Multimed Tools Appl
https://doi.org/10.1007/s11042-018-5931-7

Multimedia and geographic data integration for cultural heritage information retrieval

Erasmo Purificato[1] · Antonio M. Rinaldi[1]

Received: 8 September 2017 / Revised: 28 February 2018 / Accepted: 25 March 2018

Abstract In this paper a system providing an efficient integration between Content-Based Image Retrieval (CBIR) and Geographic Information Retrieval (GIR) is presented. Over the years, many CBIR systems have been proposed to give a solution for an efficient use of multimedia/visual contents and other issues as performance, quality of retrieval, data heterogeneity, and multimodal information integration. The aim of the proposed approach is to prove that the use of geographic data can improve the results obtained by an image matching system based only on visual data. Our framework is composed of three parts, each of them described in detail in this paper: the first part is dedicated to CBIR, with an experimental comparison of a large number of different multimedia features to choose the one to use in the system implementation; in the second part the methodology to integrate geographic and multimedia data is showed; in the last part is presented a GIR system implementation using a "points of interest" search. An Android application has been developed for the client-side using Apache Solr as server side provider for the information retrieval functionalities. An experimental evaluation is carried out to demonstrate the effective improvement given by the combination of geographic and multimedia data. Our results have been obtained using a real dataset composed of artworks located in Naples's museums.

Keywords Content-based image retrieval · Geographic information retrieval · Multimodal query · Digital cultural heritage

✉ Antonio M. Rinaldi
antoniomaria.rinaldi@unina.it

Erasmo Purificato
erasmopurif@gmail.com

[1] Department of Electrical Engineering and Information Technologies, University of Naples Federico II, Napoli, Italy

1 Introduction

Art and culture have always represented an important part in the evolution of human lives. Over the centuries, art sites have represented a very important sources of information and an interesting context for several ICT applications. In the last few years, the interest about digital cultural heritage and the preservation of historical materials for next generations has increased [16]. Generally speaking, the causes that give importance to digitalization are the following:

- Personal computers, smartphones and tablets are today the first learning and informative channel and visiting museums is possible everywhere by means of these devices;
- Digital artworks can't be destroyed and vandalized as real paintings and sculptures (e.g. Michelangelo's *Pietà*) [14].

In this context, a significant role is played by a set of technologies as *Content-Based Image Retrieval* (CBIR). Also known as Query By Image Content (QBIC) or Content-Based Visual Information Retrieval (CBVIR), CBIR is the application of computer vision to image retrieval. Its goal is to avoid the use of textual descriptions and to develop techniques for retrieving images on the basis of automatically-derived features, such as color, edge, texture and shape [49]. In this scenario, the word "content" refers to measurable visual properties of the image that represent its physical meaning [47]. As an example, we can consider a painting like *Mona Lisa*: for a visitor of the Louver museum, the content is any information about the artworks, the represented person, the author, the painting technique; for a computer the content refers to the combination of textures, shapes, colors and edges.

According to [15], there are other several issues to take into account:

- *Performance*: the quality of retrieval is considered the most critical issue in CBIR applications and significant efforts are focused on improving this aspect.
- *Volume of data*: the daily use of mobile devices and the expansion of the world-wide-web lead public image databases to an uncontrolled growth. Researchers estimate an annual growth rate of 16.2% of taken photos by mobile phones, foreseeing a frightening amount of 4.9 trillion stored images in 2017[1].
- *Semantic learning*: to smooth the issues regarding the semantic gap some directions point at learning image semantics from training data and developing retrieval mechanism to improve semantic descriptions. The semantic gap also refers to the inability of a machine to fully understand and interpret images based on automatically extracted data.
- *Heterogeneity*: it is important to understand that a digitized artwork is not the work itself, but only an image of this work, captured under specific conditions (camera position, brightness, resolution, etc.) and parameters such as quality and color depth often varying. This issue causes alterations in features extraction task.
- *Multimodal features*: the association of metadata (audio, location, text caption) with the images improve the retrieval performance, but sometimes an ambiguous recognition can be misleading and lead to a misunderstanding of the image content.
- *Concurrent Usage*: in on-line image retrieval systems, it is natural to have concurrent users. For this reason they must be efficiently designed to prevent an overloading of the host server resources.

[1] http://resourcemagonline.com/2014/12/infographic-there-will-be-one-trillion-photos-taken-in-2015/45332/ Last seen March 28, 2018.

- *User interface*: important efforts have to be performed in the design of interfaces for image retrieval using a user friendly approach.
- *Operating speed*: time is a critical aspect for web systems or mobile applications. Response time have to be low for good interactivity with end users and for this reason the implementation of such kind of systems should consider efficient algorithms for indexing and enterprize databases or advanced NoSQL technologies.
- *System evaluation*: CBIR systems need of a good process of evaluation and the design of benchmarks to perform testing requires careful appraches [42].

The concepts of *Geographic Information Retrieval* (GIR) and spatial query can be illustrated starting from a definitions given in [35]: GIR is generally considered as the specialization of Information Retrieval (IR) and Geographic Information System (GIS) that provides access to georeferenced information sources. The acronym GIS is often used for ”Geographic Information Science” (also GIScience) to refer to a research field that studies the use of geographic information systems technologies with other research domains within the broader academic discipline of geoinformatics [41]. In general, a Geographic Information System describes any information system that integrates, stores, edits, analyzes, shares and displays geographic information.

The union of the IR and GIS concepts leads to define Geographic Information Retrieval as a set of methodologies and techniques used to build a system that to index, query, retrieve and browse georeferenced data.

Spatial queries are also known as geographic query and are defined by Larson [35] as a type of queries with spatial relationships to entities geometrically defined and spatially localized. This kind of queries require that the space is represented by a well-defined coordinate system of the real world.

In this paper an approach to integrate Content-Based Image Retrieval techniques, Geographic Information Systems and Spatial queries is proposed to demonstrate the improvement of accuracy in the information retrieval process. This approach as been implemented in an Android application that make available a user mobile multimodal interface and server side functionalities based on Apache Solr.

The paper is structured as follows: in Section 2 some of the recent works published in last years about CBIR and GIR are presented. The system architecture and the proposed approach for multimodal querying are discussed in Section 3. In the same Section an overview on the mobile application is shown. The results and the methodology for the experimental evaluation, both on the multimedia features and the whole system, are discussed in Section 4. The evaluation has been performed on a real dataset composed of artworks located in Naples's museums to show the effective improvement given by the integration of geographic and multimedia data. The complete dataset is presented in Appendix A. Eventually, conclusion and future works are discussed in Section 5.

2 Related works

In this section we will present some relevant research works regarding to our context of interest. Because of the complexity of our scenario we discuss separately the literature about Content-Based Image Retrieval and Geographic Information Retrieval.

2.1 Content-based image retrieval systems

Content-based image retrieval it refers to methodologies and technologies that allow to manage and analyze digital images archives by their visual content; this research area includes different disciplines from computer science to robotics and also psychology.

CBIR systems can be organized around three broad categories [52]:

- *Search by association*: the aim of these systems is searching in large sets of pictures from undefined sources, with iterative refinements of the search results using relevance feedback techniques. The idea behind this task used some Information Retrieval system is to take the results retrieved from an initial query, collect user feedbacks, and use these information about relevance to perform a more accurate query [20]. An important feature of this kind of systems is the interactivity;
- *Search by a specific image*: also known as *image matching*, it has the purpose to retrieve a copy of the picture submitted at the search. These systems are broadly used for seeking in catalogs and domain categories as artworks [6];
- *Category search*: the target of this search is to obtain an image representative of a specific class of object. Labeling or inference are the techniques used to derive the categories [21].

