Research Statement

David A. Shamma • Centrum Wiskunde & Informatica • `aymans@acm.org`

My research is focused on understanding people and their interactions with video, images and other types of multimedia in various contexts from social communication, information retrieval, to artistic expression; this involves examining how media is created and captured, how media are stored and organized, and how media annotated and reused.

Overview

I research, design, and build systems multimedia systems and develop methods and metrics for understanding multimedia use in-the-world. *I am interested in understanding the social semantics of media which, I assert, is to be discovered as product of the affordances of the tools we use to create, annotate and share.* I seek to conduct my research in three related areas. First, I look to bridge between classical theory and modern practice in multimedia information systems. Today, with the abundance and speed of new social multimedia applications, we should bridge theory to practice to understand the world in an informed way. Second, I design and specify new engineering and computer science practices for multimedia information research. Beyond the contemporary Data Science efforts that deal with the abundant clicks and dwells times, my research aims to find stronger, often not instrumented, signals which can express greater insights into what people do. This is done as an human-centered engineering practice, where one looks to instrumenting and understanding what is know to be effective signals from theory and design. In effect, *my work relies on aspects of human-in-the-loop AI systems.* This practice also speaks to the new world of data; for example, the classic studies in multimedia are done via an analysis of the pixels, but today there is wider band of signals, from visual objects in the scene to the associated with the photos or videos (i.e. GEO, EXIF, and metadata) to the social interactions around the media objects which must be treated as first class signals. Finally, I seek to understand art theories as they apply to social computing, audience, participation, and aesthetics. This is primarily a continuation of my Ph.D. work where I am looking for the connections between art theory, computer science, and the practice of building and engineering; I believe this research is long term and holds ground breaking potential. Further, it lays a strong foundation for multidisciplinary research and education for the next generation of students. These three directions point towards the a new future of trans-disciplinary research: from bridging culture and information theory with practice, human-centered data engineering and information organization, and art theory and expression.

Beyond my research accomplishments, I have designed, built, launched, and conducted research through several prototypes and services in social computing, as well as, engineered from ground up large scale cloud based data science platforms. I built, launched, and studied a synchronized video player that was embedded in a popular instant messaging platform [16]. While an engaging product with millions of monthly sessions, I used the application, via a careful data instrumentation, to create new research into engagement, negotiated meaning [26], and virality [13]; having the right data is the key to organizing the data to find meaning. I have advocated open data sharing and released 100 million Creative Commons photos from a corporation to the broader research community [23] while

building out an internal platform for data science and photo-sharing research within the company. I seek to identify and deeply understand the meaning and motivation behind the big or small data signals and then design methods that can utilize these signals often without the need for millions of data points. This was largely evident in the design of a human-in-the-loop AI system I built to assist editors find high quality weather photos. [15] Further, I also build new experiences for interactive media and artistic performance installations to embody research concepts. Most recently, at Centrum Wiskunde & Informatica (CWI), a Dutch national academic research laboratory, I lead a team that built the sensing for a smart nightclub at the Amsterdam Dance Event in conjunction with a fashion designer [24]; which took my work into the Internet of Things and wireless sensing and real world social interactions.

While traditional content (pixels and audio) may play a role in my work, the meaning and the story of the media we share and engage with is to be found in the broader social online ecosystem. This research perspective is one that has several calls for funding (i.e. [7, p. 43]). Drawn attention to this research agenda in ACM SIG Multimedia (SIGMM) where I was Technical Program Co-Chair in 2013 and currently serve on the steering committee as coordinate the Interactive Arts Program and the Grand Challenges competition. I have published 5 journal articles, 30 conference papers, 5 workshop papers, 3 book chapters, 4 ACM or IEEE magazine articles; 3 papers have received nominations for best paper. I have organized 12 conference workshops and tutorials and, currently, 3 journal articles and 5 conference papers are in submission. Additionally, I have been awarded 6 U.S. Patents with 27 pending review. Inside an industry laboratory, I have been the PI in several academic engagements, as well as, I developed and secured new, long-term funding models inside my former institution aimed at fostering joint HCI and Computer Vision research at Stanford University. I am a Distinguished Member of the ACM and am currently serving on the steering committee for ACM Multimedia and ACM TVx (on Television Experiences). I have attended to several invite-only workshops and leadership events (e.g. NSF, the Internet Archive, and the Wikimedia foundation), as well as, served on the Berkeley iSchool data science advisory board.

