# EE-USAD: ACM MM 2018 Workshop on Understanding Subjective Attributes of Data focus on Evoked Emotions

Xavier Alameda-Pineda
Inria, Perception Team
Grenoble, France
xavier.alameda-pineda@inria.fr

Miriam Redi
Wikimedia Foundation
San Francisco, California
mredi@wikimedia.org

Nicu Sebe
University of Trento
Trento, Italy
niculae.sebe@unitn.it

Shih-Fu Chang
Columbia University
New York, New York
sfchang@cs.columbia.edu

Jiebo Luo
University of Rochester
Rochester, New York
jluo@cs.rochester.edu

## ABSTRACT

The series of events devoted to the computational Understanding of Subjective Attributes (e.g. beauty, sentiment) of Data (USAD) provide a complementary perspective to the analysis of tangible properties (objects, scenes), which overwhelmingly covered the spectra of applications in multimedia. Partly fostered by the widespread usage of social media, the analysis of subjective attributes has attracted lots of attention in the recent years, and many research teams at the crossroads of multimedia, computer vision and social sciences, devoted time and effort to this topic. Among the subjective attributes there are those assessed by individuals (e.g. safety, interestingness, evoked emotions [2], memorability [3]) as well as aggregated emergent properties (such as popularity or virality [1]). This edition of the workshop (see below for the workshop's history) is devoted to the multimodal recognition of evoked emotions (EE).

**ACM Reference Format:**
Xavier Alameda-Pineda, Miriam Redi, Nicu Sebe, Shih-Fu Chang, and Jiebo Luo. 2018. EE-USAD: ACM MM 2018 Workshop on Understanding Subjective Attributes of Data focus on Evoked Emotions. In *2018 ACM Multimedia Conference (MM '18), October 22–26, 2018, Seoul, Republic of Korea.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3240508.3243721

## 1 INTRODUCTION

The scientific scope of the EE-USAD workshop stands on three areas of the usual ACM Multimedia program: Multimedia and Vision (Understanding), Multimedia Art, Entertainment and Culture (Experience), and Social Multimedia (Engagement). The first, because many recent and highly relevant studies to the proposed to topic appeared in the computer vision community. The second, because evoked emotions are a priori strongly tied to the culture, and hence it is worth boosting research in this direction. The third, not only because many social media platforms allow for "emotional" reactions to other user's content, but also because it is worth investigating the evoked emotions from a psychological and sociological perspectives.

That being said, the proposed workshop is new to the standard program of ACM MM'18, in the sense that, to our knowledge, there have been no previous efforts to foster research on the multimodal understanding of subjective attributes of data, with special emphasis on evoked emotions. In addition, we believe our proposal is timely since several relevant studies have been sparsely published, and we now have a timely opportunity to give a more formal impulse in this direction.

Previous editions of the USAD events (see below) managed to attract several high-quality submissions. We believe that the focus on evoked emotions is highly relevant for the multimedia community. Such workshop will provide a discussion forum for researchers to exchange about the preliminary and consolidated research published in (and outside) the workshop. This forum will attract researchers to submit their studies to the proposed ACM MM'18 workshop.

## 2 HISTORY

This would be the second edition of an USAD event at ACM Multimedia, after the MUSA2 Workshop organized in 2017. Last year the workshop was very successful: we had 14 full paper submissions, out of which we accepted 8 (4 oral, 4 posters) and a substantial number of attendees. Recently, we also organized related events in neighboring communities (Special session at ICMR 2017 and a Special Issue in ACM Transactions of Multimedia). This year, we also offered a vision-specific version of the workshop to the computer vision community: the V-USAD workshop was held in conjunction with CVPR 2018 and gathered contributions addressing the problem of detecting subjective attributes in visual data.

## 3 TOPICS

Given the subjective nature of evoked emotions, many new challenges arise when attempting to automatically detect them from multimedia data, or to perform large-scale retrieval, including:

(1) Collecting huge amounts of annotation reflecting subjective judgements of evoked emotions, as opposed to binary, objective annotations.
(2) Designing features or learning representations specifically tailored for the multimodal recognition of evoked emotions


*MM '18, October 22–26, 2018, Seoul, Republic of Korea*

ACM ISBN 978-1-4503-5665-7/18/10.
https://doi.org/10.1145/3240508.3243721

(3) Reliably evaluating the accuracy of the detectors.
(4) Modeling the subjectivity and the cultural specificity of evoked emotions.
(5) Translating (social) psychology theories into computational approaches to systematically understand evoked emotions.

