DOI: 10.2352/ISSN.2470-1173.2016.11.IMAWM-458

Browsing Information Overloading Unstructured Multimedia Social Network Contents on Mobile Devices

Chang Wen Chen; State University of New York at Buffalo; Buffalo, New York, 14260-2500, USA

Abstract

This paper addresses important technical issues in browsing heterogeneous unstructured multimedia feeds on consumer mobile devices derived from social network contents. We will first present several pressing technical challenges associated with creating a browsing system that can summarize information overloading unstructured social media feeds and produce a novel GIST, namely, Graphical Intelligent Semantic Transform, for effective and visually pleasing browsing on a mobile device by the social media users. We will then illustrate innovative solutions to solving a suite of interdisciplinary problems associated with developing such a system. Preliminary results will be shown to demonstrate the feasibility of creating such a GIST for browsing information overloading social media feeds on consumer mobile devices.

1. Introduction

The recent proliferation of social media networks, including Facebook, Twitter, Flickr, YouTube, and Google+, has revolutionized everyone's life with advanced features in richer user experience, more convenient user participation, continuously updating of media content, increased scalability and enhanced openness. For example, Facebook reached the milestone of 1.5 billion monthly active users in 2015. Worldwide on Facebook, over 900 million people log onto Facebook daily (DAU) for June 2015. Every 60 seconds on Facebook, over 500 comments are posted, 293,000 statuses are updated, and 136,000 photos are uploaded [1]. At such overwhelming scales, social media is certainly having a pervasive influence on the society. Social media help people and businesses stay better connected, better informed [2][3], and also can serve as distributed social sensors and source of wisdom of crowds [4][5][6].

However, such overwhelming scales of social media triggers an unintended consequence – information overload. We are facing continuous 24/7 information streams of social media feeds, including status updates, news and information sharing, and comments and opinions. Lost and overwhelmed in the vast sea of social media is each individual user. Despite of tremendous advances in mobile devices, they are fundamentally constrained by limited browsing and access capabilities. Both ordinary users and business enterprises have been struggling to keep up with the overwhelming volume of unorganized and unfiltered information. This is particularly true when the primary mode of access to social media is now through mobile devices as we carry with us while on the move constantly. We are suffering from the inability to make decisions because our overloaded brains freeze with too much information presented on small window of mobile displays [7][8]. It is amazing to see that two contemporary applications, social media and mobile devices, are becoming a double-edged sword that affects people's wellbeing in both positive and negative ways.

It is clear that generating social media feeds for effectively browsing on mobile device with pleasant user experience is facing significant technical challenges. First, it is a real challenging task to convert the information overloading unstructured mixed media social feeds into some compact and personalized summaries and digests. The dynamic nature of these feeds needs also be carefully considered to reflect the evolution of the social media feeds over time. Second, it is also extremely challenging to render such mixed media digest and summary for display on the capacity limited mobile devices with a pleasant user experiences. Both information maximizing and aesthetically pleasing principles are required for the design of a visual presentation for effective mobile display and seamless user interaction.

2. Related Research

The research community has made several initial attempts at developing summarization and browsing systems in recent years. In this section, we present brief reviews about the related research in summarizing unstructured text and unconstrained images and videos, as well as some related research in designing aesthetically pleasing display of social media digest for mobile display.

2.1. Unstructured Text, Image, and Video Summarization

Near real-time systems have recently been developed to process short unstructured text such as tweets on Twitter. TwitterStand [3] is a system that builds a distributed news wire service from tweets. The main challenge is that the identities of the contributors or 'reporters' are not known in advance and there may be many of them. Furthermore, tweets occur as news is happening, and tend to be noisy while usually arriving at a high throughput rate. This system addressed some of the issues including removing the noise, determining tweet clusters of interest using an online method, and determining the relevant locations associated with the tweets. A system called Eddi [2] was developed to organize a user's Twitter feeds into coherently clustered trending topics for more directed exploration. Active Twitter users evaluated Eddi and found it to be a more efficient way to browse an overwhelming status update feed than the standard chronological interface. Another system deals with text-to-image summary [7], which automatically illustrate complex sentences as multimodal summaries that combine pictures, structure and simplified compressed text. Such a multimodal summaries provide additional clues of what happened, who did it, to whom and how, to people who may have difficulty reading or who are looking to skim quickly.

Video summary or in a broader sense video abstraction refers to generating a short summary of a video, which can either be a sequence of stationary images (key frames) or moving images (video skims). Various ideas and techniques have been proposed towards the effective abstraction of video contents and the early works are reviewed in a survey [10]. Note that most of the existing systems dealt with structured videos such as news, broadcasting sports, and education videos and thus take advantages of available metadata, caption, and audio information [10]. Unconstrained

DOI: 10.2352/ISSN.2470-1173.2016.11.IMAWM-458

personal or online videos pose additional challenges that achieved modest successes only in recent years [11][12].

For image summary from a collection of related images, it usually involves clustering and selecting representative images from the clusters. In [13], images related to a geographic location (e.g., The Golden Gate Bridge) are first clustered using k-means on concatenated color and texture feature vectors. SIFT features [14] are then incorporated for ranking clusters and selecting their representative images.

2.2. Web Search and Social Media Summary

Various representations of web search results have been studied comprehensively. Initially, most web search engines generate simple text excerpt, or text snippet, from selected keywords that match the query terms for users to assess their relevance. Recently, it has been shown that selected pictures in the web pages, namely image excerpt [15], can be added to text snippet, to enrich web search experiences resulting quicker relevance judgment. More advanced visual summary have been developed [16][17]. These representative text and visual summary techniques are all developed for representing individual search results and each will be presented as an ordered one-dimensional list. Recently, presentation of search results in a two-dimensional fashion has been successfully demonstrated [18]. In particular, based on the well-known Golden Triangle, or F-shaped pattern, placement of search results in a two-dimensional grid can be obtained by means of Markov chain modeling. One common feature for these schemes is that they focused on presenting collection of individual items from the search return to allow the users to access the details via web links from each item.

Several efforts have also been attempted by web application developers and multimedia researchers to summarize social media feeds. One such example is called "summify" developed by the Context Media Technologies Inc. Summify access the information and only show the user a summary of what is most relevant based on how much the specific information and stories are shared by the user's peers. In multimedia community, several pioneering studies have been carried out to represent social media blogs and stories by rich media layout such as photo books [19][20]. The generation of these rich media layout is based on interdisciplinary scientific principles. For example, the Expectation-Maximization algorithm has been designed to retrieve photos from same event to compose a coherent story [20]. Aesthetic principles, such as spatial layout rules, symmetry, and color harmony have also been employed to automatically generate the photo book layout [21]. However, these works have not considered the perception of such summarized representation and the scenarios of small display of mobile device.

2.3. Computational Aesthetics in Photos and Videos

The popularity of photo-capturing devices, including digital camera and smartphones, has resulted in an explosive growth of digital images and videos. Many such images and videos are shared via social networking sites in which people from all walks upload their photos and share their personal stories with friends through these photos. How to evaluate the visual quality of a photographic image via computational means is a challenging task. The early computational approaches to study image aesthetics [21] involve extracting visual features to classify aesthetically pleasing and displeasing images. Another attempt was also based on feature extraction in which high level semantic features, instead of bag of low-level features, are used to measure the perceptual differences between high quality professional photos and low quality snapshots [23]. These early studies use features extracted from the entire photo and therefore may not reflect the true image quality surrounding the subject regions. A more accurate characterization of image quality with semantic features has been developed focusing on the subject and its surrounding regions within a photographic image [24]. High level describable attributes have been applied to predict the aesthetics and interestingness and to select high aesthetic quality images from large collections [25].

These computational aesthetics approaches [21] have successfully been applied to individual photos and videos for a variety of tasks. Only until recently, the computational aesthetics strategy has been considered for scenarios beyond assessing individual images. In the composition of photo book layout, aesthetic principles have been employed to create more pleasant photo books using proper texts to augment several selected photos for a mixed media layout design [21]. In an attempt to enhance browsing experience of web users, a novel computational aesthetic approach has been developed to evaluate the visual quality of web pages [26]. This new approach takes into account multiple visual features, including layout, text, color, texture, and complexity features, to construct a vector for visual quality classification and quality score assignment. It is shown that this holistic approach is consistent with human evaluation of web pages.

