EA Snippets: Generating Summarized View of Handwritten Documents Based on Emphasis Annotations

Hiroki Asai and Hayato Yamana

Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan
{asai,yamana}@yama.info.waseda.ac.jp

Abstract. Owing to the recent development of handwriting input devices such as tablets and digital pens, digital notebooks have become an alternative to traditional paper-based notebooks. Digital notebooks are available for various device types. To display a list of text documents on a device screen, we often use scaled thumbnails or text snippets summarized through natural language processing or structural analyses. However, these are ineffective in conveying summaries of handwritten documents, because informal and unstructured handwritten data are difficult to summarize using traditional methods. We therefore propose the use of emphasis-based snippets, i.e., summarized handwritten documents based on natural emphasis annotations, such as underlines and enclosures. Our proposed method places emphasized words into thumbnails or text snippets. User studies showed that the proposed method is effective for keyword-based navigation.

Keywords: Digital Ink, Annotation, Summarization, Thumbnail, Snippets.

1 Introduction

Tablet and digital-pen devices that accept handwriting inputs have grown more common, and they are used as alternatives to traditional paper-based notebooks [1]. When digital notebooks replace paper-based notebooks, handwritten data will change from off-line to on-line formats. On-line handwritten data obtained by handwriting input devices, in addition to brushstroke coordinate information, include time-series data and pressure factors. As the number of on-line handwritten documents increases, the ability to search for information from such documents should be developed. One of the key ability in document search is how fast we can grasp the summary of each listed document. When we search document, we find the desired document effectively to scan through the summary list of documents called snippets, which is possible to grasp the summary of documents without scanning the contents. This paper presents a new summarized snippets of on-line handwritten documents which improve searching own or others handwritten documents, such as thumbnails and summarized text required in search systems.

Displaying scaled thumbnails (scaled pictures of original content) is an effective way to scan through lists of documents (e.g., thumbnails are effective for Web

S. Yamamoto (Ed.): HIMI 2014, Part II, LNCS 8522, pp. 20–31, 2014.

searches [3][7]). For example, Web image-retrieval services such as Google Images and Yahoo! Image Search output scaled thumbnails in a list view of the search result pages. From these thumbnails, users can see an outline of the original image. However, we cannot use traditional scaled thumbnails to understand a summary of handwritten documents since the text size is too small to read on small device screens.

In addition, text snippets, i.e., portions of original text, are commonly used in a list view of search result pages. For instance, search results from Google Web Search display text snippets, which are constructed by extracting a series of words including the query word. In handwritten documents, however, the accuracy of recognizing handwritten characters is as low as 92.77% [10], resulting in a difficulty in adopting natural language processing to summarize handwritten documents.

To the best of our knowledge, no research has been conducted on navigational views of handwritten documents, although some research does exist on search views of images and Web pages. In our previous work [2], we proposed the extraction method of emphasized words in handwritten notebooks. In this paper, we propose handwritten-document views called "EA Snippets" based on natural emphasis annotations, such as underlining and enclosures, and not based on natural language processing by using our previous work. Two types of EA Snippets are shown:

- Image EA Snippets consisting of important words or graphs, where the text is expanded for easier readability.
- Text EA Snippets consisting of both summarized text and a scaled thumbnail, where summarized text consists of important words listed in order of importance.

Furthermore, we investigate the performance of these proposed snippets types when users search for information in handwritten documents.

2 Related Work

In this section, we refer to researches on thumbnails of images and Web pages, and investigate a method to generate a thumbnail from important parts of a document, which is same approach with our method. Our proposed method generates EA Snippets consisting of important handwritten objects; however, to date there is no research concerning the generation of thumbnails for handwritten documents. Consequently, we look to apply research concerning thumbnails for images [1][4][11], and thumbnails for Web pages [9][12][13] to generate our handwritten thumbnails.

