

Improvement of Telop Recognition Quality by Integrating Web Search Results

Do Ngoc HUNG
Tokyo Institute of Technology
Ookayama, Meguro
Tokyo, 152-8552, Japan

hungdn@de.cs.titech.ac.jp

Yutaka KATSUYAMA
Fujitsu Laboratories Ltd.
Kamikodanaka, Nakahara
Kanagawa, 211-8588, Japan

katsuyama@jp.fujitsu.com

Satoshi NAOI
Fujitsu Laboratories Ltd.
Kamikodanaka, Nakahara
Kanagawa, 211-8588, Japan

naoi.satoshi@jp.fujitsu.com

Haruo YOKOTA
Tokyo Institute of Technology
Ookayama, Meguro
Tokyo, 152-8552, Japan

yokota@cs.titech.ac.jp

ABSTRACT

Many scenes in recent TV programs display rich text information on the screen as Television Opaque Projectors or telops. Such telops are useful for obtaining information about the scene, or in searching for scenes using keywords. However, the quality of recognized results for telops, especially in Japanese or Chinese, is still poor even using the latest image recognition techniques. In this paper, we propose a method of integrating web search into the recognition process, to improve the quality of telop recognition results, focusing on Japanese news TV programs. At first, the proposed method recognizes telops in the news scenes by a current image recognition method. Next, it searches news web sites for related articles based on search keywords derived from the intermediate image recognition results. Then, the searched articles are used to build a context-based dictionary for correcting errors in recognized characters. We evaluate the proposed method with actual news TV programs and news web sites. The experimental results demonstrate that retrieved web articles are sufficiently related and effective for correcting the incorrectly recognized characters, even though these search keywords are derived from poor quality image recognition results. It indicates that the integration approach is effective in improving the precision of telop recognition results which is applicable for providing searched or summarized TV scenes along with web text as integrated information.

1. INTRODUCTION

TV programs are now commonly stored and watched via such digital devices as personal computers and digital video

recorders. In such situations, it is profitable to search the stored digital content for scenes that match given keywords or to summarize a TV program by extracting key scenes. The combination of TV scenes and well matched web information is also useful as integrated information [5].

To realize such functionality, textual information on a screen known as Television Opaque Projectors or telops are very useful [3]. However, the quality of recognized results for telops is still poor even utilizing the latest image recognition techniques. Of course, nowadays, many TV companies also provide digital closed captioning as subtitles for the hearing impaired. However, these closed captions are just spoken words, and not sufficient for deriving key scenes. If the precision of telop recognition can be sufficiently improved, telops will be more effective for scene search or summarization.

There are many factors affecting the precision of telop recognition results, such as the shape, size and color of characters, the complexity of the background and especially the characteristics of the languages used. While an alphabet-based language system such as English contains a very limited number of characters, other language systems, for example Asian languages such as Japanese and Chinese, contain a huge number of characters, many of which have very similar shapes. Therefore, these Asian languages require dedicated image recognition methods. However, the quality of recognition results for dynamic environments, such as with telops, is not sufficient. [7],[3].

Moreover, the recognition process is often based on a fixed lookup dictionary, and strongly affected by the image quality, which often results in incorrect recognition of context-dependent words or generating incorrect characters. This issue belongs to the class of spelling correction problems. In [2], the problems are classified into three types, "non-word error detection", "isolated-word error correction", and "context-dependent word correction". This article emphasized the need for "context-dependent spelling correction and word recognition techniques".

In this paper, we propose a method for integrating web search into the recognition process, to improve the quality of telop recognition results, focusing on Japanese news TV programs. The proposed method consists of three steps:

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the VLDB copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Very Large Data Base Endowment. To copy otherwise, or to republish, to post on servers or to redistribute to lists, requires a fee and/or special permission from the publisher, ACM.

VLDB '08, August 24-30, 2008, Auckland, New Zealand
Copyright 2008 VLDB Endowment, ACM 000-0-00000-000-0/00/00.

