

# Visual Summarization of Web Pages\*

Binxing Jiao<sup>†‡</sup> Linjun Yang<sup>‡</sup> Jizheng Xu<sup>‡</sup> Feng Wu<sup>‡</sup>

<sup>†</sup>MOE-MS KeyLab of MCC, University of Science and Technology of China, Hefei, 230026, P. R. China

<sup>‡</sup>Microsoft Research Asia, Beijing 100190, P. R. China

bxjiao@gmail.com, {linjuny, jzxu, fengwu}@microsoft.com

## ABSTRACT

Visual summarization is an attractive new scheme to summarize web pages, which can help achieve a more friendly user experience in search and re-finding tasks by allowing users quickly get the idea of what the web page is about and helping users recall the visited web page. In this paper, we perform a careful study on the recently proposed visual summarization approaches, including the thumbnail of the web page snapshot, the internal image in the web page which is representative of the content in the page, and the visual snippet which is a synthesized image based on the internal image, the title, and the logo found in the web page. Moreover, since the internal image based summarization approach hardly works when the representative internal images are unavailable, we propose a new strategy, which retrieves the representative image from the external to summarize the web page. The experimental results suggest that the various summarization approaches have respective advantages on different types of web pages. While internal images and thumbnails can provide a reliable summarization on web pages with dominant images and web pages with simple structure respectively, the external images are regarded as a useful information to complement the internal images and are demonstrated very useful in helping users understanding new web pages. The visual snippet performs well on the re-finding tasks since it incorporates the title and logo which are advantageous on identifying the visited web pages.

## Categories and Subject Descriptors

H.5.m [Information Interfaces and Presentation]: Miscellaneous

## General Terms

Algorithms, Experimentation

\*This work was performed at Microsoft Research Asia.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIGIR'10, July 19–23, 2010, Geneva, Switzerland.

Copyright 2010 ACM 978-1-60558-896-4/10/07 ...\$10.00.

## Keywords

web page summarization, visual summarization

## 1. INTRODUCTION

Search and re-finding tasks are among the most typical applications on the Internet [4]. In order to help users accomplish these two tasks efficiently, systems usually provide a summarization of web page so that users can quickly judge whether the page is of interest. For example, the search engine usually presents the search result using the summarization of web pages, consisting of the URL, page title and a short textual snippets. When people try to re-find the web pages in the browsing history, they are likely to access the bookmarks in the web browser, where visited web pages are summarized by the url and the title. While most systems summarize web pages by text, due to its compactness and simplicity, it is difficult for users to quickly and precisely judge if one web page is of interest from the textual summarization, since it is often time-consuming for users to read a long text snippet or it usually cannot convey enough information if the text snippet is short.

A picture is worth a thousand words. Information search and re-finding tasks would become more efficient if the web pages are *visually* summarized, since it is easier for people to get a quick understanding by seeing an image than reading text. Google has recently released the “Images from the pages” feature [7] in its web search system, which summarizes the web page using the images in the page (we refer to it as *internal images* in this paper). This feature presents two images selected from the web page alongside the textual snippet to users, from which users can quickly identify the desired web pages, even without reading the textual snippets. Visual summarization can also used to enhance the bookmarks in a Web browser to help users easily accomplish the re-finding task. Apple’s web browser Safari 4 [15] has introduced a new feature called “Top Sites”, which visualizes users’ favorite web pages as a gorgeous 3D *thumbnail* wall generated from the downsized snapshot of the web pages.

Although thumbnails [15, 11] and internal images of web pages [7] are adopted by the mainstream companies due to the simplicity, they suffer from several disadvantages. First, for those web pages which have complex layout or rich content, users are difficult to see clearly anything useful from the small thumbnails. Second, although the internal images are proved to be more informative than the thumbnail [12], they are unavailable for a large amount of web pages which do not contain useful images.

To deal with the above mentioned difficulties of internal

image based visual summarization, in this paper we propose to mine images from the external for the visual summarization of web pages without suitable internal images, which is so-called external image based visual summarization. Furthermore, we perform a comparative study to investigate the applicability of the different visual summarization approaches to different kinds of web pages and for different tasks including search and re-finding. The different approaches, including thumbnails, internal images, external images, and a recently proposed new scheme called visual snippet [17], are analyzed based on user study, from which some conclusions are drawn which can provide guidance for the practical application of visual summarization for web pages.

The paper is organized as follows. In section 2, we review related work. Three kinds of technology for generating visual summarizations will be introduced in section 3, 4 and 5. Experimental results and user study are reported in Section 6 and 7. At last, we conclude this paper and point our future work in Section 8.

## 2. RELATED WORK

Thumbnail [18, 15, 6, 3, 5, 19, 20] and internal image [7, 12, 17] based visual summarizations are widely studied in research communities and commonly adopted in existing products.

