Identification of Distorted Images along with Text Recognition using SIFT algorithm

Sreelekshmi A. N
Vadasseril, Puramattom, Thiruvalla
India
sree.an1989@gmail.com

ABSTRACT: *Reading text from photographs is a challenging problem that has received a significant amount of attention. Two key components of most systems are- (i) the text detection from images; and (ii) the character recognition. Many recent methods have been proposed to design better feature representations and models for both. Scene recognition provides visual information from the level of objects and the relationship between them. The main objective of scene recognition is to reduce the 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 used for scene recognition and understanding. All of them have positive and negative aspects. In this issue one of the main difficulties is how to increase the accuracy. The negative aspects which affect the improvement in accuracy are the intrinsic relationship across different scales of the input images which are not analyzed with respect to the 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 blurred 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 PCA. To detect and describe local features of an image, SIFT algorithm is used. Semi-supervised learning framework by combining SFSMR and Multitask Model is followed. Classification is done by SVM classifier. Text recognition is achieved with the help of OCR methodology.*

Keywords: Multitask Model, Object Detection, OCR, SFSMR, Semi Supervised Learning

Received: 12 March 2014, Revised 29 July 2014, Accepted 10 August 2014

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 the improvement of pictorial information and processing of the 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 the 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 the image reconstruction by using PCA. Second module is identifying the semantic category of input image which includes downsampling, SIFT feature extraction, semi supervised learning framework by combining SFSMR and Multitask Model and SVM with models of class distribution which is used for classification purpose. Third module is the text recognition to get more information about the input image.

The main advantages of suggested method compared to existing methods are the following. First, usage of multi-resolution images generated from the same scene because these different scale images include same global spatial structures and different local features. So the different tasks can be analyzed in a joint framework. This helps to improve the performance in scene recognition. Second, the use of SIFT features to detect and describe local features in an image. Third, optimal features can be chosen along with preserving the underlying manifold structure of each feature data.

2. Literature Survey

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.

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 the 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 multi-way 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 code-word. Code-word is taken from a large vocabulary of code-words 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 code-words in each category of scenes more accurately. In recognition phase, first identify all the code-words corresponding to unknown input image. Then determine the best category model that represents the distribution of the code-words 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.

Karamiet al [9] have performed using Using SIFT Algorithm the Different Image Deformations. Based on literature survey, it has been found that many scene recognition techniques try to build an intermediate semantic representation to reduce the 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. Proposed System

The proposed project, the 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 is important 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.

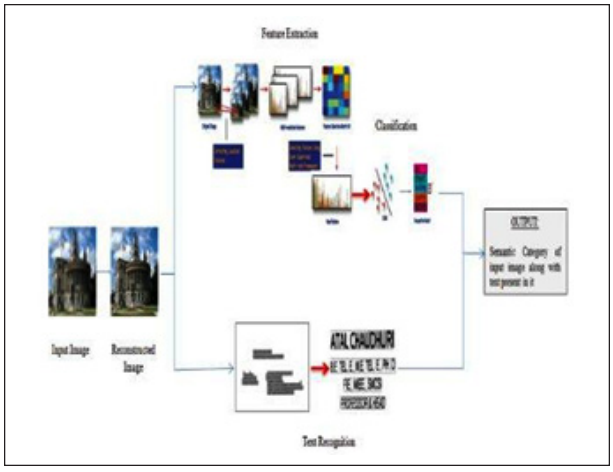

Figure 1. The Framework of Identifying semantic category of distorted image along with text recognition

Identifying semantic meaning is the challenging due to lack of accuracy. First, there exists the 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 proposed project consists of the following modules; Image reconstruction, Identify semantic category of input image and Text recognition.

3.1 Image Reconstruction

The main purpose of principal component analysis (PCA) is the analysis of data to identify patterns that represent the data well. The basic PCA approach is a linear projection technique that works well if the data is linearly separable. In the case of linearly inseparable data kernel, PCA is used. PCA includes the following tasks.

- Get the input image value in matrix format.
- Subtract the mean of dimension values from each of the data dimensions.
- Calculate the covariance Matrix.
- Calculate the eigenvectors and eigenvalues of the covariance matrix.
- Form the feature vector by choosing the components.
- Use the kernel function to map the original d-dimensional features into a larger, k dimensional; and
- Derive the new data set.

