Guest Editorial: Large-scale Multimedia Data Management: Techniques and Applications

Jason C. Hung[1] · Makoto Takizawa[2] · Shu-Ching Chen[3]

1 Introduction

Essential to many tasks in relation to multimedia research and development is the availability of a sufficiently large data set and its corresponding ground truth. However, most available data for multimedia research are either too specific, e.g., data for text retrieval; too small, e.g., face figures; nor without ground truth, such as gathering millions of un-preprocessing images from the Web for testing. While it is relatively easy to crawl and store a huge amount of data, the creation of ground truth necessary to systematically train, test, evaluate, and compare the performance of various algorithms and systems remains a challenging issue. For this reason, researchers tend to put (or re-direct) efforts into the creation of such corpus individually to carry out research on large-scale data sets. Thus, a promising trend of a united web-scale and distributed multimedia data management is urgently needed, which would benefit the entire multimedia research community.

This special issue presents and reports on the construction and analysis of large-scale multimedia data sets and resources, and provides a strong reference for multimedia researchers interested in large-scale multimedia data sets. In particular, the special issue demonstrates the emerging techniques and applications for large-scale multimedia data management.

✉ Jason C. Hung
jasonhung2013@gmail.com

Makoto Takizawa
makoto.takizawa@computer.org

Shu-Ching Chen
chens@cs.fiu.edu

[1] Department of Information Technology, Overseas Chinese University, Taichung 407, TaiwanR.O.C.

[2] Department of Advanced Sciences, Hosei University, Tokyo 184-8584, Japan

[3] School of Computing and Information Sciences, Florida International University, Miami, FL 33199, USA

This special issue contains original papers describing the latest developments, trends, and solutions, such as:

- Algorithms, techniques, framework, and models for multimedia computing
- Multimedia human-computer interaction
- Mobile and multi-device empowered multimedia
- Large-scale multimedia data management
- Ubiquitous/pervasive data for multimedia
- Social media and presence
- Cloud-based multimedia services
- Web-scale data management and analysis
- Security issues for multimedia computing
- Media/data transport, analysis, and delivery
- Data searching, browsing, and discovery
- Emerging systems, services, and middleware
- Crowd-sourcing, authoring, and collaboration
- Recent other issues in large-scale multimedia data

2 Related works

Seung-Hoon Chae et al. [3] suggested a method to auto-configure the initial contour in the level-set method. Multi-resolution analysis helped in reducing the pace of the auto-configuration process of the initial contour. In addition, the volume data of a CT image was used to prevent data loss that occurs during the MRA transformation process.

Jia Uddin et al. [32] presented a graphics processing unit (GPU)-based implementation of a Bellman-Ford (BF) routing algorithm [1] using NVIDIA's compute unified device architecture (CUDA) [19]. The Bellman-Ford (BF) algorithm computes the shortest paths from a single source vertex to all of the other vertices in a weighted graph. In the proposed GPU-based approach, multiple threads run concurrently over numerous streaming processors in the GPU to dynamically update routing information. Instead of computing the individual vertex distances one-by-one, a number of threads concurrently updated a larger number of vertex distances, and an individual vertex distance was represented in a single thread.

Kwangmu Shin et al. [28] proposed a novel stereo matching approach that was robust in controlling various radiometric variations such as local and global radiometric variations. They designed a hybrid stereo matching approach using transition of pixel values and data fitting. Transition of pixel values was utilized for the coarse stereo matching stage, and polynomial curve fitting was used for the fine stereo matching stage. Consequentially, they demonstrated that the proposed method was less sensitive to various radiometric variations, and showed an outstanding performance in computational complexity. Jianxin Liao et al. [20] presented LFFIR, a multi-feature image retrieval framework for content similar search in the distributed situation. The key idea was to effectively incorporate image retrieval based on multi-feature into the peer-to-peer (P2P) [29] paradigm. LFFIR fused the multiple features in order to capture the overall image characteristics. And then, it constructed the distributed indexes for the fusion feature through exploiting the property of locality sensitive hashing (LSH).

Oh Jung Min et al. [23] presented a database management scheme for an intelligent surveillance system utilizing multiple visual sensors and RFID readers. The objects were tracked and identified

by multiple visual sensors and RFID readers. They defined three different types of data structure to consistently store data for effective data storage. They contained global object number and identification as the common information of the same object. The global object number was uniquely assigned for each track object. The previously stored data without the common information was back-annotated when it was available in the system. Moreover, when the global object number changed because of imperfect detection and tracking, the system maintained consistency information between global object numbers for the same object by comparing their local target information or positions. The fragmented information for an object was also stitched through map information.

