International Journal of Intelligent Information and Management Science
ISSN: 2307-0692 *Volume 4, Issue 3, June 2015*

Detecting Text in Image with Distance Transform and Directional Information

Xiaolei WANG
School of Management Huaibei Normal University, Huaibei, Anhui, 235000, CHINA

Abstract: A new method to detect directly text strokes in image are proposed in this paper. Firstly the edges in image are extracted by Minimum-Spanning-Tree-like(MST-like)algorithm, the distance and directional information from the distance transform combined with low gradient amplitude threshold are applied to determine a candidate strokes region. Finally extraction of real strokes are finished by the hierarchical clustering algorithms and some other relevant features of text. Experiments showed that the method of this paper can be improved by the recall rate and precision rate of text detection, for handwritten Chinese text in the image this method also has a good effect.

Keywords: Text Detection; Distance Transform; Directional Information; MST-like; Hierarchical Clustering

1. Introduction

Character detection and localization in many applications in computer vision has an important position, such as robot navigation, video retrieval intelligent transportation and other aspects. Many scholars have done a lot of work [1] [2]. In combination of these works, text detection is generally divided into four categories: Regional connectivity method in consistent use of text color [3-4]; texture method detecting with the application of textural features by arrangement of multiple text [5-6]; detection method taking advantages of edges based on more text edges [7-8]; and corner detection method using text corners[10]. Combining these methods, one of text features is mainly applied to locate the text for these feature clusters are easier to be connected into region, for example, edge feature is applied due to intensive text edges, while corner feature is applied due to intensive corners in the text. However, in practice, these features have more prominent reactions only in small texts, and region is difficult to be formed in a large text. In recent years, text extraction with the application of stroke width information has obtained good results [1, 9, 11] mainly because of wide range of application of the consistency of the stroke width in the text and key position of stroke extraction in text localization.

In this paper, the distance and directional information from the distance transform combined with gradient amplitude are applied to extract the text strokes so as to effectively avoid the false strokes formed by character stroke edge and non-character stroke edge in stroke width method [9] [12]. Different from other methods, what are directly detected by the methods in this paper are text strokes rather than text region, which provides better data for optical character recognition in the next step.

2. Introduction to Method Used

Character strokes in natural scene image are generally with similar thickness, even handwritten Chinese characters, their strokes are also mostly similar. That is, the shortest distance from the point of center line of a stroke to both edges of it is almost changeless and the vertical lines from the point to both edges are roughly in the opposite directions, as shown in Fig.1. At the pixel point of discrete image, it is expressed that the distance from the two adjacent pixel points in the middlemost of the stroke to their respective nearest edge point are almost the same and approximately opposite in direction.

In addition, all strokes are with relatively consistent width and color and a string of texts in an image are generally with same color and relatively concentrated stroke distribution.

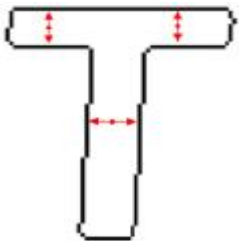


Figure 1. Diagram of the Distance from Intermediate points of a Stroke to the Edges and their Directions

Based on these characteristics, in this paper, edges in the image are first extracted, and then the points in the middle of the stroke with equal distance and in approximately opposite directions are determined by changing the distance. Then such points are used as the seed, together with the gradient information of the pixel point, to search the regions with similar color and then each region may

be identified as a stroke or not according to stroke width, color information and inter-stroke distance. In this paper, the text region is not given, because if the stroke is identified, the text will be also identified. It shall be noted that the stroke in this paper is not a stroke in the strict sense, it may be joined by a few strokes in the usual senses. For example, "T" in Fig. 1 is two strokes in usual sense, however, in this paper, they shall be only counted as one stroke because they are connected.

According to the above idea, first extract the edges of original images, change the distance of the edge image, find the adjacent points with similar distances and in substantially opposite directions, find the possible strokes according to such points and finally screen such strokes in order to identify the true strokes. The general process is as shown in Fig. 2.