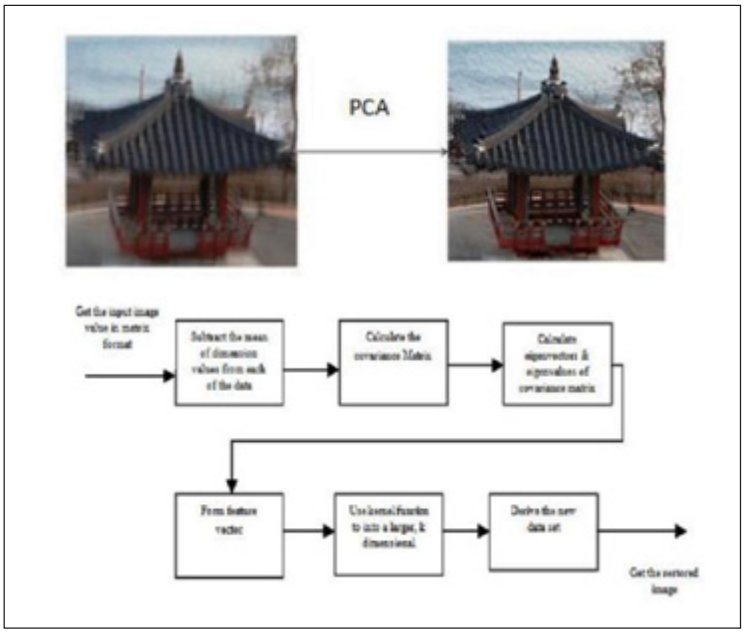

Figure 2. Image Reconstruction Workflow

3.2 Identify Semantic Category of Input Image

In order to identify the semantic category of input image, it is necessary to find the local features. To detect and describe local features in image, the SIFT algorithm is used. SVM Classifier is used for classification purpose.

SIFT includes the following sub sections; First, construct scale space, which includes two sub tasks. Take input image and generate blurred out image. Apply Gaussian blur on input image. Second, key point localization, and it also includes two sub tasks. Identify local maxima or minima of DoG images. Compare each pixel in DoG image with its own neighbors and corresponding pixels in the neighboring scale. Third one is discarding low contrast key points. Discard low contrast key points that are sensitive to noise using second order Taylor expansion. Eliminate key points with poorly determined locations and high edge

response using Hessian matrix. Fourth one is Orientation assignment. Calculate gradient magnitude and orientation of image samples. For every pixel, find gradient magnitude and orientation and finally get orientation histograms.

In SVM Classifier, Given a set of labeled training data, then it outputs an optimal hyper-plane which categorizes new examples/ inputs. Following steps are used; initially, obtain the Support Vectors (SVs) closest between each class. Then create decision hyper-plane. The Support vectors closest to each class are then identified. Classification is based on the distance of the vectors from hyper-plane. Semi supervised learning framework is performed by combining SFSMR and Multitask Model. For this process, first we need to find predicted label. Repeat it for all scales. Compute loss function that measure consistence between true and predicted labels. Use L2,1 regularization to avoid over fitting of features.

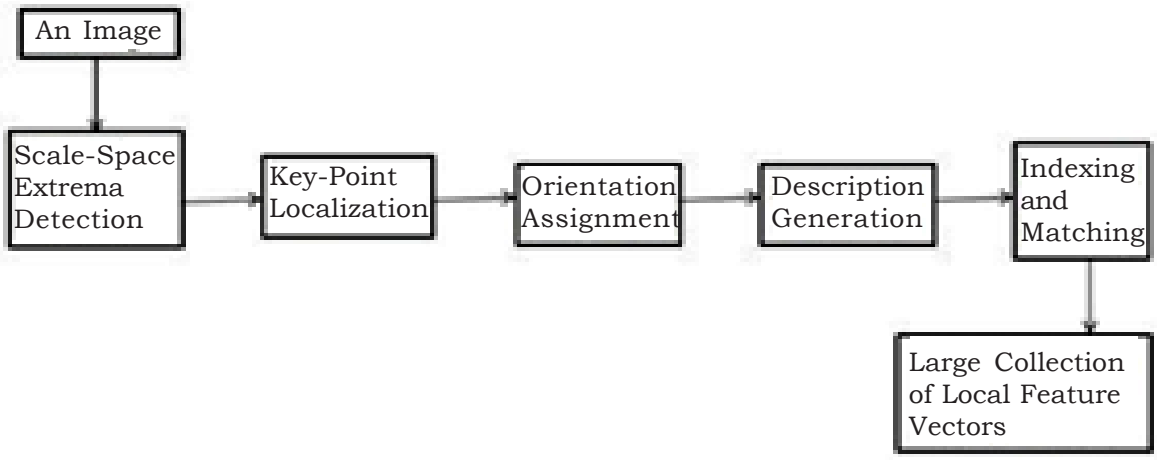

Figure 3. Detect and describe local features by using SIFT

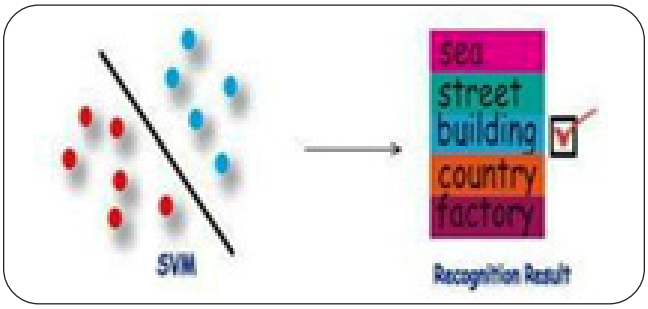

Figure 4. Classification using SVM

3.3 Text Recognition

This section includes the following task to recognize the text present in an input image. Initially the texts regions are extracted to do the skew correction. Next step is to perform the *Binarization* of regions. Then characters are passed into the recognition module and finally recognize the module. In extract text regions step, partition the input image into *m* number of blocks. Identify the 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 the skew correction for which 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 the *binarization* of regions. For this process we need to consider 8 neighbor pixels of each pixel in the text region. Find arithmetic mean of minimum and maximum intensities of text regions which is taken 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. The final module is the 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.

