Semantic Value Based Web Images Adaptation for Heterogeneous Client Networks*

Gao Yuan
Department of Computer Science and Technology, Tsinghua University
gaoy03@gmail.com

Shi Yuanchun
Department of Computer Science and Technology, Tsinghua University
shiyc@tsinghua.edu.cn

Xiao Xin
Department of Computer Science and Technology, Tsinghua University
xiaoxin00@mails.tsinghua.edu.cn

Abstract*-With the rapid development of computer and network, the Web sites with uniform contents cannot suit different networks and devices. Web content transcoding on proxy is a promising way to make Web documents adaptive to heterogeneous network conditions and browsing terminals, as it not only makes the servers maintain just one or a few versions of the content, but also reduces the data size delivered to clients. In this paper, we mainly focus on the adaptive transcoding of web images under bandwidth constraint, since compared with texts, images have much larger size, whose transcode can significantly reduce the totally delivered bytes. Firstly, the concept "semantic value" is proposed to represent the value of information provided by Web images. Then, image classification method is taken to assign the semantic value to images. Based on the semantic value assigning strategy, to optimally deliver the web contents over heterogeneous networks, an adaptive model is designed to decide proper transcoding way for each image so as to maintain as mush semantic value of transcoded Web Document as possible. Extensive experiments on 100 web pages containing 4201 images have shown the effectiveness and efficiency of our work.*

Keywords: Web, transcode, semantic value, proxy

1. Introduction

A huge number of users are Web-connected in their daily life due to the rapidly growth of Internet and pervasive computing devices. However, the ways users access the Web, both in network types and browsing terminal categories, are considerably different from each other. The Internet infrastructure can be wired such as LANs, ISDN, DSL, cable and telephone modems as well as wireless such as Bluetooth, WiMAX, CDPD and GPRS. They are differed in bandwidth, reliability and cost. The browsing terminals can be traditional PCs, laptops and handheld devices such as PDAs and mobile phones. They are differed in screen resolution, network connectivity and data processing speed.

How to make the data that spreads in the Internet fit various network conditions and browsing terminals has become a popular research subject for a few years. A well known technology to solve this problem is called content adaptation or transcoding[2]. It is applied to convert a multimedia resource from one form to another, with its basic information remained, to match the characteristics of devices and networks. It can be further divided into three categories according to where the transcoding is implemented: server-side, proxy-side and client-side.

In the server-side solutions, transcoding is done offline and the web server has to maintain several versions of the same content. It is difficult to create sufficient versions to fit all the kinds of networks and devices. In the client-side solutions, the client downloads all the contents and then applies the transcoding to the contents. This solution mainly aims to make the display of the web contents fit the screen size of the devices. However, the bandwidth is not saved. Proxy-side [1, 2] is a tradeoff of the two solutions above. The

* This research is funded by National Natural Science Foundation of China (No. 60773139).

server only needs to maintain one or a few versions, and the clients' bandwidths could also be saved. The proxy gets the Web documents from the server, applies online transcoding and delivers the result of transcoding to the client. The main focus in this solution is how to make the transcoding efficiency.

Proxy-side solution is adopted in this paper, and we bring in a concept called "semantic value", which is used to represent the amount of information that an object in a Web document can provide. We make 2 contributions in this paper:

- *Semantic value assigning strategy:* Images in a web document can play different roles, some to present the fact, some to attract viewers to click, some to decorate the document. Different roles of images give them different semantic values. Having been transcoded, images will lose some semantic values. We propose a strategy to assign the semantic value of images in the web document, in order to appropriately represent the amount of information the images provide.
- *Adaptive model:* The model is based on the optimized problem, ensures to deliver the maximum semantic values to clients with the constraints of bandwidth and viewers' patient.

The rest of the paper is organized as follows. We discuss related work in Section 2, define the strategy to assign the semantic value to Web images in Section 3, and present our adaptive models to deliver the Web images in Section 4 and evaluate them in Section 5. Section 6 concludes the paper with a discussion of future work.

2. Related work

By now, the researches on transcoding focus on two aspects: how to build the proxies and how to make the transcoding results cater to the needs of Web viewers.

