



Optical Character Recognition Using Active Contour Segmentation

Nabeel Oudah. M

Lecturer

Uruk University

nabeeloude@mail.ru

Maher Faik Esmail

Lecturer

Iraqi Cement State Company

maherfaik2000@yahoo.com

Estabraq Abdulredaa

Assistant Lecturer

Al-Israa College University

estabraq_ai_1989@yahoo.com

ABSTRACT

Document analysis of images snapped by camera is a growing challenge. These photos are often poor-quality compound images, composed of various objects and text; this makes automatic analysis complicated. OCR is one of the image processing techniques which is used to perform automatic identification of texts. Existing image processing techniques need to manage many parameters in order to clearly recognize the text in such pictures. Segmentation is regarded one of these essential parameters. This paper discusses the accuracy of segmentation process and its effect over the recognition process. According to the proposed method, the images were firstly filtered using the Wiener filter then the active contour algorithm could be applied in the segmentation process. The Tesseract OCR Engine was selected in order to evaluate the performance and identification accuracy of the proposed method. The results showed that a more accurate segmentation process shall lead to a more accurate recognition results. The rate of recognition accuracy was 0.95 for the proposed algorithm compared with 0.85 for the Tesseract OCR Engine.

Keywords: OCR, active contour, Segmentation, Automatic Character Recognition, Pattern Recognition, Tesseract OCR Engine.

التعرف الصوتي على الحروف والارقام باستخدام برنامج الماتلاب

استيرق عبد الرضا

مدرس مساعد

كلية الاسراء الجامعة

ماهر فائق اسماعيل

مدرس

شركة الاسمنت العراقية

نبيل عوده منيخر

مدرس

جامعة اوروك

الخلاصة

تحليل الوثائق من الصور الملتقطة بواسطة الكاميرا هي عملية ذات تحديات متزايدة. حيث ان هذه الصور غالبا ما تكون صور مركبة وذات جودة سيئة، تتألف من مختلف الكائنات والنصوص؛ مما يجعل عملية التحليل التلقائي معقدا. التعرف على الحروف البصرية (OCR) هي واحدة من تقنيات معالجة الصور التي لديها القدرة على اجراء عملية التعرف على الحروف والكلمات بصورة اوتوماتيكية. تقنيات معالجة الصور الحالية تحتاج إلى إدارة العديد من العوامل من اجل اجراء عملية التمييز بصورة صحيحة في مثل هذه الصور. وتعتبر عملية تجزئة الصورة (Segmentation) واحدة من هذه العوامل الأساسية المؤثرة. ناقش هذا البحث تأثير دقة عملية التجزئة (Segmentation) للحرف من صورته وتأثيرها على عملية التعرف الالي. وفقا للطريقة المقترحة، تم تصفية الصورة أولا باستخدام فلتر ويينر (Wiener filter) ومن ثم يمكن تطبيق خوارزمية active contour لأجراء عملية تجزئة الصورة. تم اختيار Tesseract OCR Engine من أجل تقييم الأداء وتحديد دقة الطريقة المقترحة. واطهرت النتائج انه كلما كانت عملية التجزئة دقيقة فان نتائج التعرف ستكون أكثر دقة. وكان معدل دقة التمييز للحروف 0,95 للخوارزمية المقترحة مقارنة مع 0,85 لخوارزمية Tesseract OCR Engine.



1. INTRODUCTION

Optical character recognition (OCR) is considered as one of the greatest useful implementations of automatic pattern recognition. Many researches and developments were applied actively for the OCR field since the mid 1950's, **Otu, 1979, Amit Choudhary, 2014**. At present, 100\$ is the cost of a rationally good OCR packages. Yet, these OCR packages are still of a limited use for only the recognition of text documents that have high printing quality or carefully written hand printed texts. there are many attempts as well to decrease the replacement error rates and refusal rates even on good quality machine printed text, because the efficient human typist, in spite of the lower printing speed is yet performing much less rates of errors.

