

Multimedia Big Data Computing

Wenwu Zhu, Peng Cui, and Zhi Wang
Tsinghua University, China

Gang Hua
Steven Institute of Technology

With the proliferation of the Internet and user-generated content, and the growing prevalence of cameras, mobile phones, and social media, huge amounts of multimedia data are being produced, forming a unique kind of big data. Multimedia big data brings tremendous opportunities for multimedia applications and services—such as multimedia searches, recommendations, advertisements, healthcare services, and smart cities. The need to compute such massive datasets is transforming how we deal with multimedia computing.

Researchers have studied some of the problems in big data computing (see the related sidebar), but multimedia big data has its own characteristics related to multimodality, real-time information, quality of experience, and so on. For example, some multimedia learning applications, games, or 3D rendering might require GPU processing. Consequently, methods for general big data might not directly apply to multimedia big data.

Compared to approaches of general text-based big data computing, multimedia big data computing faces additional compression, storage, transmission, and analysis challenges in terms of

- organizing unstructured and heterogeneous data,
- dealing with cognition and understanding complexity,
- addressing real-time and quality-of-service requirements, and
- ensuring scalability and computing efficiency.

Here, we consider these technical challenges and the related scientific problems for multimedia big data computing, introducing various research directions and emerging technologies.

The Multimedia Life Cycle

The emergence of big data computing is having a profound effect on the entire life cycle of

multimedia content. Figure 1a shows the typical multimedia life cycle, which comprises acquisition, storage, processing, dissemination, and presentation.

In recent decades, the availability of low-cost commodity digital cameras and camcorders has sparked an explosion of user-generated media content. Most recently, cyber-physical systems have started offering a new type of data acquisition through sensor networks, significantly increasing the volume and diversity of media data.¹ Riding the Web 2.0 wave and social networks, digital media content can now be easily shared through the Internet, including via social networks. The huge success of YouTube demonstrates the popularity of “Internet” multimedia; similarly, social multimedia has had great success thanks to social networks such as Facebook and Twitter.

In the early stage, media storage, processing, and dissemination were relatively small in scale—usually at the level of kilobytes. Now, the data scale is often at the terabyte or even petabyte level. The collected datasets are so large and complex that it becomes difficult to process using traditional media data processing technology.

However, multimedia big data provides great opportunities. Both the scale and richness of the data—in terms of content, context, users and crowds, and so on—provide more opportunities to build better computational models to mine, learn, and analyze enormous amounts of data. Moreover, multimedia big data algorithms require “massively parallel software running on thousands of servers distributively.”²

A typical multimedia big data computing life cycle consists of moving from data to information, from information to knowledge, from knowledge to intelligence, and from intelligence to decision, as depicted in Figure 1b. First, we need to process the collected multimedia raw data into information, creating multimedia knowledge and insight. When we combine this output with human or user knowledge, it can

Related Work in Big Data Computing

Various researchers have studied the challenges of big data computing. Xindong Wu and his colleagues presented a Heterogeneous, Autonomous, Complex, Evolving (HACE) theorem that characterizes the features of the big data revolution and proposes a big data processing model from the data mining perspective.¹ Philip Russom and his colleagues explained the concept, characteristics, and needs of big data and different offerings available in the market to explore unstructured large data.² Han Hu and his colleagues presented a literature survey and system tutorial for big data analytics platforms, aiming to provide an overall picture for nonexpert readers and instill a do-it-yourself spirit for advanced audiences to customize their own big data solutions.³ Puneet Singh Duggal and Sanchita Paul suggested various methods for catering to the problems at hand through a Map Reduce framework over the Hadoop Distributed File System.⁴ Stephen Kaisler and his colleagues have also analyzed the issues and challenges for big data analysis and design.⁵ Changqing Ji and his colleagues intro-

duced several big data processing technics from both system and application aspects.⁶

References

1. X. Wu et al., "Data Mining with Big Data," *IEEE Trans. Knowledge and Data Eng.*, vol. 26, no. 1, 2014, pp. 97–107.
2. P. Russom et al., "Big Data Analytics," *TDWI Best Practices Report, Fourth Quarter*, 2011.
3. H. Hu et al., "Toward Scalable Systems for Big Data Analytics: A Technology Tutorial," *IEEE Access*, vol. 2, 2014, pp. 652–687.
4. P.S. Duggal and S. Paul, "Big Data Analysis: Challenges and Solutions," *Proc. Int'l Conf. Cloud, Big Data and Trust 2013*, 2013.
5. S. Kaisler et al., "Big Data: Issues and Challenges Moving Forward," *Proc. 46th Hawaii Int'l Conf. System Sciences (HICSS)*, 2013, pp. 995–1004.
6. C. Ji et al., "Big Data Processing in Cloud Computing Environments," *Proc. Int'l Symp. Parallel Architectures, Algorithms and Networks (I-SPAN)*, 2012, pp. 17–23.

