International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

Identifying Semantic Category of Distorted Images along with Text Recognition

Anu E[1], Anu K S[2]

[1, 2]Department of Computer Science and Engineering, KMCT College of Engineering, Calicut

Abstract: ***Scene recognition provides visual information from the level of objects and the relationship between them. The main objective of scene recognition is to reduce semantic gap between human beings and computers on scene understanding. For example, recognize the context of an input image and categorize it into scenes (forest, seashore, building etc). Some of the applications of scene recognition are object recognition, object detection, video text detection etc. Different methods are there for scene recognition and understanding. All are having positive and negative aspects. One of the main difficulties is to increase the accuracy. The negative aspects which affect the improvement in accuracy are, intrinsic relationship across different scales of the input image are not analyzed and impact of redundant features. This paper develops a framework to overcome these limitations and provide better understanding of input image. The suggested framework includes reconstruction of distorted image and text recognition, if any, along with scene recognition to get a clear idea about the input image. Image reconstruction is done by using Non Local Mean Image Denoising method. To detect and describe local features of an image SIFT method is used. Finally classification is done by SVM classifier. Text recognition is achieved with the help of OCR Methodology.***

Keywords: Harris-Laplace detector, Image reconstruction, Non Local Mean Image Denoising, OCR, SIFT

1. Introduction

Image processing is the process of analysis and manipulation of a digitized image in order to improve its quality. Two principles of Image processing are improvement of pictorial information and processing of scene data. Enabling computers to see the way in which we see things around us is a challenging task. Generally, a learning based approach is used to perform these types of activities. A training set which would contain representative images from all categories that we need to classify will be created initially. Now these images are labeled manually to the class they belong as that of human vision. If a random image is given as an input, the machine would try to classify the image on basis of features already identified. It is necessary to figure out the most important features that are having a very strong co-relation with the class of the image. This makes learning easier, faster and an error free job. Thus, feature identification is a most important task. Once, the features are identified a classifier can be used to classify the scene.

The term recognition is used to refer to many different visual capabilities including identification, categorization and discrimination. Identification means equality on a physical level, Categorization means assigning an object to some category as humans do and Discrimination means assigning an object to one class. Recognizing the semantic category of complex scenes having content variations is a challenging task. A new framework is introduced to overcome the limitations in order to increase the accuracy. This framework includes three modules. First module is image reconstruction by using Non Local Mean Image Denoising. Second module is identifying semantic category of input image based on SIFT feature extraction. Third module is text recognition to get more information about the input image.

The main advantages of suggested method compared to existing methods are the following; First one is, Harris Interest Points and Scale Invariant Feature Transform (SIFT) descriptors are combined for fast scale invariant scene categorization. The use of SIFT features to detect and describe local features in an image while preserving the underlying manifold structure of each feature data. Second one is, image reconstruction of distorted images before the classification process. Third on is, retrieval of text information.

2. Related Works

Many scene recognition techniques try to build an intermediate semantic representation to reduce semantic gap. These methods focus on extracting low level features from single resolution image. It may fail to represent the entire scene completely. Some redundant features may reduce the accuracy of scene recognition. So a survey is required to check whether the features are useful to recognize semantic category of an input image.

Chang Cheng [2] introduced a novel outdoor scene image segmentation algorithm based on the background recognition and perceptual organization is introduced. A perceptual organization model (POM) is introduced for structurally challenging objects. This model can capture the non accidental structural relationships in the constituent parts of the objects. In POM, obtaining the geometric properties of object parts is a necessary task. The object parts may have homogenous surfaces, so the uniform regions in an image correspond to object parts. Another problem source is strong reflection. There exist some object classes with very complex structures and some parts of the objects may not strongly attach to other parts of the object. In this case, POM may not be able to piece the entire object together.

Yanfei Zhong [3] suggests scene classification is an effective tool for semantic interpretation of high spatial resolution (HSR) remote sensing image. The probabilistic topic model

(PTM) can be applied to natural scenes by using a single feature but it is inadequate for HSR images due to its complex structure. So a technique called SAL PTM is introduced to combine multiple features. In SAL PTM the complementary spectral, texture, and scale invariant feature transform features are combined in an effective manner. But in this method focus on normalization constraint of the PTM is required and structural features which are more appropriate for HSR images should be explored.

