Ivan V. Bajić 🆔, Marta Mrak 🆔, Frédéric Dufaux 🆔,
Enrico Magli 🆔, and Tsuhan Chen 🆔

Multimedia Signal Processing

A history of the Multimedia Signal Processing Technical Committee

©SHUTTERSTOCK.COM/TRIFF

Multimedia signal processing (MMSP) refers to processing of signals from multiple media—speech, audio, images, text, graphics, point clouds, etc.—often jointly. This article reviews the history of MMSP and, in parallel, the history of the MMSP Technical Committee (TC), with a focus on the last three decades (Figure 1).

Introduction

Overview of the MMSP TC

The MMSP TC of the IEEE Signal Processing Society (SPS) promotes the advancement of MMSP technology. The TC was formed in 1996. The scope of the TC includes joint processing/representation of audio–visual and multimodal information, fusion/fission of sensor information or multimodal data, integration of media, art, and multimedia technology, and analysis and feature extraction of multimodal data. Other key areas encompass virtual reality and 3D imaging, multimedia communications and networking, human–machine interface and interaction, visual and auditory quality assessment, multimedia databases, and digital libraries. In this context, the TC also serves as an incubator of technologies that lie in the gaps between traditional areas. Each year, the MMSP TC organizes the IEEE International Workshop on Multimedia Signal Processing, which attracts researchers from the SPS and related communities that work on multimedia topics. The workshop typically receives around 150 paper submissions and has more than 100 attendees from all over the world.

Historical context: the 1980s

Technological developments in the 1980s lay the foundation for the modern multimedia industry. The first CD appeared on the market in 1982 [1]. Personal computers gradually became more affordable throughout the decade and made their way into many homes. Video games, whose first prototypes appeared a few decades earlier, reached the level of popularity that made the gaming industry a notable segment of the tech sector. The first digital video coding standards, H.120 [2]

Multimedia Signal Processing

**Multimedia Signal Processing
Technical Committee**

1980's: First Digital Media Standards (CD, H.120, H.261) and WWW

1990

1991: WWW Open to Public, 2G (GSM) Cellular Communications

1992: JPEG, MPEG-1, MP3

1993: First Movie Streaming and First Live Streaming Over the Internet

1996: MPEG-2/H.262

1997: Commercial On-Demand Streaming (VXtreme and RealNetworks)

1996: MMSP TC Formed

1997: First MMSP Workshop (Princeton, NJ)

1998: Digital Terrestrial TV

1998: Second MMSP Workshop (Redondo Beach, CA)

1999: MPEG-4

1999: IEEE Transactions on Multimedia Launched

2000

2000: JPEG2000

2000: First ICME Conference (New York, NY)

2001: 3G Cellular Communications

2001: MMSP (Cannes, Fra.)

2002: ABR Over HTTP

2002: ICME (Lausanne)

2004: MPEG-4 AVC/H.264

2003: ICME (Baltimore)

2005: YouTube Launched

2004: MMSP (Siena, Italy)

2006: Amazon Unbox Launched Google Buys YouTube

2005: ICME (Amsterdam)

2006: MMSP (Victoria, BC)

2007: Netflix and Hulu Streaming

2007: ICME (Beijing)

2009: First 4G Cellular Service Launched in Norway and Sweden

2008: MMSP (Cairns, Australia), ICME (Hannover)

2009: Apple Introduces HTTP Live Streaming; Also Multi-View Coding (MVC) Extension of H.264/AVC Released

2009: MMSP (Rio de Janeiro) ICME (New York)

2010

2010: ICME Introduces Double-Blind Review, 15% Acceptance for Orals, 30% Overall

2011: MPEG-DASH

2012: AlexNet Wins ILSVRC

2013: HEVC

2013: ICME (San Jose, CA)

2014: Amazon Echo Launched

2014: MMSP (Banff, AB) ICME (Chengdu)

2015: IMT-2020 Standard Requirements for 5G Communications

2015: MMSP (Xiamen) ICME (Turin, Italy)

2016: HEVC-SCC

2017: MMSP (Luton, UK) ICME (Hong Kong)

2018: Waymo Launches First Fully Autonomous Taxi Service in Phoenix, AZ

2019: MMSP (Kuala Lumpur) ICME (Shanghai)