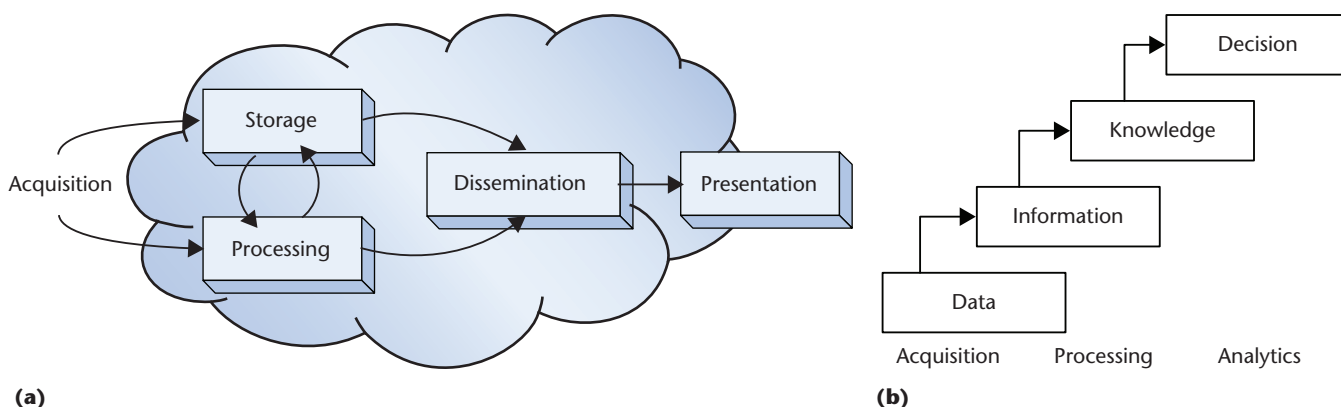


Figure 1. The emergence of big data computing is affecting the life cycle of multimedia content: (a) the typical multimedia life cycle and (b) the typical multimedia big data computing life cycle.

be used to make decisions. However, with this big data scale comes tremendous challenges.

Challenges

Compared to approaches of general text-based big data computing, multimedia big data computing faces the following fundamental technical challenges related to processing, storage, transmission, and analysis.

Unstructured and heterogeneous data. Multimedia big data is unstructured, heterogeneous, and multimodal, which makes multimedia big data representation and modeling difficult. For example, how do we turn unstructured

multimedia data into structured data? How can we represent or model multimedia big data coming from different sources or spaces (cyber, physical, and social)?

Cognition and understanding complexity.

Computers can't easily understand multimedia big data, mainly due to the semantic gap between low-level features and high-level semantics. Moreover, some multimedia big data is evolving with time and space.

Real-time and Quality of Experience (QoE) requirements. Multimedia big data applications and services are typically real time, so to

address QoE requirements, we need real-time streamed/online, parallel/distributed processing for analysis, mining, and learning.

Scalability and efficiency. Multimedia big data systems need large-scale computation, so they must optimize computation, storage, and networking/communication resources. Such systems also need online/streamed and parallel/distributed algorithms. In addition, GPU computing for multimedia big data computing brings further challenges.

Scientific Problems

The four fundamental challenges just discussed lead to four corresponding scientific problems.

Representation and modeling. How can we establish the representation and modeling for unstructured, heterogeneous, and multimodal multimedia big data?

Deep and crowd computing. How do we perform data-driven deep computing (including mining and learning) to effectively analyze data? How can we exploit crowdsourcing jointly with data-driven analysis for multimedia cognition?

Streamed or online computing. How do we perform streamed/online processing for the entire multimedia big data in a parallel/distributed way, so as to make multimedia processing, analysis, mining, and learning real-time while satisfying QoE requirements?

Computing, storage, and communication optimization. How do we design a new multimedia computing architecture to optimize computation, storage, and network/communication for multimedia big data computing? How can we efficiently use GPU-powered servers for multimedia big data computing?

Addressing the Issues

Addressing these fundamental challenges and scientific problems for multimedia big data computing will require implementing effective approaches throughout the multimedia life cycle. Next, we look at each stage of the multimedia big data computing life cycle to identify ways of addressing these issues.

Data Acquisition

In addition to acquiring multimedia data from the Internet and Internet of Things (IoT) (for example, in the form of user-generated content

and camera data), the emergence of online social networks makes it possible to collect multimedia data from individuals acting as sensors of the real world.

Raian Ali and his colleagues have proposed *social sensing*, in which users act as monitors and provide information needed at runtime.³ Even before the conceptualization of social sensing, Anmol Madan and his colleagues had already tried using collected information from users to detect epidemiological behavior change—for example, to predict a person's health status using data collected from his or her cellphone.⁴ On the other hand, the rapid development of wireless networks and mobile devices makes most collected multimedia data personal. Videos, images, and other kinds of multimedia data are increasingly correlated with individuals rather than with public groups.

Today, personal data can be collected by wearable sensors that can feed 24/7 data streams, acting as an important type of multimedia data source. This huge amount of multimedia data raises the following questions regarding data reduction (or compression) and data representation.

