# **Text Detection Based on Affine Transformation**

Xiaoyue Jiang, Jing Liu<sup>(⊠)</sup>, Yu Liu, Lin Zhang, and Xiaoyi Feng

Northwestern Polytechnical University, Xi'an, China ljing@mail.nwpu.edu.cn

**Abstract.** Text detection and recognition plays an important roles in many computer vision based systems, since text can provide explicit content information. In natural scene, variations of scale, rotation and position are the main challenges for text detection and recognition algorithms. Thus, rectified text region is required for most text recognition algorithm. In this paper, we proposed a text detection method which can provide accurate text region. With the external quadrilateral of text region, the affine parameters can be estimated. Consequently, the distorted text region can be rectified according to the affine parameters. The proposed method can provide more accurate detection result for text region. In addition, it can enhance the performance of text recognition. The experiments show the effectiveness of the proposed method.

**Keywords:** Text detection  $\cdot$  Affine transformation  $\cdot$  MSER  $\cdot$  SWT Inertia spindle

## 1 Introduction

In recent years, with the development of Internet and the popularity of digital cameras and mobile phones, images and video resources are full of people's lives. The study of understanding the content of videos or image has become hotspots in Multimedia Information technology. The text in the videos and images contains rich semantic information, which is an important clue for understanding the content of images and videos. Detecting and recognizing the text in such images or videos is of great significance in many fields such as image understanding, video content analyzing and content based image or video retrieval. For text recognition, it can meet a certain degree of application requirements, and many companies have also released some commercial software packages [1, 2]. However when it is used in a natural scene, the performance will degrade. One of the key reasons is that the accuracy of the text detection reduces, thus the overall performance of the character recognition system decreases as well. Usually, a rectangular box is used to label the text region. However, due to complex background and different view position, text in images maybe shown in different position, orientation and scale. Thus, a rectangular box always cannot mark the text region accurately. Consequently, the distorted text image will also influence the performance of text recognition algorithms. Therefore a robust text detection and system should detect text in arbitrary position and orientation accurately.

The current text detection methods can be summarized as three categories as [3], gradient-based methods, the connected region-based methods and texture based

<sup>©</sup> Springer International Publishing AG 2017

Y. Zhao et al. (Eds.): ICIG 2017, Part I, LNCS 10666, pp. 420-431, 2017.

https://doi.org/10.1007/978-3-319-71607-7\_37

methods. Always, the text area is rich of edges, while the background region has relatively less edges. Based on this feature, edges are used to determine the existence of text. Some popular edge detectors are applied for the text detection, such as Sobel edge detector, Harris corner detector. However, when the background become complex, the edge features bring a lot of false alarms [4].

Text detection based on the connected region mainly utilizes the similarity of some features in a local region, such as the character stroke width and color [5-9]. With color-based feature, it can detect the text in different directions, but it is sensitive to the change of environment. Maximum Extreme Stability region (MSER) algorithms [10-12] consider the variance of local features to determine the region that has similar features. The MSER detector is invariant to rotation, scale and affine. It can effectively detect the text in variable view and scale.

For texture feature, Yi and Tian [13] proposed a text detection algorithm based on the gradient, color and geometric characteristics of characters. They concatenated the characters with similar features into the text area. The texture feature of text area is considered to be different from that of background. Zhao et al. [14] applied the theory of wavelet transform to extract the texture information from images. They divided the image into multiple patches and classified each patch into text region and background according to the texture feature.

Images can be taken from any position, thus text in image can appear in any pose. However, text detection method only use rectangular box to label text region which always includes a lot of background region. In this paper, we proposed a text detection method to detect text and label the text region more accurately. With the quadrilateral label for text region, the inertial spindle is used to estimate the affine parameters for text region. Then text regions can be rectified according to affine transform, where the rectified text can be recognized more accurately.

## 2 Text Detection

Given an image, to localize the text region precisely is the first step to guarantee the performance of the whole text recognition system. For each character, they share some similar features, such as color and texture. Thus, we applied the region-based feature to detect the text. However, the region-based feature will also introduce some false alarms. Then geometric feature and stroke width are used further for the purification of candidate region. With all these initial character detection results, a quadrilateral is introduced to local the text region accurately. The procedure of text detection in shown in Fig. 1.



