Enhance Text Recognition by Image Pre-Processing to Facilitate Library Services by Mobile Devices

Chuen-Min Huang, Yi-Ling Chuang, Rih-Wei Chang and Ya-Yun Chen
Department of Information Management, National Yunlin University of Science and Technology, 123, Sec. 3 University Road, Douliou, Yunlin, Taiwan

Keywords: Image Pre-processing, Text Recognition, Optical Character Recognition (OCR), M-Library Services, Mobile Devices.

Abstract: Facing the popularity of web searching, libraries continuously invest in the provision of online searching and refurnish physical facilities to attract users during the past decades. In this study, we conducted a technical feasibility study to facilitate library services by applying a novel image pre-processing technique to enhance performance of OCR via mobile devices. In the binarization stage, a grayscale image is usually assigned a global threshold value to be binary, while this will not be suitable for some scenarios, such as non-uniform lightness and complicated background. Instead of segregating the grayscale image into many regions like other studies, our approach only partitioned an image into three equal-sized horizontal segments to identify the local threshold value of each segment and then restored the three segments back to the original state. The experimental results illustrate that the proposed method efficiently and effectively improves the text recognition. The accuracy rate was raised from 17.7% to 72.05% of all test images. Without counting eight unrecognizable images, the average accuracy rates of our treatment can reach 90.06%. To compare with other studies we conducted another evaluation to examine the validity of our approach. The result showed that our treatment outperforms most of the other studies and the performance achieves 74.6% in precision and 80.2% in the recall.We are confident that this design will not only bring users more convenience in using libraries but help library staff and businessmen to manage the status of books.

1 INTRODUCTION

Mary Meeker, an expert on network analysis, unveiled her annual Internet Trends report for 2013 showing robust growth in the number of broadband and mobile users. The mobile usage continued to grow, making up about 15 percent of all Internet traffic, compared with 10 percent a year before(Lawler, 2013). In addition to the growing penetration of mobile devices, the reduced price of it has made a lot of demands.

Optical Character Recognition (OCR) is frequently used to identify the text on books or in documents(Mori et al., 1992). With the explosion of mobile devices, the need of text recognition in an image, such as signs or posters is continuously rising(Minetto et al., 2011). Recently, text recognition via mobile devices has received a great deal of attention from many researchers. If text could be directly recognized from an embedded camera, it would lead to a diversity of new applications and yield enormous benefits for users. For example, as a user simply snaps a photo of a company signboard the imaging connection can instantly recognize the name and return relevant information about the company.

Many terms have been used to describe images, including pictures, graphics, drawings, photographs, digitized data, and visual resources. Here, an image is defined as a pictorial representation that conveys visual or abstract properties. An image usually consists of background context and foreground texts. Due to the variety of background components, for example, different kind of color, texture, or brightness in an image, text visibility is hence seriously affected. Notwithstanding the prominent improvement of the OCR technique at the present stage, there are still problems that would affect the result of recognition including the complicated background and non-uniform lightness situations

Huang C., Chuang Y., Chang R. and Chen Y..
Enhance Text Recognition by Image Pre-Processing to Facilitate Library Services by Mobile Devices.
DOI: 10.5220/0004818504530460
In *Proceedings of the 6th International Conference on Agents and Artificial Intelligence* (ICAART-2014), pages 453-460
ISBN: 978-989-758-015-4

(Liu and Samarabandu, 2006, Chowdhury et al., 2009, Kim et al., 2009). Some studies utilized machine-learning based method or edge detection approach to solve the problems, while those methods mainly work on the images of text with the nearly uniform background. (Raza et al., 2001, Shutao and Kwok, 2004, Ye et al., 2005). We considered text extraction from natural scene images shall be a challenging problem. To tackle the problem, we focused on the image pre-processing to deal with the stains and unwanted objects taken from the shooting. Different from general procedure of segregating the grayscale image into many regions, our approach partitioned an image into three equal-sized horizontal segments to identify the local threshold value of each segment. An adaptive threshold operation was proposed to deal with the problem of non-uniform lightness. Further, we applied the conditional connected-component to discard the unimportant or redundant background via condition setting. The result showed that the accuracy rate of recognition was raised and our treatment outperforms most of the other studies.

To test the feasibility of our experiment in a real setting, we gained a permission from the library of National Yunlin University of Science and Technology in Taiwan to access their online catalog via Z39.50 protocol from October 1st to December 31, 2011. Our system successfully connected to libraries when the user simply pressed the Internet connected photo button from their mobile devices to capture the title from a physical book or a poster. The result was displayed by indicating the location of the book in the library or the e-book stores if the book does not exist in the library. It is expected to increase the usability of library services, furthermore, promote the Internet marketing with our design.