Thumbnail is a scaled-down image which displays a snapshot of a web page as rendered in the web browser. Due to its simplicity, it is commonly adopted in existing products [18, 15, 6]. Viewzi [18] presents users the search result using not only the text snippets, but also the thumbnail of the web page, for web search. Firefox “FastDial” [6], like Safari 4 [15], visualizes bookmarks as thumbnails. Alongside the widely adoption of visual summarization of web pages in industry, many research efforts have been invested on this problem in the academia as well. Some studies are focused on the design of thumbnail. Cockburn et al. [3] generate thumbnails that consist of reduced images plus graphical icons that indicate bookmarked and frequently visited pages. Dziadosz and Chandrasekar [5] suggest that thumbnails of web pages used along with text snippets in search engine interface can help users in reducing predicting errors, at little time cost in processing time. Woodruff et al. [19, 20] design textually enhanced thumbnails, which enhance the readability of certain parts of the document within the thumbnail and display highlighted keywords transparently overlaid on the reduced document. As a result, the enhanced thumbnail takes advantages of both textual summaries and raw thumbnails. However, in the above applications, raw thumbnail has disadvantages that for those web pages which have rich content, users are difficult to see clearly anything useful from the small thumbnail. Though enhanced thumbnails [3, 19] address this problem to some extent, they are task specific because they only focus on elements relating to specific tasks (e.g., query term in the case of Woodruff et al. [19, 20] or visitation data in the case of Cockburn et al. [3]).

In order to address the problems of thumbnail based summarization, internal image based summarizations are proposed as an alternative approach since internal images usually can describe the content of the hosting web page. In order to improve the relevance judgment of web search results, Li et al. [12] proposes an internal image based scheme which extracts the *dominant images* from the web page as

“image excerpt”. The *dominant images* are first detected by a trained model based on three levels of image features. And then the most relevant one is selected based on the relationship between the query and the image’s surrounding text. Visual Snippet [17], which is an image generated by composing a *dominant image* from the web page, a watermarked logo, and salient text (e.g. title) from the web page, is an extension of “image excerpts”. These two *dominant image* based approaches are proved to be more informative for search compared to the above thumbnail based approaches. Internal image based summarizations are also adopted by industrial products, such as Google’s “Images from the pages” [7]. However, as aforementioned, for a large amount of web pages, useful internal images are unavailable, which limits the applicability of such approaches.

Some studies have involved how different summarizations perform in search and re-finding tasks. Woodruff et al. [20] conduct a study to examine the use of text summaries, raw thumbnails, as well as the enhanced thumbnails in a web search task and find that for some questions, text summaries outperform raw thumbnails while for other questions, raw thumbnails outperform text summaries. Enhanced thumbnails have better and more consistent performance than either text summaries or raw thumbnails. Teevan et al. [17] explore how text summaries, raw thumbnails and visual snippet perform in both search and re-finding tasks. In this paper, we explore not only how different summarizations perform in search and re-finding tasks, but also how different summarizations perform for different kinds of web pages.

## 3. INTERNAL IMAGE BASED VISUAL SUMMARIZATION

The images contained in a web page are usually used to describe the content of the page, which are so-called internal images and can be directly used to summarize the web page. However, a typical web page may contain a lot of images, most of which are advertisement images, decoration images and logos. Hence we need to detect the most dominant among all the images in a web page for visual summarization. In this paper, we adopt the learning-based algorithm proposed in [21] for dominant image detection, which will be described as follows.

First, we need to extract features for all the images based on the property of the image itself, the relationship of the image to the hosting web page, and the site information, as summarized in Table 1. The details of the features can be found in [21]. Then an image dominance detection model can be learned from some labeled training samples, which are represented as  $(x_{i,j}, y_{i,j})$ , where  $x_{i,j}$  is the extracted feature vector of the image  $i$  in the page  $j$  and  $y_{i,j}$  is its labeled dominance. Each image for training is assigned an important level, namely 0 (useless), 1 (important) and 2 (highly important). We regard the dominance detection problem as a ranking problem since there are multiple dominance levels. Then we use linear Ranking-SVM [9] to train the ranking model for dominance detection.<sup>1</sup>

## 4. VISUAL SNIPPET

Visual Snippet [17] is a recently proposed web page summarization approach which enriches internal dominant im-

<sup>1</sup>We directly use the trained model in [21]. Thanks to the authors for providing us the valuable model.

**Table 1: Features for dominant image detection.**

Feature kind	Feature name
Image Level	size, width/height ratio, blurriness, contrast, colorfulness, photo vs. graphic
Page Level	relative pos, relative size, relative width/height ratio
Site Level	IsInnerSite

ages by synthesizing the salient text (e.g. title), dominant image, and logo into a single image. It’s designed to capture the respective advantages of textual snippets and internal images.

