Automatic Generation of Social Media Snippets for Mobile Browsing

Wenyuan Yin †, Tao Mei ‡, Chang Wen Chen †
† State University of New York at Buffalo, Buffalo, NY, USA
‡ Microsoft Research Asia, Beijing, P. R. China
{wyin4, chencw}@buffalo.edu; tmei@microsoft.com

ABSTRACT

The ongoing revolution in media consumption from traditional PCs to the pervasiveness of mobile devices is driving the adoption of social media in our daily lives. More and more people are using their mobile devices to enjoy social media content while on the move. However, mobile display constraints create challenges for presenting and authoring the rich media content on screens with limited display size. This paper presents an innovative system to automatically generate magazine-like social media visual summaries, which is called "snippet," for efficient mobile browsing. The system excerpts the most salient and dominant elements, i.e., a major picture element and a set of textual elements, from the original media content, and composes these elements into a text overlaid image by maximizing information perception. In particular, we investigate a set of aesthetic rules and visual perception principles to optimize the layout of the extracted elements by considering display constraints. As a result, browsing the snippet on mobile devices is just like quickly glancing at a magazine. To the best of our knowledge, this paper represents one of the first attempts at automatic social media snippet generation by studying aesthetic rules and visual perception principles. We have conducted experiments and user studies with social posts from news entities. We demonstrated that the generated snippets are effective at representing media content in a visually appealing and compact way, leading to a better user experience when consuming social media content on mobile devices.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems—*Evaluation/methodology*; I.4.7 [**Image Processing and Computer Vision**]: Feature Measurement—*Feature representation*

General Terms

Algorithms, Design, Experimentation, Human Factors.

Keywords

Social Media, Snippet Presentation, Multimedia Authoring, Mobile Browsing, Responsive Design.

MM'13, Oct. 21–25, Barcelona, Spain.
Copyright 2013 ACM 978-1-4503-2404-5/13/10 ...$15.00.
http://dx.doi.org/10.1145/2502081.2502116.

(a) (b) (c) (d)

Figure 1: Example of traditional social media posts on mobile devices and the snippet generated by the proposed system: (a) twitter post, (b) post details, (c) linked article of the post, and (d) proposed snippet.

1. INTRODUCTION

Social media networks, such as Facebook, Twitter, and Google+, has continued to revolutionize people's lives since the emergence of the first social media networks two decades ago. A huge number of people and organizations are actively sharing news, experiences, and opinions via social media. The overwhelming amount of information on social media networks and the distraction caused by the large number of irrelevant social posts makes it difficult and time-consuming to find relevant information, leading to an unintended consequence: information overload.

Meanwhile, the proliferation of mobile devices and rapid development of wireless networks enable people to use their mobile devices to explore the Internet anytime and anywhere. People are spending more time than ever accessing social media content via their mobile devices. The consumption of social media on mobile devices has become pervasive.

Although the pervasiveness of mobile devices and a variety of mobile social media applications have boosted the usage of social media, the limited display size of mobile devices remains an obstacle for mobile users to peruse social posts and efficiently find interesting information. This issue can be demonstrated, as in the example in Fig.1(a), in which there is a post from the Wall Street Journal on Twitter as circled, 'Ang Lee on his Oscar win and the elusive best-picture statue: on.wsj.com/ZDJ1FW.' However, it does not stand out from the surrounding posts for users who are interested in reading it. With a small screen, mobile users have to scan the posts one by one to discover interested posts, making it a time-consuming task on mobile devices. Moreover, mobile users occasionally browse social posts in between activities (e.g., when waiting for or taking buses). It is distracting for mobile users to read the posts to find their preferred content. Because of the above challenges, it is difficult for mobile users to discover interested information from the abundance of social media content.

In addition to the mobile display constraints, the limit in the number of characters in social posts restricts the amount of information and thus hinders mobile users from obtaining important information that interests them. For example, with the 140 character limit of each post in Twitter, a user often posts a short message with additional links, as with the highlighted instance shown in Fig. 1(a). If a viewer is interested in the post, he will have to click the post for details, such as what award Ang won, as shown in Fig. 1(b). The viewer then has to click the link to get the full article as shown in Fig. 1(c) for more information, e.g., for which movie Ang was awarded.

1.1 Challenges in Social Media Presentation

The challenges when browsing social media on mobile devices can be summarized as follows.

- First, social posts in traditional text summary form are unattractive and unappealing. Moreover, the text is difficult to perceive on the limited display. Therefore, interesting posts may not stand out from the surrounding posts.
- Second, it is not convenient for mobile users to laboriously read wordy sentences in the post and linked articles in traditional forms of social posts on mobile devices to obtain important information.
- Third, The limited display of mobile devices has not been sufficiently utilized to deliver key information to mobile users. The social media content should be well organized to efficiently provide useful information in the social post summary page and thus help users obtain desired information with a quick glance.

1.2 Motivation and Contributions

The multimodality of social media with prevalent images or video clips creates a good opportunity to help mobile users quickly understand social content. Images are superior at capturing attention [16, 17, 19]. The success of adoption of images in search results presentation provides great evidence that images can indeed help quicken understanding of webpage content and relevance judgment [15, 23]. This motivates us to generate social media snippets for mobile social media browsing by utilizing the images in social posts to help mobile users more quickly understand content. In addition, it is found that usability and aesthetics are highly correlated [2]. Moreover, the research in eye tracking [19, 22] and aesthetic study [1, 13] supply us solid theoretical foundations on how people browse and perceive media. Hence, we believe it is the proper time to build an automatic social media snippet generation system based on aesthetic rules and visual perception principles to help mobile users perceive social posts in an appealing way and thus obtain key information efficiently.

In this paper, we propose a system to automatically build attractive, appealing, and informative snippets in text-overlaid image styles for browsing social media on mobile devices. To discover key information in the posts and links, social media snippet elements, including dominant images and key phrases, are first extracted via a modified TextRank algorithm [18]. In this paper, we assume that a dominant image exists and its content is relevant to the social post, which is usually valid for social media posts. Addressing the missingness of the dominant image is beyond the scope of this paper and can be addressed in the future work by introducing external image as in [10]. Then, the extracted elements are systematically selected and composed to generate snippets based on aesthetic rules and visual perception principles, under the display constraints. Compared with existing systems such as Flipboard [7], which uses fixed patterns, the innovation of the proposed system is that we incorporate well studied aesthetic rules and perception principles to automatically generate attractive, appealing and informative social media snippet with text-overlaid image as shown in Fig.1(d) by formulating it as an optimization problem.

In summary, this paper makes the following contributions:

- First, we propose a system to automatically generate social media snippet to extract and present compact and dominant elements in multimodalities. The mobile users can perceive key information quickly on limited mobile displays.
- Second, we employ human visual perception principles and aesthetic rules to actively push important information to mobile users to expedite the discovery of interesting information.
- Last, we formulate the social media snippet generation process as a constrained energy minimization problem. In addition, we develop a multi-stage optimization procedure to tackle the search complexity, rather than attempt at an expensive and integrated optimization procedure.

The rest of this paper is organized as follows. Section 2 reviews related work. Section 3 presents our proposed social media snippet generation system. Section 4 describes experiments and user studies, followed by the conclusion in Section 5.

