Expert Systems With Applications 91 (2018) 12–26

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Semantic Guided Interactive Image Retrieval for plant identification

Filipe Marcel Fernandes Gonçalves, Ivan Rizzo Guilherme, Daniel Carlos Guimarães Pedronette*

Department of Statistics, Applied Mathematics and Computing (DEMAC), State University of São Paulo (UNESP), Av. 24-A, 1515, Rio Claro, SP 13506-900, Brazil

ARTICLE INFO

Article history:
Received 1 March 2017
Revised 18 August 2017
Accepted 19 August 2017
Available online 21 August 2017

Keywords:
Interactive image retrieval
Unsupervised learning
Semantic gap
Ontology

ABSTRACT

A lot of images are currently generated in many domains, requiring specialized knowledge of identification and analysis. From one standpoint, many advances have been accomplished in the development of image retrieval techniques based on visual image properties. However, the semantic gap between low-level features and high-level concepts still represents a challenging scenario. On another standpoint, knowledge has also been structured in many fields by ontologies. A promising solution for bridging the semantic gap consists in combining the information from low-level features with semantic knowledge. This work proposes a novel graph-based approach denominated Semantic Interactive Image Retrieval (SIIR) capable of combining Content Based Image Retrieval (CBIR), unsupervised learning, ontology techniques and interactive retrieval. To the best of our knowledge, there is no approach in the literature that combines those diverse techniques like SIIR. The proposed approach supports expert identification tasks, such as the biologist's role in plant identification of Angiosperm families. Since the system exploits information from different sources as visual content, ontology, and user interactions, the user efforts required are drastically reduced. For the semantic model, we developed a domain ontology which represents the plant properties and structures, relating features from Angiosperm families. A novel graph-based approach is proposed for combining the semantic information and the visual retrieval results. A bipartite and a discriminative attribute graph allow a semantic selection of the most discriminative attributes for plant identification tasks. The selected attributes are used for formulating a question to the user. The system updates similarity information among images based on the user's answer, thus improving the retrieval effectiveness and reducing the user's efforts required for identification tasks. The proposed method was evaluated on the popular Oxford Flowers 17 and 102 Classes datasets, yielding highly effective results in both datasets when compared to other approaches. For example, the first five retrieved images for 17 classes achieve a retrieval precision of 97.07% and for 102 classes, 91.33%.

1. Introduction

The increasing image availability accessible through different technologies has demanded the development of effective retrieval and recognition methods. In this scenario, various image processing techniques have been developed and applied to digital media content (Arvor, Durieux, Andres, Laporte, 2013). Many recent advances have been made through the development of techniques that use quantitative features extracted by visual descriptors, capable of retrieving and indexing images. Most of these approaches are based on Content-Based Image Retrieval (CBIR) systems, which retrieve images by taking into account their visual content. The CBIR approaches consider various visual properties such as shape, texture, and color, extracted through global and local low-level features (Datta, Joshi, Li, & Wang, 2008; Kurtz, Depeursinge, Napel, Beaulieu, & Rubin, 2014; Lew, Sebe, Djeraba, & Jain, 2006). Recently, Convolutional Neural Networks (CNNs) have also been applied towards this goal with significant results (Hoi, Liu, & Chang, 2010; Jia et al., 2014; Razavian, Azizpour, Sullivan, & Carlsson, 2014). Therefore, the main aspects of such retrieval methods are based on feature extraction techniques by visual descriptors.

Besides the visual features, advances have been achieved in other stages of the retrieval pipeline. Approaches which exploit the user feedback through supervised learning methods have been integrated to CBIR techniques, improving the image retrieval effectiveness and adaptability to user inputs (Cheng, Jing, & Zhang, 2009; Liu, Liu, Qin, Ma, & Li, 2007b; Thomee & Lew, 2012). More recently, unsupervised learning has also attracted a lot of attention

* Corresponding author.
E-mail addresses: filipemfg@gmail.com (F.M.F. Gonçalves), ivan@rc.unesp.br (I.R. Guilherme), daniel@rc.unesp.br (D.C.G. Pedronette).