Yongzhen Huang [4] brought the idea of using genetic programming (GP) to generate composite operators and composite features from combinations of primitive operations and primitive features in object detection. The main reason for using GP is to overcome the human expert's limitations occurred in the feature synthesis. These limitations are the result of focusing only on conventional combinations of primitive image processing operations. In order to improve the efficiency of GP and a new fitness function is designed. It is based on minimum description length principle. This helps to incorporate both the pixel labeling error and the size of a composite operator into the designed fitness evaluation process.

Anna Bosch [5] introduced a hybrid discriminative approach. In this approach a set of labeled images of scenes is provided. The aim of this approach is to classify a new image into one of the categories (e.g., coast, forest, building, etc.). First discover latent topics using probabilistic Latent Semantic Analysis (pLSA). For each image a generative model from the statistical text literature is applied to a bag of visual words representation and training a multiway classifier on the topic distribution vector for each image. A novel vocabulary using dense color SIFT descriptors is introduced. The classification performance will be achieved with a discriminative classifier. But here, the images with a semantic transition between categories are not well clustered because no sufficient ambiguous images are there.

Li Fei-Fei [6] examined a bayesian hierarchical model. In this model, an input image is represented as a collection of local patches. Each patch of the input image is represented using a codeword. Codeword is taken from a large vocabulary of codewords called codebook. Image patches are detected using a sliding grid and random sampling of scales. The goal is to get a model that represents the distribution of these codewords in each category of scenes more accurately. In recognition phase, first identify all the codewords corresponding to unknown input image. Then determine the best category model that represents the distribution of the codewords of the particular image.

Jian Yao [7] provided an approach to holistic scene understanding that reasons jointly about regions, location, class and spatial extent of objects, presence of a class in the image, as well as the scene type. The aim of holistic scene understanding is recovering multiple related aspects of a scene to provide a deeper understanding of the scene as a whole. But some of the main sources of error are bad unary potentials and false negative detections.

Xiaodong Yu [8] introduced an active vision framework called active scene recognition is introduced for utilizing high level knowledge for scene recognition. The proposed approach consists of two modules. First one is a reasoning module, which is used to obtain higher level knowledge about scene and object relations, proposes instructions to the second module and draws conclusions about the scene contents. The second one is sensory module, which includes a set of visual operators. It is responsible for extracting features from images, detecting and localizing objects and actions. The sensory module does not passively process the image and it is guided by the reasoning module. The approach is based on an iterative process.

Antonio Torralba [9] suggested Contextual Priming for Object Detection. The main idea is that the context is a rich source of information about an object's identity, location and scale. The framework is based on the correlation between the statistics of low level features across the entire scene and the corresponding objects. The main problem is the lack of simple representations of context and efficient algorithms for the extracting such information from the input. One way to define the context of an object is, define in terms of previously recognized objects within the scene. Here the drawback is, it renders the complexity of context analysis. Another drawback is Color is not taken into account in this study.

Based on literature survey, it has been found that many scene recognition techniques try to build an intermediate semantic representation to reduce semantic gap. Most scene recognition methods focus on extracting low level features from single resolution image. It cannot well represent the entire scene completely. Some redundant features may reduce the accuracy of scene recognition. One of the critical problems is that, whether the features are useful to recognize the semantic category of an input image. Identifying the semantic meaning of input image after reconstructing the same in case of any distortion, helps to widen the scope of proposed system. Text analysis along with the semantic meaning of input image can provides better understanding about the scene.

3. The Framework: Identifying Semantic Category

The proposed project Identifying semantic category of distorted image along with text recognition provides better understanding of scene category. Scene category of a distorted image can be categorized along with text recognition, if any. This method provides better understanding of the input scene with high accuracy. Enabling computers to see the way in which we see things around us. Generally, a learning based approach is used to solve these types of problems. A training set which would contain representative images from all categories that we need to classify will be created initially. Now these images are labeled manually to the class they belong as that of human vision. If a random image is given as an input, the machine would try to classify the image on basis of features already identified. It is necessary to figure out the most

important features that are having a very strong co-relation with the class of the image. This makes learning easier, faster and an error free job. Thus, feature identification is a most important task. Once, the features are identified a classifier can be used to classify the scene.