2. RELATED WORK

The research community has made some attempts at developing summarization and browsing systems. For example, a system called Eddi was developed to organize a user's Twitter feeds into coherently clustered trending topics and provide interactive topic-based browsing of social status streams in Twitter [3]. TwitterStand is a news processing system built upon tweets to capture late breaking news [21]. Another system deals with text-to-image summary, which automatically illustrates complex sentences as multimodal summaries that combine pictures, structure, and simplified compressed text [24]. Adding structured compressed text to pictures was shown to be more effective than either text or images alone. In [25], the authors developed a rich-media analysis system to sense and explore events from social media and allow magazine-style event visualization anchored on event facets. Although they visualize events in a magazine style, the visualization is based on some pre-designed templates and color schemes. Hence, user perception and aesthetic issues have not been systematically considered.

Some mobile applications, such as Flipboard [7], have been successfully developed to present social media in an appealing way in a magazine style to improve the viewing experience. When displaying a social post on smartphones, the post is fit into pre-defined templates where the title or summary is placed in a fixed position, scale, and color. Although it is more visually appealing than traditional social post presentations, without the consideration of content analysis and user visual perception, it still falls short in presenting maximum allowable information under mobile displays.

In a search scenario, instead of presenting search results in a traditional text snippet form, image excerpts [15] and visual snippets [23] are explored. It has been demonstrated that these webpage summarization forms with images can indeed help users more quickly and accurately understand the content for quick relevance judgment. In [15], properly selected pictures in webpages are added to search result pages to generate a webpage summary called "image excerpt" and it has been demonstrated that image excerpt can

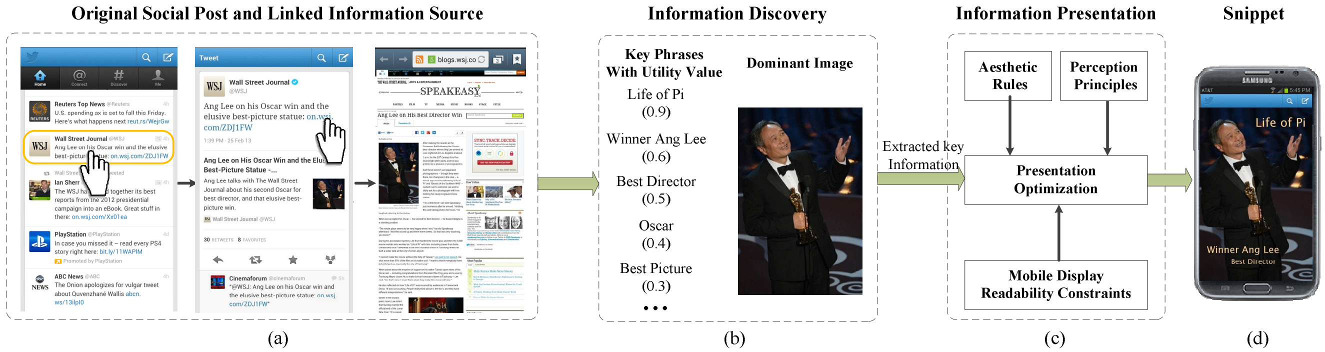


Figure 2: Social media snippet generation system overview. (a) original social post and linked information source (b) information discovery module (c) information presentation module (d) generated social media snippet.

significantly outperform traditional text snippets in terms of both search time and click numbers. Visual snippet is proposed to summarize webpages with title, image, and logo using a fixed template without consideration of how the pieces will be composed, (e.g., where to place the title on images to guarantee or enhance readability, visibility, and usability) [23]. Although these webpage summarization methods are designed for general search results presentation scenarios on desktop PCs, the successful usage of images in the summarization task in terms of visually attractiveness, ease of use, and quick understanding, demonstrate the effectiveness of images in summaries. Considering the text reading obstacle on small mobile devices, images will definitely benefit readability.

On the other hand, aesthetic studies in a variety of research areas, such as photographs [6], webpages [26], interfaces [1], and text-overlaid images [13], have drawn attention in recent decades due to their high correlation with usability. A great deal of research effort has been made to adopt the aesthetic rules in a wide range of applications, such as [12, 20]. To transform inherently dynamic social media into physical photo books to remember important moments, the authors utilize some known aesthetic principles such as the golden section and symmetry to generate appealing photo books automatically [20]. To automate magazine publishing for digital platforms, the authors develop a semi-automatic magazine layout system based on photography principles such as rule-of-space [12]. Both the aesthetic study and the successful adoption of aesthetic rules in various applications demonstrate the effectiveness of aesthetics on usability improvement.

In addition, studies on eye tracking for text-overlaid images in the advertising domain [19, 22] provide us theoretical and practical support on how people perceive text-overlaid images. In [22], the authors investigated how to improve return with various eye tracking studies such as how advertisements gain attention and how the composition of the advertisement elements influences the intended message delivery process. In [19], the authors studied how the advertisement elements (text, image, and brand) influence attention. The effect of the size of elements on attention is also studied. These studies on text-overlaid image perception and the intended message delivery process can be definitely adopted to improve information delivery for mobile users.

3. SOCIAL MEDIA SNIPPET SYSTEM

The proposed social media snippet generation system is composed of two modules: *information discovery* and *information presentation* as illustrated in Fig.2. First, information discovery is performed on the social posts to extract key phrases and dominant images as shown in Fig.2(b). Then, in the information presentation process the elements are systematically selected and composed considering visual perception principles, aesthetic rules and mobile display constraints to guarantee readability and maximize usability as shown in Fig.2(c).

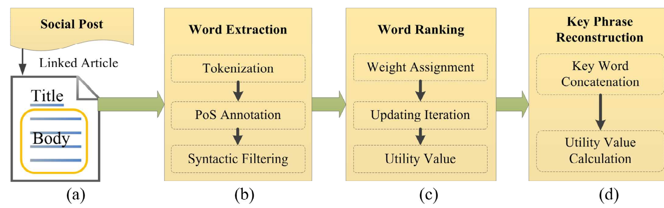


Figure 3: Key phrase extraction process: (a) original social post and linked article, (b) word extraction, (c) word ranking, and (d) key phrase reconstruction.

3.1 Information Discovery

To help mobile users obtain key information at a glance for quick content understanding, key phrases and dominant image are extracted from the social posts and linked articles in the information discovery process.

3.1.1 Dominant Image Extraction

Compared with pure text summary, images are more attractive, appealing, and informative from a phycological standpoint. Based on the study of search results presentation [15, 23], summaries with images can help users more quickly understand for faster relevance judgment. In our system, we focus on social posts in news categories with images. Based on our observations, the first image in the linked articles is usually dominant and utilized in the thumbnail as shown in Fig.1(b), so we extract the first image of the linked article as the dominant image for social media snippet generation. Learning approaches considering image position, content and surrounding texts can be adopted to refine the dominant image extraction in the future.