http://dx.doi.org/10.1016/j.eswa.2017.08.035
0957-4174/

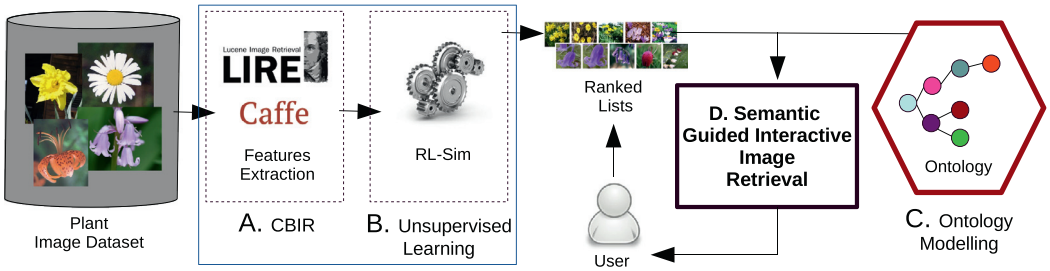


Fig. 1. Proposed retrieval approach for Plant Image Retrieval.

of the research community, once such methods exploit the dataset structure for improving the retrieval effectiveness, dispensing user interventions. In this scenario, unsupervised rank-based methods have been proposed achieving significant effectiveness gains (Bai & Bai, 2016; Bai, Bai, & Wang, 2015; Pedronette, Gonçalves, & Guilherme, 2017; Pedronette & da S. Torres, 2013; 2014).

Despite the continuous development of visual features, supervised, and unsupervised learning methods, retrieving relevant images based on the user needs still is a challenging task. The main challenge is to relate the semantic information of an image domain with the numerical values of low-level features recovered by pattern recognition algorithms. This problem refers to the semantic gap, which is defined as a lack of coincidence between the information that can be extracted from the visual content and the interpretation that the same data present to the user in a given situation (Datta et al., 2008; Smeulders, Worring, Santini, Gupta, & Jain, 2000). The semantic gap remains one of the most challenges of CBIR approaches, directly affecting the retrieval effectiveness.

On the other hand, ontologies have been widely used as a representation technique, allowing the reuse of knowledge since they transcribe a common understanding of a specific area. Ontologies declare explicit semantic, realizing significant statements and supporting the information sharing of attributes and relationships (Gruber, 1993; Guarino, 1998; Lacy, 2005). However, despite the recent advances, there is still a challenge to integrate techniques that use quantitative features with the semantics of structured knowledge representation in ontologies.

In addition, image analysis and identification tasks require specialized knowledge in many research fields, such as Systematic Botany. Traditionally, plant samples and field photographs are analyzed with many systematic descriptions, that allow the identification of organisms and their classification into groups. The identification of Angiosperms (plants with flowers and fruits) requires a vast knowledge of structures and properties of a specimen subject (Souza & Lorenzi, 2007). The identification task is even more challenging when performed solely from image sources since some plant regions are hidden. The image may not show, for example, internal structures in vegetable organs, such as the ovary. Nilsback and Zisserman state in Nilsback and Zisserman (2006) that image classification of flower branches is difficult even for humans, who need a complete knowledge of the domains. In this scenario, it is imperative the development of approaches for better representing the knowledge of many research fields in ontology structures, such that it can be interpreted and processed by both humans and machines.

In this paper, a novel interactive image retrieval approach is proposed aiming at bridging the semantic gap in plant identification tasks. The proposed approach, entitled *Semantic Interactive Image Retrieval* (illustrated in Fig. 1), consists of an automatic interactive system which combines Content Based Image Retrieval (CBIR) techniques, Unsupervised Learning, knowledge representation structured in Ontologies and interactive retrieval mechanisms. Given an image input defined by the user, the system extracts low-level features (Fig. 1A) and executes an unsupervised learning algorithm (Fig. 1B) in order to improve the retrieval results. Additionally, the system uses as an input the structured knowledge given by the ontology (Fig. 1C), which is defined by a domain specialist. The integration between the image retrieval results and the ontology knowledge constitutes the most relevant contribution of the proposed approach (Fig. 1D). The system exploits both information in order to establish a better interaction with the user, defined in terms of textual questions. A bipartite ontology graph and a discriminative attribute graph are proposed to select the most informative attributes from the ontology, capable of better discriminating the plant in the query image from those retrieved based on low-level features.