3. Need for Systematic Creation of GIST

Existing efforts in the research related to text, image, and video summarization, as well as in the presentation of summary from Web search and social media feeds, sets several pertinent foundations for developing an ultimate GIST (Graphical Intelligent Semantic Transform) for presenting the digest from unstructured multimodal social media feeds on the mobile devices. However, these existing efforts lacks of systematic design methodology to maximize both information perception and composition aesthetics of an ultimate GIST for mobile users. It is expected that an ultimate GIST shall consist of two main components: (1) a social media feeds analysis and summarization component, and (2) a graphical rendering component capable of intelligently capturing the semantics of the digest with visually pleasing representations. A suite of new algorithms needs to be developed to achieve this grand research goal. For summarization, a personalized social media digest should be richer and more informative than plain sequential textual information so that it captures the "gist" or essence of the original unstructured social media feeds. For mobile rendering of the GIST, a magazine page like GIST can be created which is not only capable of transferring maximal amount of information within one page, but also aesthetically pleasing to the mobile users. The desired dual goals in GIST creation can be fulfilled by proper partition and optimal placement of ranked items from social media summarization as well as the balanced selections of color, size, and location for the placement of pictures, texts, and other media items optimally within the magazine page like GIST for rendering on mobile devices. The principles employed in these designs are deeply rooted in theories developed in visual information perception, print advertisement design, and computational aesthetics. These multi-disciplinary principles will be creatively adapted to the design of GIST for access by mobile devices whose display and network link characteristics will also be fully incorporated.

It is therefore obvious that the conventional approaches to summarizing and ranking single media such as text, image, or video cannot be directly adopted for such a task. A systematic approach is needed to examine the users associated with the

DOI: 10.2352/ISSN.2470-1173.2016.11.IMAWM-458

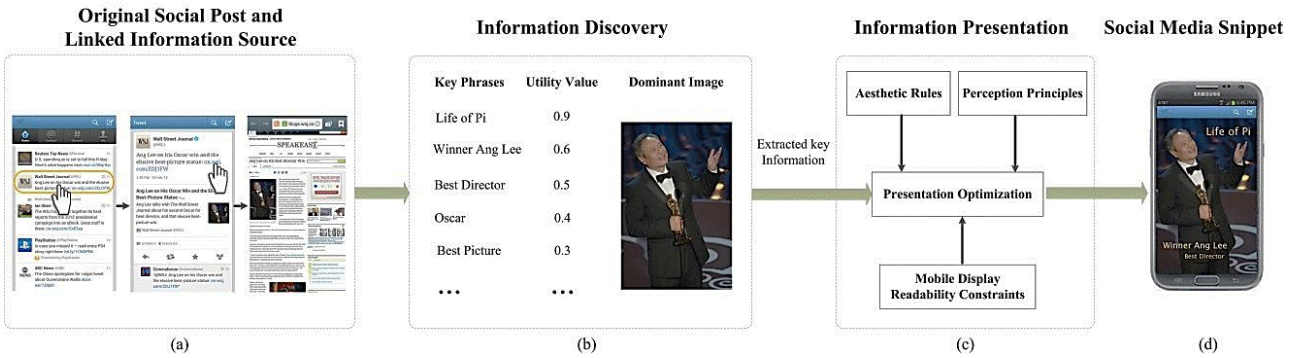

Figure 1: Overview of social media snippet. (a) Original social post and linked information source, (b) information discovery module, (c) information presentation module, (d) generated social media snippet

apparent form of social media feeds in text, image, and video in order to extract underlining relations behind the social media feeds. It is necessary to take into consideration the social relations among the users and between the feeds and the users in order to produce more personalized digest. It is also necessary to incorporate the temporal evolution of social media feeds into the summarization to capture the dynamics of the social media feeds.

4. Social Media Snippets: Initial Attempt

An initial research towards the overall objective of creating a full functional GIST has been carried out recently. Specifically, we developed a framework for automatic generation of social media snippets for efficient browsing on mobile devices [38]. In this research, the social media snippet system we have developed consists of two modules: Information discovery and information presentation as shown in Figure 3.

In this preliminary study of GIST system design [38], we developed a compact system to automatically build attractive, appealing and informative snippets in a simple text-overlaid image style for effective browsing on a mobile device. The first step towards generating social media snippet is to discover key information in the posts and links in the form of one dominant image plus several key phrases for the snippet generation in the second step. The key information discovery step has been implemented via a modified TextRank algorithm [39] which extract relevant elements from original social posts and linked information sources. Then, the second step in generating social media snippet is to design information presentation module by systematically selecting and composing the extracted elements to generate snippets based on aesthetic rules and visual perception principles considering mobile display constraints. The social media element selection and composition in this step is achieved by an optimization based on the aesthetic rules and perception principles having been proved in the literature having strong correlation with aesthetics and information perceptions in text-overlaid images.

4.1. Compact Information Discovery Module

To enable mobile users getting most important information at a quick glance, key phrases and one dominant image are extracted from social posts and linked articles during information discovery.

4.1.1. Dominant Image Extraction

Compared with pure text summaries, images are more attractive, appealing and informative from psychology study. Based on the study in search results presentation [40][41], summaries with images can help users understand content quickly for faster relevance judgment. In our system, we made experiments on social posts in news categories with images. From our observations, the first image in the linked articles is usually dominant and utilized in the thumbnail as the instance shown in Figure 5-1(b), so we extract the first image of the linked article as the dominant image for social media snippet generation.

4.1.2. Key Phrase Extraction

Key phrases can express the information concisely and accurately. Wordy sentences are difficult to be viewed 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. We develop our key phrase extraction based on the graph-based ranking algorithm TextRank [39] and contains the following stages: word extraction, word ranking and key phrase reconstruction.

Word extraction and word ranking: First, the social posts and linked articles are processed for key word extraction. As suggested in [42] and used in Visual snippet [41], when the social posts are longer than 19 characters, we will extract key phrases. We make tokenization, PoS tagged and selected nouns and adjectives as the vertices in the graph as in [39] and utilize the co-occurrence relation controlled by the distance between words to generate the edges between vertices. Two vertices are connected if they co-occur within a window of two words.

The extracted text units and their relationships are utilized to construct an undirected graph. Then, the utility value of a vertex can be defined and minimized to obtain proper word ranking. The process of word ranking follows a similar scheme as in [39] but take the importance of the title of the linked article into consideration. Instead of using unweighted graphs as in [39], we consider the superiority of the phrases in social post and title in terms of their confidence by weighting these vertices higher. The utility values of the vertices are iteratively updated until convergence. The top vertices are retained for post-processing.

Key phrase reconstruction: In this stage, multi-word key phrases are reconstructed from the potential T key words. All the potential lexical units selected are marked in the text and the key words adjacent or connected with a preposition and a conjunction are concatenated together. For example, the selected key words ‘life’ and ‘pi’ will be concatenated into a key phrase ‘life of pi’, since they are connected with a preposition. Once the key phrases are reconstructed, the utility values of key phrases are calculated by taking the sum of utility values of the key words within them. Finally, utility values of the extracted key phrases are normalized.

4.2. Generating Social Media Snippet

The multimodality of social media with prevalent images and video clips creates a good opportunity for presenting the summary

DOI: 10.2352/ISSN.2470-1173.2016.11.IMAWM-458

of social media content by a dominant image plus key text words. It has been shown that images are superior at capturing attention [43]. The success in adopting images for presenting search results provides great evidence that images can indeed help quicken understanding of webpage content and relevant judgment [16]. Based on these considerations, we developed an initial prototype of social media snippet in text-overlaid image style for mobile browsing [38]. In this preliminary research, a scheme in generating social media snippets for mobile users browsing is developed by utilizing proper images in the social posts to help the users understand the content quickly. To make sure that such snippets are consistent with the usability and aesthetics principles, the text-overlaid images are designed based on theoretical foundations on how people browse and perceive derived from research in eye tracking [44] and computational aesthetics [45].

This initial attempt aims at developing a scheme to represent social media summary with an attractive, appealing, and informative snippet in text-overlaid image style for mobile browsing. In the first step, key information discovery is carried out from the posts and links to generate snippet elements in the form of one dominant image and several key phrases. We modified the TextRank algorithm [46] to accomplish the information discovery step. We assume that we can find one dominant image whose content is indeed relevant to the social post. The properly extracted key phrases with properly designed colors and sizes are placed on top of the dominant image in such a way that the final snippet is suitable for display on mobile device screen subject to aesthetic rules and visual perception principles. This new social media snippet compares favorably over existing systems such as Flipboard [47] which use fixed layout patterns in terms of attractiveness score, appeal score, and informative score. Such a new form of social media summary is easier to show and view on various mobile devices and preferred by mobile users.

5. Towards Creating the Ultimate GIST

Although the preliminary research presented in previous section sets several pertinent foundations for developing towards the ultimate GIST, the current approach in text-overlaid *social media snippet* generation lacks of systematic design methodology to maximize both information perception and composition aesthetics. In particular, the limited layout design space is severely constrained by simplified composition with text-overlaid image to represent the social media summary. Based on the lessons learned from generating social media snippets as well as the pertinent foundations we have established, we intend to develop a holistic approach with much expanded design space to optimally represent comprehensive mixed image/video and text results obtained from multimodal summarization of unstructured social media feeds.

