GUEST EDITORS' INTRODUCTION

Integrity of Multimedia and Multimodal Data: From Capture to Use

Amit Kumar Singh, *National Institute of Technology Patna, Patna, 800005, India*

Deepa Kundur, *University of Toronto, Toronto, ON, M5S 1A1, Canada*

Min Wu, *University of Maryland, College Park, MD, 20742, USA*

Mauro Barni, *University of Siena, 53100, Siena, Italy*

With the development of Internet technologies, large volumes of multimedia and more generally, multimodal data can easily be shared among different users or entities for the purpose of diagnosis, analysis, and prediction in a variety of critical application contexts with no hard-copy backups being kept. This has serious implications for integrity, authentication, provenance, and utilization in a number of application domains. Additionally, various actions, including unlawful alteration, unauthorized distribution, copyright violation, and misuse of sensitive and private multimedia data, are growing each day producing negative impacts on the accuracy, efficiency, safety, and economics of multimodal data systems. Although there are many tools available to create, encrypt, embed, and extract data, there are relatively limited matured tools available and it remains a work-in-progress to establish integrity, authentication, and provenance. Therefore, addressing these challenges has been of interest for researchers in the field. Furthermore, various theories, techniques, and best practices developed by the community for handling multimedia data can provide useful inspirations or generalizable to multimodality data, including but not limited to multimedia. The advent of the Internet of Things (IoT), cyber-physical systems, robotics, as well as personal and wearable devices provides many opportunities for the multimedia community to develop synergies.

The guest editorial team (Amit Kumar Singh, Deepa Kundur, Min Wu, and Mauro Barni) hopes that the articles included in this special issue contribute to the state-of-the-art solutions of the integrity and utilization of multimedia and multimodal data for real-world applications. Our call for papers received an enthusiastic response with 23 submissions in this particular topic. After a rigorous review process, we accepted seven articles to form the special issue. The first three articles cover several core issues related to integrity, and the next four articles showcase multiple applications for which trustworthy and/or privacy-protected multimedia or multimodal data are important. A brief summary of each paper is introduced below.

Due to the storage issue, uncompressed videos are generally single compressed before distribution, and when videos have been tampered with by an editing tool on the malicious distributor side, a second compression occurs during the storage process. As the video forgery process commonly involves double compression, detecting recompression traces helps verify the integrity of a video. Motivated by this, Nam *et al.* present a neural network-based video forensics (referred to as DHNet) using discriminant features in "DHNet: Double MPEG-4 Compression Detection via Multiple DCT Histograms." Furthermore, the performance improvement was achieved by providing auxiliary features to dense layers. The results revealed that this work is superior to baseline methods, including CNN and non-CNN methods, in terms of various objective metrics. Aprilpyone and Kiya present a privacy-preserving image classification method that uses encrypted images and an isotropic network, such as the vision transformer, in "Privacy-Preserving Image Classification Using an Isotropic Network." The proposed solution allowed them to use a large image size and achieves high classification, robustness against attacks, and efficient compression. Recently, privacy and security of medical data is a central problem in intelligent hospital ecosystems. In "A Novel Security Framework for Medical Data in IoT Ecosystems," Bayari *et al.* discuss a novel yet simple security framework to ensure the confidentiality of medical data during the communication between IoT hops. The proposed framework utilizes the concept of chaotic pseudorandom sequence-based encryption to ensure the security. The experimental study of the proposed

1070-986X © 2022 IEEE
Digital Object Identifier 10.1109/MMUL.2022.3180122
Date of current version 6 July 2022.

framework shows that the encryption is robust, efficient, and lightweight, making it suitable for resource-constrained IoT scenarios.

