AUTOMATIC CREATION OF MAGAZINE-PAGE-LIKE SOCIAL MEDIA VISUAL SUMMARY FOR MOBILE BROWSING

Shuang Ma, Chang Wen Chen

State University of New York at Buffalo, Buffalo, NY, USA

ABSTRACT

Today, mobile users are struggling with accessing overloading and unstructured social media feeds on the severely constrained mobile display. To overcome the challenges associated with browsing social media feeds on mobile devices, we are developing an innovative scheme to automatically create and synthesize the mixed social media digest (pictures, texts and videos) into a magazine-page-like social media visual summary. Given a set of personalized social media digest, a multi-objective optimization is formulated to organize the digest into visual summary in a 9-block-partition fashion with consideration of informative delivery, aesthetic rules and visual perception principles. Each block will be optimized interactively in terms of size, position and color to best represent the overall social media digest. Extensive evaluation and analysis based on user studies demonstrate that the proposed approach is effective in presenting social media content in a visually appealing and compact way. It is expected that this visual summary will lead to much enhanced user experiences for browsing social media digest on mobile devices.

Index Terms— Magazine-page design, visual summary, info maximization, aesthetically appealing, mobile browsing

1. INTRODUCTION

The recent proliferation of social media networks, such as Facebook, Twitter, Flickr, YouTube and Google+ has revolutionized people's daily life. Meanwhile, the overwhelming scales of social media consumption triggers information overload. Recently, a summarization and browsing system that can automatically select a set of representative images to generate a visual summary has been developed [1]. However, this system does not consider visual appealing attributes in its summary. A rich-media analysis system reported in [2] is capable of visualizing the events in a magazine style. However, the visualization is based on some pre-designed templates and color schemes without systematic consideration of user perception and aesthetic issues.

The severe limitation of mobile display makes it very difficult for mobile users to efficiently find relevant information from unstructured social media feeds. Recently, there have been some attempts for developing mobile browsing applications, such as Flipboard[3] and Snippet[4]. Flipboard fits social media contents 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, it still falls short in maximizing allowable information to present on mobile displays without proper content analysis and user perception considerations. Snippet can automatically generate magazine-like social media visual summaries for efficient mobile browsing. However, it focus on the design of text-overlaid single image, which limits the presentation of social media digest to only few items.

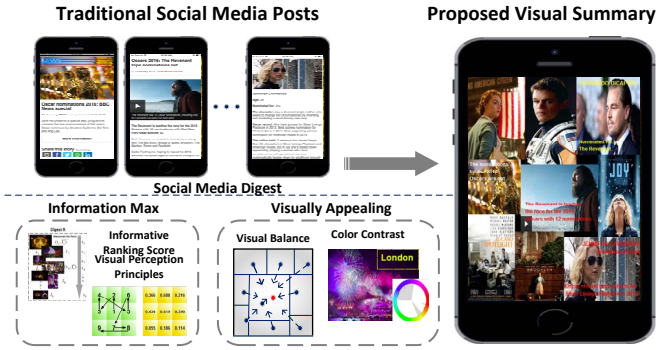

Fig. 1: Overview of the proposed system.

In recent years, aesthetic studies have been shown to have the potential to enhance user experience. A great deal of research efforts have been made to adopt the aesthetic rules in a wide range of applications, including photography recommendation[5–8], evaluation on webpages and photo quality [9][10], and text-overlaid images [11]. Furthermore, studies on eye tracking in the advertising research [12] provide us with both theoretical and practical supports on how people perceive interested information within a page. It is natural to adopt these studies to improve information delivery for mobile browsing.

The proposed system is different from previous works in that this system attempts to create the most informative and aesthetic visual summary from social media digest. In this study, we assume that the personalized social media digest as illustrated in [13] has been generated. That is, the summarization and ranking results on multimodal content-relevant

This research is supported by U.S. State of New York Empire Innovation Program.

social media feeds (images, texts and videos) are provided as the input to this system. As illustrated in Fig.1, given a set of personalized digest, we can automatically design and compose the visual summary by finding a proper partition of multiple blocks to capture and present the essence of multimedia contents in a magazine-like visually appealing page. In particular, by adopting aesthetic rules and visual perception principles, a multi-objective optimization is formulated to generate an optimal page layout under consideration of several inherent constraints. As a result, browsing such a visual summary on mobile devices is just like quickly glancing at a magazine page. To the best of our knowledge, this paper represents one of the first attempts at automatic social media presentation by applying aesthetic rules and visual perception principles.