PTC [1] combines transcoding and caching to realize multi-version Web resources. PTC only takes file size, bandwidth and proxy load into account. It aims at transporting as much data as possible. In its experiment, it only checks JPEG images and MPEG-1 videos, which have quality factors range from 0 to 100. However, researchers in Duke University have found that 74.81% of the unique images are GIF images and 24.41% are JPEG images in WWW [6]. That is to say, GIF images cannot be ignored. Canali et al [2] improve PTC on proxies cooperating in locating, adapting and delivering, but still does not take the differences between different images into account.

AMTM [3] and InfoPyramid [7] adopt the conception called "content value". This is used to describe the amount of information that a media object can provide. But it is only used to describe the differences between different types of media objects. It defines that the original videos have the highest content value, and audios, images have less content value, while texts have the lowest content value. However, differences also exist between media objects of the same type, which is not noted by the concept of "content value".

[4] and [8] study how to classify Web images according to their roles. [4] focus on news Web sites, and defines seven functional categories: story (S), preview (P), host (A), commercial (C), icons and logos (I), headings (H) and formatting (F). These seven categories can be grouped into two super classes: SPA, which is more likely to contain photographic images of regular aspect ratios, and associated with some story, and CIHF, which is more likely to be graphic, have irregular aspect ratios, and often aren't associated with a story.

Our approach is the first to utilize the idea of image classifications to assign the semantic value of images. Meanwhile we propose an adaptive model to deliver as much semantic value as possible to clients. Therefore, Web viewers can receive maximum information although the bandwidth and their patience are both limited.

3. Assignment of semantic value to the image

Nowadays, almost every website tends to use images (including photographs and graphs) to present information coupled with or instead of texts. Semantic value of an image is a measure of information that the image can represent. How to assign the semantic value of each image is what we want to present in this section.

3.1 The semantic value of same images with different presentations

Images with the same content but different attributes surely have different semantic values. Sizes, colors and saturations can influence the semantic values.

Take colors for example, the appropriate use of color can make it easier for users to absorb large amounts of information and differentiate information types and hierarchies. Researches on the effects of color in advertising show that ads using one spot of color are noticed 200% more often than black and white ads, whereas full-color ads produce a 500% increase in interest [5].

Which factor is the most important one? As shown in Figure 1, the image of an actress is used to represent a TV play. If the image is zoomed out 50% in both length and width, we cannot distinguish the person in the image, because a lot of details are lost. Meanwhile, if the image just loses the colors to be a monochrome, people can still recognize who she is, and that is to say people can still get a majority of information. And the saturation is not so important that we do not discuss it in detail. When sorted by importance in descending order, the result is size, color and saturation. It's necessary to point out that a monochrome of one image has larger data size than a zoomed out image of the same image.

Figure 1. **Two transcode methods: monochromic and zooming out.**

Now the point is when a colorful image is transcoded into a monochrome, how much semantic value is lost, and when an image is zoomed out, how much semantic value is lost. We can get this data from the following example.