With strong integrity enhanced on multimedia and/or multimodal data, the next set of articles focus on applications for which integrity and privacy are essential. Continuing in the medical application domain, we are reminded that Alzheimer's disease (AD) is a prevalent, irreversible, chronic, and degenerative disorder whose diagnosis at prodromal stage is critical. Mostly, single data modalities, such as MRI or PET, are used to make predictions in AD studies. However, metabolic and structural data fusion can provide a holistic view of AD staging analysis. Facing the aforementioned situations, in "Multimodal Fusion-Based Deep Learning Network for Effective Diagnosis of Alzheimer's Disease," Dwivedi *et al.* present a multimodal fusion-based ResNet-50 DL network that demonstrates significant performance over single modality for AD diagnosis. Evaluation of the proposed model is performed over both single modality and multimodality fused data for diagnostic classification purposes. The role of multimodal fusion in providing better detection accuracy than single-modal approaches can also be seen in mental health applications. In recent years, depression of university students has aroused general concern. Fast detection of students at risk of depression using multimedia data is a challenge. In this direction, in "Postgraduate Student Depression Assessment by Multimedia Gait Analysis," Lu *et al.* suggest an integrated gait assessment framework that contains the collection and analysis of multimedia data to assess risk of depression for postgraduate students. Compared with the conventional methods, the method increases significantly the objectivity of depression assessment in both clinical theory and practice.

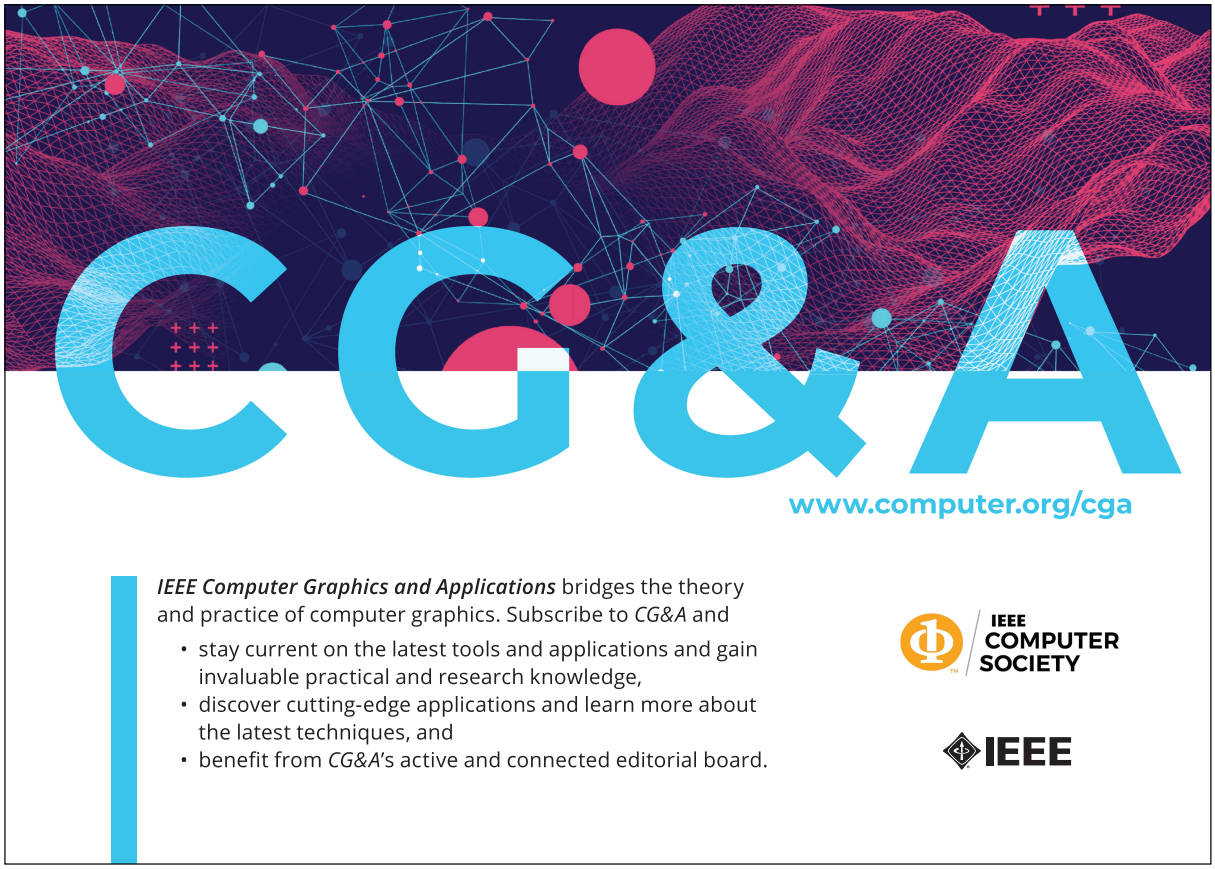
With the development of 5G networks, vehicle-to-vehicle communication—once integrity ensured—is helping to make travel safe. However, vehicle-type detection in the multimedia feed remains a problem. In order to deal with this challenge, in "Efficient Multimedia Frame-Skipping Architecture Using Deep Learning in Vehicular Networks," Ahmed *et al.* focus on the development and deployment of a deep-learning-based system for traffic congestion analysis. The model uses multiple video feeds and vehicle information to detect, classify, and count vehicles in the live traffic feed. The model is trained with a deep learning approach to align the video image and detect the object in top–down multimedia. When building on top of integrity ensured technologies, this proposed model has potential applications in intelligent transportation involving real-time data analysis and multimedia tasks. With the popularity of virtual reality (VR) games, how the game contexts influence players' head movement and the effect of head movement on VR sickness is a topic worth studying. In this direction, in "Why VR Games Sickness? An Empirical Study of Capturing and Analyzing VR Games Head Movement Dataset," Dong *et al.* have collected a head movement and heart rate dataset of 30 participants while playing different VR games. After playing VR games, the simulator sickness questionnaires were filled by the participants, and the relationship between head movement, heart rate, and VR sickness is then analyzed. Given the potentially sensitive and private nature of the head movement trajectory dataset collected, this is an interesting and emerging application that may benefit from privacy protection technologies.

