Read-Speaker Using Stick Camera

P Anand [1], J Priyadharshini[2], I Sagaya Sanjana[3], E Sankavi[4], V Sundhari[5]
[1, 2, 3, 4,, 5] Dept of Information Technology
[1, 2, 3, 4, 5] Saranathan College of Engineering, Trichy-620012, Tamil Nadu, India

Abstract- *In this paper, Detecting text in natural images, as opposed to scans of printed pages, faxes and business cards, is an important step for a number of Computer Vision applications, such as computerized aid for visually impaired and robotic navigation in urban environments. Retrieving texts in both indoor and outdoor environments provides contextual clues for a wide variety of vision tasks. In this project, we implement two processes such as text detection and text recognition. In text detection, use contrast map is then binaries by median filter and combined with Canny's edge map to identify the text stroke edge pixels based on feature extraction. The features extractors are Harris-Corner, Maximal Stable Extremal Regions (MSER), and dense sampling and Histogram of Oriented Gradients (HOG) descriptors. Then implement text recognition. The first one is training a character recognizer to predict the category of a character in an image patch. The second one is training a binary character classifier for each character class to predict the existence of this category in an image patch. The two schemes are compatible with two promising applications related to scene text, which are text understanding and text retrieval. Further we extend this concept with word level recognition with lexicon techniques with accurate results. And also recognition text in real time images, videos and mobile application images.*

I. INTRODUCTION

Text detection and recognition in images and videos is a research area which attempts to develop a computer system with the ability to automatically read from images and videos the text content visually embedded in complex backgrounds

The investigation of text detection and recognition in complex background is motivated by cutting edge applications of digital multimedia. Today more and more audio and visual information is captured, stored, delivered and managed in digital forms. The wide usage of digital media files provokes many new challenges in mobile information acquisition and large multimedia database management.

Among the most prominent are:

1. Automatic broadcast annotation: creates a structured, searchable view of archives of the broadcast content;
2. Digital media asset management: archives digital media files for efficient media management;
3. Video editing and cataloging: catalogs video databases on basis of content relevance;
4. Library digitizing: digitizes cover of journals, magazines and various videos using advanced image and video OCR;
5. Mobile visual sign translation: extracts and translates visual signs or foreign languages for tourist usage, for example, a handhold translator that recognizes and translates Asia signs into English.

II. RELATEDWORKS

In the existing method Reading words in unconstrained images is a challenging problem of considerable practical interest. While text from scanned documents has served as the principal focus of Optical Character Recognition (OCR) applications in the past, text acquired in general settings (referred to as scene text) is becoming more prevalent with the proliferation of mobile imaging devices. Since text is a pervasive element in many environments, solving this problem has potential for significant impact. For example, reading scene text can play an important role in navigation for automobiles equipped with street-facing cameras in outdoor environments, and in assisting a blind person to navigate in certain indoor environments.

In Sampling Strategies for Bag-of-Features Image Classification, Eric Nowak, Frederic Jurie-2000 [7]Bag-of-features representations have recently become popular for content based image classification owing to their simplicity and good performance. They evolved from text on methods in texture analysis. The basic idea is to treat images as loose collections of independent patches, sampling a representative set of patches from the image, evaluating a visual descriptor vector for each patch independently, and using the resulting distribution of samples in descriptor space as a characterization of the image. The four main implementation choices are thus how to sample patches, how to describe them, how to characterize the resulting distributions and how to classify images based on the result. We concentration the first issue, showing experimentally that for a representative

selection of commonly used test databases and for moderate to large numbers of samples, random sampling gives equal or better classifiers than the sophisticated multi scale interest operators that are in common use. Although interest operators work well for small numbers of samples, the single most important factor governing performance is the number of patches sampled from the test image and ultimately interest operators cannot provide enough patches to compete. The four main implementation choices are thus how to sample patches, what visual patch descriptor to use, how to quantify the resulting descriptor space distribution, and how to classify images based on the resulting global image descriptor.[7].

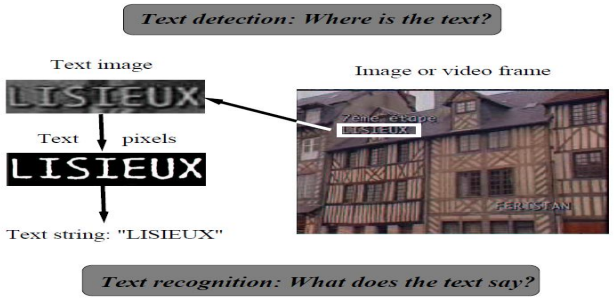


Fig 1.Global and Local Classifiers for Face Recognition by HierarchicalEnsemble

Recognizing Clothes Patterns for Blind People by Confidence Margin based Feature Combination, Xiaodong Yang, Shuai Yuan, and YingLi Tian-2011 [2], we extract both structural feature and statistical feature from image wavelet sub bands. Furthermore, we develop a new feature combination scheme based on the confidence margin of a classifier to combine the two types of features to form a novel local image descriptor in a compact and discriminative format. The recognition experiment is conducted on a database with 627 clothes images of 4categories of patterns Textons are the repetitive basic primitives to characterize a specific texture pattern. Because of the robustness to photometric and affine variations, SIFT is commonly used to capture the structural information of texture. On the other hand, it was observed that the combination of multiple complementary features usually achieves better results than the most discriminative individual feature..