2020

2020: VVC and MPEG-PCC

2020: First Virtual ICME and MMSP Workshop

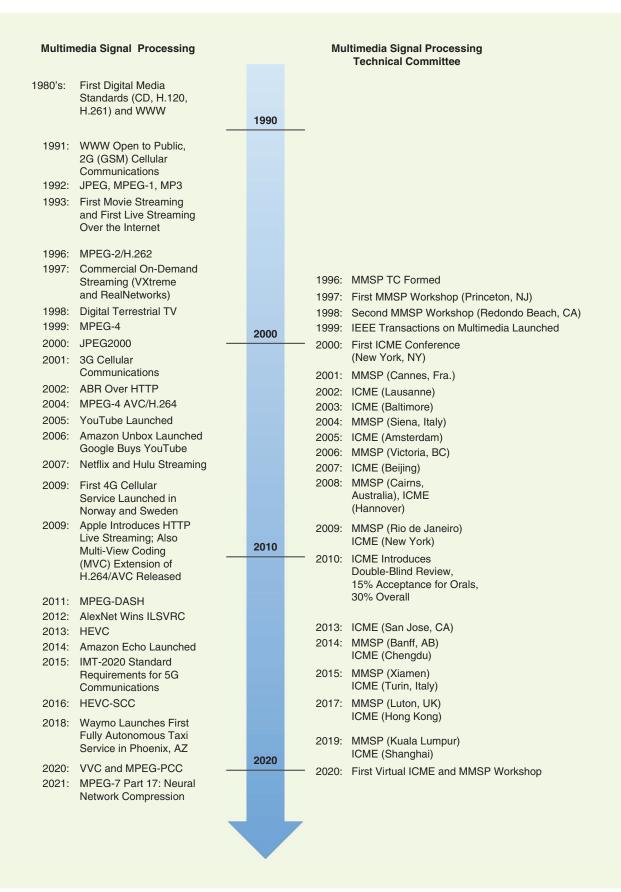2021: MPEG-7 Part 17: Neural Network Compression

FIGURE 1. MMSP timeline. See Table 1 for acronym definitions. ABR: adaptive bit rate; SCC: screen content coding.

and H.261 [3], and the first media platform World Wide Web (WWW, or simply the Web) [4] were developed in the 1980s, setting the stage for subsequent technological breakthroughs. To help the reader navigate the article, Table 1 gives a list of acronyms and their definitions, while Table 2 summarizes the multimedia standards mentioned in the text.

Developments in the last three decades

The 1990s
The 1990s were the decade of great milestones for digital multimedia. The Web was publicly released in 1991. As the first platform that enabled worldwide sharing of multimedia documents, combining text, images, graphics, audio, and video, it has since transformed the way we work, learn, shop, travel, keep in touch, and virtually all other aspects of our lives. Another pivotal event in 1991 was the launch of 2G cellular communications based on the global system for

> The scope of the TC includes joint processing/representation of audio–visual and multimodal information, fusion/fission of sensor information or multimodal data, integration of media, art, and multimedia technology, and analysis and feature extraction of multimodal data.

mobile communications (GSM) standard in Finland. Besides voice communications, 2G systems enabled short message service text messages, which later evolved to multimedia messaging service messages and lay the foundation for the myriad of today's messaging services. Cellular communications have had an equally transformational effect on our lives, providing the infrastructure over which much of today's multimedia content is being shared.

The first widely used multimedia standards were released in 1992. The Joint Photographic Experts Group (JPEG) published Part 1 of the JPEG image coding standard [5], the most popular image coding standard to date. The Moving Picture Experts Group (MPEG) issued MPEG-1 [6], the audio–visual coding standard that formed the basis of video CD and early digital cable and satellite TV. The standard also introduced MPEG-1 audio layer III [7], more commonly known as MP3, a widely popular audio format for music sharing.

Table 1. Acronyms and their definitions in alphabetical order.

Acronym	Definition
AAC	Advanced audio coding
AOM	Alliance for open media
AVC	Advanced video coding
CD	Compact disc
CDVS	Compact descriptors for visual search
DASH	Dynamic adaptive streaming over HTTP
DVD	Digital video disc
GSM	Global system for mobile communications
HDTV	High-definition television
HEVC	High-efficiency video coding
HTTP	Hypertext transfer protocol
ICME	International Conference on Multimedia & Expo
JPEG	Joint Photographic Experts Group
Lidar	Light detection and ranging
MMS	Multimedia messaging service
MP3	MPEG-1 audio layer III
MPEG	Moving Picture Experts Group
MVC	Multiview video coding
P2P	Peer-to-peer communications
PCC	Point cloud compression
RGB+D	Red, green, blue plus depth
SMS	Short message service
VCM	Video coding for machines
VVC	Versatile video coding
WWW	World wide web

Table 2. Select multimedia standards in chronological order of their release.

