マルチメディアサムネール：小画面でドキュメントを閲覧する方法

Multimedia Thumbnails：A New Way to Browse Documents on Small Display Devices

バーナ　エロル*　　カトリン　バークナー*
Berna EROL　　Kathrin BERKNER

要　　旨

アニメーションの形でスタティックなドキュメントの概要を見せる表現方法「マルチメディアサムネール」を紹介する．この新しいドキュメント表示方法のターゲットとなるのは，PDAや携帯電話，MFP等，表示画面が小さく，操作手段も限られるデバイスである．この「マルチメディアサムネール」技術では，ドキュメントから情報を抽出してアニメーションの形で再構成し，ページを切り替えたりズームした図やタイトルの部分を強調して見せたりすることが出来る．また，これらの視覚効果に加え，キーワードや各章のヘッドラインや図の説明などを同時に音声で出力することが可能である．ここでは，ドキュメントから情報を抽出するアルゴリズムやその情報を表示等で制約のあるデバイスに対して最適に再構成する方法、及び最終的な「マルチメディアサムネール」の合成方法について述べる．また，実際にユーザーを観察・調査した結果についても述べる．

※　実際に作成したアニメーションファイルを巻末のCD-ROM及びWebサイト上に添付した（movie1/movie2）。

ABSTRACT

We introduce MultiMedia Thumbnails, a new document representation that provides users an automated navigation – a guided tour – through a document. The target application for this new style of document browsing is devices with small displays and limited navigational controls, such as PDAs, cellular phones, or MFP display panels. MultiMedia Thumbnail technology extracts information from a document and transforms it to an animation that uses effects such as page flipping, zooming into a figure caption, or panning over a title. In addition to these visual effects, the audio channel of the output device speaks keywords, section headings, and figure captions. We describe the information extraction algorithm, optimization of information given the constraints of the browsing device, and synthesis of the final MultiMedia Thumbnail. We also present results from an observational user study.

＊We attached animation files of MultiMedia Thumnail in an attached CD-ROM and on the web-site of Ricoh Technical Reports(movie1/movie2).

* California Research Center, Ricoh Innovations, Inc.

1. Introduction

Devices with small displays, such as MFPs, PDAs, cellular phones, and digital cameras are increasingly being used to access documents, web pages, and images in a business environment. Browsing and viewing of documents on such devices, however, is still very difficult.

Fig.1 Today, viewing documents on small displays is very difficult (a), Multimedia Thumbnails solves this problem by using both visual animation and audio (b).

Currently this problem has limited solutions. For example, often web pages are designed for small displays. In digital cameras, the problem of browsing photos is usually solved by simply showing a low resolution version of photos and expecting the user to zoom into the picture for more details. Document viewers on PDAs employ a similar method, where the user zooms into the document and scrolls to see details. These solutions require interaction (zoom in, pan, etc.) with a device like a cellular phone that has limited input peripherals. Automatic re-flowing of text in documents and web pages is suggested by some researchers as a solution to fit them in small displays [1][2]. Moreover, an automatic navigation algorithm for photos is presented in [3][4]. However, these solutions either do not support multipage document images, or require changing the layout and appearance of the document.

In this paper, we present a new document representation called *Multimedia Thumbnails (MMNail)* that is suitable for viewing documents on small displays. An MMNail can be seen as a guided tour of a document. We animate the document pages, zoom into and pan over the most important visual elements, such as title and figures, automatically. This way, we utilize both spatial and time dimensions for presenting the documents. Moreover, the audio channel is used to communicate textual information. While document contents are shown in the visual channel, the audio channel is used to speak important keywords, figure captions, etc.

As a result, MMNails utilize both the visual and audio channel of the browsing device in order to present an overview of the document on a limited display and in a limited time-frame, while optimizing the interaction required by the user.

