World Engineering & Applied Sciences Journal 6 (3): 147-151, 2015
ISSN 2079-2204
© IDOSI Publications, 2015
DOI: 10.5829/idosi.weasj.2015.6.3.22210

Translation of Text Oriented Signboard Images from Mobile Phone Camera

[1]P. Selvi Rajendran and [2]M. Aishwarya

[1]Department of Information Technology, Tagore Engineering College, Chennai, India
[2]Software Engineer, Cognizant Technology Solutions Pvt Ltd, Chennai, India

Abstract: In this paper, a computerized translation system for Hindi signboard images is described. The framework includes detection and extraction of text for the recognition and translation of shop names into Tamil. It handles impediments brought on by different font styles and font sizes,along with illumination changes and noise effects. The main aim of the system for the translation of the texts present on signboard images captured with a cell phone camera. Two Indian languages – Hindi and Tamil are used to demonstrate this kind of system. The prototype system is created using the off-shelf components. It recognizes the texts in the images, provides access to the information in English and Tamil. The framework has been implemented in a cell phone and is demonstrated showing acceptable performance.

Key words: Signboard Translation · Text Translation · Character Recognition · Signboard Images using Mobile Phone · Cross-Lingual

INTRODUCTION

Signboard information plays a significant role within our society. Their format is usually frequently concise and direct and the data they provide is generally very useful. However, the foreign visitors might not understand the language that the signboard is written in, with the consequently loss of most that important information. The gaining momentum of portable cellular devices, the growing of these computational power and the inclusion of digital camera models on them afford them the ability to change from the classical hand dictionary translation to a brand new faster, comfortable and affordable way. In this sense, it's expected that the high percentage of the entire world population will own a cell phone by having an embedded camera, that will be all which our system needs [1].

The acquired image is first transferred to a server, which corrects the perspective distortions, detects recognizes and then translates the writing content. This translated text is sent back once again to the cell-phone in an appropriate form. We have also described here, the Hindi and Tamil OCRs which we use for the character recognition. Nevertheless, considering that the computational resources are limited, our system is founded on non expensive algorithms to be able to achieve the greatest accuracy in the cheapest processing time possible. We also propose methods to really make the recognizer efficient in storage and computation. The translated text, along with any extra information, is transmitted back once again to the user [2].

In the present system images acquired through the camera phones are generally noisy and have complex, cluttered background. Conventional scanners usually provide clean high resolution images with simple background structure. They employ orthographic (parallel) projection for the imaging. Compared to the scanners, camera-phones are of low resolution and follow a perspective projection model. Novel image understanding algorithms are hence needed to deal with the situations produced by camera based digitization. It is expected that a number of these limitations is likely to be overcome soon by the advances in hardware technology and algorithmic innovations. Reading text using a camera phone or any perspective camera is very challenging. Images from such cameras are optically distorted as a result of lens effects. With perspective distortion, character recognition became much more challenging in comparison to scanned documents [3].

Corresponding Author: P. Selvi Rajendran, Department of Information Technology, Tagore Engineering College, Chennai, India.

Fig. 1: Overview of the proposed system

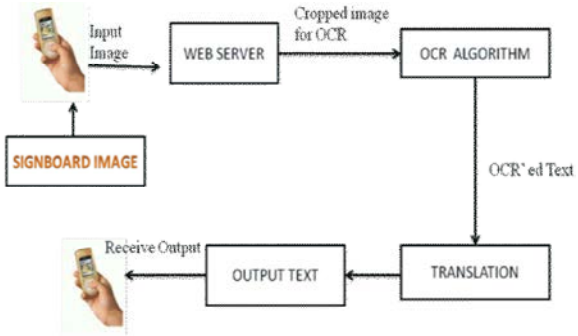

Fig. 2: System Architecture

This paper presents a computerized translation system for signboard images in Hindi language, where the content is writing usually represents a store name or direction of road The system was originally created for tourists from other states of India with little, or no familiarity with the Hindi language. The system consists of text detection, binarization, recognition and translation [4].