The need for the OCR arises when printed information should be legible for man and machine (computer) with the lack of predefined substitution inputs. The individual characteristic of the OCR over the other methods of automatic identification is that there is no need to control the information production process. Typically, an OCR system consists of the following processing steps, **Amit Choudhary, 2014**:

- ✓ Gray level scanning at an appropriate resolution, typically 300{1000 dots per inch,
- ✓ Preprocessing:
 - a- Binarization (two-level thresholding), using a global or a locally adaptive method,
 - b- Segmentation to isolate individual characters,
 - c- (Optional) conversion to another character representation (e.g., skeleton or contour curve)
- ✓ Feature extraction
- ✓ Recognition using one or more classifiers.
- ✓ Contextual verification or post processing.

Many researches were done on the OCR systems. Some of these methods can consider the binarization techniques, **Niblack, 1986, Mori, et al., 1992**, to be as text detection methods. However, when these techniques are used to process a complex scene images their accuracy shall be limited. Luo et al. from Motorola China Research Center have presented camera based mobile OCR systems for camera phones in, **Xi-Ping, et al., 2004, Xi-Ping, et al., 2005**. In, **Xi-Ping, et al., 2004**, the skew angle is assessed from a business card image which is firstly down sampled. After that, this angle is used to correct the skew text regions which are then binarized. The lines and characters within these text regions are segmented and thereafter they are recognized by passing them on an OCR system which is made of a two layer template based classifier. In, **Xi-Ping, et al., 2005**, a presentation of analogous system for Chinese-English mixed script business card images is shown. An outline of a prototype Kanji OCR to recognize Japanese texts which are machine printed and are translated into English has been presented by, **Koga, et al., 2005**. **Laine, Laine, and Olli, 2006**, made a system that just deals with the English capital letters. In the beginning, the input captured mage is skew. This image is repaired by searching for a line with the greatest number of successive white pixels through maximizing the given alignment standard. After that, segmentation is done for the image on the basis of X-Y Tree decomposition then the image is to be recognized via detecting the Manhattan distance based similarity for a set of centroid to boundary features. Yet, this is only applicable for addressing the English capital letters and the acquired precision does not reach the limit of satisfaction to apply this method in reality. The principle of the Region-based and connected component (CC) techniques, **Zhang, and Kasturi, 2008**, is the presumption of the difference between the pixel features of the text and non-text. To obtain correct results from these methods, training sets for classifiers and prior information of text position and scale are required to be



used. The application of the Hybrid methods by **Yi-Feng, et al., 2009**, showed some improvements which are a combination of region-based methods, CCs, and layout analysis methods. Lately, **Li, et al., 2008** and **Du, et al., 2009**, used the Mumford-Shah model, **Mumford, and Shah, 1989**, and Chan-Vese piecewise approximation, **Chan, and Vese, 1999**, respectively to present text line segmentation for handwritten documents.

In, **Wumo, et al., 2001**, one more work is presented to recognize Chinese script of business card images. Furthermore, the developmental researches on the OCR systems for mobile devices extend beyond the limitation of document images recognition. A work on reading LCD/LED displays has been presented by **Shen, et al., 2006**, using a camera phone. In, **Bae, et al., 2005**, a representation for a character recognition system for Chinese scripts has been done.

Based on the above mentioned studies, the OCR systems feasibility is assured besides the fact of the ability of the OCR systems to deal with the handheld devices. However, the approachable computation and accurate segmentation must be the characteristic of these systems' algorithms. They should be computationally efficient, low memory consuming and an efficient segmentation algorithm.

In this paper, we are proposing an automated extraction of characters from images using Weiner filter (as preprocessing) with active contour algorithm to obtain both the character's local and global properties so as to earn a reliable digits' detection. We aim in this research to show the effect of the segmentation process on the OCR recognition accuracy rate from applying an active contour algorithm with Weiner filter and compare the results with the Tesseract OCR Engine, **Smith, 2007, Smith, et al., 2009**, that was the HP Research Prototype and it has been developed and sponsored by Google since 2006. It is considered one of the most accurate open-source OCR engines then available and the Matlab company has built and depended it as a function in its products within the version 2014a and above.

