I-SEARCH: A Unified Framework for Multimodal Search and Retrieval

Apostolos Axenopoulos[1], Petros Daras[1], Sotiris Malassiotis[1], Vincenzo Croce[2], Marilena Lazzaro[2], Jonas Etzold[3], Paul Grimm[3], Alberto Massari[4], Antonio Camurri[4], Thomas Steiner[5], and Dimitrios Tzovaras[1]

[1] Centre for Research and Technology Hellas (CERTH/ITI),
6th Km Charilaou-Thermi, 57001 Thessaloniki, Greece
{axenop,daras,malasiot,tzovaras}@iti.gr
[2] Engineering Ingegneria Informatica S.p.a.,
Viale Regione Siciliana Nord Ovest, 7275, 90146, Palermo, Italy
{vincenzo.croce,Marilena.Lazzaro}@eng.it
[3] Hochschule Fulda, Marquardstrasse 35, 36039 Fulda, Germany
{paul.grimm,jonas.etzold}@informatik.hs-fulda.de
[4] Casa Paganini – InfoMus Research Centre, University of Genoa,
Piazza S.Maria in Passione 34, Genoa, Italy
{alby,toni}@infomus.org
[5] Google Germany GmbH, ABC-Str. 19, 20354 Hamburg, Germany
tomac@google.com

Abstract. In this article, a unified framework for multimodal search and retrieval is introduced. The framework is an outcome of the research that took place within the I-SEARCH European Project. The proposed system covers all aspects of a search and retrieval process, namely low-level descriptor extraction, indexing, query formulation, retrieval and visualisation of the search results. All I-SEARCH components advance the state of the art in the corresponding scientific fields. The I-SEARCH multimodal search engine is dynamically adapted to end-user's devices, which can vary from a simple mobile phone to a high-performance PC.

Keywords: multimodal search, multimodal interfaces, adaptive presentation.

1 Introduction

Current Internet (CI) was developed 30 years ago for serving research demands (host-to-host communications). However, it is obvious that it cannot be used today with the same efficiency, since new demanding applications rise. The number of Internet users as well as the available multimedia content of any type increase exponentially. Moreover, the increase of user-generated multimedia content and the number of mobile users will raise new challenges. Towards this direction, the Future Internet (FI) aims to overcome current limitations and address emerging trends including: network architectures, content and service mobility, diffusion of heterogeneous nodes and devices, mass digitisation, new forms of user-generated (multimodal) content

F. Álvarez et al. (Eds.): FIA 2012, LNCS 7281, pp. 130–141, 2012.

provisioning, emergence of software as a service and interaction with improved security, trustworthiness and privacy [1].

With respect to content characteristics, the content supported by FI could be: *intelligent*, i.e. able to be adapted to the user preferences, devices and access networks; *3D and haptic*, including also visual and sound features, as well as physiological or emotional user's state; *interactive*, allowing user to interact with the media objects; *cross-modal and multimodal*, thus, providing intuitive links among future media and enabling search and retrieval from one modality to another; and *collaboratively edited/filtered*, allowing editing, filtering and manipulation of content in a collaborative way. FI is expected to address several limitations of CI, with respect to content, such as *disembodied* and *non-multimodal access* to content. The lack of embodiment in CI could be faced by enhanced support of multimodality, including sound, haptics, visual, gestural, physiological, toward a deeper exploitation and integration of communication and interaction through the physical, non-verbal, full-body channels [1].

In this sense, the EU-funded project I-SEARCH aims to create a unified framework for multimodal search and retrieval, which is fully inline with the vision and objectives of FI. The search engine proposed by I-SEARCH enables retrieval of several types of media (3D objects, 2D images, sound, video and text) using as query any of the above types or their combinations. The framework provides novel multimodal interaction mechanisms to enable easy retrieval and access by users to multimedia content as well as to capture the emotional expressive and social information conveyed by both individual and groups of expert and non-expert users. Moreover, it provides novel data representations and transformations in order to support conversion of all types of conflicting and dynamic data in ways that support visualization and analysis. Finally, it provides device adaptation capabilities, addressing several types of end-user devices, such as PCs, mobile phones, PDAs and smart phones. In this paper, the overall architecture and main functionalities of the I-SEARCH framework are presented.