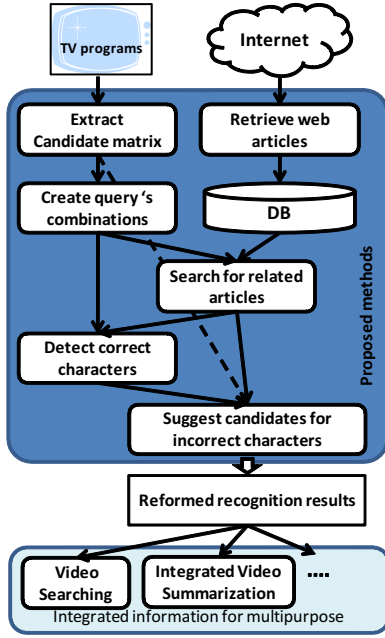


Figure 1: Process Outline

- Step1: Recognize telops in TV news screens by applying a current image character recognition method to generate search keyword combinations.
- Step2: Search recent news web sites for related articles based on the uncertain search keywords derived from the intermediate image recognition results.
- Step3: The returned articles are used to build a context-based dictionary for detecting and correcting errors in recognized characters.

We then evaluate the proposed method with actual Japanese news TV programs and news web sites in a certain time window including the air time of the TV programs. The experimental results demonstrate that retrieved web articles are sufficiently closely related, even though these search keywords are derived from the poor quality image recognition results. They also indicate that the context-based dictionary generated from these articles is effective for correcting the incorrectly recognized characters.

2. RECOGNITION WITH WEB SEARCH

We propose integrating web search into the telop recognition process to improve recognition quality. The high quality recognition results are applicable for providing scene search or key TV scene summaries with web texts as integrated information.

At first, the proposed method recognizes telops in Japanese news TV screens applying a current image recognition method for extracting a candidate matrix which is used to generate keyword combinations for web news articles. Next, it searches news web sites for related articles based on the generated search keywords and a time window including the air time of the target TV program. It then detects correct characters, and suggests candidates for incorrect characters in the candidate matrix. The outline of the proposed method is illustrated in Figure 1.

We can recursively apply the candidate-matrix improvement process, however we just focus on the effect of one level

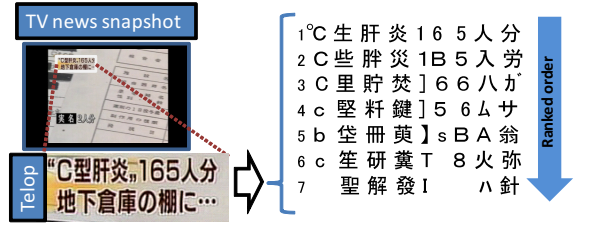


Figure 2: Candidate character matrix

application in this paper.

2.1 Image Recognition for Telops

Recently, the quality of image recognition has improved. However, there are still many incorrectly recognized characters for telops, especially those written in East Asian languages. Originally, there were no rules for the shape, size and color of telops. In the case of East Asian languages such as Japanese and Chinese, a huge number of different characters is used, many of which have very similar shapes. Moreover, the background of the same telop is dynamically changed because of movement in the original picture.

To compensate for this problem, many image character recognition systems generate an ordered list of relevant candidates for each character. We call the lists of all candidate characters generated for all positions of the telop as an intermediate recognition result the candidate character matrix. Figure 2 is an example of the candidate character matrix.

Here, we represent the candidate character matrix as

$$\begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ c_{m1} & c_{m2} & \dots & c_{mn} \end{pmatrix}$$

We make a set \mathcal{C}_k of character combinations from k elements chosen from the candidate character matrix, satisfying:

- the k elements belong to k contiguous columns.
- the elements are taken in column order.