2. PROBLEM FORMULATION

Based on the design principles as outlined above, the input to the proposed system is the personalized digest and the output is the generated visual summary, as shown in Fig.2. To generate proper input, a social media digest dataset was manually built, and the necessary preprocessing was also conducted. The details are described in the following.

Dataset building and preprocessing

We collected content-relevant social media feeds, including images, texts and videos retrieved from twitter, NY Times and Google+. Clustering of feeds has also been carried out. Here, video's key frame is used to represent link-embedded video. Therefore, each digest R in our data set is a combination of N topic-relevant items, including image I_i, text with its background images $\left(t_i, I_t^i\right)$, and video with its key frame $\left(v_i, I_v^i\right)$. In each R, these N extracted items $\Gamma\left(i\right) = \left(I_i, \left(t_i, I_t^i\right), \left(v_i, I_v^i\right)\right)$, $i = 1 \cdots N$ are ranked by a combination of post time and number of comments to identify the most important social media feeds as follows:

$$s_i = 1 - \exp\{-(\alpha \cdot time_i + \beta \cdot comments_i)\} \quad (1)$$

They are sorted in decreasing order by their ranking score, i.e. $s_1 \geq s_2 \geq s_3 \cdots \geq s_N$.

To identify and prevent possible occlusion in the most important regions, we performed saliency detection[14]. Let $O_i(x_i, y_i)$ be the center of the salient object Sal_i in I_i, which also represents the gravity center for I_i. We adopt the same scheme for text background image I_t^i and the video key frame I_v^i. Besides, considering the limit display space of the mobile devices, it is natural to adopt image cropping [15] to remove redundant information. The parameter set $((w_{i1}, w_{i2}), (h_{i1}, h_{i2}))$ represents I_i's four border lines' coordinates within page P, as shown in Fig.2.

Multi-Objective Optimization Function

As shown in Fig.2, we are able to represent each block's position and size by the parameter set. The visual summary to be created should have following properties:

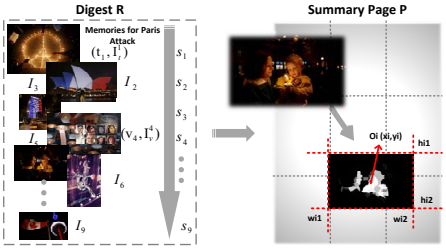

Fig. 2: Illustration of social media digest (left), and the summary page with parameter set (right).

• Multi-blocks layout. We shall initially divide the page into nine blocks based on "Rule of Thirds" principle; each block will be composed seamlessly within one page.
• Highly aesthetic display. Aesthetic composition principles employed in printed magazine design will be adopted to make the generated visual summary appealing and attractive in order to enhance user experience.
• Maximizing information perception. Human visual perception principles should be utilized so that the more important and popular feeds will be placed in more attractive locations which the users eyes tends to focus on. That is, the visual weights of the social media feeds in the generated visual summary should be consistent with their ranking scores.
• Satisfying mobile display constraints. Perceptual constraints on mobile displays are required to guarantee readability.

With these design properties, the problem can be formulated as finding a set of items and their optimal position, size and color to minimize the multi-objective cost function Z. $R = (((w_{i1}, w_{i2}), (h_{i1}, h_{i2})), O_i\left(x_i, y_i\right), \delta_i, \mathrm{a}_i), i \in [1, N]$ indicates I_i's size and position $((w_{i1}, w_{i2}), (h_{i1}, h_{i2}))$ within page P, gravity center coordinate $O_i\left(x_i, y_i\right)$, and the format indicator variable δ_i, where δ_i indicates I_i's aspect format, i.e., δ_i is 1 if I_i is landscape, 0 if I_i is portrait. Parameter a_i is the indicator variable, taking the value 1 if i_{th} item is selected and 0 otherwise. Then, we can formulate the problem as an energy function as follows:

$$\mathrm{Z} = E_{\inf}\left(R\right) + \omega_a E_{aes}\left(R\right) + \omega_v E_{vis}\left(R\right) \quad (2)$$

where $E_{\inf}$ is an information perception term. E_{aes} and E_{vis} are visual aesthetics term and visual perception term, respectively. ω_a and ω_v are weights for E_{aes} and E_{vis}.