1.1 Related Work

While the problem of retrieving one single modality at a time, such as 3D objects, images, video or audio has been extensively covered, retrieval of multiple modalities simultaneously (multimodal retrieval) has yet to yield significant results. In [10], the intra- and inter-media correlations of text, image and audio modalities are investigated in order to produce a Multi-modality Laplacian Eigenmaps Semantic Subspace (MLESS). In [11], a structure called Multimedia Document (MMD) is introduced to define a set of multimedia objects (images, audio and text) that carry the same semantics. After creating a Multimedia Correlation Space (MMCS), a ranking algorithm is applied, which uses a local linear regression model for each data point and it globally aligns all of them through a unified objective function. Within I-SEARCH, an approach for multimodal retrieval has been introduced. It is based on Laplacian Eigenmaps [12], while it has been further enhanced with large-scale indexing [13] and relevance feedback [14].

The integration of non-verbal expressive, emotional and social dimensions in multimodal queries enables novel ways users can access content. In [16] a novel paradigm for modelling and analyzing non-verbal full-body affective gestures is proposed. An approach to model and analyse full-body non-verbal social signals (entrainment, leadership) is presented in [17].

Multimodal search engines are still very experimental at the time of writing. For our work on I-SEARCH, we looked for common patterns in search-related actions. Across the Web, the pattern that is used for almost all search related actions is the text field. From big Web search engines such as Google, Yahoo, or Bing, to intranet search engines, this pattern stays the same. However, I-SEARCH cannot directly benefit from this broadly accepted pattern, as a multimodal search engine must support a large number of query types simultaneously: audio, video, 3D, image, etc. Some current search engines, even if they do not have the need for true multimodal querying, still do have the need to accept input that is not plain text. As a first example, we consider TinEye[1]. TinEye is a Web-based search engine that allows for query by image content (QBIC) in order to retrieve similar or related images. The interface allows for direct file upload, however, the requirements for a multimodal search engine like I-SEARCH are more complex. As a second example, we examine MMRetrieval [6]. It brings image and text search together to compose a multimodal query. MMRetrieval is a good showcase for the problem of designing a UI with many user-configurable options as well as multimodal aspects. For a user which is not involved within the field of information retrieval, the UI seems not necessarily clarify the meaning of all inputs in detail, especially when field-specific terms are used. Finally, we have a look at Google Search by image[2], a feature introduced in 2011 with the same UI requirements as MMRetrieval: combining text and image input. With the Search by image interface, Google keeps the text box pattern, while preventing any extra visual noise. The interface is exposed to users via a contextual menu when the camera icon is clicked.

Independently of the techniques used for querying and retrieval of multimedia databases, presentation of the results follows similar patterns as with text search. Major search engines such as Google Images and Bing Images present results as a rectangular grid or matrix of thumbnails that are ordered from left to right and top to bottom based on their ranking score. Google Videos and Youtube present results as a linear list of video surrogates containing a representative video shot plus accompanying text summary and metadata. Also, numerous interfaces have been developed for image browsing of personal collections. For example, in the PhotoMesa image browser [7], images in a directory are arranged in space filling boxes using a quantum Treemap algorithm. Clustering images by time is a popular way for organisation of personal collections [8]. In PhotoTOC [9] content based clustering is applied after time-based clustering for clusters that contain many images. Clustering based on faces was recently introduced in applications such as Google Picasa, Apple iPhoto and Flickr.

[1] http://www.tineye.com/

[2] http://www.google.com/insidesearch/searchbyimage.html

2 Overview

In multimodal search and retrieval problems, it is much more convenient to enclose multiple media types, which share the same semantics, into a media container, and label the entire container with the semantic concept, instead of labelling each media instance separately. Following this approach, in I-SEARCH, the concept of *Content Object (CO)* has been introduced to describe such rich media containers. A CO can span from very simple media items (e.g. a single image or an audio file) to highly complex multimedia collections (e.g. a 3D object accompanied with multiple 2D images and audio files). Moreover, a CO may include additional metadata related to the media, such as textual information, classification information, real-world data (location or time-based), etc. When a user refers to a CO, s/he directly refers to all of its constituting parts. A detailed description of CO is available at [1].