Data Reduction/Compression

The volume of multimedia big data must be reduced for efficient storage and communication. Multimedia data reduction refers to sampling the (massive) dataset so that it can be computed with limited computing resources. Multimedia data compression refers to reducing the raw data size for storage or communication.

Feature-transformation-based data reduction. The goal is to reduce numerical data using common signal processing or transform techniques. Compressive sensing⁵ and Wavelet Transform are two approaches for big data reduction.⁶

Analysis-aware compression. Multimedia coding can be performed prior to analysis such that the compression-aware analysis can exploit the coding. For example, you might achieve compression using a feature descriptor.⁷ During compression, you compress feature descriptions along with the images or videos. The compression, storage, and transmission of local feature descriptors of image and video applications can then later be used for computing, such as for a content-based search (or visual search).⁷

Cloud-based compression. Data compression conducted on the (sensor) client side can

effectively save storage space by compressing the data after data generation but before storage. On the cloud side, cloud-based big data compression can take advantage of data correlation and background similarity.

Data Representation

Multimedia data representation refers to a mathematical structure in which you can model the data for later analysis. Because multimedia data comes from multiple sources, it tends to have disparate representations for each source, or sometimes a common representation is needed for multimodal analysis. These interpretations are usually described by both structural and descriptive metadata. Example data representation is referred to as feature-based data representation (such as scale-invariant feature transform).

Multimedia big data representation consists of the following approaches.

Feature-based data representation. Because multimedia big data is often multimodal, sometimes we can find a common feature space to represent the data—namely, feature-based representation. Feature-selection-based data representation aims to find the best representational data among all possible feature combinations.

Learning-based representation. Today's data comes from heterogeneous sources, such as cyber-physical-social spaces, so a common explicit feature space cannot be easily found. A new approach is to find implicit “hidden space” data representation for multimodal and heterogeneous data. Such representation is also called a learning-based (rather than hand-crafted) representation.

Many machine-learning methods have been proposed to represent multimedia data. For example, deep learning represents one breakthrough in representation learning. There are a series of deep learning methods to perform multimedia data representation, such as Deep Boltzmann Machines⁸ and Deep Autoencoder.⁹ Such methods aim to learn high-level representations from low-level features using a set of nonlinear transformations and have achieved the state-of-the-art for various tasks.

Computation-oriented representation. A key factor in multimedia data representation is to reduce computation, which requires understanding the relationship between the representation

and its computational complexity. For example, in a content-based image search, hashing-based indexing methods map images represented by high-dimensional raw features into a Hamming space, where the images are represented by short binary hashing codes. By retrieving samples whose hash codes are within a small Hamming distance of the hash codes of queries, these methods can implement an efficient search. In addition, the compact hash codes can dramatically reduce the required storage space.

Data Processing and Analysis

After acquisition and storage, the next step in the multimedia life cycle is multimedia big data processing and analysis.

Multimodal Analysis

Multimedia data analysis has largely focused on how to fuse the information from the different media modalities together to form a coherent decision. However, issues arise when data entries lack data modality.

For example, many images shared on the Internet now come with geotags. Furthermore, beyond traditional RGB information, many images have additional information such as depth (taken, for example, from advanced sensors such as Kinect). However, there is a large backlog of images that do not possess additional sensing information, which present a problem for algorithms designed to operate using both the RGB and the geotags and/or depth information.

This calls for media analysis technologies that are robust despite missing data modalities, leading to the notion of cross-modality media analysis. From a machine-learning viewpoint, the problem can be formulated as follows: in the training phase, we have full access to data samples with different modalities, but in the testing phase, we might encounter data samples that are either systematically or randomly missing certain views of information. Qilin Zhang and his colleagues recently studied this problem and proposed a common latent space approach.¹⁰ As shown in Figure 2, the basic idea of this type of approach is that in the training phase, a common latent space is identified for all data modalities, and all data samples from the different modalities are projected into the same space for reasoning. Then, in the testing phase, even if some data modalities are missing, the remaining ones can still be projected into the common space for reasoning.

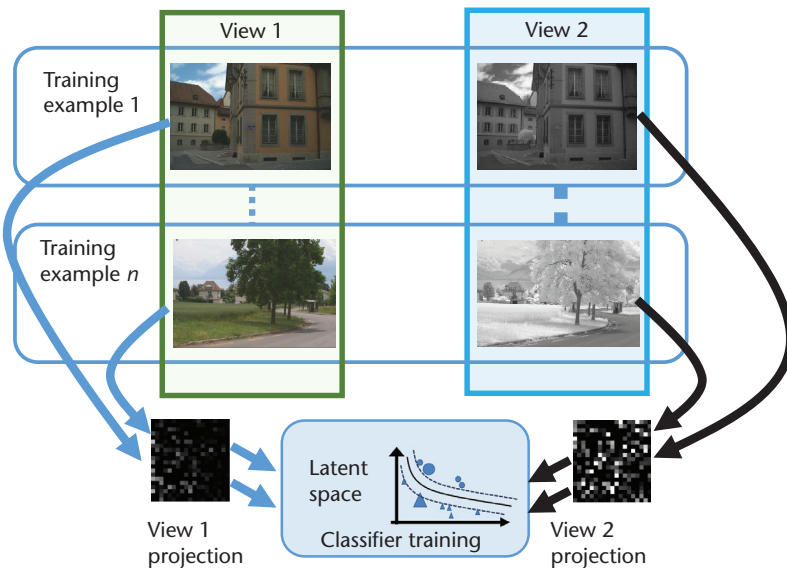


Figure 2. The common space model for cross-modal media analysis.

