Contextual In-Image Advertising

Tao Mei, Xian-Sheng Hua, Shipeng Li

Microsoft Research Asia
Vancouver skyline at night

People on beach in front of Vancouver skyline at night. From Spanish Banks.

-Added to the Cream of the Crop pool as most interesting.

This photo has notes. Move your mouse over the photo to see them.

Comments

rod kenny pro says:

Nice. The shore line draws the eye into the photo.

Posted 25 months ago. (permalink)

McPig pro says:

From the beautiful twinkling lights of Vancouver skyline to lights in Canary Wharf of London

A piece of (The Game Of Photo Association)

Posted 25 months ago. (permalink)

Rames Studios pro says:

Nice shot, i luv it.

Posted 25 months ago. (permalink)

John Sebastian Photography pro says:

Cool night shot!

Posted 25 months ago. (permalink)
Outline

- Motivation
- ImageSense
  - Framework
  - Problem formulation
  - Global and local relevance
  - Problem optimization
- Experiments
- Summary
Motivation

• Online images

- 2.8 bln image uploads by 30 mln members till 2008 Sept. (SFGate)
- 2,530 image uploads per min in 2008 Oct.
- 50 bln digital photos taken in 2007, 60 bln by 2011 (IT Facts)

• Online advertising
  “... grow to $25.9 billion this year, from $21.1 billion last year, and hit $30 billion in 2009.” (eMarketer)

• Image-driven advertising
From the perspective of industry

• Conventional advertising: treat image advertising as text advertising
  – Match ads with images only based on textual info.
  – Insert ads at a fixed position

Ads are not always contextually relevant to image content

Preserved ad blocks are intrusive and break page structure
From the perspective of research

• Webpage advertising
  – Ad keyword selection [Yih, WWW’06; Shen, SIGIR’06]
  – Ad relevance matching
    • Keyword-targeted ads (paid search ads) [Mehta, JACM’07]
    • Content-targeted ads (contextual ads) [Rib, SIGIR’05; Broder, SIGIR’05]
    • User-targeted ads (audience intelligence) [Richardson, WWW’07]

• Video advertising
  – VideoSense [Mei, MM’07]
  – vADeo [Srinivasan, MM’07]

• Image advertising (to be investigated)
  – Delivering ads inside images during image buffering [Li, MM’08]
What we want to argue

• Images are more powerful and effective *information carriers*
  – carry more information
  – more attractive than text
  – shown faster than video

• Ads are to be embedded **in-image**
  – no predefined ad blocks required
  – ads promoted by salient appearance

• Ads are to be **contextually relevant**
  – web pages are too much or too few to describe image content
  – image content and its surrounding text are more precise
What we want to propose

- **ImageSense**: a contextual in-image advertising platform
  - automatically matches ads to image content
  - finds nonintrusive ad insertion positions within images
  - seamlessly embeds relevant ads based on these positions
Outline

• Motivation
• ImageSense
  – Framework
  – Problem formulation
  – Global and local relevance
  – Problem optimization
• Experiments
• Summary
ImageSense framework

Three key Problems

1. “relevance” - how to select ads?
2. “position” - where to insert ads?
3. “appearance” - how to display ads?

Solutions

- vision-based page segmentation
- multimodal relevance matching
- image saliency detection
- auto-animation
- seamless blending and transition
Problem formulation

Objective: maximize overall relevance while minimize intrusiveness

- global textual relevance \( R_g(P, a_j) \rightarrow R_g(a_j) \)
- local textual relevance \( R_\ell(I, a_j) \rightarrow R_\ell(a_j) \)
- local content relevance \( R_c(b_i, a_j) \)

\[
\begin{align*}
\max_{(x, y)} f(x, y) &= w_g \sum_{j=1}^{N_a} y_j R_g(a_j) + w_\ell \sum_{j=1}^{N_a} y_j R_\ell(a_j) + w_c \sum_{i=1}^{N_b} \sum_{j=1}^{N_a} x_i y_j R_c(b_i, a_j) \\
\text{s.t.} \quad \sum_{i=1}^{N_b} x_i = N, \quad \sum_{j=1}^{N_a} y_j = N, \quad x_i, y_j \in \{0, 1\} \quad \text{image block: } x_i \in \{0, 1\} \\
&\text{ad: } y_j \in \{0, 1\}
\end{align*}
\]
Global and local textual relevance

Original Surrounding Text

**Keywords:**
Pyramids of Giza, Egypt

**Description:**
the most famous monuments of ancient Egypt. These massive stone structures were built around 4500 years ago on a rocky desert plateau close to the Nile. But the intriguing Egyptian pyramids were more than just tombs for kings. The mysteries surrounding their symbolism, design and purpose have inspired passionate debate. It is likely that many of these mysteries will never be solved ...

