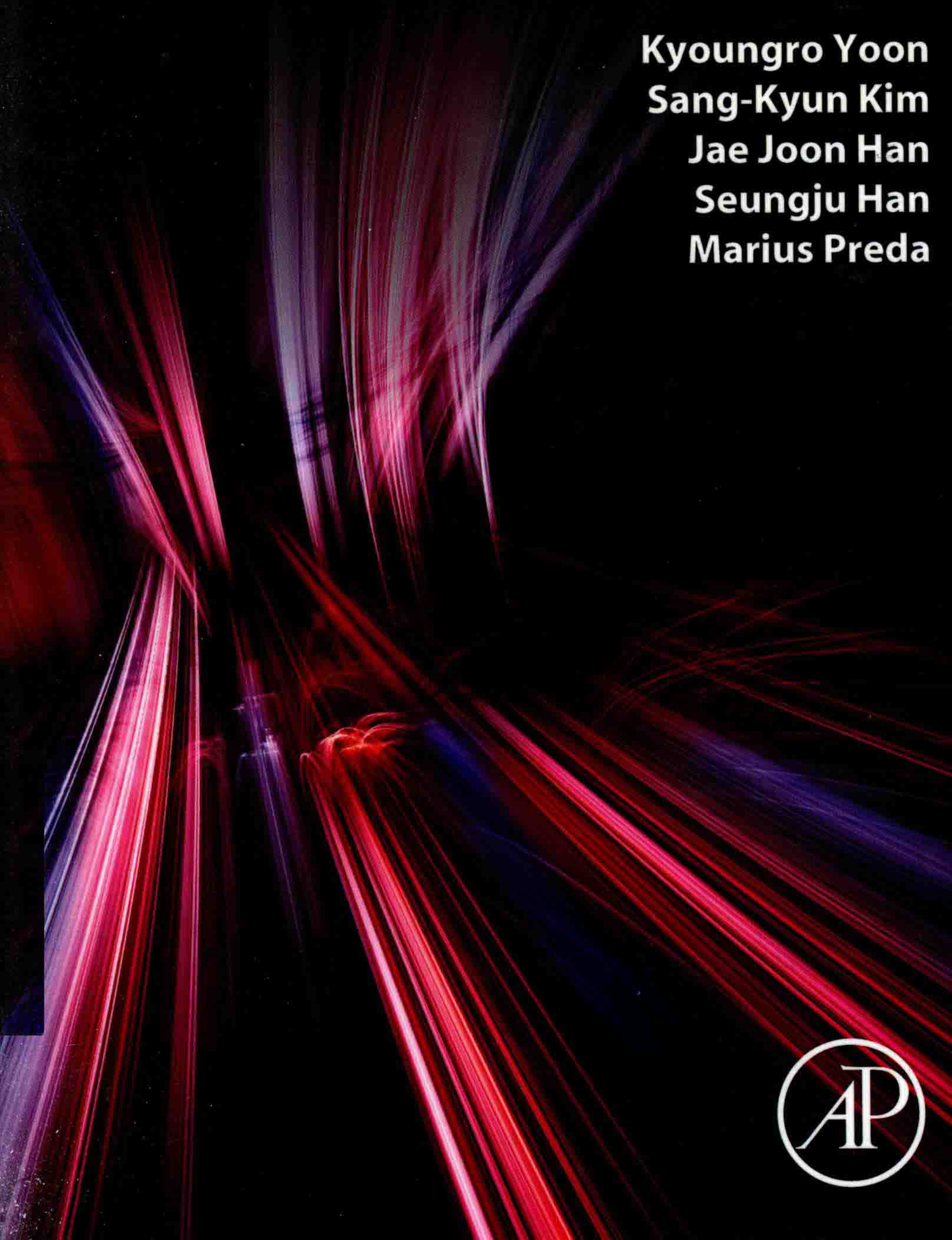
MPEG-V

Bridging the Virtual and Real World

Kyoungro Yoon
Sang-Kyun Kim
Jae Joon Han
Seungju Han
Marius Preda



MPEG-V

BRIDGING THE VIRTUAL AND REAL WORLD

KYOUNGRO YOON

SANG-KYUN KIM

JAE JOON HAN

SEUNGJU HAN

MARIUS PREDA

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier
125 London Wall, London EC2Y 5AS, UK
525 B Street, Suite 1800, San Diego, CA 92101-4495, USA
225 Wyman Street, Waltham, MA 02451, USA
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

ISBN: 978-0-12-420140-8

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress.

