Delivering Online Advertisements Inside Images

Zhiwei Li
Microsoft Research Asia
zli@microsoft.com

Lei Zhang
Microsoft Research Asia
leizhang@microsoft.com

Wei-Ying Ma
Microsoft Research Asia
wyma@microsoft.com

ABSTRACT
We present in this paper a new channel to deliver online advertisements along with Web images and show a new business model to monetize billions of Web images. The idea is intuitively inspired by image displaying processes on the Web, which typically require people to wait a few seconds before they see full resolution images. This is due to large file sizes and limited network bandwidth. To utilize idle time and the display area, we propose an innovative method for non-intrusively embedding ads into images in a visually pleasant manner. To maintain a smooth user experience, we utilize the thumbnail of the full-resolution image because it is small and visually similar to the full-resolution image. At the client side, a rendering engine first enlarges and blurs the thumbnail, and then blends the pre-chosen ads information into the enlarged image. Based on this idea, we propose three typical scenarios that can adopt the proposed image-advertising mode. More importantly, we can encourage providers of images or other users to participate in our online image ads service by tagging or annotating images. We envision revenue sharing with the providers participating in our service, and we expect that a large number of users will actively submit, tag and annotate images using the system. We have implemented a prototype image ads system, and conducted a series of experiments and user studies to evaluate such a new advertisement channel. The experimental results and user studies show that the proposed online image ad delivery is a non-intrusive ads mode, and the proposed solution is practical. This work also opens multiple new research directions ranging from multimedia to web data mining.

Categories and Subject Descriptors
D.3.3 [**Programming Languages**]

General Terms
Design, Human Factors

Keywords
Online ads, Targeting ads, Image Ads

1. INTRODUCTION
We have witnessed a fast-growing online advertising market in recent years. The U.S. market for online advertising has increased from $7.3 billion in 2003 to $16.7 billion in 2006, and has grown to $20 billion in 2007 alone [1]. This growing market has fostered the development of many companies, including the Internet search companies.

Two main factors stimulate the fast-growing online advertising market. The first is the explosive growth of the Internet, and the second is the introduction of successful online advertising modes, including banner ads, ads in search results of web search engines (e.g. Google's AdWords) [8] and ads in common web pages (e.g. Google's AdSense) [9]. Vision-based banner ads are mainly used to enhance brand names, corporate images, or new product releases, while text-based AdWords and AdSense are generally suited for transaction-targeted advertisement (i.e. click and buy). Though vision-based ads are generally more attractive than text-based ads, they have to be created manually and less efficiently by designers. Because AdWords and AdSense can effectively connect content providers and advertisers, the two modes significantly increase numbers of online advertisers, and have led to great successes in the so-called long-tail economics.

Motivated by the huge business opportunities in the online advertising market, people now are actively investigating other advertising modes that could be automatically generated by Internet services. For example, researchers have invented context-aware video ads that can be inserted into video clips using intelligent video content analysis techniques [17]. Yet due to current network bandwidth limitations, video is still not widely presented in web pages as compared to the prevalence of text and image.

As one of the most pervasive media formats on the Web, image has unique advantages in that it is attractive compared to text, and can be downloaded and displayed faster than video. As a result, images are used almost as much as text in web pages. However, current image-based ads are mostly comprised of banner ads, which must be manually designed and created. How to automatically deliver image advertisements is still a challenging problem, and has been largely left untouched. Moreover, a common drawback of current online ads modes is that ads are not in attention area (i.e. center) of web pages but at margins of web pages. Thus, even if users open web pages, they may not see the ads. Obviously, ads put at attention area of web pages can definitely attract much attention but may be intrusive.

In this paper, considering above points, we propose an innovative image advertising mode to automatically deliver image ads by an online advertising system. Typically, before a user views a full-resolution image on the Web, a few seconds are required for the image to be downloaded. During the wait, the area reserved for displaying the image is usually blank or partially filled by progressively downloaded image. Accordingly, an information channel and time are wasted. The proposed new advertising mode will utilize the wasted display area and waiting time to deliver image ads in what we believe is a non-intrusive and visually pleasant manner. In wideband networks, the downloading process may be finished immediately, but website owners can choose to hold images for a few seconds, and use display areas of images to show some targeting ads. Thus, this idea of delivering online ads inside images can be generally applied to the whole Internet.

The following sections will present a detailed description of the basic ideas, typical scenarios, system architecture and key techniques of our innovative advertising mode.

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, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
MM'08, October 26-31, 2008, Vancouver, British Columbia, Canada.
Copyright 2008 ACM 978-1-60558-303-7/08/10...$5.00.

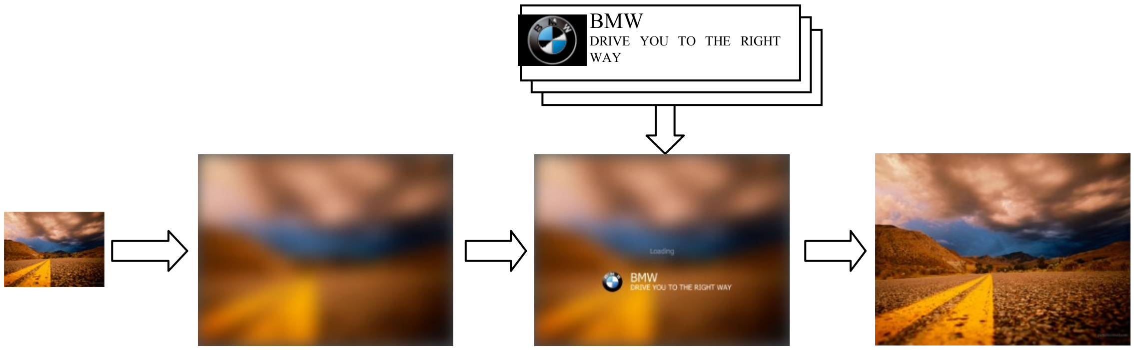

Figure 1. The client-side image ad rendering process

2. BASIC IDEA

This idea is intuitively inspired by image display processes on the Web, which typically require users to wait before viewing full-resolution versions of the image. This is due to large file sizes and limited network bandwidth. To utilize idle time and the display area, we propose an innovative method for non-intrusively embedding ads into images in a visually pleasant manner. To maintain a smooth user experience, we utilize a thumbnail of the full-resolution image because it is small and visually similar to the full-resolution image.