Researches about Content-Based Image Retrieval had a significant increasing in the last decades in terms of new directions as high-dimensional multimedia data indexing [3], relevance feedback [60], application to art and culture [28, 55], applications of CBIR to medicine [4, 11, 34], and methods for content based 3D shapes retrieval [54].

In all of the CBIR system proposed in last years pictures are represented by numeric values, known as *features* or *descriptors*. Nowadays, two approaches exist:

- *Discrete approach*: it uses text retrieval metrics and inverted filed methods. All the visual descriptors are mapped to binary features and every features in an image is consider as a word in a textual document. In this context lies *VIPER* system [53], publicly available as *GNU Image Finding Tool* (GIFT). The discrete approach has the advantage of using text information retrieval methods;
- *Continuous approach*: it is a kind of nearest neighbor classification. Feature vectors are used to represent images and compared with metrics based on distance evaluation. A ranking shows the results and the highest images are those with lowest distances. A lot of existing CBIR systems follows the continuous approach, such as *SIMBA* [51], *Blobworld* [8], *CIRES* [25], *SIMPLIcity* [56], *IRMA* [36] and *FIRE* [18].

A comparison between the two aforementioned approaches is proposed in [17].

In the last few years, CBIR systems have been consider has compete application domains for several research fields in computer science. In [27] a Content-Based Image Retrieval technique based on genetic algorithms with support vector machines and user feedbacks for image retrieval purposes is proposed using a web 3.0 architecture. Karamti et al. [30] present an image retrieval framework which includes a vectorization technique combined with a pseudo relevance model. CBIR have taken advantage of neural networks and deep learning techniques. In fact, after the study of Krizhevsky et al. [33] with the *AlexNet* results, interesting research lines have been focused on deep learning based methods [26, 29, 43], with a particular interest on convolutional neural networks (CNNs) [48, 59].

2.2 Geographic information retrieval systems

The history of modern *Geographic Information Retrieval* systems started from three events considered milestones, as described in [40]. The first one was the institution of the Canadian Geographic Information System (CGIS) in 1963, a project lasting almost a decade and in which the term ”Geographical Information System” was coined for the first time in 1966. The CGIS system was designed to allow the Canadian Federal and Provincial Administration to make a land inventory and map the location of the natural resources of the state. This system also included functionality for automated area calculations and generating vector data, such as points, lines, polygons, from raster images.

The second important event happened in the United States in 1970, when GIS technology was used for analyzing the US population census. The innovation was in Dual Independent Map Encoding software, developed from the late 1960s, which enabled digitalization of all the streets in the United States so that population data could be linked to a geographical location. The technology of this software was later combined with CGIS at the Harvard Laboratory for Computer Graphics and Spatial Analysis to create the first general-use GIS in the 1970s.

The third milestone is more recent and took place in 1995, when the United Kingdom became the first country to cover its territory with standard-scale digital maps [38]. This leads to a worldwide increasing in the use of GIS. In the same year the *Open Geospatial Consortium* (OGC)[2] was created. It is an international no-profit organization committed to making quality open standards for the global geospatial community. Its standards are also applied to sensor web and Internet of Things, GIS data processing and data sharing.

Geographic Information Systems are available to common users only since last fifteen years with Google and Microsoft that introduced, in 2005, the most used on-line GIR services. They are, respectively, *Google Maps*[3] and *Live Search Maps* (today *Bing Maps*[4]).

In the following of this section we will discuss some research works proposed in last years about GIR systems.

In [2] is described the use of a Geographic Information System to create databases of documents and federated databases, improving the handling of complex data sets and managing the information retrieval process. The system has been used as a part of an archeological experiment on the site of Piazza Armerina, Enna, Sicily, Italy. It aim is in the design of an information system able to store and organize heterogeneous data for researchers and technicians.

A framework for GIS design focused on indexing methods to efficiently process spatial keywords queries is presented in [23]. In particular, they prove that the proposed strategies improve the efficiency of answering compared with other existing techniques.

Shyu et al. [50] present a content-based multimodal Geospatial Information Retrieval and Indexing System (GeoIRIS) based on automatic feature extraction, visual content mining and high-dimensional database indexing for fast retrieval.

The huge growth of Internet leads to a new world of possibilities for GIR research and applications, such as georeferenced media or geotagged pictures posted on social networks. An extensive survey is [58] where are highlighted four significant aspects of GIR systems:

[2]http://www.opengeospatial.org. Last seen March 28, 2018.

[3]https://www.google.com/maps. Last seen March 28, 2018.

[4]https://www.bing.com/maps. Last seen March 28, 2018.

(I) Storing and browsing georeferenced data; (II) Mining semantic informations from geographic data; (III) Learning landmarks in the world; (IV) Discover geographic location of pictures.

A general-purpose geographic search engine, called *Frankenplace*, is described in [1]. It is a GIR system consisting in an interactive thematic search map that use georeferenced data to discover, organize and display documents retrieved from an user query. The particularity of this project is that each query results shows also the geographic relation between the retrieved documents. Actually the last release is dated May 2015 and in this version are indexed over five million documents.

More recently, many efforts about GIR systems have been conducted on resolving semantic ambiguities on place names [31], on extracting spatial information from the internet searches of users and from social media [22, 45] and, on identification of spatial features in textual documents [57].

2.3 Applications to cultural heritage

For centuries, our memory institutions (i.e. libraries, archives, museums) have spent their efforts on collecting and describing artifacts and social phenomena to preserve and give access to our cultural heritage. In this context, several questions related to digital libraries management and information storage and retrieval represent hard issues and exciting challenges.

Koolen et al. [32] discuss opportunities and challenges in applying modern information retrieval techniques to cultural heritage domain, with a detailed case study using the large digital collection of the Gemeentemuseum of Den Haag, Netherlands.

In the era of Internet and mobile devices, online cultural heritage exhibitions are an important way to involve people in the process of digital access to art and culture. In [37] a survey study on the current landscape and architecture of major cultural heritage online exhibitions available on the web is presented.

Mobile applications represent a fast and smart solution to have information about cultural heritage. **Smartify**[5] and **GetCOO Travel**[6] are interesting examples. Smartify is a free app that allows to scan and identify artworks, access rich interpretation and build a personal art collections in some of best museums and galleries around the world. GetCOO Travel is a smartphone application developed for tourists. They can take pictures of monuments to receive information about it.

In this paper we present a system that implement a novel multimodal query approach to combine efficiently multimedia and geographic data and improve the user satisfaction during the cultural heritage information retrieval process. In particular, the real advantage of this system is in the use of GPS device sensor, in fact, activated by the user during an artwork recognition, it allows to effectively exploit the combination between location of the considered artwork and the multimedia data provided from the image matching task. Practically, the location of an artwork, based on the geographic position of the user is used to filter the results obtained from a multimedia search, considering only the artworks sited in the place resulting from the geographic search.

[5] https://smartify.org. Last seen March 28, 2018.

[6] https://www.travel.getcoo.com. Last seen March 28, 2018.

