MVA '96 IAPR Workshop on Machine Vision Applications, November. 12-14, 1996, Tokyo, Japan

Aligning Articles in TV Newscasts and Newspapers

Yasuhiko Watanabe† Yoshihiro Okada† Tatsuhiko Tsunoda‡ Makoto Nagao‡
†Dept. Electronics and Informatics, Ryukoku University, Seta, Otsu, Shiga, Japan
‡Dept. Electronics and Communication, Kyoto University, Yoshida, Sakyo, Kyoto, Japan
e-mail: watanabe@rins.ryukoku.ac.jp

Abstract

It is important to use pattern information (e.g. TV newscasts) and textual information (e.g. newspapers) together. For this purpose, we describe a new method for aligning articles in TV newscasts and newspapers. In order to align articles, the alignment system uses words extracted from telops in TV newscasts. The recall and the precision of the alignment process are 89% and 86%, respectively.

1 Introduction

Pattern information and natural language information used together can complement and reinforce each other to enable more effective communication than can either medium alone [Feiner 91] [Nakamura 93] [Watanabe 96]. One of the good examples is a TV newscast and a newspaper. In a TV newscast, events are reported clearly and intuitively with speech and image information. On the other hand, in a newspaper, the same events are reported by text information more precisely than in the corresponding TV newscast. Figure 1 and Figure 2 are examples of articles in TV newscasts and newspapers, respectively, and report the same accident, that is, the airplane crash in which the Secretary of Commerce was killed. However, it is difficult to use newspapers and TV newscasts together without aligning articles in the newspapers with those in the TV newscasts. In this paper, we propose a method for aligning articles in newspapers and TV newscasts.

2 TV Newscasts and Newspapers

2.1 TV Newscasts

In a TV newscast, events are generally reported in the following modalities:

- image information,
- speech information, and
- text information (telops).

In TV newscasts, the image and the speech information are main modalities. However, it is difficult to obtain the precise information from these kinds of modalities. The text information, on the other hand, is a secondary modality in TV newscasts, which gives us:

- explanations of image information,
- summaries of speech information, and
- information which is not concerned with the reports (e.g. a time signal).

In these three types of information, the first and sencond ones represent the contents of the reports. Moreover, it is not difficult to extract text information from TV newscasts. It is because a lots of works has been done on character recognition and layout analysis [Sakai 93] [Mino 96]. Consequently, we use this textual information for aligning the TV newscasts with the corresponding newspaper articles. The method for extracting the textual information is discussed in Section 3.1. But, we do not treat the method of character recognition in detail, because it is beyond the main subject of this study.

2.2 Newspapers

A text in a newspaper article may be divided into four parts:

- headline,
- explanation of pictures,
- first paragraph, and
- the rest.

In a text of a newspaper article, several kinds of information are generally given in important order. In other words, a headline and a first paragraph in a newspaper article give us the most important information. In contrast to this, the rest in a newspaper article give us the additional information. Consequently, headlines and first paragraphs contain more significant words (keywords) for representing the contents of the article than the rest.

Figure 1: An Example of TV news articles (NHK evening TV newscasts; April, 4, 1996)

米商務長官ら全員の死亡確認

クロアチア最南部のドブロブニク付近で3日午後、旧ユーゴ各国を視察中のブラウン米商務長官ら乗員・乗客計33人が乗った米空軍機が墜落した事故で、クロアチア政府は4日、ブラウン長官を含む乗客ら全員の死亡を確認したと言明した。墜落当時、現場は強い風雨に見舞われていた。国防総省スポークスマンは、砲撃や爆弾テロの可能性は考えられない、と述べた。

クリントン大統領は商務省で「バルカン半島に平和を根付かせるため、米国の経済力の生かし方を探る視察で、長官はたいへん意気込んでいた。長官は私にとって最も有能なアドバイザーの1人だった」と語った。ブラウン長官は今月中旬のクリントン大統領の訪日に同行する予定だった。

今回の事故にからみ、商務省は、メアリー・グッド次官（技術担当）を長官代行に任命した。

《写真》ボスニア・ヘルツェゴビナのツズラにある空軍基地に到着、軍用のボーイング737型機から降りて兵士たちの出迎えを受けたブラウン米商務長官。この後、同じ飛行機に再び乗ってドブロブニクに向かう途中に事故が起きた＝ロイター

乗客は27人で、商務省職員や、旧ユーゴの復興に関心を寄せる米企業幹部、ニューヨーク・タイムズ紙記者らが含まれていた。

米国人はボスニアで、和平協議を推進した外交官3人が昨年夏、事故で死亡した。今年1月には、米兵2人がやはり事故で死亡した。

Figure 2: An example of newspaper articles (Asahi Newspaper; April, 4, 1996)

(a) texts in the quarter region of left side

(b) texts in the lower third region

Figure 3: Examples of texts in the TV newscast

On the other hand, an explanation of a picture in an article shows us persons and things in the picture that are concerned with the report. For example, texts in bold letters under the picture in Figure 2 is an explanation of the picture. Consequently, explanations of pictures contain many keywords as well as headlines and first paragraphs.

