedings* (Vol. 7594). Springer.
- [22] Oliva, A., Torralba, A., Guérin-Dugué, A., & Héroult, J. (1999, February). Global semantic classification of scenes using power spectrum templates. In *Proceedings of the 1999 international conference on Challenge of Image Retrieval* (pp. 9-9). BCS Learning & Development Ltd..
- [23] Vogel, J., & Schiele, B. (2007). Semantic modeling of natural scenes for content-based image retrieval. *International Journal of Computer Vision*, 72(2), 133-157.
- [24] Jindal, R., & Vatta, S. (2010). Sift: Scale invariant feature transform. *IJARIT*, 1, 1-5.
- [25] Dadi, H. S., & Pillutla, G. K. M. (2016). Improved face recognition rate using HOG features and SVM classifier. *IOSR Journal of Electronics and Communication Engineering*, 11(04), 34-44.
- [26] Herranz, L., Jiang, S., & Li, X. (2016). Scene recognition with CNNs: objects, scales and dataset bias. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 571-579).
- [27] Zhu, Y., Yao, C., & Bai, X. (2016). Scene text detection and recognition: Recent advances and future trends. *Frontiers of Computer Science*, 10(1), 19-36.