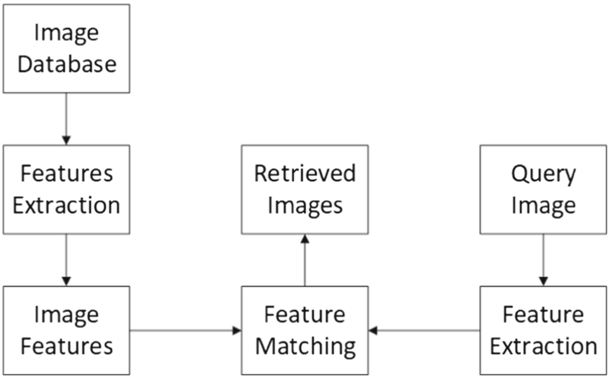

Fig. 1 General image retrieval process

In according to the use of mobile technologies for digital cultural heritage access, we implement an Android application and test our framework and system in a real use case on museums in the Italian city of Napoli.

3 The system implementation

From a methodological point of view, the proposed architecture is based on the general image retrieval process illustrated in Fig. 1.

The basic idea behind this kind of system [5] is around the feature selection and retrieval tasks. In fact, when we develop an image database, feature vectors from images are extracted and stored to allow the computation of a similarity function. This function is used to measure the distance among the features vector of the query image and all the other image vectors in the database. The obtained scores are ordered to give a rank of similarity. Moreover, a CBIR system could return a single result which represent the most relevant multimedia document with respect to the user query.

On the other hand, the proposed system architecture is in Fig. 2, where each modules is devoted to perform specific tasks in the whole retrieval process.

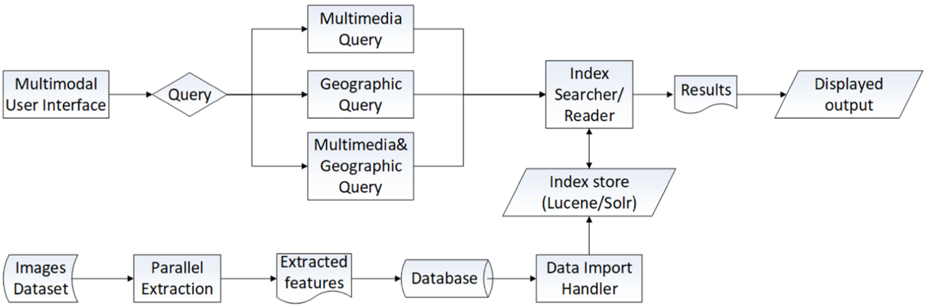

Fig. 2 The architecture of the proposed retrieval system

Multimodal user interface it allows users to select the kind of query to perform. It is implemented in the mobile application with an intuitive layout.

Query it represent the function that send data from the mobile application to the server for performing the query.

Multimedia query it is the principal task of Content-Based Image Retrieval; it aim is to implement a strategy to retrieve images based only on automatically extracted visual descriptors.

Geographic query it implements a *points of interest search*, displaying on a map a specific location that is useful and interesting for the user. Due to our context of interest (i.e. cultural heritage), we consider as points of interest for this system only the museums that are around the user location.

Multimedia&Geographic query it consists in filtering the results obtained from the CBIR functionality using the location of the considered artwork, based on the geographic position of the user. This implies that the functionality is available only for mobile devices.

Images dataset it contains the artworks considered in the test set, each of which is replicated with different size, brightness conditions or different point of view (in case of sculptures).

Parallel extraction it extracts multimedia features from the images in parallel tasks.

Database it is the storage containing the features of interests for our application.

Data import handler it provides a mechanism for importing content from a data store and indexing it in a Solr instance (*Index store*).

Index searcher/reader it has the task of performing the correct query on the Lucene index store and managing the results.

Displayed output the *Query Results* are sent to the mobile application and shown to the user.

3.1 Multimodal query strategy

Now we are in the position of describe the proposed approach for multimodal querying.

As mentioned before, the user interface is provided by an Android application designed as a client-server model. Due to the chosen functionality, client sends to server an image (*Multimedia Query*), a location (*Geographic Query*) or both (*Multimedia&Geographic Query*). The server-side system receive and manage data sent by client application to call the right method, based on data types received in the request.

In this context, we define the data structures used in the implementation of the described functionalities:

1. `InputData` is the structure created by client and sent to the server. It has two attributes:

- `base64String`: encoded string of the image uploaded or captured by user;
- `userPosition`: user geographic position.

2. `Cultural_Object` contains multimedia and spatial components and the other attributes about the artwork represented by that object:
   - `title`: name of the artwork;
   - `author`: name of the artist who produced the artwork;
   - `imageUri`: path on system where the selected/captured picture is stored;
   - `featureVector`: multimedia descriptors of the artwork represented by an alphanumeric representation of the features;
   - `place`: name of the art point of interest in which the artwork is sited;
   - `location`: geographic position of the artwork;
   - `wikiUrl`: URL of the Wikipedia web page referred to the artwork, if available.
3. `Museums_List` represents the result of the POI search and contains information about all the museums retrieved.

The implemented procedure is described in pseudocode in Table 1.

3.2 Features for content-based image retrieval

Features represent the information extracted from an image in terms of numerical values. The term "feature" is also often referred as descriptor. There are two types of features that can be extracted from images namely global and local features. *Global* features describe an image as a generalization of the entire object, and they are generally used in image retrieval, object detection and classification. Global features include contour representations, shape descriptors, and texture features. *Local* features, indeed, describe image key points of represented object and they are usually used for object recognition and identification.

In the following an overview of the features used in our work is given. In [19] is presented a comprehensive analysis of solutions proposed in the literature.

Pyramid of histograms of orientation gradients The idea is to represent an image by its local shape and the spatial layout. This descriptor [7] consists of a histogram of orientation gradients over each image subregion at each resolution level – a Pyramid of Histograms of Orientation Gradients (PHOG). The distance between two PHOG image descriptors reflects the property of images to contain similar shapes in corresponded spatial layout.

Auto color correlogram This color feature has been presented in [24]. The main characteristics of the Auto Color Correlogram feature are: spatial correlation of colors; possibility to be used to describe global distribution of local color spatial correlation; low computational effort; small size of the feature.

MPEG-7 features The Moving Picture Experts Group (MPEG) defined several multimedia descriptors in their MPEG-7 standard. In [9, 44] an overview of these multimedia content description features is presented. The MPEG-7 features used in this work are briefly illustrate below, following the description in [19].

Edge histogram descriptor This feature represents the spatial distribution of five types of edges, that is four directional edges and one non-directional edge. According to the MPEG-7 standard, the image retrieval performance is significantly improved combining the edge

Table 1 Multimodal search algorithm

```
//---------------------------------------------------------------------------
// Multimodal_Search algorithm
//
// INPUT: InputData
//
// OUTPUT: Cultural_Object for Multimedia Query and Multimedia&Geographic Query
//         Museums_List for Geographic Query
//---------------------------------------------------------------------------

Multimodal_Search(InputData)
{
   Get InputData
   if(InputData.userPosition == null) {
     Decode InputData.base64String
     Store image at Cultural_Object.imageUri
     Extract features vector from image at Cultural_Object.imageUri
     Set Cultural_Object.featureVector with the vector obtained
     Perform Multimedia Query using Cultural_Object.featureVector
     Get first result of Result_Ranking
     Set Cultural_Object.title, Cultural_Object.author, Cultural_Object.place
     Send Cultural_Object as result to client
   }
   else if(InputData.base64String == null) {
     Perform Geographic Query using InputData.userPosition
     Add all results to Museums_List
     Send Museums_List as result to client
   }
   else {
     Decode InputData.base64String
     Store image at Cultural_Object.imageUri
     Extract features vector from image at Cultural_Object.imageUri
     Set Cultural_Object.featureVector with the vector obtained
     Perform Multimedia Query using Cultural_Object.featureVector
     Add all results to Multimedia_Query_Result
     Perform Geographic Query using InputData.userPosition
     Add result to Museum
     Select all artworks in Museum and add the to Artworks_List
     For each (Cultural_Object MMObj in Multimedia_Query_Result)
       For each (Cultural_Object ArtObj in Artworks_List)
         If (MMObj.title == ArtObj.title)
           If (MMObj.author == ArtObj.author)
             Add result to Cultural_Object
     Send Cultural_Object to client
   }
}
```

descriptor with other descriptors such as the color histogram feature. This descriptor is scale invariant and supports rotation invariant and rotation sensitive matching operations.

