Telop and Flip Frame Detection and Character Extraction from TV News Articles

Yasuo Ariki
Ryukoku University,
Seta, Otsu-shi, 520-2194, JAPAN
ariki@rins.ryukoku.ac.jp

Katsumi Matsuura, Seiichi Takao
{sea,tail}@arikilab.elec.ryukoku.ac.jp

Abstract

The purpose of this study is to automatically extract telop and flip characters. It starts from the extraction of stable frame sections including the telop and flip characters. Next process is the character region extraction and here we propose application of local line density to discriminate the telop characters from other elements such as lines and symbols. Final process is the telop character extraction and we show the effectiveness of the floating adaptive three level thresholding (FATLT) which thresholds the image intensity into three levels and finally binarizes the character regions precisely even in low contrast, taking topological relation between characters and their background into consideration.

1 Introduction

We are getting much information from conventional media such as broadcasted TV news. They are now going into digital media by communication satellite and will be transmitted through an internet in near future. Under this recent evolutional situation, unstructured huge data are produced and distributed so that we can hardly retrieve our desiring information. The motivation of this study is to construct a news video database by extracting the indices from the news articles in order to retrieve them[1].

In constructing a news video database, the manual extraction of the indices from the TV news articles is so much hard work. Therefore, it is desired to automatically extract the indices. The purpose of this study is to detect telop frames as well as flip frames from TV news articles in real time and to extract the video characters included in those frames robustly and precisely for the further indexing. There have been many researches on telop extraction. Lienhart proposed a method to extract telop characters through image segmentation by split and merge algorithm and character filtering by gradient value thresholding[2]. Since we pursue a real time character extraction, frames which include telop characters are detected at first and then character regions are extracted to reduce the processing time. Finally characters are extracted from the character regions.

The telop frames and flip frames have a property that they appear gradually and stay for a few second and disappear. We propose a method to extract such a "stable section" of the telop and flip frames in real time by detecting their appearance and disappearance. Our second proposal is a method to extract the robust character regions from the telop or flip frames using character intrinsic feature "line density" which has been used in the character recognition.

Our third proposal is to extract characters from the character regions precisely. We employed a floating adaptive three level thresholding algorithm which classifies the character images into three kinds of regions and then integrates them into the binary image of the characters and the background. The effectiveness of these three proposals is shown through experiments. Hereafter, we use a term "telop" to indicate both telop and flip for abbreviation.

2 Telop frame extraction

The telop sometimes appears gradually and disappears suddenly. Our purpose is to extract the *stable telop section* composed of a sequence of frames where all the telop characters have appeared. In order to locate the *stable section* quickly, the *telop appearing frame* and *telop disappearing frame* are extracted at first. The *stable section* is extracted within the section enclosed by them.

2.1 Extraction of telop appearing frame

Each frame is divided into 8 blocks; 4 in left side and 4 in right side. There are 21 blocks in each frame; 8 of elementary blocks, 10 of 2 adjacent elementary blocks, 3 of 4 adjacent elementary blocks. On each block, the intensity histogram is computed. Then the histogram distance is computed in the corresponding blocks between the previous frame and the present frame. The blocks with the histogram distance greater than some threshold are marked[3]. The histogram distance $Diff(H_1, H_2)$ computed between the histograms H_1 and H_2 is defined as the maximum among the absolute difference as shown in Eq.(1).

$$Diff(H_1, H_2) = \max\{abs(H_1(i) - H_2(i))\}_{i=1,\cdots,M} \quad (1)$$

In the marked blocks, the *edge correspondence* is computed at every pixel between the present frame t and the next frame $t+1$. The *edge correspondence* is defined by Eq.(2) and indicates that the gradient value $E_t(x, y, n)$ in the direction of $n(1 \leq n \leq 4)$ at each pixel (x, y) is greater than some threshold and continues in the consecutive frames t and $t+1$.

