A novel algorithm of detecting document area of mobile images

Maosen Wang Shaozhang Niu
Beijing Key Lab of Intelligent Telecommunication Software and Multimedia,
Beijing University of Posts and Telecommunications, Beijing 100876, China
wmsbeijing@126.com, szniu@bupt.edu.cn

Abstract

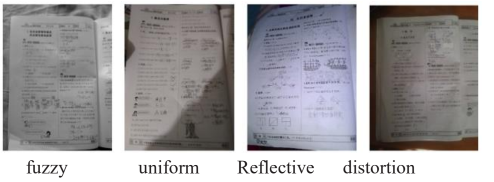
With the popularity of mobile and enhancement of mobile camera function, traditional digital cameras and scanner functions are gradually replaced by mobile camera. There are a lot of APP on mobile and able to scan paper documents. For photographic image processing, determining the boundaries of the document area from the mobile image is the first and important step. For the traditional document image obtained by the scanner, the commonly used algorithm is the Hough transform [1]. However, the mobile images are generally fuzzy, distortion because of un-uniform light source and unfixed focal length, which are introduced by the portability of mobile. Hough transform cannot solve the problem of boundary detection of mobile image very well. In addition, edge detection of using Hough transform is relatively slow, which result into being difficult to run on the mobile. In this paper, we utilize dynamic programming [2] algorithm to determine the boundary of the document, and the K-Nearest Neighbor Classification [3] algorithm to select the four edges from the candidate boundaries, and quickly segment document area. The two algorithms are efficient and have high accuracy of detecting boundary, and have been integrated in the mobile APP.

Key words: Hough Transforms, Dynamic programming, KNN, Non-maximum suppression

I. Introduction

With the improvement of mobile phone camera resolution and convenience of using of mobile, People gradually use mobile phones instead of digital cameras or scanners to scan paper documents. Area of the paper document are needed to carve out from the background image. Image segmentation is one of most important steps leading to the analysis of processed image data. The existing methods of image segmentation are divided into the following categories: thresholding method [4,5,6], region method [7,8,9], edge detection method [10,11] and other methods based on some specific theories[12,13]. The main features of the camera image are: The document area occupies a larger proportion of the background image (typically greater than 80%). The physical boundaries of the paper document are rectangular and the boundary is generally straight. The segmentation method of the document area generally adopts the edge-based segmentation method. Through the Hough transform to detect the line [14,15]. Zhengyou Zhang presented that detect the boundary of the whiteboard from the background image with Hough transform, crop out the area of the whiteboard, and refine the whiteboard area into a square [16]. The author's camera system is a high-resolution digital camera, and the camera is fixed, the image is clearer and less distorted. Mobile image quality is generally worse than the scanner or digital camera. Mobile images have the following characteristics:

1. Mobile images are not clear, often blurred with the instability of focal length since mobile phone is a handheld device and shaking during taking the picture, as shown in Fig. 1 (a).
2. Mobile Images have uneven light and shadows, because there is no uniform light when taking pictures in natural scene, especially at night, relying on indoor lighting. Especially the document image will include bright spots if you turn on the flash. Fig. 1 (b) and (c) show the cases.
3. Images of photograph documents often distorted since the document paper is a page of a book, and often not flat, the four boundaries of the document image are not straight lines.



fuzzy (a) uniform (b) Reflective (c) distortion (d)

Figure 1: image produced by a mobile phone

Detecting area of document with Hough transform, as mobile images fuzzy and noise, Hough transform produces a large number of short-term segments, especially curved borders, the whole boundary becomes multiple lines with Hough transform as show fig. 2. In addition, the computational complexity of the standard Hough transform algorithm $O(N^3)$, although the optimization, the calculation can be reduced to $O(N^2\log_2 N)$, the algorithm runs for a long time and is not efficient, it is difficult to run on the phone, the experiment proved that the Hough transform does not adapt to detecting document area of mobile images.

