



Association for
Computing Machinery

Advancing Computing as a Science & Profession



NEWS RELEASE

Contact: Jim Ormond
212-626-0505
ormond@acm.org

VOLUME, VARIETY AND USES OF MULTIMEDIA CONTENT EXPLORED AT ACM MM '19

Growing Synergy of AI and Digital Content among Topics Presented

New York, NY, October 2, 2019 –The Association for Computing Machinery’s Special Interest Group on Multimedia (SIGMM) will host the 27th [ACM International Conference on Multimedia \(ACM MM\)](#) in Nice, France, from October 21-25, 2019. Multimedia research focuses on integration of the multiple perspectives offered by different digital modalities including images, text, video, music, sensor data and spoken audio. Since 1993, the ACM MM conference has been bringing together researchers and practitioners from academia and industry to present innovative research results and discuss recent advancements.

“The use of computers to distribute and share digital content is one of the hallmarks of the way we live today,” notes Martha Larson, ACM MM 2019 General Co-chair and Professor, Radboud University, the Netherlands. “Multimedia computing has come into its own not just because of the volume and variety of content that is shared digitally, but also because of the introduction of artificial intelligence techniques that are increasingly used to analyze the content and make it available for use. The annual MM conference is the premier showcase of where our field is today and where it is headed.”

ACM MM 2019 HIGHLIGHTS

Keynote Speakers

“Using Artificial Intelligence to Preserve Audiovisual Archives: New Horizons, More Questions”

Jean Carrive, National Audiovisual Institute (INA), France

Carrive, who holds a Masters’ degree in Artificial Intelligence, Pattern Recognition and Application, as well as a Phd in Computer Science from University Pierre and Marie, Curie, Paris, will discuss his work as Deputy Head of Research and Innovation at the French National Audiovisual Institute (INA). INA’s mission is to preserve France’s radio and television archives and make them accessible to researchers working in the humanities and social sciences. Carrive will explain how, for INA’s archivists and librarians, AI’s assistance facilitates the documentation process but also poses questions about the impact of these technologies on professional practices, as well as on the scalability of these technologies over time.

“EU Data Protection Law: An Ally for Scientific Reproducibility?”

Mireille Hildebrandt, Vrije Universiteit Brussel, Belgium and Radboud University, The Netherlands

In this keynote, Hildebrandt will provide a crash course in the underlying ‘logic’ of the General Data Protection Regulation (GDPR), the new EU law governing data privacy and security, with a focus on what is relevant for inferences based on multimedia content and metadata. She will explain how the “purpose limitation principle” is the guiding rationale of EU data protection law—protecting individuals against incorrect, unfair or unwarranted targeting. In the second part of the keynote she will explain how the purpose limitation principle relates to machine learning research design, requiring keen attention to specific aspects of methodological integrity.

“Broadening Participation to Computing”

Maria Menendez-Blanco, Pernille Bjørn, University of Copenhagen

As computing becomes integrated in fundamental ways in healthcare, labor markets, and political processes, questions about who participates and makes decisions in developing digital technologies are becoming increasingly crucial and unavoidable. In their keynote, Menendez-Blanco and Bjørn will address this pressing issue by discussing gender issues in computing through their experience running FemTech.dk, an action-research project focused on creating opportunities for young women in computing. They will share insights they’ve gained through FemTech.dk, including the importance of changing computer science departments from “within,” the relevance of challenging stereotypical and narrow definitions of computer science, and the instrumentality of interactive artefacts in prompting change.

“Inventing Narratives of the Anthropocene: Microclimate Machines and Arts & Sciences Installations”

Jean-Marc Chomaz, Laboratoire d’Hydrodynamique, CNRS/École Polytechnique, France

Chomaz has been involved as a researcher and artist in “arts & sciences” projects in all disciplines (circus, theatre, design, contemporary art, music, etc.). His approach tries to give direct access to an imaginary world using scientific language and concepts not to demonstrate, but to make sense. His keynote will explore the possibility for arts & sciences research and creation to contribute to such critical thinking. Chomaz will also provide alternative narratives of the anthropocene (the current geological age—characterized by human dominance of the planet), which will be illustrated on several examples of microclimate machines and installations.

Accepted Papers (partial list)

Papers are divided into six separate topic areas: engaging users with multimedia; multimedia experience; multimedia systems; multimodal fusion; vision and language; and media interpretation. Visit the [ACM Multimedia 2019 Program Page](#) for a full list of accepted papers.

“Human-imperceptible Privacy Protection against Machines”

Z. Shen, S. Fan, Y. Wong, T.-T. Ng and M. Kankanhalli, National University of Singapore

Privacy concerns with social media have recently been under the spotlight, due to a few incidents on user data leakage on social networking platforms. With the current advances in machine learning and big data, computer algorithms often act as a first-step filter for privacy breaches, by automatically selecting content with sensitive information, such as photos that contain faces or vehicle license plate. In this paper the authors propose a novel algorithm to protect the sensitive attributes against machines, meanwhile keeping the changes imperceptible to humans.