Expansion Text

earth, shape, circle, aztec, food, pyramid, geography, triangle, egypt, ufo, rainbow, sunrise, mummy, prism, dinosaur, empire, khufu, taba, cheops, great pyramid, dahab, aswan, acient, valley of the kings, pyramid, tansik, pangea, saqqara, hurghada, greece, egyptian, africa, dubai, cyprus, jamaica, ghana

Concept Text

Outdoor, Mountain, Person

\[ T = \{ \text{original text, expansion text, concept text} \} \]

Global textual relevance:

\[ R_g(P, a_j) = sim(P^{(T)}, a_j^{(T)}) \]

Local textual relevance:

\[ R_\ell(I, a_j) = sim(I^{(T)}, a_j^{(T)}) \]

BM25 similarity:

\[ sim(q, d) = \sum_{i=1}^{n} \frac{idf(q_i) \times tf(q_i, d) \times (k + 1)}{tf(q_i, d) + k(1 - b + b \times ndl(d))} \]
Local content relevance

Original image $B_i$  

Image saliency map $S_i$  

Weight map $W_i$

Local content relevance:

$$R_c(b_i, a_j) = W_i \times (1 - S_i) \times \left(1 - d(B_i, a_j^{(V)})\right)$$

$$d(B_i, a_j^{(V)}) = \frac{1}{|B_i|} \sum_{b_i \in B_i} \sum_{k=1}^{K} \left| f(b_i)(k) - f(a_j)(k) \right|$$

Nonlinear 0-1 integer programming problem (NIP):

$$\max_{(x, y)} f(x, y) = w_g \sum_{j=1}^{N_a} y_j R_g(a_j) + w_\ell \sum_{j=1}^{N_a} y_j R_\ell(a_j) + w_c \sum_{i=1}^{N_b} \sum_{j=1}^{N_a} x_i y_j R_c(b_i, a_j)$$

$$s.t. \sum_{i=1}^{N_b} x_i = N, \sum_{j=1}^{N_a} y_j = N, x_i, y_j \in \{0, 1\}$$

Algorithm 1 The heuristic searching algorithm for Eq. (1)

1: Initialize: set the labels of all the elements in \( x \) and \( y \) as “0” (i.e., “unselected”).

2: Rank all the elements in \( x \) according to \((w_i \times (1 - s_i))\) in a descendent order, select the top \( N \) elements, and set their labels as “1” (i.e., “selected”).

3: Rank all the elements in \( y \) according to \( w_g R_g + w_\ell R_\ell \) in a descendent order, and select the top \( N'_a \) (\( N < N'_a \ll N_a \)) elements.

4: For each \( x_i \) in the top \( N \) elements in \( x \), select the unselected \( y_j \) from the top \( N'_a \) elements in \( y \) with the max\(\{w_c R_c + w_g R_g + w_\ell R_\ell\}\), and set the label of \( y_j \) as “1”.

5: Output all the pairs with \((x_i = 1, y_j = 1)\).

Computational complexity: \( C_{N_a}^N \cdot C_{N_b}^N \cdot N! \rightarrow O(N_b + N_a + N \times N'_a) \)
How to apply ImageSense

- Website
- Webpage
- WebImage

Subscribe service

Crawling/Segmentation

Image Analyzer

Image-AD Matching

Ad & ad position

Ad Database

Ad Delivery

<script type="text/javascript">
No = 85E3F7535323884617EF6FFD71389; topmargin = 0; leftmargin = 0;
</script>
<script type="text/javascript" src="http://msra-va/MSRAADFetch/AD_Webpage.js"></script>
Experiments - data

• Advertisements
  – 7,285 unique ad product logos
  – annotated by 20 subjects
  – 32,480 unique ad words