For Information on all Academic Press publications
visit our website at http://store.elsevier.com/

Typeset by MPS Limited, Chennai, India
www.adi-mps.com

Printed and bound in the United States

Publisher: Todd Green
Acquisition Editor: Tim Pitts
Editorial Project Manager: Charlie Kent
Production Project Manager: Jason Mitchell
Designer: Matthew Limbert

MPEG-V

ACKNOWLEDGMENT

This book would not be possible without the hard work of all the MPEG-V contributors that, meeting after meeting, during 3 years, built a consistent architecture supporting multi-sensorial user experiences, bringing innovative ideas and giving them a shape in terms of standard specifications. We would like to express our honor and satisfaction for working in such a challenging environment. Naming all the MPEG-V contributors would require a few pages and, probably, would not be complete, however, we would like to express special thanks to Jean Gelissen from Philipps and Sanghyun Joo from ETRI, the original initiators of the project and to Leonardo Chiariglione from Cedeo for the significant help in positioning MPEG-V in the MPEG eco-system.

AUTHOR BIOGRAPHIES

Kyoungro Yoon is a professor in School of Computer Science and Engineering at Konkuk University, Seoul, Korea. He received the BS degree in electronic and computer engineering from Yonsei University, Korea, in 1987, the MSE degree in electrical and computer engineering from University of Michigan, Ann Arbor, in 1989, and the PhD degree in computer and information science in 1999 from Syracuse University, USA. From 1999 to 2003, he was a Chief Research Engineer and Group Leader in charge of development of various product-related technologies and standards in the field of image and audio processing at the LG Electronics Institute of Technology. Since 2003, he joined Konkuk University as an assistant professor and has been a professor since 2012. He actively participated in the development of standards such as MPEG-7, MPEG-21, MPEG-V, JPSearch, and TV-Anytime and served as a co-chair for Ad Hoc Groups on User Preferences, chair for Ad Hoc Group on MPEG Query Format. He is currently serving as the chair for Ad Hoc Group on MPEG-V, the chair for Ad Hoc Group on JPSearch, and the chair for the Metadata Subgroup of ISO/IEC JTC1 SC29 WG1 (a.k.a. JPEG). He also served as an editor of various international standards such as ISO/IEC 15938-12, ISO/IEC 23005-2/5/6, and ISO/IEC 24800-2/5. He has co-authored over 40 conference and journal publications in the field of multimedia information systems. He is also an inventor/co-inventor of more than 30 US patents and 70 Korean patents.

Sang-Kyun Kim received the BS, MS, and PhD degrees in computer science from University of Iowa in 1991, 1994, and 1997, respectively. In 1997, he joined the Samsung Advanced Institute of Technology as a researcher. He was a senior researcher as well as a project leader on the Image and Video Content Search Team of the Computing Technology Lab until 2007. Since 2007, he joined Myongji University as an assistant Professor and has been an associate Professor in the Department of Computer Engineering since 2011. His research interests include digital content (image, video, and music) analysis and management, image search and indexing, color adaptation, mulsemedia adaptation, sensors and actuators, VR, and media-centric-IoT. He actively participated in the multimedia standardization activities such as MPEG-7, MPEG-21, MPEG-A, MPEG-V, as a co-chair and a project editor. He serves currently as a project editor of MPEG-V International Standards, i.e. ISO/IEC 23005-2/3/4/5, and 23005-7. He has co-authored over 40 conference and journal publications in the field of digital content management and mulsemedia simulation and adaptation. He is also an inventor/co-inventor of more than 25 US patents and 90 Korean patents.