How to identify such common space is an open question. It could be conducted, for example, via canonical correlational analysis and its variants, or using more complicated methods, such as a deep neural network. When it comes to the problem of making decisions using multimedia big data, a fundamental issue is fusing the data from the different modalities. Previous methods largely focused on either pre-fusion at the feature level or late-fusion at the decision level.

Both fusion methods have their pros and cons. Feature-level fusion lets the decision algorithm potentially benefit from the correlational information across the two different feature modalities. However, finding a way to appropriately normalize the features from the different modalities is an open issue. Decision-level fusion often learns a weighted linear combination of the decision scores from each feature modality. It does not need to deal with the feature normalization issue. However, decision-level fusion might not be able to effectively leverage the correlational information across the different modalities.

We advocate a mid-level fusion scheme at the representation level, where we learn a compact intermediate-level representation to effectively capture the correlational information across the different modalities. Intuitively, such intermediate-level representation can effectively capture the correlational information across the different media modalities while suppressing the heterogeneity. For example,

Zhenxing Niu and his colleagues proposed a visual topic network, which effectively learns an intermediate-level image representation from both visual features and sparse text tags, as illustrated in Figure 3.¹¹ This approach has shown consistent performance gain in terms of recognition accuracy, when compared with pre-fusion and late-fusion.

User-Centric Analysis

Although the rapid advances of multimedia computing technology have greatly facilitated users' information needs regarding multimedia data, it is still difficult for all multimedia applications (such as multimedia search and recommendation applications) to provide satisfactory results for users with different intentions. This is because there's a lack of understanding of users, which is more serious in multimedia applications than in nonmultimedia applications for two reasons.

First, multimedia applications often become more exploratory, and users are often interested in images or videos with a particular style, which is difficult to express and represent. Second, multimedia data is often used for entertainment when exploring a visual space, with no clear end goal.¹ How to discover users' latent intent from limited observed data is of paramount importance in improving multimedia search and recommendation performance. It resonates well with the idea that underpins *user-centric multimedia analysis*, where the user profiles, behaviors, and social networks are sensed, harnessed, and shared to adapt the results of general multimedia search and recommendation engines to be more consistent with user intent.

User intention modeling. User modeling is crucial to addressing the intention gap problem in multimedia search and recommendation. Feng Qiu and Junghoo Cho represented user interest by topics, and they proposed a method to learn user preferences from past query-click history in Google.¹² Eugene Agichtein and his colleagues proposed a method to learn the user interaction model to better predict the user's preference in terms of search results.¹³ Jaime Teevan and her colleagues explored rich models for interest modeling by combining multiple resources, such as search-related information, user-relevant documents, and emails.¹⁴ More recently, researchers have investigated user-

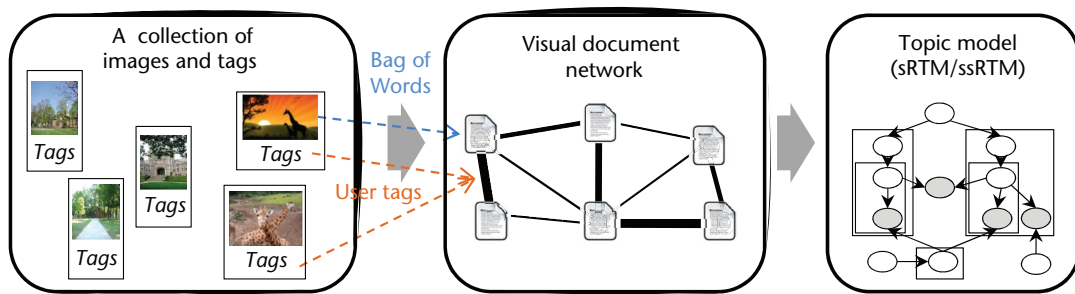


Figure 3. The visual topic network for representation-level fusion, where the sparse text tags link the different visual documents and contribute to the learning of the final representation of the images. (Source: Zhenxing Niu and his colleagues; used with permission.¹¹)

information interaction behavior patterns in social network environments.^{15,16}

The interest-modeling problem is more challenging in the image domain due to the high-dimensional space and the semantic-gap problem. Marek Lipczak, Michele Trevisiol, and Alejandro Jaimes analyzed users' favorite behavior patterns in Flickr.¹⁷ Xing Xie and his colleagues proposed detecting user interests from user-image interaction behaviors recorded by image browsing logs.¹⁸ Yun Yang and her colleagues investigated the emotion prediction problem for individual users when watching social images.¹⁹ Tags of images are mined to construct the topics and ontology to represent user preferences.¹⁵ Similar to the problem that user intentions cannot be well represented by query words in an image search,²⁰ user interests in images cannot be well represented by tags. Visual factors, such as visual style and quality, eventually play important roles in user interest formation.