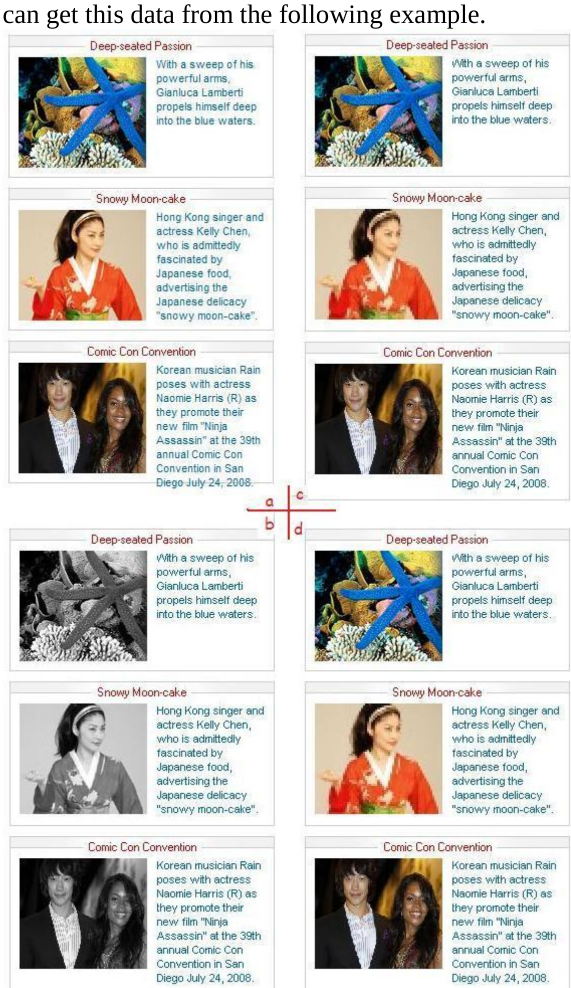

Figure 2. **A screenshot of China Daily.**

Figure 2(a) is a screenshot of http://www.chinadaily.com.cn on Jul. 28th, 2008. Figure 2(b) transcodes all the three images into grey images with the same pixels. Figure 2(c) keeps the first image as well as zooms out the next two images. Figure 2(d) keeps the first and the last images and zooms out the second image. Here, zooming out an image means reducing pixels of the image to a quarter (half in width and half in height). As the html tab can assign the show size of the image, the image will be stretched to the show size.

When the bandwidth is not enough to transport three original images but be sufficient to transport three small colored images, we can have the three transcoding results as shown in Figure 2(b), 2(c) and 2(d).

Supposed the original semantic value of the image is S (because of the same role in the Web document, the three images have the same semantic value), the semantic value of grey image with the same pixels is $a*S$, while colored images with a quarter pixels is $b*S$. Here, $0<a<1, \quad 0<b<1 \quad and \quad a>b$.

Web viewers prefer to see them in the same presentation not only because the three pictures play the same role in the Web document, but also because the small images lose so much information that the viewers cannot see the detail. So, we can conclude that the first result has the maximum semantic value of the three results. That is

$$\left.\begin{array}{l} 3aS > S + 2bS \\ 3aS > 2S + bS \end{array}\right\} \Rightarrow \left\{\begin{array}{l} 3a > 1 + 2b \\ 3a > 2 + 1b \end{array}\right.$$

More generally, if there are n images, the result is $n \bullet aS > kS + (n-k) \bullet bS \quad k = 1,2,\ldots,n-1$

Based on the observation of a lot of Web documents, n is no larger than 4. In our implementation, a is set to be 0.8 and b is set to be 0.2.

3.2 The Semantic Value of Different Images

As mentioned in the above section, the semantic value of an image is related to the role it plays in a web document. An image companied with a paragraph usually represents the truth of the text and has a high semantic value. An image just used to decorate the web site does not provide so much information and has a low semantic value. As introduced in Section 2, images found on news Web pages are categorized into 7 categories and 2 super classes [4]. SPA images are used to provide the associated text with more truths, while CIHF images are used to provide links to other web pages or decorate. So SPA images have higher semantic value than CIHF.

Figure 3. **A screenshot of China Daily**

Figure 3(a) is also a screenshot of http://www.chinadaily.com.cn on Jul. 28th, 2008. The first two images are ads, and the last image is a preview of a piece of news. Figure 3(b) transcodes all the three images into grey images with the same pixels. Figure 3(c) zooms out the two ad images and keeps the preview image. Considering the bandwidth constraint, figure 3(b) and figure 3(c) can both be the transcoding results. The preview image with full color can supply more information, while the zoomed out ads can still tell viewers what they are linked about. So viewers prefer to see figure 3(c) as the result.

Supposed the semantic value of a SPA image is S_1, and the semantic value of a CIHF image is S_2. Figure 3(c) has higher semantic value than 3(b). That is

$$aS_1 + 2aS_2 < S_1 + 2bS_2$$

More generally, if there are n SPA images and m CIHF images, the result is

$$n \bullet aS_1 + m \bullet aS_2 < n \bullet S_1 + m \bullet bS_2$$

Take $a = 0.8$ *and* $b = 0.2$, we can get

$$nS_1 > 3mS_2$$

Based on the observation of a lot of Web documents, m is not more than 3 times as n. So in our implementation, S_1 is set to be 100 and S_2 is set to be 10, in order to make grey images and small images have integral semantic values.

4. Adaptive Model

This section will describe the problem first and then present our adaptive model to solve the problem.

4.1 Problem Formulation

After assigning the semantic value to each image in a Web document, the document gets a total semantic value. Our proxy aims at maintaining as much semantic value as possible under the constraint of bandwidth and viewers' patience.