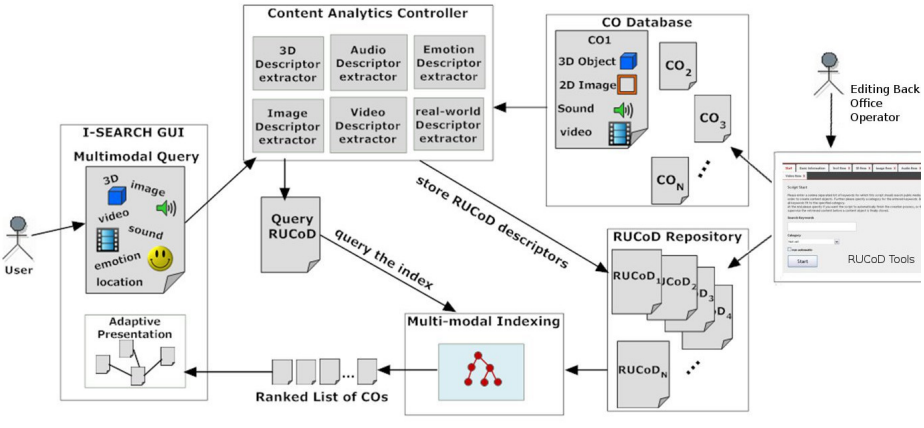

Fig. 1. Block diagram of the I-SEARCH framework

A block diagram of the I-SEARCH framework is given in Fig. 1. During the offline phase, the COs of the I-SEARCH dataset are inserted to the Content Analytics Controller (CAC). CAC is responsible for extracting low-level descriptors for each of the CO's constituting modalities. The output is a set of low-level descriptors, which are stored using a novel description format called *Rich Unified Content Description (RUCoD)*. The RUCoD format is also analysed in [1]. RUCoD descriptors are processed using a novel manifold learning framework, producing a set of multimodal descriptors, which are efficiently indexed to facilitate faster retrieval.

During the online phase, the user initiates a multimodal search session by adding one ore more modalities to the appropriate I-SEARCH interface. The interface supports text, image, video, audio and 3D queries, as well as emotional (captured by user's expressions) and real-world (user location, time) input. A query RUCoD is produced using CAC, which is used to query the multimodal index. The retrieved ranked list of COs is optimally presented to the user through the adaptive presentation component. A description of the I-SEARCH components follows.

3 Content Analytics Process

During the Content Analytics Process, an appropriate analysis is performed in order to extract descriptors from the CO's constituting modalities and store them using a multi-level structure. This structure takes into consideration: i) content-specific low-level descriptors, which characterise the type of content, ii) real-world descriptors, which associate the content with information extracted from sensors (i.e. GPS, temperature, time, weather, etc.), and iii) user-related descriptors, which encapsulate expressive, social and emotional characteristics to the semantics of these items.

The Content Analytics Controller (CAC) is the process orchestrator for low-level descriptor (LLD) extraction. As a result, LLDs are extracted for each modality within the CO and further merged into a RUCoD file. Each RUCoD is the data representation of a CO and consists of two main parts: *Header* and *Description* tags. The former includes general information edited by Content Providers during the content injection phase. The latter is representing the CO low-level features for each multimedia, real-world and user-related information. Moreover it contains the artefacts (i.e. thumbnails, key-frames, etc.) that are produced as intermediate results of low-level feature extraction phases.

Specific RUCoD Tools have been developed for content injection and RUCoD header production, which is the preliminary part of the Content Analytics process:

The *RUCoD Authoring Tool (RAT)* supports CO creation from existing media collections. It takes as input all different types of media items, real-world information and user-related information (emotional/expressive characteristics); as results a rich media representation of the Content Object is produced according to RUCoD format xml schema.