The basic idea is to display a pre-composed image ad in the same position before the full-resolution image is downloaded. To be effective and efficient, several requirements have to be satisfied.

1) The pre-generated image ads should be small in file size so that the image ads can be displayed to users before the full-resolution image is downloaded.

2) The pre-generated image ads should be visually similar to the host image so that the image ads can be displayed in a non-intrusive and visually pleasant manner.

3) The embedded ads should be relevant to the host image based on automated image classification results or user-generated annotations.

4) To be scalable, the embedded ads should be automatically selected from a large scale ads database, based on the relevance measure between ads and images. Therefore, the advertising system can involve many possible advertisers, and may duplicate the success of existing text based online ads.

To satisfy the first two requirements, we utilize a thumbnail of the full resolution image because it is small and visually similar to the full-resolution image. At the client side, a rendering engine first enlarges and blurs the thumbnail, and then blends the chosen ad information into the enlarged image. Once the full-resolution image is downloaded, the image ads will fade out and the full-resolution image will fade in. When we display an ad inside an image, it is critical to maintain the expected viewing experience when the full resolution of the image is downloaded and shown. Because of the visual consistencies of image layout and web page layouts, users are expected to experience a smooth browsing process. The rendering process is illustrated in Figure 1.

According to our current design, the types of ads information can be blended into images include:

1) Descriptive words easy to overlay on the host image. The content of text ads should match the host image content.

2) Dynamically composed ads generated based on a set of information specified by advertisers, including words, logos, or other media content, targeted audience, etc. The actual rendering of ads content will be dynamically and intelligently composed by our engine based on the host image content and layout, and the context of viewers. It is worth noting that it may lead to an "online picture advertising format" which we can later standardize.

3) Existing picture/poster-based advertisements, which can be used to replace the enlarged thumbnails with similar colors or layouts.

In case the image download time is very short, after the embedded ads fade out, we also offer a method to allow users to see the ads again if they wish. We further allow users to click on links or activate buttons in the ads if the user is interested in going to the advertiser's web site. Meanwhile, transactions can be conducted if users click on the provided links.

Though the ideas look simple, to the best of our knowledge, this is the first meaningful attempt to bring many technologies together to deliver online ads inside web images. Here we list some key technologies that will be built to enable seamless morphing, fade-in and fade-out of ad content based on color layout of images.

1) Cleverly choosing regions inside an image to show the above three types of ads. The goal is to minimize the feeling of visual interruption and disconnection, and make it visually less-intrusive and pleasant. Simple image analysis such as color, layout, and attention region detection may be sufficient to provide required knowledge for choosing ads.

2) Automatically morphing, editing and merging images to generate poster-like image ads.

3) Content-based analysis to understand images as much as possible. Typical research topics are image classification and image annotation.

We already have preliminary solutions to these problems, but there are lots of spaces for further study. Although, we have already developed automatic techniques to implement our idea, we highly encourage providers of images or other users to participate in our online image ad platform by tagging or annotating pictures. This is essentially a Web 2.0 concept, that is, letting content providers and users do the job for advertisers, and as a return they will share part of revenues. As revenues can be shared with those participating in our services, a large number of users are expected to be actively tagging or annotating pictures in our system. User-generated tagging or annotation will greatly increase the potential value of images in online advertising. Though the annotation is mainly focused to facilitate ads, the results may also be useful for improving image retrieval, as a by-product.

3. TYPICAL SCENARIOS

To turn the ideas into reality, three novel and representative scenarios for delivering online ads into images are proposed and presented in this section.

3.1 Image Ads in Search Engines

In an image search engine, search results are usually displayed as a list of small image thumbnails. If interested in an image, users can click on the image and open a detailed page to see the full-resolution image. As aforementioned, due to large file sizes and limited network bandwidth, users typically have to wait a few seconds for the selected image to be downloaded.

In this scenario, the words and dynamically composed ads can be blended on the enlarged thumbnail, and the composed ad image is displayed on top of the host image, until the full-resolution image is downloaded. To ensure a smooth user experience, picture/poster-based advertisements will not be used in this scenario since they are usually large images.

Initially, images in the search engine are automatically parsed based on surrounding text and image content analysis to generate the initial image category and image layout information. The search engine also encourages users to tag/annotate images returned in the search result. The annotation is focused to facilitate ads delivery. For example, users can tag an image to indicate where they believe it is suitable to place text or a logo, and what kinds of ads are suitable to be embedded in the particular image. Such tagging information, as well as the automatically parsed information, will be matched with the ads database to select the most relevant ads. The ads delivery engine will make the final decision based on viewer context.

To motivate users to provide annotations, revenues can be shared with people who contribute high-quality annotations to our service. However, since multiple users can possibly tag the same image, how to design an incentive strategy needs to be further investigated.

3.2 Image Ads in Photo Sharing Communities

In recent years, photo sharing communities on the web have grown in popularity due to tremendous desires to share photos. Such communities include personal photo sharing websites like www.flickr.com, blogs like spaces.live.com, and photo forums like www.photosig.com. Such web communities have attracted numerous users and accumulated a tremendous number of photographs.

To exploit these photos, we envision a scenario similar to the search engine-based scenario in that users typically see a small image thumbnail before viewing full-resolution versions of selected images. Image ads could be generated, delivered and rendered similarly to the search engine-based scenario.

However, the main difference between the two scenarios would be revenue sharing. In a photo sharing community web site, each shared photograph is owned by the user who created it, but each web page, at least its format, is defined and owned by the web site owner. Here users are content creators while web site owners are content publishers. This means that the web site owner has the privilege to deliver online ads inside each full-resolution photograph, and thus shares the revenue. Nevertheless, each user can also choose to turn off online ads if he/she cannot accept his/her photographs being temporarily modified. However, users who accept online ads can share revenues. And to make the embedded online ads more relevant to image content and thus yield more ad clicks, users can tag/annotate their photographs as described in the search engine-based scenario. We believe such a business model will attract more users and motivate them to be active in photograph sharing and tagging.

3.3 Image Ads in Individual Web Pages

Besides image search engines and photo sharing websites, a tremendous number of web pages created by various organizations or individuals contain high-resolution images. Therefore such individual web pages could also be subject to embedded online image ads.