• Online images
  – 382,371 images from www.tango.msra
  – 200,000 images from Flickr

• Evaluation
  – 1,100 ad triggering pages
    • 100 web pages from major news sites
    • 1,000 images searched by top 100 image queries
  – 86.74 words in average per page/image

party, pics, pyramids, dragon, girl, baby, hearts, zebra, spring, wallpaper, car, cat, rose, money, dog, flower, chat, butterfly, fish, graffiti, cute kittens, angel, nature, football, puppies, wedding cakes, horse
Experiments - methodology

- Different advertising strategies
  1. global textual relevance (traditional advertising)
  2. global and local textual relevance without concept text
  3. global and local textual relevance with concept text
  4. textual relevance and content relevance with equal weights
  5. textual relevance and content relevance, with different weights from IV ($w_g=0.3$, $w_l=0.5$, $w_c=0.2$)

\[
\max_{(x,y)} f(x, y) = w_g \sum_{j=1}^{N_a} y_j R_g(a_j) + w_l \sum_{j=1}^{N_a} y_j R_{\ell}(a_j) + w_c \sum_{i=1}^{N_b} \sum_{j=1}^{N_a} x_i y_j R_c(b_i, a_j)
\]

s.t. \[
\sum_{i=1}^{N_b} x_i = N, \quad \sum_{j=1}^{N_a} y_j = N, \quad x_i, y_j \in \{0, 1\}\]
Experiments - methodology

- Different advertising strategies
  
  I. global textual relevance (traditional advertising)
  II. global and local textual relevance without concept text
  III. global and local textual relevance with concept text
  IV. textual relevance and content relevance with equal weights
  V. textual relevance and content relevance, with different weights from IV
     \((w_g=0.3, w_l=0.5, w_c=0.2)\)

\[
\max_{(x,y)} f(x, y) = w_g \sum_{j=1}^{N_a} y_j R_g(a_j) + w_l \sum_{j=1}^{N_a} y_j R_l(a_j) + w_c \sum_{i=1}^{N_b} \sum_{j=1}^{N_a} x_i y_j R_c(b_i, a_j) \\
\text{s.t.} \sum_{i=1}^{N_b} x_i = N, \sum_{j=1}^{N_a} y_j = N, x_i, y_j \in \{0, 1\}
\]
Experiments - methodology

• Different advertising strategies
  I. global textual relevance (traditional advertising)
  II. global and local textual relevance without concept text
  III. global and local textual relevance with concept text
  IV. textual relevance and content relevance with equal weights
  V. textual relevance and content relevance, with different weights from IV \((w_g=0.3, w_l=0.5, w_c=0.2)\)

\[\text{II: } R_l \text{ without concept text}\]

\[
\max_{(x, y)} f(x, y) = \sum_{j=1}^{N_a} w_g y_j R_g(a_j) + \sum_{j=1}^{N_a} w_l y_j R_l(a_j) + \sum_{i=1}^{N_b} \sum_{j=1}^{N_a} w_c x_i y_j R_c(b_i, a_j)
\]

\[s.t. \sum_{i=1}^{N_b} x_i = N, \sum_{j=1}^{N_a} y_j = N, x_i, y_j \in \{0, 1\}\]
Experiments - methodology

- Different advertising strategies
  
  I. global textual relevance (traditional advertising)
  II. global and local textual relevance without concept text
  III. **global and local textual relevance with concept text**
  IV. textual relevance and content relevance with equal weights
  V. textual relevance and content relevance, with different weights from IV ($w_g=0.3, w_l=0.5, w_c=0.2$)

**III: $R_l$ with all kinds of texts**

$$\max_{(x, y)} f(x, y) = w_g \sum_{j=1}^{N_a} y_j R_g(a_j) + w_l \sum_{j=1}^{N_a} y_j R_\ell(a_j) + w_c \sum_{i=1}^{N_b} \sum_{j=1}^{N_a} x_i y_j R_c(b_i, a_j)$$

$$s.t. \sum_{i=1}^{N_b} x_i = N, \sum_{j=1}^{N_a} y_j = N, x_i, y_j \in \{0, 1\}$$
Experiments - methodology

