Towards the Creation of a Unified Framework for Multimodal Search and Retrieval

Apostolos Axenopoulos, Petros Daras, and Dimitrios Tzovaras

Centre for Research and Technology Hellas, Informatics and Telematics Institute,
6th Km Charilaou-Thermi Road Rd., 57001, Thermi, Thessaloniki, Greece
{axenop,daras,tzovaras}@iti.gr

Abstract. In this paper, a novel framework for search and retrieval of multimodal content is introduced as part of the EU-funded project I-SEARCH. The main objective of I-SEARCH is to create a unified framework for multimodal content search, i.e. to retrieve content of any media type (text, 2D images, video, audio and 3D) by using as query any of the above media, along with real-world information, expressive and social cues. The outcome will be a highly user-centric search engine, able to deliver to the end-users only the content of interest, satisfying their information needs and preferences, which is expected to significantly improve end-user's experience. The paper will present the concept of I-SEARCH, as well as its major scientific advances.

Keywords: Multimodal Content Search and Retrieval, user-centric search engine, RUCoD.

1 Introduction

Multimedia content, which is available over the Internet, is increasing at a rate faster than the respective increase of computational power and storage capabilities. Due to the widespread availability of digital recording devices, improved modeling tools, advanced scanning mechanisms as well as display and rendering devices, even over mobile environments, users are more and more empowered to live a more immersive and unforgettable experience with last-generation digital media, through experiencing audiovisual content. It is therefore now possible for users to rapidly move from a mainly textual-based to a media-based "embodied" Internet, where rich audiovisual content (images, sound, videos), 3D representations (avatars) and reconstructions, virtual and mixed reality worlds, serious games, life-logging applications, multimodal yet affective utterances (gestures, facial expressions, eye movements, etc.) become a reality.

This growth of popularity of media is not accompanied by the rapid development of media search technologies. The most popular media services in the Web are typically limited to textual search [1, 2]. However, the last years, significant efforts have been devoted, mainly by the European research community, for achieving

F. Alvarez and C. Costa (Eds.): UCMEDIA 2010, LNICST 60, pp. 26–34, 2012.

content-based search of images [6, 8, 9, 10], video [7, 11, 12, 13, 14] and 3D models [3, 15, 16, 17, 18]. Same endeavors are also lately noticed by the big players in these fields (Google image, Google SketchUp [4]).

Despite the significant achievements in multimedia search technologies, the existing solutions still lack several important features, which could guarantee high-quality search services and improved end-user experience. These features are listed below:

- A unified framework for multimodal content search and retrieval: this will enable users express their queries in any form most suitable for them, retrieve content in various forms providing the user with a complete view of the retrieved information and interact with the content using the most suitable modality for the particular user and under the specific context each time.
- Sophisticated mechanisms for interaction with content: these will exploit at best the social and collaborative behavior of users interacting with the content, which will enable them to better express what they want to retrieve.
- Efficient presentation of the retrieved results: this will optimally present to the user the most relevant results according to the query and the user preferences.

Towards this direction, the I-SEARCH project [5] aims to provide a novel unified framework for multimedia and multimodal content indexing, search and retrieval. The I-SEARCH framework will be able to handle specific types of multimedia (text, 2D image, sketch, video, 3D objects, audio and combination of the above) and support multimodal interaction means (gestures, face expressions, eye movements) along with real world information (GPS, temperature, time, weather sensors, RFID objects,), which can be used as queries and retrieve any available relevant content of any of the aforementioned types and from any end-user access device. Furthermore I-SEARCH will be able to integrate even non-verbal yet implicit, emotional cues, and social descriptors, in order to better express what the user wants to retrieve.

The proposed search engine is expected to be highly user-centric in the sense that only the content of interest will be delivered to the end-users, satisfying their information needs and preferences, which is expected to dramatically improve end-user experience. Furthermore, I-SEARCH introduces the use of advanced visual analytic technologies for search results presentation in order to facilitate their fast and easy interpretation and also to support optimal results presentation under various contexts (i.e. user profile, end-user terminal, available network bandwidth, interaction modality preference, etc.).

In the following, a description of the I-SEARCH concept and the project's main objectives is initially provided, followed by the major scientific advances proposed by I-SEARCH, such as the Rich Unified Content Description (RUCoD), Multimodal Annotation Propagation, Multimodal Interaction and Visualization.