Identifying semantic meaning is challenging due to lack of accuracy. First, there may be distorted images. Second, redundant features may exist. Third, intrinsic relationship across different scales of input image is not analyzed. So techniques for image reconstruction, detect and describe local features in an image, and identifying text, if any, present in the scene are combined to overcome the negative aspects which reduces the accuracy.

The area of the proposed system in image processing is Cybernetics. It is the science of Control & Communication in the animal and machine. This area is concerned with the study of systems of any nature which are capable of receiving, storing and processing information so as to use it for control. The system can perform Reduced Reference Image Quality Assessment. That is, the system can perform well where the perfect reference is not available, instead distorted images are given. So the system can be used as a tool to refer historical data.

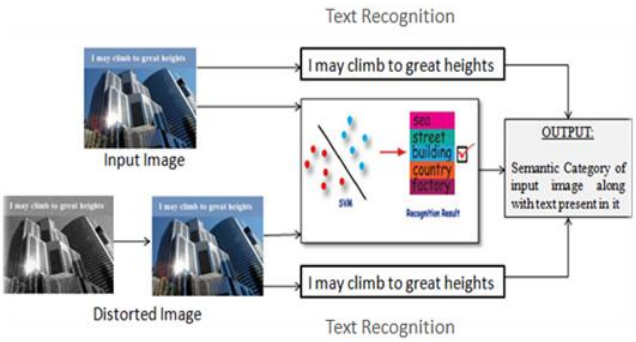

Figure 1: The Framework for Identifying semantic category of distorted image along with text recognition

Figure 1 shows the framework of the proposed system. Harris Interest Points and Scale Invariant Feature Transform (SIFT) descriptors are combined for fast scale invariant scene categorization. Along with that Non Local Mean Image Denoising (NLM), Support Vector Machine (SVM) and Optical Character Recognition are used to provide better understanding about distorted images. It is a combined frame work to reconstruct the distorted image, categorize it and along with text recognition. SIFT helps to features that are invariant to illumination, scale and rotation. The Non-Local Mean Image Denoising takes the mean of all pixels in the image, based on the similarity of these pixels with target pixel. This results in more clarity and less loss of details in the image. Support Vector Machine constructs a hyperplane for classification purpose. Optical Character Recognition helps to convert text present in images into machine-encoded text. The proposed project consists of following modules; Image Reconstruction, Feature Extraction, Classification and Text recognition.

3.1 Image Reconstruction

Non-local means is used for image denoising. Non-local means filtering takes a mean of all pixels in the image, weighted by how similar these pixels are to the target pixel. It results in high filtering clarity and less loss of detail in the image. It assumes the image contains an extensive amount of self-similarity. This self-similarity assumption is used to denoise an image. Pixels with similar neighborhoods are the base to determine the denoised value of a pixel. The following equation represents the Weight function, where Z is the normalizing term, * is convolution operator, G_d is Guassian spatial kernal and h is filtering parameter.

$$w_{ij} = \frac{1}{Z_i} \exp^{-\frac{1}{h^2} G_d * \left| I\left(N^d(i)\right) - I\left(N^d(j)\right) \right|_2^2}.$$

3.2 Feature Extraction

To extract or to detect and describe the feature points (keypoints), we use combination of Harris-Laplace detector and SIFT Descriptors. The Harris-Laplace detector combines the Harris corner detector with the Gaussian scale space representation to create a scale-invariant detector. Harris-Laplace approach detects different regions where Harris detector concentrates on corners and highly textured points. Harris-corner points have good rotational and illumination invariance in addition to identifying the keypoints of the input image. Hence these detectors extract complementary features from the input images. A corner is the intersection of two edges. An interest point is a point in an image which is having a well-defined position and it can be detected robustly. This means that an interest point can be at corner, line endings, or a point on a curve. Corner detection algorithms tests each pixel in the image to see if a corner is present based on how similar a patch centered on the pixel is to nearby overlapping patches. The similarity is measured based on the sum of squared differences (corner score) between the corresponding pixels of two patches. A Gaussian scale space representation of an image is the set of images received by convolving a Gaussian kernel of various sizes with the original image. Convolution with the Gaussian kernel smooths the input image.