- Different advertising strategies
  I. global textual relevance (traditional advertising)
  II. global and local textual relevance without concept text
  III. global and local textual relevance with concept text
  IV. textual relevance and content relevance with equal weights
  V. textual relevance and content relevance, with different weights from IV ($w_g = 0.3$, $w_l = 0.5$, $w_c = 0.2$)

\[
\text{IV: } w_g = w_l = w_c = \frac{1}{3}
\]
Experiments - methodology

• Different advertising strategies
  I. global textual relevance (traditional advertising)
  II. global and local textual relevance without concept text
  III. global and local textual relevance with concept text
  IV. textual relevance and content relevance with equal weights
  V. textual relevance and content relevance, with different weights from IV \(w_g = 0.3, w_l = 0.5, w_c = 0.2\)

\[
\max_{(x,y)} f(x,y) = \sum_{j=1}^{N_a} w_g y_j R_g(a_j) + w_l \sum_{j=1}^{N_a} y_j R_l(a_j) + w_c \sum_{i=1}^{N_b} \sum_{j=1}^{N_a} x_i y_j R_c(b_i, a_j)
\]

\[s.t. \quad \sum_{i=1}^{N_b} x_i = N, \quad \sum_{j=1}^{N_a} y_j = N, \quad x_i, y_j \in \{0, 1\}\]

V: \(w_g = 0.3, w_l = 0.5, w_c = 0.2\)
(a) images with poor tags (3.01)

(b) images with rich tags (17.93)

(c) web pages (789.8)

(d) overall (1,100 triggering pages)
Evaluation on Ad Relevance

• Observations
  – ImageSense outperforms traditional advertising, especially for the images/pages with rich tags
  – Surrounding texts and concept texts improve ad relevance
  – Local content relevance is a tradeoff between ad relevance and user viewing experience

• Computational time
  – Online ad matching: < 0.1 second
  – Image processing (800×600 pixel): < 0.1 second
  – Overall: < 0.2 second
Evaluation on User Experience

• 15 subjects involved
  – 8 female and 7 male undergraduate/graduate students
  – with different background

• Each gives a satisfaction score (1-5)
  – Ad insertion position
  – Satisfaction

<table>
<thead>
<tr>
<th>Satisfaction Score</th>
<th>Conventional Advertising</th>
<th>ImageSense</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ad position</td>
<td>3.00</td>
<td>&lt; 3.75</td>
</tr>
<tr>
<td>Overall</td>
<td>3.68</td>
<td>&lt; 3.85</td>
</tr>
</tbody>
</table>
Corbis images (http://pro.corbis.com/)
Example - Tango

Tango Photos help you preserve your memories and share your life through photographs. So why not give it a try? Please use your MSN ID to log in and upload your photos.

Photo Exhibition User Guide

Exhibition Tags
- Most Popular
- Most Original
- Best Technique

Liu xiang
Uploaded by: Stephan at 212 days ago.
12 views/ 1 comments

My Pet
Uploaded by: Stephan at 212 days ago.
Tags: little cat
15 views/ 1 comments
Live Search Images: “car”
Live Image Search Results

VIPS Segmentation
Live Image Search Results

Image Saliency Detection
Live Image Search Results

Live Search Image with ImageSense
Website: Corbis
Conclusions

- ImageSense support:
  - “contextually relevant” and “less-intrusive” advertising
    - website / webpage / webimage
  - automate ad delivery by only inserting a piece of code
  - a long-tail business model
    (for both advertisers and publishers)
Thanks!

ACM Multimedia 2008

October 27 - 31, 2008
Vancouver, BC, Canada

Venue

Pan Pacific Hotel
300 - 999 Canada Place
Vancouver, British Columbia, Canada

Hotel Information

The conference will be held in Pan Pacific Hotel, Vancouver, BC. For location and travel information, click here.

Room Booking Information

Conference attendees can book the rooms EITHER by using the E-Direct link OR by entering the Reference Code type as "Corporate and Code #: 2647490" on the hotel website.

"Guestroom Rates"
Deluxe Guestrooms (City view, Inner Harbour or Harbour Mountain) for $175.00
(This rate is guaranteed until *September 26, 2008*)
Thanks!
Implementation of ImageSense

(1) ImageSense service portal

(2) Publisher register a webpage

(3) ImageSenser provide a piece of html code

(4) Publisher’s page with ad