Standard	Description	Initial Release
H.120	First digital video coding standard	1984
H.261	Video coding standard, targeted mainly at video telephony	1988
JPEG	First digital image coding standard	1992
MPEG-1	Audiovisual coding standard	1992
MPEG-2	Audiovisual coding standard	1996
MPEG-2 Part 2 (H.262)	Video coding standard	1996
MPEG-4	Audiovisual coding standard	1999
JPEG2000	A wavelet-based image coding standard	2000
MPEG-4 Part 10 (H.264/AVC)	Video coding standard	2004
MVC	Multiview video coding, amendment to MPEG-4 Part 10 (H.264/AVC)	2009
MPEG-DASH	MPEG dynamic adaptive streaming over HTTP, a video streaming standard	2011
MPEG-H Part 2 (H.265/HEVC)	Video coding standard	2013
3D-HEVC	HEVC-based coding standard for 3D video	2015
HEVC-SCC	HEVC-based coding standard for screen content video	2016
MPEG-I Part 3 (H.266/VVC)	Video coding standard	2020
MPEG-PCC	Point cloud compression standard	2020
MPEG-7 Part 17	A standard for compression of neural network models	2021

The year 1993 was a big year for video streaming. On 22 May, the movie called *Wax or the Discovery of Television Among the Bees* became the first movie to be streamed online, at half the standard definition resolution and a frame rate of only two frames per second. The first live streaming over the Internet occurred on 24 June 1993. This was a performance by the band called *Severe Tire Damage*, streamed from the Xerox Palo Alto Research Center [9]. The video resolution was only 152×76, the frame rate only 8 to 12 frames per second, and the audio quality no better than a telephone call, but it could be seen as far as Australia [10]. It was a historic event that demonstrated the potential of streaming technology and stimulated much research and development in the decades to follow. Propelled by this success, *Severe Tire Damage* opened for the *Rolling Stones* in the second live streaming of a musical event over the Internet on 18 November 1994 [11].

In 1994, DirecTV launched the first commercial digital satellite TV service in the United States. This marked the beginning of the transition of TV from analog to digital, which continued with the introduction of digital cable TV in 1996 and digital terrestrial TV in 1998. MPEG-2 was released in 1996 [6]. It was a very popular coding standard that was used in DVDs, digital TV, and HDTV. Its Part 7, advanced audio coding, was released in 1997.

As the Internet reached more users through the 1990s, with increased capacity and higher bit rates, the battle for streaming over the Internet would heat up, especially in the second part of the decade [12]. The big players in this area were Progressive Networks (which became RealNetworks in 1997) and Microsoft, along with a number of startups, including Vivo, Xing, VDOnet, and VXtreme. Among them, they are responsible for a number of firsts, including the first audio streaming service (RealAudio, 1995), first live audio webcast of a sports game (*Seattle Mariners* versus *New York Yankees*, RealNetworks, 1995), first commercial on-demand video streaming (RealNetworks and VXtreme, 1997), as well as the most popular media players of the time.

Amid all of these developments, the MMSP TC was formed in 1996, under the leadership of the first TC chair, Tsuhan Chen. The TC organized its first workshop, the MMSP Workshop, in June 1997 in Princeton, New Jersey. The workshop attracted 95 papers, including eight demonstrations. The next two workshops were held in Redondo Beach, California, (1998) and Copenhagen, Denmark (1999). The second TC chair (1999–2001) was K. J. Ray Liu, the 2022 IEEE president.

Because of the interdisciplinary nature of multimedia, the MMSP TC has collaborated with other TCs within the IEEE SPS and other IEEE societies since its inception. A notable result of such collaboration was the launch of the *IEEE Transactions on Multimedia* in 1999. With an impact factor of 8.18,

As the first platform that enabled worldwide sharing of multimedia documents, combining text, images, graphics, audio, and video, it has since transformed the way we work, learn, shop, travel, keep in touch, and virtually all other aspects of our lives.

the journal is now considered among the top publication venues in the field of multimedia.

The 2000s

If the 1990s demonstrated the potential of multimedia technologies, the 2000s were the decade when the technology reached the level of maturity that made it not only commercially viable, but highly successful. This was aided by the development of 3G cellular communications, first commercially launched in 2001 by NTT DoCoMo in Japan. In addition, new ideas emerged both from industry and academia. One of these ideas was peer-to-peer (P2P) file sharing, pioneered by Napster.

Napster initially launched in 1999 and quickly became a very popular platform for MP3 audio file sharing, especially among college students. It was soon sued over copyright infringement [13] and had to shut down in 2001. Despite its brief existence, Napster left a lasting legacy in the multimedia world. Its P2P distribution paradigm generated enormous interest in the research community and became popular not only as a way to share files, but also in media streaming. At the same time, the music industry saw the potential for distributing content in digital form without physical media, which lay the foundation for subsequent online music stores, such as iTunes, and products such as iPod.

At the turn of the millennium, the MMSP TC was also busy launching new initiatives. The International Conference on Multimedia and Expo (ICME) was launched as a collaboration with the sister TCs in the IEEE Circuits and Systems, Communications, and Computer Societies. The first edition of the conference was held in New York in July–August 2000, and attracted over 400 papers. Since then, the ICME has established itself as a flagship IEEE conference in the field of multimedia: it has a rank of A according to the Computing, Research, and Education Association of Australia rankings and is among the top 10 venues (among both journals and conferences) in the field of multimedia, according to Google Scholar metrics.

