A video text extraction method for character recognition

CONFERENCE PAPER · OCTOBER 1999

DOI: 10.1109/ICDAR.1999.791716 · Source: IEEE Xplore

1 AUTHOR:

Osamu Hori

Toshiba Corporation

30 PUBLICATIONS **481** CITATIONS

SEE PROFILE

A video text extraction method for character recognition

Osamu Hori
Toshiba Corporation R&D Center
Multimedia Laboratory
1, Komukai Toshiba-cho, Saiwai-ku, Kawasaki, 210-8582 Japan
osamu.hori@toshiba.co.jp

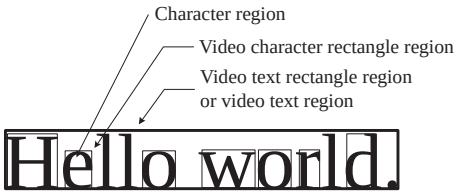

Figure 1. Explanation figure for technical terms.

Abstract

This paper presents a method to precisely extract only video character portions from a video text rectangle region in order to make a readable image for OCR. In the conventional methods, gray image binarization processing with a given threshold is employed to extract high-intensity video character regions. A video has a complex background with various kinds of intensity so that appropriate thresholds are not always obtained. The proposed method extracts reliable high-intensity regions in a video text region and then expands them in order to make the whole video character regions. The experiments show this new method to be superior to the conventional methods.

1. Introduction

A video text, so-called a caption superimposed in a video, has lately attracted considerable attention[1, 2, 3] because it contains meaningful information about video contents and is of great help as a keyword. This paper presents a method to precisely extract video character regions from a video text rectangle region in order to make a readable image for OCR. At the outset, we specify some important technical terms as shown in Figure 1. We use both "video text region" and "video text rectangle region" to indicate a bounding box rectangle region including characters and the background. "Character region" means only pixels making characters.

2. Video character segmentation by the proposed method

2.1. The abstract of the entire processing

Figure 2 describes the flow of video character segmentation from the background. The intensity distribution of characters in a video is precisely estimated to extract only character regions from the background. The first process is to eliminate the background and roughly estimate the character regions by image processing. In this process, a Sobel filter for edge detection is applied to the video text region to emphasize edges located on character contours. The edges are extracted and dilated to roughly estimate character regions. The dilated edge area mainly consists of character regions, contours, and a part of the background(See Figure 3). An intensity histogram is calculated from the region and segmented into three parts by the Otsu automatic threshold selection method[4]. The Otsu method is well known as an automatic binarization threshold selection algorithm using an intensity histogram. This method can be easily extended to a multiple segmentation method. The segmented high-intensity part of the histogram presumably comes from character regions. An intensity distribution of character regions is computed from the part. This result, however, is not precise because this roughly estimated high-intensity part includes outliers coming from the background or contours. A robust estimation is introduced to increase the accuracy of the distribution estimation. Video characters have high-intensity and are surrounded by low intensity contours in general. At first, very high-intensity parts are detected in the character regions based on the estimated average and then very high-intensity parts are expanded to the contours based on the estimated variance.

2.2. Rough estimation of high-intensity parts of a histogram

Figure 4 describes intensity histogram examples of the entire video text region and the character regions. The bro-

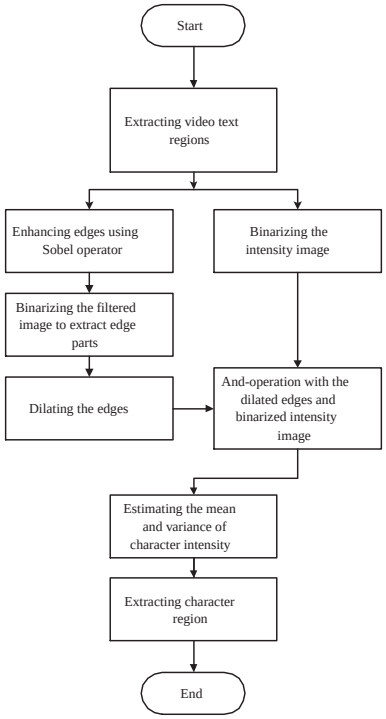

Figure 2. The flow of the proposed method.