3.1.2 Key phrase extraction

Key phrases can express information concisely and accurately. Wordy sentences are difficult to view on mobile displays, while key phrases can stand out for mobile users who are interested in the post. Therefore, it is necessary to extract key phrases and determine their importance,i.e., utility values. We extract key phrases with a modified TextRank [18] approach. The key phrase extraction process is illustrated in Fig.3, which contains the following stages: word extraction, word ranking, and key phrase reconstruction.

Word Extraction. First, the social posts and linked articles are processed for key word extraction as in Fig.3(b). After tokenization and part of speech (PoS) annotation, we construct a graph by selecting nouns and adjectives as the vertices in the graph as in [18] and utilize the co-occurrence relationship controlled by the distance between words to generate the edges between vertices. Two vertices are connected if they co-occur within a window of $N_w (N_w = 2)$ words. The extracted text units and their relationships can be represented as an undirected graph $G = (V, E)$ with the set of vertices V and set of edges E.

Word Ranking. The word ranking process is illustrated in Fig.3(c). We follow a similar method to the one in [18] but take the higher importance of the post and the title of the linked article into consideration. We consider the superiority of the words in the post and title in terms of their confidence by weighting these vertices higher. The utility values of the vertices are decided by taking into account global information recursively from the entire graph by equation 1 until convergence. Here, we set the weights of vertices from the posts and titles as 1 and others as 0.2. Finally, the words from the posts and titles and important words highly related are ranked with high utility values. For a given vertex V_i, let $S(V_i)$ be the set of vertices that link to it. In contrast to the unweighted or the weighted edge graph in [18], we construct the graph with vertex weights. The confidence value of the vertex V_i is determined by its weight w_i. The utility value of a vertex V_i is defined as

$$U(V_i) = (1-d) + d \times \sum_{V_j \in S(V_i)} \frac{w_i}{\sum_{V_k \in S(V_j)} w_k} U(V_j) \quad (1)$$

where d is the damping factor, which integrate the probability of jumping from a given vertex to another in the graph. We set d to 0.85 as in [18]. The top $\frac{1}{3}$ vertices are retained for post-processing.

Key Phrase Reconstruction. In this stage, multi-word key phrases are reconstructed from the retained key words. The workflow is shown in Fig.3(d). All the potential lexical units selected are marked in the text and the key words that are adjacent or connected with a preposition and a conjunction are concatenated together. For example, as shown in Fig.2(b), the selected words 'life' and 'pi' will be concatenated into key phrase 'life of pi,' since they are connected with a preposition. Finally, the utility values of key phrases are calculated by taking the sum of the utility values of the key words within them and normalized.

3.2 Information Presentation

Once the key phrases and dominant image are extracted from one social post and linked article, in the information presentation stage we aim to carefully select the elements and compose them into an appealing and attractive text-overlaid image to maximize the information perceived at a quick glance. This is conducted by adopting visual perception principles, aesthetic rules, and mobile display constraints.

3.2.1 Problem Formulation

Given a set of key phrases and a dominant image extracted from one social post and linked article, the aim is to generate a social media snippet in text-overlaid image form with the following properties:

- The presented region of the dominant image and the selected key phrases are representative of the social post and the linked article, summarizing its main messages to be delivered.
- The generated social media snippet should follow key aesthetic rules of text-overlaid images to make it appealing and attractive to enhance usability. As systematically investigated in text-overlaid image designs [13], visual balance shows a stronger relationship to aesthetic appeal in text-overlaid images than other factors such as symmetry. Hence, visual balance is considered in the social snippet generation process.
- Visual perception principles should be utilized so that the key phrases are placed in attractive regions that the user's eye tends to focus on.
- Visual perception principles should be utilized to highlight the most important key phrases. That is, the visual weights of the key phrases are consistent with their utility values in the social media snippets.
- Perception constraints on mobile displays are considered to guarantee readability.

The input of the problem is the original dominant image I_o and the extracted N_t key phrases $\Gamma = t_1, t_2, \cdots, t_{N_t}$ already sorted in descending order by their utility values. Hence, $u_1 \geq u_2 \geq \cdots \geq u_{N_t}$, where u_i is the utility value of the key phrase t_i. The key phrase t_i contains n_i characters. Given the string of t_i and the look-up table for the aspect ratio of the 26 characters, the aspect ratio of the key phrase t_i , r_i can be calculated. Here, we only consider one font style for the key phrase insertion. The image I is discretized into $h \times w$ grids. Each grid has $g \times g$ pixels and we set $g = 8$. The grid located at (x, y) is denoted as (x, y). Let (x_i, y_i) be the grid coordinates of the center, h_i and $\overrightarrow{c_i}$ be the height and HSV color vector of the key phrase t_i. The goal is to find the optimal set of key phrases $\Gamma' = {t_1}', {t_2}', \cdots, {t_{N_{opt}}}'$ and search for the optimal position, size, and color for each ${t_i}'$ for all $i \in [1, N_{opt}]$.

Image Preprocessing. Image pre-processing is a prerequisite to avoid cropping or occluding important regions such as a face or salient objects. We perform saliency detection [8], face detection [14] and text detection [4] on the image. The pixel level importance map is computed by combining the face map, text map, and saliency map with a max operation. To deal with the aspect mismatch of the dominant image and the mobile display, the image is cropped to the maximum allowable size with the same aspect ratio of the mobile display by maximizing importance value. Then, the image I is discretized into $h \times w$ grids. The grid importance value $Imp(x, y)$ is obtained by the average importance values of the pixels within grid (x, y). To guarantee the non-intrusiveness of text insertion, the important regions have to be preserved. Hence, the importance map is binarized by a threshold. The binarized importance map is denoted as Imp', and the binarized importance value of grid (x, y) is $Imp(x, y)'$. Once the binarized importance map is obtained, each connected region is considered as a salient object for the following optimization process.

Energy Function. The presentation optimization problem is expressed as following. We aim to find the optimal set of key phrases $\Gamma' = {t_1}', {t_2}', \cdots, {t_{N_{Opt}}}'$ and their optimal position, size, and color that minimizes the energy $E(L)$. $L = L(i), i \in [1, N_t]$ completely specifies the text selection and insertion problem onto the image I. $L(i) = (a_i, (x_i, y_i), h_i, \overrightarrow{c_i})$ indicates the center coordinate (x_i, y_i), height h_i, and HSV color $\overrightarrow{c_i}$ of key phrase t_i. a_i is the indicator variable, taking the value 1 if the key phrase t_i is present in the social media snippet and 0 otherwise.

The energy of L comprises four terms as follows:

$$E(L) = E_r(L) + w_b E_b(L) + w_a E_a(L) + w_c E_c(L). \quad (2)$$

The representativeness term E_r evaluates how representative and informative the key phrases selected from the set of extracted key phrases Γ are. The visual balance term E_b measures how well the generated text overlaid image follows the dominant, visual balance aesthetic rule. The attention attractive term E_a measures how much users are inclined to focus on the key phrases in attractive regions. The consistency term E_c represents how consistent it is between the perception on the key phrases and their importance, i.e. utility values, to highlight the most important messages and hence improve information perception. w_b, w_a, and w_c are the weights of the energy term E_b, E_a, and E_c, respectively. Each of the energy terms and mobile display constraints will be described as follows in more details.