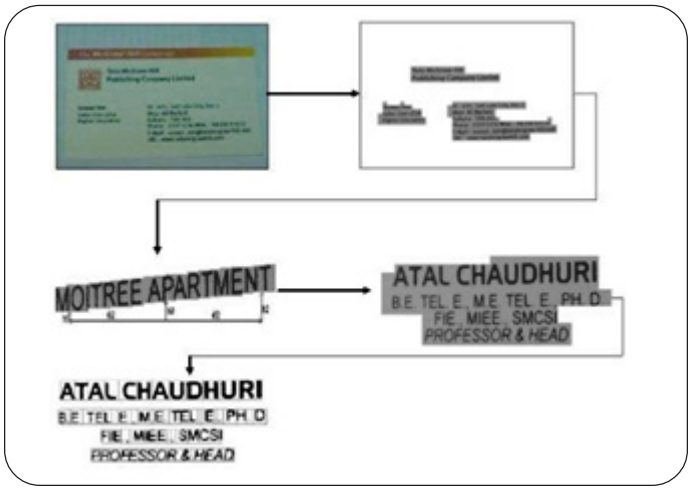

Figure 5. Text Recognition

5. Conclusion

In this paper, a framework for identifying semantic category of a distorted image along with text recognition is proposed. Initially the distorted image is reconstructed using PCA method and identify and extract key features using SIFT method. Semi-supervised learning framework by combining SFSMR and Multitask Model is performed. The Text recognition is done by using the OCR method. Thus, better understanding of scene category is achieved.

References

[1] Mollah, Faruk, Ayatullah., Majumder, Nabamita., Basu, Subhadip., Nasipuri, Mita. (2011). Design of an Optical Character Recognition System for Camera-based Handheld Devices, *IJCSI International Journal of Computer Science*, 8 (4) 1.

[2] Cheng, Chang., Koschan, Andreas., Chen, Chung-Hao., Page, David L., Abidi, Mongi A. (2012). Outdoor Scene Image Segmentation Based on Background Recognition and Perceptual Organization, *IEEE Transactions On Image Processing*, 21 (3).

[3] Zhong, Yanfei, Zhu, Qiqi, Zhang, Liangpei, (2015). Scene Classification Based on the Multi-feature Fusion Probabilistic Topic Model High Spatial Resolution Remote Sensing Imagery, *IEEE Transactions On Geoscience And Remote Sensing*, 53 (11) .

[4] Huang, Yongzhen., Huang, Kaiqi., Wang, Liangsheng., Tao, Dacheng. (2008). Enhanced biologically inspired model, *Computer Vision and Pattern Recognition*, 2008. CVPR 2008. IEEE.

[5] Annabosch, Andrewzisserma. (2008). Scene classification using a hybrid generative approach, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30 (4) (April).

[6] Fei-Fei, Li. A Bayesian Hierarchical Model for Learning Natural Scene Categories, Pasadena, CA 91125, USA

[7] Yao, Jian., Urtasun, Raquel. Describing the Scene as a Whole: Joint Object Detection, Scene Classification and Semantic Segmentation.

[8] Yu, Xiaodong., Cornelia Ferm¨ullery, Ching Lik Teoz, Yangz, Yezhou., Aloimonosz, Yiannis. Active Scene Recognition with Vision and Language, Computer Vision Lab, University of Maryland, College Park, MD 20742, USA

[9] Karami, Ebrahim., Shehata, Mohamed., Smith, Andrew. (2015). Image Identification Using SIFT Algorithm: Performance Analysis against Different Image Deformations, *In*: 2015 Conference on Newfoundland Electrical and Computer Engineering St. John's, Canada.

[10] Torralba, Antonio. (2003). Contextual Priming for Object Detection, January 15, 2003.

[11] Mollah, Faruk, Ayatullah., Majumder, Nabamita., Basu, Subhadip., Nasipuri, Mita. (2011). Design of an Optical Character Recognition System for Camera-based Handheld Devices, *IJCSI International Journal of Computer Science*, 8 (4) 1, (July).

[12] Malag´on-Borja, Luis., Fuentes, Olac. (2005). An Object Detection System using Image Reconstruction with PCA, *In*: Proceedings of the Second Canadian Conference on Computer and Robot Vision (CRV'05)

[13] Vishwanathan, S. V. N., Murty, Narasimha., M. SSVM: A Simple SVM Algorithm.