An important milestone at the turn of the millennium was the standardization of JPEG2000 [14]. This was the first coding standard based on wavelets [15]. JPEG2000 introduced tools for resolution- and quality-scalable coding and decoding, region-of-interest coding, precise rate control, and a number of other features that made it suitable for high-quality imaging applications. A related image coding approach, called *ICER*, is used for encoding and sending back images from the Mars rovers [16]. In 2004, Motion JPEG2000, an extension of JPEG2000 to video, was adopted for digital cinema applications in the film industry.

Another major milestone was the development of the MPEG-4 Part 10 advanced video coding (AVC) standard, better known as *H.264/AVC* [17], in 2003. One of the main motivations behind H.246/AVC was to support various network-based video services, such as video streaming to heterogeneous

clients. Hence, scalability also played an important role, and was materialized through the scalable extension of H.264/AVC [18], which enabled video coding and decoding at a number of resolutions, frame rates, and qualities to support a wide variety of client devices. H.264/AVC is used in Blu-ray discs and is still the most common format in online video streaming.

Although online video existed in various forms even in the 1990s, the first major video streaming service, YouTube, was launched in 2005, fueled by the development of H.264/AVC. YouTube allowed users to upload their own videos, which can then be searched and streamed to a wide audience. This quickly made YouTube very popular, leading to its purchase by Google for US$1.65 billion in 2006, less than a year after its official launch. Commercial streaming services appeared around the same time. Amazon Unbox (now Amazon Prime video) launched in 2006, followed by Netflix and Hulu streaming services in 2007. As the customers' home Internet service speeds improved, the popularity of streaming services grew, and streaming now accounts for more viewing time than cable TV in the U.S. market. These streaming services, and many more that have followed since, became successful businesses, some even launching their own production studios to create exclusive content.

> **JPEG2000 introduced tools for resolution- and quality-scalable coding and decoding, region-of-interest coding, precise rate control, and a number of other features that made it suitable for high-quality imaging applications.**

While certain forms of online games existed as far back as the 1970s, the era of massively multiplayer online gaming started in 2000s with the wider availability of fast Internet service. Gaming consoles such as Microsoft Xbox, Sony PlayStation, Nintendo, and Wii gradually became more popular, interfacing with cloud gaming platforms like Xbox Live and PlayStation Network. Increased interactivity in games and consoles' specialized hardware also incentivized the development of more sophisticated game controllers, like Microsoft Kinect, which would have a major impact on both gaming and MMSP research in the following decade.

The 2000s also saw the birth of social media, with the founding of LinkedIn in 2002, Facebook in 2004, and Twitter in 2006. A phenomenon that took the world by storm, social media allowed users to upload their own media content and share it with a circle of friends or a wider audience. Social media has since transformed marketing and market research, recruitment, the news industry, and many other aspects of our lives. It has also facilitated phenomena like trending, influencing, fake news, etc. On the technical side, the immense amount of user-supplied content ushered in the era of Big Data and set the stage for further technical developments in the coming decades. As an example, user-supplied photos and associated tags enabled Facebook to create a highly successful facial recognition system, which launched in 2010 but has since been scaled back due to ethical and privacy concerns.

The end of the decade was equally exciting in terms of technological developments. The 4G cellular service was first launched in Norway and Sweden in 2009. With increased bit rates offered to the users, the demand for online media and streaming services will rapidly increase in the next decade. The same year, Apple introduced HTTP live Sstreaming, which is currently the most popular streaming format. Also, the multi-view video coding extension of H.264/AVC was introduced.

In the meantime, the MMSP TC was busy building up the MMSP community and organizing related events. MMSP Workshops took place in Cannes, France (2001); St. Thomas, U.S. Virgin Islands (2002); Siena, Italy (2004); Shanghai, China (2005), Victoria, BC, Canada (2006); Chania, Greece (2007); Cairns, QLD, Australia (2008); and Rio de Janeiro, Brazil (2009). ICME conferences took place in New York, NY, USA (2000); Tokyo, Japan (2001); Lausanne, Switzerland (2002); Baltimore, MD, USA (2003); Taipei, Taiwan (2004); Amsterdam, The Netherlands (2005); Toronto, ON, Canada (2006); Beijing, China (2007); Hannover, Germany (2008); and again in New York, NY, USA (2009). During this period, MMSP TC chairs were K.-J. Ray Liu (1999–2001), John. A. Sørensen (2002–2003), Yu Hen Hu (2004–2005), Ingemar J. Cox (2006–2007), and Anthony Vetro (2008–2009).