The *Crawler2RUCoD script* supports creation of Collection of COs starting from a corpus of multimedia content. This strategy is an automatic creation of one CO for each media.

The *CoFetch RUCoD Tool* performs a semi-automatic creation of COs. It provides a smart way to create a RUCoD starting from keywords. CoFetch RUCoD Tool performs search on public media sources (Text, 3D, Image, Audio and Video) and creates corresponding COs.

The core of CAC process comprises a first phase of identification of multimedia content types followed by triggering of the corresponding LLD extractors. Moreover, the CAC process is responsible for merging the results of LLD extractors into valid RUCoD files. As soon as the updated RUCoDs are stored in the platform, the Search Engine Indexers are notified. Indexers are in charge of retrieving relevant COs during the online phase.

4 Multimodal Indexing

The low-level descriptors of the COs' constituting modalities are further processed to construct a new multimodal feature space. In this new feature space all COs, irrespective of their constituting modalities, are represented as d-dimensional vectors, where semantically similar COs lie close to each other with respect to a common

distance metric. The methodology, which is usually followed, is known as manifold learning, where it is assumed that the multimodal data lie on a non-linear low-dimensional manifold. The majority of manifold learning approaches is based on the computation of the *k*-nearest neighbours among all items of the dataset in order to create an adjacency matrix. In our case, the items of the dataset are COs. The *k*-nearest neighbour computation for a CO is not a trivial process, since it requires merging descriptors of heterogeneous modalities into one unified distance metric. To avoid merging of heterogeneous distance metrics, an alternative approach has been introduced in I-SEARCH [13]. The method is based on Laplacian Eigenmaps (LE) but, in our case, the creation of the adjacency matrix is modified as follows: when items i, j are neighbours, the item W_{ij} of the adjacency matrix is assigned the value 1 instead of the actual distance between i and j. A detailed description of the method is available at [13].

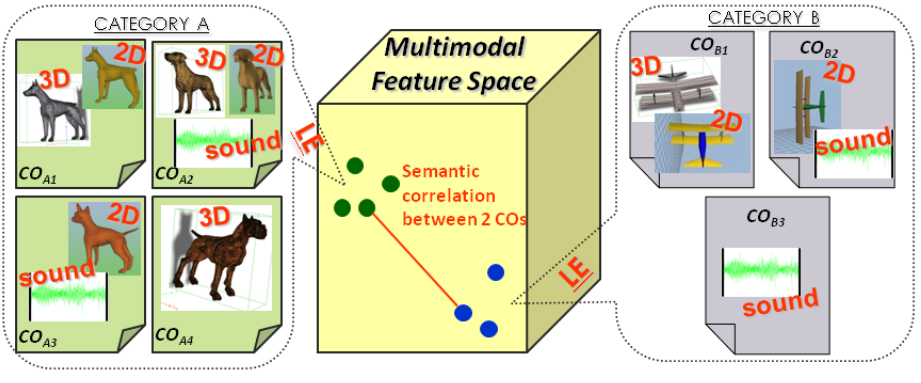

Fig. 2. Block diagram of the I-SEARCH framework

The aforementioned method relies on the calculation of all-to-all distance matrices among all objects of the dataset. However, when it comes to really large multimedia datasets, both calculation and storage of all-to-all distance matrices becomes prohibitive. Consequently, the distance matrix does not provide an efficient solution in real-life problems. On the other hand, multimedia indexing is a widely used method to speed up the nearest-neighbour search in large databases. Through indexing, there is no need to compute one-to-all distances of the query with all database objects. In I-SEARCH, a new large-scale multimedia indexing approach has been adopted to index the multimodal descriptors. The main idea of the method is that when two objects are very similar (close to each other in a metric space) their view of the surrounding world is similar as well. Thus, instead of using the distance between two objects, their similarity can be approximated by comparing their ordering of similarity according to some reference points [15].