Social-sensed multimedia search. Personalized search has been studied for many years in the text domain. The main target has been to construct accurate and complete user profiles for re-ranking the search results by measuring the distance between the search results and user profiles.²¹ More specifically, the user profiles were represented by an ontology²² and topics,¹² which are mined from the metadata, search logs, and social media. Recently, some of these techniques have been transferred into a personalized image or video search, especially for image searches in Flickr.

Considering the special characteristics of social images, Dongyuan Lu and Qiudan Li proposed a co-clustering method to discover the latent interests of users and map the Flickr search results into the latent space to measure

their matching degree. Kristina Lerman, Anon Plangprasopchok, and Chio Wong exploited user-generated metadata in the form of contacts and image annotations in Flickr to describe user interest, using them to re-rank the image search results in Flickr.²³

Merging search engine and social media has clearly become a common trend in industry. For example, Google acquired YouTube and launched Google Plus; Yahoo acquired Flickr; and Facebook put forth efforts to develop search services with a Facebook-external scope. Much could be leveraged by integrating social media platforms with multimedia search systems, as shown in Figure 4. How to discover and represent user search intention from social media and seamlessly bridge these user intentions with multimedia search systems is a research issue in need of serious attention.

Social multimedia recommendation. Content-based filtering and collaborative filtering (CF) have been widely used to help users discover the most valuable information to them. Content-based filtering introduces the basic idea of studying an item's content to address the ranking problem. With the emergence of topic modeling techniques such as Latent Dirichlet allocation (LDA), recent content-based approaches²⁴ rank candidate items by how well they match the topical interest of the user. These methods represent users and items in fine granularity.

CF methods, consisting of memory- and model-based methods, are widely used. The memory-based approaches calculate the similarity between all users based on their ratings of items.²⁵ They represent users (or items) by the item sets (or user sets), which are often unstable and can only obtain good performance for

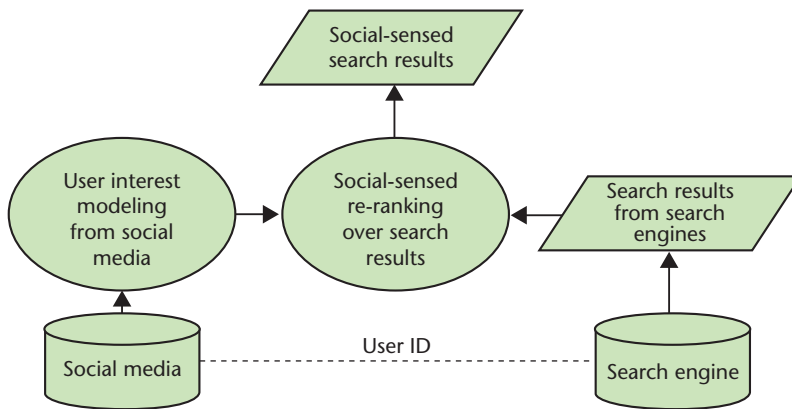


Figure 4. An illustration of the social-sensed image search.

active users or popular items. The model-based methods learn a model based on patterns recognized in the user ratings.

Several matrix factorization methods have recently been proposed.²⁶ The matrix approximation models all focus on representing the user-item rating matrix with low-dimensional latent vectors. Recognizing that influence is a subtle force that governs the dynamics of social networks, influence-based recommendation²⁷ involves interpersonal influence in social recommendation cases. Trust-based approaches²⁸ exploit the trust network among users and make recommendations based on the ratings of users who are directly or indirectly trusted.

Jiang and his colleagues proposed a probabilistic factor analysis framework, which fuses users' preference and social influence together.¹⁴ Furthermore, they have also investigated the social recommendation problem in a multiple domain setting. Most of these works are based on traditional content-based filtering or CF-based methods, and their common goal is to embed social information into traditional methods to improve the recommendation accuracy. However, few authors have targeted the problem of how to learn a new common representation for users and items in social networks, which is indeed feasible and important for boosting social recommendation performance.

Human-in-the-Loop Analysis

Multimedia big data analysis is difficult, and many algorithms and systems, if running in a purely automated fashion, would not be able to achieve the level of performance required for practical use. This has motivated researchers to explore a hybrid human-computer method for content analysis tasks. The Visipedia project²⁹

is one pioneering project in this area, where interactive human inputs and advanced computational algorithms are tightly integrated to solve the content analysis problem.