Once the Harris Laplace points are received orientation assignment will take place. To descript those points SIFT Descriptors are used. Orientation assignment (to the keypoint locations) is performed to ensure invariance to image location, scale and rotation. Now compute SIFT descriptor vector for each keypoint. This descriptor is highly distinctive. It is partially invariant to the other variations such as illumination, high dimensional viewpoint, etc. For that create gradient histogram. These histograms are computed from magnitude and orientation values of samples.

3.3 Classification

In SVM Classifier, Given a set of labeled training data, then it outputs an optimal hyperplane which categorizes new examples/inputs. Following steps are there; initially, obtain the Support Vectors (SVs) closest between each class. Then

create decision hyperplane. Support vectors closest to each class are identified. Classification is based on the distance of the vectors from hyperplane. Support Vector Machine classifies the input image into a particular category based on the extracted features. LS and OT datasets are used for training process.

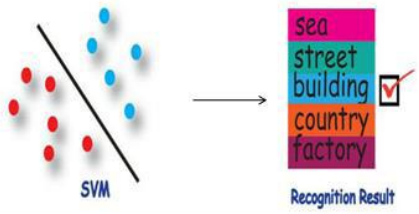

Figure 4: Classification using SVM

3.4 Text Recognition

This section includes following task to recognize the text present in an input image. Initially the texts regions are extracted then do skew correction. Perform Binarization of regions. Then characters are passed into the recognition module. Finally recognize the module. In extract text regions step, partition the input image into m number of blocks. Identify information block IB and background block BB based on the intensity variation within it. Then remove BB and non text components based on heuristically chosen rules. Now perform skew correction. For that calculate the skew angle. Consider the bottom profile of the gray shade of a text region to height in terms of pixel from the bottom edge of the rectangle. Rotate the text region with estimated skew angle. Then perform binarization of regions. For that consider 8 neighbor pixels of each pixel in the text region. Find arithmetic mean of minimum and maximum intensities of text regions. Take it as a threshold for binarization. After binarization, perform segmentation into lines and characters. Analyze horizontal histogram for segmenting regions into text lines. Use vertical histogram of each text line to identify word. Final module is recognition module. An appropriate table is created in the data base to compare the obtained values. Correlation between a template and a test pattern is calculated.

4. Implementation and Analysis

The proposed framework is implemented using Matlab. It is tested on OT and LS dataset. The OT dataset consists of 8 categories and LS dataset consist of 7 categories. These 15 categories are included for training and testing process. An image is provided as an input and classifies it based on the extracted features. If the provided image is distorted, it will be reconstructed before the classification process. After the classification process, the systems will identify the text (if any) present in the image.