The visual snippet generation algorithm consists of four steps, which are described as follows.

1. **Internal dominant image selection and processing:** The dominant image is detected by the algorithm shown in Section 3. Then the detected dominant image is auto-cropped to extract the salient area using the algorithm proposed in [13].
2. **Logo extraction:** Logo images are detected by a trained model [14] using features including the image’s location in the page, size, name and surrounding text). Then the detected logo is scaled to fit within a 120x45 rectangle while preserving the aspect ratio.
3. **Title extraction:** The title is extracted from the html source of the web page. Then the first 19 characters of the title are cropped based on the study in [10] that the leftmost 15-20 letters of a page’s title yield reasonable recognition of the page’s site.
4. **Image Composition:** The logo is made semi-transparent and overlaid on top of the dominant image, and the cropped title is placed above the image and the logo with a white background for readability.

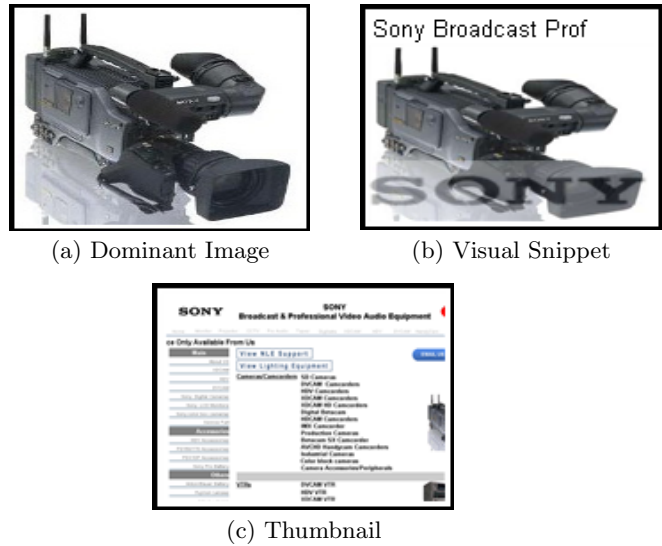
Figure 1 (a) and (b) show an example<sup>2</sup> of the dominant image selected by the algorithm in Section 3 and visual snippet generated by the algorithm as aforementioned.

## 5. EXTERNAL IMAGE BASED VISUAL SUMMARIZATION

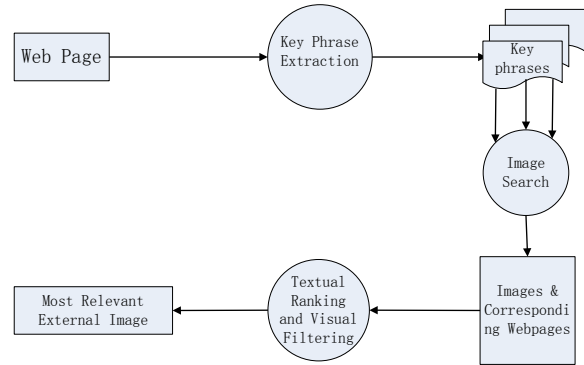
The aforementioned summarization methods including internal image and visual snippet both require that there are dominant images in the web page. However, for a large amount of web pages, the dominant images which can represent the information of the web page don’t exist. How to generate a meaningful visual summarization for such web pages is a new challenge to the web page summarization problem.

The intuitive idea to address this problem is to retrieve images which can represent visually the web page from the whole Internet. It is so-called external image based approach compared with the internal images discussed before. The idea can be briefly stated as follows. Firstly the key phrases, which can represent the salient topics in the web page, are

<sup>2</sup><http://www.expandore.com/product/sony/Proav/broadcast.htm>



**Figure 1: Dominant Image, Visual Snippet and Thumbnail Sample.**



**Figure 2: The work flow of the proposed system. The circle boxes denote operations, and the rectangle boxes denote the results after the operations.**

extracted. Then the key phrases are used as queries to query the image search engine for the images relevant to the topics of the web page. Thirdly the returned image results are reranked based on the textual similarity between the web page containing the result image and the target web page to be summarized as well as the visual similarity among the result images. Finally the top ranked image are used to summarize the target web page.

Figure 2 illustrates the work flow of our approach.

### 5.1 Key Phrase Extraction

In this paper we employ the KEX algorithm proposed in [2] to extract the key phrases from web pages.

In KEX, initial term sequences are first extracted by splitting the texts up according to phrase boundaries such as punctuation marks and dashes. Then candidate key phrases are generated using all subsequences of these initial term sequences up to a given length (4 in [2]). After these candidate key phrases are filtered using query logs, two kinds of features, including 3 structured features which are related

to the structure of HTML file and 3 text features which are related to text content, are calculated. Finally given the features and user labeled salience score, a model is trained by logistic regression. More details can be seen in [2].