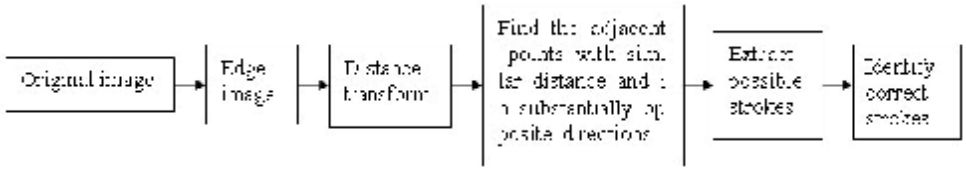


Figure 2. Process Diagram

3. Image Edge Extracting

Many algorithms may be used to extract the edge, based on text detection target, the algorithm used shall be able to avoid the loss of the edge of the stroke, while minimizing non-stroke edge piecemeal, for example, the famous Canny algorithm may form a lot of but not important edge piecemeal, which can not only reduce the difficulty of subsequent exploration, but also reduce the work of computation. An algorithm called MST-like (Minimum-Spanning-Tree-like) in application literature [13] in the paper is used to extract edges. It is an improved algorithm based on the fingerprint image segmentation method of minimum spanning tree [13], the time complexity is O (n), n is the number of pixels of the image, which can be realized rapidly. Because the minimum spanning tree can be used in image segmentation to make understanding similar to the Gestalt law [14], namely, the image segmented is more similar to people's understanding about the grouping of object items, which may result in less edge piecemeal. Because the text is relatively specific object item and this algorithm is fast, it is suitable for extracting the edge in this paper.

In MST-like algorithm, a pixel point is taken as a vertex of an image, which connects with two pixels respectively in its right neighborhood and lower neighborhood and form an edge, edge weight is the absolute value of the difference of pixel gray-values between two vertexes of the edge, then the edge larger than the set threshold value shall be removed in order to make segmentation, eliminate the generation process of minimal spanning tree and reduce calculation work, as shown in Fig.3. See literature [13] for implementation details.

This method can extract most edges of the text, because the MST-like algorithm segments the object depending on the horizontal or vertical gray value of a pixel, not considering the changes in gradient magnitude, some edges featuring smooth transition both horizontally and vertically are more likely to be lost, and the relevant solution will be described later in this paper. In short, text edges in good conditions can be gained with the minimum spanning tree method can be used to gain the text edges in good condition and can also effectively reduce influence on the edges.

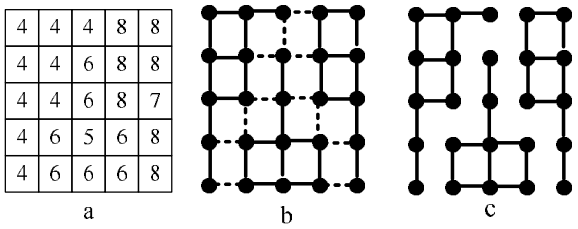


Figure 3. Diagram of Principle of Rapid Minimum Spanning Tree: (a) Original image (b) Image structured with original images, dotted line for the edge larger than a certain threshold value (c) Segmentation result after removing the edge whose weight is larger than the threshold value

4. Stroke Extracting with Application Distance transform

The image distance is the distance at which each pixel point value in binary image converted to its nearest (the distance here can be defined as per required), it is clear that the distance of non-zero pixel point after conversion must be 0. Distance transform is generally used to judge the similarity of shapes [15][16], however, most literatures only take the distance information of pixel point and neglect the directional information with the distance transform. In this paper, distance transform and directional information are used to detect the points in the

middle part of the stroke that have almost same distance to both edges and are opposite in direction.

Obviously, if the value of edge pixel point is defined as 1, adjacent pixel points at the stroke edge will have similar distance and their directions to the nearest non-zero point are substantially the same, only the adjacent pixel points in the middle part of the stroke may have similar distance and substantially opposite direction.

$s(x_1)$, $d(x_1)$ and $s(x_2)$, $d(x_2)$ are respectively used to stand for the Euclidean distance and direction vector of pixel point x_1 and x_2 to their nearest non-zero pixel point, if Formula (1) and (2) are established, pixel point x_1 and x_2 may be at the middle part of the stroke.

$$|s(x_1)-s(x_2)|\leq 1 \quad (1)$$

$$\arccos\left(\frac{<d(x_1),d(x_2)>}{\|d(x_1)\|\cdot\|d(x_2)\|}\right)>e \quad (2)$$

Where: $<\cdot,\cdot>$ stands for inner product of two vectors, $\|\cdot\|$ stands for vector length, $\arccos(\cdot)$ is arc-cosine function.