$$\begin{aligned} E_t(x,y,n) &= |I_t(x,y) - I_t(x,y,n)| \\ Edge_t(x,y,n) &= 1 \quad if \quad E_t(x,y,n) > th \\ Edge_t(x,y,n) &= 0 \quad if \quad E_t(x,y,n) \leq th \\ Ecor_t(x,y,n) &= Edge_t(x,y,n) \times Edge_{t+1}(x,y,n) \\ Ecor(x,y) &= \sqrt{\sum_n Ecor_t(x,y,n)^2} \end{aligned} \quad (2)$$

Then the number of pixels with the edge correspondence greater than some threshold is counted as *edge degree* in the marked blocks. If the present frame has at least one block with histogram distance and edge degree both greater than some threshold, it is regarded as the *telop appearing frame*.

2.2 Extraction of telop disappearing frame

The *telop disappearing frame* has a feature that the edge correspondence drops rapidly between the underlying frame and the successive frames. According to this feature, the edge correspondence is computed between the present frame and the previous frame. We denote this edge correspondence as $Ecor(previous)$. This value is computed by summing all the edge correspondence described by Eq.(2) over pixels within the marked blocks.

Then the edge correspondence is computed between the present frame and the next frame. We denote this edge correspondence as $Ecor(next)$. The edge correspondence ratio expressed by Eq.(3) is computed. If the ratio is less than some threshold, the present frame is regarded as the *telop disappearing frame*.

$$\text{Edge correspondence ratio} = \frac{Ecor(next)}{Ecor(previous)} \quad (3)$$

The *telop stable section* can be extracted as the consecutive frames whose edge correspondence ratio computed by Eq.(3) is almost one within the section enclosed by the *telop appearing frame* and the *telop disappearing frame*.

3 Character region extraction

The telop character regions are extracted according to the following processes, based on the telop features.

(1) Edge correspondence

Edge correspondence expressed by Eq.(2) is averaged at each pixel over frames within the stable section. Then the pixels with averaged edge correspondence greater than some threshold are extracted.

(2) Color difference

At the pixels extracted in (1), color difference is computed between consecutive frames and accumulated over frames within the stable section. The color difference at each pixel (x, y) between consecutive frames is defined by the following equation.

$$Color(x,y) = \frac{1}{N}\sum_{t=0}^{N-1} C_t(x,y) \quad (4)$$

$$C_t(x,y) = \sqrt{(R_{t+1} - R_t)^2 + (G_{t+1} - G_t)^2 + (B_{t+1} - B_t)^2} \quad (5)$$

$$Ecor(x,y) = \frac{1}{N}\sum_{t=0}^{N-1} Ecor_t(x,y) \quad (6)$$

Then the pixels with averaged color difference less than some threshold are extracted and marked as one. At this stage, the telop characters as well as lines, figure symbols are also extracted because they have strong edge continuation and weak color change.

(3) Local line density

To get rid of lines and figure symbols, local line density which presents the character features is computed on the extracted regions. The local line density is defined as follows using the vertical line density $LD(y)$ and the horizontal line density $LD(x)$;

$$\begin{aligned} LLD(x,y) &= \sum_i \sum_j \omega(i,j|x,y) \cdot (LD(i) + LD(j)) \quad (7) \\ LD(i) &= \sum_j (F(i,j) \oplus F(i,j+1)) \\ LD(j) &= \sum_i (F(i,j) \oplus F(i+1,j)) \end{aligned}$$

where $F(i, j)$ denotes the image with value 1 at the pixel satisfying the processes (1) and (2), value 0 otherwise. The $\oplus$ denotes the exclusive OR. The $\omega(i, j|x, y)$ is the weighing function decreasing from the center (x, y) to the peripheral line within the limited window as shown in Eq.(8). The local peaks of the local line density are extracted and the peaks greater than some threshold are determined as the telop character region.

