형태 모멘트를 이용한 텍스트 이미지 경사 측정 및 교정

Skew Estimation and Correction in Text Images using Shape Moments

주문원
성결대학교 멀티미디어학부
진성아
성결대학교 멀티미디어학부

Moon-Won Choo
Division of Multimedia, Sungkyul University
Seong-Ah Chin
Division of Multimedia, Sungkyul University

중심어 : document image processing, skew estimation, page segmentation

요 약

문서 이미지 처리에서 텍스트 블록의 수평화 프로세스는 문서 인식 솔루션을 위한 전처리 단계로서 많은 연구가 진행되고 있다. 이 논문에서는 텍스트 이미지 블록의 직교각 속성과 형태 모멘트에 후프 변환을 적용하여 경사진 텍스트 블록을 원래 문서의 텍스트와 수평화된 텍스트 이미지로 변환하는 효율적인 방식을 제안한다. 실험을 통하여 제안된 방식의 비교 성능 결과를 보인다.

Abstract

In this paper efficient skew estimation and correction approaches are proposed. To detect the skew of text images, Hough transform using the perpendicular angle view property and shape moments are performed. The resultant primary text skew angle is used to align the original text. The performance evaluations of the proposed methods with respect to running time are shown.

I. Introduction

Images convey vast amounts of information and many images contain embedded or overlaid text. The text in this image may be as meaningful as the image itself. For coding or understanding document images it is essential to identify and possibly separate text, images and graphical regions into physically distinct segments of the page in order to be able to process them properly[1],[2]. One of complexity of document processing is thought of as the skew angle detection on embedded or overlaid text images since graphical or picture objects in images may affect to the skew angle without regarding text skew. Most existing page segmentation algorithms do not handle text images with skew[4],[5]. Yu and Jain designed a skew detection algorithm using hierarchical Hough transform to connected components[6]. C.L. Tan[7], however, computes the orientation of text strings without using Hough transform whereas they calculate the angle by arctan computation on the centers of a pair of connected components. Some existing methods[10],[11,][12] are capable of detecting only a limited angle range from 5 to 45 while other methods[13],[14] can detect an arbitrary skew angles.

Two skew detections focusing on embedded text images into other objects as well as document images are proposed in this paper. In the first method, text is noted to be able to be aligned any orientation with skew angles. Highly accumulated angles might have angle distance 90o, perpendicular angle view made by vertical and horizontal stroke in text images. Perpendicular angle view sufficiently provides peak detection after mapping the image space into Hough space. In the second method, shape moments are be adapted in order to compute fast skew estimation in text images. Shape moments provide efficiently useful and fast features in detecting a skew angle. By experiments, the performance comparison is given with respect to running time complexity.

II. Methods

The first method is primarily based on recognition of perpendicular angle view in edge points from text images. In order to explore candidate skew angles, Hough transform is used to map edge image space into Hough space. Shadow vector and Threshold Projection are introduced as significant procedures. Another proposed method is using shape moments to derive the skew in text images, which is faster than the first method.

1. Skew and Perpendicular Angle View

Skew angle is defined as shown Fig. 1, where θ_a is the angle with respect to the vertical x axis. Text in images may be aligned horizontally, which results in the concept of angle view. This illustrates how text angles relate to angles detected as shown in Fig. 1. The skew angle θ_a frequently hinders extracting accurate text region in images as preprocessing steps to recognition.

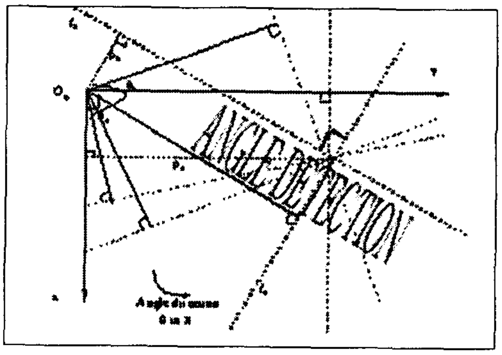


Fig. 1. Skew angle θ_a and Relation between Perpendicular Angle View and Text images

The main idea of the method is based on the principles of perpendicular angle view that text images almost carry. On the edge images, high appearance of edge points on the strokes can be seen as shown in Fig. 1. Those cross stroke edge lines has 90o angle distance called Perpendicular Angle View. These features are mainly dealt with as fundamental starting point to the next process. Even if there exist many straight lines at various angles in text images, the angle views θ_a and θ_b can be identified specifically as shown in Fig. 1. These angles can be used to transform edge image properly. Hough transform not only extracts line segments but also detects non-continuous collinear points. However, that may not guarantee to provide a way to detect an accurate skew angle. Hence, a way to detect angles θ_a and θ_b which are orthogonal is derived.