In this paper, two adjacent points conforming to Formula (1) and (2) are called intermediate points. In the formula (1), in consideration of discrete image data, if both edges exist and are substantially parallel to each other, their distances to the intermediate point generally differ by 1 or are equal. If a stroke is relatively thin, because of discrete data of the image or edge extraction, the intermediate points in the stroke may not be strictly opposite to each other, but are generally substantially opposite, thereby, Formula (2) is set with a threshold value e. It is worth noting that not all points meeting the requirements of Formula (1) and (2) in distance transform image are the intermediate point of a stroke, for example, the red point at white arrow as shown in 3(b).

Depending on the distance transform and directional information provided, it can be drawn that if the edges are approximately parallel to each other, many intermediate points may exist in the middle part; the distance of each intermediate point will not change significantly and the distance of the intermediate point is generally the maximum within the stroke region. For non-stroke region, because it is easy to remove some non-protruding edges with MST-like, some large connected region may be formed in edge image; in this region, though there are many intermediate points, for the whole connected region, the distance of each intermediate point change greatly, however the distance of pixel point rather than the intermediate point may be greatly larger that of the intermediate point, as shown in Fig.3(b). The feature provides an important basis for classification of strokes in next step.

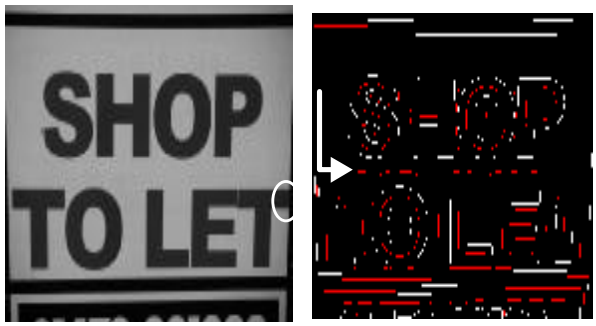
4. Identification of Stroke Region

Based on the analysis in Part 3, the intermediate points in the stroke region are with similar distance and in approximately opposite direction; in this paper, it is made to first find the possible stroke region and then screen such regions.

4.1. Identification of Possible Stroke Regions

The ultimate goal of this paper is to find the stroke, as long as some symmetrical edges of a stroke are extracted, it is easy to find the intermediate point of this stroke.

According to the definition of the distance transform, the value of Point 0 in the stroke is surely larger than 0 after distance transform and the value of the edge point is also surely 0, thereby, if a intermediate point is taken as the seed of regional growth and its distance value 0 is taken as the growth termination condition, for a full-edge stroke, in theory, all the pixels of this stroke may be found.



(a) (b)

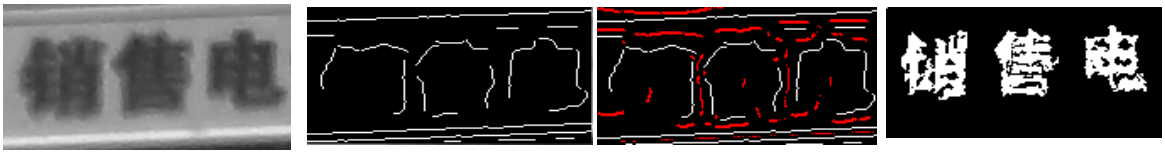
Figure 4. Intermediate points Found Based on Distance and Direction (a) Original image (b) Intermediate points (red), white arrow for the intermediate points not formed with strokes

However, due to lack of MST-like algorithm, it is easy to lose some edges with gentle transition, such as white circles as shown in Fig. 3 (a), in view of the larger gradient magnitude at the stroke edge, gradient information of pixel point shall be added during growing in the region so as to control continuous growth of the region in some broken edges and effectively extract the stroke region.

When the amplitude of the gradient is greater than a threshold value T, the region will stop growing. Considering low contrast ratio of some text edges, the threshold value can be set relatively small and it is set to be 10 in this paper. This method may cause some regions in the stroke with uneven color stop growing at certain pixels, but it does not affect the formation of the connected part of the stroke, because the growth will continue through other pixels in the strokes, so to find most of the pixels in the stroke, while losing some piecemeal formed in the stroke, and such piecemeal can be simply merged and not affect the integrity of the stroke. Though this strategy is simple, one of its important features is to extract the entire text with this strategy for some texts with low contract rate

many of whose edges may be lost with MST-like algorithm, which is mainly due to low threshold value of the gradient amplitude, the region will stop growing at the text edges, as shown in Fig. 4.

