A New Text Location Approach Based Wavelet

Weihua Li Zhen Fang Shuozhong Wang
(上海交通大学通信与信息工程学院 上海200072)

Abstract With the advancement of content-based retrieval technology, the importance of semantics for text information contained in images attracts many researchers. An algorithm which will automatically locate the textual regions in the input image will facilitate the retrieving task, and the optical character recognizer can then be applied to only those regions of the image which contain text. In this paper a new text location method based wavelet is described, which can be used to locate textual regions from complex image and video frame. Experimental results show that the textual regions in image can be located effectively and quickly.

Keywords Text location, 2-D wavelet, Morphological operation

I. Introduction

As more and more digital libraries coming out, the multimedia information and Internet technology develop very quickly, and text extraction from complex images attracts many researchers' interests. They want to apply it to web search and many other image databases retrieving systems. By far, the studies on content-based image retrieving have mainly focused on low-level features, such as color, shape and texture. The combination of low-level content-based retrieval and keywords description in manual has also been studied, but the manual description is not always reasonable and correct. It is very difficult to approach the semantic-level content-based retrieving. The textual information contained in images gives important clue to the search engine, and makes the semantic-level content-based retrieval possible.

Many efforts have been done for text location in images and videos. Fletcher, et al. [1] presented an algorithm based connected component analysis to separate text from graphics. Jain, et al. [2] detected text regions in document images by a segmentation method based on a multi-channel filtering approach to texture segmentation, in which the text is treated as a textured region. Tan, et al. [3] described a pyramid structure and connected component analysis to extract text from map images. Wu, et al. [4] proposed a text-finder system in which the texture segmentation was used to detect the text in images. Lopresti, et al. [5] combined the color clustering and connected components analysis to extract text from web images. Lienhart, et al. [6] concentrated how to detect text added to the video artificially. Sato, et al. [7] introduced a recognition-based segmentation method.

The text detection approaches can be classified into two categories. The first category is connected component based methods, which can locate text quickly but have difficulties when text is embedded in complex background or touches other graphical objects. The second category is texture classification based. However, it is hard to find accurate boundaries of text areas and false detections often exist in text-like texture areas. One of the difficulties to be faced in detecting and extracting text contained in images is the different use of character size and font combined with complex backgrounds (see Figure 1).

Our proposed text detection scheme is based on the 2-D wavelet. The text regions have obvious high frequencies in different directions. The decompositions of 2-D wavelet transform show the directional details of input image. We can find the candidate text lines by emerging obvious components along different directions. Here some techniques are used to reduce the non-text lines. Then morphological operations are applied to connect the text lines into textual regions. At last we locate the text regions with bounding boxes.

The rest of this paper is organized as follows. Section II introduces the text location method, including the text detection algorithm based wavelet, some techniques and morphological operations. Section III presents the experimental results and evaluations of the proposed approach.

II. Text location

1. Text detection based wavelet

The 2-D wavelet transform, as a "mathematic microscope", has the advantage of amplifying and polarization. If $f(x,y)$ is a two-dimensional signal, its 2-D wavelet transform can be defined as:

$$WT_f(a;b_x,b_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\Psi_{a,b_x,b_y}(x,y)dxdy \quad (1)$$

where $a>0$, b_x and b_y are real numbers that specify translation in the two directions, $\Psi(x,y)$ is a two-dimensional basic wavelet, and

$$\Psi_{a,b_x,b_y}(x,y)=\frac{1}{|a|}\Psi(\frac{x-b_x}{a},\frac{y-b_y}{a}) \quad (2)$$

This basic wavelet is scaled by a, and the scale, a, can be written as matrix form:

$$A=ar_\theta \quad (3)$$

where r_θ is the rotation factor:

$$r_\theta=\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \quad (4)$$

Then the matrix form of formula (2) and formula (1) are described as follow:

$$\Psi_{A,\bar{b}}(\bar{t})=\frac{1}{|A|}\Psi[A^{-1}(\bar{t}-\bar{b})] \quad (5)$$

$$WT_f(A,\bar{b}) = \langle f(\bar{t}),\Psi_{A,\bar{b}}(\bar{t})\rangle = \frac{1}{a}\int_{R^2} f(\bar{t}) \Psi(r_\theta^{-1}\frac{\bar{t}-\bar{b}}{a})d\bar{t} \quad (6)$$

where $a=\det A$, $t=[x, y]^T$, $b=[b_x, b_y]^T$. Therefore, we can use the 2-D wavelet transform to detect the edges of image along the direction θ.

