An Improved Text Extraction Technique based on Linear Transformation

Lama Rajab , Hiba Mohammad, Huda Karajeh and Tahani Al Khatib

Department of Computer Information Systems, King Abdullah II School for Information Technology
The University of Jordan, 11942 Amman, Jordan
Lama.rajab@ju.edu.jo, h.khadrawi@ju.edu.jo, h.karajeh@ju.edu.jo, tahani.khatib@ju.edu.jo

Abstract: Finding the textual contents in images is a challenging and promising research area in information technology. Consequently, text detection and recognition in multimedia had become one of the most important fields in computer vision due to its valuable uses in a variety of recent technical applications. In this paper, we propose an improved text detection technique which starts with employing image enhancement to remove small and thin lines from the background and to increase the contrast on the text region. The resulted enhanced image passes next into a set of morphological operations for edge extraction; later, text candidates are labeled, and Hough transform is applied for text extraction at the end of the process. The effectiveness of the algorithm is tested quantatively and qualitatively under a set of scene images from KAIST dataset. The results show that the proposed technique is robust in detecting the text from scene images with high average of Precision, Recall, and F-Score which are 88.1%, 96%, and 91.5%, respectively.

[Lama Rajab, Hiba Mohammad, Huda Karajeh and Tahani Al Khatib. **An Improved Text Extraction Technique based on Linear Transformation.** *Life Sci J* 2014;11(7):83-88]. (ISSN:1097-8135). http://www.lifesciencesite.com. 11

Key Words: Text Detection, Morphological Operations, Text Regions Labeling, Hough Transform.

1. Introduction

In the recent globalized world, digital contents are developing rapidly over both Internet and digital media which resulted in generating numerous electronic recourses. These resources are available in a variety of multimedia forms such as images, video and audio frames. Accordingly, the old fashion paper-based products such as books, letters, journals and newspapers are converting recently towards digitizing; as a result, one might find several digital images including some textual contents.

Images that contain texts might be either scene images where the text appears naturally as a part of the scene, or caption text images where the text overlays on the image (Sumathi C P et al, 2012). Working with natural (scene) images with complex backgrounds can be valuable in some computer and robotics applications. Moreover, Text detection in images such as advertisements, street or traffic signs might help in vehicle license plate recognition or in text reading programs for visually impaired person (Jung K et.al,2004).

Detecting texts in images, as well, helps in various technical applications such as keyword-based image search, automatic video logging, and text-based image indexing (Jung K et.al, 2004). The extracted text itself may include significant information about the image contents which may help in media elucidation.

Caption texts in news are examples of artificial texts that were added to the video frames in order to present a clearer understanding for the viewers. Text detection and extraction, however, is a tricky issue due to many reasons. The variety of fonts, sizes, styles, orientations and shadows might complicate the text detection process as well as the complexity in the image background (Rama G et.al,2010) and (Jain A K et.al, 1998).

In this paper we improved a text detection technique that applies image enhancement to improve the edge detection along with some well known morphological operations to locate more accurate edges on the scene image as a first step. In the second step, the technique uses the Hough transform for each connected component, and then the text is extracted by selecting those connected components whose number of peaks is greater than a discriminating threshold value.

The paper is organized as follows: section 2 reviews some previous research on text detection field. Then, in section 3, the proposed technique is presented. Section 4 presents an analysis and discussion of the technique results. Finally, the conclusion of the paper is presented in section 5.

2. Related Work

Text detection in images is an active and interesting research field on account of its valuable implementations in many technical branches. Several algorithms and studies have been proposed and developed in text detection researches.

The proposed algorithms and methods were mainly based on edge detection, projection profile analysis (Gllavata J et.al, 2003), texture segmentation, morphological operators (Rama G, et.al, 2010) and (Vaghela K and Patelothers N, 2013), color quantization and histogram techniques, artificial neural network (Coates A, et.al,2011) and wavelet transforms (Ye Q et.al, 2005), (Sumathi C P et al, 2012), (Gao J and Yang J, 2011), (Patel M et.al, 2013) and (kumar A, 2013).

In (Rama G, et.al, 2010), authors proposed a text detection algorithm that uses thresholding morphological operators for edge detection and text candidates' labeling. Similarly, Authors in (Ye Q et.al, 2003) applied some morphological operators on the image; however they also studied wavelet-based features in order to label text candidates.