The proposed approach involves various research challenges of different areas. The main contributions of the paper are summarized in the following, enumerated according to Fig. 1:

- *A. CBIR and features extraction:* (*i*) extraction of visual features using recent CBIR and deep-learning frameworks (LIRE; Lux, 2013 and Caffe; Jia et al., 2014); (*ii*) evaluation of several features and identification of the most effective features for plant image retrieval tasks;
- *B. Unsupervised learning:* (*iii*) use and evaluation of a recent rank-based unsupervised learning method (RL-Sim; Pedronette & da S. Torres, 2013) in plant image retrieval to improve the effectiveness of initial retrieved results;
- *C. Ontology modeling:* (*iv*) the development of a systematic botany ontology, which describes and conceptualizes properties and structures of Angiosperm families;
- *D. Semantic Guided Interactive Image Retrieval:* (*v*) a graph-based integration approach which combines the retrieval results information with the structured knowledge given by the ontology; (*vi*) the proposal of a semantic-guided interactive image retrieval system, in which the questions presented to the user are defined according to the most discriminative attributes of the ontology.

The proposed approach was experimentally evaluated on the two popular datasets: Oxford Flowers with 17 and 102 Classes. Experimental results demonstrated that significant effectiveness gains can be obtained through the interactive retrieval process, indicating the decrease of effects of the semantic gap. The proposed method also yields very high effectiveness results in both datasets when compared to other approaches.

The paper is organized as follows: Section 2 discusses related work and Section 3 presents the CBIR techniques used in plant image retrieval (Fig. 1A). Section 4 discusses the unsupervised distance learning method (Fig. 1B) while Section 5, the ontology modeling (Fig. 1C). Section 6 presents in details the Semantic Guided Interactive Image Retrieval (Fig. 1D). Section 7 presents the experimental evaluation and Section 8 discusses the proposed approach. Finally, Section 9 presents the conclusion and directions for future work.

2. Related work

Currently, one of the main challenges in image and multimedia retrieval research is to reduce the semantic gap (Hui, Mohamad, & Ismail, 2010). An updated review of the problem is presented in Liu, Zhang, Lu, and Ma (2007a), where the authors discuss the technical state-of-the-art approaches to reduce such gap, dividing them into five categories.

The first category uses an ontology to define high-level concepts (Manzoor, Usman, Balubaid, & Mueen, 2015; Reddy & Bandikolla, 2008). The second one uses learning methods, like supervised or unsupervised learning, to associate low-level features and input concepts of a particular query (Liu et al., 2007b; Pedronette & da S. Torres, 2013). The third technique is based on relevance feedback in recovery loop for continuous learning (Kundu, Chowdhury, & Bulo, 2015; Kwan, Welch, & Foley, 2015). The fourth one consists in generating a semantic template to support high-level retrieval of images (Manzoor et al., 2015), while the fifth category uses a textual information obtained from the Web in order to retrieve image content from Web images (Feng & Chua, 2003; Reddy & Bandikolla, 2008).

Ontologies can assist in image retrieval by supplying a semantic model based on what occurs in the image (such as objects, events, etc.); or even enabling the association of images to the same concept through the use of URIs (Halaschek-Wiener, Schain, Grove, Parsia, & Hendler, 2005). The use of ontologies provides a common standard, thus allowing other individuals to process the contents of such previously annotated images (Coto, 2008).

Various authors (Manzoor et al., 2015; Pandey, Khanna, & Yokota, 2015; Reddy & Bandikolla, 2008; Vogel & Schiele, 2007) also evaluated a collection of images and presented semantic models for CBIR systems. In Manzoor et al. (2015), Manzoor compared the low-level features of images and inferred certain concepts such as colors. Their study further evaluated whether some concept defined in the ontology features such coloring. A ranking of the most relevant images that shared those characteristics and concepts were then displayed to the user aided by other extracted features from the image and some optional textual input.

Vogel and Schiele (2007) used a semantic model of natural landscapes, also defined by images: sky, grass, sand, among others; as well as the concept position within the image (i.e.: the sky is at the top). Each image segment was analyzed separately and then compared to previously defined concepts. It was possible to determine, from an established metric, which conceptual image was referenced, depending on the number of concepts in the landscape.