Computer Vision-Based Door Detection for Accessibility of Unfamiliar Environments to Blind Persons, Yingli Tian1, Xiaodong Yang1, and Aries Arditi-2010 [3]The robustness and generalizability of the proposed detection algorithm are evaluated against a challenging database of doors collected from a variety of environments over a wide range of colors, textures, occlusions, illuminations, scale, and views. There have been many efforts to study blind navigation and way finding with the ultimate goal of developing useful travel aids for blind people but very few have met with more than limited success. The most useful and accepted independent travel aids remain the Hoover white cane and the guide dog, both of which have been in use for many years. While GPS-guided electronic way finding aids show much promise in outdoor environments, there is still a lack of orientation and navigation aids to help people with severe vision impairment to independently find doors, rooms, elevators, stairs, bathrooms, and other building amenities in unfamiliar indoor environments.

Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study, J. ZHANG AND M. MARSZA_LEK, S. LAZEBNIK [4] ,A large-scale evaluation of an approach that represents images as distributions of features extracted from a sparse set of key point locations and learns a Support Vector Machine classifier with kernels based on two effective measures for comparing distributions, the Earth Mover's Distance and the $\chi 2$ distance. We first evaluate the performance of our approach with different key point detectors and descriptors, as well as different kernels and classifiers. We then conduct a comparative evaluation with several state-of-the-art recognition methods on four texture and five object databases. On most of these databases, our implementation exceeds the best reported results and achieves comparable performance on the rest.

Finally, we investigate the influence of background correlations on recognition performance via extensive tests on the PASCAL database, for which ground-truth object localization information is available. Our experiments demonstrate that image representations based on distributions of local features are surprisingly effective for classification of texture and object images under challenging real-world conditions, including significant intra-class variations and substantial background clutter. To this end, we conduct a comprehensive assessment of many available choices for our method, including key point detector type, level of geometric invariance, feature descriptor, and classifier kernel. Several practical insights emerge from this process.

*Viewpoint invariant texture description using fractal Analysis, Yong Xu1, Hui Ji-2003l [5]*Image texture provides a rich visual description of the surfaces in the scene. Many texture signatures based on various statistical descriptions and various local measurements have been developed. Existing signatures, in general, are not invariant to 3D geometric transformations, which is a serious limitation for many applications. In this paper we introduce a new texture signature, called the multi fractal spectrum. The MFS is invariant under the bi-Lipchitz map, which includes view-

point changes and non-rigid deformations of the texture surface, as well as local affine illumination changes.It provides an efficient framework combining global spatial invariance and local robust measurements. Intuitively, the MFS could be viewed as a "better histogram" with greater robustness to various environmental changes and the advantage of capturing some geometrical distribution information encoded in the texture. Experiments demonstrate that the MFS codes the essential structure of textures with very low dimension, and thus represents a useful tool for texture classification. The MFS is globally invariant under the bi-Lipchitz transform a very general transform which includes perspective and general texture surface deformations. Furthermore, the MFS has low dimension, and is very efficient to compute. Second, its performance is similar to the top methods in traditional texture retrieval and classification on standard texture datasets, and better on the high resolution dataset.

III. THE PROPOSED METHOD

Information retrieval from video images has become an increasingly important research area in recent years. The rapid growth of digitized video collections is due to the widespread use of digital cameras and video recorders combined with inexpensive disk storage technology. Textual information contained in video frames can provide one of the most useful keys for successful indexing and retrieval of information. Keyword searches for scene text of interest within images can provide additional capabilities to the search engines. Most existing algorithms for text detection were developed to process binary document images and do not perform well on the more complex images. In past years, many different methods have been developed for text detection in color document images by taking advantage of document characteristics. For example, simple edge-based detection filters such as the canny edge detector have been proposed to detect text based on the fact that the text is brighter than the image background. Some methods also make an assumption that the text and background in a local region have relatively uniform gray levels so that the contrast information can be used to extract text. Unfortunately, these techniques are generally not applicable to the complex background found in most images.