The 2010s

During this decade, 4G communications spread throughout the world, increasing the demand for online media. Mobile screen resolutions increased sufficiently so that users could watch full HD video on their devices. Interactive media also became more popular; people could now have a reasonable videoconference on the go.

In 2011, MPEG dynamic adaptive streaming over HTTP (DASH) became an international standard. MPEG-DASH and related technologies, like Apple's HTTP live streaming, provided an incentive to consumer electronics companies to incorporate streaming apps into their devices, which in turn gave a boost to the streaming industry. Smart TVs and streaming devices like Apple TV, Amazon Fire TV, Roku, and many others, gradually started supplementing and then replacing traditional cable and satellite TV services.

Another type of application that became popular in the 2010s is mobile visual search [19], where users could take a photo of an object or a location and then retrieve additional information about it, possibly in the form of augmented reality. Audio search apps like Shazam were already established by that time, but an efficient mobile visual search required a good camera and sufficiently powerful hardware for fast feature extraction. All of it came together during the early 2010s. The MPEG compact descriptors for visual search standard [20] was released in 2015 and provided an interoperable way to compress and transmit visual features that facilitate image search and matching. While most multimedia compression standards code data for human consumption, this is a rare example of a standard for visual data coding for machine use, namely visual

search; the trend of coding for machines is becoming very popular at the time of the writing of this article.

The 2010s were a decade when immersive technologies took a big step forward. This was facilitated by improvements in sensing and display technologies over the years, but also computing infrastructure needed to process the increased amount of data required for a high-quality immersive experience. Representative technologies for 3D visual immersion include multiview video, red, green, blue plus depth (RGB+D), and point clouds, while audio counterparts include ambisonics and wave field synthesis. Haptic technologies also moved forward, finding new applications in wearable devices.

Another major event of the 2010s was a sharp rise in the popularity of deep learning with neural networks. Although the benefits of learning with many-layered models were already known in the 1960's [21] and the term *deep learning* dates back to the 1980s [22], it was the success of deep neural networks in acoustic modeling [23] and image classification [24], as well as the availability of large data sets and powerful computing infrastructure, that sparked the renewed interest in the topic, and subsequently transformed many technical fields, including MMSP. This is in part due to the ability of deep neural networks to effectively model relationships in multimodal data [25].

Among the emerging applications that were greatly facilitated by deep learning is autonomous driving, where multiple sensors—cameras, lidar, radar, microphones—collect information from the vehicle's surroundings to help it navigate the road. Processing signals from multiple modalities has traditionally been challenging. However, with the help of deep models, one can learn the complex relationships between different modalities from data, to enable their joint processing and analysis. In 2018, Waymo launched the first autonomous taxi service in Phoenix, Arizona. Another artificial intelligence (AI)/deep learning-driven trend is that of "smart" sensors and devices, such as smart speakers and cameras, whose capabilities have gone beyond capture and low-level processing of signals toward understanding and interaction with their environment.

On the video coding front, a major milestone was the 2013 release of the high-efficiency video coding (HEVC) standard, also known as *H.265* or *MPEG-H Part 2*. Beside the usual 50% coding efficiency gain over the predecessor (H.264/AVC), it was targeted at higher resolutions and allowed for higher bit-depth, thus facilitating high dynamic range display. It is currently the second most-widely used video coding format, after H.264/AVC. Despite the high adoption of standard codecs in various industries, especially by hardware developers, the research community felt that there is a strong need for royalty-free codecs. Hence, the Alliance for Open Media (AOM) was formed in 2015, with the goal of developing royalty-free video coding technology whose performance would be comparable to that of standard video codecs. Starting with Google's VP9

video codec, initially mainly used on YouTube, AOM released the AOMedia video 1 (AV1) video coding format in 2018. Royalty-free coding formats like VP9 and AV1 tend to be better supported in web browsers and streaming apps compared to standard coding formats.

Major revamping of the ICME conference took place during the 2010s. First, in 2010, ICME introduced double-blind review process, a departure from traditional single-blind review that is still common in signal processing. Moreover, the target acceptance rate was set to 30%, with the top 15% percent of papers being selected for oral presentation. Starting with 2012, ICME Workshops, which were introduced in 2009 to provide more focused satellite events and to foster new and emerging topics, have been published in separate proceedings. These innovations, and of course the hard work of many volunteers, helped the ICME become what it is today: a flagship IEEE conference in multimedia.

> As the customers' home Internet service speeds improved, the popularity of streaming services grew, and streaming now accounts for more viewing time than cable TV in the U.S. market.