Representativeness. We consider that the utility values of the key phrases represent their representativeness in terms of information deliver. Hence, the representativeness term E_r is defined as follows:

$$E_r(L) = -\sum_{i=1}^{N_t} a_i u_i. \tag{3}$$

This energy term is a function of all indicator variables $a_1, \cdots, a_{N_t}$. It will favor the most representative key phrases and also encourage the use of as many key phrases as possible.

Visual Balance. Visual balance is a key aesthetic principle in interface design [1] and it has been proven to show a strong relationship with aesthetics in text-overlaid images in [13]. Hence, this aesthetic rule is considered to guarantee the aesthetics and usability of the generated social media snippets. As in [1, 13], visual balance is achieved when the center of gravity of all objects and text areas exactly located at the center of the text-overlaid image. In other words, the center of gravity of the selected key phrases should locate at its expected coordinate $\widehat{C_t}$, so that

$$\frac{\sum_{i=1}^{N_o} w_i C_o + \sum_{j=1}^{N_t} a_j w_j \widehat{C_t}}{\sum_{i=1}^{N_0} w_i + \sum_{j=1}^{N_t} a_j w_j} = C_m. \tag{4}$$

C_m is the center of the text-overlaid image. N_o is the number of objects within the image. C_o is the center of gravity of all objects, which can be obtained by

$$C_o = \frac{\sum_{i=1}^{N_o} w_i \cdot o_i}{\sum_{i=1}^{N_o} w_i}, \tag{5}$$

where o_i is the center of gravity of the i^{th} object. The term w_i is the visual weight of the i^{th} object and it is defined as the sum of the importance values of the grids occupied by the i^{th} object Obj_i. w_j is the visual weight of the j^{th} key phrase. We consider the two dominant factors of text visual weights in text overlaid images [13]. The visual weights of key phrases in the text-overlaid image depend on the area size of key phrases and their color contrast against the background color. The visual weight of the inserted key phrase is defined as the summation of the color contrast of all grids occupied by the j^{th} key phrase as follows:

$$w_j = \sum_{x=x_j-\lfloor h_j/2 \rfloor}^{x_j+\lfloor h_j/2 \rfloor} \sum_{y=y_j-\lfloor \lceil h_j r_j \rceil /2 \rfloor}^{y_j+\lfloor \lceil h_j r_j - 1 \rceil /2 \rfloor} \triangle C(\overrightarrow{c_j}, \overrightarrow{c(x,y)}), \tag{6}$$

$\overrightarrow{c(x,y)}$ is the dominant color of grid (x, y) in the HSV space and the color contrast $\triangle C(\overrightarrow{c_x}, \overrightarrow{c_y})$ of two colors $\overrightarrow{c_x}$, $\overrightarrow{c_y}$ is defined in the HSV color space, as in [13].

From equation (4) we can obtain the expected coordinates of the center of gravity of the selected key phrases:

$$\widehat{C_t} = \frac{(\sum_{i=1}^{N_o} w_i + \sum_{j=1}^{N_t} a_j wj) C_m - \sum_{i=1}^{N_o} w_i Co}{\sum_{j=1}^{N_t} a_j w_j}. \tag{7}$$

The maximum visual balance can be achieved by minimizing the distance between the expected center of gravity of selected key phrases $\widehat{C_t}$ and the actual center of gravity C_t,

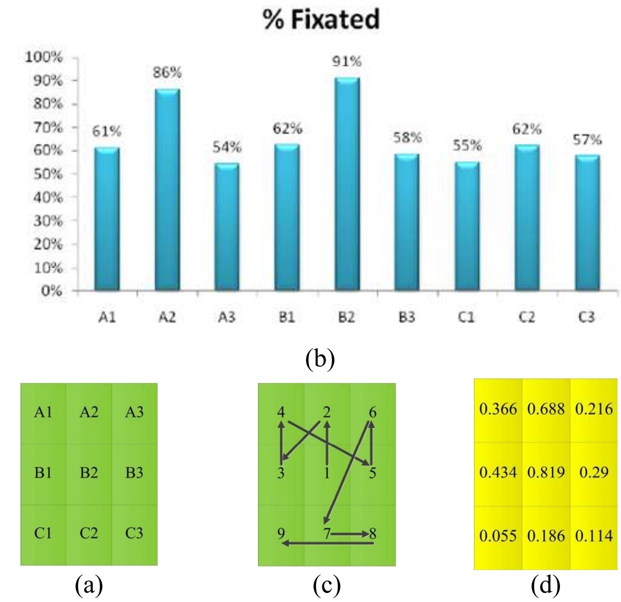


Figure 4: Eye tracking study results in [22] (a)nine areas (b)average fixation percentage (c)average scan path (d)fixation map.

$$C_t = \frac{\sum_{j=1}^{N_t} a_j w_j \cdot o_j}{\sum_{j=1}^{N_t} a_j w_j}, \tag{8}$$

where o_j is the center of gravity of the j^{th} key phrase. Since the text color is uniform, we consider $o_j = (x_j, y_j)$. Therefore, the term E_b is defined as the normalized distance between C_t and $\widehat{C_t}$ by

$$E_B(L) = \frac{d(C_t, \widehat{C_t})}{\sqrt{h^2 + w^2}}, \tag{9}$$

where $d(C_t, \widehat{C_t})$ is the distance between C_t and $\widehat{C_t}$. We have found that the term E_B depends on the position and visual weight of the selected key phrases, and the visual weights of the selected phrases further depend on their size and color contrast with the background color in the text location. Hence, this energy term is a function of all variables of L, i.e., $L(i) = (a_i, (x_i, y_i), h_i, \widehat{c_i}), \forall i \in [1, N_t]$.

Attractiveness. To measure attractiveness, we follow the eye tracking study results in [22], in which the average percentage of fixation and the average scan path on the divided nine equal areas of text-overlaid images are investigated. The areas are labeled as shown in Fig. 4(a). The average percentage for each area is demonstrated in Fig. 4(b), while the scan path using the average time for the first fixation is shown in Fig. 4(c) in which the numbers and arrows represent the average scan path from the first fixations to the last fixations with 1 to 9 respectively.

To adopt visual perception principle, we fuse the fixation percentage and fixation order to generate the fixation map as follows:

$$\begin{gathered} F(l) = Fp(l) \cdot (1 - 0.1 \cdot Fo(l)) \\ where, l = A_1, A_2, A_3, B_1, B_2, B_3, C_1, C_2, C_3. \end{gathered} \tag{10}$$

The areas fixated earlier are weighted higher. l is the index of the nine areas. $F(l)$ is the value on the fixation map of area l. $F_p(l)$ is the fixation percentage of area l as shown in Fig. 4(b) and $F_o(l)$ is the fixation order of area l as shown in Fig. 4(c). The generated fixation map is demonstrated in Fig. 4(d) with the fixation value indicated.

