Multimedia Information Retrieval in XR

Rahel Arnold
University of Basel
Basel, Switzerland
rahel.arnold@unibas.ch

Werner Bailer
Joanneum Research
Graz, Austria
werner.bailer@joanneum.at

Ralph Gasser
University of Basel
Basel, Switzerland
ralph.gasser@unibas.ch

Björn Þ. Jónsson
Reykjavik University
Reykjavík, Iceland
bjorn@ru.is

Omar S. Khan
Reykjavik University
Reykjavík, Iceland
omark@ru.is

Heiko Schuldt
University of Basel
Basel, Switzerland
heiko.schuldt@unibas.ch

Florian Spiess
University of Basel
Basel, Switzerland
florian.spiess@unibas.ch

Lucia Vadicamo
Institute of Information Science and Technologies
National Research Council
Pisa, Italy
lucia.vadicamo@isti.cnr.it

Abstract

The way we create, consume and interact with multimedia content has changed significantly in recent years with the advent of affordable recording devices and easy sharing and access in the form of mobile phones. With the imminent wave of affordable devices that enable mixed reality experiences and the large variety of devices on the market, interaction with multimedia content is expected to continue to evolve rapidly. This will also drastically affect the entire area of multimedia information retrieval in eXtended Reality (XR), for instance by novel ways to express user needs in VR, result presentation that takes the specific capabilities of XR devices into account, and/or result feedback. This tutorial on *Multimedia Retrieval in XR* discusses and demonstrates existing solutions and highlights key challenges in this evolving field.

ACM Reference Format:
Rahel Arnold, Werner Bailer, Ralph Gasser, Björn Þ. Jónsson, Omar S. Khan, Heiko Schuldt, Florian Spiess, and Lucia Vadicamo. 2024. Multimedia Information Retrieval in XR. In *Proceedings of the 32nd ACM International Conference on Multimedia (MM '24), October 28-November 1, 2024, Melbourne, VIC, Australia.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3664647.3689176

1 Motivation

eXtended Reality (XR) is changing how we consume and interact with digital content, offering novel ways to communicate, learn, and engage across various domains, including education, entertainment, healthcare, manufacturing, and cultural heritage [2]. Such immersive experiences drive the need for managing and processing large-scale data sets in real-time, handling different types of data, and enhancing user experiences. These challenges can be subsumed under the need for effective multimedia retrieval in XR.

The idea of XR spaces as a natural way for information retrieval has been around since the 1990s. However, practical limitations of systems have long hindered their success. The new wave of interest in XR technologies, together with more affordable hardware and support with AI, open new possibilities that make it worth to investigate whether and how these shortcomings could be overcome.

This tutorial provides essential knowledge and skills for developing and applying Multimedia Retrieval in XR, addressing current challenges and opportunities, and equipping participants with practical examples they can directly apply to their work. It focuses on retrieval taking place in XR, and covers a wider range of technologies and challenges, in particular also MR search – in contrast to the more pronounced 3D reconstruction and 3D retrieval focus of [5].

2 Main Topics and Challenges

In what follows, we briefly summarize the most important topics in the intersection of multimedia information retrieval and XR and we discuss the major challenges in this domain.

Introduction to XR and Multimedia Information Retrieval. The first part of the tutorial offers a concise overview of XR, highlighting key applications across various domains and challenges XR poses to multimedia information retrieval. We outline fundamental principles of multimedia information retrieval, covering essential concepts such as content-based retrieval, feature extraction, indexing, and similarity search. Examples of methods that can be employed on multimedia content commonly used in XR environments, such as 3D models, egocentric videos, and 360° videos, will also be provided.

Retrieval in XR. Media access, retrieval and analysis in XR brings novel challenges that go beyond "traditional" multimedia retrieval. This includes search paradigms beyond keyword queries (such as, for instance, query-by-example, query-by-sketch, query-by-gesture), the presentation of and interaction with results in an immersive way, and the analysis of resulting content. The tutorial

MM '24, October 28-November 1, 2024, Melbourne, VIC, Australia

ACM ISBN 979-8-4007-0686-8/24/10
https://doi.org/10.1145/3664647.3689176

(a) Object detection (b) Manual area (c) Text input (d) Result grid

Figure 1: Smartphone Application View of (MR)^2

surveys existing approaches and systems, presents concrete examples that are also be made available at the venue as demonstrators to be tested by the participants of the tutorial. As the fast-evolving MR devices build the fundamentals for media search and analysis in MR, we present a short comparison and overview of the existing hardware and the functionality used for the research prototyped on these devices.

Fundamentals and new possibilities for *query generation* including live object detection and result presentation in the MR space are demonstrated on the basis of the prototype system (MR)^2 [1] (*M*ixed *R*eality *M*ultimedia *R*etrieval) (see Figure 1).

Collection Exploration in XR. Access to XR opens up new possibilities for exploring collections based on metadata, whether that metadata is entered by users or automatically generated, providing a foundation for collection analytics. We discuss the exploration-search axis from multimedia analytics and consider the impact of XR on the exploration side of the axis.

Fundamentals and new possibilities for *integration of exploration and search* will be discussed based on the prototype system ViRMA [3] (*Vi*rtual *R*eality *M*ultimedia *A*nalytics) (see Figure 2).