The remaining part of the paper is outlined as follows. Section 2 describes and reviews the techniques applied in this study. In section 3, the research methods and experimental results are illustrated. Conclusions and future work are presented in section 4.

2 APPLIED TECHNOLOGY

2.1 Image Binarization

One critical perspective of image pre-processing is the noise filtering. To reduce noise in an image, a couple of studies have applied in the smoothing filter to solve the geometric distortion (Pei-Jun and Effendi, 2010). There are two common types of filter: one is “smoothing filter” and the other is “sharpening filter”. The principle of smoothing filter is to average the grayscale values of mask operation. Then, the average value will be a substitute for all the corresponding pixels to eliminate noise. Sharpening filter functions as image sharpening and contrast enhancement. The major purpose of image sharpening is to magnify edges or regions, consequently the contour of an image can be more significant after sharpening.

There are many kinds of filters, such as, low-pass, median filter, or high-pass. Low-pass filters and median filters are used most often for noise suppression or smoothing, while high-pass filters are typically used for image enhancement (Taisheng et al., 2012). after performing contrast enhancement. We did not adopt high-pass filters because in high brightness situation, the inner details in an image will totally disappear.

Median filter has the merit of preserving edges as wel as removing noise, while the computational effort on calculating the median of each window is huge. For mobile devices, the efficiency of calculation is a critical factor in determining how fast the application can run. For this reason we adopted the low-pass filter in the experiment.

2.2 Object Identification

Connected-component labeling is often used to detect the similar set of intensity values of connected regions in an image (Marqués and Vilaplana, 2002). The connected components can be labelled to identify objects and describe their location, height, width, or density information. Figure 1 displays the connected-component labeling. For a binary image, we can label the white pixels with a unique symbol to find interesting objects in an image.

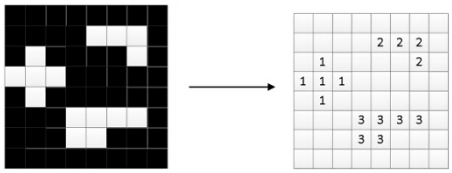


Figure 1: Connected-component labeling.

Mathematical morphology proposed by G. Matheron and J. Serra in 1989 (Besag, 1989) is used to extract the shape feature of a graph and reduce noise. It includes four operators: dilation, erosion, open, and close. Dilation and open operators are usually used to launch the edge enhancement and fill the broken pixel. Erosion and closure operators are frequently

used to carry out the noise reduction and erase the weak edge. In the operation of morphology, it will firstly define a structure element as a small set or a sub-image, and its size is often set as a 3*3 mask (Shuqing and Qiaoning, 2006). Figure 2 (a) describes the process of dilation. If the center pixel is white, the neighbors will change into white. Figure 2 (b) describes the process of erosion. If the center pixel is white, the neighbors will change into black. During the operation, if it is open it will perform the dilation first and then erosion. On the other hand, if it is close, it will firstly perform the erosion and then dilation.

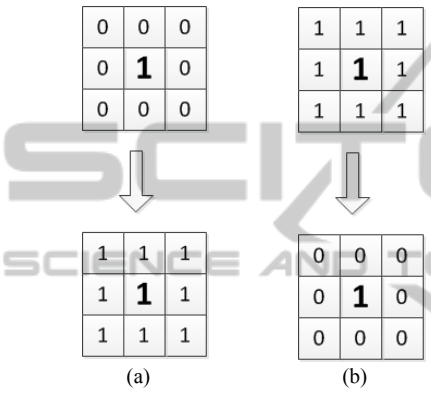


Figure 2: Dilation and Erosion of morphology.

2.3 Optical Character Recognition

Many applications about OCR have been proposed in 1970s. OCR software was first used by libraries for historic newspaper digitization projects in the early 1990's . The most successful application of OCR is applied in the fields of pattern recognition, artificial intelligence and computer vision. Even though digital cameras are fast, versatile, mobile, and convenient, they also suffer some drawbacks, such as geometrical distortions (e.g., when photographing a thick volume), focus loss and uneven document lighting(Bieniecki et al., 2007). Notwithstanding the progress of OCR technique at the present stage, the ways to improve the accuracy rate of recognition are still an intriguing issue to explore (Holley, 2009).