First, local clustering can be used to effectively handle the luminance variations of the captured images. The bounding boxes of the individual characters are obtained by connected component analysis [5].

Secondly, the writing content is recognized using direction features extracted from each character region utilizing a non-linear mesh. Finally, a database of shop names has been used to generate a translation result from the set of recognition candidates. The novel concept of this technique is in the translation scheme, by where a database of store names has been incorporated to pay for the possible incorrectness of the recognition result and probably the most probable interpretation of the term is generated by referencing the database. The block diagram of the proposed system is shown in Fig. 2.

Related Work: Recently, several} systems for text extraction from natural scene images have already been developed. Image binarization and text localizing are necessary in such systems. Popular algorithms for image binarization include k-mean clustering and thresholding. Lee *et al*. [6, 7] and Jung *et al*. [8] described approach to clustering in HVC (hue, value and chroma) color space; Lai *et al*. used RBG (red, green, blue) color space. Several} clusters is defined based on the number of the very common colors and is roughly 20. The most used thresholding algorithms include Otsu [3], Niblack and Sauvola thresholding [4].

The very first one may be the global thresholding algorithm, that may hardly be useful for natural scene images. The final two methods are local adaptive algorithms, which show higher quality for image thresholding.

Proposed Method: In the proposed system the images are captured through Camera phone. Camera phone is used to capture images with textual content. This image is then transferred to a central server, where it is processed and recognized. The recognized text is then converted to the language of user's choice and delivered appropriately. Here scanners were used for digitization in document image analysis systems. The captured images were then converted to the textual representation using Optical Character Recognizers (OCR). Cameras offer greater mobility and are easier to use.

Camera phones being ubiquitous can play an important role in serving as a cross-lingual access device. There are two possibilities in processing and accessing text using camera phones. (1) One can develop applications to under- stand the text on the camera platform. Such applications require small footprints and high efficiency to be usable on the mobile device. For most Indian languages, we do not have robust and efficient recognizers. Hence it would be better to develop the application on a server. Communication between the server and the camera-phone could happen through technologies like Multimedia Messaging Service (MMS), Infrared or Bluetooth.

Cropping: In this section, we describe the image processing modules and perspective correction process. We assume that the images of the words have a bounding rectangle. This could be the border of the sign board. For more advanced perspective correction techniques, refer to [2]. Textual image received from the cell-phone is first preprocessed for detection and isolation of boundaries for perspective correction. The received image has a resolution of. This low resolution image is first binarised using an adaptive.

Thresholding algorithm. Adaptive thresholding ensures that the system is robust to changes in lighting conditions. Illumination variation is a serious problem in any camera-based imaging systems. We assume that the image has a large white background that contains the black text. This assumption is reasonable as in many real situations this is available.

For example, signboards and number-plates have distinct white background. In the next step, we look for the presence of the bounding rectangle, as a boundary of the object or known apriori. The boundary is visible as a regular quadrilateral, which gets rectified as a rectangle after the perspective rectification. Planar images in multiple views are related by a homography [5]. Hence images from one view could be transferred to another view with the help of just a homography matrix. The homography matrix is a matrix defined only up to a scaling factor.

Hence it has unknowns. Recovery of fronto-parallel view involves the estimation of the homography to the frontal view. This could be done by using various methods using prior knowledge of the structure of the textual image [2]. A projective homography can be understood to be the product of three components – similarity, affine and projective.We are interested in removing the projective and affine components to obtain a similarity transformed (i.e., translated, rotated and/or scaled) version of the original image. When the text is surrounded by a well defined boundary, the boundary can be extracted to correct for projective deformations using the following procedure [2].

- Identify a pair of parallel lines in the text that intersect in the perspective image.
- Compute the point of intersection of the two transformed parallel lines.
- Find the equation of the line passing through these points of intersection and rectify the image by removing projective component.
- Identify a pair of perpendicular lines and remove the affine components.
- Resultant image is a frontal view (similarity transformed) version of the original image.