In our experiment, after we calculate the salience score of each candidate phrase, the top 4 key phrases with the highest scores are used as queries to the image search engine. For each key phrase, 30 results including images and corresponding web pages are retrieved for further processing.

## 5.2 Image Ranking and Filtering

We develop algorithm to rerank and filter images in the key phrase based search result based on two assumptions. First, good candidate images for visual summarization are likely to appear in web pages textually similar to the target page; Second, an image is more important if more images are visually similar to it. Thereafter, in our algorithm, we firstly rank result images based on the textual similarities between the web pages containing the images and the target web page. Then we propose an algorithm to filter out visually unimportant images.

Textual similarity between two web pages is calculated by cosine similarity based on vector space model(VSM). First, TFIDF (Term Frequency Inverse Document Frequency) score is calculated for each term in the web page. a web page is represented as a vector where each component is a TFIDF score of a particular term in the page. Finally cosine similarity [16] is adopted to calculate textual similarity between each result web page and the target web page.

Visual importance of result images is calculated using VisualRank [8]. For each image, we firstly extract local SIFT features which are a 128-dimensional vector. Then the similarity between two images is defined as the number of local features shared between them divided by the average number of local features in the two images. Finally, a graph is constructed with images as vertices and visual similarities as weights on edges. PageRank is applied on the similarity graph and then the resulted rank score vector is employed to represent the importance of each image. The images whose importance score is below a threshold will be filtered out.

Consequentially the relevance of the image to the target web page can be defined as

$$Sim(i, TW) = \begin{cases} TI(CW_i, TW), & \text{if } VRScore > Thresh \\ 0, & \text{otherwise.} \end{cases}$$

where  $CW_i$  is the  $i_{th}$  candidate web page,  $TW$  is the target web page and  $TI(CW_i, TW)$  is the TFIDF cosine similarity between  $CW_i$  and  $TW$ .  $VRScore$  is the importance computed by VisualRank.  $Thresh$  is the threshold which is set to the average of  $VRScore$  of all of the candidate images in our experiment.

## 6. EXPERIMENTAL RESULTS

We have conducted numerous experiments to evaluate the performance of various visual summarization approaches. First we introduce the dataset and our experimental setup. Then we evaluate the usefulness of external images of our system based on the labeled data in our dataset. Finally we discuss how various visual summaries perform for different kinds of web pages.

## 6.1 Experiment Setup

In order to evaluate our system, firstly we randomly selected 100 queries from the query set released by KDD CUP 2005, covering different queries including “Computers”, “Entertainment”, “Information”, “Living”, “Online Community”, “Shopping” and “Sports”. Then we used the Bing Search API [1] to retrieve the top 100 search results for each of the queries. Then the web pages and the included images were crawled for the top 100 results for each of the 100 queries. Finally, our dataset including 8412 retrieved web pages was obtained.

For each web page in our dataset, we downloaded all of the internal images and retrieved external images using the aforementioned algorithm. After all of these images were obtained, we recruited expert labelers to annotate whether the internal images were dominant and the external images were relevant to summarize the corresponding web pages. In the remaining of this paper, the web pages with at least one internal image annotated as dominant are denoted as “DIP”. Those without any dominant internal image annotated are denoted as “NDIP”. In our dataset, we notice that about 41.8% of the web pages are NDIPs, which demonstrates that internal dominant images are unavailable for a large amount of web pages.

Moreover, in order to compare the different summarization approaches, we generated the four summaries of nearly the same size so that the result is not affected by the different sizes. The settings to generate the four summaries are listed as follows:

1. **Thumbnail** (the size is of 160x120). As a tradeoff between the amount of information and the ability to be recognized, a thumbnail is created first by clipping a snapshot area of 640x480 from the web page and then resizing to 160x120.
2. **Internal Image** (the size is of about 160x120 while keeping the aspect ratio). An internal image with the highest dominance score is selected, no matter if it is a dominant image or not.
3. **Visual Snippet** (the size is of 160x120).
4. **External Image** (the size is of about 160x120 while keeping the aspect ratio).

After these summaries were generated, we asked the labelers to judge which summary, including the internal image, the external image, the thumbnail, and the visual snippet, is the best to represent the web page. If the labeler thinks all of them cannot summarize the page, he/she can select “others”.

## 6.2 Usefulness of external images

Table 2 shows the proportion of web pages which have at least one external images annotated as relevant in the entire dataset, DIPs and NDIPs, respectively. External images retrieved by our algorithm are labeled as relevant for 53.3% web pages in the entire dataset. Based on this observation, external images can be considered as a useful source for web page summarization.