## 4 PROGRAM

The program will follow the next guidelines:

- Welcome and general remarks
- Invited talk: TBA
- Oral presentation #1
- Oral presentation #2
- Invited talk: TBA
- Poster session
- Invited talk: Prof. Tat-Seng Chua
- Conclusions

There are four papers accepted at the workshop with the following titles: Depth-Aware Image Colorization Network; Artificial Empathic Memory: Enabling Media Technologies to Better Understand Subjective User Experiences; What makes natural scene memorable?; and Perceptual Similarity Ranking of Temporal Heatmaps Using Convolutional Neural Networks.

The first paper studies the bleeding problem, which remains a challenging issue in image colorization. This happens when different objects share the same color when they are nearby, leading to the boundary between objects looks unnatural. The authors study how to combine depth information into a neural network and achieve better image colorization. The reasons to integrate depth information are twofold: (1) Depth information clearly provides boundary information between objects, and (2) depth information is commonly available as the development of RGB-D cameras. Depth information was not considered in image colorization before. The proposed method is evaluated from both objective and subjective perspectives, demonstrating that better colorization results can be obtained when depth information is further considered.

The second article discusses artificial empathic memory technologies, which deal with episodic memories, which revolve around events that took place in a person's past and are typically defined by a time, place, emotional associations, and other contextual information. They form an important driver for a personâĂŹs emotional and cognitive interpretation of what is currently happening. This includes interactions with media technologies. Is is argued that current approaches for personalization of these applications are neither aware of what episodic memories are triggered in their users, nor of their emotional interpretation of those memories. This is a serious limitation, because it prevents them from correctly interpreting and anticipating users' experiences. In short, such technologies lack empathy. In this position paper, it is argued that media technologies need an Artificial Empathic Memory (AEM) of their users to address these issues for the personalization of their interactions. A psychologically inspired architecture is proposed, allowing to examine the challenges to be solved, and highlight how existing research can become a starting point for overcoming them.

The third paper deals with image memorability, in particular aiming to a clear understanding and reliable estimation of natural scene memorability remain elusive: âĂIJwhat exactly makes natural scene memorableâĂİ. Specifically, the authors first build LNSIM, a large-scale natural scene image memorability database (containing 2,632 images and memorability annotations). Then, the mine the LNSIM database to investigate how low-, middle- and high-level handcrafted features affect the memorability of natural scene. In particular, it is found that high-level features of the scene category are rather correlated with natural scene memorability. Thus, a deep neural network based natural scene memorability (DeepNSM) predictor is proposed, which takes advantage of scene category. Finally, the experimental results validate the effectiveness of DeepNSM.

The last paper studies similarity ranking. While current approaches work well on low-dimensional datasets, it becomes difficult to define similarity for more complex data types, like event sequences with multidimensional attributes. Often, its definition needs to be manually tuned according to the target domain or dataset. Visualizations are similarly manually tuned by analysts and can contain important clues about relevant features. In this paper, the authors propose the use of computer vision techniques on visualizations as a means for similarity ranking. Sequential datasets are visualized as temporal heatmaps and shown through user studies with 132 participants that humans agree in ranking results to a query based on perceptual similarity. Heat2Vec, a convolutional neural network (CNN) to learn latent representations from heatmaps using color, opacity, and position, is designed and implemented. The method is evaluated against 11 baselines using a wide range of techniques showing that Heat2Vec provides rankings consistently in line with human-annotated similarity ranking.

## 5 ORGANIZERS

Our organization committee is very diverse in terms of expertise, topics, geographic location and seniority.

- Dr. Xavier Alameda-Pineda (Research Scientist, Perception Group, Inria Grenoble)
- Dr. Miriam Redi (Research Scientist, Wikimedia Foundation, USA)
- Prof. Nicu Sebe (Professor and Head of Dept. of Information Engineering and Computer Science, University of Trento)
- Prof. Shih-Fu Chang (Richard Dicker Professor and senior Vice Dean of the School of Engineering and Applied Sciences at Columbia University)
- Prof. Jiebo Luo (Professor of the Dept. of Computer Science, University of Rochester)

## REFERENCES

[1] Xavier Alameda-Pineda, Andrea Pilzer, Dan Xu, Nicu Sebe, and Elisa Ricci. 2017. Viraliency: Pooling Local Viraliry. In *IEEE International Conference on Computer Vision and Pattern Recognition.*
[2] Brendan Jou, Tao Chen, Nikolaos Pappas, Miriam Redi, Mercan Topkara, and Shih-Fu Chang. 2015. Visual affect around the world: A large-scale multilingual visual sentiment ontology. In *ACM International Conference on Multimedia.*
[3] Aliaksandr Siarohin, Gloria Zen, Cveta Majtanovic, Xavier Alameda-Pineda, Elisa Ricci, and Nicu Sebe. 2017. How to Make an Image More Memorable? A Deep Style Transfer Approach. In *ACM International Conference on Multimedia Retrieval.*