Most recently, neural networks have been offered as an alternative method for detecting text in videos. However, training networks and adjusting parameters increase the complexity of the implementation. In contrast to more classical OCR problems, where the characters are typically mono tone on fixed backgrounds, character recognition in scene images is potentially far more complicated due to the many possible variations in background, lighting, texture and font. As a result, building complete systems for these scenarios requires us to invent representations that account for all of these types of variations. Indeed, significant effort has gone into creating such systems, with top performers integrating dozens of cleverly combined features and processing stages. Recent work in machine learning, however, has sought to create algorithms that can learn higher level representations of data automatically for many tasks. Such systems might be particularly valuable where specialized features are needed but not easily created by hand. Another potential strength of these approaches is that we can easily generate large numbers of features that enable higher performance to be achieved by classification algorithms. In this project, we'll apply one such feature learning system to determine to what extent these algorithms may be useful in scene text detection and character recognition.

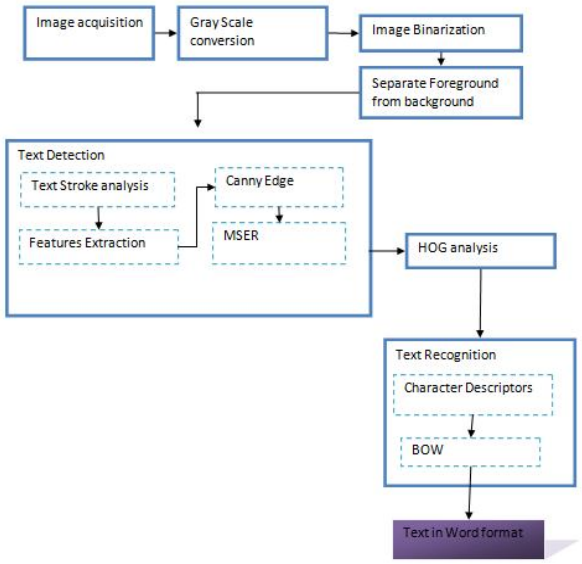
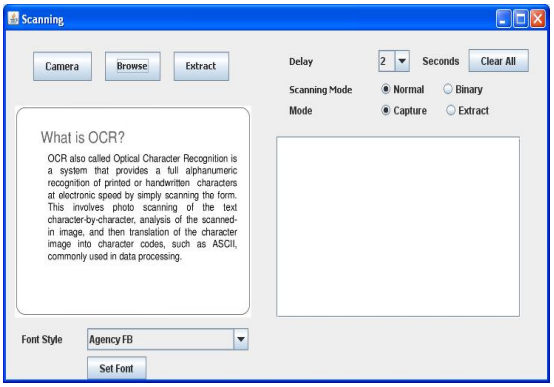


Fig 3.System Architecture

IV. WORKINGMETHODOLOGY

Image Acquisition:

Indexing images or videos requires information about their content. This content is often strongly related to the textual information appearing in them, which can be divided into two groups: Text appearing accidentally in an image that usually does not represent anything important related to the content of the image. Such texts are referred to as scene text. Text produced separately from the image is in general a very good key to understand the image. It is called artificial text. In this module, get images from users through web cameras, captured images, mobile images and so on.



Preprocessing:

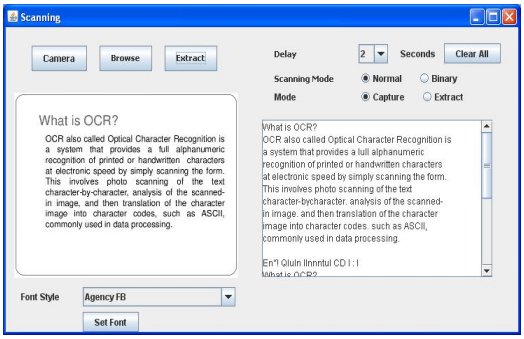
In this project we convert the RGB image into gray scale images. Then remove the noises from images by using filter techniques. The goal of the filter is to filter out noise that has corrupted image. It is based on a statistical approach. Typical filters are designed for a desired frequency response. Filtering is a nonlinear operation often used in image processing to reduce "salt and pepper" noise. A median filter is more effective than convolution when the goal is to simultaneously reduce noise and preserve edges. And implement image binarization tasks. Document Image Binarization is performed in the preprocessing stage for document analysis and it aims to segment the foreground text from the document background. A fast and accurate document image binarization technique is important for the ensuing document image processing tasks.

Text Detection:

In this project, we analyze text strokes using feature extraction algorithms. Such as Canny Edge Detection, Maximal Stable Extremal Regions (MSER),. In computer vision and image processing the concept of feature detection refers to methods that aim at computing abstractions of image information and making local decisions at every image point whether there is an image feature of a given type at that point or not. The resulting features will be subsets of the image domain, often in the form of isolated points, continuous curves or connected regions. Feature detection is a low-level image processing operation. That is, it is usually performed as the first operation on an image, and examines every pixel to see if there is a feature present at that pixel. If this is part of a larger algorithm, then the algorithm will typically only examine the image in the region of the features.