Jae Joon Han has been a principal researcher at Samsung Advanced Institute of Technology (SAIT) in Samsung Electronics, Korea since 2007. He received the BS degree in electronic engineering from Yonsei University, Korea, in 1997, the MS degree in electrical and computer engineering from the University of Southern California, Los Angeles, in 2001, and the PhD degree in electrical and computer engineering from Purdue University, West Lafayette, IN, in August 2006. Since receiving the PhD degree, he was at Purdue as a Postdoctoral Fellow in 2007. His research interests include statistical machine learning and data mining, computer vision, and real-time recognition technologies. He participated in the development of standards such as ISO/IEC 23005 (MPEG-V) and ISO/IEC 23007 (MPEG-U), and served as the editor of ISO/IEC 23005-1/4/6. He has co-authored over 20 conference and journal publications. He is also an inventor/co-inventor of three US patents and 70 filed international patent applications.

Seungju Han is currently a senior researcher at Samsung Advanced Institute of Technology (SAIT) in Samsung Electronics, Korea. He received the PhD degree in electrical and computer engineering in 2007, from the University of Florida, USA. Since 2007, he has joined Samsung Advanced Institute of Technology as a research engineer. He participated in the development of standards such as ISO/IEC 23005 (MPEG-V) and ISO/IEC 23007 (MPEG-U), and served as the editor of ISO/IEC 23005-2/5. He has authored and co-authored over 25 research papers in the field of pattern recognition and human–computer interaction. He is also an inventor/co-inventor of four US patents and 70 filed international patent applications.

Marius Preda is an associate professor at Institut MINES-Telecom and Chairman of the 3D Graphics group of ISO's MPEG (Moving Picture Expert Group). He contributes to various ISO standards with technologies in the fields of 3D graphics, virtual worlds, and augmented reality and has received several ISO Certifications of Appreciation. He leads a research team with a focus on Augmented Reality, Cloud Computing, Games and Interactive Media and regularly presents results in journals and at speaking engagements worldwide. He serves on the program committee international conferences and reviews top-level research journals.

After being part of various research groups and networks, in 2010 he founded a research team within Institut MINES-Telecom, called GRIN – GRaphics and INteractive media. The team is conducting research at the international level cooperating with academic partners worldwide and industrial ICT leaders. Selected results are showcased on www.MyMultimediaWorld.com.

Academically, Marius received a degree in Engineering from Politehnica Bucharest, a PhD in Mathematics and Informatics from University Paris V and an eMBA from Telecom Business School, Paris.

PREFACE

Traditional multimedia content is typically consumed via audio-visual (AV) devices like displays and speakers. Recent advances in 3D video and spatial audio allow for a deeper user immersion into the digital AV content, and thus a richer user experience. The norm, however, is that just two of our five senses – sight and hearing – are exercised, while the other three (touch, smell, and taste) are neglected.

The recent multitude of new sensors map the data they capture onto our five senses and enable us to better perceive the environment both locally and remotely. In the literature, the former is referred to as "Augmented Reality", and the latter as "Immersive Experience". In parallel, new types of actuators produce different kinds of multi-sensory effect. In early periods such effects were mostly used in dedicated installations in attraction parks equipped with motion chairs, lighting sources, liquid sprays, etc., but it is more and more to see multi-sensory effects produced in more familiar environments such as at home.

Recognizing the need to represent, compress, and transmit this kind of contextual data captured by sensors, and of synthesizing effects that stimulate all human senses in a holistic fashion, the Moving Picture Experts Group (MPEG, formally ISO/IEC JTC 1/SC 29/WG 11) ratified in 2011 the first version of the MPEG-V standard (officially known as "ISO/IEC 23005 – Media context and control"). MPEG-V provides the architecture and specifies the associated information representations that enable interoperable multimedia and multimodal communication within Virtual Worlds (VWs) but also with the real world, paving the way to a "Metaverse", i.e. an online shared space created by the convergence of virtually enhanced reality and physically persistent virtual space that include the sum of all Virtual Worlds and Augmented Realities. For example, MPEG-V may be used to provide multi-sensorial content associated to traditional AV data enriching multimedia presentations with sensory effects created by lights, winds, sprays, tactile sensations, scents, etc.; or it may be used to interact with a multimedia scene by using more advanced interaction paradigms such as hand/body gestures; or to access different VWs with an avatar with a similar appearance in all of them.