A grayscale image segmentation algorithm was applied in the proposed technique in (Patel M et.al, 2013), authors employed Haar wavelet transform (DWT) on the resulted image and then studied the connected components' features to find candidate text areas. Authors in (Li Y et.al, 2006) aimed to detect hand written texts by using a Gaussian window for blurring and enhancing text line structures.

The technique proposed in (Gao J and Yang J, 2001) employs multi-scale edge detection and an adaptive color modeling in the neighborhood of text candidate regions, it then performs a layout analysis as the final step for labeling text candidates. On the other hand, the technique in (Vaghela K and Patelothers N, 2013) detects the text in an image by using morphological operations, and then it labels the connected components by electing some criteria to filter non text regions, finally it impaints the text from the original image by removing the detected text.

In (Li Y and Lu H, 2012) and (Epshtein B et.al, 2010) text detection was based on evaluating and analyzing the stroke width generated from the skeleton of the text candidates. (Wu V et.al, 1997) and (Wu V et.al, 1999) use Stroke generation also in the chip generation step after segmenting texture; strokes are connected into chips if they compose connected components with certain conditions, where each of which will be tested and classified into a text or non-text regions.

In (Lee J-J et.al, 2011), the proposed system is based on the Modest AdaBoost algorithm which constructs a strong classifier from a combination of weak classifiers. However, in (Coates A, et.al, 2011), the algorithm applies some recently developed methods in machine learning which uses unsupervised feature learning for text detection and recognition.

The technique in (Zhao M et.al, 2010) applied wavelet transform to detect edges, then text candidates were obtained and refined by using two learned discriminative dictionaries, adaptive run-length smoothing algorithm and projection profile analysis.

3. Materials and Methods.

In this paper we improved a technique for automatic text detection in scene images. The proposed technique is divided into four phases which are:

- Image Enhancement.
- Edge Extraction.
- Labeling the text candidate regions.
- Text Extraction.

The flowchart for the proposed technique is shown in figure1.

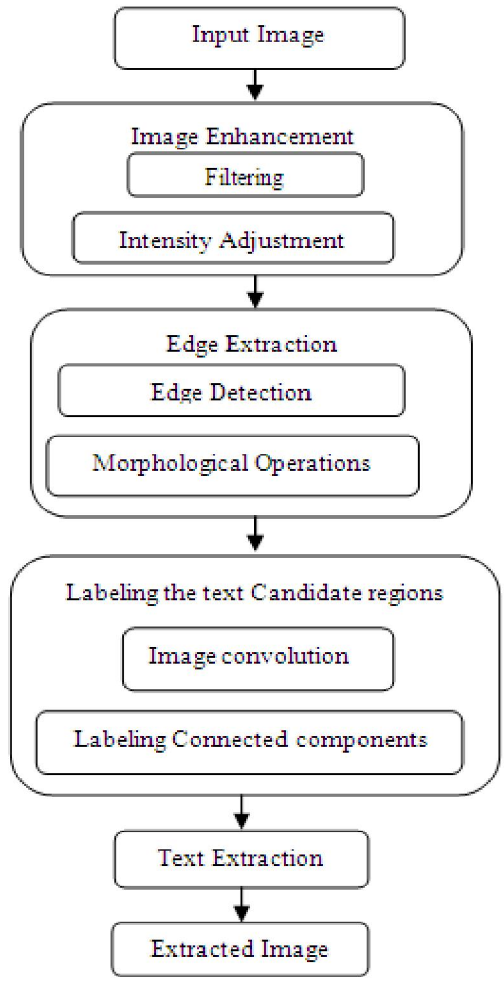


Figure 1 : Flowchart for the proposed technique.

3.1 Image Enhancement

This phase takes account of enhancing the visual appearance of scene images in order to improve the detectability of objects and edges that will be used in the Edge Extraction phase. The Image Enhancement phase includes the following steps:

Step 1: Get (Input) the colored scene image, see figure 2(a).

Step 2: Convert the input image into grayscale.

Step 3: Apply two filtering techniques: *Winner filter* and *Median filter* to remove the additive white noise.

Step 4: Apply intensity adjustment followed by *Unsharp* filter.

Step 5: Show (Output) an enhanced grayscale image as shown in image 2(b).