Besides the ICME, the MMSP TC has also been organizing MMSP Workshops, which took place in Saint Malo, France (2010); Hangzhou, China (2011); Banff, AB, Canada (2012); Pula, Italy (2013); Jakarta, Indonesia (2014); Xiamen, China (2015); Montréal, QC, Canada (2016); Luton, United Kingdom (2017); Vancouver, BC, Canada (2018); and Kuala Lumpur, Malaysia (2019). ICME conferences were held in Singapore (2010); Barcelona, Spain (2011); Melbourne, VIC, Australia (2012); San Jose, California, USA (2013); Chengdu, China (2014); Turin, Italy (2015); Seattle, Washington, USA (2016); Hong Kong (2017); San Diego, California, USA (2018); and Shanghai, China (2019). During this decade, the MMSP TC was chaired by Philip Chou (2010–2011), Oscar Au (2012–2013), Dinei Florencio (2014–2015), Enrico Magli (2016–2017), and Frédéric Dufaux (2018–2019).

The 2020s

The current decade started with an event that impacted the world in many ways: the COVID-19 pandemic. As people retreated to their homes and started working remotely, the importance of multimedia suddenly grew. The demand for streaming services spiked, and videoconferencing became the norm for business meetings and presentations, education, and simply socializing and keeping in touch with friends and family. Before the pandemic, multimedia technology was mostly driven by entertainment. Now, it has become part of the infrastructure of our society. Even as the pandemic-related restrictions get removed, the concepts of remote work and collaboration are staying.

Although the decade is still young, several important technological milestones have already occurred. The latest video coding standard, versatile video coding (VVC) [26], also known as *H.266* or *MPEG-I Part 3*, was released in 2020. Besides the usual improvement in compression efficiency over its predecessor, VVC was developed to support a broad set

of resolutions, up to 16 K, a variety of color formats, as well as 360° video. Another important standard released in 2020 was MPEG point cloud compression (PCC). Targeting applications like augmented, virtual, and mixed reality, MPEG PCC provides compression technology for video-based and geometry-based PCC [27]. A related standard that is still being developed is JPEG Pleno [28], whose goal is to provide compression support for plenoptic imaging modalities, such as light fields, holography, and point clouds.

As noted earlier, learning-based technologies are playing an increasingly important role in many areas, including MMSP. But the benefit is mutual. In 2021, MPEG released a standard for neural network compression (MPEG-7 Part 17), whose purpose is to enable compression of neural network models for efficient storage and transport. While its purpose is to compress networks rather than multimedia signals, the standard was built upon the knowledge base developed over the years in image and video compression. Neural network compression is useful in federated learning, where model weights need to be transmitted between the clients and the server during the network training process.

Broader technological trends, such as the deployment of 5G communication systems, the growing Internet of Things, and advances in AI, are opening up possibilities for "smart" homes, buildings, factories, and cities. In applications like these, automation is a necessity, since the amount of data captured and communicated is far too much for humans to take note of. For example, most of the video captured by surveillance cameras will never be seen by humans, only "seen" by machines. As a result, several standardization efforts have been initiated to create media compression formats suitable for machine use, or combined human and machine use. One of these is JPEG AI [29], whose goal is to develop learning-based compression technology that supports conventional image decoding as well as a number of image processing and machine vision tasks. The other is MPEG video coding for machines [30], which targets both machine-only and human–machine tasks. Completion of these standards will facilitate improved efficiency of many technologies already in use, such as video monitoring, autonomous navigation, and multimedia database management, and create fertile ground for new and yet-to-be-imagined applications.

Due to the pandemic, MMSP TC activities in the 2020s have mostly been virtual so far. MMSP workshops took place virtually in Tampere, Finland in 2020, again in Tampere, Finland, as a hybrid event in 2021, and virtually in Shanghai, China, in 2022. ICME was a virtual event in London, United Kingdom, in 2020, and in Shenzhen, China, in 2021, and was organized as a hybrid event in Taipei, Taiwan, in 2022. During this period, MMSP TC chairs were Marta Mrak (2020–2021) and Ivan Bajić (2022–2023).

Conclusions

Starting as a mostly entertainment-driven technology, MMSP has come a long way to become a part of the very fabric of our society. It has enabled highly successful businesses, provided critical infrastructure at the time of need, and reached virtually everyone in some form or another. The MMSP TC has been a part of that story over the last two and a half decades.

So what does the future of MMSP look like? As the saying goes, "making predictions is difficult, especially about the future." In the near term, the trends are clear: data-driven approaches in the form of AI/deep learning are pushing the boundaries of what is possible with multimedia signals, and laying the foundation for the next generation of multimedia applications, products, and services. Beyond that, who knows: perhaps quantum multimedia?

> On the technical side, the immense amount of user-supplied content ushered in the era of Big Data and set the stage for further technical developments in the coming decades.