Color layout descriptor It is the feature which effectively represents the spatial distribution of the color of visual signals in a compact form. This compactness enables visual signal matching functionality with high retrieval efficiency at very low computational costs. The outcome of this descriptor is a tiny image (8×8 pixels) that can represent the original image. Detail representation is based on coefficients of the Discrete Cosine Transformation.

Scalable color descriptor It is a color histogram in the HSV color space which is encoded by a Haar Transform. Its binary representation is scalable in terms of bin numbers and bit representation accuracy over a wide range of data rates. Retrieval accuracy increases with the number of bits used in representation. Default setting is 64.

Fuzzy color and texture histogram FCTH is a low level feature that combines, in one histogram, color and texture information [12]. FCTH feature results from the combination of three fuzzy units. Its size is limited to 72 bytes per image, rendering this visual descriptor for use in large image databases.

Color and edge directivity descriptor CEDD has been presented in 2008 by the same authors of the FCTH feature [10] and they follow the same approach. CEDD descriptor incorporates color and texture information in a single histogram, with the same color information of FCTH using two fuzzy systems that map the colors of the image in a 24-color custom palette, but its size is limited to 54 bytes per image (compared to 72 bytes of the FCTH feature). The most important attribute of the CEDD is the low computational cost needed for its extraction, in comparison with the needs of the most important visual descriptors.

Compact composite descriptor CCD are global image descriptors that capture more than one feature at the same time in a compact representation.

In [13] a combination of the two CCDs previously described, CEDD and FCTH, is proposed, called Joint Composite Descriptor (JCD). Based on the fact that the color information given by the two descriptors comes from the same fuzzy system, they assumed that joining the descriptors would result in the combining of texture areas carried by each descriptor. JCD is composed of seven texture areas, with each of those made up of 24 sub regions that correspond to color areas (the 24-bins histogram of FCTH and CEDD).

Scale invariant feature transform SIFT [39] is an approach to transform image data into coordinates relative to local features. These features are invariant to image scaling and rotation, and partially invariant to variations in brightness and 3D camera viewpoint. Furthermore, this feature is highly distinctive, that allows a single feature to be correctly matched with high probability against a large database of features, and this is the base of object recognition.

3.3 The used technologies

The extracted multimedia features and all the artworks informations are stored in a *Neo4J* database, whose logical view consists in four nodes: *Cultural Object*, *Author*, *Multimedia Representation* and *Place*.

The whole Neo4J instance is indexed into the *Apache Solr* search engine using the Neo4J *JDBC Driver* and the Solr *Data Import Handler* (DIH).

The Solr queries are computed by servlets deployed on an *Apache Tomcat* server. These servlets use *Lucene Image Retrieval* (LIRe) library for Content-Based Image Retrieval functions and of *LIRe Request Handler*, a Solr plug-in to execute multimedia queries.

An Android application has been implemented to allow users to manage the proposed system. The communication between the application and the server is based on *HTTP* protocol and, in particular, through HTTP *POST* requests.

The client-side application provides a multimodal user interface for the three main system functionalities: Multimedia Query, Multimedia&Geographic Query and Geographic Query. The server-side servlets are implemented for receiving and management of data sent by the Android application, establishing connections with Solr server, performing Solr queries and handling their results to send back to client.

Figure 3 shows a multimodal user interface example of the Android application. In particular, Fig. 3a shows the home screen of each functionality and in Fig. 3b the result of the retrieval process is shown. In this interface a map is displayed drawing a *marker* which represents the geographic location of the retrieved artwork. The *info window* on the marker contains information about the artwork. An example of results about the points of interest search is in Fig. 3c. The activity diagram in Fig. 4 has the aim to show at a glance the activities flow evolution of the client application, starting with the choice of a functionality and ending with the visualization of the results on a map or, optionally, with the web page of the artwork.

4 Experimental results

A dataset composed of paintings and sculptures images from six museums of Naples has been usd in the evaluation process:

1. Museo nazionale di Capodimonte;
2. Gallerie di Palazzo Zevallos;
3. Museo nazionale di San Martino;
4. Cappella Sansevero;
5. Museo archeologico nazionale di Napoli;
6. Museo civico di Castel Nuovo.