2.1 Thumbnails of Image

Several studies have investigated how to improve the thumbnails of pictures. Amurutha et al. [1] proposed an intelligent automatic cropping technique for pictures. Cropping is used to extract the rectangular area containing the attention objects. Prior to shrinking an image, they used Regions of Interest (ROIs) to crop objects from images. Their experiments showed that thumbnails efficiently increased the performance of context-based image retrieval (CBIR). Suh et al. [11] proposed two automatic cropping techniques; the first detects salient portions of images, while the other is a method of automatic face detection. They generated thumbnails by cropping these detected areas. Their user study shows that these methods resulted in small thumbnails that can be easily find through visual search. Avidan et al. [4] proposed an image resizing method, called Seam Carving, which supports content-aware image resizing. Seam Carving creates the energy map of an image, and then shrinks the image by removing the minimum energy path from left to right, or from top to bottom. Because Seam Carving does not discriminate between attention and other objects, the attention objects become distorted as the image shrinks.

2.2 Thumbnail of Web Pages

Other studies aimed in improving thumbnails of Web pages. Teevan et al. [12] extracted title-texts, logo images, and salient images from Web pages, and produced thumbnails by compiling these component pieces. Their experiments showed that in re-finding tasks, their thumbnails enabled users to find Web pages faster than snippets of text and traditional thumbnails. Woodruff et al. [13] proposed textually enhanced thumbnails of Web pages. These enhanced thumbnails were created by enhancing screenshots of Web pages with query words. In their study, participants searched faster using the textually-enhanced thumbnails than when using the plain thumbnails and text summaries. Lam et al. [9] proposed a thumbnail enhanced with readable text fragments. In their user study, when participants used the proposed thumbnail interface, they could find the area containing the target content in Web pages approximately 41% faster, and with 71% lower error rate, compared to traditional interfaces.

These related studies described in 2.1 and 2.2 proposed methods for detecting important objects, and producing as outputs summarized thumbnails of images and Web pages. We propose a method to detect important objects in handwritten documents by detecting emphasis annotations. Next, based on the results of previous related studies, we use the detected enhanced objects and summarize handwritten documents with EA Snippets.

3 Emphasis Annotations

In this section, we describe natural emphasis annotations used in notebooks. First, we defined some frequently used natural emphasis annotations, which are often used in notebooks. We then performed two investigations: 1) how often emphasis annotations are used in notebooks, and 2) under which situations they are utilized.

We collected 278 handwritten pages from the notebooks of eight university students in their 20s, studying such subjects as mathematics, physics, chemistry, and programming for six months to 1 year. The notebooks were written in Japanese, and our analysis shows that they include three types of natural emphasis annotations: 1) enclosing words, 2) underlined words, and 3) colored words. In addition, we found that emphasis annotations were performed 3.4 times per page on average. Table 1 shows the number of occurrences for each emphasis annotation.

Table 1. Type of emphasis annotation and the number of occurences in collected data

Emphasis Annotation	Number of occurrences
Enclosing Words	345
Underlined Words	304
Colored Words	296

We also interviewed the students to confirm when such emphasis annotations were performed. As a result, we found the following three types of situations:

1. Emphasizing important words or equations
2. Highlighting titles or topics
3. Highlighting a summary of the contents

Furthermore, the participants stated that they also emphasize titles or topics in the index area of the notebook instead of using emphasis annotations.

From the results of our survey, we found that the emphasized words indicate keywords, topics, or a general summary. We therefore assumed that we can easily understand a summary for extracting words and figures based on emphasis annotations and the index area. In addition, our previous work [2] calculated emphasis strength, which represents the importance of the emphasis annotation, each emphasis annotations in Table 1 from the questionnaire survey. We also use the emphasis strength to calculate the importance of handwritten objects.

4 Implementation

In this paper, we proposed two types of snippets based on emphasis annotations: 1) Image EA Snippets (see Fig. 1 (d)) and 2) Text EA Snippets (see Fig.2).

Our system detects both emphasis annotations and words in the title index area of notebook. Then, emphasis annotations are extracted followed by calculating emphasis scores by the method proposed in [2]. Emphasis scores represent the strength of the author's emphasis. Following the calculation of the emphasis scores, our system generates thumbnails or text snippets based on the emphasis annotations of authors.