To that end, we plan to design a framework for systematically creating Graphical Intelligent Semantic Transform (GIST) to capture and present the essence of various multimedia contents from the web and social media. A detailed illustration of the proposed system is shown in Figure 4. In the left, the system takes unstructured multimodal social media feeds, including posts, texts, images and videos, as the input. An evolutionary heterogeneous graph will be constructed that not only maps all unstructured media feeds and user information into heterogeneous graph, but also embeds temporal updates of these feeds into an evolutionary graph to capture the dynamics of the social media feeds and posts. In the middle, Figure 4 shows an example of personalized social media summary and ranking ready for generating a GIST for rendering on mobile devices. In the right, Figure 4 shows an example of final GIST that is not only capable of transferring maximal amount of information within one page, but also aesthetically pleasing to the mobile users.

5.1. Personalized Social Media Digest via Evolutionary Heterogeneous Graph

To generate a personalized social media digest, we propose to generate a summarization framework that, given a set of microblog posts and social media feeds on a certain key topic, can simultaneously select a subset of texts, images and videos to produce most informative summarization tailored to the given user from unstructured social media feeds.

The proposed research in generating the desired social media digest is fundamentally different from existing summarization approaches in two novel aspects: *personalization* and *evolution*. The personalization attribute is achieved by embedding the users and their social relations into a heterogeneous graph we propose to construct that consists of not only the posts and feeds of social media, but also the users who produce and consume these items. The evolution attribute is achieved by creating an evolutional map based on the temporal dynamics of texts, images, videos, and users, as they are continuously updated throughout the entire process.

5.1.1. Evolutionary Heterogeneous Graph Construction

In order to produce the most informative and personalized summary from unstructured social media feeds, we propose to build an *evolutionary heterogeneous graph* with multimodal nodes and edges. Such graph is able to characterize the interrelationship among various entities (i.e. users, texts, images, videos, comments and sentiments) in online social media and therefore achieve simultaneous personalization and summarization. The intuition behind the proposed research is that the social media users not only can be crowdsourced as important *sensors* to capture and share data, but also can be utilized as proper *filters* to discover the most useful information through analyzing their interaction in the social media networks. With this novel graph construction, we can formulate the summarization problem as graph-based ranking by identifying the most representative media entities with respect to a given topic query.

To initialize the construction process, we first assemble various different social media entities, including the feeds in the form of posts, texts, images, and videos, as well as both contributing and consuming users to form a collection of heterogeneous nodes. These nodes are linked by two types of connection: *interaction edges* and *time-flow edges*. In particular, the time-flow edges are introduced to characterize the evolution of the social media feeds and interactions over time.

Different from existing scheme, two additional factors will be considered when generating personalized social media digest for a given user: user's past behavior and latest current events. It is this embedding of user behavior into overall digest objective that enables the generation of more personalized digest. The inclusion of current events enables the users to be informed of possible impact of non-personal events.

Let $G = (V, E)$ be a directed graph with a collection of nodes V and edges E. These nodes correspond to multimodal social media entities, namely users, images and videos, and texts. In this research, texts shall include not only the original posts, but also user comments and sentiments. For edges, the interaction edges represent the interrelationship among various multimodal entities, while the time-flow edges reflect the evolution over time between

DOI: 10.2352/ISSN.2470-1173.2016.11.IMAWM-458

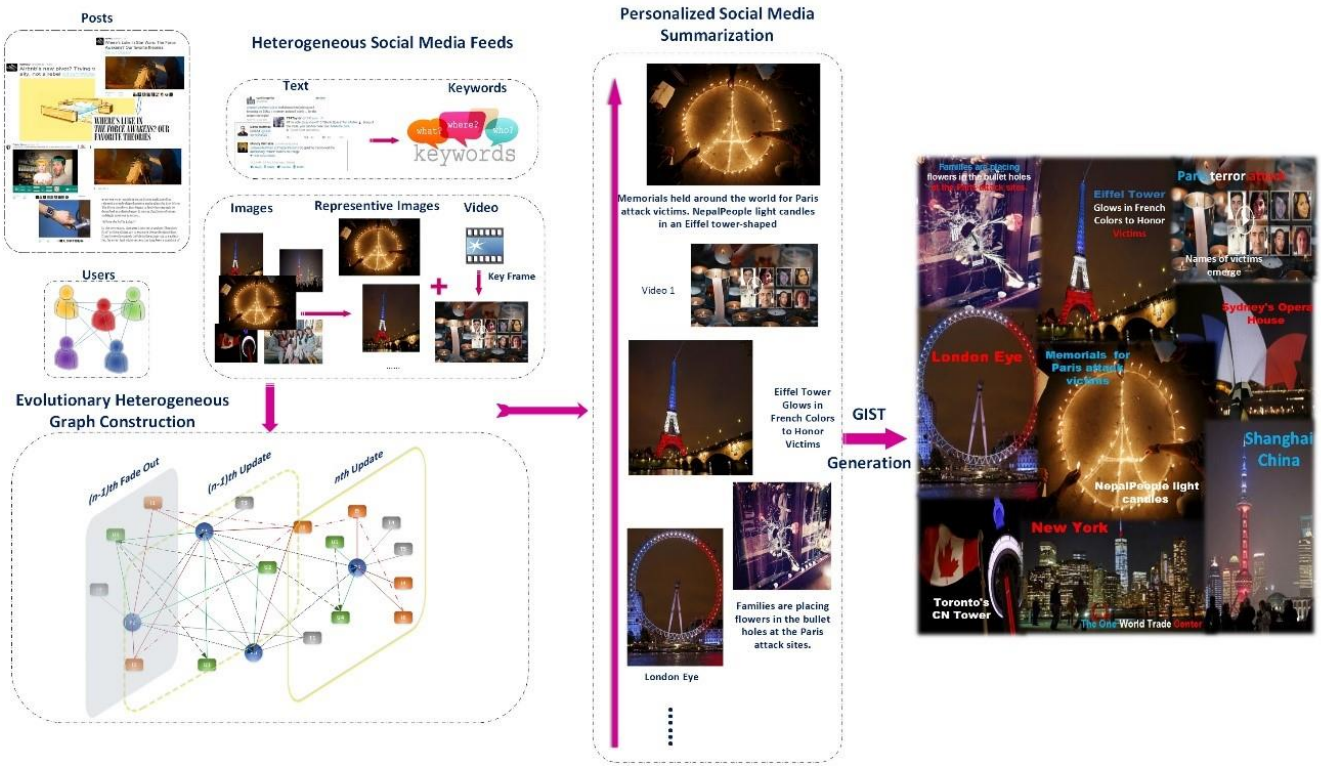

Figure 2: Overall GIST framework: Left: Unstructured social media feeds and graph construction; Middle: Ranked multimodal personalized digest; Right: Final generated GIST page for mobile device display.

a pair of posts. Figure 5 illustrates multimodal unstructured social media and corresponding evolutionary heterogeneous graph to represent the intrinsic relations among all nodes.

The following social media entities are assigned as notes in the proposed graph:

Post Nodes: For a given user, there are n posts that can be retrieved from the social media feeds. We can represent this set of posts by $P = (p_1, p_2, \dots, p_n)$. The set P consists of the original post only. The reposts will be exclude from the post nodes set in this research.

User Nodes: The type of user node is introduced for each of the n_u users if they have interacted with any of the microblog posts in the feeds of a given user. We consider several common user behavior on the social network, including uploads, repost and likes, all of which can be used to predict user's interests and popularity of a post. The set of user vertices can be represented by $= (u_1, u_2, \dots, u_{n_u})$.

Text Nodes: Microblogs and social media feeds may be grammatically incorrect and may contain spelling errors. It is therefore a significant challenge to summarize short, unstructured, and informal text of short messages. In particular, social media post may be restricted to only 140 characters and most users use abbreviations to express their sentiments towards a given social media entity. In contrast, news and blogs are usually written by professional reporters or trained writers who usually express their topics clearly. To accommodate these irregularity of text posts, we propose to extract two types of feature vectors: (1) summarizing the topics from headlines and (2) summarizing the sentiment.

Summarizing the topics from the headlines: The headlines are structured text of the original news or events, so we can use the techniques we developed for creating simplified sentences from complex sentences [37]. For topic summarization, we believe social media summarization should include the essential elements for event description in terms of subjects or objects (who), time (when), location (where), and event (what). We propose to use named identity extraction techniques (e.g., Stanford Dependency Parser) to obtain such information for an event or news item [38].