2. Functions of the angle

Generally speaking, edges are defined as the boundary between two regions with relatively distinct gray levels. Hough transform is well known for detecting straight lines in various low level image processing applications. All points at a straight line in image space can be represented by one point in which the angle θ and the distance ρ are called normal representation of a line. Hough Transform maps edge points in the image space into an accumulator array representing $H(\theta, \rho)$[3],[8],[9]. As the edge points are highly accumulated along the strokes in text, these features can be seen as maximum number of occurrences on the Hough matrix along angle direction called Shadow Vector, $P_s(\theta)$ in the equation (1). If there are more frequencies at the angle, then more points at the straight line in image space could be assumed. The angle and distance can be identified by noting this simple fact. But, multiple peaks may be detected in an accumulator array since the main angle may not necessarily to be the maximum peak. From this observation, a threshold for frequencies of angles is set up to retain the angles greater than the threshold in $H(\theta, \rho)$. The summation of the frequency over angles after thresholding called Threshold Projection Vector $P_\tau(\theta)$ in the equation(3) offers the solution of proper main text angle.

$$P_s(\theta) = \max{}_\rho H(\theta, \rho) \quad (1)$$

$$P_\rho(\theta) = \sum_{\rho=1}^{\max \rho} H(\theta, \rho) \quad (2)$$

$$P_\tau(\theta) = \sum_{\rho=1}^{\max \rho} u(H(\theta, \rho) - \tau_\theta) H(\theta, \rho) \quad (3)$$

where τ_θ, is the threshold value and $H(\theta, \rho)$, Hough matrix of the image, $1 \le \rho \le \max_\rho = \sqrt{N^2 + M^2}$ for a fixed θ with image size $N \times M$. Also

$$u(x) = \begin{cases} 1 & \text{if } (x \ge 0) \\ 0 & otherwise \end{cases} \tag{4}$$

The final goal is to find the angle associated with a straight line, which should be easily detected. To do that, the summation along angles in the Hough matrix after thresholding on the number of occurrence in the Hough matrix is needed to be computed. This implies that the angle density is high at the specific angle. This line offers many intersection points at the specified angle. In order to extract an useful angle from the Hough matrix, various one dimensional function of the angle derived from the Hough matrix is utilized. Some of the functions include the equations(2), (3) and (4).

3. Peak Detection, Angle Projection and Skew Angle

A perpendicular angle view in text is useful for detecting peaks since a perpendicular angle has a 90o difference between the two angles(θ_a and θ_b). To find the angles, a peak detection mask is implemented such as PeakMASK=[-1 2 1 0 0....0....0 -1 2 -1]T, where the total number of entries is 90. This mask extracts peaks that have a distance of 90o by investigating the shadow $P_s(\theta)$ which displays only the maximum number of occurrences at angles. Sequentially, PeakMASK is convoluted at the range of angles (0+n-1) to (90+n-1). For example, PeakMASK is applied from 0 to $\pi/2$ on the maximum number of occurrence of each angle in Fig. 2(a) to find the peak candidates. The resultant peak candidates are visible in Fig. 2 (b) at θ_p =10 which automatically implies another peak at $\theta_{p'}$ =100 due to the 90o angle distance at PeakMASK. These two peaks are used to determine

The main skew angle between the two peaks are determined by counting the number of occurrence at both angles. The results of the counts are obtained by threshold projection, $P_\tau(\theta)$ along angles at Hough matrix $H(\theta, \rho)$.

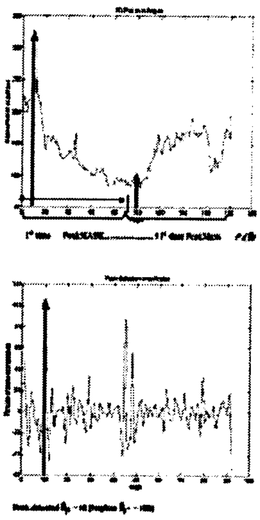


Fig. 2. (a)Maximum number of occurrence at each angle
(b) Peak detection after applying PeakMask

Hough matrix $H(\theta, \rho)$ may contain many high values at the angle θ, but none of such angles may be the maximum value. That means the maximum peak do not necessarily represent the main angle. Hence, a threshold value τ_θ is chosen from $H(\theta, \rho)$ then only the summation of the number of occurrence of those values which is greater than the threshold along the angle θ, called projection, is computed.