In the MPEG-V vision, a piece of digital content is not limited to an AV asset, but may be a collection of multimedia and multimodal objects forming a scene, having their own behaviour, capturing their context, producing effects in the real world, interacting with one or several users, etc. In other words, a digital item can be as complex as an entire VW. Since a standardizing VW representation is technically possible but not aligned with industry interests, MPEG-V offers interoperability between VWs (and between any of them and the real world) by describing virtual objects, and specifically avatars, so that they can "move" from one VW to another.

This book on MPEG-V draws a global picture of the features made possible by the MPEG-V standard, and is divided into seven chapters, covering all aspects from the global architecture, to technical details of key components – sensors, actuators, multi-sensorial effects – and to application examples.

At the time this text was written (November 2014), three editions of MPEG-V have been published and the technical community developing the standard is still very active. As the main MPEG-V philosophy is not expected to change in future editions, this book is a good starting point to understand the principles that were at the basis of the standard. Readers interested in the latest technical details can see the MPEG-V Web-site (http://wg11.sc29.org/mpeg-v/).

Marius Preda
Leonardo Chiariglione

CONTENTS

CHAPTER 1

Introduction to MPEG-V Standards

Contents

1.1 INTRODUCTION TO VIRTUAL WORLDS

The concept of a virtual world has become a part of our everyday lives so recently that we have not even noticed the change. There have been various attempts at defining a virtual world, each with its own point of view. The worlds that we are currently experiencing, from the viewpoint of information technology, can be divided into three types: the real world, virtual worlds, and mixed worlds. Conventionally, a virtual world, also referred to frequently as virtual reality (VR), is a computer-generated environment, giving the participants the impression that the participants are present within that environment [1]. According to Milgram and Kishino [1], real objects are those having actual existence that can be observed directly or can be sampled and resynthesized for viewing, whereas virtual objects are those that exist in essence or effect, but not formally or actually, and must be simulated.

Recently, Gelissen and Sivan [2] redefined a virtual world as an integration of 3D, Community, Creation, and Commerce (3D3C). Here, 3D indicates a 3D visualization and navigation for the representation of a virtual world, and 3C represents the three key factors that make a virtual

DOI: http://dx.doi.org/10.1016/B978-0-12-420140-8.00001-9

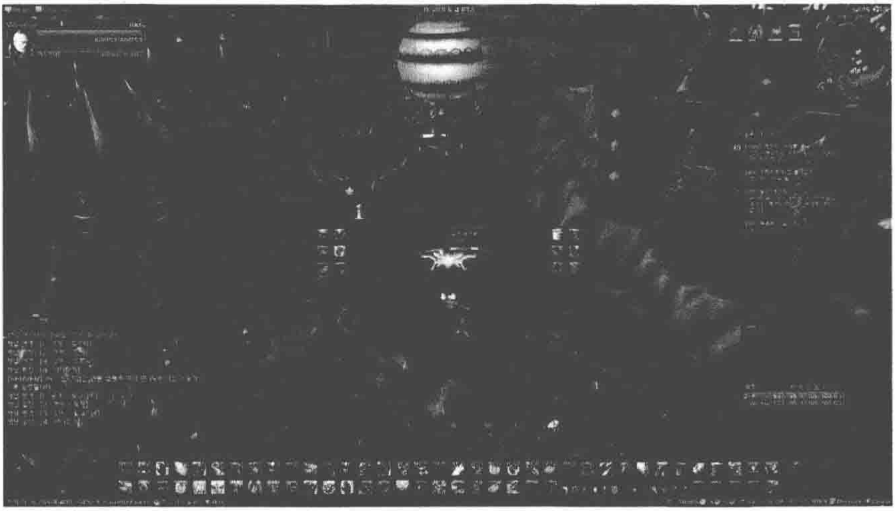


Figure 1.1 A virtual gaming world (from *World of Warcraft*).

world closer to the real world, which can be characterized by daily interactions for either economic (creation and commerce) or noneconomic/cultural (community) purposes.