An example of an MMNail for a 2 page document is shown in Fig.2. In this example, the MMNail representation shows the first

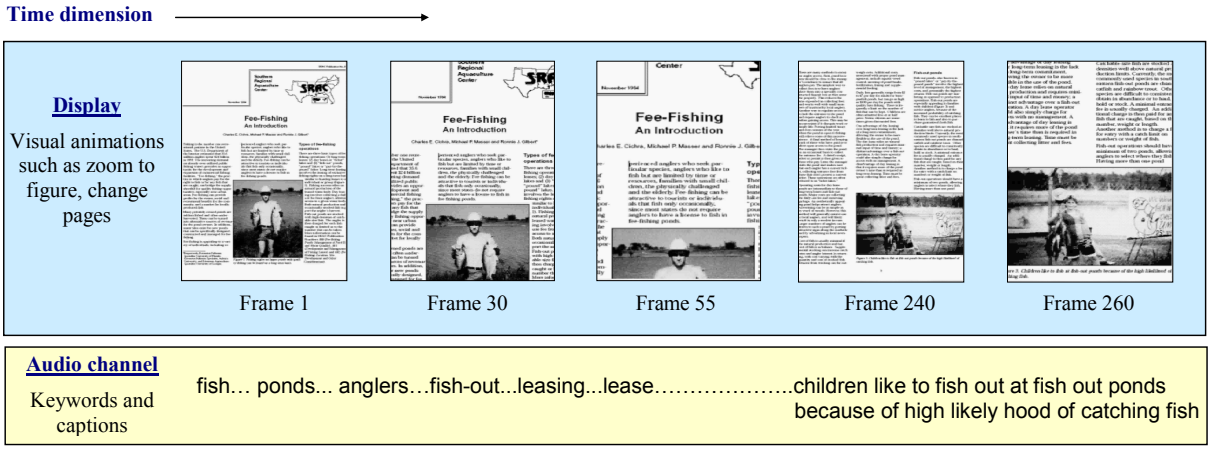

Fig.2 Multimedia Thumbnail Example

page, then automatically zooms into the title, shows the second page, then automatically zooms into the figure. The audio channel, on the other hand, first communicates the important keywords from the document and then reads out the figure captions that are too small to read on the screen.

2. Core Technology

Multimedia Thumbnails are created from electronic or scanned documents by first analyzing the document's contents to identify important visual and audible information, then optimizing the presentation and order of this information given the display and time constraints. Last, the visual and audio data is synthesized to generate a playable MMNail. This process is shown in Fig.3. In the next sections, we explain each of these processing steps in detail.

Analysis

A document and a meta data file are input of the analysis step. Currently, the system accepts PDF files as input. The associated meta data file may include extra information about a document, such as author information, creation date, etc., in the XML format. First, a preprocessing step is applied to the PDF file, which includes OCR and layout analysis. The output of the preprocessing is further analyzed to determine the visually significant features, such as title, figures, tables, graphs, and their exact locations on each page. Also, attributes such as resolution and importance are associated with this data. Besides visual information, the analysis step also determines audible document information from the document image and meta data. Examples of audible information include figure captions, keywords, authors' names, publication name, etc., that can be converted to synthesized speech. We compute the keywords of a document with a TF-IDF algorithm [5]. Attributes, such as importance and duration, are also attached to the audible information.

The duration attribute is computed by multiplying the number of characters with a constant value, which gives the duration after synthesizing the words into speech. The visual and audible importance attributes are assigned based on user studies we conducted, which are explained in more detail in Section 4.

Optimization

The optimizer receives the output of the analysis step, a characterization of the visual and audible document information, and device characteristics such as display size and available time span, and computes the optimal combination of visual and audible information. We implemented a hierarchical optimizer that works as follows. First, given the time constraint, we compute how