ACKNOWLEDGMENTS

The guest editors would like to thank all the authors who submitted their articles and the anonymous reviewers who carefully reviewed and evaluated them. They extend their sincere thanks to the Editor-in-Chief of *IEEE MultiMedia*, Prof. Shu-Ching Chen, and the Associate Editor-in-Chief in charge of special issues, Prof. Mohan Kankanhalli, for providing the opportunity and guidance to edit this special issue and the editorial staff for their continuous support in organizing the special section.

AMIT KUMAR SINGH is an assistant professor at the Computer Science and Engineering Department, National Institute of Technology Patna, Patna, 800005, India. His research interests include multimedia data hiding, image processing, biometrics, and cryptography. Singh received his Ph.D. degree in computer engineering from the National Institute of Technology, Kurukshetra. He is a Senior Member of IEEE. Contact him at amit_245singh@yahoo.com.

DEEPA KUNDUR is a professor and chair at The Edward S. Rogers Sr. Department of Electrical & Computer Engineering, University of Toronto, Toronto, ON, M5S 1A1, Canada. Her research interests include interface of cybersecurity, signal processing, and complex dynamical networks. Kundur received her Ph.D. degree in electrical and computer engineering from the University of Toronto. She is a Fellow of IEEE and Canadian Academy of Engineering, and a Senior Fellow of Massey College. Contact her at dkundur@ece.utoronto.ca.

MIN WU is a professor and the associate dean of engineering and a distinguished scholar-teacher at the University of Maryland, College Park, MD, 20742, USA. Her research interests include information security and forensics, multimedia signal processing, and applications of data science and machine learning in health and the IoT. Wu received her Ph.D. degree in electrical engineering from Princeton University. She is a Fellow of IEEE, AAAS, and U.S. National Academy of Inventors. Contact her at minwu@umd.edu.

MAURO BARNI is a full professor at the Department of Information Engineering and Mathematics, University of Siena, 53100, Siena, Italy. His research interests include digital image processing and information security, with particular reference to the application of image processing techniques to copyright protection (digital watermarking) and authentication of multimedia (multimedia forensics). Barni received his Ph.D. degree in informatics and telecommunications from University of Florence, Italy. He is a Fellow of IEEE. Contact him at barni@dii.unisi.it.