Research Highlights

Here I detail research contributions that I have made as well as include forthcoming work. I position this work as three interwoven threads: *(i)* understanding social semantics, *(ii)* metrics and protocols for engagement, *(iii)* creative acts.

Understanding Social Semantics

The current scale of social networking sites is tremendous. With them, the number of photographs and videos in the world has grown rapidly. In 2011, a back of the envelope calculation[1] estimated that 10% of all photographs in the world were taken in the past 12 months. The growing popularity of the cameraphone allows one to capture any number of photographs for a variety of reasons: a friend, a landmark, a receipt, or where you parked. Of these photos, many are edited, shared, commented on, or annotated as our favorites. Instagram[2], a popular photo sharing site, alone sees 60 million photos daily and 1.6 billion likes. The amount of photographic and social data that is created daily creates a rich landscape for much of today's current research in HCI and Multimedia.

While classical computer vision and multimedia research seeks to understand what is in these images and videos, it tends to ignore the social signals around the media. These signals can not

---

[1] http://bit.ly/1vcDWke
[2] http://bit.ly/1zkLrKJ

only aid the classical research approaches, these signals may reveal an entirely new meaning or new semantics that would be invisible to just looking at the visual (pixels) or audio content. Further, I assert this claim of social semantics around media was a product of the interface and ecosystem that encapsulated the media [20].

Perhaps the largest study I undertook looking at social understanding involved studies of video sharing in synchronous environments [16, 9]. To start, I built and deployed a small instant-messenger based tool called Zync which predated Google Hangouts but was a similar deployment as a plugin for a larger instant message platform. With a small team, we examined how video in a shared context could reveal the socially negotiated genre of the video (i.e. comedy, film, sports) as well as give insights into the overall virality of the video [26, 13, 25]. The research, which was a mixture of qualitative and quantitative methods, was not a typical "big data" techniques but rather sought to model the *meaning* of the signals in the data. Ultimately, we were able to predict genre and virality using rather small data traces (under 20 shared sessions per video). I extended this work into studies of engagement in online photo communities [1] and asking what role does content play when formally modeled with community signals.

Overall it is an approach I prefer and embody in my work: understand the data people make and use it not only for prediction but also for explanation [14]. More so, my position is that multimedia is more than the legacy visual signals, multimedia encompasses the plurality of signals that are available in and around media in the world today which includes the rich social signals therein. Towards this end, I have lead various efforts in SIGMM, growing this perspective on social multimedia within the broader community. Most recently, as 2013 Technical Program Co-chair of ACM Multimedia, I brought several new areas into the technical program, including social signals, music, geographic information, and crowd-sourcing to name a few. As General Co-chair of Creativity and Cognition, we are similarly expanding the technical calls.

Metrics and Protocols for Engagement

Part of my expansion of multimedia signals involves defining metrics and protocols for interaction [16, 3, 4] to support the new frameworks and requirements that are introduced. One such example comes from identifying events in multimedia. For example, in photography, Henri Cartier-Bresson, known as the father of photojournalism, said "To me, photography is the simultaneous recognition, in a fraction of a second, of the significance of an event." [2, pp. 1–14] However, finding *when* a photo took place is a non-trival matter, so we developed new techniques for time-stamp disambiguation [22] as 40% of the photos on Flickr have conflicting timestamps between the GPS time and the photo-taken time.