2. METHODOLOGY

In our current work we present an automatic character recognition algorithm that is used to recognize English characters found in image/graphics embedded text documents taken by camera like plate of car's numbers images.

Commonly, color images can be obtained by the modern handheld devices. Color pixels are the component of a color image. These color pixels combines three basic color components i.e. red (r), green (g) and blue (b). The range of values for all these color components is 0-255. Thus, the corresponding gray scale value (x, y) for every single pixel that is also lies between 0-255. Through the scanning process, a digital image of the original document is captured.

The proposed OCR system has the components as can be shown in **Fig. 1**. Firstly, an optical scanner is used to convert the analog document into digits. When the detection of the text containing areas is done, then these symbols can be processed, noise elimination process using Wiener filter is performed in order to make features extraction process in the next step easier. After that, a segmentation process is performed to extract every symbol. Every symbol is then identified through the comparison between the extracted features and the descriptions of the symbol classes gained during a former learning phase. In the end, the words and numbers of the original text are reconstructed by using contextual information.

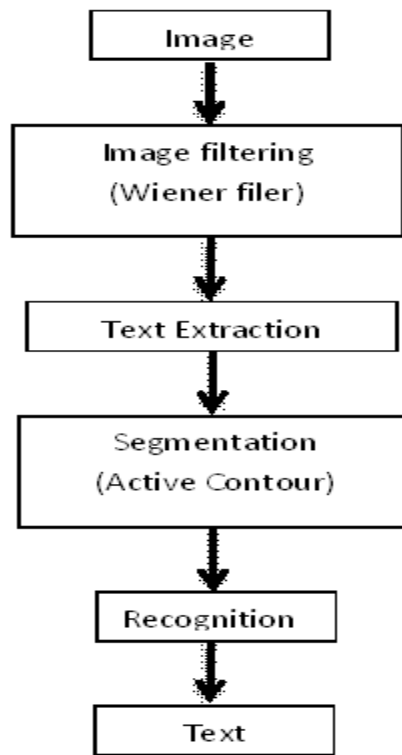


Figure 1. Block diagram of the proposed algorithm.

2.1 Preprocessing

A definite amount of noise can be found in the image that is produced by the scanning process. Smoothing involves both filling and thinning procedures whereas the filling process is done to remove small breaks, gaps and holes in the digitized characters whilst the thinning process will decrease the line width. The Wiener filter would be used in this work to filter out the noise from the corrupted images. The Wiener filter is the mean square error optimal stationary linear filter for images degraded by additive noise and blurring. In other words, it minimizes the overall mean square error in the process of inverse filtering and noise smoothing.

2.2 Active Contour

It can be defined as a parametric curve which attempts to move into a place to make its energy decrease and spline guided by external constraint forces and influenced by image forces that pull it toward features such as lines and edges, **Esmail, et al., 2013, Tiilikainen**. A general edge-detector that can be defined by a positive and decreasing function, depending on the gradient of the image, such that, **Esmail, et al., 2013, Xu, et al., 1998**:

$$E_{snake}^* = \int_0^1 E_{int} V(s) + E_{img} V(s) + E_{con} V(s) ds \quad (1)$$

From Equation (1) we can find that there are three terms to consist the snake energy. E_{int} is the first term



which is referred to the internal energy of the snake. E_{img} is the second term and is referred to the image forces. Lastly, the term E_{con} is referred to the external constraint forces. Another term is found here. It is the term E_{ext} and it is used to denote the external snake forces which are made up from both the sum of the image force, E_{img} and the external constraint force, E_{con} . The internal energy of the snake is given by, **Esmaille, et al., 2013, Xu, et al., 1998:**

$$E_{int} = \frac{1}{2} \left(\alpha(s) \|V_s(s)\|^2 + \beta(s) \|V_{s,s}(s)\|^2 \right), \tag{2}$$