This type of hybrid human-computer system is further catalyzed by crowdsourcing, where cheap online human labors can be exploited with a small fee on a per-input basis. Figure 5 provides an illustration of such hybrid human-computer systems. There are several issues that need to be carefully modeled when considering crowdsourcing-based human inputs. First of all, human input from crowdsourcing could be very noisy, so it should be carefully modeled. Often, inputs from several online workers are solicited to enable consensus-based analysis or to model the quality of the workers.

The second question is how and when human input should be engaged. It cannot be too frequent, because the cost would be high and the response time long. In this regard, uncertainty- and confidence-based reasoning would be critical, because it naturally serves as the measure for soliciting human input as needed.

The third issue to be addressed is to close the loop, where the human inputs should feed back to the computational methods to improve them. From a learning viewpoint, online learning and, more broadly, the notion of a life-long learning system applies here. Some studies have researched these three issues in the context of collaborative active learning in crowdsourcing for visual content analysis.^{30,31}

Distribution and Systems

With big data, multimedia distribution and delivery can exploit the “intelligence” of content and users, so new systems technologies are appearing for multimedia systems.

Data-Driven Edge-Network Multimedia Distribution

Recent studies have found that edge-cloud resources can improve the performance of social media delivery compared to traditional content delivery.³² Meanwhile, as multimedia content is generated by social network users and personal devices, data is also stored and processed at the edge of the network. Distribution over the edge network can naturally meet the demand of processing the data at different geo-distributed edge datacenters.

In recent years, online social network has greatly changed content delivery—that is, the

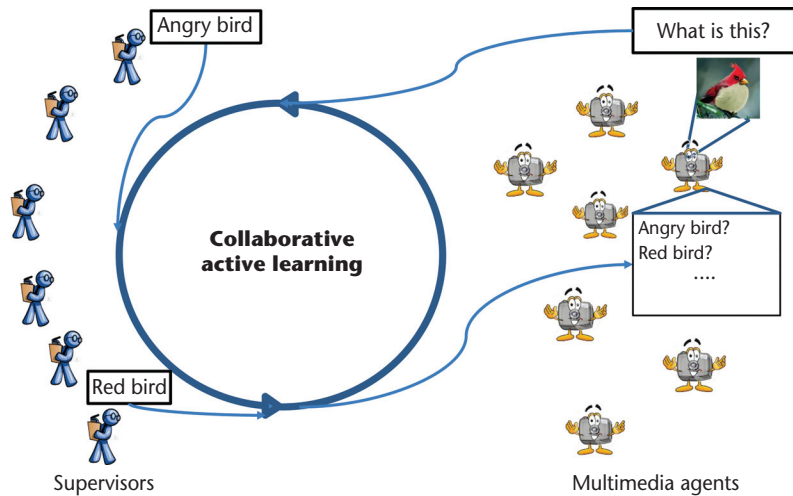


Figure 5. Illustration of a human in the loop multimedia big data analysis system exploiting crowds.

distribution of social contents has shifted from a “central-edge” manner to an “edge-edge” manner. Eytan Bakshy and his colleagues studied the social influence of people in the online social network, observing that some users can be very influential in social propagation.³³ Haitao Li, Haiyang Wang, and Jiangchuan Liu studied the content sharing in an online social network and observed the skewed popularity distribution of content and the *power-law* activity of users.³⁴ Giovanni Comarola and his colleagues investigated response time of social contents using collected traces and observed factors that affect the response time in social propagation.

As online social networks are affecting *dissemination* for all types of online contents, conventional content delivery paradigms need improvement using social information. Josep Pujol and his colleagues designed a social partition and replication middleware where users’ friends’ data can be co-located in datacenter servers.³⁵ Salvatore Scellato and his colleagues investigated using social *cascading* information for content delivery over the edge networks.³⁶ The possibility of inferring social propagation according to users’ social profiles and behaviors has also been investigated,³² allocating network resources at edge-cloud servers based on propagation predictions (see Figure 6).

Using a data-driven approach, a new trend is to exploit machine-learning techniques to learn user and content intelligence³⁷—for example, in the form of mining the user’s quality of experience, which in turn can make multimedia distribution more efficient.

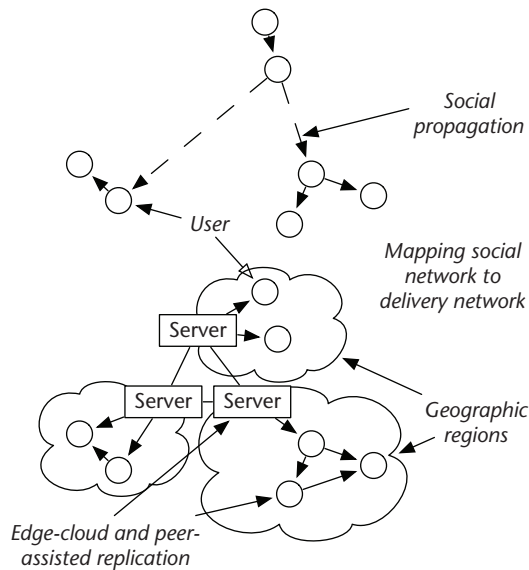


Figure 6. Edge-network social multimedia content distribution.