The goal of this energy term is to place the selected key phrases onto attractive regions and moreover, the more important key phrases should be inserted into more attractive areas that are fixated earlier for a longer period so that the key phrases are able to be viewed quickly for expediting information deliver. The energy term E_a is defined as follows:

$$\begin{aligned} E_a(L) &= 1 - \frac{\sum_{i=1}^{N_t} a_i u_i \frac{Att(x_i,y_i)-Min_{Att}}{Max_{Att}-Min_{Att}}}{\sum_{i=1}^{N_t} a_i u_i} \\ where, Min_{Att} &= \min_{1\le x\le h, 1\le y\le w} Att(x,y), \\ Max_{Att} &= \max_{1\le x\le h, 1\le y\le w} Att(x,y). \end{aligned} \tag{11}$$

$$Att(x,y) = F(l), \text{ if } (x,y) \in l, \tag{12}$$

where $Att(x,y)$ is the attractiveness value of position (x,y). By minimizing E_a, the attractiveness of the key phrases is maximized, while the utility value u_i is taken into account to insert more important key phrases into more attractive areas. The term $Max_{Att} - Min_{Att}$ is to make E_a range from 0 to 1. The term E_a is a function of the indicator variables $(a_1, \cdots, a_{N_t})$ and the positions of the key phrases $(x_1,y_1), \cdots, (x_{N_t}, y_{N_t})$.

Consistency. The more important phrases with higher utility values should be composed with larger visual weights. That is, the visual weights of the key phrases have to be proportional to their utility values. Hence, the energy term E_c is defined by

$$E_c(L) = \frac{\sum_{i=1}^{N_t} a_i |w_i - k u_i|}{hw}, \tag{13}$$

where k is the coefficient that will be discussed later. The energy term E_c is normalized by the total area of the image since the largest possible value of w_i is the area occupied by the key phrase t_i in the text-overlaid image. We have found that E_c depends on the visual weights of the key phrases, and thus depends on their size, color, and positions as described in equation 6. Hence, E_c is a function of all variables of L, i.e. $L(i) = (a_i, (x_i, y_i), h_i, \overrightarrow{c_i}), \forall i \in [1, N_t]$.

Constraints. The optimization of $E(L)$ is done under certain constraints, especially considering the small display and perception issues on mobile devices as listed below:

1. *Minimum font size.* To guarantee the readability of the key phrases on mobile displays, the font size has to be larger than a threshold T_s, which is three times the size of the grid.
2. *Maximum number of characters.* Considering the information perception on mobile displays, long sentences or too many words in the text-overlaid images will hinder the information capture of mobile users. The maximum number of characters is set to 39 according to the study in [11].
3. *Non-intrusiveness.* If the insertion of key phrases occludes salient objects or faces, the information perception on both the objects and the texts will be degraded and unappealing.
4. *Text area to image area ratio.* Too large a text area makes the text-overlaid image messy and unappealing. The principle in text-overlaid images, such as the cover pages of magazines that designers usually follow, indicated that the optimal text area to image area ratio is 1:4.
5. *Non-overlapping.* The key phrases inserted into the image I should not overlap with each other.
6. *Non-border crossing.* The key phrase insertion should not cross the border of image I.
7. *Color harmony.* The notion of color harmony has been studied by artists and scientists [9]. Eight color schemes have been discovered to quantify color harmony [5]. Due to the computation complexity, instead of searching through the whole color space using the color schemes, we carefully select from existing colors along with black and white for the key phrases coloring.
8. *Text size consistency constraint.* As discovered in [19], the attention to text elements in text-overlaid images is in proportion to their size.

In the following subsection, an optimization procedure is defined that searches efficiently in the allowed space to obtain the solution with low energy, but since the algorithm is approximate, it is not necessarily the global minimum.

3.2.2 Energy Minimization

The search space for optimization of the energy $E(L)$ defined in the previous subsection, is the entire space of L, i.e. $L(i) = (a_i, (x_i, y_i), h_i, \overrightarrow{c_i}), i \in [1, N_t]$, which results in a huge search space for the optimization. By analyzing the dependency characteristics of the energy terms, a heuristic but effective approach is developed in which the key phrase selection, and the position, size, and color of the key phrases are optimized in sequence. The optimization steps are described as follows.

Key phrase selection. The key phrase selection step in the sequence of optimizations, addresses the E_r term in the energy function equation 2. With the utility values and the number of characters of the N_t key phrases, the goal is to maximize the total utility values of the selected phrases under the maximum number of characters constraint, that is,

$$\max_{a_1, \cdots, a_{N_t}} \sum_{i=1}^{N_t} a_i u_i, s.t. \sum_{i=1}^{N_t} a_i n_i \le T_c. \tag{14}$$

The key phrase selection is converted to a 0-1 knapsack problem which can be solved efficiently by dynamic programming. Hence, the set of indicator variables $a_1, \cdots, a_{N_t}$ are solved in this step.

Position optimization. The energy term E_b and E_c both rely on the visual weights of the key phrases in the text-overlaid image, and their visual weights depends on their size and color contrast with their local surroundings as defined in equation (6). Therefore, the positions of the selected key phrases have to be solved first. Their positions are determined by the energy term E_b and E_a, and thus can be obtained by optimizing the term $w_b E_b(L) + w_a E_a(L)$ in the energy function (2). Here, we set $w_b = w_a = 0.5$. However, without determining the size and color of the selected key phrases, E_b is not able to be calculated. Considering the consistency energy term E_c, the term $w_b E_b(L) + w_a E_a(L)$ is optimized by assuming

$$w_i = k u_i, \forall i \in [1, N_t] \; s.t. \; a_i = 1. \tag{15}$$

In this way, the optimal positions of the selected key phrases $(x_i, y_i), \forall i \in [1, N_t] \; s.t. \; a_i = 1$ can be calculated under the non-intrusiveness, non-border crossing, and non-overlapping constraints with the assumption that $h_i = T_s, \forall i \in [1, N_t] \; s.t. \; a_i = 1$. At this step, taking the text area to image area ratio constraint, we set $k = \frac{0.2hw}{\sum_{i=1}^{N_t} a_i u_i}$, so that the selected phrases can split the optimal allowable text area according to their utility values. Note the maximum possible value of the visual weight of a key phrase is the area occupied by the key phrase in the text-overlaid image according to the definition of visual weight in equation 6, since the color contrast $\triangle C$ ranges from 0 to 1.

Size and color optimization. The size optimization together with the following color optimization step will minimize the energy term E_c. Since visual weights depends on the key phrase size and color contrast with local backgrounds, once the positions of the phrases are obtained from the position optimization, the visual weights can be directly achieved via size and color optimization. In this way, the size and color of each selected key phrase can be solved under the minimum font size, non-intrusiveness, color harmony, text area to image area ratio, and text size consistency constraints. It is possible that the initial solution would violate the non-overlapping or non-border crossing constraints. In that case, we iterate the position optimization as well as the size and color optimization several times, until all of the constraints are satisfied. Usually, the solution can be obtained within ten iterations.

3.2.3 *Complexity Analysis*

In this section we analyze the complexity of the proposed scheme. If we take the size of the key phrase set, the location, the size, and the color of the key phrases into consideration, the total search space of the proposed problem is $O(2^{N_t}) \cdot O((hw)^{N_t}) \cdot O(h^{N_t}) \cdot O(256^{3 \cdot N_t})$, which makes it almost unfeasible to carry out a brute-force search, even if the number of selected phrases is small. The proposed staged solution greatly reduces the complexity to $O(N_t^2) + O((hw)^{N_{Opt}}) + O(h^{N_{Opt}}) + O(C^{N_{Opt}})$, where C is the color space of image I. In real applications, N_{Opt} is usually no more than 3 and hence the search complexity is reduced significantly.