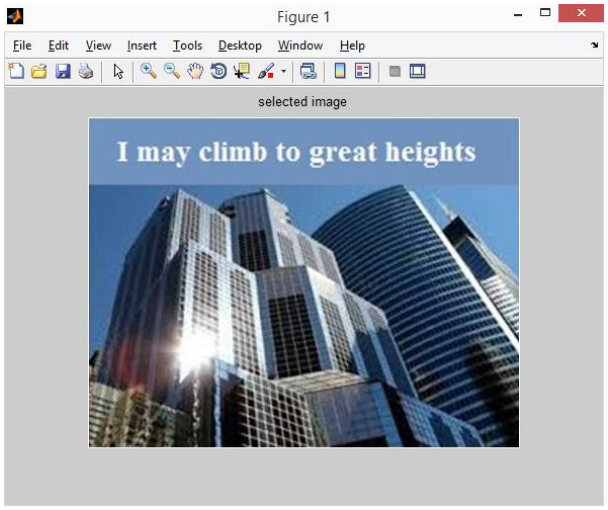

Figure 1: Original Image

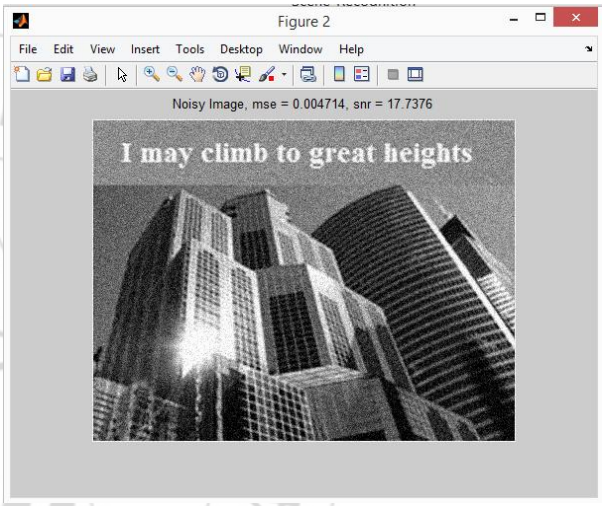

Figure 2: Noisy Image

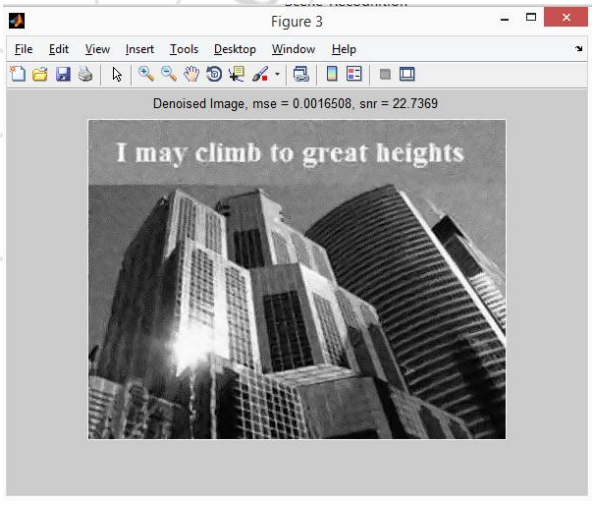

Figure 3: Denoised Image

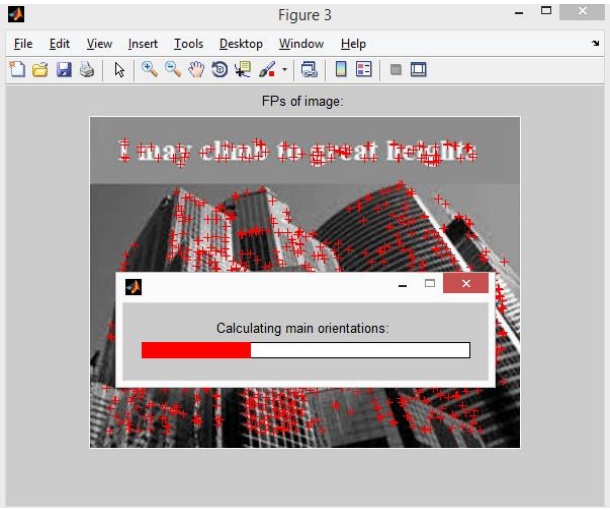

Figure 4: Feature Extraction

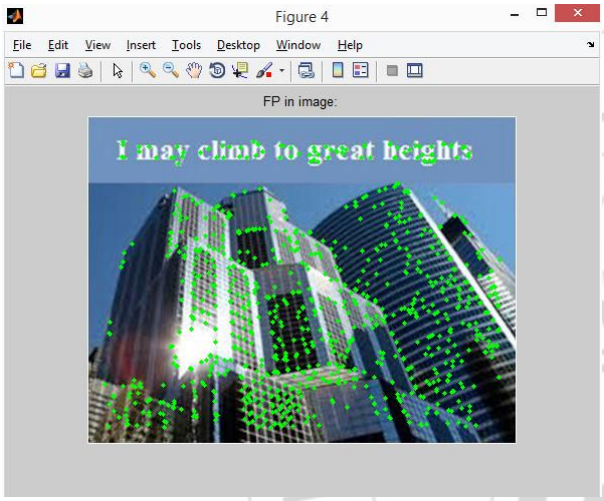

Figure 5: Key Points

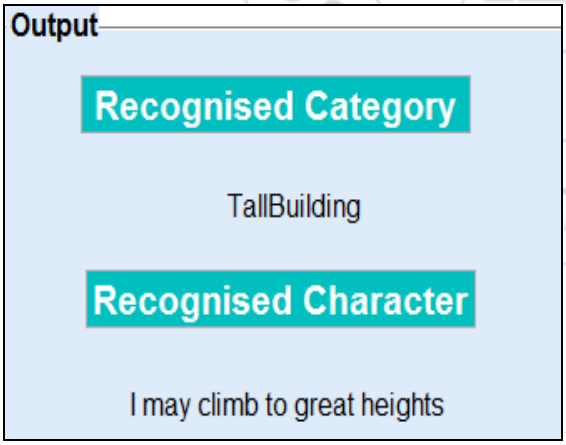

Figure 6: Output

By comparing with the existing similar methodologies, the proposed system performs more efficiently. It results in good accuracy in image categorization.

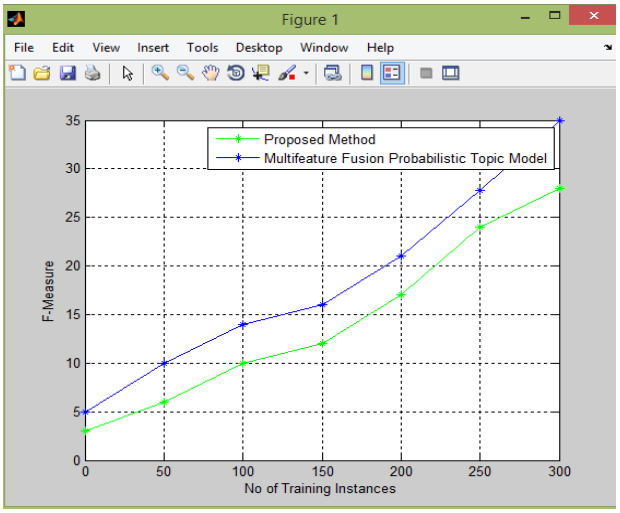

Figure 7: Comparison Graph

A comparison graph is plotted with F-Measure (is a measure of a test's accuracy) against No of Training Instances. The proposed system is compared with Multifeature Fusion Probabilistic Topic Model. Here the proposed method performs well. It has an increased accuracy rate compared with the other systems.

5. Conclusion

Harris Interest Points and Scale Invariant Feature Transform (SIFT) descriptors are combined for fast scale invariant scene categorization. Along with that Non Local Mean Image Denoising (NLM), Support Vector Machine (SVM) and Optical Character Recognition are used to provide better understanding about distorted images. It is a combined frame work to