3. ENERGY TERMS AND CONSTRAINTS OF MULTI-OBJECTIVE FUNCTION

According to the multi-objective cost function described in Sction 2, a visual summary can be created by minimizing three energy terms under a set of constraints. In the equation defined above, $E_{\inf}$ measures how informative the generated summary is; E_{aes} measures how well it follows the aesthetic rules; E_{vis} measures how much users are inclined to focus on the important items in attractive locations within P. Each of the energy term and constraints will be illustrated in detail in the following.

(a) (b)

Fig. 3: (a) Average scan path (left) and fixation map (right); (b) Hue wheel of text with local surrounding.

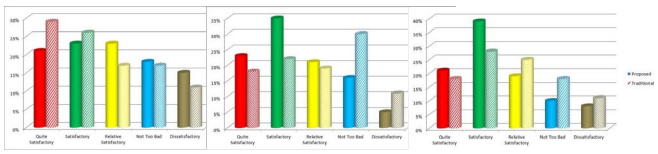

Fig. 4: Results of subjects evaluation corresponding traditional social media posts.(left) informative evaluation; (middle) aesthetic evaluation; (right) overall evaluation.

3.1. Informative and Popular Attributes

The ranking score for each item can be considered as representing its informative and popular attributes. That is, a higher score indicates a more important social media feed. Hence, the ranking score for each feed is used to define the information perception term E_{inf}:

$$E_{\text{inf}}(R) = -\sum_{i=1}^{N} a_i s_i \tag{3}$$

This energy term is a function of all indicator variables $a_1, \cdots a_N$. It will favor the most informative digest items.

3.2. Visual Perception

The average percentage of fixation and average scan path on the divided nine equal partitions for a single page layout are investigated in an eye tracking study [12], as illustrated in Fig.3. Intuitively, the highest ranked digest item should be placed at the location of first fixation. The next ranked items should be placed following the scan path of the decreasing fixation percentage. It is assumed that the amount of information perceived by mobile users for a certain social media feed should be proportional to its ranking score. Thus, it is natural to compute the overall information perception based on these two ranking lists as follows:

$$E_{vis}(R) = 1 - \frac{\sum_{i=1}^{N} a_i s_i \cdot d\left(per\left(x_i, y_i\right), Min_{per}\right)}{\sum_{i=1}^{N} a_i s_i} \tag{4}$$

where $per(x_i, y_i)$ is the visual perception value of position (x_i, y_i). $Min_{per} = \min_{(x,y)\in P} per(x, y)$. By minimizing E_{vis}, the visual perception of the selected items is maximized. The term E_{vis} is therefore capable of placing the selected digest items onto attractive regions by taking the ranking score s_i into account. $d(\cdot)$ is the normalized distance between $per(x_i, y_i)$ and Min_{per}.

3.3. Visual Aesthetics

It is very important to incorporate computational aesthetics into our system to enhance user experience and improve information reception efficiency. From the systematically investigated magazine page design, visual balance is a key aesthetic principle. In addition, to maintain readability, the text should be poped out from the local surrounding θ_i as shown in Fig. 3b. θ_i is the region that contains the text, which is defined as follow:

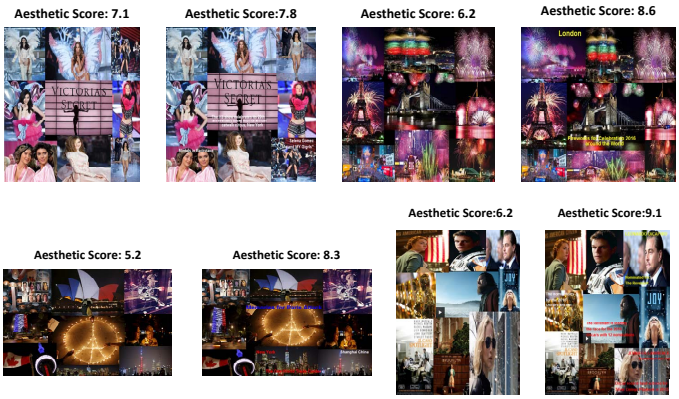

Fig. 5: Intermediate and final optimization results.

$$\theta_i = \left\{ \hat{I}_i\left(x, y\right) \mid |x - x_t| = \mu h_i,\ |y - y_t| = \nu w_i \right\} \tag{5}$$

where μ and ν are parameters that control the size of θ_i. $\hat{I}_i(x, y)$ is the synthesized image using I_t^i. h_i and w_i are height and width of I_t^i, respectively. Here, we adopt the *Complementary* color scheme for text color selection. Based on the overall consideration, the aesthetic term is defined as:

$$E_{aes}(R) = \omega_b \widetilde{e}_b(C, \widehat{C}) + \omega_c \widetilde{e}_c\left(\theta_i, T_{m,\alpha}\right) \tag{6}$$