4. EXPERIMENTS

In order to evaluate the performance of the proposed social media snippet generation system, we performed experiments and user studies on a large social media dataset. Due to privacy concerns, we collected 1,000 news social posts on various topics, including economics, politics, fashion, sports, and science and technology from nine entities on Twitter, including ABC news, The New York Times, USA TODAY, ABC World News, Wall Street Journal, NBC News, The White House, FoxNews Science and Technology, and CNN. The linked articles were obtained together with the collected posts. Some of the posts contained no images in their linked articles for which related external images could be utilized in the social media snippet generation by semantic analysis. For now, our aim is to validate the effectiveness of the proposed social media snippet presentation form and the developed social media snippet generation approach, so those posts without images are excluded. The valid dataset contains the remaining 683 social posts with their linked articles.

4.1 Evaluation of Key Phrase Extraction

We randomly sampled 100 social posts with their linked articles from the valid dataset to label their key phrases. Since key phrase extraction is subjective, each social post along with the linked article was labeled by three people independently. For each post, each person was required to extract at least the ten most important key phrases from the post and the linked article. The key phrases voted by at least two people were considered the ground truth.

In order to evaluate the performance of key phrase extraction, we compared the developed approach with two other methods. The first one was an intuitive rule-based method, which scores the words in the graph vertices based on frequency. The second one was the key phrase extraction method proposed in [18], which computes the utility value using an unweighted graph with uniform vertices weights. All three methods process the social posts and linked articles with the same tokenization and syntactic filtering procedure to build the graph vertices, but they rank the vertices in different ways. As in [18], the number of selected key vertices T was set to a third of the total number of vertices in the graph. The three methods had the same number of key vertices selected and processed by the same key phrase reconstruction method as shown in Fig. 3. The precision, recall, and F-measure of the proposed key phrase extraction method along with the two comparison methods are presented in Table 1.

Table 1: The precision, recall, and F-measure of key phrase extraction with the frequency-based approach (frequency), the unweighted graph-based approach in [18] (unweighted), and the proposed weighted graph-based approach (proposed).

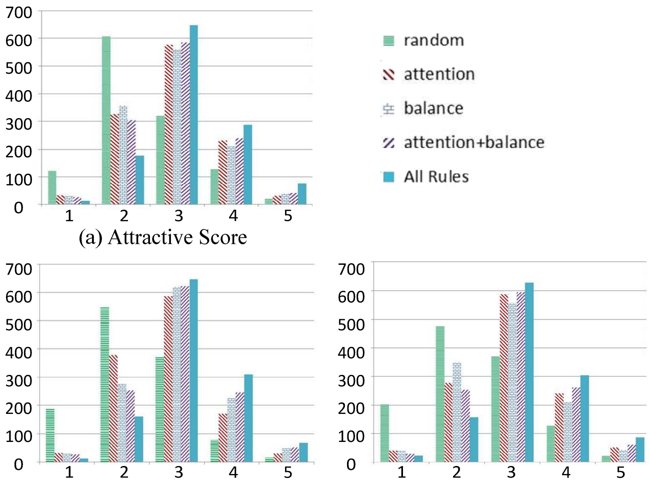
	Frequency	Unweighted [18]	Proposed
Precision	23%	37%	58%
Recall	35%	32%	34%
F-measure	28%	34%	43%

We found that by considering the superior confidence value of the words appearing in the post and titles, the precision of the developed weighted graph-based method was significantly improved compared with the frequency-based and unweighted graph-based approach. The recall was similar to the frequency-based and unweighted graph-based approach, since some key words were not connected with or far from the highly weighted vertices and thus could be discarded. However, in the social media snippet generation system, precision matters more than recall, because it is important to discover the most key phrases for the snippet generation, while the discovered phrases are not able to be shown all together due to mobile display constraints. In addition, the proposed approach achieved the highest F-measure compared with the other two methods.

4.2 Evaluation of Information Presentation

In order to evaluate whether the adoption of the aesthetic rules and visual perception principles is proper as well as to validate the proposed optimization process, we conducted a user study. We randomly selected 100 social posts along with their linked articles from the collected 683 posts. Once the data was ready, it was parsed to extract key phrases with their utility values and dominant images. Then, the extracted key phrases were selected by maximizing their total utility value under the number of characters constraint as illustrated in the key phrase selection process. We generated five snippets with the same required display aspect ratio in the following ways with the set of selected key phrases for comparison. All of the snippets were generated under the basic non-intrusiveness, non-overlapping, and non-border crossing constraints.

1. Snippet Set 1: We generated the snippets by randomly setting position, size, and color of the phrases.
2. Snippet Set 2: We generated the snippets by setting equal font height and the same color for the phrases while optimizing position based solely on the attractiveness term E_a.
3. Snippet Set 3: We generated the snippets by setting equal font height and the same color for the phrases while optimizing position based solely on the visual balance term E_b.
4. Snippet Set 4: We generated the snippets by setting equal font height and the same color for the key phrases while optimizing position based on the visual balance and attractiveness term $w_b E_b(L) + w_a E_a(L)$ with equal weights.
5. Snippet Set 5: We generated the snippets via optimization based on all aesthetic rules including visual balance and color



(a) Attractive Score

(b) Appealing Score

(c) Informative Score

Figure 5: The score distribution on the five sets of social media snippets based on different aesthetic and perception principles: (a) score distribution on attractiveness, (b) score distribution on appeal, and (c) score distribution on informativeness.

harmony as well as visual perception principles such as visual consistency and attractiveness as described in the energy minimization subsection in information presentation.

12 subjects with age ranging from 24 to 32 were invited, of which seven were females. To make the comparison fair, for each social post, the corresponding five social snippets were presented at the same time on the screen of a desktop PC with the same size of a smartphone model, i.e., 4.8 inch Samsung Galaxy S3. The original social posts and links were also presented. The subjects were allowed to click the linked webpage for more information. During the study, the order of the five snippets was shuffled so that the subjects were able to give fair rates. The subjects were asked to grade the five snippets for each social post according to the following aspects.

1. How attractive is the snippet, i.e., how good is the text on the image at drawing their attention?
2. How visually appealing is the snippet?
3. How informative is the snippet in terms of guiding user focus to the most important information?

The subjects were requested to rate the snippets with an integer from 1 (definitely no) to 5 (definitely yes). The score distribution of the 12 subjects on the five set of snippets were demonstrated in Figure 5. Some visual examples of the five generated social media snippets along with their average scores on the three questions are shown in Fig. 8.

Regarding the attractiveness in the Question 1, the randomly generated snippets had low rates, since the randomly selected position, size, and color for key phrases were sometimes unsuitable and hence made it difficult to find the texts. By placing the texts into regions that people tend to focus on via optimized attention attractiveness term E_a, the score of the snippet set 2 was higher than that of randomly generated ones on average. Comparing the score of set 2 and set 3, the number of rates at 1 and 5 were similar, and the number of rates at 3 and 4 of set 2 were larger than set 3 while the number of rates at 2 was smaller. Hence, the attractiveness of set 2 outperformed set 3 on average. With regards to the adoption of consistency perception rules, set 5 outperformed sets 2, 3, and 4.