Streamed and In-Memory Multimedia Processing

Parallel data processing has been the mainstream of designing efficient data-processing platforms so that data could be processed in a distributed and parallel manner, improving the throughput of data processing. MapReduce (<http://mapreduce.sandia.gov>) is the most representative paradigm. Today’s multimedia data is streamed in nature—that is, the data is generated and updated over time. An effective multimedia data-processing system must process the data in a streamed manner. Storm (<https://storm.apache.org>) is one such system. It creates

In-memory multimedia data processing is critical for real-time multimedia applications.

a topology to form several “pipes” for data to pass through for processing along the way.

Another trend in multimedia big data is “in-memory” processing—that is, the data is processed in the memory instead of on hard disks, significantly reducing the processing latency. In general data processing, several in-memory paradigms have been invented, such as Berkeley’s Spark (<https://spark.apache.org>). These types of systems often implement a “cache” strategy, which can move data from hard disks to memory for repeated actions, to reduce the cost of accessing hard drivers. In the context of multimedia data processing, in-memory image and video processing is in demand.

The approaches for multimedia big data computing have a broad range of application scenarios, such as healthcare and medical applications, social media, satellite imaging, IoT, and smart cities. Future research will focus on the scale and complexity problems encountered in multimedia big data computing. For analytics, we need effective and efficient algorithms that address issues of scale and complexity. Taking a data-driven approach, such as with deep learning, will still be effective for multimedia big data analytics. However, with new developments in artificial intelligence and human-computer interaction, we see potential in systematically integrating knowledge- and data-driven approaches. For example, jointly considering deep learning and wisdom of the crowd (crowdsourcing) will be a promising future direction. In terms of systems, determining how to jointly optimize computing, storage, and communication/networking will require further research. **MM**

Acknowledgments

We thank Daixin Wang and Shaowei Liu from Tsinghua University for their contributions on representation learning and suggestions.

References

1. W. Wolf, “Cyber-Physical Systems,” *Computer*, vol. 42, no. 3, 2009, pp. 88–89; doi: 10.1109/MC.2009.81.
2. A. Jacobs, “The Pathologies of Big Data,” *Comm. ACM*, vol. 52, no. 8, 2009, pp. 36–44.
3. R. Ali et al., “Social Sensing: When Users Become Monitors,” *Proc. 19th ACM SIGSOFT Symp. 13th European Conf. Foundations of Software Eng.*, 2011, pp. 476–479.
4. A. Madan et al., “Social Sensing for Epidemiological Behavior Change,” *Proc. 12th ACM Int’l Conf. Ubiquitous Computing*, 2010, pp. 291–300.
5. H.T. Kung, “Big Data and Compressive Sensing,” *Contemporary Computing*, Springer, 2012, p. 6.
6. M.K. Jeong et al., “Wavelet-Based Data Reduction Techniques for Process Fault Detection,” *Technometrics*, vol. 48, no. 1, 2006, pp. 26–40.
7. L.-Y. Duan et al., “Compact Descriptors for Visual Search,” *IEEE MultiMedia*, vol. 21, no. 3, 2014, pp. 30–40.
8. N. Srivastava and R. Salakhutdinov, “Multimodal Learning with Deep Boltzmann Machines,” *Proc. Neural Information Processing Systems Conf. (NIPS)*, 2012, pp. 2231–2239.
9. F. Feng, X. Wang, and R. Li, “Cross-Modal Retrieval with Correspondence Autoencoder,” *ACM Int’l Conf. Multimedia*, 2014, pp. 7–16.
10. Q. Zhang et al., “Can Visual Recognition Benefit from Auxiliary Information in Training?” *Computer Vision—ACCV 2014, Lecture Notes in Computer Science*, vol. 9003, 2015, pp. 65–80.
11. Z. Niu et al., “Semi-Supervised Relational Topic Model for Weakly Annotated Image Recognition in Social Media,” *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 4233–4240.
12. F. Qiu and J. Cho, “Automatic Identification of User Interest for Personalized Search,” *Proc. 15th Int’l Conf. World Wide Web*, 2006, pp. 727–736.
13. E. Agichtein et al., “Learning User Interaction Models for Predicting Web Search Result Preferences,” *Proc. 29th Ann. Int’l ACM SIGIR Conf. Research and Development in Information Retrieval*, 2006, pp. 3–10.
14. J. Teevan, S.T. Dumais, and E. Horvitz, “Personalizing Search via Automated Analysis of Interests and Activities,” *Proc. 28th Ann. Int’l ACM SIGIR Conference on Research and Development in Information Retrieval*, 2005, pp. 449–456.
15. M. Jiang et al., “Social Recommendation Across Multiple Relational Domains,” *Proc. 21st ACM Int’l Conf. Information and Knowledge Management*, 2012, pp. 1422–1431.