Visual Balance

Based on [18], visual balance is achieved when the gravity center of all blocks coincides with the center of the whole page. The maximum visual balance can be achieved by minimizing the distance between the expected center, i.e. the center of the page $\widehat{C}\left(\frac{p_x}{2}, \frac{p_y}{2}\right)$ and the actual center C. This way, we define the visual balance term as:

$$\widetilde{e}_b(\cdot) = d\left(C, \widehat{C}\right) / \sqrt{{p_x}^2 + {p_y}^2} \tag{7}$$

where $d\left(C, \widehat{C}\right)$ is the distance between C and $\widehat{C}$. p_x and p_y represent page width and height, respectively. The actual center C can be obtained by

$$C = \frac{\sum_{i=1}^{N} a_i s_i O_i}{\sum_{i=1}^{N} a_i s_i} \tag{8}$$

We found that the term depends on position, size and ranking score of the selected items.

Color Contrast

Followed by [16], $T_{m,\alpha}$ represents a m-type template under a certain rotating angle α, where $m \in \{L, I, Y, X\}$, $\alpha \in [0, 2\pi)$. The color contrast term $\widetilde{e}_c\left(\theta_i, T_{m,\alpha}\right)$ is defined as follows:

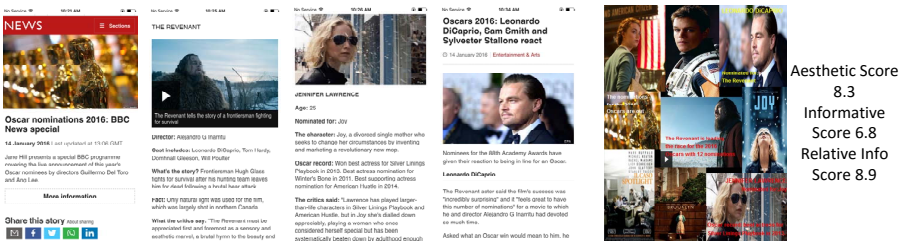

Fig. 6: Illustration of traditional social media posts(left four columns) and their corresponding visual summary.

$$\widetilde{e}_c(\cdot) = \frac{\left(\sum_t H^H_{\theta_i}(t) \cdot T_{m,\alpha}(t)\right)^2}{\max(H^H_{\theta_i}) \cdot \sum_t T_{m,\alpha}(t)},\ 0 \le t \le BIN \qquad (9)$$

where $H^H_{\theta_i}(t)$ is the histogram of the H channel of θ_i. BIN and $\max(H^H_{\theta_i})$ are the number of bins and maximum value of the histogram, respectively. $T_{m,\alpha}(t) = \mathbb{1}(t \in \mathbb{Q})$ is the t_{th} bin value of the harmonic scheme, $\mathbb{Q}$ is the gray sector.

3.4. Constraints

The set of constraints are illustrated in detail as follows:

• Seamlessly tiled display. To achieve the seamless synthesis, each block's four border edges should be adjacent to its neighbor block. In addition, all the blocks should cover the whole page P, i.e. $\sum_{i=1}^{9} area_i = P_x \times P_y$.

• Non-distortion. In order to prevent distortion during the optimization process, the updated images should keep the same aspect ratio, i.e. $\delta_i{}' = \delta_i$.

• Text readability. The font size of text has to be larger than a threshold th_{size}. Besides, considering the information perception on mobile displays, the maximum number of text characters should also satisfy a different threshold th_{num}.

• Non-intrusiveness. The text t_i should not occlude the salient objects or faces within its background image I^i_t.

• Non-overlapping. The texts inserted into its background image should not overlap with each other.

4. PERFORMANCE EVALUATION AND DISCUSSION

To validate the proposed system, we conducted subjective evaluations by comparing the magazine-page-like visual summaries created by the proposed scheme against the traditional social media posts. 20 subjects including 12 males and 8 females with age ranging from 20 to over 35 were invited to browse 30 randomly selected visual summaries as well as their source social media posts. The questionnaire for the subjective study are:

• How visually appealing is the visual summary as compared with the direct social posts?

• How informative is the visual summary as compared with the direct the social posts?

• What is your overall rating?