4.1 Text/Non-text Classification

To detect handwritten diagrams and emphasis annotations, our system classifies all the input strokes into either text strokes or non-text strokes by applying an SVM. We use the following four stroke features as inputs to SVM after reducing the noise of handwritten strokes by using Gaussian filter:

1. Stroke length
2. Stroke curvature

3. Long side length of the stroke's bounding box
4. Number of crossing other strokes

4.2 Emphasizing/Graph Classification

After text/non-text classification, non-text strokes are further classified into emphasis strokes and graph strokes. Emphasis strokes consist of both underlined and enclosing strokes. On the other hands, graph strokes consist of non-emphasis strokes like diagrams and illustration.

Here, a stroke is classified as an underlined stroke when the height of the stroke's bounding box located under the word's bounding box, is within the height of the word's bounding box. Specifically, non-text strokes satisfying the following two conditions are categorized as underlined:

1. Shape condition

$$\begin{cases} 2W_{WordAve} < W_{Stroke} \\ H_{WordAve} > H_{Stroke} \end{cases}$$

2. Neighborhood character count condition

When two or more neighborhood characters satisfy the following conditions:

$$\begin{cases} \min(X_{Stroke}) < X_{WordG} < max\ (X_{Stroke}) \\ Y_{WordG} - H_{WordAve} < min\ (Y_{Stroke}) \\ \max(Y_{Stroke}) < Y_{WordG} \end{cases}$$

Variables $\mathrm{H_{WordAve}}$ and $\mathrm{W_{WordAve}}$ are the average height and width of the characters in the page. Variables $\mathrm{H_{Stroke}}$ and $\mathrm{W_{Stroke}}$ are the height and width of the target stroke. Variables $\mathrm{X_{Stroke}}$ and $\mathrm{Y_{Stroke}}$ are the sets of x- and y-coordinates of the target stroke. Variables $\mathrm{X_{WordG}}$ and $\mathrm{Y_{WordG}}$ are the x- and y-coordinates of the median point of the characters' bounding box.

Conversely, the enclosing stroke is extracted if its bounding box encloses the word's bounding box. Specifically, non-text strokes satisfying the following two conditions are categorized as enclosing:

1. Shape condition

$$\begin{cases} 2W_{WordAve} < W_{Stroke} \\ \frac{1}{2}H_{WordAve} < H_{Stroke} \end{cases}$$

2. Comprehension character count condition

The bounding box of the target stroke contains the center point of character, and the number of characters in the bounding box of the target stroke is greater than or equal to

$$\max\left(2, \frac{\mathrm{S_{Stroke}}}{\alpha \mathrm{S_{WordAve}}}\right)$$

Variable S_{Stroke} represents the bounding box area of the target stroke. $S_{WordAve}$ is the average bounding box area of the characters in the page. Variable α is the threshold of the character's density, which we set to 6.0 to maximize detecting accuracy.

4.3 Recognizing Emphasized Words

After emphasizing/graph classification, we detect which part of the text is emphasized by the author, and which patterns of emphasized expression are present in the text. First, our system splits text strokes into character groups. We use .NET Ink Analyzee for the character grouping. After grouping, we detect underlined and enclosed words by using the spatial relationships between character groups, and the underlined and enclosing strokes we extracted from non-text strokes.

Underlined words are located above the underline stroke. Thus, our method detects underlined words by extracting the words satisfying the following conditions:

$$\begin{cases} \min(X_{Underline}) < X_{WordG} < \max(X_{Underline}) \\ \min(Y_{Underline}) < Y_{WordG} \\ Y_{WordG} < \max(Y_{Underline}) + \frac{3}{2} H_{WordAve} \end{cases}$$

Variables $X_{Underline}$ and $Y_{Underline}$ represent the sets of x-and y-coordinates of the underline strokes we extracted from non-text strokes. Conversely, enclosed words are located within the area enclosed by the enclosing stroke. Thus, our method detects enclosed words by extracting words whose median points are within the bounding box of the enclosing stroke.