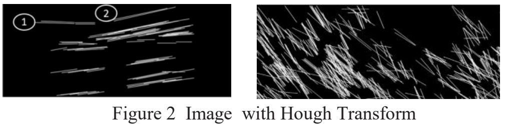


Figure 2 Image with Hough Transform

II. Dynamic Programming Approach

In this paper, the dynamic programming algorithm is applied to detect edge of the document region. The dynamic programming algorithm is introduced by R. E. Bellman. In computer vision, dynamic programming has a lot of practical applications, Amini showed on the example of active contours how Dynamic programming can be utilized to perform energy minimization [18]. Milan Sonka uses dynamic programming to detect lines [19], Nicolas Merlet utilizes dynamic programming to detection of road direction in satellite images

ISBN 978-1-5090-4897-7

[20]. Kathrin uses dynamic programming to achieve target segmentation in medical images [21]. Jose A. Rosado-Tora proposed a dynamic programming algorithm based on the cost function under polar coordinates to automate the segmentation of objects [22].

Taking the k-segment graph of Fig. 3 as an example, the multi-phage graph is divided into (v1, v2, v3, v4, v5) , and the edge of document area is transformed into the longest path problem.

1. A max cost path from S to T in the k-phase graph can be thought of as a result of making some decision in the k-2 phase.
2. The i-th decision determines which node in V_{i+1} is on this path, where $1 \le i \le k-2$.
3. The principle of optimality is established for multi-phase graphs.

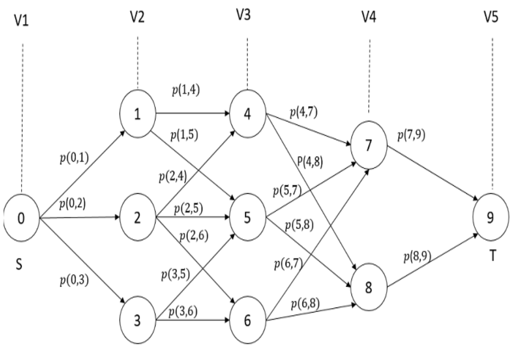


Fig 3 Multi-stage dynamic programming

Set C (i, j) is the cost of point j in i-th phase.

$$C(i,j) = \max_{\substack{i \in v_{i+1} \\ j,l \in E}} \{C(i-1,l) + g(l,j)\}$$

Where

$$g(l,j) = \begin{cases} p(i,j) < i,j > \in E \\ 0 < i,j > \notin E \end{cases}$$

Backwards:

1. Step: k=2

$$C(2,1) = p(0,1)$$

$$C(2,2) = p(0,2)$$

$$C(2,3) = p(0,3)$$

2. Step: K=3

$$C(3,4) = \max(C(2,1) + p(1,4), C(2,2) + p(2,4))$$

$$C(3,5) = \max(C(2,1) + p(1,5), C(2,2) + p(2,5), C(2,3) + p(3,5))$$

$$C(3,6) = \max(C(2,2) + p(2,6), C(2,3) + p(3,6))$$

…

By recursive k = 5 to find the maximum cost, through backtracking, seeking a maximum path.

For an N-line, M-column image, the upper boundary of the document area should be one of the longest paths, from the first column to the last column. If one path is the longest in the set in 1 ... M, it also is the longest in the set of 1 ... K (K <M). We utilize cost function to represent the possibility that this point is the edge of document area, and define cost function for one point as follows:

$$C(x_k^{m+1}) = \max_i [C(x_i^m) + g^m(i,k)] + D(x_k^{m+1}) \qquad (1)$$

Where $C(x_k^{m+1})$ is the cost value of the point(k,m+1), and $g^m(i,k)$ is the cost between nodes x_i^m and x^{k+1}, so the complete optimization solution is defined as the below function:

$$\max[C(x^1,x^2,\dots,x^M)] = \max_{k=1,\dots,n} [C(x_k^M)] \qquad (2)$$

Where x_k^M is the point of the Kth row and the last column, and M is the width of the image, $C(x^1,x^2,\dots,x^M)$ denotes the cost of a path between the first column and the last column in Figure 5.