In this equation the measure of the elasticity is given by the first-order term $\|V_s(s)\|^2$ (The first derivatives) and the measure of the curvature is obtained from the second-order term $\|V_{s,s}(s)\|^2$ (The second derivatives). The first and second derivatives of the contour represent these energy terms and called “Elastic forces” and “Bending forces”, respectively. The total snake energy is controlled by the coefficients $\alpha(s)$ and $\beta(s)$ in such a way that the snake is to be more elastic and less rigid by reduction or rising and vice versa. The parameters $\alpha(s)$ and $\beta(s)$ in front of each term represents weighting functions. In general, values of these weighting functions are constants for all snaxels. Selecting an appropriate set of these constants creates one of difficulties of the snake. They have large impact in snake’s behaviors and totally control the performance of deformation process. Each object in an image requires different set of constants value for snake to perform well. The one way to solve this problem is to make the snake dynamically change these values to suitable values during deformation process. However, it requires a computer to recognize shapes or topologies of an object in an image automatically. Therefore, the solution is left for further improvement of the snake. Currently, these parameters are up for a user to select at the initialization process.

The snake is pulled to the closest image edge via the image forces E_{img} . Performance of the image energy can be done depending on the following:

$$E_{img} = w_{int}E_{int} + w_{edge}E_{edge} + w_{term}E_{term} \tag{3}$$

The image forces that are formed by a linear combination of line, edge and termination energy terms will lead to the formation of the energy term. These energy terms are computed from the image in which the Eline and Eedge can be performed as:

$$E_{line} = I(x, y), \tag{4}$$

$$E_{edge} = -|\nabla I(x, y)|^2, \tag{5}$$

where, $I(x,y)$ represents the image intensity. The term E_{term} can be defined as the curvature of the level contour in Gaussian smoothed image. The adjustment of the weights, w can lead to the creation of a wide range of snake behavior.

The external constraints forces E_{con} are written as, **Esmaille, et al., 2013, Guerrero, et al., 2003:**

$$E_{con} = (x_i - x_{i-1})^2 + (y_i - y_{i-1})^2 \tag{6}$$



Processing of the E_{con} is achieved by allowing the user to introduce a “volcano icon” that aids in pushing the snake faraway. The benefit of this action is to push the snake out of an undesired local minimum. In case when there is discrimination, the representation of the contour is done by N points $P_1, P_2, P_3, \dots, P_N$ whereas $P_1 = (x_i, y_i)$ and the first derivative is approximated by a finite difference, **Tang, 1982, Esmail, et al., 2013**.

2.3 Template-matching and correlation techniques

The difference between these methods and the other methods is that there is no actual extraction of features. Instead of that, a direct matching is done for the matrix containing the image of the input character with a set of prototype characters that represents each probable class. A computational process is done for the space between the pattern and each prototype. The best matching class of the prototype is then assigned to the pattern. The hardware implementation of this method is simple and easy. It was applied in many commercial OCR machines. Nevertheless, this method is affected by noise and style variations and it lacks the ability of processing rotated characters.

3. RESULTS AND DISCUSSION

In order to do the OCR performance evaluation there are no standardized test sets exist for character recognition and as the performance of an OCR system is highly dependent on the quality of the input. This makes it difficult to evaluate and compare different systems. Still, recognition rates is usually used to assess the performance of an OCR system, and usually presented as the percentage of characters correctly classified. Accordingly, the following criteria can be applied:

$$\text{Recognition rate} = CR \times 100 / TT \quad (7)$$

where CR , the number of correctly recognized characters, and TT , the total number of tested characters.

In order to show the effect of the segmentation process over the recognition process accuracy, the proposed need to be compared with another algorithm. Based on that, Tesseract OCR Engine that was built and depended as a function by the Matlab company within the 2014a version was selected for the purpose of that comparison because of its accurate performance with low computational time. The performance and accuracy of the proposed algorithm were measured subjectively and objectively using English sample text and car’s plate images (with noised images and different images’ resolution and quality).