3.2 Edge Extraction

In this phase an effective morphological edge detection scheme is applied on the enhanced image in order to find the edges and lines. This phase includes the following steps:

Step 1: Apply *Sobel* filter on the enhanced image resulted from phase 1 to get the edge map for the image.

Step 2: Apply a set of morphological operations on the edge map from step 1 as follows:

2.1 Apply *open* operation on the edge map.

2.2 Apply *thresholding* on the resulted image from step 2.1 to remove minor non-text regions.

2.3 Apply the *dilation* operation with the suitable structured element to connect isolated edges of each detail region.

2.4 Apply *open* operation on the dilated image.

2.5 Apply *close* operation on the dilated image.

2.6 Find the *average* image of the two images resulted from steps (2.4) and (2.5).

The output from this phase is shown in figure 2(c).

3.3 Labeling the text candidate regions

In Labeling the text candidate regions phase, a Two-dimensional image convolution is applied on the edge extracted image from phase 2; then the connected components in the resulting image are labelled as shown in figure 2(d).

3.4 Text Extraction

Text extraction phase involves extracting relevant text from the image. Hough transform is a technique which can be used to isolate features of particular shapes within an image. In the proposed technique we apply Hough transform on each connected component, and then the number of peaks for each connected component is computed.

Through this research, we applied many experiments on text images; most of them showed that the number of peaks in non-text regions is higher than any text region since text regions are more homogenous. Accordingly, all connected components with number of peaks greater than a certain threshold are eliminated from the image since they are assumed to be non-text regions. The final step in the text extraction phase is the (filling) operation which results in producing the extracted image as shown in figure 2(e).

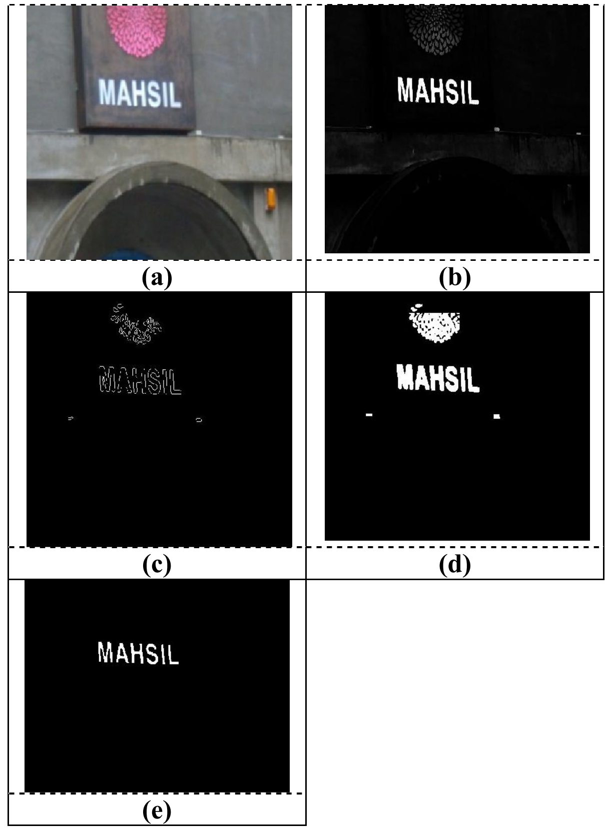


Figure 2: An outdoor image from KAIST dataset in all phases using the proposed technique. (a) The outdoor image. (b) The enhanced image after the image enhancement phase (c) The image after the edge extraction phase. (d) The image after labeling the text candidate regions. (e) The extracted image.

4. Results and Analysis

An evaluation of our proposed text detection algorithm is presented in this section. Based on the used material and experiments, we will discuss the performed measurements and the results analysis in the subsequent sections.

4.1 Materials and Methods

To demonstrate the performance of the proposed algorithm, different types of images are tested from the KAIST dataset. This dataset is developed by the Artificial Intelligence and Pattern Recognition Lab in Korean Advanced Institute of Science and Technology. KAIST dataset consists from scene text images that are arranged in different subsets; each one is captured in a different environments: indoors, outdoors, light, shadow, and night. Dataset images are captured either with a high resolution camera or with mobile phone; all have been resized to 640×480. In addition, is classified based on the language: English (text and numbers), Korean, and mixed of English and Korean language. We have selected a set of scene text images of English language with different properties such as (font size, colour, alignment, and orientation) from this dataset to test the performance of our proposed technique under different conditions.