In this scenario, a web page owner can submit the web page's URL to our system as well as insert a piece of javascripts code in their pages. The javascripts code will replace the owner-specified image in the web page with a Flash container, in which ads will be dynamically rendered. Note that our system caches a small thumbnail of the specified image speed up the delivery of image ads. Similar to the search engine-based scenario, the owner can tag or annotate the image to help our system choose more relevant online ads.

Thereafter, when a user browses this page, the Flash container will first display a pre-generated image ad before the full-resolution image is downloaded. Revenues caused by ads clicks can be shared with page owners.

4. IMPLEMENTATION

In this section, we present the details of the implementation of image ads.

4.1 SYSTEM OVERVIEW

To effectively deliver online image ads, we designed a platform which integrates many technologies in a meaningful way, including content-based image analysis, text mining, large-scale indexing, user-powered tagging and annotation, and client-side rendering. The architecture of the platform is shown in Figure 2.

Logically, the system can be divided into four engines, i.e. image processing, ad processing, image-ad matching, and image ad delivering, as illustrated in Figure 2.

1) The image processing engine is responsible for image collecting, parsing and indexing. Meanwhile, users are allowed to tag or annotate images in the image database and share the increased revenues generated by more accurate ad-image matching.

2) The ad processing engine allows advertisers to submit their ads and bid on keywords and images. The parsed ads are stored and indexed in the ads database. This part also maintains a log database to store ad click-through data and a revenue database for advertisers to access their business intelligence statistics.

3) The image-ad matching engine matches ads with images based on the ad and image parsing results, and maintains an index for efficiently delivering online image ads.

4) The image ad delivery engine is responsible for retrieving matched ads, rendering composed image ads, and collecting user behavior information.

Each engine and corresponding techniques will be described in detail in the following sections. Note that we did not implement all modules mentioned in the following sections, and lots of solutions can be further improved. We only implemented indispensable components to make a running system.

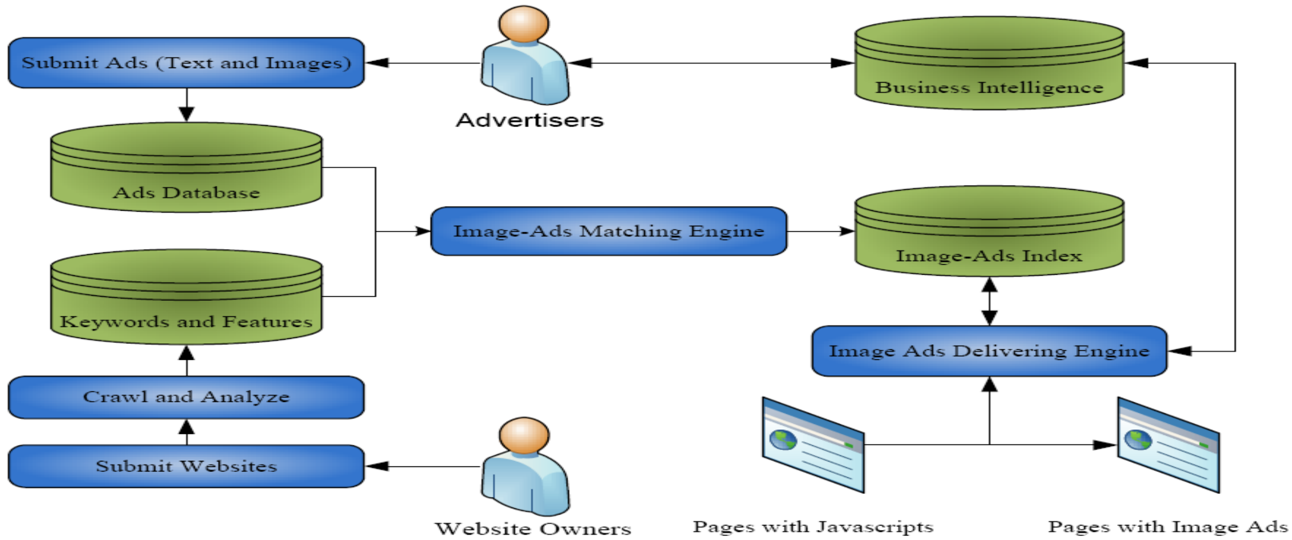

Figure 2: Architecture of the proposed image ads system

4.2 Image Processing Engine

This module is the major module to process image contents and construct text descriptions to images. The objectives of image processing include:

1) Collecting images from the Web. In section 5, we will present our solution in which website owners can join the image ads service. We only crawl pages and images from the registered websites.

2) Parsing images to get image descriptions and visual information, which will be used to match relevant ads and render ads on images.

3) Providing an interface to web users who can help to annotate images to make money and share revenues generated by more relevant ad-image matching, and better rendering effects. Figure 3 shows the work flow of the image processing engine.

The image processing engine maintains a self-service portal to enable image and website owners to submit their images as candidate ad-hosting images, and a set of editing tools for image providers to label and edit images.

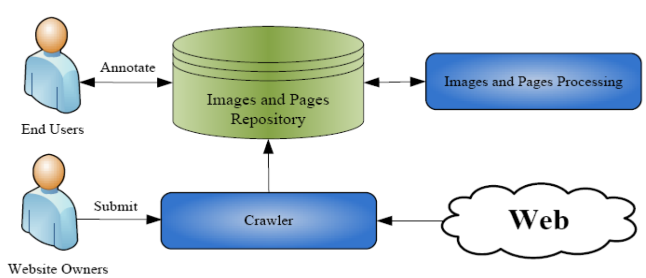

Figure 3: Workflow of image processing engine

For each input image, the engine computes both visual features and textual features necessary for matching ads. Visual features like color layout and image qualities are useful for rendering ads, while textual features (as well as image category information) are useful for finding relevant ads. Content-based image analysis is a core technology in the whole system. To effectively utilize image analysis results, a standard schema or format for those features could be defined, like the content description standard MPEG 7 [12].