Moreover, the external images are labeled as relevant for 59.9% of DIPs and 44.0% of NDIPs, respectively. We can see that surprisingly the proportion of web pages for which the

**Table 2: Proportions of web pages with annotated external images.**

	Total	Page with Relevant External Image	Proportion
Entire Dataset	8412	4480	53.3%
DIP	4894	2932	59.9%
NDIP	3518	1548	44.0%

**Table 3: Statistics of the best summarization.**

Best Summarization	NDIP	DIP
Internal	0	44.01%
External	34.51%	14.69%
Thumbnail	36.24%	27.99%
VisualSnippet	2.67%	13.24%
Others	26.58%	0.06%

external images are useful in DIPs is even larger than that in NDIPs. We argue it is because that web pages containing dominant images can be regarded as more appropriate for visual summarization, which, hence, can also be summarized well by a relevant image retrieved from the external. In addition, the external images are useful for 44.0% web pages of NDIPs, which states that the external images are a valuable complement to internal images for visual summarization of web pages.

### 6.3 Comparison of various visual summaries

Based on the labelers’ judges on which summary is the best for each of web pages, we conduct the statistics on the proportion of web pages for which the various summaries are regarded as the best, in DIPs and NDIPs, respectively, as shown in Table 3.

We assume that web pages which have the same kind of best summarizations are likely to have common characteristics. To explore these characteristics, we looked through our labeled data, and we also had an interview with our labelers to know their thoughts about this problem. Finally some conclusions about the summarization problem were drawn from our observation and the labelers’ suggestions.

#### 6.3.1 Dominant Image and Visual Snippet as the best summarization

The dominant image can only be adopted for DIP, and it is voted as the best one for 44.01% of DIP. Besides, visual snippet which is also a dominant image based summarization is voted as the best for 13.24% of DIP. Therefore, dominant image based summarizations are the best summarization for over a half DIPs. Summarizing web pages using dominant images is reasonable because dominant images can present the ideas of the web pages and often they are impressive to attract users’ attention. However, thumbnail and external image based summarization are voted as the best summarization for 27.99% and 14.69% of DIPs respectively. We will explain why these two summarizations outperform dominant image based summarization in the next two sections.

We also notice that for most web pages of DIP, the visual snippet is worse than the dominant image. This observation seems anti-intuition because a visual snippet contains more

information (title and logo) than a single internal image. To explain this observation, we discussed with the labelers and found that some weakness slightly let down the appropriateness of visual snippet. Specifically, the title information sometimes doesn’t describe the content of the dominant image and this would mislead the users; more importantly, visual snippet is composed of three kind of components, which seems a little rough-and-tumble. Therefore, though visual snippet is promising for representing web pages, current implementation limits its applicability.

#### 6.3.2 Thumbnail as the best summarization

We explain some characteristics of those web pages which are best summarized by thumbnails. By observing the web pages which can be best represented by thumbnails, we categorize these pages into the following categories including “pages in small size”, “pages with several images in snapshot area”, “pages with salient and clear image/text in snapshot area”, “pages with logo in snapshot area”, “pages with simple page structure”, and “pages in well-known web site”. Among those pages for which thumbnails are not good summaries, most of them contain a large amount of text. Since texts are usually unable to be seen clearly in a thumbnail with small size, it is difficult for users to recognize or recall these pages containing many text in the body.

For 27.99% of DIP, the thumbnail based summarization outperforms all of the other summarizations. Based on our observations, users tend to choose thumbnail as the best summarization for “pages with salient and clear image/text in snapshot area”. Since dominant images are contained and can usually be recognized in the thumbnail, users tend to choose thumbnails as better summarizations because thumbnail can contain more information (e.g. structure of web page) than a single dominant image.

For 36.24% of NDIP, the thumbnail based summarization outperforms all of the other summarizations. This is partly because sometimes external images are not that relevant to the target web page. On the other hand, for NDIP, thumbnail based summarization is preferential for “pages with logo in snapshot area” and “pages in well-known web site”.

#### 6.3.3 External Image as the best summarization

For 34.51% of NDIP, the external images are voted as the best summarization. For all of these web pages, external images are annotated as relevant by labelers. However, 50.3% of external images are not relevant to web pages which are best summarized by other summarizations. Therefore, the relevance of external image is the crucial factor to select external image based summarization for NDIPs. However, though useful, external images often suffer from reliability problem, i.e., external images retrieved from the web are not always that relevant to the target web pages, which is demonstrated in Section 6.2. For NDIPs, external images with tremendous relevance scores can be adopted as reliable summarizations, otherwise, thumbnails are better, especially for “pages with logo in snapshot area” and “pages in well-known web site”.

For 14.69% of DIP, the external images outperform the internal images. This is because for some web pages about popular topics, there are relevant images across the web. This suggests that external images, as well as internal images, should be uniquely taken into consideration when we

calculate the dominance of image based summarizations in our future work.