Summarizing the sentiments: The sentiments include the comments of the public, friends, or local social network from the microblog posts. For sentiment summarization, we propose to exploit 'like', 'dislike', and all the abbreviations that can convey emotions, e.g., LOL/OMG/U2/Gr8. Such simplified sentences can be displayed as pseudo-headlines in the social media digest. In addition, topics will be categorized using ONP ontology [8] (technology science, sports, health, pop culture, etc.). After we extract the feature vector for each text feeds, a text vertex is added for each of n_t texts used within the social media feeds, $T = (t_1, t_2, \dots, t_{n_t})$.

Image & Video Nodes: Given a collection of related images: $I = (i_1, i_2, \dots, i_{n_i})$, an image node is added to the graph for each image that is attached to any of the posts. We define two types of image links to capture different relationship as: image-to-post link and image-to-image link. Specifically, an image node will link to its parent post via image-to-post link and to another image via image-to-image link. The image-to-image link are measured by the visual similarities between images. We propose to employ the homogeneous and heterogeneous message propagation (H2MP) [39] to summarize image and video nodes. This scheme extends affinity propagation that works over both homogeneous and heterogeneous relations. With this approach, we can find a few

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.11.IMAWM-458

image exemplars to represent the image set semantically as well as visually by exploiting both visual and textual information associated with images.

For video feature extraction, we propose to adopt the scheme of heterogeneity image patch (HIP) index [40]. This scheme provides a new entropy-based measure of the heterogeneity of patches within any picture. Finally, we can extract the key frame from the collected video stream.

The following two types of edges will be assigned to the proposed graph:

Interaction Edges: Multiple social interactions and activities are captured through edges added between post nodes and their surrounding attributes (nodes) (i.e. users, texts, and images & videos). The direction shall point from the parent nodes to its descendent. For instance, a tweet is posted by a user u_i and reposted by another user u_j, then the edge direction will be defined as $u_i \rightarrow u_j$. Note that, a user can reposts from others, a text can be associated to multiple posts, and a post can be liked by many users. We propose to represent the interaction relationship between multimodal entities by adjacency matrices.

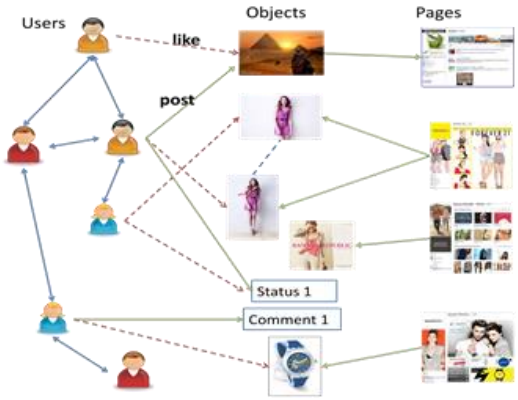

Figure 3: Multimodal unstructured social media and their links

Time-flow Edges: Edges will be introduced among post nodes when they are published within a close time interval. The adjacency matrix for posts is defined as:

$$M_{tf} = \begin{cases} 1 & \text{if } \left|T(p_i) - T(p_j)\right| \ll T_h \\ 0 & \text{otherwise} \end{cases}$$

where T_h is a threshold controlling the time interval proximity of two posts. The direction of the time-flow edges is consistent with the time order for each post nodes. We focus on a portion of the time-evolving graph with a fixed time window centered at a target event, which can simplify the graph to enhance the efficiency of the algorithm.

5.1.2. Generating Multimodal Social Media Digest from Evolutionary Heterogeneous Graph

Given the evolutionary heterogeneous graph constructed according to the guidelines described in the previous Section, an graph-based ranking algorithm can be applied to obtain a collection of representative multimodal entities derived from the social media feeds, posts, texts, images and videos, for a given mobile device user. In this research, we propose to adopt the manifold ranking [41] to accomplish this task. This algorithm computes a relevance score for each nodes with respect to a query based on the intrinsic graph structure. Consequently, we can obtain a ranked list of candidate texts, images, and images associated with videos. These highly ranked entities are selected as representative entities to produce a desirable multimodal summarization. The heterogeneous graph shall be updated continuously as shown in Figure 4.

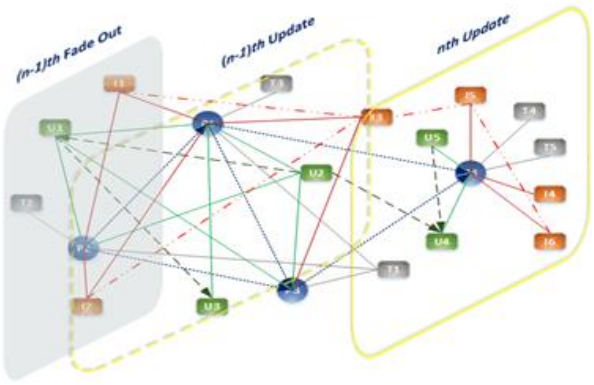

Figure 4: Evolutionary updates in heterogeneous graph

Notice that the proposed evolutionary heterogeneous graph by including users and related info as well as temporal info enjoys two unique advantages over existing approaches in ranking and summarization:

1. Empowering the graph-based ranking and summarization to obtain more personalized results through creatively designed embedding of user interaction and user past behavior into the graph.
2. Enabling the evolutionary update of timely results from the heterogeneous graph that truly reflect the inherent dynamics of social media feeds that evolve continuously over time.

Graph Updating: The updating process shall be activated in either time-orientated or interests oriented fashion. At a given time, the system generating the digest shall focus on a window of the time-evolving graph centered at this time for a given mobile user. As the time progresses, new nodes and edges should be appended to the graph while the old ones phasing out of the current time window should be removed. In addition, when the users need to know more detail information for a certain post or event, or their interests change over time, the system shall be able to move the window over the evolutionary graph to a new center for an updated graph and the nodes and edges within this new graph will be updated subsequently. We propose to adopt node and edge aggregation approach, which define a "Factor-Item bipartite graphs" to fulfil the node and edge aggregation, with an improvement for this algorithm to become more compact and effective [43].

Graph Diversification: The manifold ranking we adopt in this research takes into account relevance and importance of the data, but without considering the diversity. However, as we aim at generating the most informative digest, diversity is very important in order to cover different aspects among heterogeneous social media entities. To address this problem, we propose to use a greedy algorithm which explicitly penalizes redundancies and promotes diversity [44]. The central idea of the diversification algorithm is to decrease the overall ranking scores of those nodes which have their information already conveyed by the more informative ones in temporal proximity. After this step of diversification, the nodes with the highest overall scores will be selected for final summarization of the social media feeds.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.11.IMAWM-458

5.2. GIST Design Principles

Given its focus on news feeds, TwitterStand provides a simple interface with the left pane showing the result of online clustering on tweets with their corresponding geographic focus in the right pane [3]. Eddi, on the other hand, presents a richer topic-oriented browsing interface: clockwise from upper right is the tag cloud and timeline tabs, the topic dashboard, and the navigation list (with a scroll bar) on the left hand side. Neither of them displays any images or videos on the primary interface. In contrast, Google News incorporates thumbnails of images and videos along with the top story.

We propose GIST as a novel representation for the personalized digest generated as outlined in Section 4.1. It is capable of (1) maximizing information perception by proper placement of digest items; (2) composing the ranked digests into highly aesthetic display; and (3) incorporating user, network link, and terminal contexts for robust and adaptive delivery.

5.2.1. Maximizing Information Perception

An important objective of GIST design is to maximize information perception of the personalized digest by the users. This will be particularly true when the personalized digest is delivered to mobile terminals in which the display size is relatively small. Several existing approaches have only implicitly considered such objective. In this research, we explicitly establish this as the most desired capability when designing GIST.

For social media feeds, information can come in an unstructured fashion and in a variety of form in text, image, video and implicit context. The personalized digest we generate shall provide the results of simultaneous clustering and ranking. Unlike the web search results, we believe it is more appropriate to represent the digest from social media feeds with structured layout for easy understanding. This is also consist with the principle of information maximization and allows for more rigorous evaluation of GIST design. Given a single page layout, we shall design a magazine like page and shall initially divide the page into nine blocks based on the "Rule of Thirds" guideline in the visual art composition [45]. We shall demonstrate in the next that information maximization can be accomplished by designing an algorithm to place a given item at certain location of the display and another scheme to change the size of each block in accordance with the relative information entropy of the digest items.

Information Maximization (Objective Criterion): It has been recognized that digests from social media feeds need to follow the 4W's rule (Who, What, When, and Where). This is largely true when we design simultaneous clustering and ranking to generate the digest. These 4W's also constitute the main building blocks for designing the information maximizing composition of GIST. We need to figure out which block should these ordered digest items be placed based on human visual perception principles. The score we obtained for each media item can be considered as relative information entropy because a higher score indicates a more important digest item. To maximize the information perception, intuitively the highest ranked digest item (highest score) should be placed at the location of first fixation. The next ranked items should be placed in such an order that they follow the scan path of the decreasing fixation percentage.

