


Chapter 2

Content-Based Video Retrieval With Temporal Localization Using a Deep Bimodal Fusion Approach

G. Megala

 <https://orcid.org/0000-0002-8084-8292>
Vellore Institute of Technology, India


P. Swarnalatha

Vellore Institute of Technology, India

S. Prabu

Pondicherry University, India

R. Venkatesan

 <https://orcid.org/0000-0002-4336-8628>
SASTRA University, India

Anantharajah Kaneswaran

University of Jaffna, Sri Lanka

ABSTRACT

Content-based video retrieval is a research field that aims to develop advanced techniques for automatically analyzing and retrieving video content. This process involves identifying and localizing specific moments in a video and retrieving videos with similar content. Deep bimodal fusion (DBF) is proposed that uses modified convolution neural networks (CNNs) to achieve considerable visual modality. This deep bimodal fusion approach relies on the integration of information from both visual and audio modalities. By combining information from both modalities, a more accurate model is developed for analyzing and retrieving video content. The main objective of this research is to improve the efficiency and effectiveness of video retrieval systems. By accurately identifying and localizing specific moments in videos, the proposed method has higher precision, recall, F1-score, and accuracy in precise searching that retrieves relevant videos more quickly and effectively.

DOI: 10.4018/978-1-6684-8098-4.ch002

INTRODUCTION

Multimedia information systems are becoming more crucial due to the growth of internet access, big data, and high-speed networks as well as the increasing need for multimedia information with visualization. Multimedia data, however, needs a lot of processing (Megala et al., 2021) and storage (Megala & Swarnalatha, 2022). Therefore, there is a requirement for effective extraction, archiving, indexing, and retrieval of video content from a huge multimedia database. The video has emerged as one of the most prevalent methods to share information because it is visual and powerful. Many people around the world have easy access to it. Media administrators find it hard to use video material for storage and search. Prominent web browsers today often skip searches that are heavy on content in facilitate subtitles that contain basic information regarding the videos being searched. As an alternative to traditional techniques of keyword search, users on online platforms desire to look up precise videos in almost real-time.

Video moment localization and content-based video retrieval using deep bimodal fusion is an emerging research field that aims to develop advanced techniques for analyzing and retrieving video content. With the proliferation of digital video content, the need for efficient and effective video retrieval systems has become increasingly important in a wide range of applications, including entertainment, education, and surveillance.

The process of video moment localization involves identifying and localizing specific moments in a video, such as a particular scene or event. This can be a challenging task, as videos can contain a wide range of visual and auditory information, making it difficult to accurately identify specific moments of interest.

Content-based video retrieval, on the other hand, involves retrieving videos that contain similar content to a given query video. Objects that occurred in the video or images are identified and are saved as a bag of visual features. Efficient object detection methods (Megala & Swarnalatha, 2023) are used to perform depth prediction along spatial and temporal features. These bag of features are more helpful in the retrieval process. This process requires the development of accurate models for analyzing video content and identifying similarities between videos.

To address these challenges, researchers have turned to deep learning techniques, particularly deep bimodal fusion. This approach involves integrating information from both visual and audio modalities to develop more accurate models for analyzing and retrieving video content. By combining information from both modalities, researchers can develop more robust and accurate models for identifying specific moments in videos and retrieving relevant videos based on content.

In this work, we describe a deep bimodal fusion (DBF) method for recognizing a person's obvious personality from movies, which addresses this issue and yields better results than previously published research. The DBF framework's structure is depicted in Figure 1.

Overall, the use of deep bimodal fusion techniques in video moment localization and content-based video retrieval has the potential to revolutionize the way we analyze and retrieve video content, making it easier and faster to find relevant videos in a wide range of applications.

The structure of this chapter is as follows: related works on video retrieval followed by the proposed method, experimental analysis, and conclusion.

9 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/content-based-video-retrieval-with-temporal-localization-using-a-deep-bimodal-fusion-approach/325933

Related Content

A Based-Rule Method to Transform CIM to PIM into MDA

Yassine Rhazali, Youssef Hadiand Abdelaziz Mouloudi (2016). *International Journal of Cloud Applications and Computing* (pp. 11-24).

www.irma-international.org/article/a-based-rule-method-to-transform-cim-to-pim-into-mda/159848

Sustaining Adaptation to the New Normal: Tenacity and Originality in Higher Education Marketing After the COVID-19 Pandemic

Redouan Ainous (2024). *Driving Transformative Technology Trends With Cloud Computing* (pp. 342-357).

www.irma-international.org/chapter/sustaining-adaptation-to-the-new-normal/353364

Cloud-Based Computing Architectures for Solving Hot Issues in Structural Bioinformatics

Dariusz Mrozek (2015). *Emerging Research in Cloud Distributed Computing Systems* (pp. 294-314).

www.irma-international.org/chapter/cloud-based-computing-architectures-for-solving-hot-issues-in-structural-bioinformatics/130278

Predictive Modeling for Imbalanced Big Data in SAS Enterprise Miner and R

Son Nguyen, Alan Olinsky, John Quinnand Phyllis Schumacher (2018). *International Journal of Fog Computing* (pp. 83-108).

www.irma-international.org/article/predictive-modeling-for-imbalanced-big-data-in-sas-enterprise-miner-and-r/210567

A Comparative Study of Cloud GPU Offerings for AI/ML Engineers

Sanjay P. Ahuja, Madhuri Golanakondaand Sandeep Reddivari (2026). *International Journal of Cloud Applications and Computing* (pp. 1-15).

www.irma-international.org/article/a-comparative-study-of-cloud-gpu-offerings-for-aiml-engineers/406737