2.4 Mobile Library

With the trend toward digital libraries, paperless is one of the ways of resource conservation, for example, “The Library without a Roof Project” initiated by South Alabama University Library in November, 1993 was the first systematic effort to connect a PDA to library online public access catalogs, commercial online databases, and the Internet (Foster, 1995). It has been well recognized and implemented in many aspects (Rong-Yuh, 2002).

In recent years, the rapid development of mobile devices provides a good platform for the expansion of library services for the readers. A recent investigation of how library users access and interact with information when they are on the move found that the most needed services include opening hours, contact information, location map, library catalog, and borrowing records (Mills, 2010). Therefore, using WAP Web technology to design and implement a more convenient mobile library system become possible.

3 EXPERIMENT AND RESULT

To test the feasibility of our experiment in a real setting, we gained a permission from the library of National Yunlin University of Science and Technology in Taiwan to access their online catalog via Z39.50 protocol from October 1st to December 31, 2011 and obtained part of the layout of the floor plan and bookshelves of the libary. Figure 3 shows the system architecture of this study. Special attention will be paid on image pre-processing in the pattern recognition module.

3.1 Communication Module

This model is activated as users press an “upload” button to submit images from their mobile devices.

3.2 Pattern Recognition Module

Upon receiving the image from communication module, the color image will be converted into grayscale to reduce the computational complexity as well as memory requirements. After that, a series of steps including image binarization as Figure 4 and text extraction as Figure 5 will be processed. The OCR dynamic link library designed by Microsoft Office Document Imaging was applied in this experiment.

3.2.1 Grayscale Transformation

To convert a grayscale image into binary, the original image will be applied a threshold value T as the cutting point, then change the pixels into black or white based on the grayscale value of each pixel as (1).

$$\text{Pixel Color} = \begin{cases} Grayscale\ value > T, white \\ Grayscale\ value \leq T, black \end{cases} \quad (1)$$

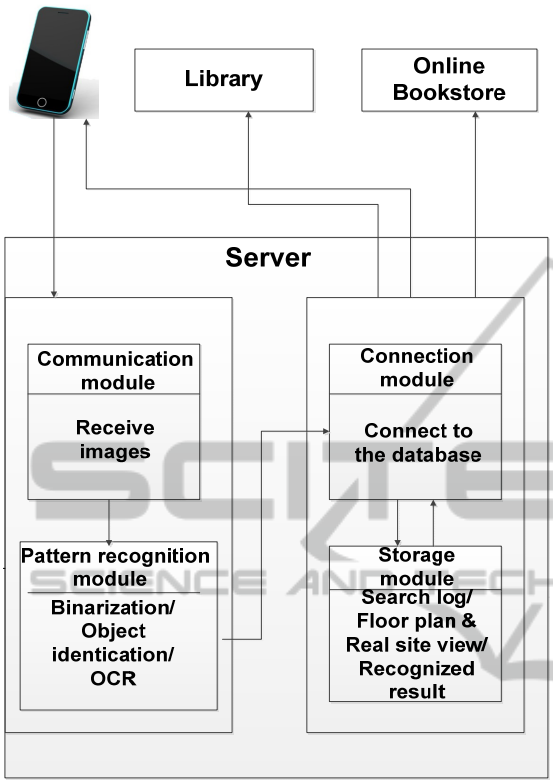


Figure 3: System architecture.

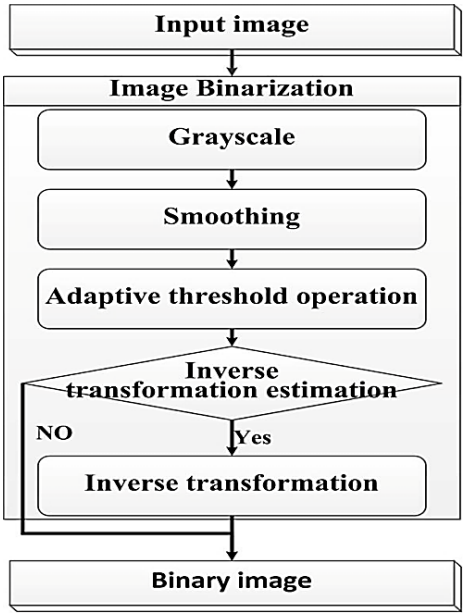


Figure 4: Image binarization.