reconstruct the distorted image, categorize it and along with text recognition. SIFT helps to features that are invariant to illumination, scale and rotation. The Non-Local Mean Image Denoising takes the mean of all pixels in the image, based on the similarity of these pixels with target pixel. This results in more clarity and less loss of details in the image. Support Vector Machine constructs a hyperplane for classification purpose. Optical Character Recognition helps to convert text present in images into machine-encoded text. Different types of distortion can be merged to the proposed frame work as future enhancement.

References

[1] Xiaoqiang Lu, Xuelong Li, Fellow, IEEE, and Lichao Mou "Semi Supervised Multitask Learning for Scene Recognition" IEEE TRANSACTIONS ON CYBERNETICS, VOL. 45, NO. 9, SEPTEMBER 2015
[2] Chang Cheng, Andreas Koschan, Member, IEEE, Chung-Hao Chen, David L. Page, and Mongi A. Abidi "Outdoor Scene Image Segmentation Based on Background Recognition and Perceptual Organization" IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 21, NO. 3, MARCH 2012
[3] Yanfei Zhong, Senior Member, IEEE, Qiqi Zhu, Student Member, IEEE, and Liangpei Zhang, Senior Member, IEEE "Scene Classification Based on the Multifeature Fusion Probabilistic Topic Model High Spatial