Fig. 1. The flow chart of text detection



**Fig. 2.** Bounding box for text region. (a) bounding box for each character from text detection, (b) maximal external bounding box, (c) adjusted bounding box.

In order to detect all the possible text regions, region-based feature, MSER [10] is applied. The principle of MSER is based on watershed algorithm. The increment of the threshold is similar to rising the water level. As the water rises, some shallow valleys will be flooded. If you look down from the sky, the earth is divided into two parts, land and water, which is similar to the binary image. In image, the connect regions that has low variance in features will be merged together. Its mathematical definition is as follows:

$$\Delta q_i = |Q_{i+\Delta} - Q_{i-\Delta}| / |Q_i| \tag{1}$$

where  $Q_i$  denotes a certain connection region when the threshold is i,  $\Delta$  is a minimal change of the gray scale threshold, and  $q_i$  is the changing rate of the region  $Q_i$  when the threshold value is i. When  $q_i$  is a local minimum, then  $Q_i$  is the maximum stable extremum region. MSER is robust to affine transformation, thus it can effectively detect the text in the case of perspective transformation and scale change.

Although the MSER can pick out most of the text, but it can also include some parts without text. Non-text area can be discarded by examining the properties of the candidate region. Due to the geometric properties of text region, the width-to-height ratio, eccentric and Euler number are applied. Eccentricity is defined as the ratio between the main axis and the auxiliary axis of the image area. The Euler number is the total number of objects in the image minus the total number of holes in those objects.

Besides geometric feature, the stroke width is another robust feature. The text areas tend to have a smaller stroke width variation than non-text areas. Thus we applied stroke width detector (SWT) [15] to calculate the stroke width for candidate regions. The variance of stroke width is used to determine the existence of text.

### **3** External Bounding Box Estimation

After text detection, each candidate character will be labeled by a rectangle box. Always text region is represented by the maximal external box of all the candidate text region, as shown in Fig. 2(b). However, it includes many background and cannot represent the orientation of the text region, which is useful for further affine transform. In this section, we will introduce the proposed method for accurate external bounding box estimation as shown in Fig. 2(c).



**Fig. 3.** Three different cases to determine the external bounding box. (a) the condition of quadruple extreme points; (b), (c) the condition of triple extreme points and the external bounding box for it; (d), (e) two possible conditions for twin extreme points, (f) the external bounding box for twin extreme points case.

In order to determine the affine parameters, we need to find an accurate bounding box to represent the contour of text region. To determine the bounding box, the four corner points of the quadrilateral should be decided according to the distribution of characters. Since the bounding box of each character has presented the position of each character, the corners of the bounding box for each character are used to determine the external quadrilateral. In fact, we find the extreme points from all the corners of character bounding box, naming the maximal and minimal horizontal position as  $X_{\min}, X_{\max}$ , and maximal and minimal vertical position as  $Y_{\min}, Y_{\max}$ . According to the distribution of these extreme points, there are three different cases to determine the external quadrilateral.

#### 3.1 Quadruple Extreme Points

In this case, the four extreme points are not coincided together. That is, each extreme point represents one corner of external bounding box, as shown in Fig. 3(a). Then the external quadrilateral can be determined directly by these four points. With the external bounding box, the parameters of affine transform can be estimated. The details about affine transform will be introduced in Sect. 4.

#### 3.2 Triple Extreme Points

In this condition, two extreme value appears in one points, and the other two extreme value appears in two different points. Thus, the four extreme value can determine three

points of external bounding box, as shown in Fig. 3(b). With three points, a triangle can be determined. Then the forth point should be determined by the distribution of characters. First an external rectangular can be found for the three extreme points defined triangle by drawing lines parallel to axis through the triangle points. Then the region inside the rectangular is divided into four sub-regions, and there are three sub-regions outside the triangle. Then the forth point should lie in the region where most characters stay. Here the density of character can be described by the density of corner points of character bounding boxes. Then the character density of these three sub-regions is calculated and compared. The corner point of the sub-region that has the maximum character is selected as the forth point for the external quadrilateral. Figure 3 (c) shows a solution for triple extreme points.

### 3.3 Twin Extreme Points

In this condition, the four extreme points can only determine two corner points because of the coincidence of extreme points, as shown in Fig. 3(d), (e). Then we need to determine another two points for the external quadrilateral, which is based on the distribution of characters. In fact, the inertia spindle of the detected characters represents the main orientation of text region. Thus we propose to use the direction of inertia spindle to determine the other two points of the external bounding box. For each of the extreme points, two lines that are parallel and appendicular to the inertia spindle can be set, respectively. Then the intersection of two sets of lines can determine the other two corner points of the external bounding box, as shown in Fig. 3(f).

