



UvA-DARE (Digital Academic Repository)

From Content-Based Image and Video Retrieval With Relevance Feedback to Multimodal Large Language Models and Beyond: A Journey Through Modalities and Semantic Levels

Rudinac, Stevan

DOI

[10.1145/3746263.3764110](https://doi.org/10.1145/3746263.3764110)

Publication date

2025

Document Version

Final published version

Published in

MA-LLM '25

License

Article 25fa Dutch Copyright Act (<https://www.openaccess.nl/en/policies/open-access-in-dutch-copyright-law-taverne-amendment>)

[Link to publication](#)

Citation for published version (APA):

Rudinac, S. (2025). From Content-Based Image and Video Retrieval With Relevance Feedback to Multimodal Large Language Models and Beyond: A Journey Through Modalities and Semantic Levels. In *MA-LLM '25: Proceedings of the ACM MM 2025 Workshop on Multimedia Analytics with Multimodal Large Language Models : October, 27-31, 2025, Dublin, Ireland* (pp. 13). Association for Computing Machinery. <https://doi.org/10.1145/3746263.3764110>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)
You will be contacted as soon as possible.



From Content-Based Image and Video Retrieval With Relevance Feedback to Multimodal Large Language Models and Beyond: A Journey Through Modalities and Semantic Levels

Stevan Rudinac

s.rudinac@uva.nl

University of Amsterdam

Amsterdam, The Netherlands

Abstract

The multimedia community has made a long journey from early content-based image and video retrieval systems, which often facilitated interactive search, exploration, and learning using relevance feedback and similar techniques, to modern foundation models, such as multimodal large language models (MLLMs). During that period, multimedia analysis techniques have improved dramatically, maturing from the low semantic levels of colours, textures, and shapes, over semantic concepts, actions, and events to increasingly abstract and complex multimedia content understanding. With the advent of foundation models, the traditional search and recommendation paradigms have been increasingly replaced with multimodal multi-turn conversational search. In this keynote, I will reflect on this journey by showcasing representative works, including those by our team, advocating that the trends surrounding modern MLLMs have been a product of gradual evolution, rather than a revolution. In addition, I will illustrate how MLLMs can be applied to advance business and society by injecting domain knowledge from the other disciplines and facilitating improved interaction with the user. Finally, I will share some thoughts on the potential of recent research developments aiming to improve performance while significantly reducing the size, complexity and cost of MLLMs.

CCS Concepts

• **Information systems** → **Multimedia and multimodal retrieval; Users and interactive retrieval.**

Keywords

Multimedia Information Retrieval, Interactive Learning, Interactive Search and Exploration, Multimodal Conversational Search, Multimodal Large Language Models

ACM Reference Format:

Stevan Rudinac. 2025. From Content-Based Image and Video Retrieval With Relevance Feedback to Multimodal Large Language Models and Beyond: A Journey Through Modalities and Semantic Levels. In *Proceedings of the ACM MM 2025 Workshop on Multimedia Analytics with Multimodal Large Language Models (MA-LLM '25)*, October 27–31, 2025, Dublin, Ireland. ACM, New York, NY, USA, 1 page. <https://doi.org/10.1145/3746263.3764110>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

MA-LLM '25, Dublin, Ireland.

© 2025 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2045-1/2025/10

<https://doi.org/10.1145/3746263.3764110>

1 Biography

Stevan Rudinac is an associate professor of artificial intelligence for business with the University of Amsterdam Business School. He received the PhD degree in computer science from the Delft University of Technology, Delft, The Netherlands, in 2013. He has worked as a researcher with the University of Belgrade, Eindhoven University of Technology, and the Netherlands Forensic Institute. He conducts research on multimedia information retrieval, i.e., facilitating access to large multimedia collections and gaining insight from them, by jointly analyzing visual content and the heterogeneous information associated with it, ranging from text and automatically generated metadata to information about users and their social network. His research focuses on urban computing, business, and security applications. Stevan is a scientific director of Multimedia Information Retrieval for Business, an interfaculty lab between the University of Amsterdam (UvA) Faculty of Economics and Business and the Informatics Institute of the Faculty of Science, which is a part of broader UvA's initiative focusing on AI Technology for People. In addition, he is a Co-Principal Investigator in UvA Data Science Centre HAVA-Lab focusing on pioneering research in Human-Aligned Video AI. Over the years he has taken various organizational and community service roles, such as most recently Technical Program Chair of the 2023 International Conference on Multimedia Modeling (MMM'23), General Chair of MMM'24 and Technical Program Chair of the 33rd ACM International Conference on Multimedia (ACM Multimedia 2025).