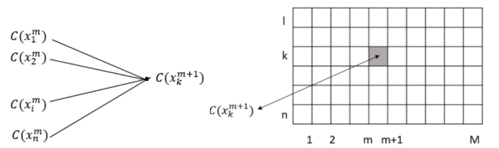


Figure 5 Dynamic planning energy calculation method

Analyzing the upper edge of the document area in the mobile phone image, it has three features:

1. Close to the horizontal line

As shown in Figure 6, the point $P_{i,j}$ has 8 neighborhoods, but only three paths can arrive to the point $P_{i,j}$, they are $P_{i-1,j-1}$ to the $P_{i,j}$, $P_{i,j-1}$ to $P_{i,j}$, and $P_{i+1,j-1}$ to $P_{i,j}$.

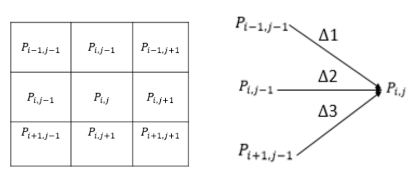


Fig 6: the neighborhood of the pixel

2. The boundary is continuous.

Due to the noise and artifacts in the mobile images, the document boundary may be broken, but is continuous in the overall trend. The calculation of the energy value at each point is based on the values of the neighbor points in the previous column according to the formula (1), so it is a left-to-right accumulation process. For each column, the points with local maximum value are candidate points of the boundary.

3. The boundary must be a foreground in the binarized image

In (1) $D(x_k^{m+1})$ represents the additional cost value. It is defined as:

$$D(P_{i,j}) = \begin{cases} 0 & \text{if } P_{i,j} = 0 \\ \theta & \text{if } P_{i,j} = \theta \ (\theta > 0) \end{cases} \qquad (3)$$

Calculate the energy of each pixel point by the following formula

$$C(P_{i,j}) = \max_{l} \left[C(P_{l,j-1}) + g^m(l,k) \right] + D(P_{i,j}) \quad (l = i-1, i, i+1) \quad (4)$$

Where $g^m(i,k)$ denotes the cost from a pixel point of the previous column to this point, and three paths $g(P_{i,j}, P_{i-1,j-1})$, $g(P_{i,j}, P_{i,j-1})$, $g(P_{i,j}, P_{i+1,j-1})$ are defined as follows:

$$\begin{cases} g(P_{i,j}, P_{i,j-1}) = \Delta 1 \\ g(P_{i,j}, P_{i,j-1}) = \Delta 2 \\ g(P_{i,j}, P_{i,j-1}) = \Delta 3 \end{cases} \quad (5)$$

The parameters (Δ1, Δ2, Δ3) in the formula (5) represent the possibility of approaching the horizontal line that is the upper boundary of the document.

set Δ1=Δ3, Δ2=2*Δ1.

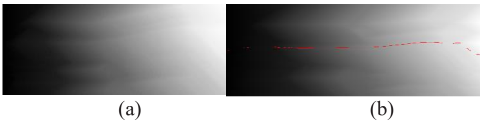
The parameters (θ) in the formula (3) , set θ=1.

A. Algorithm: Determine the boundary by DP

The width of the image is M (m = 1, 2, ..., M) and the height is N(n=1,2,⋯,N),

1. Specify initial costs of all nodes in the first Column, The cost values of all pixels in first column are initialized as $C(P_{i,1})(i = 1,2,\dots n) = D(P_{i,1})$
2. The cost function of the point $P_{i,j}$ is calculated according to:
   $C(P_{i,j}) = \max_{l} \left[C(P_{l,j-1}) + g^m(l,k) \right] + D(P_{i,j}) \quad (l = i-1, i, i+1)$
3. Repeat step 2 for m-1 times with j = 2……M.
4. Repeat step 3 to calculate the points of all rows with i = 1, 2, ..., N.

A candidate path can be obtained by finding the local maximum of energy function C(P_(i,M)) i=1,2,…n in the last column of pixels. The path in Figure 7 was got through backtracking from one point that has local maximum.