Since content-based image analysis is still a challenging problem and automatic visual feature extraction may not be satisfactory, the engine also provides an interactive service to encourage users to tag and annotate images. In this way, we expect to obtain more accurate image annotation results and thus find more relevant ads. As a return, users can share revenues with us. This idea is coherent with the popular Web 2.0 concept. For those images we cannot get annotations, we use a search based image annotation technique to automatically generate keywords for them [18]. The basic idea of search based image annotation approach is to find visually similar images of a test image on the web, and use the surrounding text of these images to annotate the test image. This approach sufficiently leverages rich images and web pages resources on the web. As a result, it can be applied to annotate web images.

To facilitate image-ad matching, we need to categorize images to eight manually selected categories, i.e. *landscape*, *city*, *sports*, *technique*, *products*, *health*, *entertainment*, and *adults*. These categories are deliberately selected for our applications. Image categorization is still an open problem, and its performance is far from satisfactory [13]. However, the accuracy of text categorization is good enough for our applications [11]. Thus, in practice, we build classifiers to categorize web pages, and use the categories of web pages as categories of images.

The main components and functionalities of the image processing engine are summarized as follows:

1) Internet service for accepting images into our system
    a) Self-service portal
    b) Editing tools for image providers
2) Image analysis engine
    a) Compute features from images necessary for matching with ads
    b) Schema or standard format for those features
    c) Classify images to some pre-defined categories
3) Internet service for users to tag and annotate our image collection

a) Editing tools for web users to participate and add values to our network

b) Reputation ranking

c) Accounting and revenue sharing mechanism

4.3 Ad Processing Engine

The module is the major module to manage ads, and support business intelligence (BI) applications. The objectives of ad processing include:

1) Providing a web service for advertisers to bid on keywords or images, and submit ads. To avoid delivering of ads to improper images, advertisers also can specify what keywords or categories of images they want to avoid.

2) Generating business intelligence statistics for advertisers by analyzing click-through log.

3) Analyzing ads to prepare data for image-ad matching engine.

4) Fraud detection by analyzing click-through logs.

Figure 4 shows the workflow of the ad processing engine. Through the self-service portal, advertisers can bid on keywords and submit ads, use a set of editing tools to compose and edit ads, and preview effects of image ads. In our current implementation, an ad consists of four components, i.e. logo of the company or a small image of a product, company name, a short ad slogan or a product name, and some keywords or descriptions. For example, an ad slogan of Microsoft is "Our passion, Your potential".

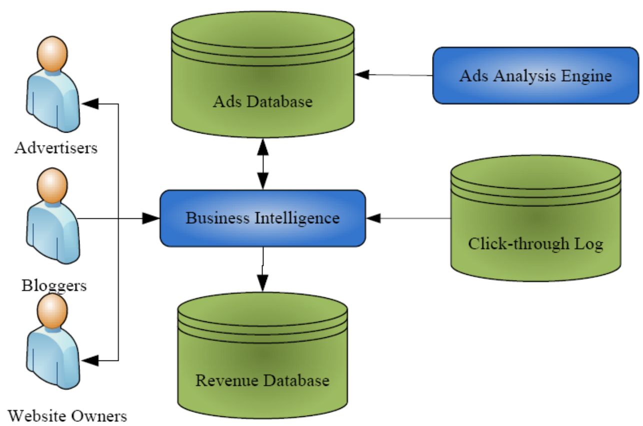

Figure 4. Workflow of Ad processing engine.

For each submitted ad, the ad processing engine computes ad features necessary for matching images. The features are used to find semantically relevant images and visually matched images to compose artistic ad posters. To effectively utilizing ad analysis results, a standard schema or format for those features could be defined. In current implementation, all text submitted by advertisers are used as features of ads.

Similar to other online advertising services, the system also has a business intelligence engine to analyze user responses to the delivered image ads and generates business intelligence statistics for advertisers to use in understanding the impacts of their submitted ads. This module also can help system administrators to detect fraud. Fraud detection in online advertisements is still an open problem [10].

The main components and functionalities of the ad processing engine are summarized as follows:

1) Internet service for advertisers to submit ads

a) Self-service portal

b) Editing tools for advertisers to create, compose, edit, and publish ads

c) Preview effects of image ads

2) Ad analysis engine

a) Compute features from ads necessary for matching images

b) Schema or standard format for ad content (words, picture, multimedia, etc.)

3) Business intelligence (BI) engine for advertisers

a) Collect user behavior data (i.e. how users respond to our embedded ads)

b) Derive relevant statistics

4.4 Image-Ad Matching

The image-ad matching engine processes the output of image processing engine and ad processing engine, and generates an efficient image-ads index. This index will be used by the ad delivery service to guarantee online access to image ads in real time. The workflow of the image-ad matching engine is shown in Figure 5.

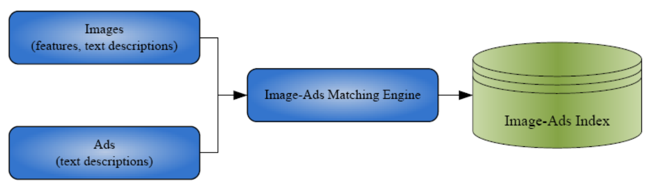

Figure 5. Workflow of Image-Ad matching engine.

The core technology in the image-ad matching engine is the similarity measure to match ads and images. A good similarity measure is critical for finding relevant ads for each host image, and is potentially capable of increasing click rates. Some researchers reported increased click rates of Google AdWords-like text ads with the improvements of relevance of ads and web pages [2, 3].

In our implementation, both images and ads are described by a few numbers of keywords, and these short "documents" are represented by vectors in the vector space model, which is very popular in text retrieval field [16]. Each element of the vector corresponds to a word in a document, and its value is the term-frequency (TF) of the word. Due to the well-known synonymy and polysemy problem, direct similarity measure, like the *cosine* similarity, may fail [5]. For example, two words "girl" and "lady" have similar meanings, but their *cosine* similarity is 0. In text retrieval literatures, a widely acknowledged solution is to project documents to some latent semantic space, and measure their similarities in that space. We also adopt this approach in our solution. We trained a Latent Dirichlet Allocation (LDA) model on 1 million Wikipedia pages with topic numbers 100 (t=100) [7]. To train the model, we built a distributed computing environment by using a cluster consisting of 10 high performance workstations. The variational topic mixture θ (a t-dimensional vector, where t is the number of hidden topics) of each document is deemed as a new representation of the document. θ is drawn from a Dirichlet distribution:

$$p(\theta|\alpha) = \frac{\Gamma(\sum_{i=1}^{t} \alpha_i)}{\prod_{i=1}^{t} \Gamma(\alpha_i)} \theta_1^{\alpha_i - 1} \cdots \theta_t^{\alpha_t - 1}$$

where $\Gamma(x)$ is the Gamma function, and α_i are Dirichlet parameters. Intuitively, the topic mixture θ of a document is the coordinates of the document in the latent semantic space [6, 7]. Besides the capabilities on dealing with synonymy and polysemy problem, another important advantage of this approach is that it can get very compact representations of documents, i.e. *t*-dimensional vectors, which are critical for web-scale applications. Once we get the compact representations of both ads and images, we can use *cosine* function to measure their similarities.