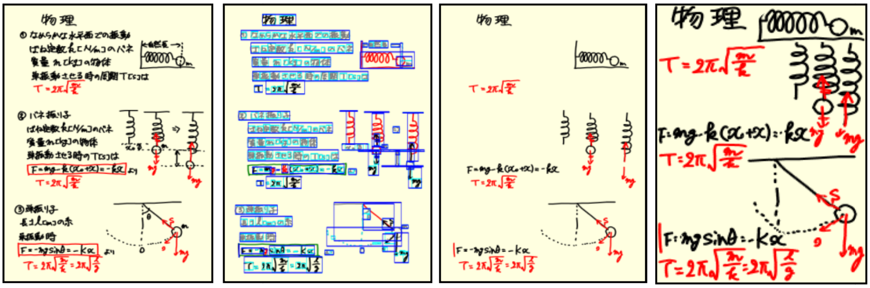
4.4 Calculating Emphasized Scores

When a word is classified as an emphasized word, we calculate its emphasis score, indicating the importance of the word. First, our system groups strokes by the kind of handwritten object to avoid displaying handwritten strokes discretely. We sequentially check strokes ordered in a time series. Two strokes adjacent in a time series are grouped together if the distance between the center gravities of the strokes is less than the threshold value and the type of emphasis of the adjacent strokes is the same. After grouping handwritten strokes, the score is calculated based on the emphasis strength, calculated from questionnaire survey in [2], each group.

4.5 Generating Image EA Snippets

Compared to traditional scaled thumbnails for images, we should take into account when used for handwritten documents:

- The text in a scaled thumbnail is too small to read.
- The amount of text, i.e., the amount of information, in a scaled thumbnail is not reduced compared to the original data. Due to this, the cognitive load of understanding contents is not reduced.



(a) Original data (b) Grouping (c) Object Selection (d) Zooming objects

Fig. 1. The process of generating proposed Image EA Snippets

Therefore, we propose “Image EA snippets” summarizing the intended emphasis of authors. Our proposed method summarizes the contents of handwritten data based on emphasis, such as underlines and enclosings, and increases the size of text in the contents of the thumbnail. Fig.1 shows the process to generate Image EA Snippets.

First, our system groups handwritten strokes, then calculate their emphasis scores by applying the method proposed in [2] (see Fig.1 (b)). Second, the number of text stroke groups is decreased by removing the groups whose emphasis scores are under a threshold (see Fig.1 (c)). Note that non-text stroke groups, such as diagrams, are not removed. The adopted threshold is the maximum value satisfying the following condition;

$$\sum_{n=1}^{N_{SG}} S_{group}(n) B_{thres}(n) < \beta S_{org}$$

where N_{SG} represents the number of stroke groups, and $S_{group}(n)$ returns the area of the bounding box of the nth stroke group. If the emphasis score of the nth stroke group is more than the threshold, $B_{thres}(n)$ returns one, otherwise it returns zero. The scaling rate of the thumbnail is denoted by β, and the area of the original contents is denoted by S_{org}.

Finally, the stroke groups are reallocated and expanded by using the Seam Carving method [4]. Using this method, we can scale down a handwritten document by removing blank spaces, removing contents below the threshold, and maintaining the alignment of stroke groups. Fig. 1(d) shows our proposed thumbnails scaled by the Seam Carving method. From this thumbnail, we can understand the summary of the contents.

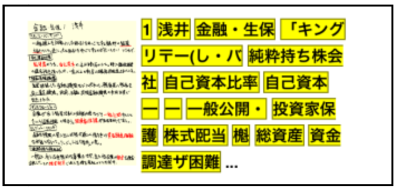


Fig. 2. Text EA Snippets generated by our method

4.6 Generating Text EA Snippets

Here, we present a method to generate "Text EA Snippets" based on the intended emphasis of authors. First, our method applies a handwritten recognition method to text stroke groups. Specifically, we use the .Net Ink Analyzer for handwritten recognition. Next, we sort text stroke groups by their emphasis scores, and clip at a maximum the 80 top-ranked words. Finally, our method displays scaled thumbnail to help users understand the layout and graphs of the contents in addition to the 80 top-ranked words. Fig.2 shows our proposed Text EA Snippet summarized by the emphasis annotations. From the text snippet, we can understand the keywords in the contents.