Reddy and Bandikolla (2008) presented an image retrieval approach by using textual information and Web image characteristics of the 2007 Cricket World Cup. An ontology was created with the concepts related to the championship. In their study, the authors evaluated pictures of different websites and extracted their low-level features, in addition to evaluating image labels. Then, if an image had an annotation, such as the name of a given cricket player, it could be inferred on the ontology that the player was the captain of a certain team.

Our bibliography survey found a lot of studies that only addressed the analysis of low-level plant images characteristics (Caballero & Aranda, 2010; Goëau et al., 2013; Kebapci, Yanikoglu, & Unal, 2009; Nilsback & Zisserman, 2006). However, there are few studies that addressed the issues mentioned in the semantic analysis of plant images (Walls et al., 2012). Much of this is due to the complexity of such images, since the plants have small structures and/or internal flower components that are not clearly shown in flower branches (Nilsback & Zisserman, 2006).

To reduce the semantic gap, improve the effectiveness of image retrieval and assist researchers interested in identifying Angiosperm families, we developed the proposed approach. We addressed the difficulties associated with plant identification by simply analyzing the low-level features of an image by proposing a Semantic Guided Interactive Image Retrieval, which employs an innovative integration system that combines Content Based Image Retrieval (CBIR), unsupervised learning, ontology information and interactive image retrieval mechanisms.

3. CBIR and features extraction

This section presents a formal definition of the image retrieval model and describes the techniques used to extract the low-level features from the images.

3.1. Image retrieval model

A general image retrieval model is considered for defining our approach. Let $\mathcal{C} = \{img_1, img_2, \ldots, img_n\}$ be an image collection, where each image represents a plant species and n is the size of the collection. Let $\rho(i, j)$ denotes a distance function between two images img_i and img_j, according to a given visual feature.

Based on the distance function ρ, a ranked list τ_q can be computed in response to a query image img_q, which also defines a plant species. The top positions of ranked lists are expected to contain the most similar images with regard to the query. The ranked list $\tau_q = (img_1, img_2, \ldots, img_{n_s})$ can be defined as a permutation of the subset $\mathcal{C}_s \subset \mathcal{C}$, which contains the most similar images to query image img_q, such that and $|\mathcal{C}_s| = n_s$. A permutation τ_q is as a bijection from the set $\mathcal{C}_s$ onto the set $[n_s] = \{1, 2, \ldots, n_s\}$. For a permutation τ_q, we interpret $\tau_q(i)$ as the position (or rank) of image img_i in the ranked list τ_q.

Based on each image feature, a distance matrix A can be computed, containing the distances among all images of the collection. We can also take every image $img_i \in \mathcal{C}$ as a query image img_q, in order to obtain a set $\mathcal{T} = \{\tau_1, \tau_2, \ldots, \tau_n\}$ of ranked lists for each image of $\mathcal{C}$. The objective of the unsupervised learning step consists in exploiting the contextual information encoded in the distances and the ranked lists for improving the retrieval results. Formally, it can be defined as function f_r which computes a new and more effective distance matrix $\hat{A} = fr(A, \mathcal{T})$.

3.2. Visual features

Various distinct visual properties are considered in the feature extraction process. The descriptors were made available through the LIRE (Lucene Image Retrieval) framework (Lux, 2013; Lux & Chatzichristofis, 2008). The framework consists in a recent open source Java library for CBIR, built based on indexing structures provided by the Apache Lucene textual retrieval engine. The library allows the extraction of image features, its storage and indexation for later retrieval (Lux, 2013; Lux & Chatzichristofis, 2008). Various recent techniques involving global and local features are available (Lux & Chatzichristofis, 2008). After indexing the images dataset, the distance between each pair of images is computed, such all images are compared to each other (Lux & Chatzichristofis, 2008). According to Lux and Marques, in Lux (2013), various metrics may be applied to compute the distance between images.

Convolutional Neural Network (CNN) features were also considered using the Caffe framework (Jia et al., 2014). CaffeNet was trained to recognize 1000 object categories and the features from the 7th fully connected layer (fc7) were used. The input images were resized to 256×256 pixels and the feature vectors have 4096 dimensions. Features were considered in the Euclidean space (L2 distance function).