4.2 Procedure

The proposed approach is implemented and evaluated against the ground truth data under a set of images that is selected from different data set under different conditions and with different properties. The availability of ground truth data gives us a good opportunity to compare the proposed text detection results to the ground truth text area qualitatively and quantitatively. Figure 3 shows an example on a scene image and its ground truth from KAIST dataset.

4.3 Quantitative Metric

The effectiveness and the robustness of the proposed text detection technique are measured quantitatively by calculating three metrics: *Precision*, *Recall,* and *F-Score*. These metrics are calculated based on computing the number of corresponding matched text between the ground truth and detected text area in the image. Therefore, we need to calculate three measurements firstly which are true positive *tp*, false positive *fp*, and false negative *fn*. True positive *tp* represents the number of pixels that are truly classified as text, and false positive *fp* represent the number of pixels that are falsely classified as text while it is actually a background. False Negative *fn* represents the number of pixels that are falsely classified as background while it is actually a text. Depending on these measurements, *Precision*, *Recall,* and *F-Score* are calculated as follows:

$$Precision = \frac{tp}{tp + fp} \ldots\ldots\ldots..(1)$$

$$Recall = \frac{tp}{tp + fn} \ldots\ldots\ldots\ldots(2)$$

$$F - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \ldots..(3)$$

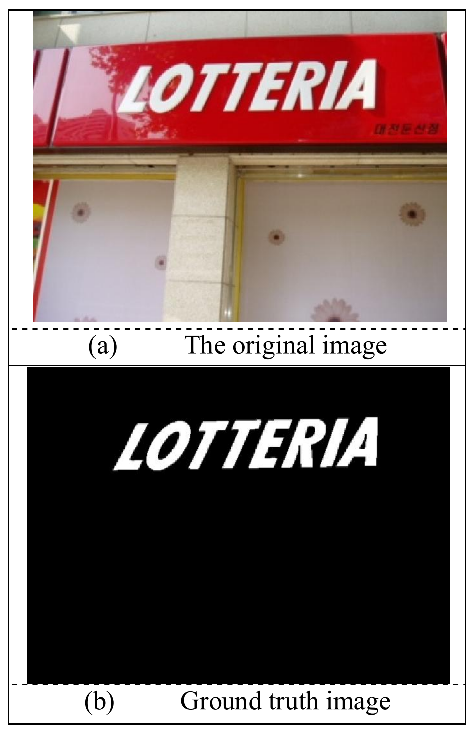
4.4 Results and discussion

The proposed technique is tested on a set of selected scene images from the KAIST dataset with different properties such as colour, orientation, alignment, and font size and type that captured in different environment. A part of these scene images are depicted in Figure 4 with its detected text area. This Figure illustrates in each row the output of the proposed technique on different environment: light, outdoor, indoor, and night. The first row shows the performance of the proposed technique on two images that affected with a light, while the second, third and the fourth rows show the performance of the proposed technique on two images that captured in outdoor, indoor, and night, respectively.

Noticeably, the obtained results from the proposed technique are accurate and consistent under different environment. As we show in figure 4 (b), the image is affected with a strong light, however, the proposed technique detects the large characters accurately and the majority of the small ones (except I and T).

Additionally, and as shown in Figure 4, the proposed technique proved to be robust in detecting the text accurately from these images with different properties (colour, orientation, alignment, and font size and type). Unfortunately, the proposed technique fails in detecting some part of text in dark spots in the lighten image, it also falsely detects the highlighted spots in the image as text. Generally, the proposed technique fails in detecting the text area in shadow images due to the inconsistency of light in it.

The Precision, Recall, and the F-Score measurement that obtained from comparing the output image from the proposed technique and the ground truth image is calculated for a set of images from KAIST dataset . The average of Precision, Recall, and F-Score on this set of images is 88.1%, 96%, and 91.5% respectively. Table 1 shows the precision, Recall, and F-Score for each image that is shown in Figure 3.



(a) The original image

(b) Ground truth image

Figure 3: one of the KAIST images and its ground truth.

Table 1: The Precision, Recall and the F-score for the images shown in Figure 4.