In this way, keywords in a newspaper article are distributed unevenly. In other words, keywords are more frequently in the headline, the explanation of the picture, and the first paragraph. In addition, these keywords are shared by the newspaper article with TV newscasts. For these reasons, we align articles in TV newscasts and newspapers using the following clues:

- location of keywords in each article,
- frequency of keywords in each article, and
- length of keywords.

Figure 4: An example of title texts

3 Aligning Articles in TV Newscasts and Newspapers

3.1 Extracting Nouns from Telops

An article in the TV newscast generally shares many words, especially nouns, with the corresponding article in the newspaper. Making use of these nouns, we align articles in the TV newscast and in the newspaper. For this purpose, we extract nouns from the telops as follows:

Step 1 Extract texts from the lower third region and left/right quarter regions of the TV images (Figure 3) by hands. It is because these regions contain more nouns which represent the names of persons and things than the other regions. When the text is a title, we describe it. It is not difficult to find title texts because they have specific expression patterns, for example, an underline (Figure 4 and a top left picture in Figure 1). In addition, we describe the following kinds of information:

- size of each character
- distance between characters

Step 2 Divide the texts extracted in Step 1 into lines. Then, segment these lines at the point where the size of character or the distance between characters changes. For example, the text in Figure 3 (b) is divided into "沖縄県 (Okinawa Prefecture)", "大田 (Ohta)", and "知事 (Governor)".

Step 3 Segment the texts by the morphological analyzer JUMAN [Matsumoto 96].

Step 4 Extract nouns from the results of the morphological analysis if the last word is a noun. It is because a text the last word of which is not a noun is mostly a quotation of a speech. For example, the last word in Figure 5 is not a

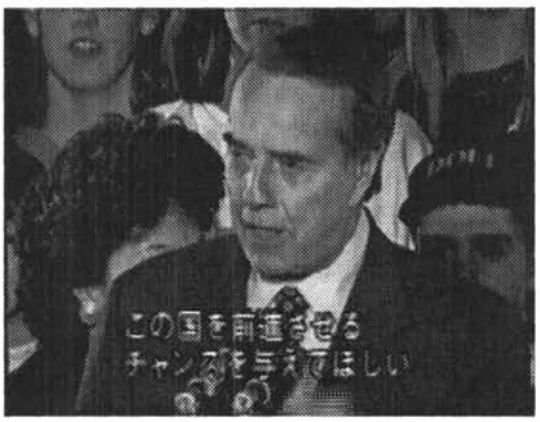

Figure 5: An example of a quatation of a speech

noun but an adjective. A quotation of a speech is used as the additional information and may contain inadequate words for aligning articles. In consequence, we don't use words in quotations of a speech for aligning articles.

3.2 Extraction of Layout Information in Newspaper Articles

For aligning with articles in TV newscasts, we use newspaper articles which are distributed in the Internet. The reasons are as follows:

- articles are created in the electronic form, and
- articles are created by authors using HTML which offers embedded codes (tags) to designate headlines, paragraph breaks, and so on.

Taking advantage of the HTML tags, we divide newspaper articles into four parts:

- headline,
- explanation of pictures,
- first paragraph, and
- the rest.

The procedure for dividing a newspaper article is as follows.

1. Extract a headline using tags for headlines.
2. Divide an article into the paragraphs using tags for paragraph breaks.
3. Extract paragraphs which start " 《写真(picture)》 " as the explanation of pictures.
4. Extract the top paragraph as the first paragraph. The others are classified into the rest.

		i			
		1	2	3	4
j	1	8	4	4	2
	2	4	2	2	1

i : the part of a newspaper
j : the part of a TV newscast

$$i = \begin{cases} 1 & : \text{title} \\ 2 & : \text{explanation of pictures} \\ 3 & : \text{first paragraph} \\ 4 & : \text{the rest} \end{cases}$$

$$j = \begin{cases} 1 & : \text{title} \\ 2 & : \text{the rest} \end{cases}$$

Table 1: The weight $w(i,j)$

3.3 Procedure for Aligning Articles

Before aligning articles in TV newscasts and newspapers, we chose corresponding TV newscasts and newspapers. For example, an evening TV newscast is aligned with the evening paper of the same day and with the morning paper of the next day. We aligned articles within these pairs of TV newscasts and newspapers.