With Tesseract OCR Engine, **Smith, 2007, Smith, et al., 2009**, the images were filtered, binarized, clipped and resized. Lines of text were then extracted from the images. The font size was identified; segmentation was performed on each line to segment characters taking in consideration the characteristics of English Verdana font’s templates. MATLAB (R2014a/64-bit) is used to implement the proposed OCR algorithm. On the other side, the same images have been processed by our proposed methods wherein the first step, the image had been filtered using the wiener filter then had been segmented using active contour algorithm. After that, the recognition process had been applied with the same template that is used in the Tesseract OCR Engine. The hypothesis here is that the accurate

segmentation will lead to the less error rate and high recognition accuracy performance. **Fig. 2** shows the active contour performance to do the segmentation process for the letters. **Fig. 3 and 6** show the results of the proposed algorithm and the Tesseract OCR Engine.

Regarding **Fig. 3**, the proposed algorithm and Tesseract OCR Engine could recognize the connected letters successfully and there is no difference in the both algorithms accuracy's results.

Fig. 4 shows the result of the proposed algorithm and Tesseract OCR Engine for the noised image by the salt and pepper wherein the proposed algorithm could overcome the noise effects while the Tesseract OCR Engine showed approximately 99% error rate with all tested images and it could not overcome the noise effect whereas this test illustrates clearly the effect of the segmentation algorithm on the OCR accuracy and performance. Since the Tesseract OCR Engine wasn't able to overcome the salt and pepper noise type which is considered as a simple type of noise, for this reason there was no need to perform more tests with another type of noise.

As can be seen in **Fig. 5b**, and **5b** the proposed algorithm showed more accuracy than the Tesseract OCR Engine (see 4c, and 5c) wherein some errors can be seen with the results of the Tesseract OCR Engine.g., the symbol (“) in the **Fig. 5c** instead of 1, and Z instead of 2 in the **Fig. 6c**). On the same context, an error occurred with the proposed algorithm as can be seen in **Fig. 6b** wherein the letter (I) was not recognized correctly. Table 1 shows the recognition rate and rejection rate for the proposed algorithm and the Tesseract OCR Engine with noised images and with different images' resolution and quality. According to the calculations in Table 1, we can estimate that the error rate was 0.05 and 0.15 for the proposed algorithm and the Tesseract OCR Engine respectively.

4. CONCLUSION

Nowadays, optical character recognition is most effective for constrained matter that is documents generate under various control. This paper has discussed the effect of segmentation accuracy on OCR wherein the active contour with Weiner filter are proposed to be applied in the segmentation process and compared the obtained results with the Tesseract OCR Engine . The results have been shown that more segmentation accuracy results more recognition accuracy. Based on the simulation results, the proposed algorithm has achieved high recognition rate and consequently low error rate results.