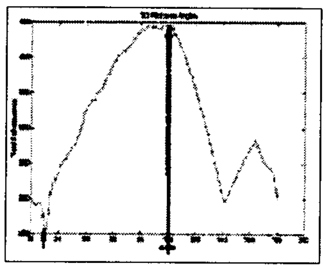


Fig. 3. Threshold Projection on Peak candidates: $\theta_p = 10$ and $\theta_{p'} = 100$ in Fig. 2

By this process, the peak over threshold projection $P_{\tau}(\theta)$, can be picked and mainly skew angle can be properly computed. By choosing the threshold value using the number of occurrences, angle θ_a could be correctly selected, since text lines hold many vertical strokes la as shown in Fig. 1.

Higher values at the Hough matrix imply more intersection points at sinusoidal curves indicating a line existence in the image space. The higher peak between θ_p and $\theta_{p'}$ in the angle projection matrix $P_{\tau}(\theta)$ directly represents the mainly skew angle in the image as shown Fig. 3 : $P_{\tau}(\theta_{p'} = 100)$ is higher than, $P_{\tau}(\theta_p = 10)$. Thus, θ_s is set to 100. By this procedure, the mainly skew angle θ_s , which is composed of the x axis and θ_a can be computed.

4. Shape Moments

If the shape moments are known, the skew direction θ is calculated as follow :

$$\theta = \frac{1}{2}\tan^{-1}\left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}}\right), \qquad (5)$$

where

$$\mu_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - x_c)^p (y - y_c)^q f(x, y)\,dxdy$$

and x_c and y_c are the region's center of gravity. Shape moments are calculated on the edge images. This method is more efficient than the first method with respect to running time complexity and simplicity of procedures.

III. Horizontal alignment of images

For horizontal alignment, all points in the edge matrix is transformed by taking two steps as follows. At first, the combined transformation matrix, $T_{(Mcos\theta, 0)}R_{(\pi/2-\theta)}$ is computed. Then this combined matrix is used for transformation, which saves computing time by time complexity O(n). The image matrix is rotated by $(\pi/2 - \theta)$ radians counter clockwise and translated $Mcos\theta$ units in the x direction. Then the transformed image matrix size has been changed from N to $Nsin\theta + Mcos\theta$ and from M to $Msin\theta + Ncos\theta$ respectively as shown in Fig. 4. The new N and M is denoted as, $N' = Nsin\theta + Mcos\theta$ $M = Msin\theta + Ncos\theta$. M' can also be detailed as $M' = \sqrt{N^2 + M^2}\cos(\tan^{-1}(M/N) - \theta)$. In the case of $\pi/2 < \theta \leq \pi$, the edge matrix is rotated by the counter clock wise by $\theta - \pi/2$ as same as $R_{(\pi/2-\theta)}$ and translated to y axis with $Ncos(\pi - \theta)$).

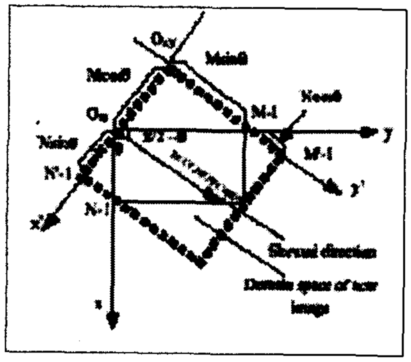


Fig. 4. Relationship of the new image axes(x',y') to the original x axes(x,y) and the position of the new origin Ox'y' with the skew angle θ.

The algorithm performs the transformation of the skew image as below :

```
Algorithm : Transformation of skew image : f'(x', y')
Input : N,M, ,θ ,f(x, y)
Output : f' (x', y')
% New image size
if    (θ ≤ π/2)
          N' = Nsinθ    +    Mcosθ ;
          M' = Msinθ    +    Ncosθ ;
else  if   (θ > π/2)
          N'=  Ncos(θ -π/2)  +   Mcos(π-θ)
          M'=  Ncos(π-θ)     +   Mcos(θ -π/2)
```

```
        end if
    end if
    f'(N', M')=0 ;

    % Edge transformation
    for x=0 to N-1
        for y=0 to M-1
            if    (θ ≤ π/2)
                        x'  = xsinθ - ycosθ + Mcosθ  ;
                        y'   =  xcosθ + ysinθ  ;
                        f(x', y')= f(x, y) ;
            else  if   (θ > π/2)
                        x'  = xsinθ - ycosθ ;
                        y'   =  xcosθ + ysinθ- Ncos(π-θ) ;
                        f'(x', y')= f(x, y) ;
            end if
        end if
    end for
    end for
```