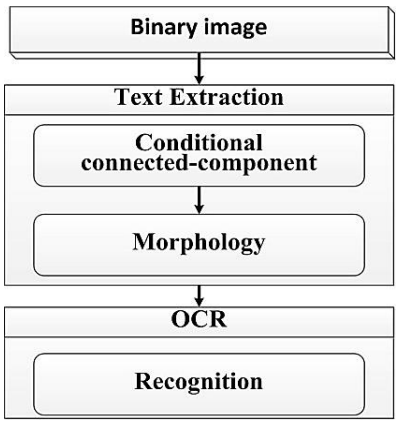


Figure 5: Text extraction.

3.2.2 Smoothing

Due to impaired words usually share the same texture with its background context, we conducted the optical low-pass filter (OLPF) to reduce the noise. LPF will blur the broken characters as Figure 5.

Figure 6: Image smoothing.

3.2.3 Adaptive Threshold Operation

Due to the adoption of a global threshold value to be binary may not be suitable for some scenarios, such as non-uniform lightness and complicated background. In (Kaur, 2013), a face detection algorithm dynamically changes the threshold to distinguish different faces and proved it a better grayscale transformation method.

Different from their studies, our research used the concept of divide and merge to identify local thresholds. Instead of segregating the gray image into many regions (Jian et al., 2009), our approach only partitioned an image into three equal-sized horizontal segments (top, middle, and bottom as Figure 6) to identify the local threshold value of each segment.

3.2.4 Inverse Transformation

The stage is to standardize binarization template to have a black background and white foreground. Inverse transformation (IT) estimation is conducted if the result happened to be opposite. If the total number of pixels in white is greater than black, we will conduct the transformation estimation as (2)(2). Where P'_i indicates the value of pixels after IT, and P_i is the original value. After processing, the three segments will be restored to its original state.

$$P'_i = \begin{cases} 0, & if\ P_i = 255 \\ 255, & if\ P_i = 0 \end{cases} \quad (2)$$

Figure 7: Three-segmented image.

3.2.5 Conditional Connected Component

In many cases, the pixels in white (S_1) may span two to three segments, hence there are three kinds of merging circumstances stated as follows:

- S_1 is the top segment: (1) top segment only (2) the merge of top and middle segments (3) the merge of all three segments.
- S_1 is the bottom segment: (1) bottom segment only (2) the merge of the bottom and middle segments (3) the merge of all three segments.
- S_1 is the middle segment: (1) middle segment only (2) the merge of top and middle segments (4) the merge of the middle and bottom segments (3) the merge of all three segments.

To determine whether the identified components need to be merged and to what extent, we configured a set of formulas ($< 5\% * \frac{1}{3} * T_i$, $< 15\% * \frac{1}{3} * T_i$, $< 25\% * \frac{1}{3} * T_i$, $< 35\% * \frac{1}{3} * T_i$, $< 45\% * \frac{1}{3} * T_i$) with 10, 20, 30, and 40 images to test the difference before and after the merging, respectively. Ti represents the total number of pixels of image i. Using the formula $< 5\% * \frac{1}{3} * T_i$, we gain the least number of connected components after segment combination whereas it also neglects valid text region. On the other hand, using the formula $< 45\% * \frac{1}{3} * T_i$, it makes little effect of the change of the components before and after the segment combination. Finally we identified $15\% \times \frac{1}{3} \times T_i$ as the optimal threshold to process the connected components.

After the above processing, we successfully captured the text region. To eliminate invalid objects (too large or too small, width is far longer than the height, etc.), we devised three filtering rules as (3), (4), (5) based on the shape of fonts. Any object satisfying any one of the following criteria is eliminated.

$$\text{Font height} \times \text{Font width} > 250 \quad (3)$$

$$0.5 < \frac{Font\ height}{Font\ width} < 1.8 \quad (4)$$

$$0.5 < \frac{Font\ height}{Font\ height} < 1.5 \quad (5)$$

3.2.6 Morphological Enhancements

In this part, we executed the erosion and dilation to cope with the non-textual noise. Erosion was conducted to reduce image noise and magnified the difference between foreground texts and background context. Due to the side effect of erosion may shrink the size of the connected components, we resumed it by the way of dilation. After executing morphological enhancement, we successfully erased the unnecessary stains and made the texts clearer as Figure 8. The purified texts are ready to process OCR.

Figure 8: Morphology execution.

3.3 Connection & Storage Module

In this phase, we built the floor plan with a 3D virtual reality of the library presented as Figure 9. Google SketchUp was the tool to model the graph. A personal account database was created to record users' query history. It was designed to reduce the query time by referring to the query logs. We also provided a book order link to a price comparison webpage. In addition, users can also select the "book

recommendation" function, if the needed book does not exist in the library.