On the other hand, the threshold value T is set low and will not increase the number of regions significantly, because if a non-text region has edges with low values, the region will stop growing and will not increase the number of regions, and it is also the reason for taking the intermediate point as the seed. In theory, the number of the segmented regions will not exceed the number of the intermediate points to the most, so many pixel points in the image, due to the gradient amplitude which may stop growing and are not counted in the region growth, will not be affected by the low threshold T value. However, because the intermediate points are widespread in the stroke region, non-strokes region is not included in the region growth and will not affect the extraction of the stroke region.



(a) (b) (c) (d)

Figure 5. (a) Character with smooth transition (b)Edges extracted with FMST (c) Edges and intermediate points detected (d) Stroke region with gradient magnitude value to control growth

4.2. Strokes Regional Screening

After region growth with the above method, many non-stroke regions are included and shall be screen. Based on the characteristics used to classify the stroke in this paper: (1) The change in the distance on the intermediate point of the stroke is relatively small while the change in distance on the non-stroke intermediate point is relatively large, and the distance of other Point 0 may be greatly larger that of the intermediate point. (2) Colors of the strokes of the text are also the same; (3) the widths within a same stroke region will not differ greatly each other. Stroke regions may be screened with such characteristics with the following steps:

Step 1: Hierarchical clustering algorithm: take *Average-linkage* distance of average gray value of the region as the similarity, cluster the extracted regions, if the difference between the average gray value of different classes is greater than the threshold value T1, stop class pooling.

Step 2: For each class, if the shortest distance from a member to the other members of the class is two times the largest intermediate point in these two classes, cancel the member.

Step 3: If a member of a class completely include an isolated member of the other class, cancel the isolated member.

In Step 1, hierarchical clustering algorithm is used mainly because it can stop classification according to the similarity of average gray values so as to automatically determine the number of clusters. In Step 2, the minimum spatial distance refers to the minimum distance between the edges of two classes and the reason for determining the datum as twice is because the minimum distance between the strokes of a character will generally be not more than [9], On the other hand, due to the maximum intermediate point distance used, for some texts with small contrast and stop growing with gradient magnitude, one character region usually has only one connected region, through this data, it is easy to avoid deleting this character, as the examples shown in Figure 4. In Step 3, the background of some texts is well regulated, a stroke region will be generally formed, but these pseudo-stroke regions formed in the background are completely included in a real text region, as shown in Fig.3, annular regions in “O” and “P”.

5. Obtaining Experiment Parameters

The selection of experiment threshold is very important for correctly and effectively extracting character stroke. Important thresholds used in this paper are obtained from experiment.

(1) Parameter e for angular separation of intermediate point. This parameter threshold is mainly used to distinguish whether the direction of intermediate point is opposite. Since this parameter is mainly used for finding out the area of a stroke, according to the practice of extracting stroke area with intermediate point as seed, in one stroke, the stroke cannot be lost as long as you can select one point. Therefore, the principle of selecting this parameter is that one stroke shall have the intermediate point to the greatest extent, but the internal intermediate point shall be as less as possible. The experiment randomly selects 50 pictures. It can be seen that there are a total of 872 strokes after extracting edge. Parameter e ranges from 140° to 170° at intervals of 10°. The stroke having intermediate point is correct. With interior points of stroke as less as possible is regarded as the evaluation index. The experimental result is as shown in Table 1.

Table 1. Experimental data for selecting parameter e

Type \ Degree	1400	1500	1600	1700
Number of strokes with Intermediate points	872	871	803	542
Number of Interme-diate points	16332	10985	5632	3128

Therefore, $e = 150^{\circ}$ is selected in the experiment. The experiment also finds out that, when e increases, it is in line with two sides of stroke edge, it has little influence on the extraction of intermediate point of stroke. The edge presenting arc will have a large influence. And the thinner the stroke is, the less rapid for extracting the intermediate point will be. This is mainly due to that the direction of point dividing the unparallel edge from the middle will easily generate big deviation. However, for the edge of thin strokes, when calculating the angle, since the distance is little, the generated deviation will be big.

