DEFINING AESTHETIC PRINCIPLES FOR AUTOMATIC MEDIA GALLERY LAYOUT FOR VISUAL AND AUDIAL EVENT SUMMARIZATION BASED ON SOCIAL NETWORKS

Thomas Steiner[1†], *Ruben Verborgh*[2]

[1]Universitat Politècnica de Catalunya
Llenguatges i Sistemes Informàtics (LSI)
08034 Barcelona, Spain
tsteiner@lsi.upc.edu,
gabarro@lsi.upc.edu

Joaquim Gabarro[1], *Rik Van de Walle*[2]

[2]Ghent University – IBBT,
ELIS – Multimedia Lab,
B-9050 Ledeberg-Ghent, Belgium
ruben.verborgh@ugent.be,
rik.vandewalle@ugent.be

ABSTRACT

In this paper, we present and define aesthetic principles for the automatic generation of media galleries based on media items retrieved from social networks that—after a ranking and pruning step—can serve to authentically summarize events and their atmosphere from a visual and an audial standpoint.

Index Terms— event summarization, media galleries, social networks, media item ranking, media layout, aesthetics

1. INTRODUCTION

Mobile devices such as smartphones, together with social networks, enable people to create, share, and consume media items like videos or images. They accompany their owners almost everywhere and are thus omnipresent at all sorts of events. Given a stable network connection, event-related media items and microposts are published on social networks during events and afterwards. Ranked media items stemming from multiple social networks can serve to create authentic media galleries that illustrate events and their atmosphere. A key feature for this task is the semantic enrichment of media items and associated microposts and the extraction of **visual**, **audial**, **textual**, and **social** features. Based on this set of features, additional **aesthetic** features can be defined and exploited to obtain appealing and harmonic media galleries.

2. RELATED WORK

While enormous efforts have been made to extract those features from media items and microposts on social networks in *isolation*, to the best knowledge of the authors, remarkably less initiatives concern the extraction and the application of all those features *in combination* for *all* types of media items, including microposts. In [1], Sandhaus *et al.* consider visual and aesthetic features for the automatic creation of photo books. Obrador *et al.* use visual and aesthetic features for a category-based approach to automatically assess the aesthetic appeal of photographs [2]. In [3], Knees *et al.* use audial and textual features for the automatic generation of music playlists. Choudhury *et al.* show in [4] how social and textual features can be used to achieve precise detection results of named entities and significant events in sports-related microposts. In [5], Davidson *et al.* show how visual, textual, and social features can be used for personalized video recommendations. A service called Storify [6] lets users manually combine microposts, images, videos, and other elements onto one page for the purpose of storytelling or summarizing an event, and share stories permanently on the Web. Finally, social networks present images and videos often in grid-like galleries[1], sometimes scaled based on the amount of comments.

3. MEDIA ITEM RANKING CRITERIA

In this section, we describe several feature categories that can serve to rank media items retrieved from social networks. We assume (and are working on) media item extractors that, given event-related search terms, extract raw binary media items and associated microposts from multiple social networks.

Visual This category regards the contents of images and videos. We distinguish *low-* and *high-level* visual ranking criteria. High-level criteria are, *e.g.*, logo detection, face recognition, and camera shot separation. Low-level criteria are, *e.g.*, size, resolution, duration, geolocation, and time. Via OCR, contained characters can be treated as textual features.

Audial This category regards the audio track of videos. *High-level* ranking criteria are the presence or absence of silence, music, speech, or a mixture thereof. Similar to visual features before, audial *low-level* features are the average bit rate, volume, possibly distorted areas, *etc*. Through audio-transcription, speech can be converted to a textual feature.

Textual This category regards the microposts that accompany media items. Typically, microposts provide a description of media items. Using named entity disambiguation tools, textual content can be linked to LOD cloud concepts [7].

Social This category regards social network effects like shares, mentions, view counts, expressions of (dis)likes, user diversity, *etc*. Prior work [8] allows us to not only examine

[†] T. Steiner is partially EC-funded (Grant No. 248296 FP7 I-SEARCH).

[1] http://twitpic.com/904yka/full

978-1-4673-0726-0/12/$31.00 © 2012 IEEE

these effects on a *single* social network, but in a *network-agnostic* way across multiple social networks.

Aesthetic This category regards the desired outcome after the ranking, *i.e.*, the media gallery that illustrates a given event and its atmosphere. Studies exist for the aesthetics of automatic photo book layout [1], photo aesthetics *per se* [2], video and music playlist generation [5, 3], however media gallery composition requires mixing video *and* image media items.