Beyond single source metrics, I conducted the first work to look at modern social streams, like Twitter, during media events on television. Here I looked towards the reification of status updates (tweets) as they related to large scale media events (i.e. elections[3] and award shows) [11, 12]. Here, I identified the structure of tweets as conversational media and described how they relate to live events. While the terms and text of the tweets played a role in identifying what was happening in the event, signals in the Twitter medium (i.e. `#hashtags` and `@mentions`) alone could surface sub-event onsets to identify something is happening. We took this work further to identify topics; going beyond trends, we identified what was part of the event's agenda as well as what was the lasting and conversational topics that surfaced during the event. These temporally salient topics often would be discarded by vector-space analysis as many times they would be judged insignificant by the conventional methods.

---

[3]Google Scholar list this work in the top 10 cited HCI papers over the last 5 years.

Creative Acts

My investigation of online communities negotiating meaning in images stemmed from my thesis work [18]. I was one of the first who positioned network data as an artistic medium instead of capricious fodder [6]. Even simple signals, such as web frequency could be used as computational models of familiarity [17] to be reused by software agents in autonomous performances. While the thrust of my work on understanding online communities is represented in my other work, I still maintain a strong commitment to the arts in my role as Arts Co-editor for SIGMM as well as in my technical papers [8, 21].

In particular, I believe there is much more to academically learn about the act of performance. It is during a performance where meaning is attempting communication. To understand how and why performances conduct this communication, I conducted some qualitative research examining DJs who were live casting their sets [10], as well as, built physical interfaces for data manipulation [19, 5] as well as the aforementioned large-scale sensing event [24]. I believe there is still much work to do understanding performances and the role information plays in them from complex like signals online to conventional vinyl music; part of the rich future of research in computer-mediated communication lies in performance and the creative act.

References

[1] BAKHSHI, S., SHAMMA, D. A., AND GILBERT, E. Faces engage us: Photos with faces attract more likes and comments on instagram. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2014), CHI '14, ACM, pp. 965–974.

[2] CARTIER-BRESSON, H. *The Decisive Moment*. Simon and Schuster, New York, 1952.

[3] CORTEZ, J., SHAMMA, D. A., AND CAI, L. Device communication: A multi-modal communication platform for internet connected televisions. In *EuroITV '12: Proceedings of the 10th international interactive conference on Interactive TV & Video* (July 2012).

[4] DE SÁ, M., SHAMMA, D. A., AND CHURCHILL, E. F. Live mobile collaboration for video production: design, guidelines, and requirements. *Personal and Ubiquitous Computing* (2013), 1–15.

[5] GROTH, P., AND SHAMMA, D. A. Spinning data: Remixing live data like a music DJ. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA, 2013), CHI EA '13, ACM, pp. 3063–3066.

[6] MIRAPAUL, M. Art unfolds in a search for keywords. *The New York Times* (17 June 2004), E5. Vol. CLIII, No. 52883, Circuits Section.

[7] PRESIDENT'S COUNCIL OF ADVISORS ON SCIENCE AND TECHNOLOGY. Designing a digital future: Federally funded research and development in networking and information technology. `http://bit.ly/2eNWe9n`, December 2010.

[8] SHAMMA, D. A. Autonomous expressionism: a framework for installation directed network arts. *International Journal of Arts and Technology 2*, 1 (2009), 62–81.

[9] SHAMMA, D. A., BASTEA-FORTE, M., JOUBERT, N., AND LIU, Y. Enhancing online personal connections through the synchronized sharing of online video. In *CHI'08 extended abstracts on Human factors in computing systems* (2008), ACM, pp. 2931–2936.

[10] SHAMMA, D. A., CHURCHILL, E. F., BOBB, N., AND FUKUDA, M. Spinning online: a case study of internet broadcasting by djs. In *C&T '09: Proceedings of the fourth international conference on Communities and technologies* (New York, NY, USA, 2009), ACM, pp. 175–184.

[11] SHAMMA, D. A., KENNEDY, L., AND CHURCHILL, E. F. Tweet the debates: Understanding community annotation of uncollected sources. In *WSM '09: Proceedings of the international workshop on Workshop on Social Media* (Beijing, China, 2009), ACM.