(a) (b)

Figure 7: (a) the energy map (b) A path backtracking

It's possible that there are multiple points that has local maximum values $C(P_{i,M})$ $i = 1,2,\dots n$, in the last column. since multiple candidate boundaries exist in one image. Utilizing non-maximum suppression algorithm to get the local maximum, then we backtrack from this point to get the candidate path. As shown in Figure 8 below, two paths are possible upper boundary of the document.

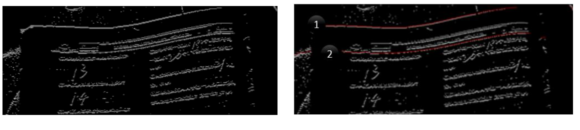


Fig 8: candidate paths

Through regional division , Determine the four boundaries of the document area upper(a), the bottom (b), the left (c), the right (d) as showed Fig. 9.

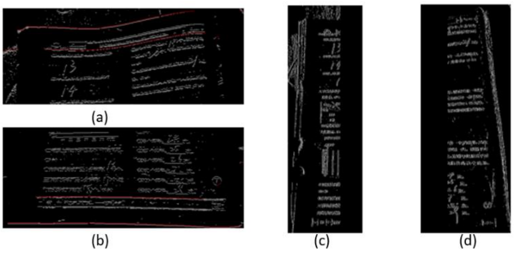


Fig 9 . four boundaries of document area

There are about 2 to 5 candidate edges in each aspect of the document area. There may be a lot of quadrilaterals, how to form a quadrilateral from the candidate edges, and there are many strategies. ZhengYou Zhang presented that quadrilateral has the following characteristics in the whiteboard image scanning and image enhancement [2]:

1. Parallel: The opposite two edges should have quite opposite or similar orientation ($180^0 \pm 30^0$).
2. Distance: The opposite edges should be quite far from each other.
3. Angle: The angle between two neighboring edges should be close to $90^0 \pm 30^0$.
4. Continuity: The four edges of consisting the quadrangle should be continuous.
5. Area: The area surrounded by the quadrangle should be large enough.

According to the five conditions, the KNN classifier is designed to screen out the real edges.

Table.1 experiment result(1)

Parallel	Distance	Angle	continuity	Area	Result
170	20	87	4	80	1
176	30	88	3	67	1
160	25	78	4	87	1

According to the test data of 5000 samples, the accuracy rate of classification is 97.9%.

Ⅲ. Experiments

Through testing the images, that contain fuzzy or uneven illumination or distortion of the image from the camera, the boundaries of the document area can be found by dynamic programming algorithm, and an example of the experiment is listed in fig. 9.

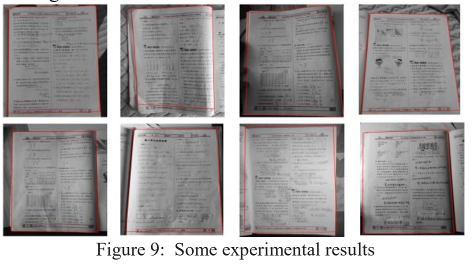


Figure 9: Some experimental results

 ISBN 978-1-5090-4897-7

1. Algorithm test on PC.

PC configuration:

CPU: I3 Processor I3-6100, Memory 8G.

The description of evaluation indicator

1. The success rate of detecting four edges: four sides are founded and form a quadrilateral from the background image.
2. The accuracy of determining the document area: Is the quadrilateral area the document area?
3. Time cost: the time of detecting four boundaries

Table.2 experiment result(2)

Quality	**edges detected**	**Success rate**	**document area**	**Success rate**	**Time cost**
3800	3600	94.70%	3500	92.10%	80ms

2. Spend time on mobile phone

Mobile Phone Model: Lenovo A60 (2013).

For the application on mobile phones, the speed requirement is very important because the response of the user will be slow if it takes much time on searching the edges, which is not a good experience. The rate of detecting edges integrated DP algorithm can reach to 7 frames / sec or more, which was obtained by testing on some common smart phones. Comparison of similar products such as Youdao cloud notes[23], time cost decreased to 85%.