“Towards Automatic Face-to-Face Translation”

Prajwal Renukanand, Rudrabha Mukhopadhyay, Jerin Philip, Abhishek Jha, Vinay Namboodiri and C.V. Jawahar, Indian Institute of Technology (IIT)

In light of the recent breakthroughs in automatic machine translation systems, the authors propose a novel approach of what they term as "Face-to-Face Translation." As today's digital communication becomes increasingly visual, they argue that there is the need for systems that can automatically translate a video of a person speaking in language A into a target language B with realistic lip synchronization. The authors then move towards "Face-to-Face Translation" by incorporating a novel visual module, LipGAN for generating realistic talking faces from the translated audio.

"Eye in the Sky: Drone-Based Object Tracking and 3D Localization"

Haotian Zhang, Gaoang Wang, Zhichao Lei, Jenq-Neng Hwang, University of Washington

Drones, or general UAVs, equipped with a single camera have been widely deployed to a broad range of applications, such as aerial photography, fast goods delivery and most importantly, surveillance. Despite the great progress achieved in computer vision algorithms, these algorithms are not usually optimized for dealing with images or video sequences acquired by drones, due to various challenges such as occlusion, fast camera motion and pose variation. In this paper, a drone-based multi-object tracking and 3D localization scheme are proposed based on the deep learning based object detection. The system deployed on the drone can not only detect and track the objects in a scene, but can also localize their 3D coordinates in meters with respect to the drone camera.

"Matching Images and Text with Multi-modal Tensor Fusion and Re-ranking"

Tan Wang, Xing Xu, Yang Yang, Heng Tao Shen, Jingkuan Song, University of Electronic Science and Technology of China; Alan Hanjalic, Delft University of Technology

A major challenge in matching images and text is that they have intrinsically different data distributions and feature representations. The authors propose a novel framework that achieves remarkable matching performance with acceptable model complexity. Specifically, in the training stage, they propose a novel Multi-modal Tensor Fusion Network (MTFN) to explicitly learn an accurate image-text similarity function with rank-based tensor fusion rather than seeking a common embedding space for each image-text instance. Then, during testing, we deploy a generic cross-modal re-ranking (RR) scheme for refinement without requiring additional training procedure.

"Aligning Linguistic Words and Visual Semantic Units for Image Captioning"

Longteng Guo, Jing Liu, Hanqing Lu, Chinese Academy of Sciences; Jinhui Tang, Nanjing University of Science and Technology; Jiangwei Li, Wei Luo, Huawei Devices

Image captioning attempts to generate a sentence composed of several linguistic words, which are used to describe objects, attributes, and interactions in an image, denoted as "visual semantic units" in this paper. Based on this view, the authors propose to explicitly model the object interactions in semantics and geometry based on Graph Convolutional Networks (GCNs), and fully exploit the alignment between linguistic words and visual semantic units for image captioning. Particularly, the authors construct a semantic graph and a geometry graph, where each node corresponds to a visual semantic unit, i.e., an object, an attribute, or a semantic (geometrical) interaction between two objects.

"Deep Fusion Network for Image Completion"

Xin Hong, Chinese Academy of Sciences; Pengfei Xiong, Renhe Ji, Haoqiang Fan, Megvii Technology

Deep image completion usually fails to harmonically blend the restored image into existing content, especially in the boundary area. This paper handles with this problem from a new perspective of creating a smooth transition and proposes a concise Deep Fusion Network (DFNet). The authors qualitatively and quantitatively compare their method with other state-of-the-art methods on two well-

known datasets: Places2 and CelebA. The results show the superior performance of DFNet, especially in the aspects of harmonious texture transition, texture detail and semantic structural consistency.

Workshops

For full descriptions, visit the [MM 2019 Workshops Page](#).

- 1st Workshop on Structuring and Understanding of Multimedia heritAge Contents (SUMAC 2019)
- 2nd International Workshop on Multimedia Content Analysis in Sports (MMSports 2019)
- 2nd Workshop on Multimedia for Accessible Human Computer Interfaces (MAHCI 2019)
- 9th International Audio/Visual Emotion Challenge and Workshop (AVEC 2019)
- Fourth International Workshop on Multimedia for Personal Health and Health Care (HealthMedia 2019)
- 1st International Workshop on Fairness, Accountability, and Transparency in MultiMedia (FAT/MM 2019)
- 5th International Workshop on Multimedia Assisted Dietary Management (MADiMa 2019)
- 1st International Workshop on AI for Smart TV Content Production, Access and Delivery (A14TV 2019)
- 1st International Workshop on Multimodal Understanding and Learning for Embodied Applications (MULEA 2019)
- 1st International Workshop on Search as Learning with Multimedia Information (SALMM 2019)