16. P. Cui et al., "Who Should Share What? Item-Level Social Influence Prediction for Users and Posts Ranking," *Proc. ACM Int'l Conf. Information Retrieval (SIGIR)*, 2011, pp. 185–194.
17. M. Lipczak, M. Trevisiol, and A. Jaimes, "Analyzing Favorite Behavior in Flickr," *Advances in Multimedia Modeling*, Springer, 2013, pp. 535–545.
18. X. Xie et al., "Learning User Interest for Image Browsing on Small-Form-Factor Devices," *Proc. SIGCHI Conf. Human Factors in Computing Systems*, 2005, pp. 671–680.
19. Y. Yang et al., "User Interest and Social Influence Based Emotion Prediction for Individuals," *Proc. 21st ACM Int'l Conf. Multimedia*, 2013, pp. 785–788.
20. P. André et al., "Designing Novel Image Search Interfaces by Understanding Unique Characteristics and Usage," *Human-Computer Interaction—INTERACT*, Springer, 2009, pp. 340–353.
21. P.A. Chirita et al., "Using ODP Metadata to Personalize Search," *Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval*, 2005, pp. 178–185.
22. A. Sieg, B. Mobasher, and R. Burke, "Web Search Personalization with Ontological User Profiles," *Proc. 16th ACM Conf. Information and Knowledge Management*, 2007, pp. 525–534.
23. K. Lerman, A. Plangprasopchok, and C. Wong, "Personalizing Image Search Results on Flickr," *Intelligent Information Personalization*, 2007.
24. K. Stefanidis, E. Pitoura, and P. Vassiliadis, "Managing Contextual Preferences," *Information Systems*, vol. 36, no. 8, 2011, pp. 1158–1180.
25. B. Sarwar et al., "Item-Based Collaborative Filtering Recommendation Algorithms," *Proc. ACM Int'l Conf. World Wide Web (WWW)*, 2001, pp. 285–295.
26. Y. Koren, R. Bell, and C. Volinsky, "Matrix Factorization Techniques for Recommender Systems," *Computer*, vol. 42, no. 8, 2009, pp. 30–37.
27. J. Leskovec, A. Singh, and J. Kleinberg, "Patterns of Influence in a Recommendation Network," *Advances in Knowledge Discovery and Data Mining*, Springer, 2006, pp. 380–389.
28. M. Jamali and M. Ester, "Trustwalker: A Random Walk Model for Combining Trust-Based and Item-Based Recommendation," *Proc. 15th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining*, 2009, pp. 397–406.
29. S. Branson, P. Perona, and S. Belongie, "Strong Supervision from Weak Annotation: Interactive Training of Deformable Part Models," *Proc. IEEE Int'l Conf. Computer Vision (ICCV)*, 2011, pp. 1832–1839.
30. G. Hua et al., "Collaborative Active Learning of a Kernel Machine Ensemble for Recognition," *Proc. IEEE Int'l Conf. Computer Vision (ICCV)*, 2013, pp. 1209–1216.
31. C. Long, G. Hua, and A. Kapoor, "Active Visual Recognition with Expertise Estimation in Crowdsourcing," *Proc. IEEE Int'l Conf. Computer Vision (ICCV)*, 2013, pp. 3000–3007.
32. Z. Wang et al., "Propagation-Based Social-Aware Replication for Social Video Contents," *Proc. ACM Int'l Conf. Multimedia*, 2012, pp. 29–38.
33. E. Bakshy et al., "Everyone's an Influencer: Quantifying Influence on Twitter," *Proc. ACM Int'l Conf. Web Search and Data Mining (WSDM)*, 2011, pp. 65–74.
34. H. Li, H. Wang, and J. Liu, "Video Sharing in Online Social Network: Measurement and Analysis," *Proc. ACM Network and Operating System Support on Digital Audio and Video Workshop (NOSSDAV)*, 2012, pp. 83–88.
35. J.M. Pujol et al., "The Little Engine(s) that Could: Scaling Online Social Networks," *ACM SIGCOMM Computer Comm. Rev.*, vol. 40, no. 4, 2010, pp. 375–386.
36. S. Scellato et al., "Track Globally, Deliver Locally: Improving Content Delivery Networks by Tracking Geographic Social Cascades," *Proc. 20th Int'l Conf. World Wide Web*, 2011, pp. 457–466.
37. A. Balachandran et al., "Developing a Predictive Model of Quality of Experience for Internet Video," *Proc. ACM SIGCOMM 2013 Conf.*, pp. 339–350.

Wenwu Zhu is a full professor at Tsinghua University, China. Contact him at wwzhu@tsinghua.edu.cn.

Peng Cui is an assistant professor at Tsinghua University, China. Contact him at cui@tsinghua.edu.cn.

Zhi Wang is an assistant professor at Tsinghua University, China. Contact him at wangzhi@sz.tsinghua.edu.cn.

Gang Hua is an associate professor at Steven Institute of Technology. Contact him at ganghua@gmail.com.

 Selected CS articles and columns are also available for free at <http://ComputingNow.computer.org>.