Similarly, the calculation of inertia spindle of text region also can be based on the corner points of character bounding box. Let C represents a set of corners  $(x_i, y_i)$ , i = 1, 2, ..., N where N is the total number of corners of the detected character bounding boxes. The centroid point of all the corners is defined as  $(\bar{x}, \bar{y})$ . Then the moment of inertia for corner points C is defined as

$$G_c = \sum_{i=1}^{N} \left[ (xi - \bar{x})^2 + (yi - \bar{y})^2 \right]$$
(2)

The moment of inertia for corner set C against the straight line L with angle  $\boldsymbol{\theta}$  is defined as

$$G_{\theta} = \sum_{i=1}^{N} \left[ (xi - \bar{x}) \sin \theta - (yi - \bar{y}) \cos \theta \right]^2$$
(3)

The inertia spindle is defined as the angle  $\hat{\theta}$  which make the moment of inertial  $G_{\theta}$  minimal as

$$\hat{\theta} = \arg\min G_{\theta} \tag{4}$$

Then let the derivative of inertia moment  $G_{\theta}$  equal to zero, as  $G'_{\theta} = 0$ 

$$G'_{\theta} = 2 \sum_{i=1}^{N} \left[ (xi - \bar{x}) \sin \theta - (yi - \bar{y}) \cos \theta \right] \cdot \left[ (xi - \bar{x}) \cos \theta - (yi - \bar{y}) \sin \theta \right]$$
  
= 
$$\sum_{i=1}^{N} \left[ (xi - \bar{x})^2 - (yi - y)^2 \right] \sin 2\theta - 2 \sum_{i=1}^{N} (xi - \bar{x}) (y_i - \bar{y}) \cos 2\theta$$
(5)

then

$$\tan 2\hat{\theta} = \frac{2\sum_{i=1}^{N} (xi - x)(yi - y)}{\sum_{i=1}^{N} [(xi - \bar{x})^2 - (y_i - \bar{y})^2]}$$
(6)

Let  $m_{20} = \sum_{i=1}^{N} (x_i - x)^2$ ,  $m_{02} = \sum_{i=1}^{N} (y_i - y)^2$ ,  $m_{11} = \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$ , then Eq. (5) can be written as,

$$\tan 2\hat{\theta} = \frac{2m_{11}}{m_{20} - m_{02}} \tag{7}$$

Because

$$\tan 2\theta = \frac{2\tan\theta}{1-\tan^2\theta} \tag{8}$$

Then bring Eq. (9) into Eq. (3), we can get

$$m_{11}\tan^2\hat{\theta} + (m_{20} - m_{02})\tan\hat{\theta} - m_{11} = 0$$
(9)

$$\tan \hat{\theta}_{1,2} = \frac{-(m_{20} - m_{02}) \pm \sqrt{(m_{20} - m_{02})^2 + 4m_{11}^2}}{2m_{11}}$$
(10)

For the two orientation  $\hat{\theta}_{1,2}$ , the one makes the second derivative of  $G_{\theta}$  greater than zero is selected as the direction of inertia spindle. Then the inertia spindle can be described as the line passing through centroid point with the direction of  $\hat{\theta}$ 

$$y_i - \bar{y} = (x_i - \bar{x}) \tan \theta \tag{11}$$

Let  $\rho = \tan \hat{\theta}$ , then the other two points of the external bounding box can be determined by the intersection of parallel and perpendicular lines to direction  $\rho$  through extreme points. Equations 14–17 are the equations for four lines.

$$y_{1} = \rho(x - x_{\max}) + y_{\min}; \quad y_{2} = -\frac{1}{\rho}(x - x_{\max}) + y_{\min}$$
  

$$y_{3} = \rho(x - x_{\min}) + y_{\max}; \quad y_{4} = -\frac{1}{\rho}(x - x_{\min}) + y_{\max}$$
(12)