Image Caption	Precision	Recall	F-Score
A	88.1%	99.9%	93.6%
B	73.9%	98.7%	84.5%
C	84.3%	88.0%	86.2%
d	96.1%	98.1%	97.1%
e	100.0%	80.2%	89.0%
f	87.5%	99.4%	93.0%
g	96.6%	96.6%	96.6%
h	84.5%	100.0%	91.6%

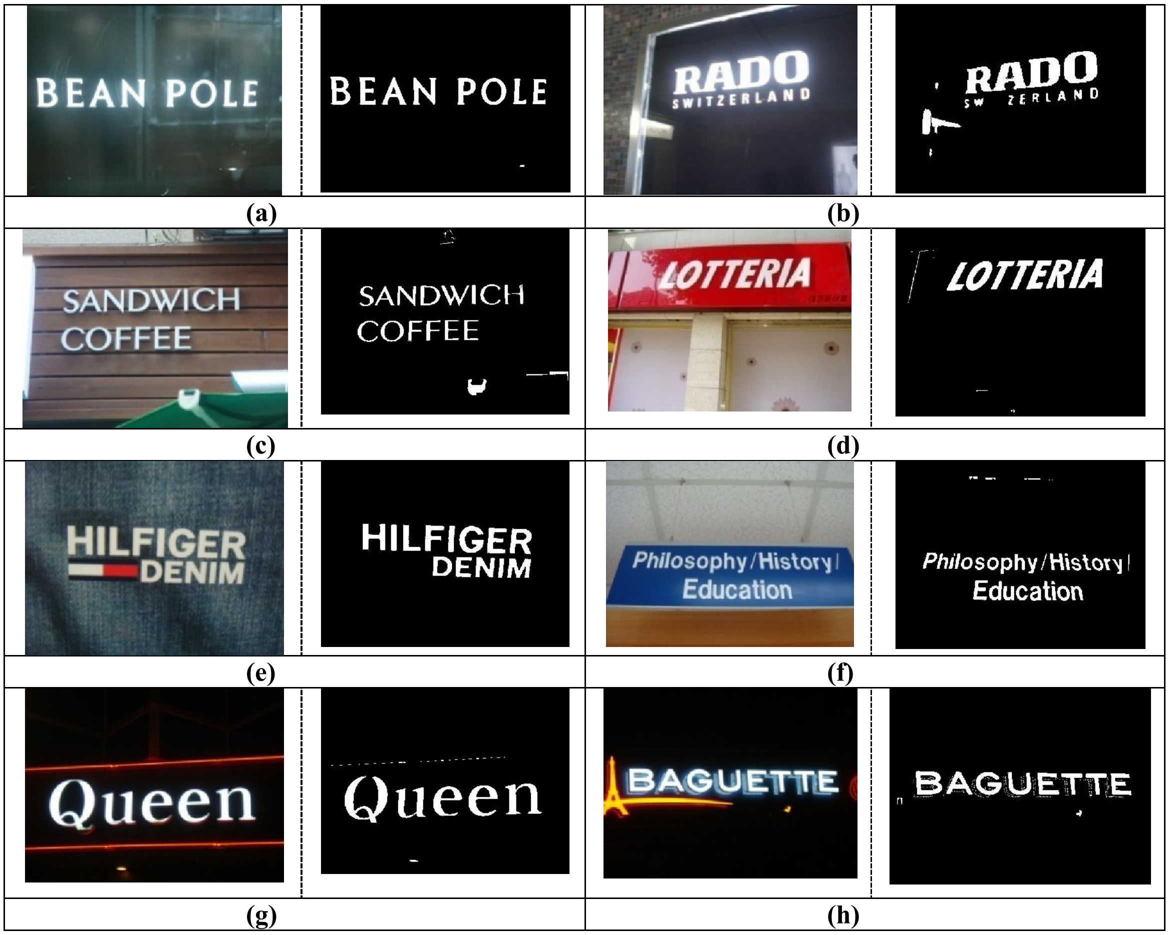


Figure 4: shows a set of images from KAIST dataset with detected text areas using the proposed technique. (a) and (b) are two images that captured on a light environment. (c) and (d) are two images that were captured on outdoor. (e) and (f) are two images that were captured on indoor environment. (g) and (h) are two images that were captured at night.

5. Conclusion

In this paper, we propose an improved text detection technique which is based on linear transformation for text extraction. The improvement is represented by using pre-processing steps to adjust the edges on the image before detecting text using morphological operation. A Hough linear transformation is used to extract the text regions from the resulted candidate text from the previous steps.