Figure 9: 3D virtual reality.

3.4 Evaluation and Result Analysis

Forty book titles were used as the test cases. The accuracy rate is defined as the proportion of recognized characters actually correct. Results showed that except for 8 extremely complex images as Table 1, our treatment successfully identified a part or full text. Table 2 displays a snapshot of the text recognition result with and without the treatment. The average accuracy rates of text recognition are presented in Table 3. The average accuracy rate was raised from 17.7% to 72.05% of all test images with the treatment. Without counting the eight unrecognizable images, the average accuracy rates of our treatment can reach 90.06%.

To compare with other studies (Wai-Lin and Chi-Man, 2011, Kim et al., 2004), we conducted another evaluation to examine the validity of our approach. We selected 90 images with the size of 640 * 480 from the ICDAR2003 Robust Reading Competition. The deliberately selected samples contain the features including different font-size, non-uniform illumination, low contrast, or complicated background. In the second evaluation, we adopted precision and recall as the measurement criteria. A total of 895 text objects was captured from the 90 images. Precision measures the proportion of recognized text objects actually correct, whereas recall measures the proportion of correct objects actually recognized. Table 4 displays the performance comparison. Our treatment outperforms most of the other studies and the performance achieves 74.6% in precision and 80.2% in the recall. Through the test cases, we are confident that a careful implementation of image pre-processing is essential to enhance the accuracy rate of text recognition of nature scene images.

Table 1: Complex images -2 samples.

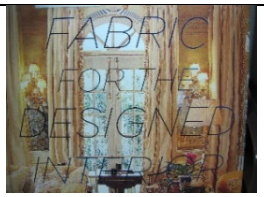
 Complicated context	 Obscure texts

Table 2: Result of text recognition – 1 sample.

	Without treatment: RES_D____ LAYOUI Accuracy rate: 52.94%
	With treatment: RESIDENTIAL LAYOUT Accuracy rate: 100%

Table 3 Average accuracy rates.

Treatment	Yes	No
40 book titles	72.05%	17.7%
32 book titles	90.06%	22.13%

Table 4: A comparison with other studies.

	P	R
Ashida	0.55	0.46
HW David	0.44	0.46
Wolf	0.3	0.44
Chan	0.58	0.59
Kim	0.64	0.83
Our approach	0.75	0.80

* the digits are the rounded value.

4 CONCLUSIONS

In this study, we conducted a series of experiments of image pre-processing to deal with the unwanted objects taken from the shooting. Different from general procedure of segregating the grayscale image into many regions, our approach partitioned an image into three equal-sized horizontal segments to identify the local threshold value of each segment. An adaptive threshold operation was proposed to deal with the problem of uneven lightness. Further, we applied the conditional connected-component and morphological enhancement to identify

components and cope with the non-textual noise.

We used the library of National Yunlin University of Science and Technology in Taiwan as a testing site to apply this technique for searching books via mobile devices. One of the main advantages of this mechanism is that none the existence of a standard database is needed. So it can be applied to different types of images. The results illustrate that the proposed method effectively improves the text recognition. The accuracy rate was raised from 17.7% to over 72.05%. Without counting the eight unrecognizable images, the average accuracy rates of our treatment can reach 90.06%.

To compare with other studies we conducted another evaluation to examine the validity of our approach. The result shows that our treatment outperforms most of the other studies and the performance achieves 74.6% in precision and 80.2% in the recall. Our future work will focus on optimizing the current recognition results by exploiting new approaches for segmentation and new types of features for better noise attenuation and correction of text skew orientation. We are confident that this design will not only bring users more convenience in using libraries but help library staff and businessmen to manage the status of books.

ACKNOWLEDGEMENTS

This research is partly supported by National Science Council, Taiwan, R.O.C. under grant number NSC 101-2221-E-224-056.

REFERENCES

Besag, J. 1989. Digital Image Processing: Toward Bayesian Image Analysis. Journal Of Applied Statistics, 16, 395-407.

Bieniecki, W., Grabowski, S. & Rozenberg, W. 2007. Image Preprocessing For Improving Ocr Accuracy. International Conference On Perspective Technologies And Methods In Mems Design.

Chowdhury, S. P., Dhar, S., Das, A. K., Chanda, B. & Mcmenemy, K. Robust Extraction Of Text From Camera Images. Proceedings Of The 10th International Conference On Document Analysis And Recognition, 2009 Barcelona, Spain. 1635445: Ieee Computer Society, 1280-1284.