[12] SHAMMA, D. A., KENNEDY, L., AND CHURCHILL, E. F. Peaks and persistence: modeling the shape of microblog conversations. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work* (New York, NY, USA, 2011), CSCW '11, ACM, pp. 355–358.

[13] SHAMMA, D. A., KENNEDY, L., AND CHURCHILL, E. F. Viral actions: Predicting video view counts using synchronous sharing behaviors. In *ICWSM 11: Proceedings of the International Conference on Weblogs and Social Media Data* (Barcelona, Spain, July 2011), AAAI Press.

[14] SHAMMA, D. A., KENNEDY, L., AND CHURCHILL, E. F. Watching and talking: Media content as social nexus. In *Proceeding of the 2nd ACM international conference on Multimedia Information Retrieval* (Hong Kong, June 2012), ACM.

[15] SHAMMA, D. A., KENNEDY, L., LI, J., THOMEE, B., JIN, H., AND YUAN, J. Finding weather photos: Community-supervised methods for editorial curation of online sources. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (New York, NY, USA, 2016), CSCW '16, ACM, pp. 86–96.

[16] SHAMMA, D. A., AND LIU, Y. *Social Interactive Television: Immersive Shared Experiences and Perspectives*. Information Science Publishing, Hershey, PA, USA, 2009, ch. Zync with Me: Synchronized Sharing of Video through Instant Messaging, pp. 273–288.

[17] SHAMMA, D. A., OWSLEY, S., BRADSHAW, S., AND HAMMOND, K. J. Using web frequency within multi-media exhibitions. In *Proceedings of the 12th International Conference on Multi-Media* (2004), ACM Press.

[18] SHAMMA, D. A., OWSLEY, S., HAMMOND, K. J., BRADSHAW, S., AND BUDZIK, J. Network Arts: Exposing cultural reality. In *Alternate track papers & posters of the 13th international conference on World Wide Web* (2004), ACM Press, pp. 41–47.

[19] SHAMMA, D. A., SCHEIBLE, J., AND SHEPPARD, R. M. Graffiti dance: interaction of light, information, and environment. In *C&C '09: Proceeding of the seventh ACM conference on Creativity and cognition* (New York, NY, USA, 2009), ACM, pp. 479–480.

[20] SHAMMA, D. A., SHAW, R., SHAFTON, P. L., AND LIU, Y. Watch what i watch: using community activity to understand content. In *MIR '07: Proceedings of the international workshop on Workshop on multimedia information retrieval* (New York, NY, USA, 2007), ACM, pp. 275–284.

[21] SHAMMA, D. A., SHEPPARD, R. M., AND SCHIBLE, J. Human-to-dancer interaction: Designing for embodied performances in a participatory installation. In *DIS '10: Proceedings of the 8th ACM conference on Designing interactive systems* (Århus, Denmark, 2010), ACM, pp. 356–359.

[22] THOMEE, B., MORENO, J. G., AND SHAMMA, D. A. Who's time is it anyway?: Investigating the accuracy of camera timestamps. In *Proceedings of the ACM International Conference on Multimedia* (New York, NY, USA, 2014), MM '14, ACM, pp. 909–912.

[23] THOMEE, B., SHAMMA, D. A., FRIEDLAND, G., ELIZALDE, B., NI, K., POLAND, D., BORTH, D., AND LI, L.-J. Yfcc100m: The new data in multimedia research. *Commun. ACM 59*, 2 (Jan. 2016), 64–73.

[24] WONG, K. Byborre & Red Bull explore the future of clubbing. Hypebeat: `http://bit.ly/2eNLmbh`, October 2016.

[25] YEW, J., AND SHAMMA, D. A. Know your data: Understanding implicit usage versus explicit action in video content classification. In *IS&T/SPIE Electronic Imaging* (January 2011).

[26] YEW, J., SHAMMA, D. A., AND CHURCHILL, E. F. Knowing funny: Genre perception and categorization in social video sharing. In *Proceedings of the 2011 annual conference on Human factors in computing systems* (New York, NY, USA, 2011), CHI '11, ACM, pp. 297–306.