Myeongsu Kang et al. [16] showed that the formant synthesis process using multiple pairs of digital resonators and band-pass filters was accelerated with the power of a general-purpose graphics processing unit (GPGPU) [2, 8, 25, 27, 30, 31]. This research compared the performance of the proposed GPGPU-based parallel approach with the CPU-based sequential approach in order to validate the effectiveness of the proposed massively parallel method.

Xavier Jerald Punithan et al. [24] proposed a two-player non-cooperative zero-sum game with incomplete information for dynamic intrusion signature configuration (DISC), where the various lengths of an intrusion signature had been activated in a time-shared manner. After formulating the problem into the game theoretic approach, they found the optimal strategy for DISC in the S-IDS. To the best of our knowledge, this work was the first of its kind that analyzes the optimal DISC strategy against the various mutants of intrusion packets.

ChangWon Jeong et al. [12] suggested a medical image information system environment using data synchronization methods. They designed as synchronization methods using detection of creation image data on components of system. Also, they used the cloud-computing environment, which reduced the number of high-latency image transmissions. Finally, they showed the data synchronization process of the system with imaging application services based on a cloud-computing service.

Seung-Won Jung et al. [14] proposed a simple but effective method for obtaining an all-in-focus (AIF) color image from a database of color and depth image pairs. Since the defocus blur was inherently depth-dependent, the color pixels were first grouped according to their depth values. The defocus blur parameters were then estimated using the amount of the defocus blur of the grouped pixels. Given a defocused color image and its estimated blur parameters, the AIF image was produced by adopting the conventional pixel-wise mapping technique. In addition, the availability of the depth image disambiguated the objects located far or near from the in-focus object and thus facilitates image refocusing.

Shingchem D. You et al. [34] studied the accuracy of detecting singing segments using the hidden Markov model (HMM) classifier with various features, including Mel frequency cepstral coefficients (MFCC) [26], linear predictive cepstral coefficients (LPCC), and linear prediction coefficients (LPC). Simulation results showed that detecting singing segments in a soundtrack was more difficult than detecting them among pure-instrument segments. In addition, combining MFCC and LPCC yields higher accuracy. The bootstrapping technique has only limited accuracy improvement to detect all singing segments in a soundtrack.

Shuai Liu et al. [22] used fractal image encoding into the compression because of its high compression ratio by extracted and analyzed the loss in the fractal encoding. To solve the most important problem in fractal image encoding method, which was its high computational complexity and long encoding time, they first used statistical analysis to the fractal encoding method. They created its box-plot to find the distributional of loss value. Then, they partitioned them to several parts and map them to the given model. After that, they presented a novel method to save the loss and maintain the quality in image compression.

Ing-Jr Ding et al. [7] explored the well-known HMM pattern recognition method with the support of the Kinect device to classify the human's active gestures where a user adaptation scheme of MAP+GoSSRT that enhances MAP by incorporating group of states shifted by referenced transfer (GoSSRT)

Weina Fu et al. [9] proposed a novel method, which was suitable for applying on relatively high-resolution videos that moving objects can be distinguished from their color and shape information. This method matched and tracked multiple moving objects in video by extracting and combining multi-features. With the background re-construction method, the moving objects were separated as sub images from the background; they first extracted some valuable features from each sub image, especially the topological information. Then, features were applied to a strong classifier which was accumulated with weak feature classifiers. After that, by the initial matching of moving objects, they extracted their kinematical features to reinforce the matching method.

Hao Chen et al. [4] introduced the continuous development of information technology. Various multimedia data were constantly emerging and presented the characteristics of autonomous and heterogeneous; how to integrate and analysis data more correctly and efficiently has become a challenging problem. Firstly, in order to improve the quality of the integrated data, two real-time threads combined with data adapter were used to monitor and refresh necessary updates from heterogeneous data efficiently. Once the original data had been updated, the real-time data will be loaded into the data center soon. Secondly, a data reverse cleaning method was proposed to improve the data quality. It used the data source tree that built in the data integration process to find the location of the original data quickly after reverse cleaning. Finally, a data accuracy assessment algorithm was designed for data quality assessment, which was based on Bayesian network and the path condition algorithm.

Hyosook Jung et al. [15] proposed an approach to introduce the Semantic Web to novice users. To this end, they had built an easy-to-use system that helps users create simple RDF documents and construct a small-scale Semantic Web-like environment. Their system could take an input that a user provided and created an RDF document and all that the user needed to do was to define a string for the RDF document according to the grammar. Users could also define simple rules using the grammar and practice programming using RDF documents.

Gelan Yang et al. [33] proposed a wavelet-energy based new approach for automated classification of MR human brain images as normal or abnormal. SVM was used as the classifier, and biogeography-based optimization (BBO) was introduced to optimize the weights of the SVM. The study offered a new means to detect abnormal brains with excellent performance.