Virtual worlds can also be divided into gaming and nongaming worlds. A virtual gaming world is a virtual world in which the behavior of the avatar (user) is goal-driven. The goal of a particular game is given within its design. *Lineage* [3] and *World of Warcraft* [4] are examples of virtual gaming worlds. Figure 1.1 shows a screen capture from *World of Warcraft*. In contrast, a nongaming virtual world is a virtual world in which the behavior of the avatar (user) is not goal-driven. In a nongaming virtual world, there is no goal provided by the designer, and the behavior of the avatar depends on the user's own intention. An example of a nongaming virtual world is *Second Life* by Linden Lab, a captured image of which is shown in Figure 1.2 [5].

A virtual world can provide an environment for both collaboration and entertainment [6]. Collaboration can mainly be enabled by the features of the virtual world, such as the 3D virtual environments in which the presence, realism, and interactivity can be supported at a higher degree than in conventional collaboration technology, and avatar-based interactions through which the social presence of the participants and the self-presentation can be provided at a higher degree than in any other existing environment.

Figure 1.2 A nongaming virtual world (from *Second Life*).

1.2 ADVANCES IN MULTIPLE SENSORIAL MEDIA

1.2.1 Basic Studies on Multiple Sensorial Media

Along with the sensations associated with 3D films and UHD display panels, the development of Multiple Sensorial Media (MulSeMedia), or 4D media, has received significant attention from the public. 4D content generally adds sensorial effects to 3D, UHD, and/or IMAX content, allowing audiences to immerse themselves more deeply into the content-viewing experience. Along with the two human senses of sight and hearing, sensorial effects such as wind, vibration, and scent can stimulate other senses, such as the tactile and olfaction senses. MulSeMedia content indicates audiovisual content annotated with sensory effect metadata [7].

The attempts to stimulate other senses while playing multimedia content have a long history. Sensorama [8,9] which was an immersive VR motorbike simulator, was a pioneer in MulSeMedia history. As a type of futuristic cinema, Sensorama rendered sensorial effects with nine different fans, a vibrating seat, and aromas to simulate a blowing wind, driving over gravel, and the scent of a flower garden or pizzeria. Although Sensorama was not successful in its day, its technology soon became a pioneer of current 4D theaters and the gaming industry.

The significance of olfactory or tactile cues has been reported in many previous studies [10–14]. Dinh et al. [10] reported that the addition of tactile, olfactory, and auditory cues into a VR environment increases the

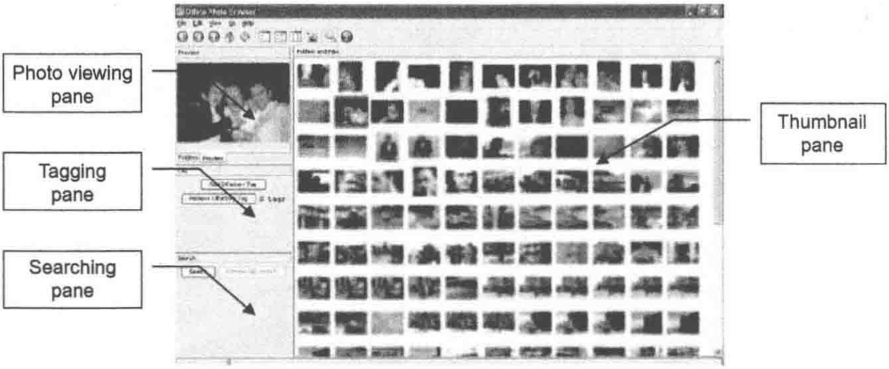


Figure 1.3 Search and retrieve based on odor [13].