Solve the Eq. 17, the coordinates of the other two corners of the external bounding box can be calculated as,

$$\left(\frac{\rho^2 x_{\min} + \rho(y_{\min} - y_{\max}) + x_{\max}}{\rho^2 + 1}, \frac{\rho^2 y_{\min} + \rho(x_{\max} - x_{\min}) + y_{\max}}{\rho^2 + 1}\right)$$
(13)

$$\left(\frac{\rho^2 x_{\max} + \rho(y_{\max} - y_{\min}) + x_{\min}}{\rho^2 + 1}, \frac{\rho^2 y_{\max} + \rho(x_{\min} - x_{\max}) + y_{\min}}{\rho^2 + 1}\right)$$
(14)

## **4** Affine Transformation

After the external bounding box is calculated, the parameters for affine transform should be estimated first, then the affine transform will be applied to rectify the distorted text region.

#### 4.1 Affine Parameter Estimation

According to imaging principle, the original text will be projected to image plane. For two parallel lines  $l_1, l_2$  in image plane, where  $l_i : a_i x + b_i y + c_i = 0 (i = 1, 2)$ . The text plane  $(r_1, s_1, t_1)$  can be represented as

$$\begin{pmatrix} r_1 \\ s_1 \\ t_1 \end{pmatrix} = \frac{1}{\sqrt{a_1^2 + b_1^2}} \begin{pmatrix} b_1 \\ -a_1 \\ 0 \end{pmatrix}$$
(15)

If  $l_1, l_2$  are not parallel in the image plane, and the vanishing point is  $(x_1, y_1)$ , then the original text plane  $(r_1, s_1, t_1)$  can be represented as

$$\begin{pmatrix} r_1 \\ s_1 \\ t_1 \end{pmatrix} = \frac{1}{\sqrt{x_1^2 + y_1^2 + f^2}} \begin{pmatrix} x_1 \\ y_1 \\ f \end{pmatrix}$$
(16)

where f is the focus length of the camera. Then with two set of lines in the image plan, for example the two sets of lines of the external bounding box, the original text plane can be determined, as

$$\begin{pmatrix} r_3 \\ s_3 \\ t_3 \end{pmatrix} = \begin{pmatrix} r_1 \\ s_1 \\ t_1 \end{pmatrix} \times \begin{pmatrix} r'_2 \\ s'_2 \\ t'_2 \end{pmatrix}$$
(17)

where  $(r'_2, s'_2, t'_2)$  is determined by another pair of lines in image plane.

### 4.2 Affine Correction

With the estimation for the original text plane, we can rectify the distorted text in the image plane. Figure 4 shows the relationship between the text plane and image plane. The mapping from a point  $(X_t^0, Y_t^0, Z_t^0)$  on the text plane  $O_t X_t Y_t Z_t$  to a point  $(x_i, y_i)$  of the image plane can be written as follows:



Fig. 4. The projection of text to image

$$\begin{pmatrix} x_t^0 \\ y_t^0 \end{pmatrix} = \begin{pmatrix} f & 0 \\ 0 & f \end{pmatrix} \begin{pmatrix} \left( \frac{X_{it}^0 + X_0}{Z_{it}^0 + Z_0} \right) \\ \left( \frac{Y_{it}^0 + Y_0}{Z_{it}^0 + Z_0} \right) \end{pmatrix}$$
(18)

where

$$\begin{pmatrix} X_{it}^{0} \\ Y_{it}^{0} \\ Z_{it}^{0} \end{pmatrix} = \begin{pmatrix} r_{1} & r_{2} & r_{3} \\ s_{1} & s_{2} & s_{3} \\ t_{1} & t_{2} & t_{3} \end{pmatrix} \begin{pmatrix} X_{t}^{0} \\ Y_{t}^{0} \\ Z_{t}^{0} \end{pmatrix}$$
(19)

In order to reconstruct the front view of the text plane, we can use the affine transformation to correct.

$$\begin{pmatrix} x_{it}^{0} \\ y_{it}^{0} \end{pmatrix} = \begin{pmatrix} f & 0 \\ 0 & f \end{pmatrix} \begin{pmatrix} \left( \frac{X_{t}^{0} + X_{0}}{Z} \right) \\ \left( \frac{Y_{t}^{0} + Y_{0}}{Z_{0}} \right) \end{pmatrix}$$
(20)