### 6.3.4 Summary

As a conclusion, after the comparison of web pages which are best represented by different visual summarizations, we propose several guidance of selecting summarization type here.

For NDIP, external images are useful summarizations, but they also suffer from reliability problem. Thumbnail based summarizations are also good for NDIP, especially for “pages with logo in snapshot area” and “pages in well-known web site”.

For DIP, if the dominant images are contained and can be recognized in the thumbnail, thumbnail is the best choice for summarization; otherwise dominant images are more reliable. Besides, for a web page of DIP, if an exclusive dominant image is found and the title information agrees with the content of the dominant image, visual snippet tends to outperform a single dominant image. Actually, an automatic learning algorithm for selecting the best summarization is our future work.

## 7. USER STUDY

We have conducted a user study to compare how participants use different summarizations to understand and re-find web pages. The goal of this study is to explore how different summarization approaches (thumbnail, internal image, external image and visual snippet) support a *understanding* or a *re-finding* task.

### 7.1 Set-up

51 participants were invited to conduct our user study, all of whom have at least an Bachelor’s degree and range in age from 20 to 35 years old. Among them, 56% participants are male. Besides, the participants are all heavy web users, covering a large range of careers, such as computer engineers, researchers, designers, and students.

In this study, the participants were asked to complete a two-phase study. The first phase involves *understanding* the contents of web pages that have never been visited by the participants. The second phase involves *re-finding* the web pages that have been visited during the first phase.

The 11 web pages in this study are chosen from a range of different categories, for example amazon for online sales, wikipedia for information provider. 2 of the 11 web pages are NDIPs. A simple descriptive name extracted from the url for each web page is listed below the x axis in figure 3 and 4.

#### 7.1.1 Phase 1: Understanding

In this phase, the participants were asked to perform 11 tasks about the understanding of web pages. For each task, each participant was given a web page and one of the four visual summaries corresponding to the web page. The participants were asked to strictly perform the following two steps. First they guessed the content of the web page based solely on the provided visual summarization without reading the web page. The second step was to read the web page and to rate a score ranging from -1 to 3 based on whether participants have guessed the content of the web page from the visual summary. Specifically, score -1 means that the summarization misleads the participant; score 0 means that the

participant cannot understand the summarization; score 1 means the summarization can help users grasp a rough idea on what the page is about and score 3 means the summarization can help users exactly understand the content of the web page; score 2 is the between of 1 and 3.

We note that the order in which the different summarization methods are provided to users would affect the judges of the subsequent summaries. For example, if a participant sees the internal image first, she may get more from the thumbnail which may include the internal images, so that she would obtain a better understanding of the page from the thumbnail. In order to avoid such a bias, the participants were divided into 4 groups. Each group was provided a single visual summarization type.

#### 7.1.2 Phase 2: Re-finding

In several hours after participants completed phase 1, they were asked to accomplish the second phase of the user study. In this phase, we explored how the various summarization types would affect the recall of the web pages visited in phase 1. Participants were asked to re-find the web page based solely on one visual summary. In this phase, after we specified a web page to be re-found, the participants were asked to select an image, which can remind them of the web page, from an image collection including a single kind of visual summary for all of the web pages.

The participants were also divided into four groups; each of them was exactly the same as that in phase 1. Since a visual snippet is partly based on internal dominant images, we did not supply these two summarization types at the same time. Therefore, two groups were provided visual snippets; the other two groups were provided internal images.

In order to evaluate the performance of each summarization type for this phase, we define the error rate of a summarization type ( $e(s)$ ) as

$$e(s) = \frac{\sum I\{GuessedWebpage \neq DesiredWebpage|s\}}{NumberofParticipants},$$

where  $I\{GuessedWebpage \neq TargetWebpage|s\}$  is an indicator which equals to 1 when the web page guessed by a participant based on a summarization type ( $s$ ) is not the desired web page.

## 7.2 Results and Discussion

### 7.2.1 Understanding Task

For each summarization method, we average the scores by different participants for each of web pages, respectively. The result is shown in Figure 3, from which we can draw the following conclusions.

From the figure, we can draw several conclusions about the performance of different summarizations for this understanding task.

First, none of the summarization types always perform the best for any kind of web pages. For example, internal images perform well in DIP. However, they get a score below 0 for NDIP (e.g. coldwar<sup>3</sup>). Therefore, we analyze the performance of summarizations for DIP and NDIP respectively.