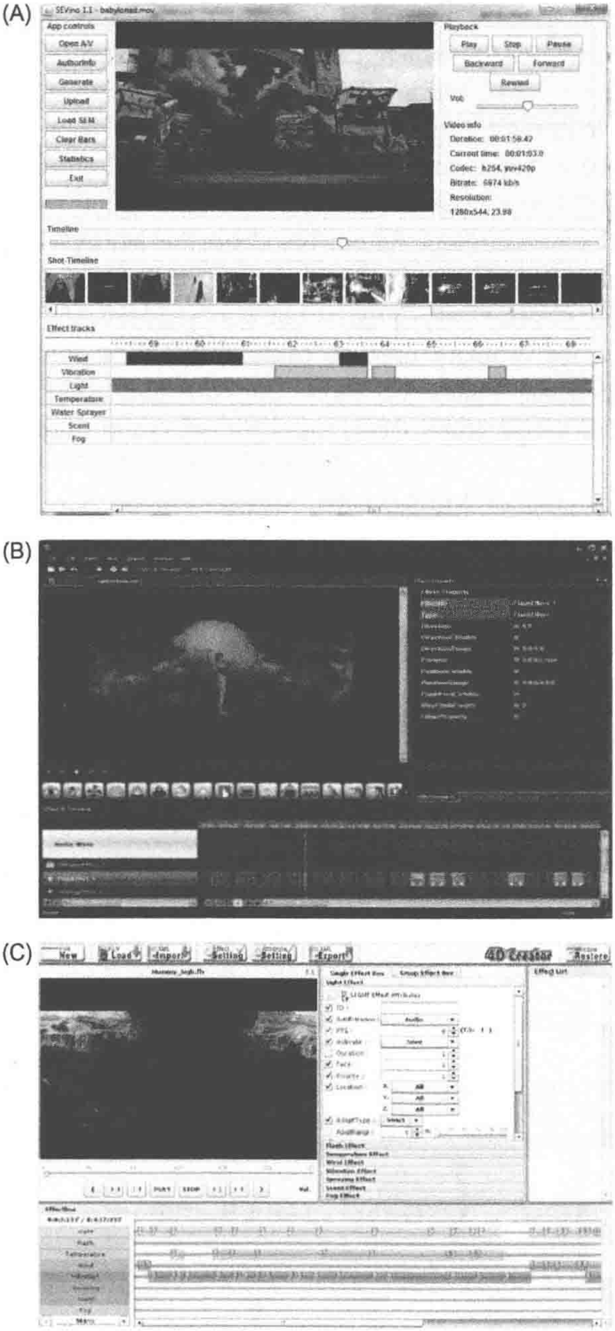
user's sense of presence and memory of the environment. Bodnar et al. [11] reported that the olfactory modality is less effective in alarming users than the other modalities such as vibration and sound, but can have a less disruptive effect on continuing the primary task of the users. Ryu and Kim [12] studied the effectiveness of vibro-tactile effects on the whole body to simulate collisions between users and their virtual environment.

Olfactory cues can be used to evoke human memories. Brewster et al. [13] presented a study on the use of smell for searching through digital photo collections, and compared text- and odor-based tagging (Figure 1.3). For the first stage, sets of odors and tag names from the user descriptions of different photos were generated. The participants then used these to tag their photos, returning two weeks later to answer questions regarding these images. The results showed that the performance when using odors was lower than that from simple text searching but that some of the participants had their memories of their photos evoked through the use of smell.

Ghinea and Ademoye [14] presented a few design guidelines for the integration of olfaction (with six odor categories) in multimedia applications. Finally, Kannan et al. [15] encompassed the significance of other senses incorporated in the creation of digital content for the packaging industry, healthcare systems, and educational learning models.

1.2.2 Authoring of MulSeMedia

The difficulties in producing MulSeMedia content mainly lie in the time and effort incurred by authoring the sensory effects. For the successful industrial deployment of MulSeMedia services, the provisioning of an easy and efficient means of producing MulSeMedia content plays a critical role. Figure 1.4 shows examples of the authoring tools used to create digital content with sensorial effects.



Duration: 00:01:58.42
Current time: 00:01:03.0
Codec: h264, yuv420p
Bitrate: 6974 kb/s
Resolution: 1280x544, 23.98

Figure 1.4 Authoring tools for sensorial effects: (A) SEVino by Waltl et al. [18,19], (B) RoSEStudio by Choi et al. [16], and (C) SMURF by Kim [17].