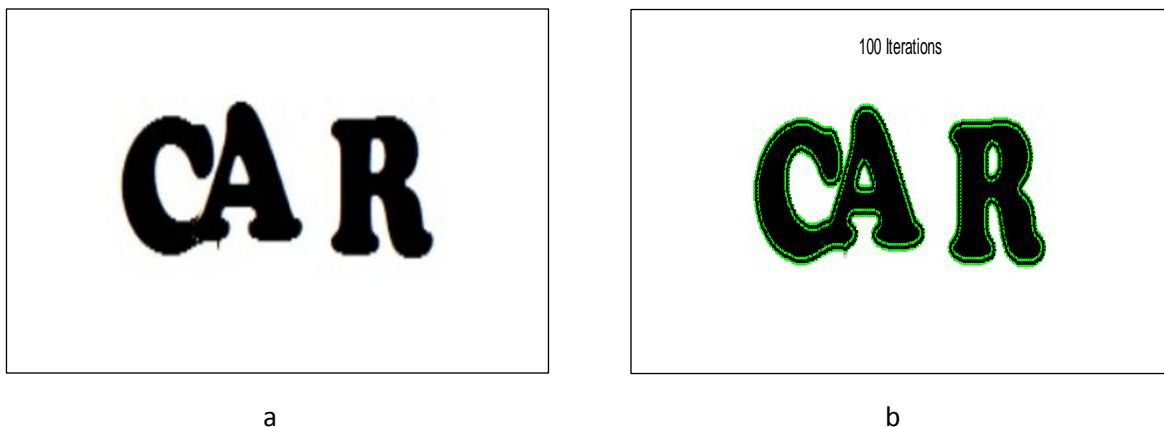


Figure 2. Active contour segmentation, a- original image, b- segmented image.

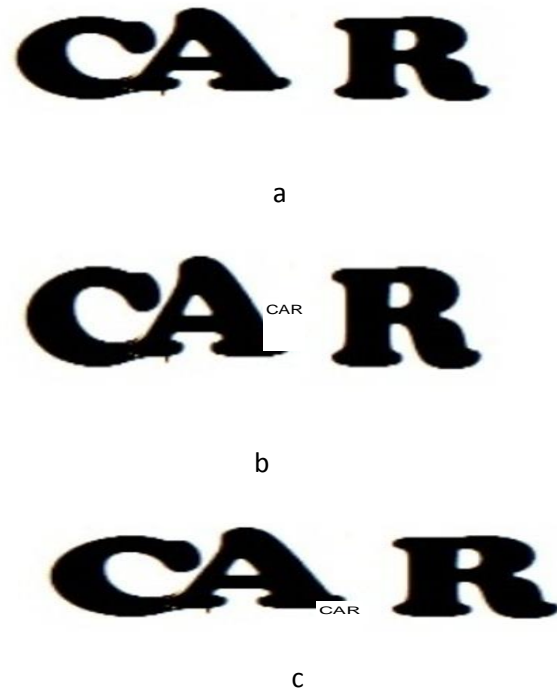


Figure 3. the comparison result between the proposed algorithm the Tesseract OCR Engine, a. the original image, b. the proposed algorithm, c. the Tesseract OCR Engine.



a



b



c

Figure 4. the comparison results between the proposed algorithm the Tesseract OCR Engine, a. the original image, b. the proposed algorithm, c. the Tesseract OCR Engine.

Table 1. The result of Recognition Rate and Error Rate for the proposed Algorithm and Tesseract OCR Engine

	Connected Letters		Noised Images		Different images quality	
	Proposed	Matlab	Proposed	Matlab	Proposed	Matlab
Character Recognition Rate	98%	98%	98%	1%	95%	90%
Error rate	2%	2%	2%	99%	5%	10%



a



b



c

Figure 5. the comparison result between the proposed algorithm the Tesseract OCR Engine, a. the original image, b. the proposed algorithm, c. the Tesseract OCR Engine.



a



b



c

Figure 6. the comparison result between the proposed algorithm the Tesseract OCR Engine, a. the original image, b. the proposed algorithm, c. the Tesseract OCR Engine.