5 Multimodal Interfaces

5.1 The I-SEARCH Graphical User Interface and Multimodality

With the I-SEARCH project, we aim at the creation of a multimodal search engine that allows for both multimodal in- and output. Supported input modalities are audio, video, rhythm, image, 3D object, sketch, emotion, social signals, geolocation, and text [5]. Each modality can be combined with all other modalities in an enhanced version of the search box pattern. The graphical user interface (GUI) of I-SEARCH is not tied to a specific class of devices, but rather dynamically adapts to the particular device constraints like varying screen sizes of desktop and mobile devices like cell phones and tablets. Fig. 3 gives an impression of what this adaptive behaviour looks like in practice and how multimodal queries are assembled i.e. on a mobile device (Fig. 4). The I-SEARCH GUI is implemented with the objective of sharing one common code base for all possible input devices.

Fig. 3. Automatic adaption of I-SEARCH GUI to different devices and screens

It uses a JavaScript-based component called UIIFace [4], which enables the user to interact with I-SEARCH via a wide range of modern input modalities like touch, gestures, or speech. Therefore it provides an adaptive algorithm for gesture recognition along with support for novel input devices like Microsofts Kinect in a web environment. The GUI also provides a WebSocket-based collaborative search tool called CoFind [4] that enables users to search collaboratively via a shared results basket, and to exchange messages throughout the search process. A third component called pTag [4] produces personalized tag recommendations to create search queries, filter results and add tags to retrieved result items.

One important goal of I-SEARCH is to hide this complexity from the end-user through a consistent and context-aware user interface based on standard HTML5, JavaScript, and CSS, with ideally no additional plug-ins like Flash required. We aim at sharing one common code base for both device classes, mobile and desktop, with the user interface getting progressively enhanced [3] the more capable the user's Web browser and connection speed are. Search engines over the years have coined a

common interaction pattern: the search box. We enhance this interaction pattern by context-aware modality input toggles that create modality query tokens in the I-SEARCH search box. Below within Fig. 5, three example modality query tokens for audio, emotion, and geolocation, can be seen.

Fig. 4. Multimodal query consisting of geolocation, video, emotion and sketch

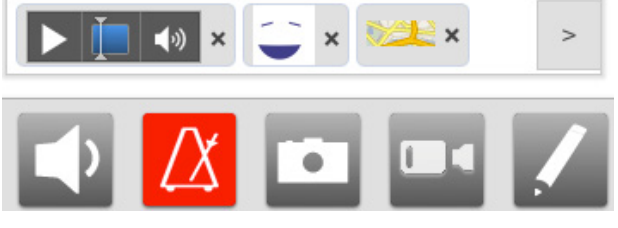

Fig. 5. Multimodality-enhanced search box pattern with query

5.2 Expressive and Emotional Interfaces

Main innovations proposed in I-SEARCH include (i) the extraction of expressive and emotional information conveyed by a user to build a query, and (ii) the possibility to build collective queries, i.e. queries resulting from a social interaction – verbal as well as non verbal – among a group of users. The I-SEARCH platform includes real-time algorithms for the analysis of non-verbal emotional behaviour expressed by full-body gesture, algorithms for the analysis of the social behaviour in a group of users, and methods to extract data from sensors for accessing real-world information. In the following we sketch a couple of use cases to explain the need for automated analysis techniques of non-verbal emotional and social behaviour.

To describe this type of interfaces, we sketch up a couple of use cases, which are also studied in I-SEARCH: a) *Individual multimodal search of music content* and b) *Social multimodal search of music content*.

According to the first use case, a professional user is looking for music material that share common features. This research aims at discovering unexpected filiations and similarities across music artworks. The target group can vary from professional users/music experts to end-users/music lovers. *Multimodal input* includes text (words, phrases, tags, etc.), audio files/clips (query by example), gestures captured via a video camera or accelerometers embedded in mobile devices and real-world information (e.g. the GPS position of the user). *Typical search and retrieval tasks are the following*: search for a list of audio files having the same rhythm of a pattern specified by the user (via tapping with a finger on a table/microphone, clapping her hands or moving her arms in the air) but also sharing the same emotional features (e.g. similar level of arousal) of the captured user movements or attitude.