IV. Experimental Results

For the evaluation of the proposed methods, a number of experiments were carried out. Image files were captured from a scanner and digital camera. The experiments were conducted in a MATLAB environment under Pentium4 Compaq Desktop workstation. In this experiments it was found that 10 to 12 point size of text should be captured with a minimum of 400dpi resolution since lower resolution cause significant data loss especially with small type fonts. This proposed methods successfully detect accurate skew angle in any arbitrary angle. In Fig. 6, (a), (c) and (e) are skew input images and (b), (d) and (f) rotated and extracted images applying transformations. Table 1 shows the experimental results in terms of parameters such size, resolution, skew angle error denoted by es which can be computed by difference between the real skewed angle θ_r and the detected skew angle θ_s, and the comparison between two proposed methods with respect to running time in sec.

Table 1. Running Time comparisons(documents images of Fig. 6).

Trials	Size	DPI	Hough (sec)	O'*Gorman & Clowes's*	Shape *Moments (sec)*	θ_s	θ_r	e_s
Fig. 5(a)	580X434	300	11.92	1.89	0.93	59	60	1
Fig. 5(c)	532X262	96	26.65	6.89	3.01	119	120	1
Fig. 5(e)	425X337	300	33.75	17.61	6.06	30	30	0

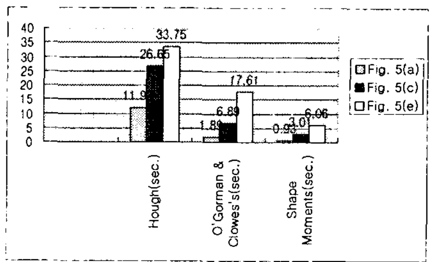
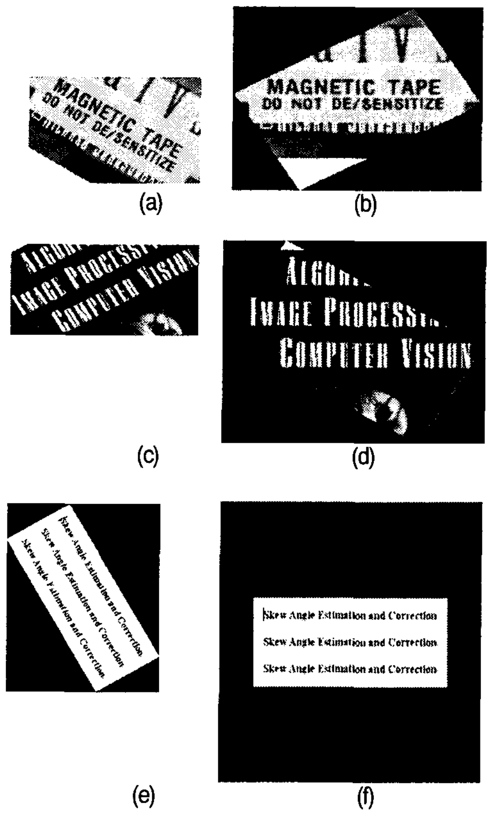
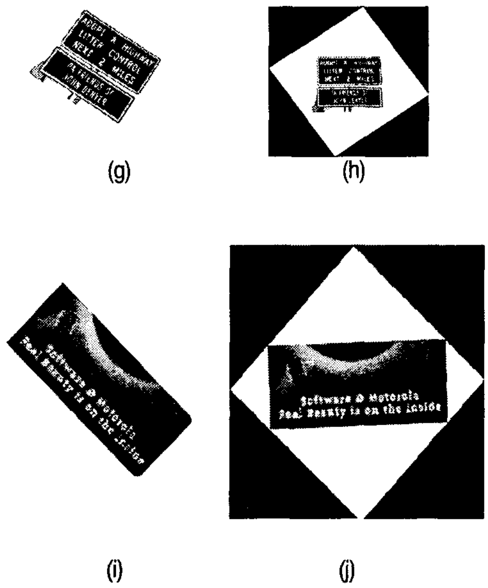


Fig. 5. Running time comparison: Hough, O'Gorman and Shape moments in sequence.