A fast and effective pruning algorithm is designed to extract Maximally Stable Extremal Regions (MSERs) as character candidates using the strategy of minimizing regularized variations. Distance weights and clustering threshold are learned simultaneously using the proposed metric learning algorithm; character candidates are clustered into text candidates by the single-link clustering algorithm using the learned parameters.

Text Recognition:



After detecting text region in the image, from that text region text is extracted from the image using character descriptors and structure configuration. These methods used to convert images with text into editable formats and processes input images with text and get editable documents like TXT file. It employs four types of keypoint detectors, Harris detector (HD) to extract keypoints from corners and junctions, MSER detector (MD) to extract keypoints from stroke components, Dense detector (DD) to uniformly extract keypoints, and Random detector (RD) to extract the preset number of keypoints in a random pattern. At each of the extracted keypoints, the HOG feature is calculated as an observed feature vector in feature space. HOG is selected as local features descriptor because of its compatibility with all above keypoint detectors. In the process of feature quantization, the Bag-of-Words (BOW) Model and Gaussian Mixture Model (GMM) are employed to aggregate the extracted features. BOW is applied to keypoints from all the four detectors. GMM is applied to those only from DD and RD, because GMM-based feature representation requires fixed number and locations of the keypoint all character patch samples, while the numbers and locations of keypoints from HD and MD depend on character structure in the character patches.

In both models, character patch is mapped into characteristic In text retrieval application, the query character class is considered as an object with fixed structure, and we generate its binary classifier according to structure modeling. Character structure consists of multiple oriented strokes, which serve as basic elements of a text character. From the pixel-level perspective, a stroke of printed text is defined as a region bounded by two parallel boundary segments. Their orientation is regarded as stroke orientation and the distance

between them is regarded as stroke width. Histogram as feature representation. Then their corresponding character classifiers are invoked to confirm the character classes. If most of the queried characters exist, the text retrieval application will provide positive response, otherwise provide negative response.

V. CONCLUSION

We have proposed a scene text recognition method using multiple images. Scene text detection is used to extract bounding boxes of text strings from the first frame as initial location of a text region to be tracked. Text tracking further tracks the bounding box to images. Feature extraction is then applied to each tracked bounding box. We employ GMM and Bow model to configure text words, where the output score of feature extraction is used as node features and the frequency of neighboring features are used as edge features. The combination of text extraction and tracking is able to improve efficiency in practical applications. The experimental results validate the effectiveness of our proposed detection based scene text recognition in multiple scene images. We perform groups of experiments to evaluate the effectiveness of our proposed algorithm, including text classification, text detection, and character identification. The evaluation results on benchmark datasets demonstrate that our algorithm achieves the state-of-the-art performance on scene text classification and detection, and significantly outperforms the existing algorithms for character identification.

REFERENCES

[1] Eric Nowak, Fredericjurie“Sampling Strategies for Bag-of-Features Image Classification’’,2000.
[2] Xiaodong Yang, Shuai Yuan, and YingLiTian-,”Recognizing Clothes Patterns for Blind People by Confidence Margin based Feature Combination”vol. 22, no. 8, august 2011.
[3] Yingli Tian1, Xiaodong Yang1, and Aries Arditi"Computer Vision-Based Door Detection for Accessibility of Unfamiliar Environments to Blind Persons", -2010
[4] J. Zhang and m. Marsza_lek, s. Lazebnik. Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study, -2006.
[5] ,Manoj Kumar*, Prof. Sagun Kumar sudhansuWavelet Based Texture Analysis and Classification with Linear Regration Model -2012.
[6] T. de Campos, B. Babu, and M. Varma, “Character recognition in natural images,” in Proc. VISAPP, 2009.
[7] B. Epshtein, E. Ofek, and Y. Wexler, “Detecting text in natural scenes with stroke width transform,” in Proc. CVPR, Jun. 2010, pp. 2963–2970.
[8] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, “Object detection with discriminatively trained part-based models,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 9, pp. 1627–1645, Sep. 2010.
[9] T. Jiang, F. Jurie, and C. Schmid, “Learning shape prior models for object matching,” in Proc. CVPR, Jun. 2009, pp. 848–855.
[10] S. Kumar, R. Gupta, N. Khanna, S. Chaudhury, and S. D. Johsi, “Text extraction and document image segmentation using matched wavelets and MRF model,” IEEE Trans. Image Process., vol. 16, no. 8, pp. 2117–2128, Aug. 2007