Authors

Ivan V. Bajić (ibajic@ensc.sfu.ca) received his Ph.D. degree in electrical engineering from Rensselaer Polytechnic Institute. He is a professor of engineering Science at Simon Fraser University, Burnaby, BC V5A 1S6, Canada, and the current Chair of the Multimedia Signal Processing (MMSP) Technical Committee. He was an associate editor of the *IEEE Transactions on Multimedia* and served on the organizing and program committees of the main conferences in the field, having won several service awards in these roles. His group's research has received awards at IEEE International Conference on Multimedia and Expo 2012, IEEE International Conference on Image Processing 2019, and IEEE MMSP 2022. He is a Senior Member of IEEE.

Marta Mrak (marta@ieee.org) received her Dipl. Ing. and M.Sc. electrical engineering degrees from the University of Zagreb, Croatia; and her Ph.D. degree from Queen Mary University of London, U.K. She is a senior AI research engineer at Helsing, W1T 3BL London, U.K. She has participated in the work of the Multimedia Signal Processing Technical Committee (MMSP TC) since 2014 and was MMSP TC chair from 2020–2021. During that time, she was a lead engineer at BBC R&D, where she ran various projects, ranging from video compression fundamentals to new content experiences powered by machine learning. Within the TC, her main contributions included serving as a general chair of IEEE International Conference on Multimedia and Expo (ICME) 2020 and lead Technical Program Committee chair for IEEE ICME 2019. She is a Senior Member of IEEE.

Frédéric Dufaux (frederic.dufaux@centralesupelec.fr) received his M.Sc. degree in physics and Ph.D. degree in electrical engineering from École Polytechnique Fédérale de Lausanne in 1990 and 1994, respectively. He is a

CNRS research director at Université Paris-Saclay, CNRS, CentraleSupélec, 91190 Gif-sur-Yvette, France. He was vice general chair of ICIP 2014, general chair of Multimedia Signal Processing (MMSP) 2018, and technical program co-chair of ICIP 2019 and ICIP 2021. He served as chair of the IEEE Signal Processing Society MMSP Technical Committee from 2018–2019. He is chair of the International Conference on Multimedia and Expo Steering Committee for 2022–2023. He was a founding member and the chair of the European Association for Signal Processing Technical Area Committee on Visual Information Processing from 2015 to 2021. He is a Fellow of IEEE.

Enrico Magli (enrico.magli@polito.it) received his Ph.D. degree from the Politecnico di Torino, Italy, in 2001. He is a professor with the Politecnico di Torino, 10129 Torino, Italy. He is a senior associate editor of *IEEE Journal on Selected Topics in Signal Processing*. He is a Fellow of the European Lab for Learning and Intelligent Systems Society for the advancement of artificial intelligence in Europe, and has been an IEEE distinguished lecturer from 2015 to 2016. He was the recipient of the IEEE Geoscience and Remote Sensing Society 2011 Transactions Prize Paper Award, the IEEE ICIP 2015 Best Student Paper Award (as senior author), the IEEE ICIP 2019 Best Paper Award, and the IEEE Multimedia 2019 Best Paper Award. He is a Fellow of the IEEE.

Tsuhan Chen (dprtchen@nus.edu.sg) is the deputy president for research and technology and distinguished professor at National University of Singapore, Singapore 119077. He also serves as the Chief Scientist of AI Singapore, a national program in artificial intelligence. He founded the Technical Committee on Multimedia Signal Processing in the IEEE Signal Processing Society, which later evolved into founding of the *IEEE Transactions on Multimedia* and the IEEE International Conference on Multimedia and Expo, joining efforts from multiple IEEE societies. He was appointed the editor-in-chief for *IEEE Transactions on Multimedia* from 2002 to 2004. He is a Fellow of IEEE.

References

[1] "Compact disc." Wikipedia. Accessed: Jun. 24, 2022. [Online]. Available: https://en.wikipedia.org/wiki/Compact_disc

[2] *H.120: Codecs for Videoconferencing Using Primary Digital Group Transmission*, ITU-T H.120, International Telecommunications Union, Geneva, Switzerland, 1984.

[3] *H.261: Video Codec for Audiovisual Services at p x 384 Kbit/s,* ITU-T H.261, International Telecommunications Union, Geneva, Switzerland, Nov. 1988. [Online]. Available: https://www.itu.int/rec/T-REC-H.261-198811-S/en

[4] D. H. Johnson, "Signal processing and the world wide web," *IEEE Signal Process. Mag.*, vol. 12, no. 5, pp. 53–57, Sep. 1995, doi: 10.1109/79.410440.

[5] *Information Technology – Digital Compression and Coding of Continuous-Tone Still Images – Requirements and Guidelines*, ITU-T T.81, International Organization for Standardization, Geneva, Switzerland, Sep. 1992.

[6] T. Sikora, "MPEG digital video-coding standards," *IEEE Signal Process. Mag.*, vol. 14, no. 5, pp. 82–100, Sep. 1997, doi: 10.1109/79.618010.

[7] P. Noll, "MPEG digital audio coding," *IEEE Signal Process. Mag.*, vol. 14, no. 5, pp. 59–81, Sep. 1997, doi: 10.1109/79.618009.