Several global (color, texture) and local descriptors besides the CCN-Caffe were evaluated. This study presents only those de-

scriptors that have achieved higher effective results in plant image retrieval tasks: Auto Color Correlation (ACC) (Huang, Kumar, Mitra, Zhu, & Zabih, 1997) and Border/Interior Pixel Classification (BIC) (Stehling, Nascimento, & Falcão, 2002) as color descriptors; Speeded Up Robust Features (SURF) (Bay, Ess, Tuytelaars, & Van Gool, 2008) as a local descriptor, which is based on Bag of Visual of Words; and CNN-Caffe.

4. Unsupervised distance learning

The unsupervised distance learning step is performed by the RL-Sim Algorithm (Pedronette & da S. Torres, 2013), which is a recently proposed re-ranking and rank aggregation method used for improving the effectiveness of general image retrieval tasks. The RL-Sim Algorithm (Pedronette & da S. Torres, 2013) exploits contextual information encoded in the similarity between ranked lists aiming to improve the effectiveness in retrieval tasks. In general, if two images are similar, their ranked lists should be similar as well. In this way, ranked lists represent a relevant source of information, since they establish a relationship among a set of images contained in ranked lists, instead of only between pairs of images (in distance functions).

4.1. Ranking contextual distance measure

In this section, the RL-Sim Algorithm (Pedronette & da S. Torres, 2013) is described by using a ranking contextual distance measure based on similarity/dissimilarity of ranked lists.

The ranking contextual distance measure is iteratively learned in an unsupervised setting, by incorporating the contextual information provided by rank correlation measures.

Let us consider the neighborhood set $\mathcal{N}(i,k)$ of an image img_i, which contains the k most similar images to img_i, according to a given distance (say ρ defined by the image descriptor). The set $\mathcal{N}(i,k)$ can be obtained by the well-known k-Nearest Neighbor approach, where the cardinality of the set is denoted by $|\mathcal{N}(i,k)| = k$.

Let $d(\tau_i, \tau_j, k)$ denote a rank correlation measure between ranked lists τ_i and τ_j, considering their top-k positions given by the sets $\mathcal{N}(i)$ and $\mathcal{N}(j)$. The rank correlation measure considered is based on the intersection between ranked lists (Pedronette & da S. Torres, 2013). A non-iterative contextual distance measure $\rho_c(img_i, img_j)$ based on the comparison of ranked lists τ_i, τ_j can be defined as follows:

$$\rho_c(img_i, img_j) = d(\tau_i, \tau_j, k) \tag{1}$$

Based on the conjecture that the contextual distance measure ρ_c represents a more effective distance between images, the distance among all images in a collection can be recomputed based on this measure. Therefore, a new set of ranked lists can be obtained, such that the contextual distance can also be recomputed and the process can be repeated in an iterative way. Let $^{(t)}$ denote the current iteration and let $\tau_i^{(t)}$ denote the ranked list at iteration t. Let $\rho_c^{(0)}$ be the contextual distance at first iteration, which is equal to the distance defined by the image descriptor, such that $\rho_c^{(0)}(img_i, img_j) = \rho(img_i, img_j)$ for all images $img_i, img_j \in \mathcal{C}$. The iterative contextual measure is defined as:

$$\rho_c^{(t+1)}(img_i, img_j) = d(\tau_i^{(t)}, \tau_j^{(t)}, k) \tag{2}$$

It is expected that the effectiveness of the distance measure improves along iterations, so non-relevant images are moved out from the first positions of the ranked lists. In this way, the size of the neighborhood k can be increased for considering more images along iterations. Therefore the contextual measure can be redefined as:

$$\rho_c^{(t+1)}(img_i, img_j) = d(\tau_i^{(t)}, \tau_j^{(t)}, k+t) \tag{3}$$

After a given number of T iterations, a new distance $\hat{\rho}$ is computed based on contextual distance measure ρ_c:

$$\hat{\rho}(img_i, img_j) = \rho_c^{(T)}(img_i, img_j) \tag{4}$$

Finally, using the distance $\hat{\rho}$, a new distance matrix can be computed such $\hat{A}_{ij} = \hat{\rho}(img_i, img_j)$, providing more effective retrieval results.