The subjects were requested to rate with a satisfaction score ranging from 1 to 10. To better evaluate the informative aspect, we conducted a relative informative evaluation. Specifically, subjects need to give rating within a same time duration for a certain visual summary and its source posts. Based on an analysis of SimilarWeb data [17] in 2015, time duration for mobile browsing ranging from 3s to 7s for different areas, where 2.55s for news and media. Thus, we limit the time in 5s for subjects rating. An example is shown in Fig. 6, in which original source social posts and the generated visual summary are compared. We can see that, the relative info score is 8.9 which is much better than the absolute score 6.8. This is because one needs to browse many more direct social posts in order to perceive the same amount of information delivered by the new visual summary. Therefore, within same time period, browsing the proposed summary is much more effective than browsing the original social posts.

From Fig.4, we can see that, comparing with the traditional social posts, there are nearly 80% subjects who consider the proposed visual summary is more visually appealing and satisfied with its overall performance. Fig.5 shows sample results of the generated visual summaries. Each set has two images, one represents intermediate result of optimization while the other represents final result. Based on the proposed multi-objective function, we can see the dynamics of the optimization process in which the position and size of the nine blocks continuously change to minimize the overall energy while still satisfying the constraints. The aesthetic scores for each of them have been increased to achieve optimization. This proves that the proposed optimization scheme is efficient and practical in social media visual summary design.

5. CONCLUSION

In this paper, we propose an effective visual summary creation system to automatically convert a set of personalized social media digests into a magazine-like page for mobile browsing. In particular, aesthetic rules and perception principles are employed to optimize the initially generated layout under several constraints. As a result, the visual summary is not only capable of transferring maximal amount of information within one page, but also aesthetically pleasing to the mobile users for efficient mobile browsing under the display size constraints. Extensive experimental results have shown that the proposed visual summary is effective in representing social media feeds and enhancing user experience on mobile browsing.

References

[1] J. Xu and T. Lu, "Seeing the Big Picture from Microblogs: Harnessing Social Signals for Visual Event Summarization," in Proceedings of ACM IUI, 2015.

[2] Z. Wang, P. Cui, L. Xie, H. Chen, W. Zhu, and S. Yang, "Analyzing social media via event facets," in Proceedings of ACM Multimedia, 2012.

[3] Flipboard. http://www.flipboard.com.

[4] W. Yin, T. Mei, C. W. Chen, "Automatic Generation of Social Media Snippets for Mobile Browsing,"in Proceedings of ACM Multimedia, 2013.

[5] W. Yin, T. Mei, C. W. Chen, and S. Li, "Socialized mobile photography: learning to photograph with social context via mobile devices," IEEE Transactions on Multimedia, vol.16, no.1, pp.184-200, 2014.

[6] W. Yin, T. Mei, C. W. Chen, "Crowdsourced learning to photograph via mobile devices," in Proceedings of IEEE International Conference on Multimedia & Expo, pp.812-817, 2012.

[7] S. Ma, Y. Fan, C. W. Chen, "Finding your spot: a photography suggestion system for placing human in the scene," in Proceedings of IEEE International Conference on Image Processing, 2015.

[8] S. Ma, Y. Fan, C. W. Chen, "Pose Maker: A Pose Recommendation System for Person in the Landscape Photographing," in Proceedings of ACM Multimedia, 2015.

[9] 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 WSDM, 2011.

[10] W. Yin, T. Mei, C. W. Chen, "Assessing photo quality with geo-context and crowdsourced photos," in Proceedings of IEEE International Conference on Visual Communications and Image Processing, 2012.

[11] C. lai, P. Chen, S. Shih, Y. Liu and J. 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, 2010.

[12] 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.

[13] C. W. Chen, "Browsing Information Overloading Unstructured Multimedia Social Network Contents on Mobile Devices," in Proceedings of Conf. on Imaging and Multimedia Analytics in a Web and Mobile World, 2016.

[14] C. Yang, L. Zhang, H. Lu, X. Ruan, and M. H. Yang, "Saliency detection via graph-based manifold ranking," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2013.

[15] L. Zhang, M. Song, Q. Zhao, X. Liu and J. Bu, "Probabilistic graphlet transfer for photo cropping," IEEE Transactions on Image Processing, 2013.

[16] D. Cohen-Or, O. Sorkine, R. Gal, T. Leyvand and Y. Q. Xu, "Color harmonization," ACM Trans. Graph, 2006.

[17] https://www.clickz.com/clickz/column/2388915/why-mobile-web-still-matters-in-2015

[18] 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, 2006.