Assume the bandwidth is B, and the tolerance limitation of viewers is T, then the client can receive totally $B*T$ data. We use S to denote $B*T$. Supposed there are n images in the web document, the proxy's aim can be described as following:

$$\max \quad \sum_{i=1}^{n} V_i$$

$$s.t. \quad \sum_{i=1}^{n} S_i \le S$$

Here, V_i stands for the semantic value of the i[th] image, and S_i stands for the data size of the i[th] image.

We proposed two adaptive methods to solve this problem.

4.2 Greedy Adaptive Method

As described in the last section, SPA images of any presentation types have higher semantic value than CHIF images of any presentation types. So the proxy will choose SPA images first. Consider the constraint of S, the proxy should then choose smallest data size in the same classification first. So a greedy algorithm can be a solution. In our greedy adaptive method, all images in the same Web document will be divided into two classifications according to their original semantic value. Total data size of each classification is calculated. S_1, S_2, S_3 stand for the data size of original images, grey images and zoomed out color images in SPA classification respectively, while S_4, S_5, S_6 stand for the data size of the three presentations of images in CHIF classification. In each classification, the images are sorted according to the data size of each image in ascending order.

Table 1. **SPA images deliver strategy**

$S>S_1$	original images
$S_2 < S < S_1$	grey images
$S_3 < S < S_2$	zoomed out color images

S< S_3	zoomed out color images

S is first compared with S_1, S_2 and S_3 to determine which form of SPA images should be delivered. The strategy is illustrated by Table 1. When S is smaller than S_3, the proxy chooses as many SPA images as possible to deliver.

Supposed SPA images been choosen to deliver are S' in data size. Then we compare S-S' with S_4, S_5, and S_6 to determine which form of CHIF images should be delivered. This strategy is shown by Table 2.

Table 2: **CHIF images deliver strategy**

S-S'>S_4	original images
S_5 <S-S'< S_4	grey images
S_6 <S-S'< S_5	zoomed out color images
S-S'< S_6	zoomed out color images

4.3 Greedy adaptive method modified with area

Based on the observation of a lot of Web documents, generally speaking, there are several SPA images with large area (larger than 500 pixels), as well as many SPA images with small area (smaller than 500 pixels). Large area usually means large data size. However, large SPA images and small SPA images both have the same semantic values. If we make large SPA images give up some bandwidth to small SPA images, we can get more semantic value. This is the main point of greedy adaptive method modified with area.

Supposed that the data size of SPA images with large area is S_7, and the data size of SPA images with small area is S_8, we can get that S_1=S_7+S_8. If S is less than S_1, but larger than S_8, the proxy can deliver grey SPA images with large area and original SPA images with small area. In the greedy adaptive model, the proxy delivers all SPA images of grey form and loses more semantic value.

5. Evaluation

In this section, two adaptive methods described in last section are evaluated. In order to reveal the efficiency of the model, they are compared with random model. In random model, images are not classified, and just handled in the order of being visited. A random number is generated to decide which form of images should be delivered.

We collect 4201 images from 100 Web pages in 4 classifications, which are illustrated as Table 3. We choose www.chinadaily.com.cn to be the representative of the news web sites, and www.sohu.com to be the representative of the composite web sites. As for education web sites, we check the home pages of 25 universities in China. We use home pages of 10 companies such as IBM, HP, Nokia, and Motorola and so on to represent commerce web sites. In news and composite web sites, SPA images have larger data size than CIHF, while SPA images have smaller data size than CIHF in education and commerce web sites.

Table 3. **Distribution of the 4201 images from 100 Web pages**

Classification	Num of websites	Num of images	SPA images(bytes)	CIHF images (bytes)
news	25	866	8980374	2207586
composite	25	1506	3599982	2180998
education	25	1032	3224866	3940308
commerce	25	797	1774732	2736807
total	100	4201	17579954	11065699

The result of the experiment is shown in Figure 4. Both greedy methods perform better than random one, which means our methods work indeed. Due to the character of education and commerce web pages, when the bandwidth is small, greedy method modified with area performs worse than greedy method. That is because SPA images that have large data size are first considered in greedy method modified with area. Even if they are transcoded to small data size, they will still be kept in this method and make SPA images occupy more bandwidth. However, when the bandwidth become larger and larger, greedy method modified with area performs better and better.