Foster, C. 1995. Pdas And The Library Without A Roof. Journal Of Computing In Higher Education, 7, 85-93.

Holley, R. 2009. How Good Can It Get? Analysing And Improving Ocr Accuracy In Large Scale Historic Newspaper Digitisation Programs. D-Lib Magazine.

Jian, Y., Yi, Z., Kok-Kiong, T. & Tong-Heng, L. Text Extraction From Images Captured Via Mobile And Digital Devices. Proceedings Of The Ieee/Asme International Conference On Advanced Intelligent Mechatronics(Aim), 14-17 July 2009. 566-571.

Kaur, A. 2013. Mingle Face Detection Using Adaptive Thresholding And Hybrid Median Filter. International Journal Of Computer Applications In Technology, 70, 13-17.

Kim, E., Lee, S. & Kim, J. 2009. Scene Text Extraction Using Focus Of Mobile Camera. The 10th International Conference On Document Analysis And Recognition.

Kim, K. C., Byun, H. R., Song, Y. J., Choi, Y. W., Chi, S. Y., Kim, K. K. & Chung, Y. K. Scene Text Extraction In Natural Scene Images Using Hierarchical Feature Combining And Verification. Proceedings Of The 17th International Conference On Pattern Recognition(Icpr) 23-26 Aug 2004. 679-682.

Lawler, R. 2013. Mary Meeker's 2013 Internet Trends: Mobile Makes Up 15% Of All Internet Traffic, With 1.5b Users Worldwide [Online]. Available: Http://Techcrunch.Com/2013/05/29/Mary-Meeker-2013-Internet-Trends/.

Liu, X. & Samarabandu, J. Multiscale Edge-Based Text Extraction From Complex Images. Proceedings Of The Ieee International Conference On Multimedia And Expo, 9-12 July 2006 Toronto,Ontario,Canada. 1721-1724.

Marqués, F. & Vilaplana, V. 2002. Face Segmentation And Tracking Based On Connected Operators And Partition Projection. Pattern Recognition, 35, 601-614.

Mills, K. 2010. M-Libraries: Information Use On The Move. In: Neeham, G. & Ally, M. (Eds.) M-Libraries 2: A Virtual Library In Everyone's Pocket. London: Facet Publishi.

Minetto, R., Thome, N., Cord, M., Stolfi, J., Precioso, F., Guyomard, J. & Leite, N. J. 2011. Text Detection And Recognition In Urban Scenes. Ieee International Conference On Computer Vision Workshops (Iccv Workshops).

Mori, S., Suen, C. Y. & Yamamoto, K. 1992. Historical Review Of Ocr Research And Development. Proceedings Of The Ieee, 80, 1029-1058.

Pei-Jun, L. & Effendi 2010. Adaptive Edge-Oriented Depth Image Smoothing Approach For Depth Image Based Rendering. Ieee International Symposium On Broadband Multimedia Systems And Broadcasting (Bmsb).

Raza, M. U., Ullah, A., Ghori, K. M. & Haider, S. Text Extraction Using Artificial Neural Networks. Proceedings Of The International Conference On Networked Computing And Advanced Information Management, June 2001 Gyeongju, Gyeongsangbuk-Do, South Korea. 134-137.

Rong-Yuh, H. The Design And Implementation Of Mobile Navigation System For The Digital Libraries. Proceedings Of The Sixth International Conference On Information Visualisation 2002. 65-69.

Shuqing, Z. & Qiaoning, Y. Microarray Images Processing Based On Mathematical Morphology. Proceedings Of The 2006 8th International Conference On Signal Processing, 16-20 2006 2006.

Shutao, L. & Kwok, J. T. Text Extraction Using Edge Detection And Morphological Dilation. Proceedings Of The International Symposium On Intelligent Multimedia, Video And Speech Processing, 20-22 Oct 2004. 330-333.

Taisheng, L., Xuan, Z. & Chongrong, L. 2012. An Improved Adaptive Image Filter For Edge And Detail Information Preservation. International Conference On Systems And Informatics (Icsai).

Wai-Lin, C. & Chi-Man, P. Robust Character Recognition Using Connected-Component Extraction. Proceedings Of The Seventh International Conference On Intelligent Information Hiding And Multimedia Signal Processing (Iih-Msp), 14-16 Oct 2011. 310-313.

Ye, Q., Huang, Q., Gao, W. & Zhao, D. 2005. Fast And Robust Text Detection In Images And Video Frames. Image Vision Comput, 23, 565-576.