Besides the relevance, the image-ad matching engine considers lots of context information, such as country, language, user ages and tabu. For example, ads of BMW should not be delivered to images with Benz cars. All the offline matching and composing information is stored in an efficient index for real time access by the client-side rendering engine.

The main functionalities of the image-ads matching engine are summarized as follows:

1) Define a similarity measure between ads and images

2) Efficient indexing scheme (web-scale and distributed)

3) Modeling users and context

4.5 Image Ads Delivery Engine

The tasks for the image ad delivery engine include:

1) Ranking image ads according to user requests (i.e. ad types, image URL, and account info)

2) Rendering image ads in a visually pleasant manner.

The workflow of image-ads delivery engine is shown in Figure 6. At the client side (e.g. in Internet Explorer), when a user requests to browse a full-resolution image registered in our system, the image ad delivery engine will first connect to the back-end system to retrieve a set of relevant candidate ads. Since the image-ad matching and pre-composing information has been stored and indexed, the retrieval process is expected to be very fast. Based on the current image display context, relevant candidate ads are ranked at the server side and the top matched ads will be sent to the client side.

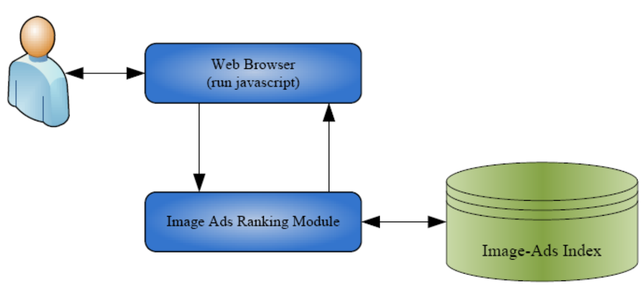

Figure 6. Workflow of Image-Ad rendering engine

Then the client side rendering engine renders the image ad at the same position where the full-resolution image will be displayed after being downloaded. The rendering effects may include, but are not limited to, text and logo overlay, poster overlay, sound effects, animation, and user interaction. With the image ad fading out after the full-resolution host image is downloaded, the client-side rendering container will place several hidden buttons on the image. When the mouse is moved over the image, the buttons will display to let users replay the ad animation, or go to the ad website.

The rendering of image ads is performed by a Flash container which is embedded in web pages by javascrips (as mentioned in Section 3.3). It is overlapped on the targeted image. The rendering of ads consists of two phases. The first phase is the downloading time. At this time, because the client-side only has a very small thumbnail, our solution is to enlarge the small thumbnail and overlap the matched ad on the enlarged image, as shown in Figure 1. After the full-resolution image is downloaded to the client-side, it will fade in smoothly. In the second phase, we will deal with the ads rendering problem on the full-resolution image. Because in most of viewing time, the user is facing with this image, the second phase rendering is very critical for image ads. When the user moves mouse over the image or clicks the replay button on the image (i.e. a Flash), the ad will be displayed again. Thus, we have to determine what rendering schema we should adopt and where to draw the ad in the second phase.

To render an ad in a schema as shown in Figure 7 on a full-resolution image, we first detect a proper region in the image, and then compute a distinguishable color to draw text. A design guide to find a proper region is that the ad should not overlap on the main object of the image. Thus, an experimental solution is to partition an image to 3x3 grids, and compute the "complexity" of each block. We first compute the gray-scale histogram of each block, and then use entropy to measure the complexity of each block:

$$\mathrm{H} = -\sum_{\mathrm{i}=0}^{255} \mathrm{p(i)} \log \mathrm{p(i)}$$

where $\mathrm{p(i)}$ is the ratio that the number of pixels taking value *i* to the number of all pixels in the block. The bigger the *H* value is, the more complex the block will be. Ad will be rendered to the block with the minimum *H* value. Because the computation of this algorithm is very simple, it can be applied to deal with web-scale image collections, which is very critical for our system. An alternative method is to always use the center of the image to render the ad. Figure 7 shows a comparison of the two approaches.

A by-product of this approach is the dominant intensity of each block. The dominant intensity of a block is defined as the intensity that maximum pixels take. It is used to determine the intensity of ad text by this simple rule,

$$tc = \begin{cases} 255 - dc & dc < 64 \; or \; dc > 192 \\ dc + 128 & 64 \leq dc \leq 128 \\ dc - 128 & 128 < dc \leq 192 \end{cases}$$

where *tc* stands for the intensity of ad text, and *dc* is the dominant intensity of the block where the ad will be rendered. By design, the ads text will be monochromatic. In this way, we can guarantee the necessary contrast of the foreground text and background image. Moreover, we notice the background (i.e. the rendering block) for ad rendering is usually not uniform, which will affect the visual effect of image ads. Thus, we overlap a 25% transparent mask on the rendering block. From Figure 7, we can see the ads are very clear. It is worth noting that the analysis of image contents is performed in the image processing engine after we crawl them from the web. The rendering block and dominant intensity of an image are important meta-data and are computed off-line.

Actually, the presented rendering schema is only one experimental schema. Other schemas including some kinds of

Figure 7: Comparision of two rendering effects

animation are more attractive. However, how to use image analysis techniques to automatically choose the best schema and corresponding parameters is not solved yet. Apparently, the current solution is far from "good-looking" for wide range of images and there is a large improvement space in this step. However, it may be the most applicable one due to its extremely low computational cost.

The main components and functionalities of the image ad delivery engine are summarized as follows:

1) Ad delivering engine
   a) The process of getting ads into host images during the image consumption process.
   b) Client-side javascripts
2) Ad rendering engine
   a) Control the actual display of ads inside images
   b) Animations
   c) Enable replay and dynamic linking.