Ruijun Liu et al. [21] proposed a novel feature encoding method called label constrained sparse coding (LCSC) for visual representation. The visual similarities between local features were jointly considered with the corresponding label information of local features. This was achieved by combining the label constraints with the encoding of local features. In this way, they could ensure that similar local features with the same label were encoded with similar parameters. Local features with different labels were encoded with dissimilar parameters to increase the discriminative power of encoded parameters. Besides, instead of optimizing for the coding parameter of each local feature separately, they jointly encoded the local features within one sub-region in the spatial pyramid way to combine the spatial and contextual information of local features. They applied this label constrained sparse coding technique for classification tasks on several public image datasets to evaluate its effectiveness.

Dong-yuan Ge et al. [10] proposed a new approach for binocular vision system calibration and 3D re-construction. While the system was calibrated, the sum of square distances between the

vector coordinates recombined with the coordinates of feature points in the world frame and those in image frame to the fitted hyperplane was taken as an objective function. An orthogonal learning neural network was designed, where a self-adaptive minor component extracting method was adopted. When the network comes to equilibrium, the projective matrixes for the two cameras were obtained from the eigenvectors of the autocorrelation matrix corresponding to the minimum eigenvalues, so the calibration of the binocular vision system was achieved. As for 3D re-construction, an autocorrelation matrix was obtained from feature point coordinates in image planes and calibration data, and an orthogonal learning network was designed. After the network was trained, the autocorrelation matrix's eigenvector corresponding to the minimum eigenvalues was obtained, from which the 3D coordinates are obtained also. The proposed approach was a novel application of minor component analysis and orthogonal learning network in binocular vision system and 3D re-construction.

Ching-Nung Yang et al. [18] show a new data hiding scheme for verifying the embedding rate during the embedding and extracting phases. The proposed research "Hamming+3," which is a reasonably acceptable steganography method, shows better performance than "Hamming+1." Woogyoung Jun et al. [13] proposed the duplicate video detection for large-scale multimedia that is based on block histogram. It uses a dynamic matching algorithm for fast and real-time process of large-scale data. Ran Choi et al. [5] proposed a method that uses lattice block pattern to be tested in a white plaster sphere with 14-cm diameter to reconstruct 3D surface curvature. Junchul Chun et al. [6] shows an idea for 3D face pose estimation by a robust real-time tracking of facial features. This research detects facial the facial region and major facial features by Haar-like feature and AdaBoost learning algorithm. Finally, Jin-Mook Kim et al. [17] suggested a novel model of combined risk probability map generation to predict crime frequency. It analyzes risk in collective residential areas using urban spatial information, and then two risk probability maps are generated by the information based on terrain.

References

1. Bellman R (1958) On a routing problem. Q J Appl Math 16:87–90
2. Belloch JA, Gonzalez A, Martinez-Zaldivar FJ, Vidal AM (2011) Real-time massive convolution for audio applications on GPU. J Supercomput 58:449–457
3. Chae S-H, Moon H-M, Chung Y, Shin J, Pan SB (2014) Automatic lung segmentation for large-scale medical image management. Multimed Tools Appl. doi:10.1007/s11042-014-2201-1
4. Chen H, Ouyang Y, Jiang W (2015) An optimized data integration model based on reverse cleaning for heterogeneous multi-media data. Multimed Tools Appl. doi:10.1007/s11042-015-2683-5
5. Choi R, Cho C-S (2015) An efficient approach for obtaining 3D surface curvature using blocked pattern projection. Multimed Tools Appl. doi:10.1007/s11042-015-2902-0
6. Chun J, Kim W (2014) 3D face pose estimation by a robust real time tracking of facial features. Multimed Tools Appl. doi:10.1007/s11042-014-2356-9
7. Ding I-J, Chang C-W (2015) An adaptive hidden Markov model-based gesture recognition approach using Kinect to simplify large-scale video data processing for humanoid robot imitation. Multimed Tools Appl. doi:10.1007/s11042-015-2505-9
8. Divya UJ, Kim HS, Lee J, Kim JI (2013) Fractal based method on hardware acceleration for natural environments. J Convers 4(3):6–12
9. Fu W, Zhou J, Ma Y (2015) Moving tracking with approximate topological isomorphism. Multimed Tools Appl. doi:10.1007/s11042-015-2519-3
10. Ge D-Y, Yao X-F, Lian Z-T (2015) Binocular vision calibration and 3D re-construction with an orthogonal learning neural network. Multimed Tools Appl. doi:10.1007/s11042-015-2845-5