5 User Study

5.1 Collecting On-line Handwritten Data

Compared to traditional paper-based notebooks, digital notebooks enable us to collect on-line handwritten data consisting of a time series of strokes, pressure, and writing speed. Using digital notebooks, we can analyze handwritten data in more detail because we have more information than using traditional off-line handwritten data. Our method uses on-line handwritten data to detect the intended emphasis of authors. Hence, we have developed an experimental system for Windows (using the pen tablet WACOM Cintiq 12WX) to collect on-line handwritten documents.

We have developed our system in Visual C# equipped with a pen tablet device to enter inputs by handwriting. We collected 42 pages (consisting of 38,416 handwritten strokes) of on-line handwritten notebook data written by eleven university students majoring in computer science. We gave them a document containing common topics and current events, and informed them about the important words in the documents. Participants were instructed to create a note summarizing the documents. Note that participants were not forced to follow any format, i.e., participants could emphasize important words using any emphasis expression they wanted, and were allowed to use the notebook in any way they chose.

5.2 Recognition of Emphasized Words

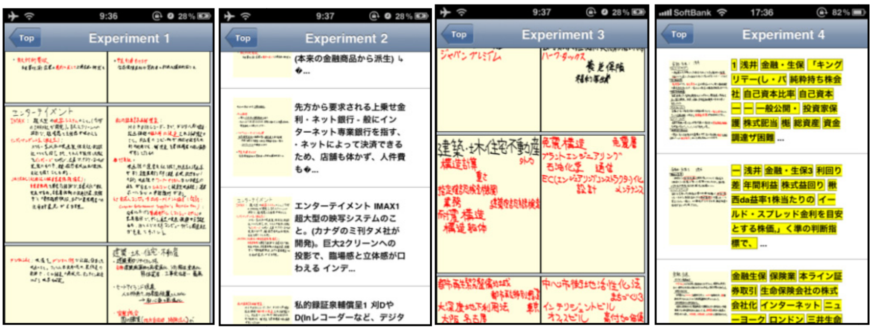
First, we evaluated the recognition performance of our detection method for emphasized words. Here, words in the title index and colored words were successfully detected from on-line handwritten data by using their color and their written area.

We investigated 38,416 strokes contained in the handwritten documents. The manual classification of the documents resulted in 16 enclosings and 72 underlines. Our method detected all the enclosings in the documents with no errors. Conversely, our system detected underlines with 85.71% precision rate, and 96.43% recall rate. We found that text written by hand above the ruled line was falsely recognized as underline. In addition, some underlines could not be detected, because the underline was located far from handwritten text.

5.3 Search Performance

We conducted a user study to compare the search time required for handwritten documents using both traditional thumbnails and our proposed EA Snippets. To measure the search performance, we developed an evaluation application that shows various views of handwritten documents and operates on various device types. We used an iPhone 3GS, and a screen capture of our experimental system is shown in Fig.3. In this study, we compared the following four view types:

1. Traditional Scaled Thumbnails, which are reduced versions of the original image (see Fig.3 (a)).
2. Traditional Head Text Snippet + Scaled Thumbnail, which are generated by recognizing the first 80 characters of handwritten text in a document. A scaled thumbnail is also presented (see Fig.3 (b)).
3. Proposed Image EA Snippets, which are summarized based on their emphasis scores (see Fig.3 (c)).
4. Proposed Text EA Snippets + Scaled thumbnails, which are also summarized based on their emphasis scores (see Fig. 3 (d)).