4.2. Rank aggregation

The RL-Sim Algorithm (Pedronette & da S. Torres, 2013) can also be used for combining different visual features, which can provide complementary visual information. Each visual feature gives rise to a distance matrix composing a set of matrices $\{A_1, A_2, \ldots, A_p\}$.

The RL-Sim Algorithm (Pedronette & da S. Torres, 2013) combines the set of matrices in a unique matrix A_c using a multiplicative approach. Each position of the combined matrix is computed as follows:

$$A_{c_{ij}} = \prod_{l=1}^{p}(1 + A_{l_{ij}}). \tag{5}$$

Given a combined distance matrix A_c, a new set of ranked list is computed and submitted to the original unsupervised distance learning algorithm.

5. Ontology modeling

The ontology modeling was defined with a specific vocabulary that represents concepts related to morphological structures, relationships and constraints of each Angiosperm family in a level that was enough to differentiate all of the studied plants.

The modeling of the Angiosperm ontology was developed using Protege (Knublauch, Fergerson, Noy, & Musen, 2004) in OWL (Web Ontology Language), following the Methontology procedures (Fernández-López, Gómez-Pérez, & Juristo, 1997) and APG III definitions (Angiosperm Phylogeny Group), as well as existing Systematic Botany bibliography (Souza & Lorenzi, 2005; 2007).

The whole process of modeling the Angiosperm families involved a thorough analysis of plant parts and structures. The concepts were categorized into classes based on their common features. A class is defined by a series of properties. The basic condition for belonging to a given class is to have all of those properties.

Several specific relationships (object properties and datatype properties) were also developed to relate classes defined in the ontology. This process allowed to determine the domain – which holds the relationship; and the range – which would be the target classes of said property; thereby increasing the expression of semantic relationships in the ontology.

The modeling also follows the pattern described in Botany literature, in which a particular structure is specific to an organ. For example, it is commonly said that a flower has a pistil, but the pistil is part of the gynoecium (Fig. 2). Thus, a flower indirectly has a pistil, since a flower has a gynoecium, which in turn presents a pistil. Still, a flower only has a pistil if it is female or hermaphrodite, since the pistil presence obligatorily requires a gynoecium.

The Fig. 2 exemplifies part of the ontology modeling in the *Ranunculaceae* family – which features an apocarpous gynoecium. The *Flower* class has an object property "HasFlowerStructure" (whose domain is *Flower* and the range is *FlowerStructure*), which relates to the *Flower Structure* class. The *Gynoecium* class, which is a *Floral Whorls*, means that *Gynoecium* is a *Floral Whorls* subclass. *Ranunculaceae* family particularly has an *Apocarpous Gynoecium*, which is implemented as a specific type of gynoecium as a *Gynoecium* subclass.

The developed ontology defines 250 classes, 45 object properties and 1 datatype property for Oxford Flowers 17 Classes.