(g) (h)

(i) (j)

Fig. 6. Sample experimental images: (a),(c),(e),(g), and (i) are input images with the skew corrected images(b),(d),(f),(h) and (j) respectively.

V. Conclusion

A novel skew angle detection algorithms using perpendicular angle and shape moments are presented. This work is focusing on figuring out the skew angle especially on embedded text images into any objects. The approach proposed here seems to be promising since the proper principles to identify orthogonal angle view is followed. Also it shows acceptable experimental results. In conclusion, shape moments provide the more efficient running time comparatively.

References

[1] Rasshid. Malik and Seongah CHIN, "Extraction of Text in images," IEEE International Conference On Information Intelligence, ICIIS99, pp. 534-537, Oct. 1999.

[2] Kamran Etemad, David Doermann, and Rama Chellappa, "Multiscale Segmentation of unstructured Document Pages Using Soft Decision Integration," IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 1, 19, No. 1, pp. 92-96, Jan. 1997.

[3] Robert M. Haralick, Linda G. Shapiro, Computer and Robot Vision, Vol. 1, Addison Wesley, 1992.

[4] Anil K. Jain and B. Yu, "Document Representation and Its Application to Page Decomposition," IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 20, No. 3, pp. 294-308

[5] H. Baird, "Anatomy of a Versatile Page Reader, " Proc. IEEE, Vol. 80, pp. 1,059-1,065, 1992.

[6] B. Yu and A. Jain, "A Robust and Fast Skew Detection Algorithm for Generic Documents," Pattern Recognition, Vol. 29, pp. 1,599-1,629, 1996.

[7] C. L. TAN and P. O. NG, "Text Extraction Using Pyramid," Pattern Recognition, Vol. 31, No.1, pp. 63-72, 1998.

[8] R. Duda and P. Hart "Use of the Hough Transformation to Detect Lines and Curves in Pictures," Graphics and Image Processing, Vol. 15, No. 1, pp. 11-15, Jan. 1972.

[9] Rafael C., Gonzalez Richard E. Woods, Digital Image Processing, Addison Wesley, 1993.

[10] Avanindra and S. Chaudhuri, "Robust Detection of Skew in Document Images," IEEE Trans. on Image Processing, Vol. 6, No. 2, pp. 344-349, 1997.

[11] B. Gatos, N. Papamarkos, and C. Chamzas, "Skew Detection and Text Line Position Determination in Digitized Documents," Pattern Recognition, Vol. 30, No. 9, pp. 1505-1519, 1997

[12] U. Pal and B.B. Chaudhuri, An Improved Document Skew Estimation Technique," Pattern Recognition Letters, Vol. 17 No. 8, pp. 899-904, 1996.

[13] B. Yu, and A. Jain, "A Robust and Fast Skew Detection Algorithm for Generic Documents," Pattern Recognition, Vol. 29, No. 10, pp. 1599-1629, 1996.

[14] Oleg Okun, Matti Pietikainen and Jaakko Sauvola, "Robust Document Skew Detection Based on Line Extraction," Proc.of the 11th Scandinavian Conf. On Image Analysis, June 7-11, pp. 457-464, 2002

주 문 원(Moon-Won Choo) 정회원

1981년 ~ 1985년 : San Jose State. 대학교 Mathematics학과 (이학사)
1986년 ~ 1987년 : New York Tech. 대학원 computer science학과 (이학석사)
1988년 ~ 1990년 : 삼성전자 기흥연구소 (시스템 소프트웨어연구원)
1991년 ~ 1996년 : Stevens Tech. 대학교 computer science전공(이학박사)
1997년 ~ 현재 : 성결대학교 멀티미디어학부 교수
<관심분야> : 컴퓨터 시각, 영상 처리, 감성 공학

진 성 아(Seong-Ah Chin) 정회원

1991년 : 전북대학교 수학과(이학학사)
1993년 : 전북대학교 전산학(이학석사)
1995년 : (미) City University of New York 대학원 수료
1995년 ~ 1999년 : (미) Stevens Institute of Technology (Computer Science , Ph. D)
1997년 ~ 1999년 : Teaching Assistant, SIT
2000년 ~ 2001년 : 서강대학교 영상대학원 미디어공학과 연구교수
2001년 ~ 현재 : 성결대학교 멀티미디어학부 교수
<관심분야> : Virtual Reality, Computer Vision, Document Analysis