The performance of the proposed technique is tested under a set of images from KAIST dataset which were captured in different environments (light, outdoor, indoor, and night). In addition, the quantative metrics (Precision, Recall, and F-Score) are used to test the accuracy of the detection rate for the proposed technique. The results show that the proposed technique is robust in detecting the text from scene images with diverse properties and characteristics and which were captured in different environments. Moreover, the proposed technique extracts the text with high average of Precision, Recall, and F-Score which are 88.1%, 96%, and 91.5%, respectively.

References:

1. Coates A,Carpenter B, Case C, Satheesh S, Suresh B, Tao Wang, Wu D J and Ng A Y. **Text detection and character recognition in scene images with unsupervised feature learning**. 2011 International Conference on Document Analysis and Recognition (ICDAR), IEEE. 2011; 440 – 445.
2. Epshtein B, Ofek E and Wexler Y. **Detecting Text in Natural Scenes with Stroke Width Transform**. Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. 2010; 2963 – 2970.
3. Gao J and Yang J. **An Adaptive Algorithm for Text Detection from Natural Scenes**. Computer Society Conference on Computer Vision and Pattern Recognition, IEEE. 2001; 2: II-84.
4. Gllavata J, Ewerth R, and Freisleben B. **A Robust Algorithm for Text Detection in Images**. proceedings of 3rd of international symposiums on image and signal processing and analysis. 2003; 2: 611-616.
5. Jain A K and Yu B. **Automatic text location in images and video frames**. Pattern Recognition. 1998; 31(12): 2055-2076.
6. Jung K, In Kim K and Jain A K. **Text Information Extraction in Images and Video: A Survey.** Pattern Recognition, Elsevier. 2004; 37(5): 977-997.
7. kumar A, Ahuja P and Seth R. **Text extraction and recognition from an image using image processing in Matlab**. Conference on advances in communication and control systems, 2013; 2(4): 429-435.
8. Lee J-J, Lee P-H, Lee S-W, Yuille and Koch C. **AdaBoost for Text Detection in Natural Scene.** 2011 International Conference on Document Analysis and Recognition (ISDAR), IEEE. 2011; pg 429- 434.
9. Li Y, Zheng Y, Doermann D, and Jaeger S. **A New Algorithm for Detecting Text Line in Handwritten Documents**. Tenth International Workshop on Frontiers in Handwriting Recognition, La Baule (France). 2006; 1: 35-40.
10. Li Y and Lu H. **Scene Text Detection via Stroke Width**. The 21st International Conference on Pattern Recognition (ICPR 2012), IEEE. 2012; 681-684.
11. Patel M and Dodiya A. **Detect and localize text in natural images using wavelet transforms.** International Journal of Science and Research (IJSR). 2013; 2(5): 190-194.
12. Rama Mohan Babu G, Srimaiyee P and Srikrishna A. Text **extraction from heterogonous Images using mathematical morphology**. Journal of theoretical and applied information technology JATIT. 2010; 16(1):39-47.
13. Sumathi C P, Santhanam T and Devi G G. **A survey on various approaches of text extraction in images**. International journal of computer science and engineering survey (IJCSES). 2012; 3(4): 27-42.
14. Vaghela K and Patelothers N. **Automatic text detection using morphological operations and inpainting**. International journal of innovative research in science, engineering and technology IJIRSET. 2013; 2 (5): 1333-1336.
15. Wu V, Manmatha R and Riseman E M. **Finding Text in Images**. Proc. of ACM International Conference on Digital Libraries. 1997; 3-12.
16. Wu V, Manmatha R and Riseman E M. **TextFinder: An Automatic System to Detect and Recognize Text In Images**. IEEE transactions on pattern analysis and machine intelligence. 1999; 21(11): 1224-1229.
17. Ye Q, Gao W, Wang W and Zeng W. **A robust text detection algorithm in images and video frames**. In Information, Communications and Signal Processing, 2003 and Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on, IEEE. 2003; 2: 802-806.
18. Ye, Q, Huang Q, Gao W and Zhao D. **Fast and robust text detection in images and video frames**. Image and Vision Computing. 2005; 23(6), 565-576.
19. Zhao M, Li S and Kwok J. **Text detection in images using sparse representation with discriminative dictionaries**. Image and Vision Computing. 2010; 28(12): 1590-1599.

4/8/2014