REFERENCES

- Amit C., “A Review of Various Character Segmentation Techniques for Cursive Handwritten Words Recognition”, *International Journal of Information & Computation Technology*, ISSN 0974-2239 Volume 4, Number 6, pp. 559-564, 2014.
- Bae. S., Kim K. K., Chung Y. G. and Yu W. P., “Character Recognition System for Cellular Phone with Camera”, *Proceedings of the 29th Annual International Computer Software and Applications Conference*, vol. 1, pp. 539-544, 2005.
- Chan T. and Vese., “An active contour model without edges”, in *Proceeding of the Second International Conference on Scale-Space Theories in Computer Vision (Scale-Space '99)*, Corfu, Greece, September pp. 26–27, 1999.
- Du X., Pan W., and Bui T. D., “Text Line Segmentation in Handwritten Documents Using Mumford-Shah Model, *Pattern Recognition*”, vol. 42, pp. 3136-3145, 2009.
- Esmail M.F., Marhaban M.H., Mahmud R., Saripan, M.I., “Cross -sectional area calculation for arbitrary shape in the image using star algorithm with Green's theorem”. *IEEJ Transactions on Electrical and Electronic Engineering* 8(5):497-504(IF:0.343), 2013.
- Guerrero J., Salcudean S. E., McEwen J. A., Masri B. A., Nicolaou S., “Deep Venous Thrombosis Screening System Using Numerical Measures,” in *Proc. 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Cancun, pp. 894-897, 2003.
- Koga M., Ryuji M., Tatsuya K., Toshikazu T., Masahiro Y. and Teruyuki Y., “Camera-based Kanji OCR for Mobile-phones: Practical Issues”, *Proceedings of the Eighth International Conference on Document Analysis and Recognition*, pp. 635-639, 2005.
- Laine M. and Olli S. N., “A standalone OCR system for mobile camera-phones”, *Personal, Indoor and Mobile Radio Communications, 2006 IEEE 17th International Symposium*, pp.1-5, Sept. 2006
- Li Y., Zheng Y., Doermann D., and Jaeger S., “Script-Independent Text Line Segmentation in Freestyle Handwritten Documents”, *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 30, no. 8, pp. 1313 – 1329, 2008.
- Mori S., Suen C. Y., and Yamamoto K., “Historical review of OCR research and development,” *Proceedings of the IEEE*, vol. 80, pp. 1029-1058, July 1992. 3.
- Mumford D. and Shah., “Optimal approximation by piecewise smooth functions and associated variational problems”, *Comm. Pure Appl. Math.*, 42, 577-685, 1989.
- Niblack W., “An Introduction to Image Processing”, Prentice-Hall, Englewood Cliffs, NJ, pp. 115-116, 1986.
- Otsu N., “A threshold selection method from grey level histogram”, *IEEE Trans SMC*, 9, pp. 62-66, 1979.
- Shen, H. and Coughlan, J., “Reading LCD/LED Displays with a Camera Cell Phone”, *Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop*, 2006.



- Smith R., An Overview of the Tesseract OCR Engine, Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) Vol 2, pp. 629-633, 2007.
- Smith, R., D. Antonova, and D. Lee. Adapting the Tesseract Open Source OCR Engine for Multilingual OCR. Proceedings of the International Workshop on Multilingual OCR, (2009).
- Tang G. Y., “A Discrete version of Green's theorem,” IEEE Trans. Pattern Anal Mach Intell, vol. 4, pp. 242-249, 1982.
- Tiilikainen. P., “Comparative Study of Active Contour Snakes,” unpublished.
- Wumo P., Jianming J., Guangshun S., Wang Q. R., “A System for Automatic Chinese Business Card Recognition”, ICDAR, pp. 577-581, 2001.
- Xi-Ping L., Jun L. and Li-Xin Z., “Design and implementation of a card reader based on build-in camera”, International Conference on Pattern Recognition, pp. 417-420, 2004.
- Xi-Ping L., Li-Xin Z., Gang P., Jun L. and Bai-Hua X., “Camera based mixed-lingual card reader for mobile device”, International Conference on Document Analysis and Recognition, pp. 665-669, 2005.
- Xu C., Prince J. L., “Snakes, Shapes, and Gradient Vector Flow,” IEEE Trans. Image Process, vol. 7, pp. 359-369, 1998.
- Yi-Feng P., Xinwen H., Cheng-Lin L., “Text Localization in Natural Scene Images based on Conditional Random Field”, 10th International Conference on Document Analysis and Recognition, Barcelona, Spain, 2009.
- Zhang J. and Kasturi R., “Extraction of text objects in video documents”, Recent progress, in Proceeding of the Eighth IAPR Workshop on Document Analysis Systems (DAS'08), pages 1–13, Nara, Japan, 2008.