Based on an eye tracking study [33], for a single page layout with "Rule of Thirds" partition, the first fixation is at the center of layout while the subsequent fixations follow an average scan path as shown in Figure 5 (left and center). With such a scan path based on the order of fixation time, the amount of information perceived by the user for a given digest item is proportional to the product of the score of this item and the fixation time expended on this item when the layout is viewed. The overall information perception is thus proportional to the sum of products of score and fixation time for all 9 blocks of display.

The computation of overall information perception shall be based on these two rank ordered lists. The first one is the rank order of digests from the simultaneous clustering and ranking algorithm presented in Section 4.1 while the second one is from the rank order of fixation time of each block that follow a scan path as shown above. The intuitive strategy of digest item placement that matches the order of two lists can be shown to also maximize the overall information perception. To prove this rigorously, let $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)$ be a sequence of non-negative numbers with $\sum_{i=1}^{N} \alpha_i = 1$. The Ordered Weighted Averaging (OWA) operator OWA^{α} associated with α can be defined as [46][47]:

$$\text{For any } t = (t_1, t_2, \dots, t_N),\ OWA^{\alpha}(t) = \sum_{j=1}^{N} \alpha_j t_{\sigma(j)},$$

where for any $j, t_{\sigma(j)} \geq t_{\sigma(j+1)}$. Furthermore, if α is an ordered sequence such that $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_N \geq 0$, then, we will have:

$$OWA^{\alpha}(t) = \max \sum_{j=1}^{N} \alpha_j t_{\sigma(j)}$$

This indeed proves that OWA achieves its maximum when both α and t are rank ordered. In the proposed placement algorithm for information maximization, we will normalize the rank scores of digest items as well as the fixation percentage of the display blocks. However, this normalization shall not change the order for both lists under consideration. When the number of digest items is greater than nine, it may need to be truncated because of there are only nine blocks for placement.

Perception Maximization (Subjective Criterion): The information maximization algorithm we proposed above can be considered as an objective criterion for placement design. We plan to exploit a subjective criterion to complement the objective criterion-based design. This is a perceptually driven approach trying to design a systematic scheme to change the relative size of each block for possibly improved perception. In particular, we conjecture that the information perception can be enhanced when more important digest items are displayed with larger display size. For a fixed size layout, this means that less important items will need to be displayed with smaller sizes. We will need very careful balancing among all items in order to achieve the overall perception enhancement. Since there is no rigorous proof can be executed for such subjective criterion, we shall rely on extensive user study to evaluate such placement design via eye tracking experiments in order to verify its effectiveness in maximizing information perception.

In order to design a proper eye tracking experiment to quantify the subjective criterion in perception maximization, we plan to take advantage of the intrinsic invisible grid partition of the display and to model the transition of eye gaze from one block to the next as a Markov chain. The transition probably of the Markov chain can be represented by percentage of fixation time during a given pass of eye tracking study. Such Markov chain model has been successfully applied to optimize user utility in 2D search results presentation [17]. The information perception rate on a given personalized digest can be measured by the product between relative information entropy calculated from ordered ranking and the time of fixation during eye tracking study. We assume that the rate of information perception for a given digest item is inversely proportional to relative fixation time and the overall rate of

DOI: 10.2352/ISSN.2470-1173.2016.11.IMAWM-458

information perception is simply the sum of these product terms. Explicitly, the eye tracking experiments shall be able to provide adequate evidence that certain variation of summary page design can maximize information perception.

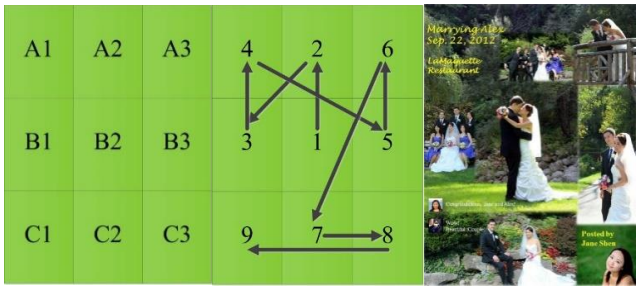

Figure 5: Illustration of "Rule of Thirds" partition (left); corresponding average fixation scan path from [33] (center); Descriptive example of a social media feeds summary for a wedding event (right).

5.2.2. Computational Aesthetics for Personalized Digest Rendering

In addition to information maximizing placement design as shown above, it is also very much desired to incorporate computational aesthetics into the design of the novel GIST we are creating in this research. It is well-known that visually pleasing presentations can significantly enhance user experience and improve information reception efficiency [48]. Since the final digest from social networking feeds is in the form of mixed media strategically placed on individual block of digest layout, the computational aesthetics approach developed for individual images and videos [21][23][24][25] cannot be simply applied to such design. A holistic approach capable of evaluating the aesthetic measures of mixed media layout, especially a balanced mixture of text and images, needs to be developed.

Adopting aesthetic composition principles in printed advertisement: A well-designed aesthetically pleasing composition of mixed media layout should look very attractive and may resemble those visually appealing print advertisement. This is because printed advertisements, especially those published in high end magazines, usually contain well-designed arrangement of multiple components in brand, pictures, and texts. The goal of advertise design has been to capture and maintain the visual attention of the customers with aesthetically appealing advertisement [32][49]. It is therefore natural to adopt several aesthetic composition principles employed in printed advertisement design in this research in order to develop a computational aesthetics approach to personalized digest rendering. These principles have been derived from extensive investigation of how human visual attention would react to certain arrangement of images, texts and contexts (brand information) and its sensitivity to the change of size and placement of these primitive components [32].

Several conclusions from the attention capture and transfer study in advertising [32][49] can indeed be adopted to enhance the composition of personalized digest via computational aesthetics. We propose to implement three confirmed attributes to boost the aesthetical appealing of the personalized digest rendering: (1) Centrality bias in overall visual attention [49], (2) Pictorial superiority effect on baseline attention [32], and (3) Text superiority effect on incremental attention [32]. To understand these attributes, baseline attention is defined as the attention devoted to an element in the layout, independent of its spatial size and other factors, and usually caused by visual pop-up effect whereas incremental attention is defined as the extra amount of attention beyond baseline due to the increases in spatial size.

The first attribute suggests that it is extremely important to place aesthetically appealing element in the center of layout because this item is key to the overall visual attention. Subsequent placement of digest items shall follow the ordered fixation scan path. For mobile device electronic rendering, we shall carry out extensive experiments to either verify or to change the fixation scan path. The second attribute suggests that pictorial digest draws significant amounts of baseline attention and is expected to be preferred form of presentation for attention capturing. However, we notice that certain text digest may be more superior in transferring the information in a more compact fashion. In particular, when this attribute is considered in conjunction with the third attribute to achieve superior effect on incremental attention, the aesthetically appealing layout for the social media digest should have a balanced handling between pictorial elements and text elements. This is because, comparing with a pictorial element, it is relatively easy to increase the spatial size of a text element to achieve the desired incremental attention to complement the well-designed baseline attention. We shall carry out a serious of eye-tracking experiments to strike a proper balance in picture and text compositions.

Color as a basic aesthetic primitive for visually appealing design: It is well known that human vision system is organized as a parallel modular system. It processes different attributes of a visual scene in different regions of the brain, including color, form, motion, depth, and spatial organization. It has been suggested that color, as one basic dimension of visual aesthetics, serves an important role in aesthetic experiences [48]. Following the well-known research with subjective experience and objective rationale of color, we shall investigate three design principles to achieve visually pleasing effects: (1) Limiting the number of strong colors, (2) Application of complementary contrast, and (3) Creating global and local colorfulness [48][26]. These three design principles will be applied to the selection of theme colors for the overall composition, selection of complementary colors for text digest or a combined text/image digest, and selection of multiple color pairs for individual blocks and whole layout.

One important guideline in color layout is to limit the number of strong colors [50]. Such underlining restriction is rooted in the modular nature of human visual system because excessive number of colors may not be properly processed at one glance. More importantly, when a particular color in visual pattern is not a salient feature of a dominant object in an image, the observer's visual attention may be distracted and the aesthetic impression is therefore missed [48]. It is thus necessary to compute the dominant colors of the highly ranked digest items and limit the number of strong colors when composing the digest layout. Another frequently adopted color guideline is the application of complementary contrast when presenting text digests superimposed over blurred image background or complementary color compositions. Such color pairs are located opposite to each other in the color circle [51]. In particular, the aesthetic appeal will be greatly enhanced if the complementary colors are among the theme colors. One possible design for text digest item is to select a pair of complementary colors for text digests and their corresponding image or non-image background. Well-designed size and color selections for the text digests are expected to deliver both high baseline attention and high incremental attention.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.11.IMAWM-458

Once the color-based design for individual blocks is complete, we will merge them to form a coherent composition. The goal of this step is to ensure that the color transitions around the boundaries between individual blocks are soft and seamless. A color controlled natural and seamless image composition scheme we have recently developed [52] can be adopted to process the block boundaries to achieve the desired seamless transitions. A fast algorithm shall be developed to in order achieve low complexity requirement in real-time generation of social media digest compositions. Once the composition step is complete, we shall calculate the colorfulness of such composition within local blocks as well as for whole layout. According to a recent research in webpage visual quality evaluation [26], colorfulness is an important visual attribute, in addition to hue and brightness, for evaluating webpage's aesthetics and attractiveness. An example of completed GIST composition is shown in Figure 5 (Right).