(2) Threshold T_1 in hierarchical clustering. This classification threshold is mainly based on artificial judgment. The principle is that the text area with the same color shall be classified into a class as much as possible. 30 figures are randomly selected. There are 92 text areas are judged as in the same type. The threshold uses from 10 to 30, respectively, with the interval of 5. The accuracy is shown in Table 2.

Table 1. Experimental data for selecting threshold T_1

Threshold T1	10	15	20	25	30
Number of correct time	61	78	85	73	42
Accuracy (%)	66.3	84.9	92.4	79.3	56.5

According to the above-mentioned result, in the experiment $T_1 = 20$ is selected. Big threshold will easily tend to merge different types into the same type, which will cause difficult to subsequent analysis. When the threshold takes 20, the main error is that the same type is divided into different types. This mainly is that the same type of character is mistakenly classified due to uneven light. But such classification mistake does not have a big influence on the final stroke extraction, because although the same type of stroke is not classified into the same type, it participates in the subsequent stroke screening process.

6. Experimental Results and Analysis

In order to verify the effectiveness of methods in this paper, database provided by ICDAR 2005[17] and pictures taken are taken as the testing data. Taking precision rate and recall rate as the evaluation indicators, text detection is defined as:

Precision rate = number of detected texts/ number of all detected texts

Recall rate= number of retrieved texts / number of texts that should be detected

The experimental results are as shown in Table 3.

Table 3. Experimental Results

Method	Precision rate	Recall rate
Methods in this paper	0.79	0.83
Literature [9]	0.73	0.60

Tests show that compared with algorithm in literature [9] based on stroke width, the recall rate of algorithm in this paper is higher, mainly because that literature [9] looks for the stroke width by searching for another edge of the stroke in the vertical direction of the stroke edge. For stroke edge with low text contrast, or with more series noise in the middle of the stroke, very inconsistent widths are often formed, which produces a more serious negative impact on the analyzed stroke. With the methods in this paper, the stroke can be obtained with fewer edges via the intermediate point and gradient magnitude having lower threshold, less affected by noise of stroke's intermediate point. This is mainly because it is rare when intermediate point occurs and the distance of intermediate point is larger than other point in the region, while the text performance is more prominent. As is shown in Fig. 5(b), extraction of edge by canny algorithm used in literature [9] will loss all these characters. By the methods in this paper, edges will be much fewer, which shows that MST-like can effectively reduce the edge. Note that some edges in Fig. 5(c) are incomplete, for example upper margin of character "P", but a correct result can be obtained finally. It is mainly because low gradient amplitude thresholds are added as the condition for stopping regional growth, also compensating for deficiency of edge extraction. There a point to be explained here that result in Fig. 5(d) is produced after filling in a small black hole on the extracted stroke region, and these black holes are generated when region grows due to low threshold T1.

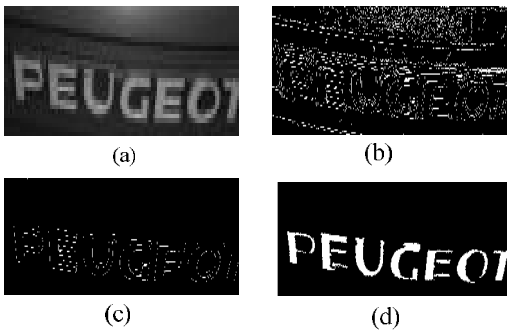


Figure 6. (a) Original figure (b) Edge extracted by literature [9] (c) Edges extracted in this paper (d) Final results of methods in this paper

To test the effectiveness of methods in this paper for detecting handwritten Chinese characters, some self-shot pictures are used to conduct the experiment. The results are good. Fig. 6 shows the effect picture of an example, which shows that this method is also applicable for extracting text with handwritten Chinese characters.