The text lines contained in image can be detected in different directions. In this paper we detect the horizontal, diagonal and vertical lines from image using 2-D wavelet transform with $\theta=0°,45°,90°$, and the haar basis is used. The experimental results show that all of the textual lines are detected and some other non-textual lines are also detected as text. Some techniques are used to select the correct regions from the candidates in the following part.

2. Selection of text regions

In order to eliminate some obvious non-text pixels in the three outputs from the wavelet transform, another three matrices are constructed. Since the directional frequencies of text regions are high, the largest K coefficients in magnitude in the every output are replaced with 1 and all other coefficients are replaced with 0. Then the three new matrices are emerged into one matrix to select the pixels on all three directions.

We define a rectangular window, which is used to scan over the matrix to remove isolated points by setting a threshold for the minimum number of non-zero points in such a moving window. The window size should be chosen in accordance with the characters to be located. In this paper the window size we chose is 10×20.

To locate the text regions in image, the pixels remaining in the matrix should be grouped. Here the morphological operation is used to make the located targets connected. Morphological operation is a nonlinear filtering method, which can resolve many problems in image processing, such as noise restriction, feature extraction, edge detection and image segmentation. The basic morphological operations are dilation and erosion, defined as:

Erosion: $X=E\otimes B=\{(x,y):B(x,y)\subset E\}$ (7)

Dilation: $Y=E\oplus B=\{(x,y):B(x,y)\cap E\neq\varnothing\}$ (8)

Erosion operation is a process of cleaning up the boundary pixels, so it can remove the noise and make objects smaller. Dilation operation is a process of adding pixels to boundary, so it can make small holes in object disappear and make some closed objects connected. The erosion and dilation operations can be combined into two operations, closing and opening operation. The process of erosion followed by dilation is called opening. And the process of dilation followed by erosion is called closing. After applying a set of dilation and erosion operations on the matrix, we get the satisfied result that the text pixels are connected into textual regions, see Figure 2.

Then we can calculate the bounding boxes according every connected textual region in image, and locate the correct textual areas, as depicted in Figure 3.

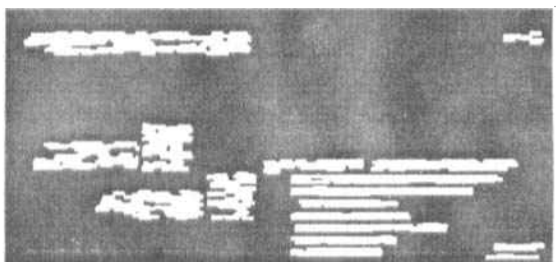
III. Experimental results

In the experiments, we evaluate our method using advertising images, medical images and video frames. Nearly all the textual regions are located in the end with our method. The isolated characters in medical images, which cannot be located using other methods, are also located with our proposed method. The evaluation result of the proposed text location approach is listed in Table 1.

Table 1 the evaluation result of our text location approach

Images	Characters	Detected characters	False alarms
21	1147	1105 (96.34%)	12 (1.05%)

Figrue1 An advertising image scanned from Newspapen with size 373×795 pixels



Figrue2 The grouping result using morphological operations

Figrue3 The final located regions with our method

(下转第114页)

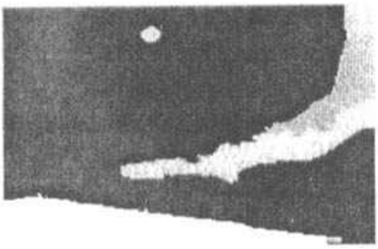


图12 Moscheni,Bhattacharjee[14]