The alignment process consists of two steps. First, we calculate reliability scores for an article in the TV newscasts with each article in the corresponding newspapers. Then, we select the newspaper article with the maximum reliability score as the corresponding one. If the maximum score is less than the given threshold, the articles are not aligned.

As mentioned earlier, we calculate the reliability scores using these kindes of clue information:

- location of words in each article,
- frequency of words in each article, and
- length of words.

If we are given a TV news article x and a newspaper article y, we obtain the reliability score by using the words $k(k = 1 \cdots N)$ which are extracted from the TV news article x:

$$SCORE(x,y) = \sum_{k=1}^{N} \sum_{i=1}^{4} \sum_{j=1}^{2} w(i,j) \cdot f_{paper}(i,k) \cdot f_{TV}(j,k) \cdot length(k)$$

where $w(i,j)$ is the weight which is given to according to the location of word k in each article. We fixed the values of $w(i,j)$ as shown in Table 1. As shown in Table 1, we divided a newspaper article into four parts: (1) title, (2) explanation of pictures, (3) first paragraph, and (4) the rest. Also, we divided texts in a TV newscasts into two: (1) title, and (2) the rest. It is because keywords are distributed

the number of the articles in the TV newscasts	160
the number of the corresponding article pairs	114
the number of the pairs of aligned articles	118
the number of the correct pairs of aligned articles	102

Figure 6: The results of the alignment

unevenly in articles of newspapers and TV newscasts. $f_{paper}(i,k)$ and $f_{TV}(j,k)$ are the frequencies of the word k in the location i of the newspaper and in the location j of the TV news, respectively. $length(k)$ is the length of the word k.

4 Experimental Results

To evaluate our approach, we aligned articles in the following TV newscasts and newspapers:

- NHK evening TV newscast, and
- Asahi Newspaper (distributed in the Internet).

We used 160 articles of the evening TV newscasts in this experiment. As mentioned previously, articles in the evening TV newscasts were aligned with articles in the evening paper of the same day and in the morning paper of the next day. Figure 6 shows the results of the alignment. In this experiment, the threshold was set to 100. We used two measures for evaluating the results: recall and precision. The recall and the precision are 89% and 86%, respectively. We may say that our approach is effective, because the precision and recall are relatively high.

One cause of the failures is abbreviation of words. For example, "信用金庫 (shinyo-kinko)" is abbreviated to "信金 (shinkin)". In our method, these words lower the reliability scores. To solve this problem, we would like to improve the alignment performance by using dynamic programming matching method for string matching.

In this experiment, we didn't align the TV news articles of sports, weather, stock prices, and foreign exchange. It is because the styles of these kinds of TV news articles are fixed and quite different from those of the others. From this, we concluded that we had better align these kinds of TV news articles by the different method from ours. As a result of this, we omitted TV news articles the title text of which had the special underline for these kinds of TV news articles. For example, Figure 7 shows a special underline for a sports news.

Figure 7: An example of a sports news article

5 Conclusion

In this paper, we propose a new method for aligning articles in TV newscasts and newspapers. The obtained results contributes to the information retrieval and multimedia. For example, the results are useful for the construction of the image retrieval system which uses the explanation texts of image data as clues for content based retrieval.

Acknowledgement

The data for this paper are obtained from NHK evening TV newscasts and Asahi Newspaper. We would like to express our gratitude to NHK and Asahi Newspaper for permission to use the data.

References

[Feiner 91] Feiner, McKeown: Automating the Generation of Coordinated Multimedia Explanations, IEEE Computer, Vol.24 No.10, (1991).

[Nakamura 93] Nakamura, Furukawa, Nagao: Diagram Understanding Utilizing Natural Language Text, 2nd International Conference on Document Analysis and Recognition, (1993).

[Matsumoto 96] Matsumoto et al: JUMAN Manual version 3 (in Japanese), Nagao Lab., Kyoto University, (1996) [1].

[Mino 96] Mino: Intelligent Retrieval for Video Media, Journal of Japan Society for Artificial Intelligence Vol.11 No.1, (1996).

[Sakai 93] Sakai: A History and Evolution of Document Information Processing, 2nd International Conference on Document Analysis and Recognition, (1993).

[Watanabe 96] Watanabe, Nagao: Diagram Understanding for Pictorial Book of Flora Using Integration of Pattern Information and Natural Language Information, ECAI 96 workshop on "Processes and Representations between Vision and Natural Language", (1996).

[1] The source file and the explanation (in Japanese) of Japanese morphological analyzer JUMAN can be obtained using anonymous FTP from
`ftp://pine.kuee.kyoto-u.ac.jp/pub/juman/juman3.0.tar.gz`