For the 9 web pages which are DIP, internal images achieve an average score of 1.62 while the average scores for the thumbnail, the visual snippet, and external image

<sup>3</sup><http://www.coldwar.org/articles/90s/BorisYeltsin.asp>

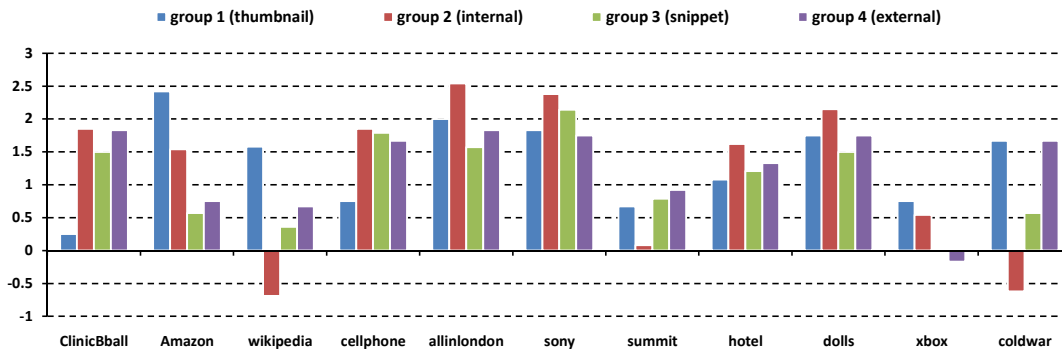


Figure 3: Average scores for all of the web pages in phase 1.

Table 4: Error rate for various summaries.

Summary	Error Rate
Internal Image	0.175
Visual Snippet	0.076
Thumbnail	0.143
External Image	0.152

are 1.27, 1.24 and 1.29, respectively. Moreover, for a large majority of DIPs (6 out of 9), dominant image achieves the best result. For the other three DIPs for which internal images are not the best, based on our interview with the participants we find that it is mainly because the participants do not understand the contents of these images due to the limited expertise of the participants on these topics. In conclusion, for DIP, dominant internal images are regarded as the best summarization.

For the 2 web pages which are NDIP in our study, thumbnail and external image based summarizations achieve relatively high average scores of 1.62 and 1.17 respectively, while the average scores for internal image and visual snippet are -0.65 and 0.46 respectively. The average score for internal image illustrates that the non-dominant internal images selected from NDIP would mislead users (e.g. coldwar in the study).

### 7.2.2 Re-finding Task

We draw the error rates of various summarizations ( $e(s)$ ) for all of the web pages in Figure 4. From this figure, the error rates of some summarizations (e.g. internal image, thumbnail) vary among different web pages, i.e. this re-finding task is also web page dependent. However, we can draw some common conclusions for this specific task.

We average the error rate on all of the web pages for every summarization type and the result is shown in Table 4.

The visual snippet performs better than, or comparably with other summarizations. Especially, the dominant internal images are worse than the visual snippet although the main informative visual part of the visual snippet is the dominant internal image. Thus, the additional two components of the visual snippet: title and logo information are useful for identifying different visited web pages.

The thumbnail based summarization outperforms visual snippet for NDIPs since the visual snippet is generated from non-dominant images in this case. However, it is worse than

visual snippet for most of DIPs, which indicates that although users re-find the web page via visual information, a short but descriptive text (e.g. cropped title) is also important.

The external image based summarization is not suitable for this task. Intuitively, an image which the users do not see in the target web page before will not recall the users of the page. Besides, external images suffer from reliability problem, i.e., retrieved external images are sometimes not that relevant to the web pages. Therefore, external image based summarization is a useful complementary for NDIPs in understanding task, however, it is not suitable for re-finding task, even for NDIPs.

## 8. CONCLUSION AND FUTURE WORK

In this paper, we present a careful study on comparing the various visual summarization methods for web pages, including thumbnails, internal images, and visual snippets. An extension to internal image based visual summarization is proposed, which selects the representative images not from within the web page, but from the external entire Internet.

The experimental results and user studies suggest that the various summarization approaches have respective advantages on different types of web pages and for different tasks. While thumbnails are simple and effective for web pages with simple structure or clear image and text with large size, internal images can summarize the web page well if it includes dominant images. The external images are a useful source for those web pages without dominant images and can be regarded as a valuable complement to internal images. Visual snippets are extremely effective in helping users identifying the visited web pages in the re-finding task.

The above conclusions suggest two future work. First, since the internal and external images have respective advantages, we will develop an integrated framework which does not select the internal images and external images separately, but selects the representative images from a combined pool of internal and external images. Second, we can develop an algorithm to automatically select the most appropriate visual summaries from the four types of candidates based on the type of the web page and the tasks for which the summaries are used, so that the user satisfaction is improved over each of the individuals.

## 9. REFERENCES

- [1] Bing API. <http://www.bing.com/developers>.