5.2.3. Beyond GIST Rendering: Adaptive Delivery and Clickable Interface

Once the composition of the personalized digest is complete, this magazine like GIST summary page will be delivered to the mobile users for their rendering and additional interactions. The ultimate goal of delivery is that the transport of such graphical page should be robust against the bandwidth limited and dynamic wireless links between the social media server and the mobile terminals as well as adaptive to both user and terminal contextual environments. Furthermore, this picturesque summary page should also retain the desired interface with clickable interactions to allow the users to easily access additional information about any of the digest items in the summary page.

Efficient transmission of clickable summary page over wireless connections: The transmission of GIST summary page we have composed is fundamentally different from transporting simple photos or images over Internet and wireless channels. This is because such summary page also contains the embedded clickable links to access more detailed information on any given digest items. Conventionally, the viewing component of such a complex graphics is executed on the client device, rendering the display updates received from the distant application server and enabling the user interaction. Existing transmission and remote display frameworks are not suitable for GIST consisting of digest items with diverse graphical characteristics and with embedded clickable links. We propose to design a scheme that can transfer to the client, in addition to the binary encoded digest items, semantic information about the characteristics and metadata associated with each digest [53]. This semantic delivery scheme can significantly reduce interaction latency and mitigate the bursty remote display traffic pattern resulting a robust transmission of summary pages.

Adaptive delivery of summary page with heterogeneous user/terminal contexts: As mobile devices are becoming the preferred means for most people to connect to social networks, the delivery of social media feeds to mobile users needs to accommodate the heterogeneity of mobile networks, the diversity of mobile displays, and peculiarity of terminal usage contexts. Furthermore, the spatial resolution of the summary page needs to be properly adjusted to match the actual mobile display size to avoid unnecessary waste of data transmission. Media adaptation approaches we have developed [54][55] will be applied at the gateway nodes to reduce the display size of the page to match the actual rendering dimension of the mobile devices. To combat different channel impairments among heterogeneous mobile links, we shall leverage rich experiences of the wireless multimedia [56][57] to design a joint source and channel coding strategy to allocate bit budgets between the encoding of the summary page and the application of error correction coding for optimal and adaptive delivery of summary page. Finally, we plan to address the Quality-of-Experience (QoE) issues for the mobile users when the mobile devices are viewed under a variety of usage contexts. The QoE for viewing image like GIST summary pages can be substantially influenced by whether the viewing context is indoor or outdoor and whether the user movement context is walking, riding a high speed vehicle, or simply stationary. A series of new QoE measures specifically developed for mobile media viewing, named mobile JND (Just Noticeable Difference) model [58][59], can be applied to generate matching bitstreams for delivery when the instantaneous viewing context can be properly identified. For smartphones, there are embedded sensors capable of detecting indoor-outdoor scenario (illumination sensor) as well as user movement patterns (accelerometers) [59].

Embedding clickable links in GIST summary page for hierarchical information access: The picturesque GIST summary page we created shall contain clickable interactions to allow the users to access links to acquire additional information beyond the digest items appeared in the GIST summary page. The embedded clickable links can be accomplished by creating HTML image maps for a GIST page [60][61]. Essentially, we shall make use of the "Rule of Thirds" partition of the summary page so that every digest item is associated with one coordinate to indicate the center of the "hot-spot" for embedding links. The circle shaped "hot-spot" is commonly selected for its compact representation as it needs only centroid coordinate and radius to specify a clickable area [62]. The click to access linked information will be activated whenever the cursor moves within any hot-spot area. For touch-screen displays that are commonly used in smartphones and mobile devices, clickable access shall be implemented by touching the previously defined "hot-spot" area. Once such HTML image maps are generated for the GIST summary page, all information regarding these clickable links will become the associated metadata and transmitted along with image or text components for the purpose of clickable rendering.

6. A Preliminary GIST Design Example

The holistic approach to creating ultimate GIST for browsing social media feeds as described in Section 5 would need massive implementation efforts and considerable system development. In this section, we present a preliminary GIST design example assuming that the personalized social media digest has been generated by either some existing schemes such as RankComplete algorithm [63] successfully developed for bibliography networks and visual information networks. The GIST design illustrated in this section focuses on the graphical presentation of the digest for mobile display applying the design principles outlined in Section 5.

6.1. A Multi-Objective Optimization Formulation

Based on the overall considerations as described in Section 5, we formulate the problem of GIST design as a multi-objective non-linear optimization in a constrained joint-state-space. The input of the proposed design includes the summarized and ranked multi-modal social media digest items R_i. Each term of R_i includes pre-processing images I_i, texts with their background images (t_i, I_t^i), and relevant videos with their key frames (v_i, I_v^i). The digest items form $\Gamma = (I_1, I_2, (t_3, I_t^3), (v_4, I_v^4), \cdots (t_{N_t}, I_t^{N_t}))$ which is a vector of multi-model elements already sorted in decreasing order based on their ranking score, i.e. $s_1 \geq s_2 \geq s_3 \cdots \geq s_{N_t}$. Note

DOI: 10.2352/ISSN.2470-1173.2016.11.IMAWM-458

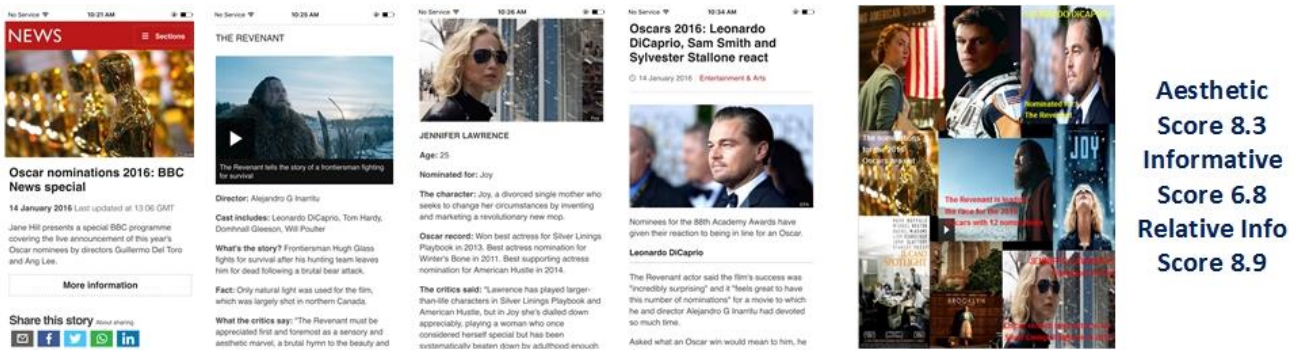

Figure 6: Illustration of multi-page traditional social media posts (left four columns) and their corresponding GIST created from these posts based on the proposed schemes as described in Section 6.

that, these ranking score will be used as a measure of information entropy to achieve information maximization as illustrated below.

Let $O_i(x_i, y_i)$ be the gravity center for i_{th} image, i.e. the center of the salient object Sal_i in I_i. Note that, we adopt the same scheme for text background image I_t^i and the video key frame I_v^i. $\left((w_{i1}, w_{i2}), (h_{i1}, h_{i2})\right)$ represents the coordinates of four border lines of the frame I_t^i within page P. Therefore, for each item R_i, we are able to represent their position and size by all these variables. The GIST design goal is to find an optimal set of position and size for each of them to create a novel representation like magazine page which is capable of (1) maximizing information perception by proper placement of digest items; (2) composing the ranked digests into highly aesthetic display.