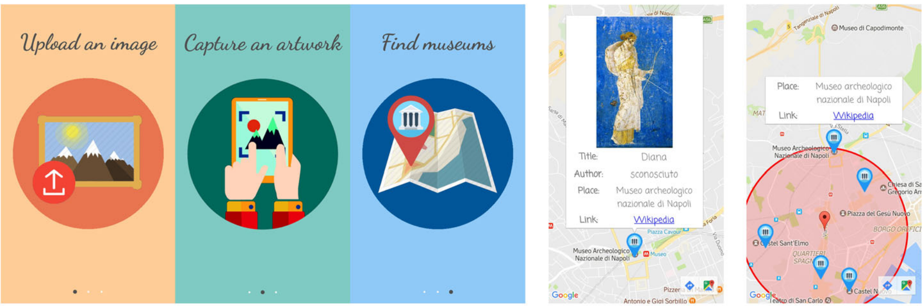

Fig. 3 Android application example

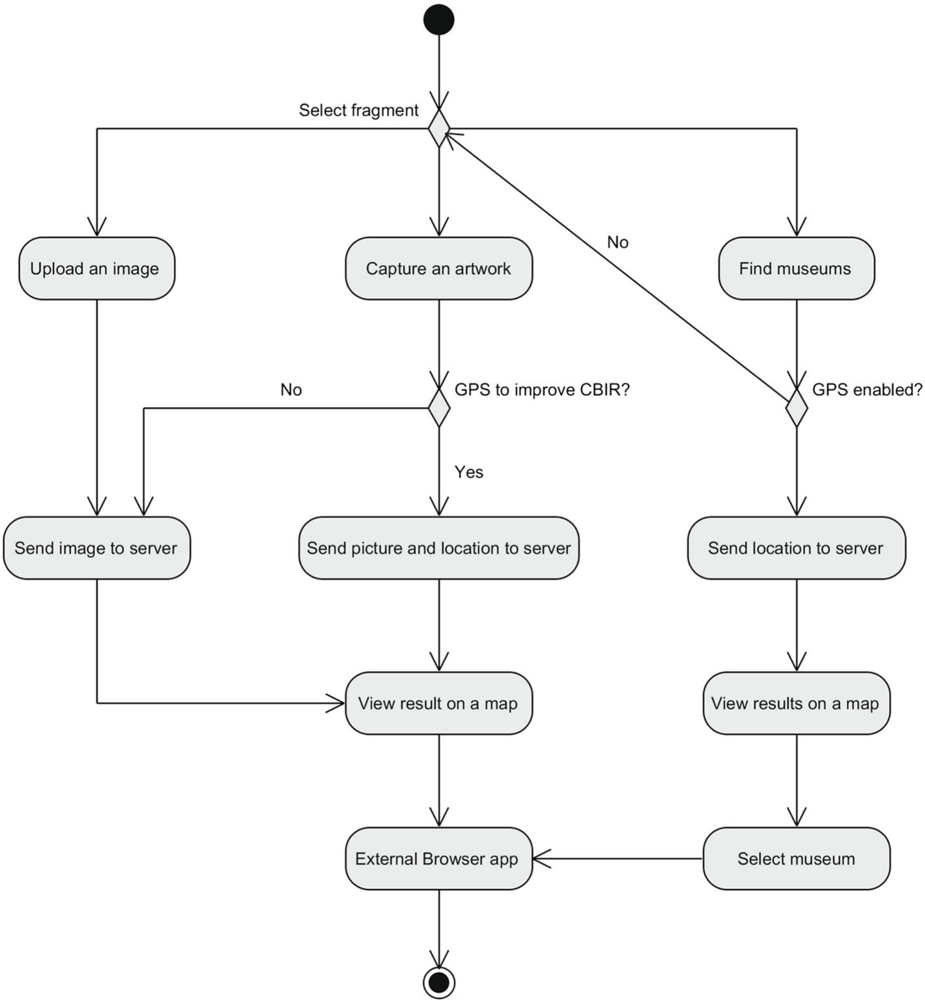

Fig. 4 Activity diagram of the client application

The complete artworks list is shown in Appendix A.

4.1 Evaluation of image descriptors

In the evaluation process the features described in Section 3.2 have been extracted from each dataset image and indexed into a Solr core using the *ParallelSolrIndexer* class provided by LIRe plug-in. Then, the *Multimedia Query* illustrated in Section 3.1 is used following the steps listed below:

- a single feature is selected and set as parameter of the features extraction;
- the extraction is performed on an image at a time, representing one of the image in the dataset, with possible variations in scale, brightness or viewpoint. The image is loaded in the system and the response value is passed as parameter in *search by feature* query.

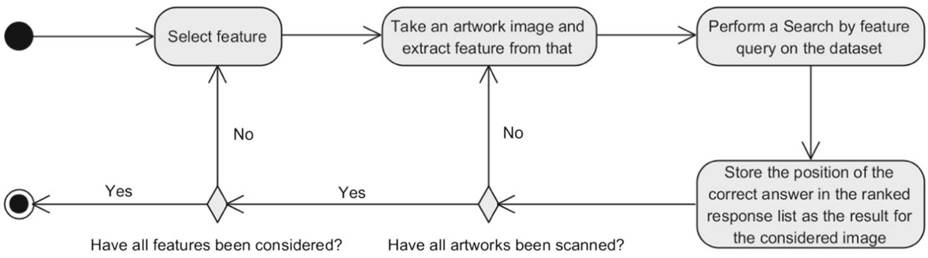

Fig. 5 Activity diagram – Features evaluation process

In this phase results are still taken as a ranked list and the position of the correct answer in this ranking is stored to later evaluate the effectiveness of the considered feature;
– once all the images are scanned, the process comes back to the first step, selecting a different feature and so on to all of them.

In Fig. 5 the activity diagram of the evaluation process is shown.

The ratings of features have been calculated from two points of view, both based on the position of the correct answer in the ranked result list:

1. In the first evaluation, the percentage of the exactly matched artworks is considered, that is the artworks correctly retrieved and resulted in first position in the ranking, and in this way the accuracy of each feature is estimated. This test had the aim to choose the feature to use in the system in the *Multimedia Query* functionality, in which only multimedia descriptors are considered;
2. The second evaluation is used to find the best feature to use in the integration with geographic data. For this aim, four thresholds based on the position in the ranked results list are considered and a score for each threshold is set:

   – 1st (artwork correctly retrieved): 3 points;
   – 2nd – 5th: 2 points;
   – 6th –10th: 1 point;
   – over 10th: 0 points.

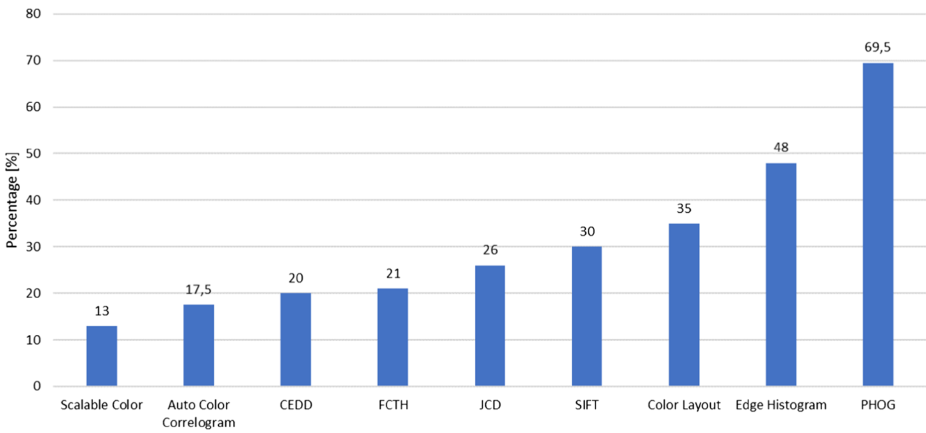

Fig. 6 Features evaluation – Exact matching test

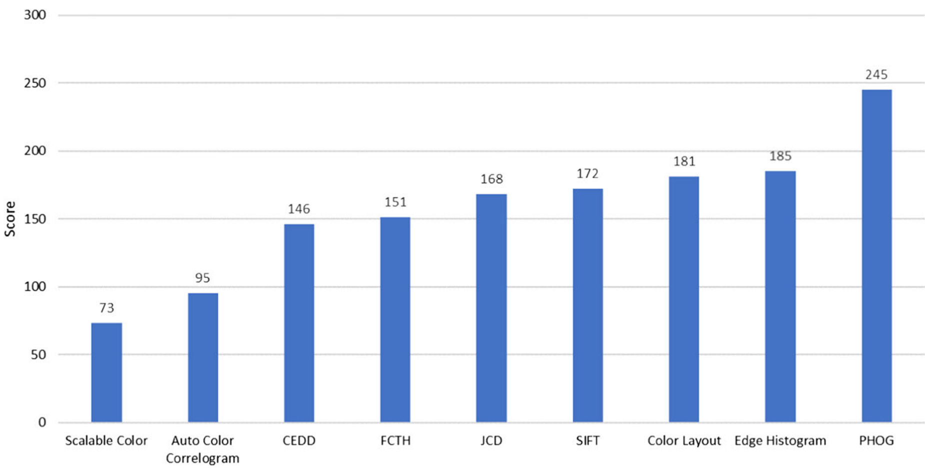

Fig. 7 Features evaluation – Score test

The results of the tests are shown, respectively, in Figs. 6 and 7. The trend of the two graphics is the same and they display significant performances of the PHOG feature. we decided to use this descriptor for the implementation of the Android application in the image features extraction task.

4.2 System evaluation

The test strategy proposed in [46] has been used in order to have a reliable evaluation of the whole system. To evaluate the *accuracy* of the retrieval system, its outputs must be compared with a ground truth, which is determined by humans. In particular, the developed application has been tested ”on the field” in museums that granted the access in their rooms for the testing. These museums have been: Museo nazionale di Capodimonte, Museo nazionale di San Martino and Gallerie di Palazzo Zevallos.

For each artwork considered in the dataset and located in these places, both the *Multimedia Query* and the *Multimedia&Geographic Query* are tested. Some paintings or sculptures has been correctly recognized only after a crop operation, because of their position that makes difficult to take a good picture immediately. So, for these artworks, testing the *Multimedia&Geographic Query* was not possible, due to the impossibility in the use the camera and the GPS sensor at the same time.