Mathematically, \mathcal{C}_k is described by the following equation

$$\mathcal{C}_k = \{c_{i_1,j} \cdot c_{i_2,j+1} \dots c_{i_k,j+k-1} | 1 \leq i_1, i_2, \dots, i_k \leq m, 1 \leq j \leq n - k + 1\},$$

where $c_{i_1,j} \cdot c_{i_2,j+1}$ denotes the concatenation of characters $c_{i_1,j}$ and $c_{i_2,j+1}$, n is the number of recognized characters, m is the maximum number of candidates of each recognized character's position, and $c_{i,j}$ is an element of the candidate matrix.

2.2 Retrieval of Related Web Articles

Using the combinations created in the previous step, we calculate the degree of relationship of all the retrieved web articles to the TV program by the *Point Indicator* $pi(A)$, where A is an article. To correctly detect relevant articles, we focus on the frequency of the combinations in the articles, the rank order of each character of the combinations in the candidate list and the co-occurrence of each combination with others in article. $pi(A)$ is described by

$$pi(A) = \sum_{i=1}^N \left(tf.idf(w_i) \cdot \sum_{l=1}^k rank(w_i^l) \cdot \sum_{j=1, \neq i}^N C(w_i | w_j) \right),$$

where $w_i \in \mathcal{C}_k$, $N = |\mathcal{C}_k|$ and w_i^l is the l th character of combination w_i . The function $rank(w_i^l)$ assigns points for w_i^l depending on whether it is highly ranked in its candidate list or not. For example, we can use $rank(w_i^l) = 10$ if w_i^l is ranked first and $rank(w_i^l) = 1$ otherwise. The co-occurrence function $C(w_i|w_j)$ counts the number of times w_i and w_j co-occurs in the article. A simple version of $C(w_i|w_j)$ is

$$C(w_i|w_j) = \begin{cases} 1 & \text{if } w_i \& w_j \text{ co-occurs in the article} \\ 0 & \text{otherwise} \end{cases}$$

In preparation, we retrieved articles from different web sites hourly and stored them in a database. We apply the point indicator function to all articles published on web sites in a 48-hour window around the time of the TV broadcast. The result is, for each web site, a ranked list of articles based on the relationship with the TV program.

2.3 Error Character Detection

We adopt the k -gram approach [4] to detect error characters. The first ranked article is used to build a context-based dictionary, which contains all extracted k -grams of those articles (k - is also the number of characters of each combination in \mathcal{C}_k). We do the matching of k -gram created from the first ranked characters in CCM to the dictionary’s k -gram and eliminate the non co-occurring character strings. The remaining strings are considered to be correctly recognized. All of their characters are labeled correct and we leave the other characters in the recognition results unlabeled.

2.4 Correct Character Suggestion

Having the results from the previous step, we know whether a recognized contiguous string of characters is correct or not based on its label. In this step, we first detect the candidates for the unlabeled characters which are in neighboring (prefix or postfix) positions to labeled ones. For the prefix position of a contiguous labeled character string, we create a set of “postfix combinations” $\mathcal{B}_k = \{s_1, \dots, s_i \dots, s_k\}$, which is extracted from the contiguous characters by deleting the last character one by one until only one is left. The resulting set will contain all n -grams of contiguous labeled characters of decreasing length where the order of the characters in those n -grams is preserved. We then retrieve all characters having s_i as a postfix in the top articles we retrieved in the previous step and in the recognition result. The best relevant candidate for that position is the one that maximizes the ranking function $R(c)$,

$$R(c) = \sum_{i=1}^k [\alpha_i \cdot Fa(c.s_i) + \beta_i \cdot Fr(c.s_i)],$$

where, $k = |\mathcal{B}_k|$, $c.s_i$ is the concatenation of the character c with the string s_i and α_i and β_i are weight parameters. In the simplest case, we set $\alpha_i = \beta_i = 1$. The terms $Fa(c.s_i)$ and $Fr(c.s_i)$ are the frequencies of the string $c.s_i$ within searched articles and within the recognized result, respectively.