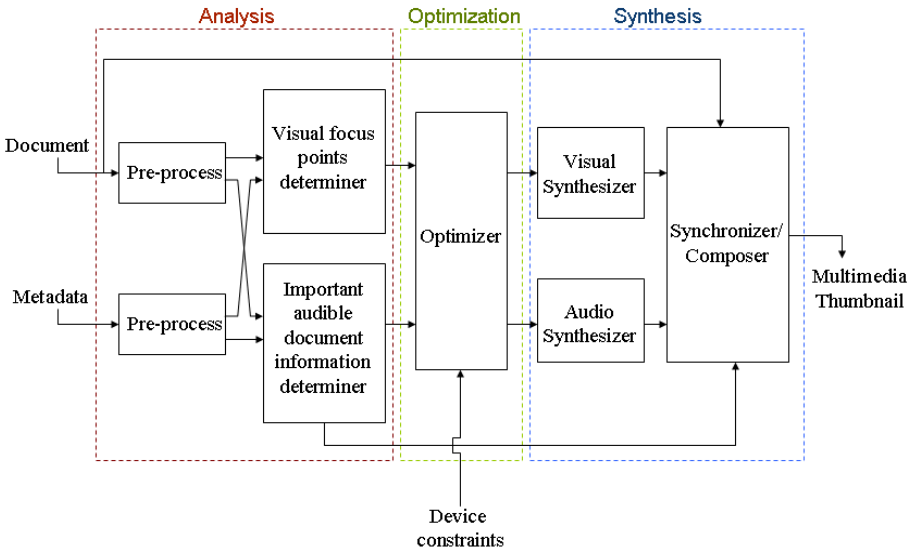

Fig.3 Processing steps for generating Multimedia Thumbnails

many pages can be shown to the user. If there is any time left, the optimizer allocates time for zooming into the title of the document. Because the visual channel is occupied during this time, the audio channel is used to deliver keywords. The number of keywords to be included is determined by sorting the keywords based on their importance attributes and performing linear packing considering their duration attributes. If there is any time left after page flipping and title zooming, the optimizer sorts the captions of the figures based on their duration attributes. As many figures as possible are selected from this list to fill the remaining time. These figures are zoomed in, and during the occupation of the visual channel, the synthesized caption is "spoken" in the audio channel. After optimization, the optimizer orders the selected visual and audio segments with respect to the reading order.

Synthesis

The synthesizer composes the Multimedia Thumbnail by executing the multimedia processing steps determined in the optimizer. These steps include visual animations such as page flipping, pan, and zoom to certain locations on a page, and speech synthesis. Visual animations are implemented in Flash using ActionScript 2.0. Speech synthesis is implemented using the AT&T Natural Voices Text-to-Speech SDK [6]. After obtaining visual and audio streams, synchronization is performed using Action Script to obtain a playable MMNail.

3. User Interface

Fig.4 Flash interface – document browsing

A document browser interface that displays the thumbnail of the first page of each document is shown in Fig.4. The interface is implemented in Flash 6.0, and is compatible with Windows and Macintosh operating systems and PDAs running the Pocket PC OS.

Fig.5 Flash interface – document viewing

When a user selects a document thumbnail in order to view the MMNail representation, automated navigation is activated in the interface given in Fig.5. The user has control over playback with the "control bar", which he can use to start, stop, go backward and forward in the MMNail timeline.

4. Observational User Study

Document browsing behavior depend on the task at hand. For example, a user looking for a document written by a specific author will most likely never browse past the first page, whereas a user looking for a specific figure may browse the entire document. We performed a user study in order to better understand how people browse documents given specific browsing tasks. Ten paper documents were shown to the users one week in advance. Then, the following two tasks are given to them:

1. Search task: Shown several documents, identify the previously seen documents.
2. Understanding task: Given a limited time, try to understand a new document in order to answer some questions about its contents.

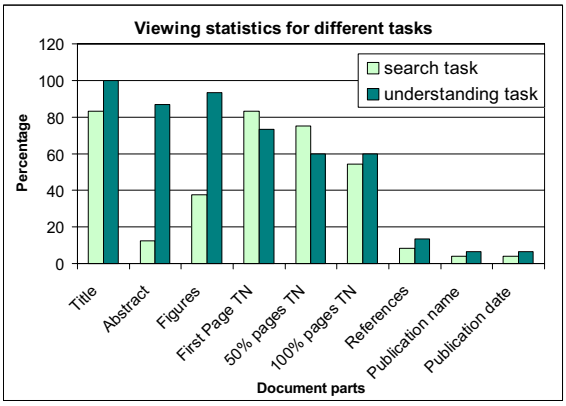

Fig.6 Percentage of users who viewed different parts of the documents for search and understanding tasks.