[8] J. Markoff, "Cult film is a first on internet," *NY Times*, May 1993. [Online]. Available: https://www.nytimes.com/1993/05/24/business/cult-film-is-a-first-on-internet.html

[9] K. Savetz, N. Randall, and Y. Lepage, *MBONE: Multicasting Tomorrow's Internet*. New York, NY, USA: Wiley, 1996.

[10] "Streaming media." Wikipedia. Accessed: Jul. 1, 2022. [Online]. Available: https://en.wikipedia.org/wiki/Streaming_media

[11] N. Strauss, "Rolling stones live on internet: Both a big deal and a little deal," *NY Times*, Nov. 1994. [Online]. Available: https://www.nytimes.com/1994/11/22/arts/rolling-stones-live-on-internet-both-a-big-deal-and-a-little-deal.html

[12] D. Rayburn. "The early history of the streaming media industry and the battle between Microsoft and Realnetworks." Seeking Alpha. Accessed: Jul. 3, 2022. [Online]. Available: https://seekingalpha.com/article/3957046-early-history-of-streaming-media-industry-and-battle-microsoft-and-realnetworks

[13] R. Stern, "Napster: A walking copyright infringement?" *IEEE Micro*, vol. 20, no. 6, pp. 4–5, Nov./Dec. 2000, doi: 10.1109/40.888696.

[14] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The JPEG 2000 still image compression standard," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 36–58, Sep. 2001, doi: 10.1109/79.952804.

[15] O. Rioul and M. Vetterli, "Wavelets and signal processing," *IEEE Signal Process. Mag.*, vol. 8, no. 4, pp. 14–38, Oct. 1991, doi: 10.1109/79.91217.

[16] A. Kiely and M. Klimesh, "The ICER progressive wavelet image compressor," Jet Propulsion Lab., Nat. Aeronaut. Space Admin., Washington, DC, USA, IPN Prog. Rep. 42-155, Nov. 2003. Accessed: Oct. 1, 2022. [Online]. Available: https://ipnpr.jpl.nasa.gov/progress_report/42-155/155J.pdf

[17] T. Wiegand and G. J. Sullivan, "The H.264/AVC video coding standard [Standards in a Nutshell]," *IEEE Signal Process. Mag.*, vol. 24, no. 2, pp. 148–153, Mar. 2007, doi: 10.1109/MSP.2007.323282.

[18] H. Schwarz and M. Wien, "The scalable video coding extension of the H.264/AVC standard [Standards in a Nutshell]," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 135–141, Mar. 2008, doi: 10.1109/MSP.2007.914712.

[19] B. Girod et al., "Mobile visual search," *IEEE Signal Process. Mag.*, vol. 28, no. 4, pp. 61–76, Jul. 2011, doi: 10.1109/MSP.2011.940881.

[20] *"Information technology - Multimedia content description interface - Part 13: Compact descriptors for visual search,"* ISO/IEC 15938-13:2015, Aug. 2015.

[21] A. G. Ivakhnenko, "Heuristic self-organization in problems of engineering cybernetics," *Automatica*, vol. 6, no. 2, pp. 207–219, Mar. 1970, doi: 10.1016/0005-1098(70)90092-0.

[22] R. Dechter, "Learning while searching in constraint-satisfaction-problems," in *Proc. AAAI Nat. Conf. Artif. Intell.*, Aug. 1986, pp. 178–183.

[23] G. Hinton et al., "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012, doi: 10.1109/MSP.2012.2205597.

[24] O. Russakovsky et al., "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Apr. 2015, doi: 10.1007/s11263-015-0816-y.

[25] D. Ramachandram and G. W. Taylor, "Deep multimodal learning: A survey on recent advances and trends," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 96–108, Nov. 2017, doi: 10.1109/MSP.2017.2738401.

[26] B. Bross et al., "Overview of the versatile video coding (VVC) standard and its applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 10, pp. 3736–3764, Oct. 2021, doi: 10.1109/TCSVT.2021.3101953.

[27] S. Schwarz et al., "Emerging MPEG standards for point cloud compression," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 9, no. 1, pp. 133–148, Mar. 2019, doi: 10.1109/JETCAS.2018.2885981.

[28] P. Astola et al., "JPEG Pleno: Standardizing a coding framework and tools for Plenoptic imaging modalities," *ITU J., ICT Discoveries*, vol. 3, no. 1, Jun. 2020, Art. no. 10.

[29] *"White paper on JPEG AI scope and framework v1.0,"* ISO/IEC JTC 1/SC29/WG1 N90049, 2021.

[30] *"Use cases and requirements for video coding for machines,"* ISO/IEC JTC 1/SC29/WG2 N00190, Apr. 2022.

SP