Riding the Multimedia Big Data Wave

John R. Smith
IBM T. J. Watson Research Center
Kitchawan Road
Yorktown Heights, NY 10532 USA
jsmith@us.ibm.com

ABSTRACT

Across multiple generations of information technology that have dealt with structured and unstructured data, the explosion of multimedia data is creating the biggest wave of all. Huge volumes of multimedia – images, video and audio are being generated and consumed daily. Currently, multimedia makes up 60% of internet traffic, 70% of mobile phone traffic and 70% of all available unstructured data. To give specific examples, Web users are uploading 72 video-hours to YouTube per minute, and on an average day, social media users post 300 hundred million photos to Facebook. Consumers using mobile phones and digital cameras are taking 500 billion photos per year, or 78 per person on the planet [1]. Specialized domains are participating too. Medical institutions are acquiring one billion radiological images per year, and cities are installing hundreds of millions of video cameras worldwide for safety, security and law enforcement. Industries across life sciences, petroleum exploration, astronomy, insurance, retail and many others are faced with huge and growing volumes of multimedia data.

Multimedia clearly is "big data". But, it it big data not just because there is a lot of it. Multimedia is big data because increasingly it is becoming a valuable source for insights and information. Multimedia data can tell us about things happening in the world, point out places, events or topics of interest (memes), give clues about a person's preferences and even capture a rolling log of human history [1, 2]. However, the challenge with multimedia big data is that images, video and audio require much more sophisticated algorithms for content analysis than previous waves of structured and unstructured data. This is spurring on a tremendous amount of research on efficient and effective techniques for "bridging the semantic gap" to enable large-scale multimedia information extraction and retrieval [3, 4].

In this talk we present a perspective across multiple industry problems, including safety and security, medical, Web, social and mobile media, and motivate the need for large-scale analysis and retrieval of multimedia data. We describe a multi-layer architecture that incorporates capabilities for audio-visual feature extraction, machine learning and semantic modeling and provides a powerful framework for learning and classifying contents of multimedia data. We discuss the role semantic ontologies for representing audio-visual concepts and relationships, which are essential for training semantic classifiers [5]. We discuss the importance of using faceted classification schemes in particular for organizing multimedia semantic concepts in order to achieve effective learning and retrieval [6]. We also show how training and scoring of multimedia semantics can be implemented on big data distributed computing platforms to address both massive-scale analysis and low-latency processing [7]. We describe multiple efforts at IBM on image and video analysis and retrieval, including IBM Multimedia Analysis and Retrieval System (IMARS), and show recent results for semantic-based classification and retrieval. We conclude with future directions for improving analysis of multimedia through interactive and curriculum-based techniques for multimedia semantics-based learning and retrieval.

SIGIR'13, July 28–August 1, 2013, Dublin, Ireland.
ACM 978-1-4503-2034-4/13/07.

Categories and Subject Descriptors

H.2.4 [**Systems**]: Multimedia databases; I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding—*Video analysis*; I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods—*Semantic networks*; I.2.6 [**Artificial Intelligence**]: Learning—*Concept learning*

Keywords

Multimedia information retrieval, video analysis, content-based search, machine learning, semantic modeling

Biography

Dr. John R. Smith is Senior Manager, Intelligent Information Management Dept, IBM T. J. Watson Research Center. He received his Ph.D., Electrical Engineering from Columbia University in 1997. He currently leads IBM's research in multimedia information retrieval including image and video content extraction, multimedia content-based search, video event detection and retrieval and social media analysis. Dr. Smith is currently principal investigator for IBM Multimedia Analysis and Retrieval

System (IMARS), which has been recognized by multiple awards including a Wall St. Journal innovation award. Dr. Smith is a long-time participant in the NIST TRECVID video retrieval evaluation and co-led the development of the Large Scale Concept Ontology for Multimedia (LSCOM), which has been incorporated into the TRECVID evaluation. From 2000-2004, Dr. Smith served as Chair, ISO/IEC MPEG Multimedia Description Schemes Group and led the development of multiple parts of the MPEG-7 Multimedia Metadata Standard and MPEG-21 Digital Framework Standard. While a student with Prof. Shih-Fu Chang at Columbia University, Dr. Smith conducted some of the earliest work on content-based image retrieval (VisualSEEk) and Web image/video search (WebSEEk), which has been highly influential for researchers and practitioners. Dr. Smith has published more than two hundred papers (>14K citations, h-index of 55, i-index of 166). Dr. Smith is currently a member of ACM SIGMM, Fellow of IEEE and Editor-in-Chief of *IEEE MultiMedia*.

1. REFERENCES

[1] John R. Smith. History made everyday. *IEEE MultiMedia*, 18(2), July-Sept 2011.

[2] L. Xie, A. Natsev, J. R. Kender, M. Hill, and J. R. Smith. Visual memes in social media: Tracking real-world news in youtube videos. In *Proc. of the 19th ACM Intl. Conf. on Multimedia*, pages 53–62. ACM, November 2011.

[3] John R. Smith. Minding the gap. *IEEE MultiMedia*, 19(2), January-March 2012.

[4] A. Hanjalic, R. Lienhart, W.-Y. Ma, and J. R. Smith. The holy grail of multimedia information retrieval: So close or yet so far away? *Proc. IEEE*, 96(4):541–547, 2008.

[5] M. Naphade, J. R. Smith, J. Tesic, S.-F. Chang, W. Hsu, L. Kenendy, A. Hauptmann, and J. Curtis. Large-scale concept ontology for multimedia. *IEEE MultiMedia*, 13(3), July-Sept 2006.

[6] John R. Smith. Just the facets. *IEEE MultiMedia*, 20(1), January-March 2013.

[7] R. Yan, M. O. Fleury, M. Merler, A. Natsev, and J. R. Smith. Large-scale multimedia semantic concept modeling using robust subspace bagging and map-reduce. In *Proc. of the 1st ACM Workshop on Large-Scale Multimedia Retrieval*. ACM, October 2009.