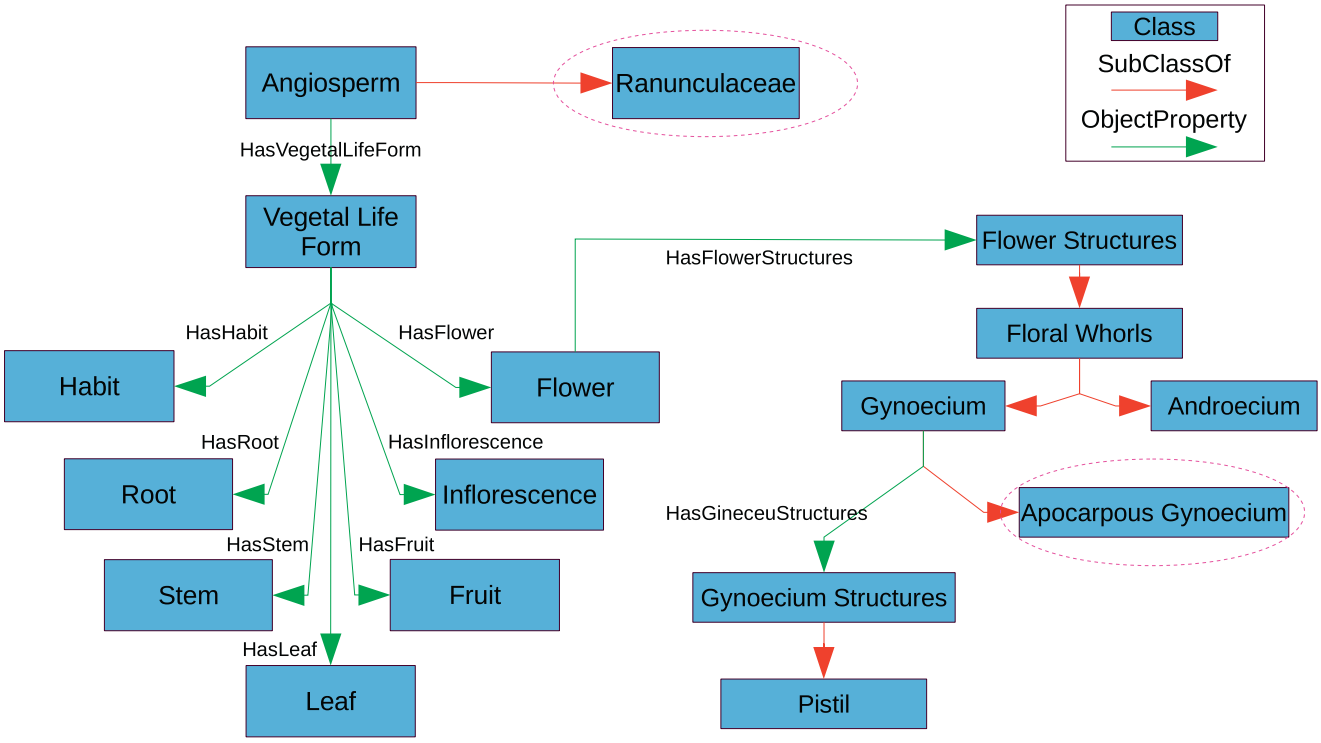


Fig. 2. Ontology modeling for Ranunculaceae family.

While for the 102 Classes dataset, the ontology defines 429 classes, 66 object properties and 2 datatype properties. Constraint and class relationships with their respective properties and specific attributes allowed to model the basic features of each studied family and thus defined concepts linked to the ontology domain.

6. Semantic Guided Interactive Image Retrieval

In this section, we describe the main contribution of this paper. This approach represents the combination of the low-level feature information with the knowledge modeling process, by guiding a user with discriminative questions in order to improve the results retrieved.

Since the mid-1990s, interactive mechanisms have been used in image retrieval systems to optimize a similarity metric and to iteratively correct errors made by the CBIR system (Kundu et al., 2015). Interactive Image retrieval information is then used to modify the weights of the combination to reflect different feature relevance (Giacinto, 2007). When interactive methods are used, the search is considered an iterative process in which the original query is refined interactively, to progressively obtain more accurate results (Arevalillo-Herraez & Ferri, 2013).

One way to bridge the gap between the low-level features of the images and the high-level semantic concepts is through user interaction (Guan & Qiu, 2007). The main challenge of the proposed approach is to identify relevant information related to the query image and visual retrieval results, in order to define the interaction with the user. In this way, our method allows the selection of most discriminative attributes for distinguishing families present in ranked lists. As a result, the retrieval and classification results can be improved supported by the user responses. Fig. 3 illustrates the organization of the proposed interactive approach.

The problem was split into five stages, according to Fig. 3 and discussed as follows:

1. The first step consists in retrieving the entire information about Angiosperm families from the modeled ontology, represented in terms of a *Bipartite Ontology Graph* (Fig. 3A, detailed in Section 6.1);
2. The second one computes a *Family Rank-Based Histogram* of Angiosperm families based on the retrieved results, aiming at identifying the most frequently families (Fig. 3B, Section 6.2);
3. The third step consists in combining both semantic and low-level visual information into a *Discriminative Attribute Graph*, which provides a structure for determining the most discriminative attributes for identification purposes (Fig. 3C, Section 6.3);
4. The fourth step performs the selection of the most discriminative attribute by exploiting the graph structure constructed in the previous stage (Fig. 3D, Section 6.4);
5. Once an attribute is selected, a question is formulated and showed to the user. Given the answer provided by the user, the retrieval result is then updated and improved for the next iteration (Fig. 3E, Section 6.5).