5. JOINING IMAGE ADS SERVICE

The Image ads service will be run by a league of a service provider, advertisers, website owners, and individual users. If the service runs smoothly, all partners of the project will earn money.

5.1 Websites Join the Service

We designed a simple way for website owners. Even a website with only a few web pages/images can join the project. As long as it has a number of active users, it will benefit from the project. According to our design, to join the project, the website owners have to do two things.

5.1.1 Register as a Member of the Service

Login the service website and submit URLs of their websites. Once the service provider got these URLs, it will begin to crawl their web pages, and analyze both pages and images to extract features. These features will be used to detect dominant images [4], analyze image layout and matching images and ads. Website owners also can select what kinds of ads they would like to deliver to their sites, and styles and frequencies of image ads. If no selection is made, our system will automatically select ads according to our matching algorithm.

5.1.2 Modify Web Pages

Add a simple javascript in headers of their web pages. When the end user opens a web page on the site, the embedded javascript code will run and replace the dominant image of the page with the rendering container (i.e. a Flash). By inserting a javascript function in the tag of an image, website owners can also manually specify host images of ads. For some websites (e.g. forum sites), whose pages are generated from some pre-defined templates, web site administrators only need to modify these templates.

5.2 Individual Users Join the Service

Individual users can register as members of the service, and login the image ads website to browse and annotate images. If their annotations lead to a better relevance matching of images and ads, and result in higher click rates of ads, they will share revenues from the service. By our design, users not only can provide keywords for images, but also can propose rendering schemas and decide corresponding parameters for images.

From the perspective of research, this mechanism is a very promising approach to collecting image annotation keywords. Google Image Labeler [14] is a similar service for image annotation, in which users label images like playing a game [15]. However, it does not have a clear business model to stimulate users to continuously label images. Usually, users will feel tedious after play the game for several times. Different from Google's Image Labeler, people can earn money from labeling images in the proposed service. In the future, we will explore to combine the game model with our business model to attract more people to join the labeling work.

6. EMPIRICAL STUDY

To evaluate the proposed approach of online image ads, we conducted three experiments. The first experiment is designed to optimize the cost of extra bandwidth for image ads delivery. The second one is designed to evaluate the performance of image-ad matching approach. The last one is a subjective user study.

6.1 Bandwidth Cost of Image Advertisement

A direct concern to our new ads delivering mode is the cost of bandwidth. Thus, we designed this experiment to check the extra cost caused by ads, and try to find a way to minimize the cost and meanwhile preserve a good user experience. The extra bandwidth cost of ads consists of three parts: a Flash container, a thumbnail image and the needed ad information (i.e. a logo and some ad keywords), as shown in Table 1.

Most of web browsers will cache part of previously downloaded contents. When a user opens a web page, a browser first checks whether contents needed to render the web page have been downloaded before in the local cache files. If an object has been downloaded before, it will not be downloaded again. Thus the Flash container only needs to be downloaded once. By carefully optimizing, we reduce the Flash container to only 10k bytes.

Table 1. Extra bandwidth cost of image ads

Object	Size	Download Times
Flash Container	10k bytes	1 time
Thumbnail Image	2-15k bytes	1 time per page view
Ads	3-5k bytes	1 time per page view

Table 2. Results of user study about thumbnail size

	24x24	32x32	48x48	64x64	80x80	100x100
#Good	0	1	2	3	3	4
#Fair	1	0	2	2	2	1
#Bad	4	4	1	0	0	0

Thus, the extra cost of bandwidth only consists of the thumbnail image and the ad itself. The total size is no more than 20K bytes. To further reduce the cost, we test the user experience under different sizes of thumbnail images by a user study. We deliver thumbnails from 24x24 pixels to 120x120 pixels to the rendering engine, and enlarge them to compose image ads. Five students were asked to grade the synthesized image ads. Table 2 shows the details of the results. When the thumbnail is very small, i.e. smaller than 32x32 pixels, the enlarged image even cannot preserve the original color of the full-resolution image. Once the thumbnail is larger than 64x64 pixels, the user experience is acceptable. Thus, in our implementation, we deliver 64x64 pixels thumbnail image to end users. The average file size of such thumbnails is only 5k bytes.

6.2 Performance of Image-Ad Matching

Due to the expensive cost of analyzing image contents, we do not use image annotation techniques to generate keywords for images. As aforementioned we extract "surrounding text" of an image as its description, and use this description to find relevant ads.

We manually collected ads of 30 top companies in various domains, e.g. Xeorx and Cocacola, and use their articles on Wikipedia as their descriptions. To construct a ground truth, for each company, we use its company name as the query to search top 10 relevance pages from Live Search. Because some pages do not have meaningful images, we remove these web pages. In this way, we finally constructed a dataset of 207 web pages. Supposing ads will be delivered to the dominant image in each page, we use the DOM (document object model) tree-based method to analyze the page layout and extract surrounding text of its dominant image.

The LDA model was trained with 1 million web pages crawled from Wikipedia. We trained three LDA models with topic numbers from 60 to 150. Both descriptions of ads and images are projected to the latent semantic spaces spanned by LDA topics. Their coordinates in this space are new representations of ads and images. Given a new representation of an image, we use cosine similarity to rank ads.

We implemented three baseline algorithms. The first algorithm is the cosine similarity in vector space model [16], in which a document is represented by term-frequency of words. This algorithm is the most popular algorithm in information retrieval applications. The second algorithm is Latent Semantic Indexing (LSI) [5], in which documents are projected to a latent space by performing singular value decomposition (SVD) [5]. The last algorithm is probabilistic latent semantic indexing (pLSI) [6]. This algorithm is a probabilistic version of LSI. The three algorithms are termed as cosine, LSI and pLSI respectively. The only parameter of LDA, pLSI and LSI is the number of topics. Figure 8 shows the results of the four approaches under different topic numbers. The accuracy of LDA is significantly better than others. The best accuracy of LDA model is obtained when the number of latent topics is 120.