In the text plane,  $Z_t^0 = 0$ . Since the origin point  $(X_0, Y_0, Z_0)$  of  $O_t X_t Y_t Z_t$  and  $O_{it} X_{it} Y_{it} Z_{it}$  is mapped to the same point  $(x_0, y_0)$  in the image panel, we can get

$$x_0 = f \frac{X_0}{Z_0}, y_0 = f \frac{Y_0}{Z_0}$$
(21)

Then the correction can be done by Eq. (22)

$$x_{it}^{0} = x_{0} + f \frac{\begin{vmatrix} x_{0} - x_{t}^{0} & r_{2}f - t_{2}x_{t}^{0} \\ y_{0} - y_{t}^{0} & s_{2}f - t_{2}y_{t}^{0} \end{vmatrix}}{\left| r_{1}f - t_{1}x_{t}^{0} & r_{2}f - t_{2}x_{t}^{0} \\ s_{1}f - t_{1}y_{t}^{0} & s_{2}f - t_{2}y_{t}^{0} \end{vmatrix}}, y_{it}^{0} = y_{0} + f \frac{\begin{vmatrix} r_{1}f - t_{1}x_{t}^{0} & x_{0} - x_{t}^{0} \\ s_{1}f - t_{1}y_{t}^{0} & y_{0} - y_{t}^{0} \end{vmatrix}}{\left| r_{1}f - t_{1}x_{t}^{0} & r_{2}f - t_{2}x_{t}^{0} \\ s_{1}f - t_{1}y_{t}^{0} & s_{2}f - t_{2}y_{t}^{0} \end{vmatrix}}$$
(22)

That is for each point on the image plane, it can be mapped to a new position according to the affine transform. In order to make the image more consistent, the bilinear interpolation method is applied to reconstruct the rectified text images.



**Fig. 5.** Rectified results. (a) bounding box for quadruple extreme points case. (b), (c) bounding box for triple extreme points case. (d) bounding box for twin extreme points case. (e), (f), (g) rectified results for (a), (d), (b), respectively.

# 5 Experimental Results

We used the MSRA-TD500 database [16] to test the performance of the algorithm. This dataset contains 500 indoor and outdoor natural images. The indoor images are mainly signs, doorplates and caution plates while the outdoor images are mostly guide boards and billboards in complex background. The resolutions of the images vary from  $1296 \times 864$  to  $1920 \times 1280$ . Some of the rectified results in different cases are shown in Fig. 5. We can see the proposed method can provide accurate detection for text region in different conditions. With the external quadrilateral of text region, the parameters of affine transform can be estimated precisely and consequently distorted images can be rectified. For Fig. 5(d), the corner points mainly lies in the middle of text region. Thus the detected text region is shifted. However, the inertia spindle still provides the main direction of the test, and consequently it can be rectified as shown in Fig. 5(f).

In order to quantify the effectiveness of the detection algorithm, the commonly used accuracy rate (P), recall rate (R) and F-score (Eq. 23) are applied

$$P = \frac{tp}{tp + fp} \times 100\%, \ R = \frac{tp}{tp + fn} \times 100\%$$
(23)

$$F = 2\frac{P \times R}{P + R} \tag{24}$$

where tp is the area of the text region correctly detected in the image, fp is the non-text area in the image that is erroneously detected, that is, the false detection area. fn represents the text area that is not detected in the scene image. F score is the harmonic mean of precision and recall rate. The proposed method (TD-Affine) is test on the MSRA-TD500 database, it achieved precision rate of 0.58 and recall rate of 0.62. Compared with other methods as shown in Table 1, the proposed method keeps a high recall rate which is achieved by the MSER method. Then geometric and stroke width feature are used to remove false alarms. The final text region is determined due to different distributions of character. Since we try to keep the main direction of text region, but not the exact test region, therefore the precision rate still can be improved. Keeping the main direction is useful for the estimation of affine transformation. Figure 6 shows more results of affined rectified images. For the first two rows, all the text are in the same direction, then the proposed method can re-project the text to the frontal view. But when text are in different directions, as the third row, the global inertia spindle is balanced among local regions. Thus the proposed method cannot adjust each local region. Similarly, when the background is complex, as the forth row, it will also influence the correct estimation of inertia spindle for the text region. According the wrongly estimated affine parameters results in the wrongly rectification.

In order to shown the effectiveness of affine correction, we also performed text recognition experiments on the dataset. We used the ABBYY Finereader [17] to recognize the text. Table 2 shows the performance of text recognition with and without affine correction. From the results, we can see that the proposed affine parameter estimation and correction method can improve the text recognition rate. The proposed method mainly consider the main direction of text region, even when the detection region is not so accurate, the parameters of affine transform still can be accurately estimated. Consequently, the distorted images can be rectified precisely.