11. Hsu B, Sosnick-Perez M (2013) Real-time GPU audio. Commun ACM 56(6):54–62

12. Jeong C-W, Kim W-H, Lypengleang S, Jeong Y-S, Joo S-C, Yoon K-H (2015) The development of a medical image information system environment using data synchronization based on cloud computing. Multimed Tools Appl. doi:10.1007/s11042-015-2506-8

13. Jun W, Lee Y, Jun B-M (2015) Duplicate video detection for large-scale multimedia. Multimed Tools Appl. doi:10.1007/s11042-015-2724-0

14. Jung S-W, Park JH, Jeong Y-S (2015) All-in-focus and multi-focus color image reconstruction from a database of color and depth image pairs. Multimed Tools Appl. doi:10.1007/s11042-015-2535-3

15. Jung H, Yoo S, Kim D, Park S (2015) A grammar based approach to introduce the Semantic Web to novice users. Multimed Tools Appl. doi:10.1007/s11042-015-2898-5

16. Kang M, Islam S, Islam R, Kim J-M (2014) Accelerating the formant synthesis of haegeum sounds using a general-purpose graphics processing unit. Multimed Tools Appl. doi:10.1007/s11042-014-2297-3

17. Kim D-H, Kim J-M, Jeong Y-S, Park K-R (2015) A risk probability-map generation model on multimedia services environment. Multimed Tools Appl. doi:10.1007/s11042-014-2441-0

18. Kim C, Yang C-N (2014) Data hiding based on overlapped pixels using hamming code. Multimed Tools Appl. doi:10.1007/s11042-014-2355-x

19. Lalami ME, El-Baz D, Boyer V (2011) Multi GPU implementation of the simplex algorithm. IEEE Int Conf High Perform Comput Commun, Banff:179–186

20. Liao J, Yang D, Li T, Qi Q, Wang J, Sun H (2015) Multimed Tools Appl. doi:10.1007/s11042-015-2892-y

21. Liu R, Chen Y, Zhu X, Hou K (2015) Image classification using label constrained sparse coding. Multimed Tools Appl. doi:10.1007/s11042-015-2626-1

22. Liu S, Zhang Z, Qi L, Ma M (2015) A fractal image encoding method based on statistical loss used in agricultural image compression. Multimed Tools Appl. doi:10.1007/s11042-014-2446-8

23. Oh J-M, Moon N, Hong S (2015) Trajectory based database management for intelligent surveillance system with heterogeneous sensors. Multimed Tools Appl. doi:10.1007/s11042-015-2725-z

24. Punithan XJ, Kim J-D, Kim D, Choi Y-H (2015) A game theoretic model for dynamic configuration of large-scale intrusion detection signatures. Multimed Tools Appl. doi:10.1007/s11042-015-2508-6

25. Ranjan R, Gan WS (2014) Fast and efficient real-time GPU based implementation of wave field synthesis. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, Florence, 4–9

26. Rocamora M, Herrera P (2007) Comparing audio descriptors for singing voice detection in music audio files. Proc. of 11th Brazilian Symposium on Computer Music:1–10

27. Savloja L, Valimaki, V, Smith III JO (2010) Real-time additive synthesis with one million sinusoids using a GPU. In: 128th AES Conventions, London, pp. 115

28. Shin K, Kim D, Chung K (2015) Visual stereo matching combined with intuitive transition of pixel values. Multimed Tools Appl. doi:10.1007/s11042-015-2962-1

29. Stoica I, Morris R, Karger D, Kaashoek MF, Balakrishnan H (2001) Chord: a scalable peer-to-peer lookup service for internet applications. The 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM'01), San Diego, USA, 149–160

30. Trebien F, Oliveira MM (2009) Realistic real-time sound re-synthesis and processing for interactivevirtual worlds. Vis Comput 25:469–477

31. Tsai PY, Wang TM, Su A (2010) GPU-based spectral model synthesis for real-time sound rendering. In: Proceedings of the 13th International Conference on Digital Audio Effects, Graz, pp. 1–5

32. Uddin J, Jeong I-K, Kang M, Kim C-H, Kim J-M (2014) Accelerating IP routing algorithm using graphics processing unit for high speed multimedia communication. Multimed Tools Appl. doi:10.1007/s11042-014-2013-3

33. Yang G, Zhang Y, Yang J, Ji G, Dong Z, Wang S, Feng C, Wang Q (2015) Automated classification of brain images using wavelet-energy and biogeography-based optimization. Multimed Tools Appl. doi:10.1007/s11042-015-2649-7

34. You SD, Wu Y-C, Peng S-H (2015) Comparative study of singing voice detection methods. Multimed Tools Appl. doi:10.1007/s11042-015-2894-9