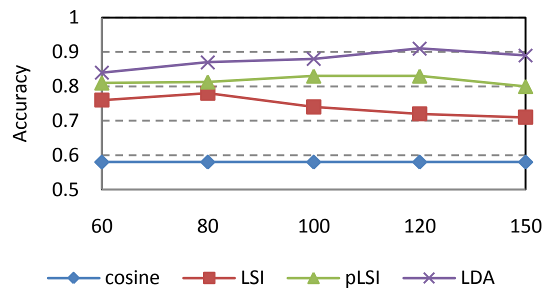

Figure 8. Performance of different image-ad matching algorithms

6.3 User Study

To compare the proposed new ads mode with existing commercial ads models, e.g. banner ads and Google AdSense-like text ads, one should conduct comprehensive user studies. However, such kind of user study is very difficult to conduct because different ads modes have different characteristics and functionalities, and a fair comparing environment is difficult to build. To simplify the situation, in this paper we only conduct a preliminary study on whether users will accept image ads, and what are non-intrusive manners to deliver image ads. In our future work, we will design more comprehensive user studies to compare the strength and weakness of these ads modes.

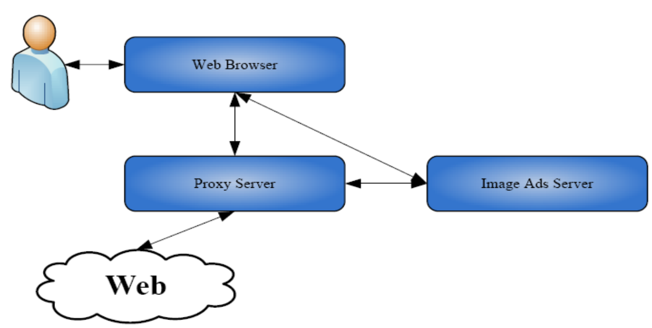

Figure 9. Architecture of our experimental system.

6.3.1 System Design and Settings

To insert image ads to any web pages, we use a trick to hack web pages. We built a proxy server and force the web browser to use it. When the proxy server receives a page (URL) request, it first requests the page from the web, and then inserts a piece of javascripts code in the HTML header of the page. We use this hack technique to simulate the process that a website administrator manually modifies a web page. Figure 9 shows the structure of our experimental system.

Seven students were invited to participate in our user study for three days. Usually, we asked them to see some ads, and rate each ad in three levels: *Good*, *Fair*, and *Bad*. The average numbers of ratings are used to evaluate the user experience of the shown ads.

6.3.2 Relevant Ads vs. Irrelevant Ads

We asked each participant view 20 pages with image ads. In this study, a participant can start from any web page, and he or she is limited to view only 20 pages, i.e. 20 image ads. The participant can either follow links in pages or randomly jump to any web pages. Among the 20 image ads, 10 image ads are relevant ads got by our image-ad matching engine, while the rest 10 ads are randomly selected from 30 ads in our database. We randomly permutated the 20 image ads given to each user to make sure the comparison is fair. Table 3 shows the average numbers of the three ratings.

Table 3. Experimental results of relevant ads vs. irrelevant ads

	Relevant Ads	Irrelevant Ads
#Good	5.7	4.9
#Fair	3.8	4.2
#Bad	0.5	0.9

According to the table, we find users only slightly prefer relevant image ads. This experimental result is much different from our expectation. Intuitively, we think the user will much prefer relevant ads than irrelevant ads. This result could be explained in two ways. First, we only have 30 ads for 30 top companies in our database. All of these companies have lots of products or services. The relevant or irrelevant judgment is not that obvious. For example, for a sports web image, there may be no difference to deliver ads of Nike, Cocacola, or even Nokia. The goal of these ads is to build brand names or corporate images rather than advertise information of specific products or services. Second, image ads may be not sensitive to relevance, at least, not as sensitive as that of text ads. Image ads may have better adaptability than text ads, which is very useful for companies who not only want to advertise product information but also want to build brand names and corporate images.

Thus, only based on this study, it is insufficient to draw a conclusion that relevant image ads are better than irrelevant image ads. Some researchers reported users might prefer to click relevant ads [2, 3] in text-based ads. However, such conclusion is difficult to check in our environments. Different ads purposes may have different relevance requirements to image ads, i.e. building brand names or advertising product information. It will be an important research topic in our future work after we get more ads data and sufficient click-through log of image ads by deploying a testing system.

Table 4. Experimental results of relevant ads vs. irrelevant ads

	Good Schema	Bad Schema
#Good	5.7	3.2
#Fair	3.8	3.9
#Bad	0.5	2.9

6.3.3 Good Rendering Schema vs. Bad Rendering Schema

In the second user study, we deliver image ads to end users in two different rendering schemas, i.e. the well designed schema as described in section 4.5 and an ordinal schema. In the second schema, we randomly picked a region in an image to show ads, and the color of text is fixed to be white. Table 4 shows the experimental results.

Obviously, according to Table 4, the good schema brought users more pleasure than the bad schema. We can safely conclude that a good ad rendering schema is very useful for image ads. However, we cannot conclude that our rendering schema is a good schema because the evaluation is very subjective.

Table 5. Experimental results of two rendering schemas

	Ads at Center	Our Approach	By Designer
#Good	5.2	5.7	8.4
#Fair	2.1	3.8	0.9
#Bad	2.7	0.5	0.7

To further test the performance of the proposed ad rendering method, we conducted a similar study. In this study, users were asked to view 30 image ads, among which 10 ads were shown at the center of images (as shown in the left image in Figure 7), 10 ads were shown in blocks selected by the proposed algorithm (as shown in the right image in Figure 7), and the rest 10 ads were rendered by schemas designed by our designers for these images. This study tests the effects of different rendering parameters to the same rendering schema, and different rendering schemas. Table 5 illustrates the experimental results. For the first two approaches, we did observe the improvement of user experience, but the improvement is not significant. By further analyzing the ratings of individual results, we find the ratings from different users vary greatly. It is obvious some users like showing ads at the center of images, and some users do not care where the ads are. This result indicates the difficulty to find a generally acceptable rendering parameters. However, the ads effects designed by our designers significantly outperformed the other two approaches. This fact indicates the importance of artistically visual effects for image ads. Inspired by the artistically designed schemas, we are investigating a new solution in which we ask experienced designers to design a series rendering schemas and provide necessary descriptions. We believe a well designed algorithm is capable of choosing a right rendering schema to match the image and ad.