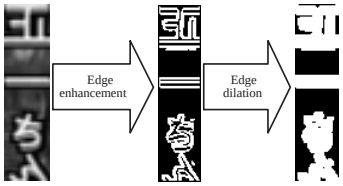

Figure 3. Rough estimation of video text region using edge dilation.

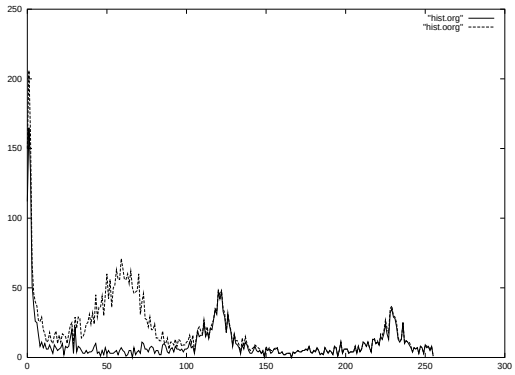

Figure 4. The broken line indicates the intensity histogram of the entire text region. The solid line indicates the intensity histogram of the character region.

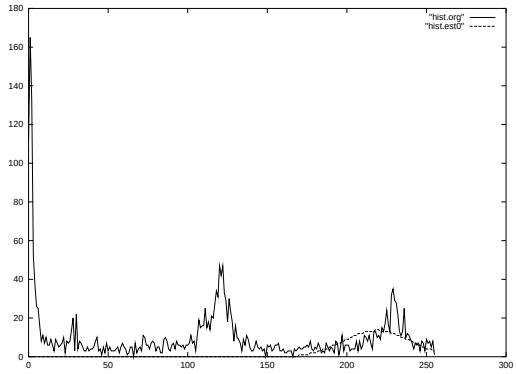

Figure 5. Normal distribution estimation of the high-intensity part.

ken line indicates the intensity histogram of the entire region. On the other hand, the solid line indicates the intensity histogram of the character region with three peaks because almost all the background region is eliminated by the edge dilation process. The high-intensity part with a peak is roughly extracted from the character region histogram using the Otsu method. Two thresholds segmenting the histogram into three parts are determined as 64 and 167 respectively in Figure 5. Three parts mean character, contour, and the background regions. The average and variance of the normal distribution are calculated from the high-intensity part. The broken line indicates the estimated normal distribution in Figure 5. This estimation does not fit the real histogram in the high-intensity part because the data includes irrelevant data in the outskirts of the distribution. Another approach is to use the EM method[5] for estimating three mixed normal distributions. This data, however, includes many outliers and so the EM method is not always a good approach. In the case of estimating one normal distribution in the high-intensity part, robust estimation is workable because other data from the background and contours can be dealt with as outliers and eliminated.

2.3. Intensity estimation using robust estimation

2.3.1 Example of character intensity distribution estimation using M-estimation

A character high-intensity part is segmented by the Otsu method from an intensity histogram. The initial value is calculated using this part. Figure 5 shows the initial value. This value is modified by the M-estimation[6]. The result is shown in Figure 6. The character intensity distribution is precisely estimated without the influence by noise in the outskirts of the histogram peak.