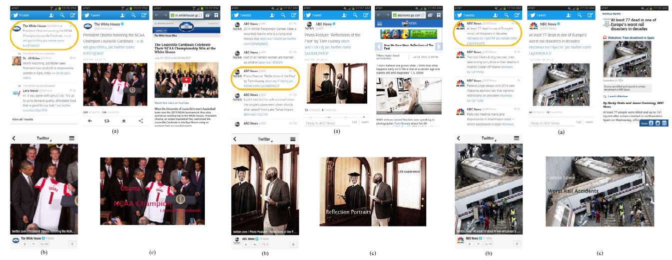


Figure 7: Examples of (a) the original social posts, details and linked articles, (b) Flipboard results, and (c) generated social media snippets with the proposed system.

Hence, besides the introduction of the attention attractiveness perception rule, the visual weights consideration by optimizing text size and color improved the attractiveness further.

Regarding the appeal in the Question 2, a large portion of randomly generated snippets were rated with 2 and 3, due to the unsuitable color introduced for text insertion, while a small portion were rated 4 or 5 probably because the random solution happened to introduce the proper color at suitable positions, helping them to look appealing. The adoption of aesthetic rule visual balance made the appeal score of set 3 and 4 higher than set 1 and 2 on average. The harmonic color introduced in set 5 further improved the appeal score compared to set 3 and 4.

For the informative judgment in the Question 3, the randomly generated position, size, and color may have degraded the eye capturing process on important phrases, hence the score concentrated at 2 and 3. The position optimized by attention attractiveness term made the informative score of set 2 much higher than set 1. The proper size and color optimization on the consistency term made the important key phrase salient, and hence insured the score of set 5 was higher than set 2.

4.3 Evaluation of Social Media Snippet

To validate the effectiveness of the proposed social media snippet presentation form and the developed system, we compared the generated social media snippets by the proposed system with the traditional Twitter post presentation and the summarization form with Flipboard on smartphones.

Another randomly selected 100 social posts and their linked articles were processed through the information discovery and information presentation modules as described in Section 3. Another 12 subjects were invited to browse the original social post, the Flipboard results, and the generated social media snippets on a Samsung Galaxy S3 smartphone. For each post, the snippet, the Flipboard result, and the original social post were displayed by flipping the pages. The subjects were allowed to browse forward and backward for comparison. They were also allowed to access the linked webpages of the social post. The subjects were asked to rate the generated snippet, the Flipboard result, and the corresponding original post by answering the following questions:

1. How attractive is the social post presentation as a whole?
2. How visually appealing is the social post presentation?
3. How good is the information delivery for the social post presentation?
4. How good is the readability of the social post presentation?
5. What is your overall evaluation on the snippet?

The subjects were requested to rate with a satisfaction score with the scale of 1–5 (higher score indicating higher satisfaction). The

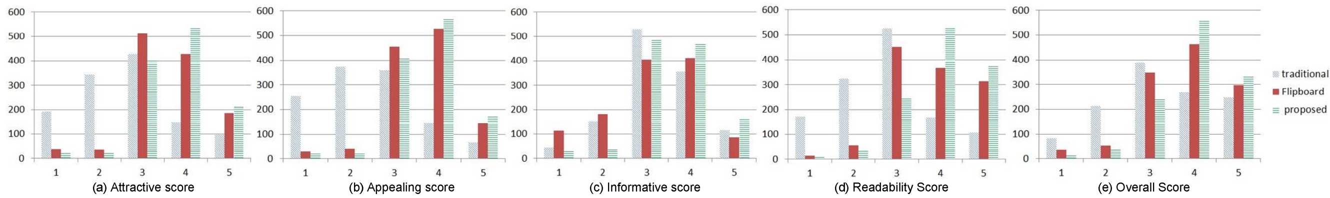


Figure 6: The score distribution of the social posts in the traditional form, Flipboard style, and the social media snippet.

score distributions of the 12 subjects on the 100 social posts in the traditional presentation form, Flipboard results, and the generated social media snippets are presented in Fig. 6. Some examples of the original social posts, the corresponding Flipboard results, and the generated social media snippets are shown in Fig. 7. We found that the social media snippets generated by the developed system significantly outperformed the traditional presentation form and Flipboard results in terms of attractiveness score, appeal score, and informative score. They were easier to read on mobile devices and preferred by mobile users.

5. CONCLUSION

To overcome the mobile social media browsing obstacles under mobile environment constraints, we have proposed a social media snippet generation system to summarize easy-to-be perceived important information for mobile users with optimized text-overlaid images. To generate the snippets, key information from social posts and embedded links as well as dominant images are extracted and systematically selected and organized. By taking visual perception principles, aesthetic rules, and mobile display constraints into consideration, we optimized the layout of the extracted elements. Extensive experiments and user studies show that the generated social media snippets indeed improve attractiveness, appealingness, and informativeness for mobile browsing and hence are much preferred by mobile users.

In the future, external images can be utilized to generate social media snippets for posts without images by analyzing post content and relationships among posts and other sources online. In addition, topic discovering can be adopted to generate social media snippet collages to facilitate mobile social media browsing in facets considering post timestamp, relationship and importance, as well as perception principles and aesthetic rules for collages. Moreover, social content and mobile user preference need to be analyzed to rank the social posts for better snippet generation based on perception rules.

6. REFERENCES

[1] M. Bauerly and Y. Liu. Computational modeling and experimental investigation of effects of compositional elements on interface and design aesthetics. *Int. J. Human-Computer Studies*, 64(8):670–682, Aug. 2006.

[2] T. Ben-Bassat, J. Meyer, and N. Tractinsky. Economic and subjective measures of the perceived value of aesthetics and usability. *ACM Trans. Computer-Human Interaction*, 13(2):210–234, June 2006.

[3] M. S. Bernstein, B. Suh, L. Hong, J. Chen, S. Kairam, and E. H. Chi. Eddi: interactive topic-based browsing of social status streams. In *Proceedings of ACM symposium on User interface software and technology*, pages 303–312, 2010.

[4] X. Chen and H. Zhang. Text area detection from video frames. In *Proceedings of IEEE Pacific Rim Conference on Multimedia*, pages 222–228, 2001.

[5] D. Cohen-Or, O. Sorkine, R. Gal, T. Leyvand, and Y.-Q. Xu. Color harmonization. *ACM Trans. Graph.*, 25(3):624–630, July 2006.

[6] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Studying aesthetics in photographic images using a computational approach. In *Proceedings of European conference on Computer Vision*, pages 288–301, 2006.

[7] Flipboard. `http://www.flipboard.com`.

[8] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. In *Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[9] J. Itten. *The Art of Color: The Subjective Experience and Objective Rationale of Color*. A VNR book. Wiley, 1974.

[10] B. Jiao, L. Yang, J. Xu, and F. Wu. Visual summarization of web pages. In *Proceedings of ACM SIGIR conference on Research and development in information retrieval*, pages 499–506, 2010.