The second use case deals with collaborative music retrieval by a group of users. More specifically, four friends at a party wish to dance together, and to accomplish this they search some music pieces resonating with their (collective) mood. They do not know in advance the music pieces they want, and they use the I-SEARCH tool collaboratively to find their music, and possible associated videos. *Multimodal input* includes audio or video clip of a favourite singer (query by example), text-based keywords, rhythmic queries (using hands, clapping, full-body movement), gestures, entrainment /synchronization and dominance/leadership among users, measured by on-body sensors and/or environment video cameras. *Typical search and retrieval tasks include the following*: Iterative search for audio files (as well as the video clips or images that are associated with them) by periodically performing a query for a new music piece similar to the one currently been played and having a location in the valence/arousal plane close to the position obtained from the movements of the dancers.

6 Adaptive Presentation

The proposed visualisation framework is based on a hierarchical conceptual organization of the dataset. According to this conceptual organizations the result of each query may be diverse enough to be organized in several topics and associated sub-topics, while each sub-topic (at the bottom of the hierarchy) may be specific enough to be mapped to a continuous similarity space designating a variability of a single object along some important dimensions. We argue that such organization is very suitable for explorative browsing of dataset and is diverse enough to cover a vast range of data, information needs, and browsing tasks. To achieve the proposed organization, we automatically augment the results of the multi-modal search engine with analytics information. In particular, given a mutual similarity matrix among results documents we perform hierarchical clustering by means of spectral clustering algorithm. For each resulting group of results we subsequently perform a dimensionality reduction or transformation algorithm (e.g. minimum spanning trees) that maps documents on 2D “similarity space”.

We use Treemaps, Hyperbolic Trees and classical tree-like structures interchangeably to navigate the user to specific groups of results. To avoid cluttered

displays of documents with similar coordinates we employ a fast thumbnail placement algorithm that is similar to those employed for placing labels on a cartographic map.

For visual multimedia content, such as images, video, 3D objects, an iconic or pictorial representation of the item, such as an image thumbnail, provides a summary of the object descriptive enough for the user to make relevance judgments. While generation of such pictorial representations is straightforward for inherently pictorial media, it is more difficult with media that are inherently non-visual and/or have a strong temporal dimension such as audio and video. For visualisizing audio we compute spectral features from the audio samples which are subsequently mapped to a 5-dimensional space. These five parameters are finally used for drawing parametric shapes which are used as representative thumbnails. For videos we employ a story-board based visualisation using indicative key frames.