Tamil and Hindi Ocrs: A pattern classification system which recognizes a character image is at the heart of any OCR. We employ a Support Vector Machine based classifier for the character recognition [7]. Feature extraction [8] is done using PCA (Principal Component Analysis) for building a compact representation. When samples are projected along principal components, they can be represented in a lower dimension with minimal loss of information.

Recognition using a SVM classifier involves the identification of an optimal hyperplane with maximal margin in the feature space [8]. The training process results in the identification of a set of important labeled samples called support vectors. Discriminant boundary is then represented as a function of the support vectors. The support vectors are the samples near the decision boundary and are most difficult to classify. We build a Directed Acyclic Graph of binary classifiers where is the number of classes. Every component passes through classifiers before being classified. More information regarding the Directed Acyclic Graph approach for character recognition can be found in [7].

In [7], we analyzed the performance of the SVM-based OCR with different Kernel functions for the classification. It was observed that the accuracy of the OCR did not vary significantly with the increase in complexity of the Kernels. We exploit this advantage in building an efficient SVM-OCR which needs much smaller representation for the storage.

With a direct SVM-OCR, the model files (support vectors and Lagrangians [8]) occupy large amount of space; usually of the order of several hundreds of megabytes. Since higher order kernel functions did not improve the accuracy, we restricted our classifier to linear kernels. A single decision boundary is stored in the model file for each node in place of the support vectors. This resulted in a reduction of the size of the model file by a factor of 30. This reduced model file could be ported to a PDA or to devices where storage space is precious. We are presently exploring the option of porting the OCR into a camera-phone.

After classification of the sequence of components, class labels are converted to the UNICODE output specific to a language.

Hindi OCR: Hindi is written in Devanagari script and is spoken by the largest number of people in India. Though various dialects of Hindi exists the underlying script do not vary. The complex script employed by Hindi presents the OCR research community with novel challenges. Words appear together as a single component and characters are often slightly modified by the matras. In [7], we discuss the details of the OCR built for two languages: Hindi and Telugu. During preprocessing, the Sirorekha is

identified and used to separate the upper matras from the remaining word. Horizontal projection profiles are then used to separate the lower matras. The Sirorekha itself is removed to obtain three different zones of the word –upper, middle and lower zones. During training of the classifier system, the upper, lower and the middle components are trained separately. Since the vowel modifiers are separated from the main characters, the total number of classes to be recognized is reduced. During testing, a component is sent to the appropriate classifier depending on its position. Since the upper, lower and middle zones of the word have separate classifiers, a component from one part will not be confused with a component from a different part increasing the accuracy.

Tamil OCR: Tamil is one of the Dravidian languages and is spoken in Tamil Nadu, a southern state of India. We extend our earlier work [7] for the cross lingual access of Hindi and Tamil text. For this, we have developed a Tamil OCR whose details are given here. Tamil shares the same structure of the script as the other south Indian languages with minor differences. The Tamil script is different from Hindi as characters are not connected with vowel modifiers and the vowel modifiers themselves may appear before or after the actual character. The characters in Tamil are formed from two sets of base characters. The first set consists of 12 characters (called uyirezhuthu). These form the vowels of the language. The other set (called meiyezhuthu) consists of 18 characters which form the consonants. The combinations of these produce 216 different characters. Some of the characters that are formed are shown in Figure 3. The columns represent vowels and the rows different characters and their modified forms. These modified characters may be composed of different connected components. These different components may also appear before or after the character.

In some cases, the vowel modifiers appear as a part of the character itself. However the number of different components produced are limited. The reduction in the number of characters in the language, is due to the absence of stronger forms of some consonants.