As a result, the GIST design example can be expressed as finding an optimal set of items and their optimal position and size that minimizes the multi-objective energy $\mathrm{E}(R)$. Each digest item $\mathrm{R}(i)$, $i \in [1, N_t]$ shall completely specifies the items selection and designing within P. A given digest item consists of five tuples $R_i = \left(\left((w_{i1}, w_{i2}), (h_{i1}, h_{i2})\right), O_i(x_i, y_i), \delta_i, \mathrm{a}_i\right)$ including I_i 's size and position $\left((w_{i1}, w_{i2}), (h_{i1}, h_{i2})\right)$ within page P, gravity center coordinate $O_i(x_i, y_i)$, and the format indicator variable δ_i and indicator variable a_i . δ_i is 1 if the I_i is landscape, 0 if it is portrait. This is used to avoid distortion during the optimization process. a_i shall take the value 1 if the item R_i is selected in the GIST and 0 otherwise. The overall optimization can be formulated as an energy minimization with the following energy terms:

$$\mathrm{E}(R) = E_e(R) + \omega_b E_b(R) + \omega_p E_p(R)$$

where $E_e(R)$ is measures the information entropy, that evaluates how informative for an item selected from the set of digest items Γ is. The visual balance term $E_b(R)$ measures how well the design of magazine-like page follows the visual balance aesthetic rules. The visual perception term $E_p(R)$ measures how much users are inclined to focus on important items in the attractive regions within p. The parameter ω_b and ω_p are weights for $E_b(R)$ and $E_p(R)$, respectively.

6.2. Preliminary Results and Discussions

To validate the effectiveness of the preliminary GIST design, we have performed simulations based on considerable amount of social media posts retrieved from twitter, BBC news, Facebook and Google+. We also conducted subjective evaluations by comparing the preliminary GIST generated based on the principles outline in Section 5 against the traditional plain social media posts. We recruited 20 subjects including 12 females and 8 males with age ranging between 20 and over 35 to rate the design results. We randomly selected 30 generated GIST from their relevant source posts. The subjects were invited to browse both the original social posts and the generated GIST and were asked to rate the generated GIST by answering the following questions:

- How visually appealing is the GIST and the social posts?
- How informative is the GIST and the social posts?
- What is your overall rating for the GIST?

The subjects were requested to rate with a satisfaction score range from 1 to 10. For the informative evaluation, we also conducted an evaluation of relative information. This is because when showing a single page of GIST and complete social posts, some subjects may consider that the traditional posts deliver more information since more detailed information is shown in much more texts. However, it is unfair to just evaluate the preliminary GIST design this way because one needs to spend much more time in traditional way to absorb the same overall information contained in one-page GIST. Thus, we also asked subjects to rate the informative score by limiting their browsing within the same time period. An example is showed in Fig. 6 in which the original source social posts and the generated GIST are compared. We can see that, the information score for the GIST is only 6.8, but its relative information score is 8.9 by considering the traditional browsing to perceive the same amount of information in GIST. Therefore, viewing GIST on mobile is a much more efficient.

The scores of the subjective evaluation are presented in Fig. 7. We can see clearly from (c) that, nearly 80% of the subjects are satisfied with the GIST generated using the proposed approach. Fig. 8 shows intermedia and final optimization results of the GIST generation. Each set has two images, one represents intermediate result of optimization while the other represents final result of the optimization. Based on the proposed energy function, we can see the dynamics of the optimization process in which the position and size of the nine blocks have been changed to minimize the overall energy while still satisfying the constraints. The aesthetic scores for each of them have been increased to achieve optimization.

7. Concluding Remarks

In this paper, we have presented a suite of novel solutions to solving several interdisciplinary problems associated with creating a GIST for browsing unstructured social media feeds by the mobile users. Major innovations of the proposed research rest in its holistic exploitation of an array of inter-dependent information extraction and information presentation attributes in order to generate an ultimate GIST for mobile browsing. We illustrated

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.11.IMAWM-458

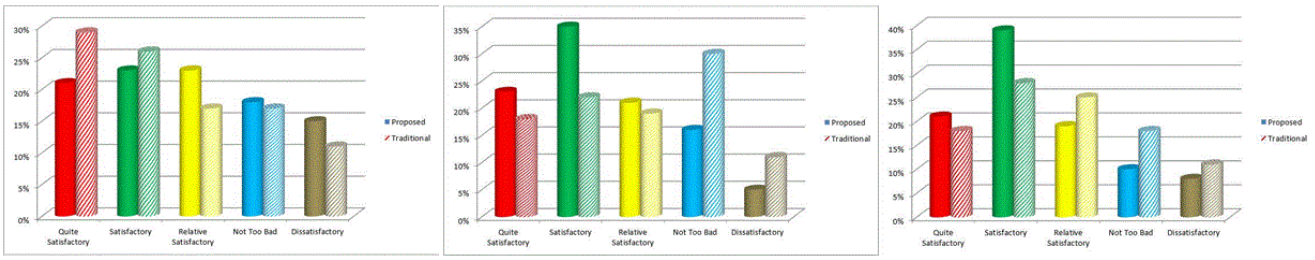

Figure 7: Results of subjective evaluation of the GIST generation and corresponding traditional social media posts. Left: informative evaluation results; Middle: aesthetic evaluation results; Right: overall evaluation results.

Figure 8: Illustration of intermediate and final optimization results. Left: GIST optimization for Victoria Secret; Right: GIST optimization for New Year

details towards developing a systematic approach to address major challenges in creating both informative and visually pleasing multimedia presentations from social media feeds with severe constraints: heterogeneous information type, personalized social context, dynamic content generation, aesthetic presentation, robust delivery, and effective rendering on mobile devices.

8. Acknowledgement

The author would like to thank several collaborators for their valuable contributions to the research presented in this paper. In particular, the author would like to thank Prof. Jiebo Luo of University of Rochester, Dr. Tao Mei of Microsoft Research Asia, Dr. Wenyuan Yin of Yahoo Research, and Miss Shuang Ma of State University of New York at Buffalo for carrying out related joint research and for their permission to use sample results in this paper. This work is supported in part by NSF Grant 1405594.

References

[1] https://zephoria.com/top-15-valuable-facebook-statistics/

[2] M. S. Bernstein, B. Suh, L. Hong, J. Chen, S. Kairam, E. H. Chi, "Eddi: interactive topic-based browsing of social status streams," *Proceedings of UIST'2010*.

[3] J. Sankaranarayanan, H. Samet, B. Teitler, M. D. Lieberman, and J. Sperling, "TwitterStand: News in tweets," *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2009.

[4] M. Naaman, J. Boase, and C. Lai, "Is it Really About Me? Message Content in Social Awareness Streams," *Proceedings of CSCW*, 2010.

[5] X. Jin, A. Gallagher, J. Han, J. Luo, "Wisdom of Social Multimedia: Using Flickr for Prediction and Forecast," *Proceedings of ACM Multimedia Conference*, 2010.

[6] X. Jin, C. Wang, J. Luo, X. Yu and J. Han, "LikeMiner: A System for Mining the Power of 'Like' in Social Media Networks," *Proceedings of 2011 ACM SIGKDD*, San Diego, Aug. 2011.

[7] http://www.business2community.com/infographics/information-overload-users-burnt-social-media-infographic-0876828#MrUZrrzZ2uQ3yICl.97

[8] http://www.adweek.com/socialtimes/social-media-overload/488800

[9] N. UzZaman, J. P. Bigham and J. F. Allen, "Multimodal Summarization of Complex Sentence," *Proceedings of International Conference on Intelligent User Interfaces (IUI),* Palo Alto, California, 2011.

[10] B. T. Truong and S. Venkatesh, "Video abstraction: A systematic review and classification," *ACM Trans. Multimedia Comput. Commun. Appl.*, Vol. 3, No. 1, February 2007.

[11] C. Yang, J. Yuan, and J. Luo, "Towards Scalable Summarization of Consumer Videos via Sparse Dictionary Selection," *IEEE Trans. on Multimedia*, Special Issue on Object and Event Classification in Large-Scale Video Collections, 14(1): 66-75, February 2012.

[12] B. Zhao, L. Fei-Fei and E. P. Xing, "Online Detection of Unusual Events in Videos via Dynamic Sparse Coding," *Proceedings of IEEE CVPR,* 2011.

[13] L. Kennedy, M. Naaman, S. Ahern, R. Nair, and T. Rattenbury. How Flickr helps us make sense of the world: context and content in community-contributed media collections. *Proc. ACM MM* 2007.

[14] David Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, 2004.

[15] Z. Li, S. Shi and L. Zhang. Improving relevance judgment of web search results with image excerpts. *In Proc. WWW '08,* 2008

[16] J. Teevan, E. Cutrell, D. Fisher, S. M. Drucker, G. Ramos, P. Andr′e, and C. Hu. Visual snippets: summarizing web pages for search and revisitation. *In Proc. CHI '09,* pages 2023–2032, New York, 2009.

[17] B. Jiao, L. Yang, J. Xu, F. Wu: Visual summarization of web pages. *SIGIR 2010*: 499-506

[18] F. Chierichetti，R. Kumar，P. Raghavan, Optimizing Two-Dimensional Search Results Presentation, *WSDM'11.*

[19] P. Sandhaus, M. Rabbath, I. Erbis and S. Boll, "Blog2Book: transforming blogs into photo books employing aesthetic principles," *ACM MM'10*, October 25–29, 2010, Firenze, Italy.