Results are organized in the following way:

– Tables 2, 3 and 4 show the summary of the tests in each single museums to make results immediately visible;

Table 2 Evaluation summary – Museo nazionale di Capodimonte

	Number	Percentage
Analized artworks	17	85%
Multimedia Query positive (including crop)	7	41.18%
Multimedia&Geographic Query positive	12	70.59%

Table 3 Evaluation summary – Gallerie di Palazzo Zevallos

	Number	Percentage
Analized artworks	18	90%
Multimedia Query positive (including crop)	7	38.89%
Multimedia&Geographic Query positive	15	83.33%

Table 4 Evaluation summary – Museo nazionale di San Martino

	Number	Percentage
Analized artworks	14	87.5%
Multimedia Query positive (including crop)	7	50%
Multimedia&Geographic Query positive	10	71.43%

Table 5 Global evaluation summary

	Number	Percentage
Analized artworks	49	87.5%
Multimedia Query positive (including crop)	21	42.86%
Multimedia&Geographic Query positive	37	75.51%
Difference after integration	+16	+32.65%

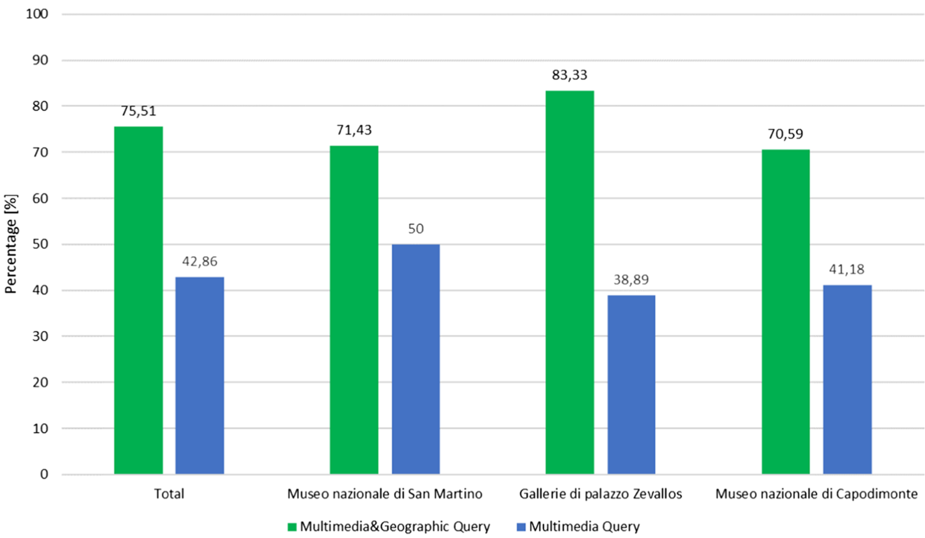

Fig. 8 Global system evaluation

– results of global evaluations are shown in Table 5 and report the percentage of correct retrieval for *Multimedia Query* and the comparison between the two queries to prove the improvement given by addition of geographic data. In this last evaluation, the artworks recognized after the crop operation are considered as correctly retrieved in both the queries.

The improvement given by the integration between geographic and multimedia data is clearly visible in the bar histogram in Fig. 8 and quantifiable with an improvement of 32.65% considering the correctly retrieved artworks of *Multimedia Query* and *Multimedia&Geographic Query*.

5 Conclusion and discussion

In this paper a system providing an efficient integration between multimedia and geographic data to retrieve information in the cultural heritage domain is proposed.

With this aim, three functionalities has been designed and developed: a *Multimedia Query*, a *Geographic Query* and a *Multimedia&Geographic Query*. The first of these query is used to evaluate the performances of various visual descriptors, in order to determine the best feature in the system implementation. The results obtained from the tests carried out on a dataset composed of one-hundred artworks located in Naples's museums show that the PHOG (Pyramid of Histograms of Orientation Gradients, [7]) feature can be considered the best possible choice.

The *Geographic Query* has been designed as a points of interest search and its development allows users to find the museums that are located close to him/her, using the device GPS sensor.

The *Multimedia Query* and *Multimedia&Geographic Query* functionalities, evaluations have been carried out ”on the field” in three Naples's museums. The results are displayed in terms of accuracy, both for each single museum and for global evaluation, and prove a real improvement given by the integration between Geographic Information Retrieval and Content-Based Image Retrieval. The values in Table 5 and in Fig. 8 show that the correct retrieval percentage obtained by the *Multimedia Query* is about 43% (this value is quite low because of the difficult conditions in which the pictures had to be taken), while the percentage obtained by the *Multimedia&Geographic Query* is about 75.5%, demonstrating that the addition of geographic data to multimedia data leads to a significant increase of the system accuracy percentage in term of 32.65%.

Appendix A: Dataset

The used dataset is show below togheter with the place of interest, the artwork name and the author.

– *Museo nazionale di Capodimonte*

1. Atalanta e Ippomene - Guido Reni;
2. Bottega del macellaio - Joachim Beuckelaer;
3. Carlo di Borbone visita il Papa Benedetto XIV nella coffee-house del Quirinale - Giovanni Paolo Pannini;
4. Crocifissione - Masaccio;

5. Danae - Tiziano Vecellio;
6. Ercole al bivio - Annibale Carracci;
7. Flagellazione di Cristo - Caravaggio;
8. Fondazione di Santa Maria Maggiore - Masolino da Panicale;
9. Giuditta che decapita Oloferne - Artemisia Gentileschi;
10. La Zingarella - Correggio;
11. Lucrezia romana - Parmigianino;
12. Madonna col Bambino e due angeli - Botticelli;
13. Napoleone I, imperatore - François Gerard;
14. Partenza di Carlo per la Spagna vista da terra - Antonio Joli;
15. Partenza di Carlo per la Spagna vista dal mare - Antonio Joli;
16. Pietà - Annibale Carracci;
17. Ritratto di Ferdinando IV - Anton Raphael Mengs;
18. Ritratto di Galeazzo Sanvitale - Parmigianino;
19. Ritratto di Paolo III - Tiziano Vecellio;
20. Trasfigurazione di Cristo - Giovanni Bellini.

– *Gallerie di palazzo Zevallos*

1. A Posillipo - Anton Sminck van Pitloo;
2. Adorazione dei Magi - Maestro dell'Annuncio ai pastori;
3. Agar e Ismaele nel deserto confortati dall'angelo - Francesco Solimena;
4. Autoritratto - Francesco Paolo Michetti;
5. Coppia di popolani in costume partenopeo - Anton Sminck van Pitloo;
6. Giuditta che decapita Oloferne - Louis Finson;
7. Il concerto - Gaspare Traversi;
8. Il coro della chiesa di Santa Maria Donnaregina Nuova - Domenico Battaglia;
9. La Zingara - Vincenzo Gemito;
10. Martirio di Sant'Orsola - Caravaggio;
11. Museo - Paolo Vetri;
12. Napoli via Toledo, impressioni di pioggia - Carlo Brancaccio;
13. Profeta - Vincenzo Gemito;
14. Ratto di Elena - Luca Gioradano;
15. Sansone e Dalila - Artemisia Gentileschi;
16. Tentazione - Gaetano Esposito;
17. Veduta di Napoli con largo di palazzo - Gaspar van Wittel;
18. Veduta di Roma con piazza del Popolo - Gaspar van Wittel;
19. Veduta di Roma con piazza Navona - Gaspar van Wittel;
20. Vitello - Anton Smick van Pitloo.