Multimedia Grand Challenge

[The Multimedia Grand Challenge](#) presents a set of problems and issues from industry leaders and top academic institutions, geared to engage the multimedia research community in solving relevant, interesting and challenging questions for multimedia on a three-to-five-year vision. This year, there are seven different challenges, including: AI Meets Beauty Perfect Corp. Half Million Beauty Product Image Recognition Challenge; BioMedia: Multimedia in Medicine; iQIYI Celebrity Video Identification Challenge; Live Video Streaming; Relation Understanding in Videos; Social Media Prediction; and Content-based video relevance prediction: Hulu.

Brave New Ideas (partial list)

The Brave New Ideas Track showcases innovative papers that open up new vistas for multimedia research in high impact applications. BNI papers stimulate activity towards addressing new, long term challenges of interest to the multimedia research community.

“Neural Storyboard Artist: Visualizing Stories with Coherent Image Sequences”

Shizhe Chen, Qin Jin, Renmin University of China; Bei Liu, Jianlong Fu, Microsoft Research Asia; Ruihua Song, Pingping Lin, Xiaoyu Qi, Microsoft Xiaolce; Chunting Wang, Jin Zhou, Beijing Film Academy

A storyboard is a sequence of images to illustrate a story containing multiple sentences, which has been a key process to create different story products. In this paper, the authors tackle a new multimedia task of automatic storyboard creation to facilitate this process and inspire human artists. Inspired by the fact that our understanding of languages is based on our past experience, the authors propose a novel inspire-and-create framework with a story-to-image retriever that selects relevant cinematic images for inspiration and a storyboard creator that further refines and renders images to improve the relevancy and visual consistency.

“HyperLearn: A Distributed Approach for Representation Learning in Datasets with Many Modalities”

Devanshu Arya, Stevan Rudinac, Marcel Worring, University of Amsterdam

Multimodal datasets contain an enormous amount of relational information, which grows exponentially with the introduction of new modalities. The authors introduce a hypergraph-based model for data representation and deploy Graph Convolutional Networks to fuse relational information within and across modalities. Their approach provides an efficient solution for distributing otherwise extremely computationally expensive or even unfeasible training processes across multiple-GPUs, without any sacrifices in accuracy.

“Moment-to-Moment Detection of Internal Thought during Video Viewing from Eye Vergence Behavior”

Michael Xuelin Huang, Google; Jiajia Li, Guangdong University of Technology; Grace Ngai, Hong Va Leong, The Hong Kong Polytechnic University; Andreas Bulling, University of Stuttgart

Despite the close link between the eyes and the human mind, only few studies have investigated vergence behavior during internal thought and none has studied moment-to-moment detection of internal thought from gaze. While prior studies relied on long-term data analysis and required a large number of gaze characteristics, the authors describe a novel method that is user-independent, computationally light-weight and only requires eye vergence information readily available from binocular eye trackers.

“Learning Subjective Attributes of Images from Auxiliary Sources”

Francesco Gelli, Tat-Seng Chua, National University of Singapore; Tiberio Uricchio, Alberto Del Bimbo, Università degli Studi di Firenze; Xiangnan He, University of Science and Technology of China

Recent years have seen unprecedented research on using artificial intelligence to understand the subjective attributes of images and videos. These attributes are not objective properties of the content but are highly dependent on the perception of the viewers. The authors note that users or organizations often interact with images in a multitude of real-life applications, such as the sharing of photographs by brands on social media or the re-posting of image microblogs by users. They argue that these aggregated interactions can serve as auxiliary information to infer image interpretations. To this end, they propose a probabilistic learning framework capable of transferring such subjective information to the image-level labels based on a known aggregated distribution.

About ACM SIGMM

[SIGMM](#) is ACM’s Special Interest Group on Multimedia—the community of researchers and practitioners dedicated to building next-generation technologies and applications around multimedia. SIGMM hosts several vibrant premiere conferences including ACM Multimedia (with 600+ participants annually), ICMR on multimedia retrieval, and MMSys on multimedia systems. The community also takes pride in its publications including the flagship journal *ACM TOMCCAP* and the affiliated *Springer Multimedia Systems Journal (MMSJ)*.

About ACM

[ACM, the Association for Computing Machinery](#), is the world’s largest educational and scientific computing society, uniting computing educators, researchers and professionals to inspire dialogue, share resources and address the field’s challenges. ACM strengthens the computing profession’s collective voice through strong leadership, promotion of the highest standards, and recognition of technical excellence. ACM supports the professional growth of its members by providing opportunities for life-long learning, career development, and professional networking.

###