(a) Traditional (b) Head text (c) EA Thumbnail (d) EA Text

Fig. 3. Screen shot of experimental system

On the same screen, four pages are displayed together for 1) and 3). On the other hand, 2.5 pages are displayed for 2) and 4). The goal of this study is to verify which snippets enable us to find information more easily.

We conducted three types of evaluation in our user study. On the first study (described in “1)”), we performed the comparison of the search time of four snippets types to answer the fill-in-blank question, on condition that the keyword of the question is included in the proposed view. After the first study, we also conducted the additional studies in addition to the first study because the first study leaves the two questions; a) we did not consider the situation in which the users searched document by using proposed snippets which is not include the keyword of the question, and b) there is no consideration of document's author, that is to say we did not consider the difference of the performance searching in own documents or other's documents. We conducted the two additional user studies to evaluate the two questions (described in "2)" and "3)").

Search Performance (Emphasized Words Include Search Keywords). First, we evaluate the search performance. We invited twenty participants to participate in our user study, including eleven who were authors of the collected handwritten documents. All participants were university students in their 20s, two of them women. We performed the user study using the four snippets types shown in Fig. 3, and measured the time required to finish answering the questions from each view. In each experiment, all participants were given twenty pages of handwritten documents each from the collected data, along with five questions. All participants were given the same questions and handwritten documents. The participants were required to answer the questions by navigating using the views generated from the documents. The questions were fill-in-the-blank types, and the answers were written directly on the original handwritten documents. In addition, the keywords of each question were indicated using emphasis annotations.

Fig.4 shows the average search time and the standard error. In addition, we performed a Kruscal-Wallis test, and conducted a pairwise comparison of the results. The results show that, on average, our proposed Image EA Snippets result in the best search time among the four snippets types. Compared with the traditional scaled thumbnails, we found that our Image EA Snippets enable users to search 42% faster ($p < 0.001$) on average. On the other hand, compared to traditional head text snippets, our Text EA Snippets also enable users to search faster on average, although the difference is not statistically significant ($p > 0.1$). Moreover, the results show that our Image EA Snippets help users search faster than do our Text EA Snippets ($p < 0.0001$).

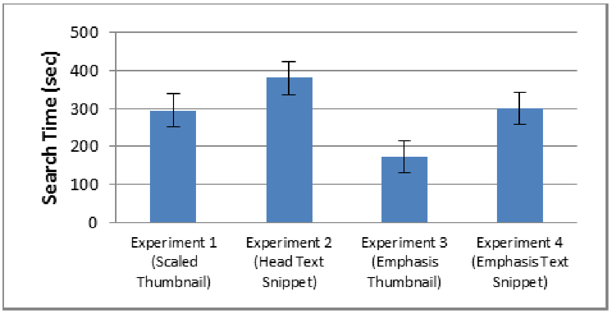


Fig. 4. Results of search performance in user study

Search Performance (Emphasized Words Include No Search Keywords). For the above task, the users could search for answers based on the emphasized keywords. In other words, we did not consider situations in which the users searched for information not based on the emphasized words. We therefore conducted another user study ($N = 10$) for searches not based on the emphasized keywords. Herein, N is the number of subjects used. The results show that the proposed Image EA Snippets enable users to search 24% faster on average than traditional scaled thumbnails, although the difference is not statistically significant ($p > 0.1$), while Text EA Snippets enable only a

9% speed increase ($p > 0.1$). After this study, some participants commented that they could find the page if they can imagine the keyword from the words which include in the snippet. From these results, we found that our proposed snippets is effective when we can imagine the information we want to know from the words or graph showing in snippet.

Comparison of the User's Own Notes and the Notes of Others. In addition, we also investigated the difference between searching one's own notes and the notes of others ($N = 4$). The results show that, using traditional scaled thumbnails, the users found the answer 206% slower on average for notes other than their own ($p < 0.05$). In contrast, there are no statistical differences between searching one's own notes or the notes of another student for Image EA Snippets and Text EA Snippets ($p > 0.1$). From these result, we found that our proposed snippets is effective for navigating pages of handwritten documents to find information regardless of the authors.