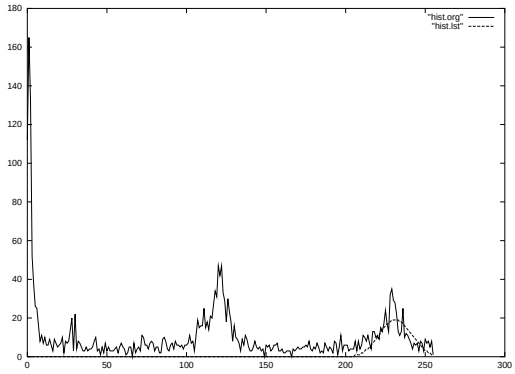

Figure 6. Result of character intensity distribution estimation by M-estimation

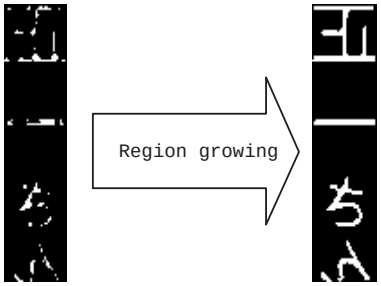

Figure 7. Extraction examples of the proposed method

2.3.2 Character region segmentation using the estimated character intensity distribution

At first, stable high-intensity pixels, which may be parts of characters, are extracted from dilated edge regions. The threshold is $m + \sigma$, where the average m and variance σ are estimated by M-estimation. These high-intensity regions are expanded to the entire character regions. The neighbors of initial high-intensity regions are checked. If the intensity of neighbors is higher than $m - 3\sigma$, the neighbors are concatenated to the initial intensity regions. This operation is repeated until no concatenation occurs. This processing depends on the assumption that only parts of characters in dilated edge regions have higher intensity than $m + \sigma$. The entire character pixels have higher intensity than $m - 3\sigma$ and the contours between the characters and the background have lower intensity than $m - 3\sigma$. All character regions are successfully extracted by this processing if these conditions stand. Figure 7 shows the result of character extraction by the proposed method. 3σ was determined heuristically according to our experiments.

3. Experimental examples

3.1. A simple video text example

Figure 8 shows an example of dealing with a simple video text with the monotonous intensity. This data is processed

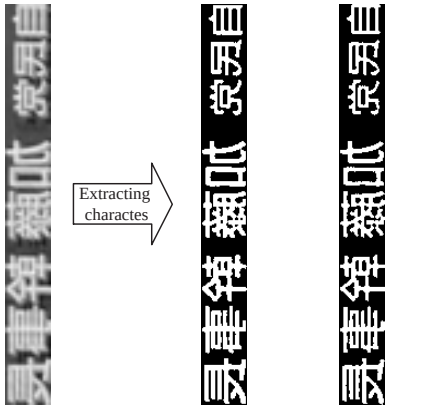

Figure 8. Results of a simple video text processed by two methods.

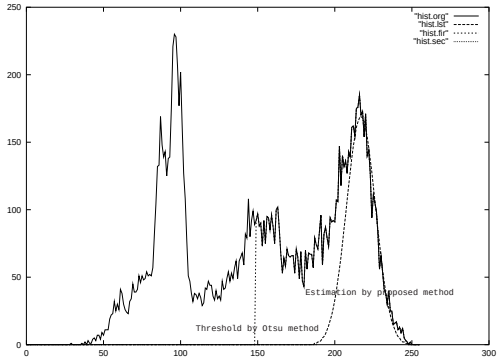

Figure 9. The histogram of the character regions masked by dilated edge regions and the estimated distribution result.

by the Otsu method and the proposed method. The leftmost figure is the original image. The rightmost figure is a processed image by the proposed method. In this case, the Otsu method and the proposed method successfully segment the characters from the background. The original image has two kinds of intensity, but the neighbors of contours have the medium intensity, so that the histogram shows almost three peaks. Figure 9 shows the intensity histogram of the video text regions masked by dilated edge regions. The selected binary threshold by the Otsu method and the estimated character intensity distribution are indicated in the Figure.

3.2. An example of a video text including three kinds of intensity: from characters, contours, and the background.

Figure 10 shows an example of dealing with a video text consisting of three kinds of intensity; from characters, contours, and the background. As shown in Figure 11, the Otsu method fails to segment the characters because this

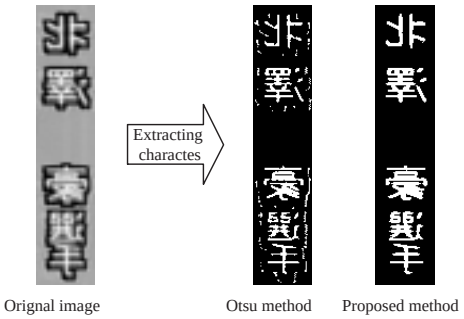

Figure 10. Results of an image with three kinds of intensity processed by two methods.