For example the pronunciations ‘ka’ and ‘ga’ have the same underlying script in Tamil whereas they are represented by different characters in Hindi. The basic language consists of about different classes to recognize. Hence Tamil is simpler to analyze owing to the lesser number of classes present. For the purpose of recognition, characters are treated in their entirety. Features from images are extracted using PCA and classified using SVM’s under the DAG architecture similar to Hindi. Post-processing was done to differentiate between characters that were frequently misclassified. In some cases, positional information of connected components was also used to resolve ties. Such an example is given in Figure 3a. The order of the connected components alone does not suffice in this case. The positions of the two connected components are also used to resolve the ambiguity.

Translation: Given the OCR result, the system translates it word by word, looking in a 21,000 word thesaurus for the desired term. Refer the system architecture in Fig. 1.1. The system was tested with images of city names prepared in both Hindi and Tamil. The city names were printed in fonts Naidunia for Hindi and TM-TTValluvar for Tamil. When images were taken from various angles close to the frontal view, the recognition accuracy of the system was close to 100%. The accuracy of the system was limited only by the performance of OCR for small variations in angle from the frontal view.

To estimate the robustness of the system towards arbitrary angles, we moved the camera around the image with increasing angle from the frontal view. The distance of the camera from the text image was kept constant. For angles approximately upto about 60%, the recognition accuracy was close to 100%.

Comparative Analysis

Text Detection: Text detection methods that operate on compressed images are fast by nature [5]. Yu, Hongjiang and Jain [1] present good examples of how this kind of methodologies can work by just using simple operations (such as sums and comparisons) in a short amount of time, as our system does. Other methods [6] showed the potential of this text detection approach against other promising algorithms that usually require non affordable processing time and make a lot of assumptions about the text nature. All these considerations, plus the fact that most digital images are stored in a compressed form made us decide to base our system on this kind of ideas.

Text Extraction: Following the classification [7], our text extraction methodology belongs to a clustering-based algorithm, which is effective for natural scene images. Threshold based approaches [8] were studied as well, applied either on the luminance images or on each color channel independently, but no satisfactory results were achieved, even if they worked with local thresholds after the text segmentation.

Fig. 3: Formation of characters in Tamil

Fig. 3a: Positional Information

CONCLUSION

In this paper, we proposed a system to translate signboard images taken with a mobile phone camera from Indian Languages to English. Since the computational resources of these devices are limited, we had to use fast, simple and accurate possible algorithms to work in the most common situations. Our system shows some characteristics that make it interesting and deserve further research:

Other systems like Chinese-English translation have been proposed, but no research has been found for Indian Language to English translation of outside signboard texts.

The whole system shows a high robustness under uneven illumination conditions.

All the operations in our system are easy to understand, implement and maintain.

It can work in a very short amount of time which ensures its viability.

REFERENCES

1. Doerman, D., J. Liang and H. Li, 2003. Progress in camera-based document image analysis, Proceeding International Conference of Document Analysis and Recognition, pp: 606-616.
2. Jagannathan, L. and C.V. Jawahar, 2005. Perspective correction methods for camera based document analysis, Proc. First Int. Workshop on Camera-based Document Analysis and Recognition (CBDAR), Seoul, pp: 148-154.
3. Mirmehdi, M., P. Clark and J. Lam, 2001. Extracting low resolution text with an active camera for ocr, in Proc. IX Spanish Symposium on Pattern Recognition and Image Processing, pp: 43-48.
4. Clark, P. and M. Mirmehdi, 2001. Estimating the orientation and recovery of text planes in a single image, Proc. 12th British Machine Vision Conf., pp: 421-430.
5. Hartley, R. and A. Zisserman, 2000. Multiple View Geometry in Computer Vision. Cambridge University Press.
6, Li, H., D. Doerman and O. Kia, 2000. Automatic text detection and tracking in digital videos, IEEE Transaction on Image Processing, 9(1): 147-167.
7. Jawahar, C.V., MNSSK Pavan Kumar and S.S. Ravikiran, 2003. A bilingual OCR system and its applications, Proc. Int. Conference of Document Analysis and Recognition, pp: 408-413.
8. Duda, R.O., P.E. Hart and D.G. Stork, 2002. Pattern Classification. John Wiley and Sons.