For unlabeled characters in the postfix position of labeled contiguous strings, we go through similar steps. We detect postfix and prefix characters in turn until all unlabeled characters have suggested relevant candidates.

3. EVALUATION

Table 1: TV news program data

Broadcast time	Nov.12 th ~Nov.23 rd
Number of news	9
Length	4’30”~9’50”

Table 2: Web article data

Published time	Nov.11 th ~Nov.24 th
Number of articles	1440

3.1 Experiment Environment

We retrieve web articles hourly from three news web sites: Asahi [9], Yomiuri [10] and Ann [8] News. To be able to detect the exact publishing time of the articles, we stored the articles’ title, content and also the time when it was retrieved and the names of the source web sites into database. In case of Japanese news web sites, at any time, the latest articles of forty eight hours are kept on web. In our experiments, we used the articles that were retrieved at the time of twelve hours after the broadcasted time of each TV news.

We used an image character recognition system developed by Fujitsu Laboratories Ltd. to recognize telops [1]. It generates the candidate character matrix of on-screen text of experimental video data. As input data, we recorded the daily news programs of NHK, Asahi, NTV and Fuji Television. These news programs often have a length varying from 30 minutes to one and a half hours. They also contain various types of content such as news, commercials, weather forecasts, etc. We manually cut the program into smaller video clips in which each clip contains the same type of content. Table 1 and 2 show the number and length of news programs as well as the number of articles for experiments.

3.2 Accuracy of Web Article Retrieval

Since the process of building the context-based dictionary for error correction uses the top ranked web articles based on uncertain keywords, the accuracy of the ranked results is influential for the error correction results. To evaluate the ranking of web articles, we used different types of input data, different definitions for the set of character combination k -grams (\mathcal{C}_k), and changed the scope in each web article where the points were calculated. We built bigrams (\mathcal{C}_2), trigrams (\mathcal{C}_3) and semantic bigrams, which are subsets of bigrams, where every element is a meaningful word. Using these combinations, we created different queries and estimated the article relationship degree based on both content and title of articles and on the title only. We used nine video news clips in our experiments, for each, we searched for related articles from three news web sites (Asahi, Yomiuri, Ann news). Totally, we did twenty seven experiments for each type (Table 3).

Since the title of an article usually represents a summary of important information within the article, searching titles for keywords should usually be influential for the correct ranking. However, in Table 3, the ratio of correct articles ranked first by title (\mathcal{C}_2 @title) is 55%. This indicates that the title of an article is not sufficient for ranking articles related to a telop. One reason is that there is a high possibility of describing the same event using different words, terms and different grammars. Another reason is that the generated candidate matrix still contains incorrect keywords. However, the combination of title and text content with bigram (\mathcal{C}_2 @content+title) provides a rather good ratio of correct articles (83%) even using uncertain keywords. This is useful

Table 3: Result of relevant article detection

	No of experiments	Correct article in 1 st rank
Word@content+title	27	69%
\mathcal{C}_2 @content+title	27	83%
\mathcal{C}_3 @content+title	27	55%
\mathcal{C}_2 @title	27	55%

Table 4: Result of error correction

Video	Total# characters	#Wrong characters	#Fixed characters	Fixed ratio
Video1	59	24	5	20.8%
Video2	25	3	2	66.7%
Video3	34	28	1	3.5%
Video4	71	25	7	28%
Video5	15	5	1	20%
Video6	43	15	3	20%
Average				26.5%

enough to detect and correct incorrectly recognized characters in the candidate character matrix.

3.3 Error Correction Ratio

Based on the evaluation of related web article retrieval for the telop, we selected the combination of title and text content with bigram (\mathcal{C}_2 @content+title) to search for related web articles and use these results in the evaluation of error correction.