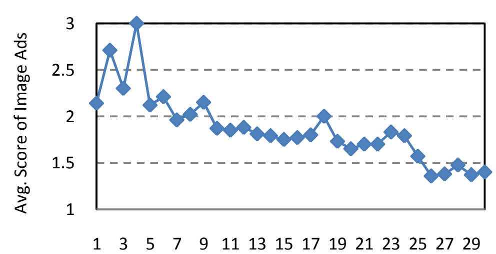

Figure 10. Average ratings to 30 continuous shown images ads

6.3.4 Delivering Frequency of Image Ads

In psychology study, scientists found a fact that people will be "tired" facing with a scene for a long time, even if the scene is beautiful [15]. Thus, we also concern the reaction of users when they continuously see lots of images with ads, even if the visual effects are good. To simulate the user behaviors, within the three days, we asked the seven participants to view 30 image ads continuously, and rate each image ads to three levels. The experimental results are shown in Figure 10. The horizontal axis is the order of image ads, and the vertical axis is the average ratings to each image ads.

From this experiment, we observed the same fact as obtained in psychology. Users will feel tedious to images ads which are continuously shown in web pages. This fact indicates that,

1) We should use a mechanism, e.g. session, to record the ads viewing of the same user, and stop to show image ads to him or her after 3-5 image ads have been shown.

2) We should vary the rendering schemas in an image ad view session. It may refresh the minds of users.

7. CONCLUSION

We have presented an innovative idea that allows ads to be embedded into images in a non-intrusive and visually pleasant manner. Our ads mode intelligently leverages the wasted information channel when users are waiting to see a full-resolution image being downloaded. To maintain a smooth user experience, we choose a semantically relevant ad and blend the ad on the enlarged thumbnail image in a visually consistent manner. Based on this idea, we propose three typical scenarios that can adopt the proposed image-advertising mode, and designed the system architecture to implement the online image advertising system. Importantly, the system creates a new business model that can encourage users to tag or annotate photos and share the increased revenues. We believe this innovative advertising mode will greatly increase the online advertising market and bring remarkable revenues to online business.

8. OPEN PROBLEMS

In most cases, researches are driven by applications. Currently, the most important application of web images is image search engine. However, the poor query share of image search engines definitely cannot embody the value of web images and stimulate researches on these important multimedia contents. Although web images are as pervasive as web pages, their value is greatly under estimated. Through this application, we probably find a way to monetize numerous web images. Though the proposed solution is still preliminary, we expect that this work can inspire more research on web images. From the perspective of research, we point out some new research directions of web image.

8.1 New Image Labeling Model

Labeling of images is very critical for research. Most existing labeling models either pay labelers (i.e. revenue) or attract users by games (i.e. pleasure) [15]. Obviously, these models have defects. We believe a good motivating mechanism should cover both sides at the same time: revenue and pleasure. Existing online ads business models only consider maximizing the profit of advertisers and service providers. However, in our model, we provide business intelligence to involve end users and motivate them label images. In this model, a user is probably willing to label as many as possible images if this can bring him or her as much as possible money. This model targets at maximizing the revenue of all participants. The complete business model to connect advertisers, service provider and end users is not only a very interesting research topic, but also a very healthy ecosystem.

8.2 Automatic Image Annotation

Automatic image annotation has been proven a very challenging research problem [18]. Most existing approaches try to learn models from small scale training set. As a result, these models cannot be applied to wide-semantic-range images, like web images. Combined with this new labeling model, we see a light to solve the web image annotation problem by leveraging numerous web users. In such an innovative annotation approach, not only technical factors but also commercial factors are considered.

8.3 Annotation Refinement

Using surrounding text as annotations of images is an effective and efficient approach to annotate images. However, due to the complexity of web page layout, it is difficult to precisely extract surrounding text of images. Thus, surrounding text is very noisy. Effective and efficient annotation refinement approaches will be extremely useful for this project.

8.4 Images and Text Collage

As shown in our user study, good ad rendering effects are very important for end users. This is because good rendering approaches will lead better advertisement effects, and more revenue. Both graphics and computer vision researchers could be involved to study images and text collage techniques. Some work in the computer graphics community may be adapted to address this problem [19].

9. REFERENCES

[1] Online advertising to hit $16.7 billion this year, July 18, 2006, *http://www.emarketer.com/ Article.aspx?1004078*

[2] M. Ciaramita, V. Murdock, and V. Plachouras. Semantic associations for contextual advertising. *IJECR* 9(1), 2008

[3] J. Carrasco, D. Fain, K. Lang and L. Zhukov. Clustering of bipartite advertiser-keyword graph. *ICDM*, 2003.

[4] Z.W. Li, S.M. Shi, and L. Zhang. Improving Relevance Judgment of Web Search Results with Image Excerpts. *WWW*, 2008

[5] S. Deerwester, Susan Dumais, G. W. Furnas, T. K. Landauer, R. Harshman. Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science*, 41

[6] T. Hofmann, Probabilistic Latent Semantic Indexing, *SIGIR*, 1999

[7] D. Blei, A. Ng, M. Jordan. Latent Dirichlet allocation. *JMLR*

[8] Google AdWords. http://adwords.google.com/

[9] Google AdSense. https://www.google.com/adsense/

[10] N. Daswani, C. Mysen, V. Rao, S. Weis, K. Gharachorloo, and S. Ghosemajumder. Online Advertising Fraud. *http://www.google.com/adwords*

[11] T. Joachims. Learning to Classify Text using Support Vector Machines: Methods, Theory, and Algorithms. Kluwer, 2002

[12] ISO/IEC MPEG-7 Standard. http://www.chiariglione.org/mpeg/standards/mpeg-7

[13] Y. Gao, and J. Fan. Semantic image classification with hierarchical feature subset selection. *MIR*, 2005

[14] Google Image Labeler. http://images.google.com/imagelabeler/

[15] L. Ahn and L. Dabbish. Labeling Images with a Computer Game. *ACM CHI*, 2004

[16] C. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008.

[17] T. Mei, X.-S. Hua, L.J. Yang, S.P. Li. VideoSense - Towards Effective Online Video Advertising. ACM MM, 2007

[18] X.-J. Wang, L. Zhang, F. Jing, W.-Y. Ma, AnnoSearch: Image Auto-Annotation by Search. CVPR, 2006

[19] A. Agarwala1, et al. Interactive Digital Photomontage. SIGGRAPH, 2004