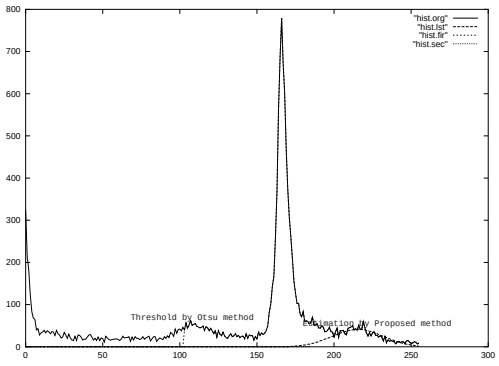

Figure 11. The histogram of the character regions masked by dilated edge regions and the estimated distribution result.

method assumes the histogram of the image has two peaks. The proposed method successfully extracts the character regions.

3.3. Evaluation by OCR

We evaluated our method using two 15-minute news broadcast programs, received and digitized. The video data are encoded by MPEG-2 format at 720 $\times$ 480 resolution. The programs including 440 characters and 615 characters, respectively, were processed by the two methods: the Otsu method and the proposed method. Our own OCR developed for a print document reader was used. Table 1 and 2 show the results of OCR recognition. The results depend on kinds of news programs. Video text images in Sample 1 have histograms with three peaks. On the other hand, video text images in Sample 2 have histograms with two peaks but character image intensity is not high enough. Some results have many errors because our OCR is very sensitive to noise and binarized images. Figure 12 shows an example of failure of character region extraction by the proposed method. In this case, the backgound includes a high-intensity object and characters have low intensity in comparison to the background. But these characters have high-saturation.

Table 1. Sample 1 Results of OCR recognition.

Method	Correct	Rate
Otsu	62	14%
New	420	96%

Table 2. Sample 2 results of OCR recognition.

Method	Correct	Rate
Otsu	264	43%
New	312	51%

4. Conclusion

This paper described a novel method to precisely segment only character regions from the complex background in videos for OCR data entry. Further work is required to extend this method to low intensity character extraction by using saturation information.

References

[1] Y. Ariki and T. Teranishi. Indexing and Classification of TV News Articles Based on Telop Recognition. In *Proceedings of International Conference on Document Analysis and Recognition*, pages 422–427, 1997.

[2] Toshio Sato, Takeo Kanede, Ellen K. Hughes, Michael Smith, and Shin'ichi Satoh. Video OCR for digital news archive. In *Proceedings of International Workshop on Content-Based Access of Image and Video Database*, pages 52–60, 1998.

[3] Jae-Chang Shim, Chitra Dorai, and Ruud Bolle. Automatic Text Extraction from Video for Content-Based Annotation and Retrieval. In *Proceedings of International Conference on Pattern Recognition*, pages 618–620, 1998.

[4] Jun Otsu. A threshold selection method from gray-scale histograms. *IEEE Trans. Syst Man Cybern.*, SMC-9(1):62–66, 1986.

[5] R. A. Redner and H. F. Walker. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Reviews*, 26(2):195–239, 1984.

[6] P. J. Huber. *Robust Statistics*. Wiley, New York, 1981.

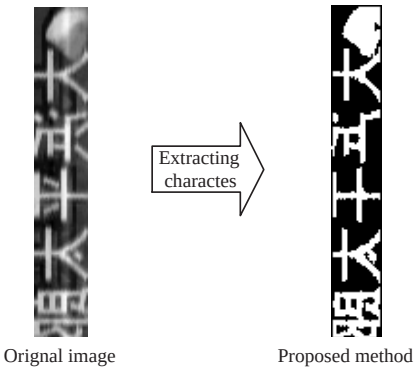

Figure 12. Example of a failure