Of the nine video clips used in the previous step, three have no errors in the image recognition results. Therefore, we used the remaining six video clips to evaluate error correction using the context-based dictionary generated from the related web article.

Table 4 lists the total number of characters in telops in each video clip, the number of characters incorrectly recognized by the image recognition process, the characters fixed by the proposed method, and the fraction of incorrectly recognized characters repaired by this method. The average repair rate is 26.5% for these six video clips. It performed better than the error correction rate for speech recognition using web information reported by [6] at 19.9%.

4. CONCLUSIONS

In this paper, we propose a method for integrating web search into the recognition process to improve the quality of telop recognition results. At first, the proposed method recognizes telops in Japanese news TV programs using a current image recognition method to generate a candidate character matrix. The generated candidate character matrix is used to create keyword combinations for searching for related web news articles. Then, it builds a context-based dictionary from these web articles to correct errors in the image recognition results by detecting erroneous characters and suggesting relevant candidates for undetected characters.

We also evaluated the proposed method using actual Japanese news TV programs and news web sites. The experiment indicated that the combination of title and text content with bigrams is good for retrieving related web articles using the recognized telop information, giving 83% correct articles. It also indicated that the context-based dictionary built from the retrieved related articles are effective for improving the quality of recognized results, and the average character re-

pair is 26.5%, outperforming a previously reported result.

The evaluation results indicate that the integration approach is effective in improving the precision of telop recognition results. The recognition results are also applicable for providing integrated information from web texts together with scene search results or TV scene summaries.

In future work, we plan to do a more detailed evaluation using a large number of video clips. Since we estimate that the recursive application of the proposed process, which recursively use repaired recognition result as new input query for new web articles search is effective for improving the accuracy of the correction, we plan to implement a recursive version. We also plan to extend the target of application not only to news programs but to TV programs in general. Application of the recognition process to providing integration of the TV clip with matched web information is also planned.

5. ACKNOWLEDGMENTS

This work is partially supported by a Grant-in-Aid for Scientific Research of MEXT Japan (#19024028), Tokyo Institute of Technology 21 COE Program “Framework for Systematization and Application of Large-Scale Knowledge Resources”, and CREST of JST (Japan Science and Technology Agency).

6. REFERENCES

- [1] Y. Katsuyama, H. Bai, H. Takebe, and K. Fujimoto. A study for caption character pattern extraction. PRMU2007 239, IEICE, 2008.
- [2] K. Kukich. Techniques for automatically correcting words in text. *ACM Computing Surveys*, 24(4):377–439, 1992.
- [3] H. Kuwano, Y. Taniguchi, H. Arai, M. Mori, S. Kurakake, and H. Kojima. Telop-on-demand: video structuring and retrieval based on textrecognition. In *Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on*, volume 2, pages 759–762, 2000.
- [4] J. H. Lee and J. S. Ahn. Using n-grams for korean text retrieval. In *SIGIR '96: Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 216 – 224, 1996.
- [5] H. Miyamori, Z. Stejic, T. Araki, M. Minakuchi, Q. Ma, and K. Tanaka. Towards integration services for heterogeneous resources: An integrated search engine forweb content and tv programs. In *SKG '06. Second International Conference on Semantics, Knowledge and Grid*, pages 19–19, 2006.
- [6] H. Nishizaki and Y. Sekiguchi. *Word Error Correction of Continuous Speech Recognition Using WEB Documents for Spoken Document Indexing*. Springer Berlin/Heidelberg, 2006.
- [7] T. Sato, T. Kanade, E. Hughes, and M. Smith. Video ocr for digital news archives. In *CAIVD '98: IEEE Workshop on Content-Based Access of Image and Video Databases*, pages 52–60, 1998.
- [8] The ANN NEWS. <http://www.tv-asahi.co.jp/ann/news/web/index.html>.
- [9] The Asahi Shinbun. <http://www.asahi.com>.
- [10] The Yomiuri Shinbun. <http://www.yomiuri.co.jp>.