In total, nine users participated to the study and 6 documents were shown to each of them for the above two tasks. The users browsed the documents on a small (PDA-size) display with limited manual navigation. The users' navigation behaviors were recorded and analyzed in order to understand which document parts they viewed during browsing.

Fig.6 shows the percentage of users who viewed various document parts. As can be seen from the figure, viewing "abstracts" and "figures" of documents is task dependent. As expected, very few users read the abstract of a document for the search task, but the majority of the users read the abstract for the understanding task.

Table 1 Questionnaire results regarding to the importance of different parts of the document for the search and understanding tasks.

Doc Part	Search Task Average Score	Understanding Task Average Score
Title	9.4	10
Figures	9.4	8.8
Abstract	7.5	10
Figure captions	3.1	7.5
Thumbnail of first page	6.9	5.6
Thumbnail of %50 pages	6.3	5
Publication name	5.7	6.1
Publication date	5	6.1
Author names	5	6.7
Thumbnail %100 pages	3.1	4.3
References	1.8	5

The participants were also asked in a questionnaire to assign importance scores (between 1 to 10) for viewing various document parts for the search and understanding tasks. The results are presented in Table 1. Users generally agreed that the title, figures, and abstract were the most important document parts for both tasks. An interesting result was that the figure captions were important for the understanding task (average score: 7.5), but not very important for the search task (average score: 3.1).

Table 2 Questionnaire results regarding to the importance of different audible parts of the document.

Doc Part	Ave
Figure captions	8.9
Title	6.1
Keywords	5
Authors	4.4
Publication name	4.4
Number of pages	4.4
Publication date	3.9

The participants were also given a questionnaire regarding to the importance of hearing different types of document information through the audio channel. User scores are presented in Table 2. Additional user comments pointed out that usefulness of the audio depends on the quality of the synthesized speech. Particularly for very short audio segments, such as keywords, understanding the audio content was considered to be difficult. The questionnaire results of this study were used to determine the visual importance and audible importance of different parts of the document and to design on the hierarchical optimizer which was presented in Section 2.

Table 3 Questionnaire results for MMNails

	Ave	StdDev
Animation	7.2	3.6
Audio	7.1	2.7

The participants were interviewed regarding the usefulness of

the visual animation and the audio in MMNail examples. The results indicate that both elements received a score of 7 out of 10 in terms of usefulness, as shown in Table 3.

From our user study, the overall impression was that people have very different preferences about viewing documents. Some people like the visual overview provided by thumbnails, others want to read the title and parts of the text. Some people read beginnings of paragraphs, some people endings of paragraphs. Therefore, it is important to design an MMNail generator which can be customized for personal preferences as well as for the end application.

5. Future Applications

Fig.7 Multimedia Thumbnail application examples.

Multimedia Thumbnails enables many applications. Using MMNail representations, a user can easily browse and view documents on their handheld devices before printing them. Another possible MFP application is to use an MMNail representation to review scanned document quality. For example, after a user scans a document, MMNail analysis could be performed to find the potentially problematic areas of a document in terms of scan quality, such as text with small fonts and graphics. Then, the MFP console can be used to display these less than ideal areas to the user by automatically zooming in and panning over. The user could then decide whether the quality is acceptable or if another scan is needed with different settings.

References

1) T. M. Breuel, W. C. Janssen, K. Popat, H. S. Baird, “Paper to PDA”, Proceedings of the International Conference on Pattern Recognition, 2002.

2) K. Berkner, E. L. Schwartz, C. Marle, “SmartNails – Image and Display Dependent Thumbnails,” Proceedings of SPIE, vol. 5296, pp. 53-65, San Jose, 2004.

3) M-Y. Wang, X. Xie, W-Y. Ma, H-J. Zhang, “MobiPicture – Browsing Pictures on Mobile Devices,” International Conference of ACM Multimedia, Berkeley, Nov. 2003.

4) X. Xie, H. Liu, S. Goumaz, W. Ma, “Learning User Interest for Image Browsing on Small-form-factor Devices, ” ACM Conference on Human Factors in Computing Systems CHI, 2005.

5) G. Salton, Automatic Text Processing, Addison-Wesley, 1989.

6) AT&T Natural Voices Text to Speech SDK at http://www.naturalvoices.att.com/products/