$$\begin{aligned} \omega(i,j|x,y) &= \frac{1}{d(i,j|x,y)^2} \quad (8) \\ d(i,j|x,y) &= \sqrt{(x-i)^2 + (y-j)^2} \end{aligned}$$

4 Character extraction

4.1 Floating adaptive three level thresholding

The final stage is character extraction from the character regions extracted in the previous section. The conventional method will be binarization of the extracted character regions. However, the telop character is usually noisy due to video signal. Furthermore the contrast between the telop characters and the background is sometimes low depending on the background. This is completely different from handwritten characters or printed characters. Therefore the conventional binarization algorithm is difficult to be applied.

We employed a floating adaptive three level thresholding algorithm (FATLT) to binarize the telop character regions [4]. The FATLT segments the original image into three parts; character part, background part and the other part based on deviation image produced between the original image and the its moving averaged image. Fig.1 shows how the segmentation is performed by FATLT. In Fig.1(a), the intensity of the original image $I(x, y)$ along the x scanning direction is drawn by a solid line and its moving averaged intensity $\overline{I(x, y)}$ is shown by a dotted line.

The deviation image $\Delta I(x, y)$ is shown by the solid line in Fig.1(b). By thresholding this deviation image into three level, $B(x, y)$ image and $D(x, y)$ image are produced as shown in Eq.(9) to Eq.(12). The $B(x, y)$ image corresponds to character part and $D(x, y)$ image corresponds to the background part. Here σ is the standard deviation of $\Delta I(x, y)$. The parameter k is experimentally defined.

$$\overline{I(x,y)} = \sum_i \sum_j I(x+i, y+j) \quad (9)$$

$$\Delta I(x,y) = I(x,y) - \overline{I(x,y)} \quad (10)$$

$$B(x,y) = 1 \quad if \quad \Delta I(x,y) \geq k\sigma \quad (11)$$

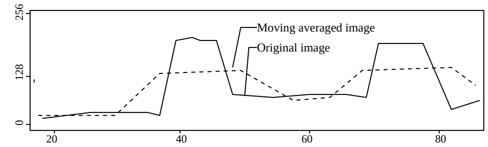
$$D(x,y) = -1 \quad if \quad \Delta I(x,y) \leq -k\sigma \quad (12)$$

4.2 Character Extraction Process

The character extraction process using FATLT is summarized as follows;

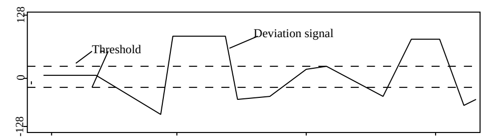
(1) Two kinds of rectangular mask windows are applied to the original image $I(x, y)$; one is horizontally long and the other is vertically long. Using this mask windows, two kinds of moving averaged images $\overline{I(x, y)}$ defined by Eq.(9) are produced. Fig.2(a) shows the example of the original image.

(2) Two kinds of $B(x, y)$ and $D(x, y)$ images are produced. Fig.2(b) and (c) show $B(x, y)$ and $D(x, y)$ images produced by x-directional rectangular mask. Small noises are removed on these images.



X-scanning direction

(a) Intensity of original image and moving averaged image



X-scanning direction

(b) Deviation image

Figure 1. Concept of floating adaptive three level thresholding

(3) Characters are extracted as the regions "1" (white) on the $B(x, y)$ image, if their surroundings are the region "-1" (white) on the corresponding $D(x, y)$ image.

(4) By AND operation of the two kinds of images produced in two different directions at (3), the final telop characters are obtained as the binary image. Fig.2(d) shows an example of the binary image obtained by the FATLT.

(5) By computing the horizontal peripheral distribution, the telop characters are separately extracted.