Evaluating XR search/exploration. We give a brief overview of evaluation paradigms and campaigns for interactive multimedia retrieval [8], and review the success of XR-based retrieval systems. We also discuss the specific challenges for evaluating XR-based systems [6].

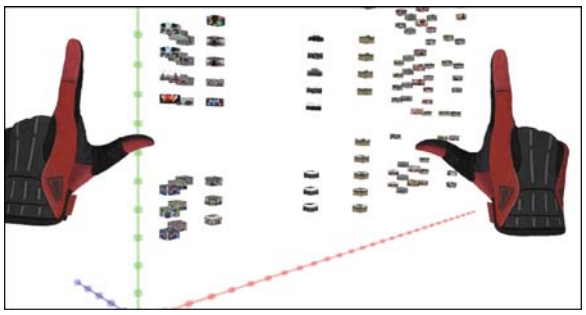
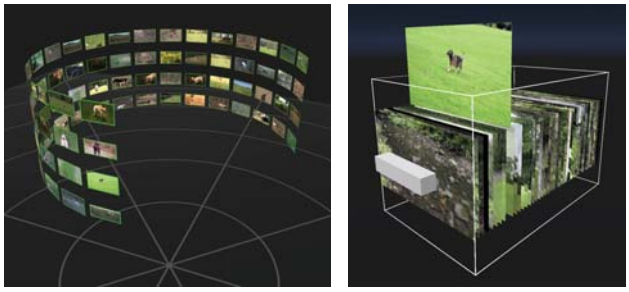


Figure 2: Collection Exploration in ViRMA (from [4]).



(a) Cylindrical grid results display (b) Multimedia drawer segment view

Figure 3: vitrivr-VR User Interface (from [7])

Research challenges. We outline open challenges and future research topics, aiming to motivate young researchers to address this field in their work.

Demonstrators. The tutorial showcases the vitrivr-VR [7] (see Figure 3), (MR)^2 and ViRMA systems; selected systems are also accessible to the participants for a hands-on experience.

Acknowledgements. This work was partly supported by the XR mEdia eCOsystem (XReco) project (Horizon Europe contract no. 101070250 and Swiss State Secretariat for Education, Research and Innovation (SERI) contract no. 22.00268), the Social and hUman ceNtered XR (SUN) project (Horizon Europe contract no. 101092612), and Icelandic Research Fund (IRF) grant 239772-051.

References

[1] Rahel Arnold and Heiko Schuldt. 2024. Multimedia Retrieval in Mixed Reality: Leveraging Live Queries for Immersive Experiences. In *2024 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR)*. 289–293. https://doi.org/10.1109/AIxVR59861.2024.00048

[2] Ralf Doerner, Wolfgang Broll, Paul Grimm, and Bernhard Jung (Eds.). 2022. *Virtual and Augmented Reality (VR/AR): Foundations and Methods of Extended Realities (XR)*. Springer International Publishing. https://doi.org/10.1007/978-3-030-79062-2

[3] Aaron Duane and Björn Þór Jónsson. 2022. ViRMA: Virtual Reality Multimedia Analytics. In *International Conference on Multimedia Retrieval (ICMR 2022)*. ACM, Newark, NJ, USA, 211–214. https://doi.org/10.1145/3512527.3531352

[4] Aaron Duane and Björn Þór Jónsson. 2022. ViRMA: Virtual Reality Multimedia Analytics at Video Browser Showdown 2022. In *MultiMedia Modeling, Proceedings Part II – 28th International Conference (MMM 2022) (Lecture Notes in Computer Science, Vol. 13142)*. Springer, Phu Quoc, Vietnam, 580–585. https://doi.org/10.1007/978-3-030-98355-0_58

[5] Maria Pegia, Sotiris Diplaris, Stefanos Vrochidis, Heiko Schuldt, Florian Spiess, Rahel Arnold, and Werner Bailer. 2024. Multimedia Retrieval in and for XR. In *Proceedings of the 2024 International Conference on Multimedia Retrieval, (ICMR 2024)*. ACM, Phuket, Thailand, 1324–1325. https://doi.org/10.1145/3652583.3658421

[6] Florian Spiess, Ralph Gasser, Silvan Heller, Heiko Schuldt, and Luca Rossetto. 2023. A Comparison of Video Browsing Performance between Desktop and Virtual Reality Interfaces. In *Proceedings of the 2023 ACM International Conference on Multimedia Retrieval (ICMR 2023)*. ACM, Thessaloniki, Greece, 535–539. https://doi.org/10.1145/3591106.3592292

[7] Florian Spiess, Luca Rossetto, and Heiko Schuldt. 2024. Exploring Multimedia Vector Spaces with vitrivr-VR. In *MultiMedia Modeling, Proceedings, Part IV – 30th International Conference (MMM 2024) (Lecture Notes in Computer Science, Vol. 14557)*. Springer, Amsterdam, The Netherlands, 317–323. https://doi.org/10.1007/978-3-031-53302-0_27

[8] Lucia Vadicamo, Rahel Arnold, Werner Bailer, Fabio Carrara, Cathal Gurrin, Nico Hezel, Xinghan Li, Jakub Lokoc, Sebastian Lubos, and Zhixinand others Ma. 2024. Evaluating Performance and Trends in Interactive Video Retrieval: Insights from the 12th VBS Competition. *IEEE Access* (2024). https://doi.org/10.1109/ACCESS.2024.3405638