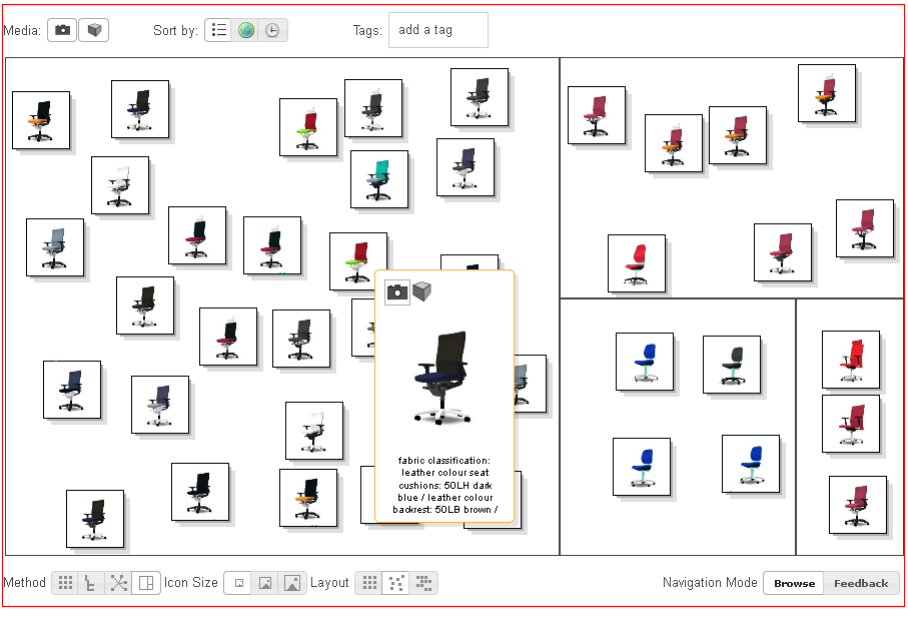

Fig. 6. Prototype of the I-SEARCH result visualisation interface. http://vision.iti.gr/sotiris/isearch/index.html

In any case we favor a hierarchical visualization of documents using three levels of detail. At the first level single thumbnails are presented aiming at fast but crud relevant judgments. The second level presents a more detailed view of the item both in content and resolution. Finally the third stage involves downloading and previewing the item in its original form. For documents containing several modalities a stacking metaphor is used at the lowest level of detail, with the most relevant modality on the top while for higher levels of detail the user may switch among different modalities by means of a menu. If real world information is available, then additional "views" are possible. Currently our system supports geographic information (latitude-longitude coordinates) and temporal information (single

time-stamp for each document). This allows rearranging the document thumbnails to reflect spatial or temporal relationships instead of document similarity.

7 Conclusions

A novel approach for multimodal search was presented in this article. I-SEARCH allows easy retrieval of multiple media types simultaneously, namely 3D objects, images, audio and video, using as queries combinations of different media types, text, real-world information, expressions or emotions captured from the user with simple devices. Several innovative solutions, which were developed within I-SEARCH, constitute the proposed search and retrieval framework: a) a method for multimodal descriptor extraction and indexing able to index COs irrespective of their constituting modalities; b) a dynamic graphical use interface (GUI), enhanced with multimodal querying capabilities; c) methods for analysing non-verbal emotional behaviour expressed by full-body gestures and translating this behaviour to multimodal queries; d) adaptive presentation of the search results using visual analytics technology. The multimodal search engine is dynamically adapted to end-user's devices, which vary from a simple mobile phone to a high-performance PC. The framework will be extended, including more functionalities, such as personalisation, relevance feedback, annotation propagation and personalised recommendation exploiting social tagging.

The technologies implemented within I-SEARCH can potentially influence the FI architecture and related frameworks. The outcomes of I-SEARCH can contribute to Future Internet Public Private Partnership (FI-PPP) [19], which aims to advance Europe's competitiveness in FI-related technologies and to support the emergence of FI-enhanced applications of public and social relevance, more specifically to FI-WARE Core Platform [18]. FI-WARE is expected to deliver an integrated service infrastructure, building upon elements (called Generic Enablers) which offer reusable and commonly shared functions making it easier to develop FI applications in multiple sectors. Since multimedia/multimodal search has not yet been adopted by FI-WARE, it can be proposed as a Generic Enabler of the FI-WARE core platform.

Acknowledgements. This work was supported by the EC-funded project I-SEARCH (http://www.isearch-project.eu/isearch/).

References

1. Daras, P., Axenopoulos, A., Darlagiannis, V., Tzovaras, D., Le Bourdon, X., Joyeux, L., Verroust-Blondet, A., Croce, V., Steiner, T., Massari, A., Camurri, A., Morin, S., Mezaour, A.D., Sutton, L., Spiller, S.: Introducing a Unified Framework for Content Object Description. International Journal of Multimedia Intelligence and Security, Special Issue on Challenges in Scalable Context Aware Multimedia Computing 2(3-4), 351–375 (2011), doi:10.1504/IJMIS.2011.044765