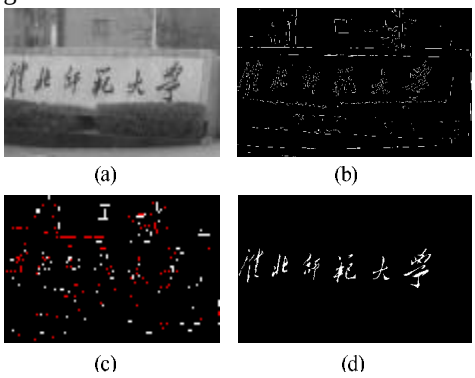


Figure 7. Example for identification of text with handwritten Chinese characters
(a) Original figure (b) Edge image (c) Intermediate point image (d) Final result

7. Conclusion

In this paper, on the basis of edge extraction by MST-like algorithm, the application of distance and directional information from the distance transform combined with low gradient amplitude threshold to extract text stroke improves the effect of stroke extraction. The results showed that application of the minimum spanning tree edge extraction combined with low gradient amplitude threshold in the text extraction process can supplement each other. On the one hand, MST-like algorithm can reduce the edge interference, on the other hand, the use of low gradient magnitude will not lose too many strokes in the region growing process. Different from other methods, methods in this paper will not locate text region but obtain the text strokes directly in which way to better use the optical character recognition system.

References

[1] Hong gang Zhang, Kaili Zhao. Text extraction from natural scene image: A survey [J], Neurocomputing, Vol.122(12). 2013: 310–323

[2] Yuan Haidong, Ma Huadong, Huang Xiaodong, Video Text Detection and Localization Based on Gradient and Roughness [J]. Chinese Journal of Electronics. Vol. 36 No. 8.2008:1660-1664

[3] P.Shivakumara,T.Q.Phan,C.L.Tan, Agradient difference based technique for video text detection, 10th International Conferenceon Document Analysis and Recognition, ICDAR'09, pp.156–160.

[4] Q.Ye, J.Jiao, J.Huang, H.Yu, Text detection and restoration in natural scene images, Journal of Visual Communication and Image Representation .18(6) (2007)504–513.

[5] Ye Q X,Huang Q M,Gao W,et al. Fast and robust text detection in images and video frames[J]. Image and VisionComputing, 2005, 23(6) :565-576.

[6] Tebourbi R, Belhadj Z. A texture based multispectral images indexing[C]//12th IEEE International Conference on Eletronics, Circuits and Systems,Gammarth,2005:1-4.

[7] P.Shivakumara,W.Huang,C.L.Tan, Anefficient edge based technique fo rtext detection invideoframes, in:2008 Eighth IAPR International Workshopon Document AnalysisSystems, DAS'08, IEEE,2008, pp.307–314.

[8] X. Liu,J.Samarabandu, Anedge-based text region extraction algorithm for indoor mobile robot navigation, in:2005 IEEE International Conferenceon Mechatronics and Automation,vol.2, IEEE, 2005, pp.701–706.

[9] Boris Epshtein Eyal Ofek. Detecting Text in Natural Scenes with Stroke Width Transform. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, p2963-2970, 2010.

[10] Xu Zhao, Kai-Hsiang Lin Text From Corners: A Novel Approach to Detect Text and Caption in Videos. IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL.20, NO.3, MARCH 2011 :790-799.

[11] Jerod J. Weinman. Toward Integrated Scene Text Reading. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 36, NO. 2, FEBRUARY 2014,p375-387

[12] Dinh V, Chun S,Cha S,et al.An efficient method for text detection in video based on stroke width similarity[C]//ACCV 2007,2007

[13] Cha B,Suetake N,Kawano H,et al.Fast minimum-spanningtree-like image segmentation[C] //Proceedings of the 4th International Conference on Natural Computation, 2008:152-156

[14] C.T.Zahn. "Graph Theoretical Methods for Detecting and Describing Gestalt Clusters". IEEE Trnas. on Computers, 20(1): 1997: 68–86.

[15] JIANG Ruyi, Liu Hongtao, Hu Wen, et al., Application of Improved Image Distance Transform Method in Lane Detection, Journal of Shanghai Jiaotong University, 2012, 23(9):1450-1454.

[16] Y. Lee and K. Grauman. Shape Discovery from Unlabeled Image Collections. IEEE Conf. on Computer Vision and Pattern Recognition, Miami, FL,USA, June, 2009: 2254-2261.

[17] ICDAR 2005 text locating competition results, Eighth International Conference on Document Analysis and Recognition, 2005. Proceedings. pp 80-84(1).