[11] 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.

[12] M. Kuhna, I.-M. Kivelä, and P. Oittinen. Semi-automated magazine layout using content-based image features. In *Proceedings of ACM Multimedia*, pages 379–388, 2012.

[13] C.-Y. Lai, P.-H. Chen, S.-W. Shih, Y. Liu, and J.-S. Hong. Computational models and experimental investigations of effects of balance and symmetry on the aesthetics of text-overlaid images. *International Journal of Human-Computer Studies*, 68(1-2):41 – 56, 2010.

[14] S. Z. Li, L. Zhu, Z. Zhang, A. Blake, H. Zhang, and H. Shum. Statistical learning of multi-view face detection. In *Proceedings of European Conference on Computer Vision*, pages 67–81, 2002.

[15] Z. Li, S. Shi, and L. Zhang. Improving relevance judgment of web search results with image excerpts. In *Proceedings of World Wide Web*, pages 21–30, 2008.

[16] T. Mei and X.-S. Hua. Contextual internet multimedia advertising. *Proceedings of the IEEE*, 98(8):1416–1433, 2010.

[17] T. Mei, X.-S. Hua, and S. Li. Contextual in-image advertising. In *Proceedings of ACM Multimedia*, pages 439–448, 2008.

[18] R. Mihalcea and P. Tarau. Textrank: Bringing order into text. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*, pages 404–411, 2004.

[19] R. Pieters and M. Wedel. Attention capture and transfer in advertising: Brand, pictorial, and text-size effects. *Journal of Marketing*, 68(2):36–50, 2004.

[20] M. Rabbath, P. Sandhaus, and S. Boll. Automatic creation of photo books from stories in social media. *ACM Trans. Multimedia Comput. Commun. Appl.*, 7S(1):1–18, Nov. 2011.

[21] J. Sankaranarayanan, H. Samet, B. E. Teitler, M. D. Lieberman, and J. Sperling. Twitterstand: news in tweets. In *Proceedings of ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 42–51, 2009.

[22] M. Smit. Optimising advertising roi on print in a rapidly changing economic and media landscape - the contribution of eye tracking now and into the future. In *Proceedings of Southern African Marketing Research Association Conference*, 2012.

[23] 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 *Proceedings of SIGCHI Conference on Human Factors in Computing Systems*, pages 2023–2032, 2009.

[24] N. UzZaman, J. P. Bigham, and J. F. Allen. Multimodal summarization of complex sentences. In *Proceedings of International Conference on Intelligent User Interfaces*, pages 43–52, 2011.

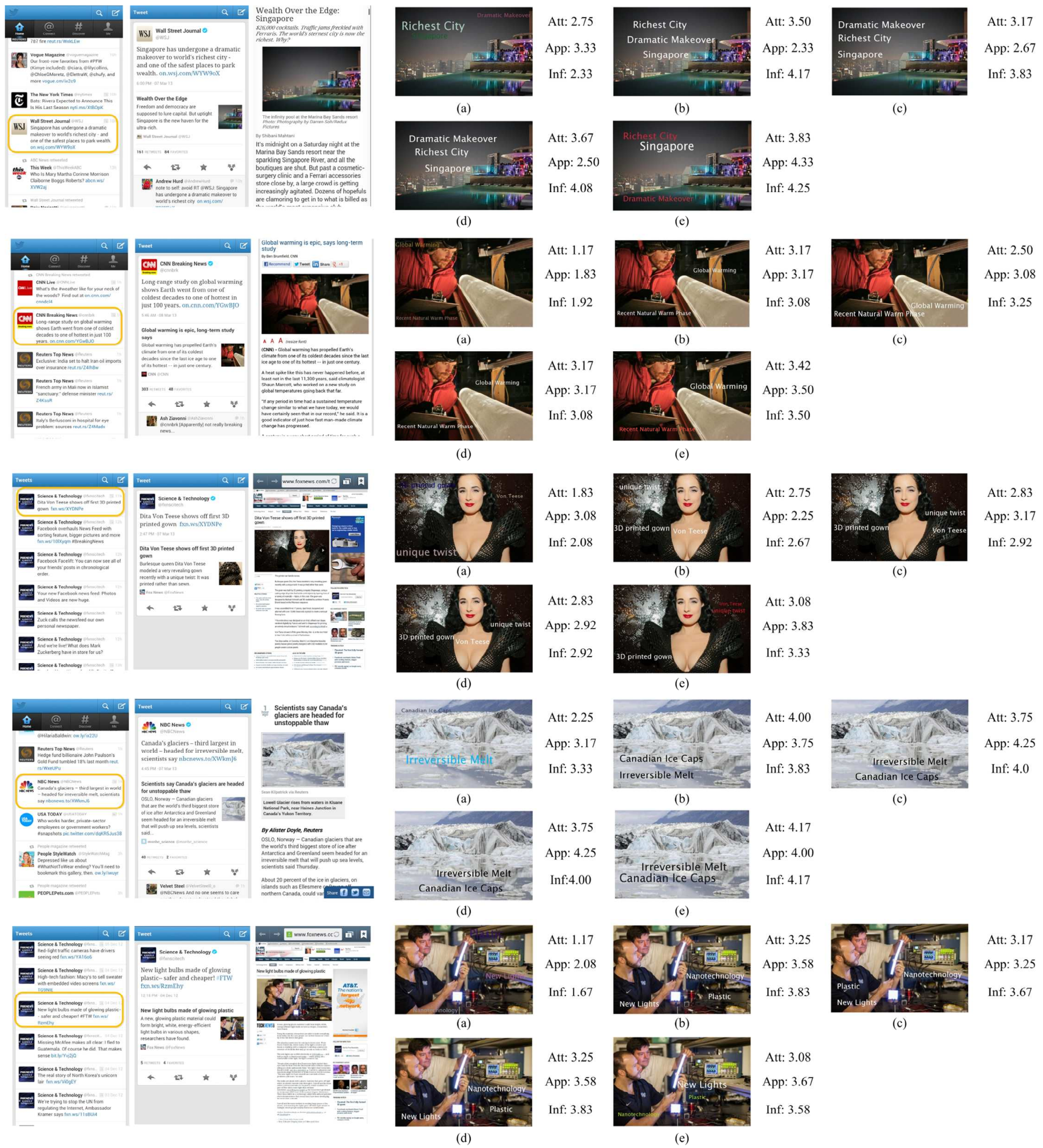


Figure 8: Visual examples of the original social media posts, details, and linked articles, as well as (e) the generated social media snippets with the proposed system and five other approaches, i.e., (a) random setting position, size and color of key phrases, (b) optimization based on solely attractiveness term, (c) optimization based on solely visual balance term, (d) optimization based on the attractiveness and visual balance.

[25] Z. Wang, P. Cui, L. Xie, H. Chen, W. Zhu, and S. Yang. Analyzing social media via event facets. In *Proceedings of ACM Multimedia*, pages 1359–1360, 2012.

[26] O. Wu, Y. Chen, B. Li, and W. Hu. Evaluating the visual quality of web pages using a computational aesthetic approach. In *Proceedings of ACM International Conference on Web Search and Data Mining*, pages 337–346, 2011.