– *Museo nazionale di San Martino*

1. Da Frisio a Santa Lucia - Eduardo Dalbono;
2. Darsena di Napoli - Gaspar van Wittel;
3. La grotta di Posillipo - Gaspar van Wittel;
4. La principessa di Sant'Antimo - Francesco Hayez;
5. Napoli dal mare - Gaspar van Wittel;
6. Presunto ritratto di Masaniello - Onofrio Palumbo;

7. Prevetariello - Antonio Mancini;
8. Ritratto di Eduardo Scarpetta nelle vesti di Felice Sciosciammocca - Augusto Emilio Fabri;
9. Ritratto di Masaniello - Onofrio Palumbo;
10. San Girolamo - Jusepe de Ribeira;
11. Tavola Strozzi - unknown;
12. Tribunale della vicaria - Carlo Coppola.
13. La Giumenta - Giuseppe Renda;
14. Madonna col Bambino e San Giovannino - Pietro Bernini;
15. Malatiello - Vincenzo Gemito;
16. San Martino divide il mantello con il povero - Pietro Bernini.

– *Cappella Sansevero*

1. Ritratto di Vincenzo di Sangro - Carlo Amalfi.
2. Altare Maggiore (La Deposizione) - Francesco Celebrano;
3. Amor divino - unknown;
4. Cristo Velato - Giuseppe Sanmartino;
5. Decoro - Antonio Corradini;
6. Disinganno - Francesco Queirolo;
7. Dominio di se stessi - Francesco Celebrano;
8. Educazione - Francesco Queirolo;
9. Liberalità - Francesco Queirolo;
10. Macchine Anatomiche - Giuseppe Salerno;
11. Pudicizia velata - Antonio Corradini;
12. Santa Rosalia - Francesco Queirolo;
13. Sincerità - Francesco Queirolo;
14. Soavità del giogo coniugale - Paolo Persico;
15. Zelo della religione - Fortunato Onelli.

– *Museo archeologico nazionale di Napoli*

1. Diana - unknown;
2. Flora - unknown;
3. Leda - unknown;
4. Medea - unknown;
5. Ritratto di Paquio Proculo - unknown;
6. Venditrice di amorini - unknown.
7. Afrodite Sosandra - Calamide;
8. Atlante Farnese - unknown;
9. Battaglia di Isso (mosaic) - unknown;
10. Cave canem (mosaic) - unknown;
11. Doriforo - Policleto;
12. Ercole Farnese - Glycon ateniese;
13. Hermes in riposo - unknown;
14. Testa di cavallo (Testa Carafa) - Donatello;
15. Venere di Capua - unknown.

– *Museo civico di Castel Nuovo*

1. Abramo e i tre angeli - Pacecco De Rosa;
2. Adorazione dei Magi - Marco Cardisco;
3. Cesare Mormile e la rivolta napoletana del 1547 - Vincenzo Marinelli;
4. Ingresso di Garibaldi a Napoli - Franz Wenzel Schwarz;
5. Le stragi di Altamura - Michele Cammarano;
6. Madonna con Bambino e San Mauro Abate - Francesco Solimena;
7. Morte di San Giuseppe - Paolo De Matteis;
8. San Nicola in Gloria - Luca Giordano;
9. Veduta della nuova strada della Riviera di Chiaia - Carlo Vanvitelli.
10. Carlotta d'Asburgo a Miramare - Francesco Jerace;
11. Guappetiello - Francesco Jerace;
12. Il pescatorello - Vincenzo Gemito;
13. Testa di fanciullo - Vincenzo Gemito.

References

1. Adams B, McKenzie G, Gahegan M (2015) Frankenplace: interactive thematic mapping for ad hoc exploratory search. In: Proceedings of the 24th international conference on world wide web. ACM, pp 12–22
2. Agnello F, Corsale R, Franco V, Lo Brutto M, Midulla P, Orlando P, Villa B (2003) Cultural heritage and information systems, an investigation into a dedicated hypertext. Int Archives Photogrammetry Remote Sens Spatial Inform Sci 34(5/W12):7–12
3. Ai L-F, Jun-Qing Y, He Y-F, Guan T (2013) High-dimensional indexing technologies for large scale content-based image retrieval: a review. J Zhejiang University Sci C 14(7):505–520
4. Akgül CB, Rubin DL, Napel S, Beaulieu CF, Greenspan H, Acar B (2011) Content-based image retrieval in radiology: current status and future directions. J Digit Imag 24(2):208–222
5. Basanth Kumar HB (2015) An overview on content-based image retrieval
6. Bhatti A, Butt SM, Butt MM (2014) Visual feature extraction for content-based image retrieval. Sci Int 26(1):1–5
7. Bosch A, Zisserman A, Munoz X (2007) Representing shape with a spatial pyramid kernel. In: Proceedings of the 6th ACM international conference on image and video retrieval. ACM, pp 401–408
8. Carson C, Belongie S, Greenspan H, Malik J (2002) Blobworld: image segmentation using expectation-maximization and its application to image querying. IEEE Trans Pattern Anal Mach Intell 24(8):1026–1038
9. Chang S-F, Sikora T, Purl A (2001) Overview of the mpeg-7 standard. IEEE Trans Circuits Syst Video Technol 11(6):688–695
10. Chatzichristofis SA, Boutalis YS (2008) Cedd: color and edge directivity descriptor: a compact descriptor for image indexing and retrieval. In: International conference on computer vision systems. Springer, pp 312–322
11. Chatzichristofis SA, Boutalis YS (2010) Content based radiology image retrieval using a fuzzy rule based scalable composite descriptor. Multimed Tool Appl 46(2-3):493–519
12. Chatzichristofis SA, Fcth YSB (2008) Fuzzy color and texture histogram-a low level feature for accurate image retrieval. In: Ninth international workshop on Image analysis for multimedia interactive services, 2008. WIAMIS'08. IEEE, pp 191–196
13. Chatzichristofis S, Boutalis Y, Lux M (2009) Selection of the proper compact composite descriptor for improving content based image retrieval. In: Proceedings of the 6th IASTED international conference, vol 134643, p 064
14. Chen C-C, Wactlar HD, Wang JZ, Kiernan K (2005) Digital imagery for significant cultural and historical materials. Int J Digit Libr 5(4):275–286
15. Datta R, Li J, Wang JZ (2005) Content-based image retrieval: approaches and trends of the new age. In: Proceedings of the 7th ACM SIGMM international workshop on multimedia information retrieval. ACM, pp 253–262
16. Datta R, Joshi D, Li J, Wang JZ (2008) Image retrieval: ideas, influences, and trends of the new age. ACM Computing Surveys (Csur) 40(2):5