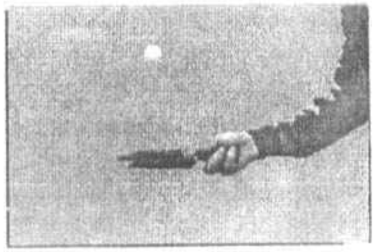


图13 本文的分割结果

结论 边缘在视频序列的运动分割中有着重要的意义,如果图像中的边缘都是可见的,那么只要简单地提取图像的边缘就可以得到很精确的分割结果。本文阐述了如何利用两帧图像之间的运动跟踪获取边缘的运动标记,从而确定区域的运动层次和区域标记的过程。运动跟踪,初始化运动和区域增长等方面的改进使得在两帧之间对边缘进行了很好的运动标记,绝大多数的样点都能匹配成功,区域增长的结果也与实际情形很一致,从而最终得到了比较令人满意的分割效果。

参考文献

1 Dempster A P,Laird H M,Rubin D B. Maximum likelihood from incomplete data via the EM algorithm. Journal of Royal Statistical Society Series B,1977,39:1~38

2 Moscheni F. Dufaux F. Region merging based on robust statistical testing. In:SPIE Proc. Visual Communications and Image Processing, Orlando, Florida, USA, March 1996

3 徐静波. MPEG-4标准与多媒体通讯. 信息工程学院学报,1999,8(4)

4 Smith P, Drummond T,Cipolla R. Edge tracking for motion segmentation and depth ordering. Department of Engineering University of Cambridge

5 Smith P A. Edge-based Motion Segmentation. Jesus College Aug. 2001

6 Hsu S, Anandan P. Peleg S. Accurate computation of optical flow by using layered motion representations In:Intl. Conf. on Pattern Recognition,Oct. 1994

7 Ayer S,Sawhney H S. Layered representation of motion video using robust maximum-likelihood estimation of mixture models and MDL encoding. Intl. Conf. on Computer Vision, Cambridge, June 1995

8 Sinclair D. Voronoi seeded colour image segmentation. AT&T Laboratories Cambridge, submitted for publication, 1999

9 Mech R, Wollborn M. Automatic segmentation of moving objects. MPEG98/m3187, 1998

10 Weiss Y, Adelson E H. A unified mixture framework for motion segmentation; Incorporating spatial coherence and estimating the number of models. IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, June 1996

11 季白杨,陈纯,钱英. 视频分割技术的发展. 计算机研究与发展, 2001,38(1)

12 Canny J. A computational approach to edge detection. IEEE Trans on Pattern Analysis and Machine Intelligence,1986

13 Dufaux F,Moscheni F,Lippman A. Spatio-temporal segmentation based on motion and static segmentation. In: Proc. Intl. Conf. on Image Processing, Washington DC, USA, Oct. 1995

14 Moscheni F,Bhattacharjee S. Robust region merging for spatio-temporal segmentation. In: Proc. Intl. Conf. on Image Processing Lausanne, Switzerland, Sept. 1996

(上接第106页)

IV. Discussion

This paper describes a new text location approach based wavelet. The experimental results show that our method can locate and extract text from complicated background accurately and effectively, and is not heavily dependent upon the image contrast, font type and character size. And the computation in this method has a high speed. However, there is a problem in this method that the threshold used in experiment is set by manual, which will be solved in our future studies.

Reference

1 Fletcher L A, Kasturi R. A Robust algorithm for text string separation from mixed text/graphics images. IEEE transaction on Pattern Analysis and Machine Intelligence, 1988,10(6):910~918

2 Jain A K,Bhattacharjee S. Text Segmentation Using Gabor Filters for Automatic Document Processing. Machine Vision and Applications,1992,5:169~184

3 Tan C L, NG P O. Text extraction using pyramid. Pattern Recognition,1998,31(1):63~72

4 Wu V,Manmatha R, Riseman E M. *Textfinder: An automatic* system to detect and recognize the text in images. IEEE transaction on Pattern Analysis and Machine Intelligence,1999,12(11):1224~1229

5 Loprestie D,Zhou J. Locating and Recognizing Text in WWW Images. Information Retrieval,(Kluwer Academic Publishers.), 2000,2:177~206

6 Lienhart R,Effelsberg W. Automatic text segmentation and text recognition for video indexing. ACM/Springer Multimedia System, 2000,8:69~81

7 Sato T, Kanade T, Hugesd E K, Smith M A, Satoh S. Video OCR: Indexing digital news libraries by recognition of superimposed caption. ACM multimedia Systems, 1999,7(5):385~395