4. MEDIA GALLERY AESTHETICS

Definition A media gallery in our context is a compilation of images or videos retrieved from social networks that are related to a given event. Given a set $M = \{m_1, ..., m_n\}$ of media items related to a certain event, and given a ranking formula f, the subset $M' \subset M$ is the result after the application of f to M: $f(M) = M'$. Each media item m_i can either be an instance of video or image. For each point t_x on a timeline T, the state of the media gallery at t_x is defined for each media item m_i as a set S_x of n tuples $s_{x,i}$, where $s_{x,i} = \langle \mathit{left}, \mathit{top}, \mathit{width}, \mathit{height}, \mathit{alpha}, \mathit{z\text{-}index}, \mathit{animation}, \mathit{start}, \mathit{playing}, \mathit{volume} \rangle$. The first 6 properties are defined as in CSS, the $\mathit{animation}$ property allows for the definition of CSS transitions and transformations as defined in [9, 10], the start property defines the start time in a video. A schematic media gallery at t_x can be seen online[2].

Audial aesthetics We recall the purpose of our media galleries: to illustrate an event and its atmosphere. Audial aesthetics thus consist of aspects like volume level normalization, avoiding multiple videos playing music in parallel, smooth transitions, *etc.* We remark that through selective mixing of audio tracks of event-related videos, "noise clouds" very characteristic for the event atmosphere can be observed.

Visual aesthetics Visual aesthetics are determined by the composition, *i.e.*, the relation of images to videos *globally*, *per coherent scene*, and per *point in time*. In order not to overcharge the perceptive capacity of viewers, the number of visible (moving) media items at a time should be limited. Depending on the event, a consistent or a contrasty overall appearance of items may be desired, also for transitions.

5. PRELIMINARY RESULTS AND CONCLUSION

After introducing media item ranking criteria as well as aesthetic audial and visual principles for media galleries, we performed some first experiments. A manual evaluation revealed positive results on a set of media items that were collected for events in recent history (among others the *Costa Concordia* disaster in Italy, the *Consumer Electronics Show* in Las Vegas, the global *Blackout SOPA* campaign). Especially when combining mixed content types, *i.e.*, videos and images, users expressed they preferred media items to crossfade smoothly rather than having sharp contrasts between transitions. In consequence, we will put special emphasis on shot detection with video content to ensure a harmonic holistic perception of mixed content in media galleries. In the coming months, we will apply those principles to a large collection of media items related to events, and, via automatic multivariate blind tests, measure user engagement for different feature configurations on both desktop and mobile.

6. REFERENCES

[1] Philipp Sandhaus, Mohammad Rabbath, and Susanne Boll, "Employing Aesthetic Principles for Automatic Photo Book Layout," in *Proceedings of the 17th International Conference on Advances in Multimedia Modeling – Volume Part I (MMM 2011)*, 2011, pp. 84–95.

[2] Pere Obrador, Michele Saad, Poonam Suryanarayan, and Nuria Oliver, "Towards Category-Based Aesthetic Models of Photographs," in *Proceedings of the 18th Int. Conference on Advances in Multimedia Modeling – Volume Part I (MMM 2012)*, pp. 63–76. Springer, 2012.

[3] Peter Knees, Tim Pohle, Markus Schedl, and Gerhard Widmer, "Combining Audio-based Similarity with Web-based Data to Accelerate Automatic Music Playlist Generation," in *Proceedings of the 8th ACM Int. Workshop on Multimedia Information Retrieval*, New York, NY, USA, 2006, MIR '06, pp. 147–154, ACM.

[4] Smitashree Choudhury and John Breslin, "Extracting Semantic Entities and Events from Sports Tweets," in *Making Sense of Microposts*, 2011, pp. 22–32.

[5] James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, et al., "The YouTube Video Recommendation System," in *Proceedings of the 4th ACM Conference on Recommender Systems*, New York, NY, USA, 2010, RecSys '10, pp. 293–296, ACM.

[6] Kelly Fincham, "Review: Storify (2011)," *Journal of Media Literacy Education*, vol. 3, no. 1, 2011.

[7] Thomas Steiner, Ruben Verborgh, Joaquim Gabarro, et al., "Adding Meaning to Facebook Microposts via a Mash-Up API and Tracking its Data Provenance," in *Proceedings of the 7th Int. Conference on Next Generation Web Services Practices*, 2011, pp. 342–345.

[8] Houda Khrouf, Ghislain Atemezing, Giuseppe Rizzo, Raphaël Troncy, et al., "Aggregating Social Media for Enhancing Conference Experience," in *Proceedings of the 1st Int. Workshop on Real-Time Analysis and Mining of Social Streams*, 2012, (accepted for publication).

[9] Dean Jackson, David Hyatt, Chris Marrin, and L. David Baron, "CSS Transitions Module Level 3," Tech. Rep., W3C, 2012, `http://www.w3.org/TR/css3-transitions`.

[10] Simon Fraser, Dean Jackson, David Hyatt, Chris Marrin, et al., "CSS Transforms," Tech. Rep., W3C, 2012, `http://www.w3.org/TR/css3-transforms/`.

[2] `http://twitpic.com/9je27h/full`