17. de Vries AP, Westerveld T (2004) A comparison of continuous vs. discrete image models for probabilistic image and video retrieval. In: 2004 international conference on Image processing, 2004. ICIP'04, vol 4. IEEE, pp 2387–2390
18. Deselaers T, Keysers D, Ney H (2004) Fire–flexible image retrieval engine: imageclef 2004 evaluation. In: Workshop of the cross-language evaluation forum for european languages. Springer, pp 688–698
19. Deselaers T, Keysers D, Ney H (2008) Features for image retrieval: an experimental comparison. Inform Retrieval 11(2):77–107
20. Frederix G, Caenen G, Pauwels EJ (2000) Panoramic, adaptive and reconfigurable interface for similarity search. In: 2000 international conference on Image processing, 2000. Proceedings, vol 3. IEEE, pp 222–225
21. Giuca A-M, Seitz KA, Furst J, Raicu D (2012) Expanding diagnostically labeled datasets using content-based image retrieval. In: 2012 19th IEEE international conference on Image processing (ICIP). IEEE, pp 2397–2400
22. Golubovic N, Krintz C, Wolski R, Lafia S, Hervey T, Kuhn W (2016) Extracting spatial information from social media in support of agricultural management decisions. In: Proceedings of the 10th workshop on geographic information retrieval. ACM, p 4
23. Hariharan R, Hore B, Li C, Mehrotra S (2007) Processing spatial-keyword (sk) queries in geographic information retrieval (gir) systems. In: 19th international conference on Scientific and statistical database management, 2007. SSBDM'07. IEEE, pp 16–16
24. Huang J, Ravi Kumar S, Mitra M, Zhu W-J, Zabih R (1997) Image indexing using color correlograms. In: 1997 Proceedings of the IEEE computer society conference on Computer vision and pattern recognition, 1997. IEEE, pp 762–768
25. Iqbal Q, Aggarwal JK (2002) Cires: a system for content-based retrieval in digital image libraries. In: 7th international conference on Control, automation, robotics and vision, 2002. ICARCV 2002, vol 1. IEEE, pp 205–210
26. Irtaza A, Jaffar AM, Aleisa E, Choi T-S (2014) Embedding neural networks for semantic association in content based image retrieval. Multimed Tool Appl 72(2):1911–1931
27. Irtaza A, Jaffar MA, Muhammad MS (2015) Content based image retrieval in a web 3.0 environment. Multimed Tool Appl 74(14):5055–5072
28. Ivanova K, Dobreva M, Stanchev P, Totkov G (2012) Access to digital cultural heritage: innovative applications of automated metadata generation plovdiv university publishing house paisii hilendarski
29. Ji Wx, Wang D, Hoi SCH, Pengcheng W, Zhu J, Zhang Y, Li J (2014) Deep learning for content-based image retrieval A comprehensive study. In: Proceedings of the 22nd ACM international conference on multimedia. ACM, pp 157–166
30. Karamti H, Tmar M, Visani M, Urruty T, Gargouri F (2017) Vector space model adaptation and pseudo relevance feedback for content-based image retrieval. Multimed Tool Appl:1–27
31. Kim J, Vasardani M, Winter S (2017) Similarity matching for integrating spatial information extracted from place descriptions. Int J Geogr Inf Sci 31(1):56–80
32. Koolen M, Kamps J, de Keijzer V (2009) Information retrieval in cultural heritage. Interdisc Sci Rev 34(2-3):268–284
33. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105
34. Kumar A, Kim J, Cai W, Fulham M, Feng D (2013) Content-based medical image retrieval: a survey of applications to multidimensional and multimodality data. J Digital Imag 26(6):1025–1039
35. Larson RR (1996) Geographic information retrieval and spatial browsing. Geographic information systems and libraries: patrons, maps, and spatial information [papers presented at the 1995 Clinic on Library Applications of Data Processing, April 10-12, 1995]
36. Lehmann TM, Güld MO, Deselaers T, Keysers D, Schubert H, Spitzer K, Ney H, Wein BB (2005) Automatic categorization of medical images for content-based retrieval and data mining. Comput Med Imaging Graph 29(2):143–155
37. Liew CL (2005) Online cultural heritage exhibitions: a survey of information retrieval features. Program 39(1):4–24
38. Longley P (2005) Geographic information systems and science. Wiley, New York
39. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110
40. Maliene V, Grigonis V, Palevičius V, Griffiths S (2011) Geographic information system: old principles with new capabilities. Urban Des Int 16(1):1–6
41. Mark DM (2003) Geographic information science: defining the field. Found Geographic Inform Sci 1:3–18

42. Müller H, Müller W, Squire DM, Marchand-Maillet S, Pun T (2001) Performance evaluation in content-based image retrieval: overview and proposals. Pattern Recogn Lett 22(5):593–601
43. Ng JY-H, Yang F, Davis LS (2015) Exploiting local features from deep networks for image retrieval. arXiv:1504.05133
44. Ohm J-R (2001) The mpeg-7 visual description framework—concepts, accuracy, and applications. In: Computer analysis of images and patterns. Springer, pp 2–10
45. Rasuli B, Asadi S (2017) Web geographic information retrieval: Literature review and conceptual model
46. Rinaldi AM (2011) Automatic web pages hierarchical classification using dynamic domain ontologies. Int J Knowl Web Intell 2(4):231–256
47. Rinaldi AM (2014) A multimedia ontology model based on linguistic properties and audio-visual features. Inf Sci 277:234–246
48. Ruigang F, Li B, Gao Y, Wang P (2016) Content-based image retrieval based on cnn and svm. In: 2016 2nd IEEE International Conference on Computer and Communications (ICCC). IEEE, pp 638–642
49. Shete DS, Chavan MS, Kolhapur K (2012) Content based image retrieval. Int J Emerging Technol Advan Eng 2(9):85–90
50. Shyu C-R, Klaric M, Scott GJ, Barb AS, Davis CH, Geoiris KP (2007) Geospatial information retrieval and indexing system—content mining, semantics modeling, and complex queries. IEEE Trans Geoscience Remote Sens 45(4):839–852
51. Siggelkow S, Schael Mx, Burkhardt H (2001) Simba—search images by appearance. In: Joint pattern recognition symposium. Springer, pp 9–16
52. Smeulders AWM, Worring M, Santini S, Gupta A, Jain R (2000) Content-based image retrieval at the end of the early years. IEEE Trans Pattern Anal Mach Intell 22(12):1349–1380
53. Squire D, Muller W, Muller H, Raki J (1998) Content-based query of image databases, inspirations from text retrieval: inverted files frequency-based weights and relevance feedback
54. Tangelder JWH, Veltkamp RC (2008) A survey of content based 3d shape retrieval methods. Multimed Tool Appl 39(3):441
55. Tsai C-F (2007) A review of image retrieval methods for digital cultural heritage resources. Online Inf Rev 31(2):185–198
56. Wang JZ, Li J, Wiederhold G (2001) Simplicity: semantics-sensitive integrated matching for picture libraries. IEEE Trans Pattern Anal Mach Intell 23(9):947–963
57. Zaila YL (2017) GEIR: a full-fledged geographically enhanced information retrieval solution. PhD thesis alma
58. Zheng Y-T, Zha Z-J, Chua T-S (2011) Research and applications on georeferenced multimedia: a survey. Multimed Tool Appl 51(1):77–98
59. Zheng L, Yi Y, Tian Q (2017) Sift meets cnn: a decade survey of instance retrieval. IEEE Trans Pattern Anal Mach Intell
60. Zhou XS, Huang TS (2003) Relevance feedback in image retrieval: A comprehensive review. Multimed Syst 8(6):536–544

Erasmo Purificato received the MSc degree from the Department of Electrical Engineering and Information Technologies, University of Naples, Federico II, Napoli, in 2017. His interests are about information retrieval, artificial intelligence and data mining.

Antonio M. Rinaldi has graduated in Computer Engineering at the University of Napoli Federico II and he received his PhD in Computer Science and Systems in 2005 from the same university, with a dissertation on the use of ontologies and semantic relatedness for information retrieval on the web. In the last years, he has been involved in several research and industrial projects related to information retrieval, multimedia database integration and geographic information systems. His fields of interest are: knowledge representation and retrieval, Big Data, multimedia database and GIS. Currently, he is an Assistant Professor at DIETI Dipartimento di Ingegneria Elettrica e delle Tecnologie dellInformazione and he is the Director of the IKNOS-LAB Intelligent and Knowledge Systems, LUPT, University of Napoli Federico II.