Algorithm	Precision	Recall	F-score
TD-Affine (proposed)	0.58	0.62	0.60
TD-ICDAR [16]	0.53	0.52	0.52
Epshtein et al. [9]	0.25	0.25	0.25
Chen and Yuille [6]	0.05	0.05	0.05

Table 1. Comparison of text detection results on MSRA-TD500

	Precision	Recall	F-score
Before affine	0.76	0.75	0.75
After affine	0.82	0.84	0.83

Table 2. Text recognition results on MSRA-TD500



Fig. 6. More results for affine rectification. The left column shows the character detection results, the right column shows the corresponding rectified results.

# 6 Conclusion

Text always provide important context information for images. In order to realize the text detection and recognition in natural environment, this paper proposed an affine transform based test detection algorithm. Since text detection is always an essential prerequisite for text recognition, providing high quality images for recognition is

necessary. Thus we not only try to detect the location of text but also try to estimate the distortion of text region. In this paper, we utilize the MSER and SWT algorithms to detect the text area, and then combines the inertial spindle with some known extreme points to find the four vertices of the outer quadrilateral of the text area. Affine parameters can be obtained based on the four vertices, which will be used for affine rectification for distorted text. As a result, the text recognition rate can be improved significantly with the rectified text input.

## References

- 1. Ou, W., Zhu, J., Liu, C.: Natural scene text positioning. J. Chin. Inform. Process. 18(5), 42–47 (2004)
- Maio, D., Maltoni, D., Cappelli, R., et al.: Fingerprint verification competition. Trans. Pattern Anal. Mach. Intell. 24(3), 402–412 (2002)
- 3. Ye, Q., Doermann, D.: Text detection and recognition in imagery: a survey. IEEE Trans. Patten Anal. Mach. Intell. **37**(7), 1480–1500 (2014)
- Jiang, X., Lian, J., Xia, Z., Feng, X., Hadid, A.: Fast Chinese character detection from complex scenes. In: The Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA 2016) (2016)
- Wu, H., Zhao, Y., Li, X., Zou, B., Zou, R.: Natural scene text detection based on color clustering. J. Central South Univ. (Sci. Technol.) 6(7), 1672–7207 (2015)
- Chen, X., Yuille, Y.: Detecting and reading text in natural scenes. In: Proceedings of CVPR (2004)
- Wang, W., Fu, L., Gao, W., Huang, Q., Jiang, S.: Superimposed text detection method based on stroke character. Chin. J. Image Graph. 12(019), 1354–1361 (2015)
- Liu, Y., Xue, X., Lu, H., Guo, Y.: A video character detection algorithm based on edge detection and line feature. J. Comput. 03(18), 0254–4164 (2015)
- 9. Epshtein, B., Ofek, E., Wexler, Y.: Detecting text in natural scenes with stroke width transform. In: Computer Vision and Pattern Recognition, pp. 2963–2970. IEEE (2010)
- Matas, J., Chum, O., Urban, M., et al.: Robust wide—baseline stereo from maximally stable extremal regions. Image Vis. Comput. 22(10), 761–767 (2004)
- Tang, Y., Bu, W., Wu, X.: Multi-level MSER natural scene text detection. J. Zhejiang Univ. (Eng. Sci.) 50(6), 135–1140 (2016)
- 12. Dai, J.: MSER-based text detection method. Tianjin Normal University (2014)
- 13. Yi, C., Tian, Y.: Text string detection from natural scenes by structure-based partition and grouping. IEEE Trans. Image Process. **20**(9), 2594–2605 (2011)
- Zhao, M., Li, S., Kwok, J.: Text detection in images using sparse representation with discriminative dictionaries. Image Vis. Comput. 28(12), 1590–1599 (2010)
- Song, W., Xiao, J.: Natural scene text detection method based on stroke width transform. Comput. Eng. Appl. 49(9), 190–192 (2013)
- Yao, C., Bai, X., Liu, W., Ma, Y., Tu, Z.: Detecting texts of arbitrary orientations in natural images. In: IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, pp. 1083–1090 (2012)
- 17. Richards, M.: ABBYY FineReader 9. APC (2008)