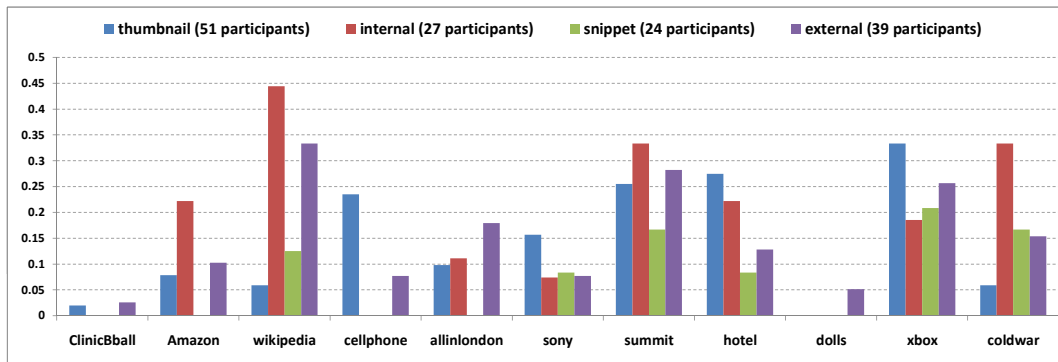


Figure 4: Error rate for all of the web pages in Phase 2.

- [2] M. Chen, J.-T. Sun, H.-J. Zeng, and K.-Y. Lam. A practical system of keyphrase extraction for web pages. In *CIKM '05: Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 277–278, New York, NY, USA, 2005. ACM.
- [3] A. Cockburn, S. Greenberg, B. McKenzie, M. Jasonsmith, and S. Kaasten. Webview: A graphical aid for revisiting web pages, 1999.
- [4] A. Cockburn and B. Mckenzie. What do web users do? an empirical analysis of web use. *International Journal of Human-Computer Studies*, 54:903–922, 2000.
- [5] S. Dziadosz and R. Chandrasekar. Do thumbnail previews help users make better relevance decisions about web search results? In *SIGIR '02: Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 365–366, New York, NY, USA, 2002. ACM.
- [6] FastDial. <https://addons.mozilla.org/en-us/firefox/addon/5721>.
- [7] Google. <http://www.google.com>.
- [8] Y. Jing and S. Baluja. Visualrank: Applying pagerank to large-scale image search. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(11):1877–1890, May 2008.
- [9] T. Joachims. Optimizing search engines using clickthrough data. In *KDD '02: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 133–142, New York, NY, USA, 2002. ACM.
- [10] S. Kaasten, S. Greenberg, and C. Edwards. How people recognize previously seen web pages from titles, urls and thumbnails. In *Proceedings of Human Computer Interaction*, pages 247–265, 2001.
- [11] T. Kopetzky and M. Mühlhäuser. Visual preview for link traversal on the world wide web. In *WWW '99: Proceedings of the eighth international conference on World Wide Web*, pages 1525–1532, New York, NY, USA, 1999.
- [12] Z. Li and L. Zhang. Improving relevance judgment of web search results with image excerpts. In *WWW '08: Proceedings of the 17th international conference on World Wide Web*, pages 21–30, April 2008.
- [13] Y.-F. Ma and H.-J. Zhang. Contrast-based image attention analysis by using fuzzy growing. In *MULTIMEDIA '03: Proceedings of the eleventh ACM international conference on Multimedia*, pages 374–381, New York, NY, USA, 2003. ACM.
- [14] T. Maekawa, T. Hara, and S. Nishio. Image classification for mobile web browsing. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, pages 43–52, New York, NY, USA, 2006. ACM.
- [15] Safari4. <http://www.apple.com/safari>.
- [16] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. *Inf. Process. Manage.*, 24(5):513–523, 1988.
- [17] J. Teevan, E. Cutrell, D. Fisher, S. M. Drucker, G. Ramos, P. André, and C. Hu. Visual snippets: summarizing web pages for search and revisitation. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*, pages 2023–2032, New York, NY, USA, 2009. ACM.
- [18] Viewzi. <http://www.viewzi.com/search/webscreenshot>.
- [19] A. Woodruff, A. Faulring, R. Rosenholtz, J. Morrison, and P. Pirolli. Using thumbnails to search the web. In *CHI '01: Proceedings of the 19th international conference on Human factors in computing systems*, pages 198–205, New York, NY, USA, 2001. ACM.
- [20] A. Woodruff, R. Rosenholtz, J. B. Morrison, A. Faulring, and P. Pirolli. A comparison of the use of text summaries, plain thumbnails, and enhanced thumbnails for web search tasks. *J. Am. Soc. Inf. Sci. Technol.*, 53(2):172–185, 2002.
- [21] Q. Yu, S. Shi, Z. Li, J.-R. Wen, and W.-Y. Ma. Improve ranking by using image information. In *ECIR'07: Proceedings of the 29th European conference on IR research*, pages 645–652, 2007.