Resolution Remote Sensing Imagery" IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 53, NO. 11, NOVEMBER 2015
[4] Yongzhen Huang,Nat Lab of Pattern Recognition, Chinese Acad of Sci Beijing ; Kaiqi Huang LiangshengWang DachengTao "Enhanced biologically inspired model" Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference june 2008
[5] Annabosch,&andrewzisserma"scene classification using a hybrid generative approach" ieee transactions on pattern analysis and machine intelligence, vol. 30, no. 4, april 2008
[6] Li Fei-Fei California Institute of Technology Electrical Engineering Dept. Pietro Perona California Institute of Technology Electrical Engineering Dept "A Bayesian Hierarchical Model for Learning Natural Scene Categories" Pasadena, CA 91125, USA
[7] Jian Yao TTI Chicago, Sanja Fidler University of Toronto, Raquel Urtasun TTI Chicago "Describing the Scene as a Whole: Joint Object Detection, Scene Classification and Semantic Segmentation"
[8] Xiaodong Yu_, Cornelia Ferm¨ullery, Ching Lik Teoz, Yezhou Yangz, Yiannis Aloimonosz "Active Scene Recognition with Vision and Language" Computer Vision Lab, University of Maryland, College Park, MD 20742, USA
[9] Antonio Torralba Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA "Contextual Priming for Object Detection" Received August 15, 2001; Revised October 21, 2002; Accepted January 15, 2003
[10] Ayatullah Faruk Mollah1, Nabamita Majumder, Subhadip Basu and Mita Nasipuri "Design of an Optical Character Recognition System for Camera-based Handheld Devices" IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 4, No 1, July 2011
[11] Luis Malag´on-Borja and Olac Fuentes "An Object Detection System using Image Reconstruction with PCA" Proceedings of the Second Canadian Conference on Computer and Robot Vision (CRV'05) 0-7695-2319-6/05 $ 20.00 IEEE
[12] S.V.N. Vishwanathan, M. Narasimha Murty Dept. of Comp. Sci. and Automation, Indian Institute of Science, "SSVM : A Simple SVM Algorithm
[13] Jair Cervantes , Xiaoou Li , Wen Yu Department of Computer Science. CINVESTAV "Multi-Class SVM for Large Data Sets Considering Models of Classes Distribution"
[14] Dan Ventura \SVM Example" March 12, 2009
[15] G.Vamvakas, B.Gatos, N. Stamatopoulos, and S.J.Perantonis Computational Intelligence Laboratory, Institute of Informatics and Telecommunications, National Center for Scientific Research" A Complete Optical Character Recognition Methodology for Historical Documents" The Eighth IAPR Workshop on Document Analysis Systems
[16] Westley Evans "Image Denoising with the Non-local Means Algorithm" CS766 Fall 2005
[17] Sachin Chachada ; Signal and Image Processing Institute, Ming Hsieh Dept. of Electrical Engineering, University of Southern California, Los Angeles, 90089-2564, USA ; Byung Tae Oh ; Namgook Cho ; San A. Phong "Extension of Non-Local Means (NLM) algorithm with Gaussian filtering for highly noisy images
[18] Deng ; Dept. of Electron. Eng., La Trobe Univ., Bundoora, Vic., Australia ; L. W. Cahill "An adaptive Gaussian filter for noise reduction and edge detection"
[19] A. Buades ; Dept. Matematiques i Informatica, UIB, Palma de Mallorca, Spain ; B. Coll ; J. -M. Morel "A non-local algorithm for image denoising" 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
[20] deepak raghuvanshiv dept. of digital communication rkdfist, bhopal (m.p.) shabahat hasan dept. of digital communication rkdfist, bhopal (m.p.) mridula agrawal dept. of digital communication ies-ips academy,indore (m.p.) "Analysing Image Denoising using Non Local Means Algorithm"

Author Profile

Anu E is pursuing her M.Tech degree in Computer Science and Engineering from KMCT College of Engineering, Calicut University. She obtained her B.Tech Degree in Computer Science and Engineering from Amrita School of Engineering, in 2011

Anu K.S is Assistant Professor, Department of Information Technology, KMCT College of Engineering, Calicut University. Her research focuses on image processing and data mining. She obtained AMEI in Computer Science & Engineering from IEI in 2007. She completed her M.Tech degree in Computer Science & Engineering from NMAMIT college, Nitte in 2010.