(a) Original image

(b) B-image (x-direction)

(c) D-image (x-direction)

(d) Binarized image

Figure 2. Example of telop character extraction by FATLT

5 Experimental results

5.1 Extraction of stable section

We carried out the experment to extract the *stable section* of the telop frames for the NHK TV news broadcasted at 12:00 and 23:00 on August 20 in 1998. The thresholds were determined according to the preliminary experiments. The experimental result is shown in Table1. The image size is 320 × 240 pixels and the images are compressed by 75% motion JPEG.

We evaluated the extraction results of the *telop stable section* by the *recall* and *precision rate* expressed in Eq.(13) and in Eq.(14).

$$Recall = \frac{z}{x} \quad (13)$$

$$Precision = \frac{z}{y} \quad (14)$$

Here, x and y indicate the number of *stable sections* which human extracted by his eyes and the number of stable sections which a computer extracted respectively. On the other hand, z is the number of stable sections which are included in both human extraction and the computer extraction.

Table 1. Extraction of stable section

	Recall	Precision
12:00	66.6(46/69)	34.8(46/132)
23:00	83.0(113/136)	35.7(113/316)
Total	77.6(159/205)	35.5(159/448)

5.2 Character region extraction

We further carried out the experment to extract the telop character regions from the stable sections for the same material. The thresholds were determined according to the preliminary experiments. The experimental results are shown in Table2. We evaluated the extraction results of the telop character regions by the *recall* and *precision rate*.

Table 2. Experimental result of character region extraction

	Character inclusion rate	Recall	Precision
12:00	96.5 (701/726)	89.5(77/86)	53.1(77/145)
23:00	92.3 (1670/1808)	85.8(140/163)	55.1(140/254)
Total	93.5 (2371/2534)	87.1(217/249)	54.4(217/399)

In the table, the *character inclusion rate* indicates how many characters are really included in the extracted character regions. Therefore even though the *recall rate* of the telop character regions is low, the successive character extraction and character recognition have a chance to keep the high rate, if the *character inclusion rate* is high.

5.3 Character extraction

We carried out the extraction experment of the telop characters from their character regions of the same material. The result was evaluated by an *extraction ratio* which is defined as the ratio of the number of the correctly extracted characters to the total number of the characters included in the extracted regions. The extraction rate was 80.2% (1903/2371). Through character region extraction and character extraction, the total extraction rate was 75.0% (93.5% × 80.2%). There are several reasons for 19.8% failure of character extraction. The main reason is extraordinary low contrast and smear for some characters. Characters with small size tend to be failed in character extraction.

6 CONCLUSION

We described the automatic extraction of flip and telop characters. It starts from the extraction of stable frame sections which include the telop and flip characters. There we proposed the extraction of the disappearing frames in addition to the appearing frames. Next was the character region extraction. There we proposed the local line density to discriminate the characters from other elements such as lines and symbols. Final is the telop character extraction. Here we showed the effectiveness of the floating adaptive three level thresholding which thresholds the image intensity into three levels. Taking topological relation between characters and their background into consideration, the character regions with low contrast can be binarized.

As the experimental results, we obtained 77.6% *recall rate* for the stable frame extraction. In the telop character region extraction, the *character inclusion rate* was 93.5% while their *recall rate* was 87.1%. The character *extraction rate* for the characters included in the extracted character regions was 80.2%.

References

[1] Y. Ariki and T. Teranishi, “Indexing and Classification of TV News Articles Based on Telop Recognition", ICDAR97, pp.422-427, 1997.

[2] R. Lienhart and F. Stuber, “Automatic text recognition in digital videos", in Image and Video Processing IV, Proc. SPIE 2666-20, 1996.

[3] H.Kuwano, S.Kurakake and K.Odaka, “Telop Detection Method for Content-Based Video Data Retrieval", PRMU96-98, pp.39-46, 1996. (in Japanese)

[4] K.Miyamoto, M.Tamada, I.Fujita, Y.Hayama and S.Eiho, “Extraction of Character String Region by a Correlation", IEICE, Vol.J81-D-II, No.9, pp.2052-2060, 1998.