6. Conclusions and future work

In this work, to achieve the goal of adapting Web images for heterogeneous client networks, we not only introduce the conception of *semantic value*, which can represent the value of information provided by images, but also present the assignment of semantic value to the images. An adaptive model of proxy is established to optimally deliver the web contents under the constraint of bandwidth. An experiment with 4201 images from 100 Web pages shows the effective and efficiency of our work. This work can be extended to include the relative position of images and user preference.

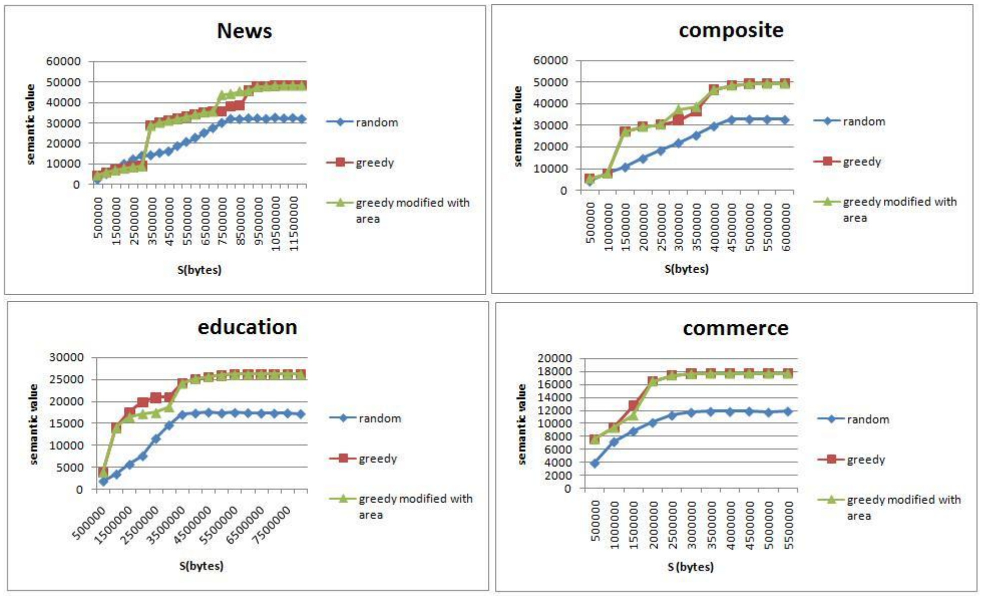

Figure 4. **Experiment result of 4201 images in 4 classifications of Web sites**

References

[1] A. Singh, A. Trivedi, K. Ramamritham, and P. Shenoy. "PTC: Proxies that Transcode and Cache in Heterogeneous Web Client Environments," World Wide Web, vol. 7,no. 1, pp.7-28, Nov, 2004.

[2] C. Canali, V. Cardellini, and R. Lancellotti. "Content adaptation architectures based on squid proxy server," World Wide Web, vol. 9, no. 1, pp. 63-92, Nov, 2005.

[3] C. Liao, Y. Shi, and G. Xu. "AMTM—an Adaptive Multimedia Transport Model," SPIE proceedings series, vol. 4210, pp. 141-149. 2000.

[4] J. Hu, A. Bagga. "Categorizing Images in Web Documents", IEEE Multimedia, vol. 11, no. 1, pp. 22-30, 2004.

[5] Suzanne Watzman. "Visual Design Principles for Usable interfaces", The Human-Computer Interaction Handbook, Lawrence Erlbaum associates, publishers, London, 2003.

[6] S. Chandra, A. Gehani, C. S. Ellis, and A. Vahdat. "Transcoding Characteristics of Web Images", Proceedings of SPIE, vol. 4312, pp. 135-149, 2001.

[7] J. R. Smith, R. Mohan, and C. Li. "Transcoding internet content for heterogeneous client devices", Proceedings of the 1998 IEEE International Symposium on Circuits and Systems, vol. 3, pp. 599-602, May 1998.

[8] T. Maekawa, T. Hara, and S. Nishio "Image classification for mobile web browsing", Proceedings of the 15th international conference on World Wide Web, pp. 43-52, 2006.