[20] M. Rabbath, P. Sandhaus and S. Boll, "Automatic Creation Of Photo Books From Stories In Social Media," *TOMCCAP*, 2011
[21] P. Sandhaus, M. Rabbath and S. Boll, "Employing Aesthetic Principles for Automatic Photo Book Layout," In *Proceedings of The 17th International Conference On Multimedia Modeling*, 2011
[22] D. Joshi, R. Datta, E. Fedorovskaya, J. Li, J. Z. Wang, J. Luo, "Computational inference of aesthetics, mood, and emotion in images," *IEEE Signal Processing Magazine*, 28(5): 94-115, 2011.
[23] Y. Ke, X. Tang and F. Jing, "The design of high-level features for photo quality assessment," *Proc. of CVPR 2006*, pp. 419-426
[24] Y. Luo and X. Tang, "Photo and video quality evaluation: focusing on the subject," *Proc. of ECCV 2008*, pp. 386-399
[25] S. Dhar, V. Ordonez and T. Berg, "High level describable attributes for predicting aesthetics and interestingness," *Proc. CVPR2011*.
[26] C. Wu, Y. Chen, B. Li, W. Hu, "Evaluating the visual quality of web pages using a computational aesthetic approach," *Proc. 4th ACM WSDM2011*, February, 2011, Hong Kong, China.
[27] W. Yin, T. Mei and C. W. Chen, "Automatic generation of social media snippets for mobile browsing," *Proc. of ACM MM 2013*, pp. 927-936, October 2013
[28] R. Mihalcea, and P. Tarau, "TextRank: Bringing Order into Texts," in *Proc. of the Conference on Empirical Methods in Natural Language Processing*, 2004.
[29] Z. Li, S. Shi, L. Zhang, "Improving relevance judgment of web search results with image excerpts," in *Proc. of International Conference on World Wide Web*, 2008.
[30] J. Teevan, E. Cutrell, D. Fisher, S.M. Drucker, G. Ramos, P. André, C. Hu, "Visual snippets: summarizing web pages for search and revisitation," in *Proc. of International Conference on Human Factors in Computing Systems,* pp. 2023-2032, 2009.
[31] S. Kaasten, S. Greenberg, and C. Edwards, "How people recognize previously seen Web pages from titles, URLs and thumbnails," in *Proc. of HCI*, pp. 247-265, 2002.
[32] R. Pieters and M. Wedel, "Attention capture and transfer in advertising: Brand, pictorial, and text-size effects," *Journal of Marketing*, Vol. 68, No. 2 (Apr., 2004), pp. 36-50
[33] Marilu Smit, "Optimising advertising ROI on Print in a rapidly changing economic and media landscape - the contribution of eye tracking now and into the future," *Proc. of Southern African Marketing Research Association (SAMRA) conference*, August 2012.
[34] M. Bauerly and Y. Liu, "Computational modeling and experimental investigation of effects of compositional elements on interface and design aesthetics," *International Journal of Human-Computer Studies*, Vol. 64, No. 8, pp. 670-682, August 2006
[35] R. Mihalcea and P. Tarau, "TextRank: Bridging order into text," *Proc. of Conference on Empirical Methods in Natural Language Processing*, pp. 404-411, 2004
[36] Flipboard. http://www.flipboard.com
[37] U. Naushad, P. Jeffrey and F. A. James, "Multimodal Summarization of Complex Sentence. Proceedings of International Conference on Intelligent User Interfaces (IUI), Palo Alto, California, 2011.
[38] Z. Youjie, L. Jiebo. Geo-Location Inference on News Articles via Multimodal pLSA. Proceedings of ACM Conf. on Multimedia, Nara, Japan, October 2012.
[39] J. Wang, H. Xu, H. X-S Hua and S. Li, "Hybrid Image Summarization. Proceedings of ACM Multimedia, 2011.
[40] C. Dang and H. Radha, "Heterogeneity Image Patch Index and Its Application to Consumer Video Summarization," *IEEE Transactions on Image Processing*, 2014.
[41] D. Zhou, J. Weston, A. Gretton, etc., "Ranking on data manifolds," *Proc. of Int. Conf. on Neural Information Processing Systems*, 2003
[42] Z. Luming, X. Yingjie, M. Kuang, etc.. An Effective Video Summarization Framework toward Handheld Devices. *IEEE Transactions on Industrial Electronics,* 2015.
[43] S. Lee, M. Kim and S. Lee, "Constructing compact and effective graphs for recommender systems via node and edge aggregations," *Proc. Int. Conf. on Expert Systems with Applications*, 2014.
[44] X. Jiejun, L. Tsai-Ching. Seeing the Big Picture from Microblogs: Harnessing Social Signals for Visual Event Summarization. Proceedings of ACM Conf. on Intelligent User Interfaces, 2015.
[45] Bert P. Krages, "*The Art of Composition,"* Allworth Communications, Inc., 2005
[46] R. R. Yager, "On ordered weighted averaging aggregation operators in multi-criteria decision making," *IEEE Transactions on Syst., Man and Cybernetics*, 18, 183-190, 1988
[47] N. Usunier, D. Buffoni, P. Gallinari, "Ranking with Ordered Weighted Pairwise Classification," *Proceedings of the 26th International Conf. on Machine Learning*, Montreal, Canada, 2009
[48] G. Peters, "Aesthetic primitives of images for visualization," Proc. of 11th International Conf. on Information Visualization (IV'07), 2007
[49] A. S. Atalay, H. O. Bodur, and D. Rasolofoarison, "Shining in the center: Central gaze cascade effect on product choice," *Journal of Consumer Research*, Electronically published May 03, 2012
[50] W. Lidwell, K. Holden, and J. Butler*, Universal Principles of Design*, Beverly, MA: Rockport Publishers, 2003
[51] Johannes Itten, *The art of color: the subjective experience and objective rationale of color*, Translated by Ernst van Haagen, New York: Van Nostrand Reinhold, 1973
[52] W. Yang, J. Zheng, J. Cai, S. Rahardja, C. W. Chen, "Natural and seamless image composition with color control", *IEEE Trans. on Image Processing*, vol. 18, no. 11, pp. 2584-2592, Nov. 2009.
[53] P. Simoens, B. Joveski, L. Gardenghi, I. J. Marshall, B. Vankeirsbilck, M. Mitrea, F. Preteux, F. De Turck, B. Dhoedt, "Optimized mobile thin clients through a MPEG-4 BiFS semantic remote display framework," *Multimedia Tools and Applications*, Vol. 61, No. 2, pp. 447-470, Nov. 2012
[54] W. Yin, J. Luo, C. W. Chen, "Event-based Semantic Image Adaptation for User-centric Mobile Display Devices," *IEEE Transactions on Multimedia*, Vol. 13, No. 3, pp. 432-442, June 2011.
[55] Y. Wang, H. Li and C. W. Chen, "An attention-information-based spatial adaptation framework for browsing videos via mobile devices," *Journal of Applied Signal Processing* special issue on Video Adaptation for Heterogeneous Environments, Vol. 2007, Article ID 25415, 12 pages, doi:10.1155/2007/25415.
[56] Z. Li, Q. Sun, Y. Lian and C. W. Chen, "Joint source-channel-authentication resource allocation and unequal authenticity protection for multimedia over wireless networks," *IEEE Trans. on Multimedia,* Vol.9, No.4, pp.837-850, June, 2007
[57] X. Zhu and C. W. Chen, "A joint layered scheme for reliable and secure mobile JPEG-2000 streaming," *ACM Trans. Multimedia Computing, Communications and Applications*, Vol. 8, No. 3, Article 30, July 2012
[58] J. Xue and C. W. Chen, "Study on Perception of Mobile Video with Surrounding Contextual Influences," *Proc. IEEE QoMEX 2012* (Quality of Multimedia Experiences) July 2012
[59] J. Xue and C. W. Chen, "Mobile JND: Environment Adapted Perceptual Model and Mobile Video Quality Enhancement," *Proc. ACM Multimedia Systems*, Chapel Hill, NC, February 2012
[60] J. S. Guedalia, "Method and system for HTML-driven interactive image client," *US Patent 6,121,970,* 2000
[61] J. S. Guedalia, "Method and system for HTML-driven interactive image client," *US Patent 6,356,283,* 2002
[62] http://www.javascriptkit.com/howto/imagemap.shtml
[63] L. Cao, A. D. Pozo, X. Jin, J. Luo, J. Han and T. S. Huang, "RankComplete: Simultaneous ranking and clustering of information networks," Neurocomputing, Vol. 95, No. 15, Pages 98–104, October 2012