Table.3 experiment result(3)

	Accuracy rate	Frame Rate
Algorithm of this paper	780/1000=78%	7.02fps
Algorithm of Youdao	720/1000=72%	6fps

IV. Conclusion

Experiments show that dynamic programming algorithm can solve detecting document area of mobile images, compared with the Hough transform algorithm, the dynamic programming algorithm has the advantages of less parameters, faster speed and higher accuracy. The algorithm is integrated in mobile APP. Since February 2016 on the line, more than 500,000 people used and gave good feedback, which proved that the algorithm is stable and efficient.

Acknowledgements

This work is supported by National Natural Science Foundation of China (No.61370195, U1536121).

References

[1]PVC Hough - U.S. Patent, No. 3069654 Method and means for recognizing complex patterns 1962.
[2]Bellmann R. Dynamic Programming. Princeton University Press, Princeton, NJ, 1957.
[3]Altman, N. S. An introduction to kernel and nearest-neighbor nonparametric regression". The American Statistician. 46 (3): 175–185. 1992.
[4]N Otsu A threshold selection method from gray-level histograms OHTSU Nobuyuki IEEE Trans. Syst., Man, Cybern. SMC-9(1), 62-66, 1979
[5]Sahoo P.K., et al. Survey of thresholding techniques Computer Vision, Graphics, and Image Processing, 41(2):233-260, 1998.
[6]Rosenfeld A. and Kalk A. C. Digital Picture Processing. Academic Press, New York, and edition, 1982
[7] Schettini R. A segmentation algorithm for color images. Pattern Recognition letters, 14:4999-506, 1993.
[8]Vlachos T. and Constantinides A. G. Graph-theoretical approach to color picture segmentation and contour classification. IEEE Proceedings Communication, Speech and Vision, 1993, 140:36-45.
[9] Pavlidis T. Structural Pattern Recognition. Spring Verlag, Berlin 1997.
[10]Demi M. Contour tracking by enhancing corners and junctions. Computer Vision and Image Understanding, 1996,63:118-134.
[11]Heijden F. v. d. Edge and line feature extraction based on covariance models. IEEE Transaction on Pattern Analysis and Machine Intelligence, 1995, 17:69-77.
[12]Law T., Itoh H., and Seki H. Image filtering, edge detection, and edge tracing using fuzzy reasoning. IEEE Transaction on Pattern Analysis and Machine Intelligence, 1996, 18:481 -491.
[13]An Optimal Fuzzy System for Edge Detection in Color Image Using Bacterial Foraging Algorithm IEEE Transactions on Fuzzy systems, February 2017, VOL. 25, NO. 1.
[14]Duda R.O. and Hart P. E. Using Hough transform to detect lines and curves in pictures. Communications of the ACM, 1972, 15(1):11-15.
[15]Xu L. and Oja E. Randomized Hough transform (RHT): Basic mechanisms, Algorithms, and computational complexities. CVGIP – Image understanding, 1993,57:131-154.
[16] zhengyou Zhang, Li_Wei He Microsoft Corporation, One Microsoft Way, Redmond, WA 98052-63999, USA Available online 15 June 2006
[17]Sun Feng-Rong Fast Hough Transform Algorithm Chinese J. Computers OCT 2001,Vol. 24 No. 10.
[18] AMIR A. Weymouth Using Dynamic Programming for solving variational problems in vision, IEEE Transaction Pattern Analysis and Machine Intelligence September 1990,VOL. 12 NO. 9.
[19] Milan Sonka et al. Image Processing, Analysis and Machine Vision. ISBN: 978-0-412-45570-4 (Print) 978-1-4899-3216-7
[20] Nicolas M Josiane Z. New Prospects in Line Detection by Dynamic Programming IEEE Transactions on Pattern Analysis and Machine Intelligence archive 1996 ,Volume 18 Issue 4,.
[21] kathrin Ungru, XiaoyiJiang Dynamic Programming Based Segmentation in Biomedical Image Computational and Structural Biotechnology Journal 15(2017)255-264
[22] Jose A. Rosado-Toro, Dynamic Programming Using Polar Variance For Image Segmentation IEEE Transactions on Image Processing, December 2016,VOL .25, NO. 12.
[23] Youdao cloud notes, http://note.youdao.com/