Discussion. After the user study, we discussed with the participants our proposed snippets. Some of them commented that they could not understand what was written in traditional scaled thumbnails, because characters were too small to read. Conversely, they could guess the contents in our proposed Image EA Snippets, and our proposed image snippets often helped them in searching for information in handwritten documents. Participants also reported that if the exact search keyword was not included in the thumbnail, they had trouble determining the contents of the thumbnail. On the other hand, some participants reported that Text EA Snippets occasionally did not help them understand the summary of the handwritten documents, because the accuracy of the handwritten character recognition was low. In addition, some of them said that they often looked scaled thumbnail only in the text snippet. We believe that the removal of these limitations could improve the searching speed.

6 Conclusion

In this paper, we discussed the ineffectiveness of traditional thumbnails in information retrieval when targeting handwritten documents, and presented a new approach, i.e., detecting natural emphasis annotation. We proposed the use of EA Snippets summarized by emphasis annotations. In the user study, we conducted that our proposed snippets are effective for navigating in handwritten documents. In addition, we found that thumbnails are more effective than text snippets for searching handwritten documents because handwritten data are hard to recognize that results in defective structural analysis.

References

1. Amurutha, I.S., Shylaja, S.S., Natarjan, S., Balasubramanya Murthy, K.N.: A smart automatic thumbnail cropping based on attention driven regions of interest extraction. In: Proc. ICIS 2009, pp. 957–962. ACM (2009)

2. Asai, H., Yamana, H.: Extraction of Emphasized Words from On-line Handwritten Notebooks DBSJ Journal. DBSJ 10(1), 67–72 (2011) (in Japanese)
3. Aula, A., Khan, R.M., Guan, Z., Fontes, P., Hong, P.: Comparison of visual and textual page previews in judging the helpfulness of web pages. In: Proc. WWW 2010, pp. 51–60. ACM (2010)
4. Avidan, S., Shamir, A.: Seam carving for content-aware image resizing. In: Proc. SIGGRAPH 2007, Article 10. ACM (2007)
5. Brandl, P., Richter, C., Haller, M.: NiCEBook: supporting natural note taking. In: Proc. CHI 2010, pp. 599–608. ACM (2010)
6. Chen, M., Sun, J.-T., Xeng, H.-J., Lam, K.-Y.: A Practical system of keyphrase extraction for web pages. In: Proc. CIKM 2005, pp. 277–278. ACM (2005)
7. Dziadosz, S., Chandrasekar, R.: Do thumbnail previews help users make better relevance decisions about web search results? In: Proc. SIGIR 2002, pp. 365–366. ACM (2002)
8. Hinckley, K., Zhao, S., Sarin, R., Baudisch, P., Cutrell, E., Shilman, M., Tan, D.: InkSeine: In Situ search for active note taking. In: Proc. CHI 2007, pp. 251–260. ACM (2007)
9. Lam, H., Baudisch, P.: Summary thumbnails: readable overviews for small screen web browsers. In: Proc. CHI 2005, pp. 681–690. ACM (2005)
10. Nakagawa, M., Bilan, Z.: On-line Handwritten Japanese Characters Recognition Using a MRF Model with Parameter Optimization by CRF. In: Proc. ICDAR 2011, pp. 603–607. IEEE (2011)
11. Suh, B., Ling, H., Bederson, B.B., Jacobs, D.W.: Automatic thumbnail cropping and its effectiveness. In: Proc. UIST 2003, pp. 95–104. ACM (2003)
12. Teevan, J., Cutrell, E., Fisher, D., Drucker, S.M., Ramos, G., André, P., Hu, C.: Visual snippets: summarizing web pages for search and revisitation. In: Proc. CHI 2009, pp. 2023–2032. ACM (2009)
13. Woodruff, A., Faulring, A., Rosenholtz, R., Morrison, J., Pirolli, P.: Using thumbnails to search the Web. In: Proc. CHI 2001, pp. 198–205. ACM (2001)