2. Daras, P., Alvarez, F.: A Future Perspective on the 3D Media Internet. Towards the Future Internet - A European Research Perspective (January 2009) ISBN: 978-1-60750-007-0
3. Champeon, S.: Progressive Enhancement and the Future of Web Design, http://www.hesketh.com/thought-leadership/our-publications/progressive-enhancement-and-future-web-design
4. Etzold, J., Brousseau, A., Grimm, P., Steiner, T.: Context-Aware Querying for Multimodal Search Engines. In: Schoeffmann, K., Merialdo, B., Hauptmann, A.G., Ngo, C.-W., Andreopoulos, Y., Breiteneder, C. (eds.) MMM 2012. LNCS, vol. 7131, pp. 728–739. Springer, Heidelberg (2012)
5. Steiner, T., Sutton, L., Spiller, S., et al.: I-SEARCH – A Multimodal Search Engine based on Rich Unified Content Description (RUCoD). Submitted to the European Projects Track at the 21st International World Wide Web Conference. Under review, http://www.lsi.upc.edu/~tsteiner/papers/2012/isearch-multimodal-search-www2012.pdf
6. Zagoris, K., Arampatzis, A., Chatzichristofis, S.A.: www.MMRetrieval.Net: a multimodal search engine. In: Proceedings of the Third International Conference on SImilarity Search and APplications (SISAP 2010), pp. 117–118. ACM, New York (2010)
7. Bederson, B.B.: PhotoMesa: A Zoomable Image Browser Using Quantum Treemaps and Bubblemaps.. In: ACM Symposium on User Interface Software and Technology, UIST 2001, CHI Letters, vol. 3(2), pp. 71–80 (2001)
8. Graham, A., Garcia-Molina, H., Paepcke, A., Winograd, T.: Time as essence for photo browsing through personal digital libraries. In: Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2002), pp. 326–335. ACM, NY (2002)
9. Platt, J.C., Czerwinski, M., Field, B.A.: PhotoTOC: automatic clustering for browsing personal photographs. In: Proceedings of the 2003 Joint Conference and the Information, Communications and Signal Processing Fourth Pacific Rim Conference on Multimedia, December 15-18, vol. 1, pp. 6–10 (2003)
10. Zhang, H., Weng, J.: Measuring Multi-modality Similarities Via Subspace Learning for Cross-Media Retrieval. In: Advances in Multimedia Information Processing, PCM (2006)
11. Yang, Y., Xu, D., Nie, F., Luo, J., Zhuang, Y.: Ranking with Local Regression and Global Alignment for Cross Media Retrieval. ACM MM, Beijing, China (2009)
12. Axenopoulos, A., Manolopoulou, S., Daras, P.: Multimodal Search and Retrieval using Manifold Learning and Query Formulation. In: ACM International Conference on 3D Web Technology, Paris,France, June 20-22 (2011)
13. Daras, P., Manolopoulou, S., Axenopoulos, A.: Search and Retrieval of Rich Media Objects Supporting Multiple Multimodal Queries. Accepted on IEEE Transactions on Multimedia
14. Axenopoulos, A., Manolopoulou, S., Daras, P.: Optimizing Multimedia Retrieval Using Multimodal Fusion and Relevance Feedback Techniques. In: Schoeffmann, K., Merialdo, B., Hauptmann, A.G., Ngo, C.-W., Andreopoulos, Y., Breiteneder, C. (eds.) MMM 2012. LNCS, vol. 7131, pp. 716–727. Springer, Heidelberg (2012)
15. Gennaro, C., Amato, G., Bolettieri, P., Savino, P.: An Approach to Content-Based Image Retrieval Based on the Lucene Search Engine Library. In: Lalmas, M., Jose, J., Rauber, A., Sebastiani, F., Frommholz, I. (eds.) ECDL 2010. LNCS, vol. 6273, pp. 55–66. Springer, Heidelberg (2010)
16. Glowinski, D., Dael, N., Camurri, A., Volpe, G., Mortillaro, M., Scherer, K.: Towards a Minimal Representation of Affective Gestures. IEEE Transactions on Affective Computing 2(2), 106–118 (2011)
17. Varni, G., Volpe, G., Camurri, A.: A System for Real-Time Multimodal Analysis of Nonverbal Affective Social Interaction in User-Centric Media. IEEE Transactions on Multimedia 12(6), 576–590 (2010)
18. http://www.fi-ware.eu/
19. http://www.fi-ppp.eu/