6.1. Bipartite Ontology Graph

A graph-based approach is proposed with the objective of representing the semantic knowledge encoded in the ontology. In this way, a Bipartite Ontology Graph (BOG) is proposed for defining the relationships among each Angiosperms family and the attributes from biological structures which compose them. Fig. 3A illustrates the proposed graph-based approach. Based on the graph representation, the semantic knowledge encoded in the ontology can be exploited by the retrieval process, guiding the interactions with the user.

Formally, the Bipartite Ontology Graph can be defined as an undirected graph $G_o = (V_o, E_o)$. Let $F = \{f_1, f_2, \ldots, f_r\}$ be a set of Angiosperms families being analyzed. Let $A_t = \{a_1, a_2, \ldots, a_m\}$ be a set of attributes which represent plant properties modeled by the ontology. The graph nodes V_o are defined as a union of such sets, $V_o = F \cup A_t$.

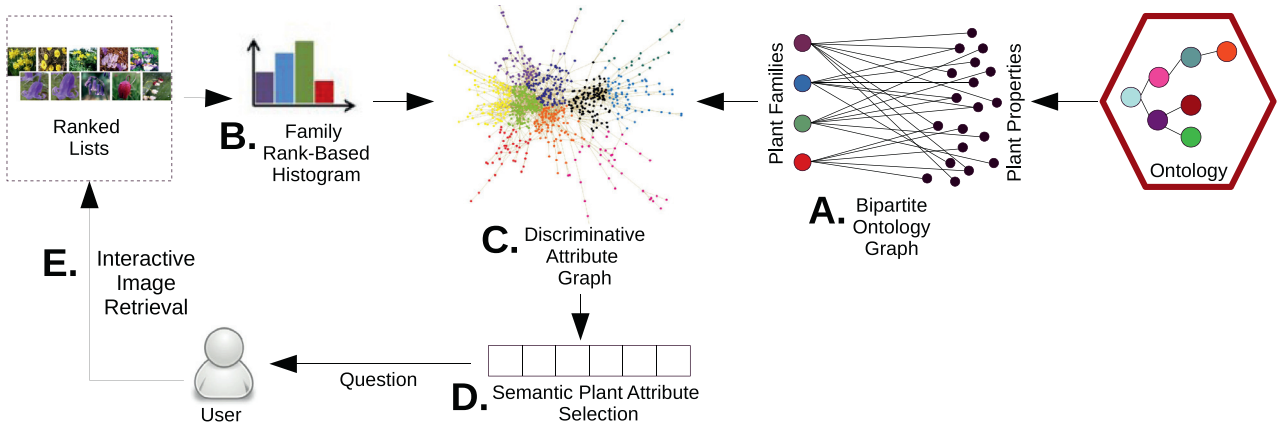


Fig. 3. Semantic Guided Interactive Image Retrieval for plant identification.

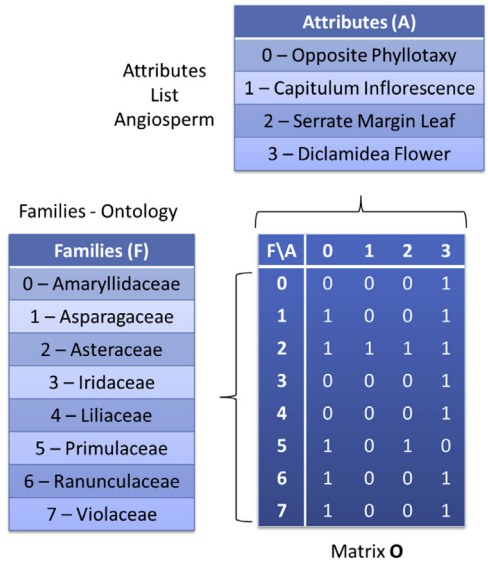


Fig. 4. Computation of adjacency matrix **O** – Bipartite Ontology Graph.

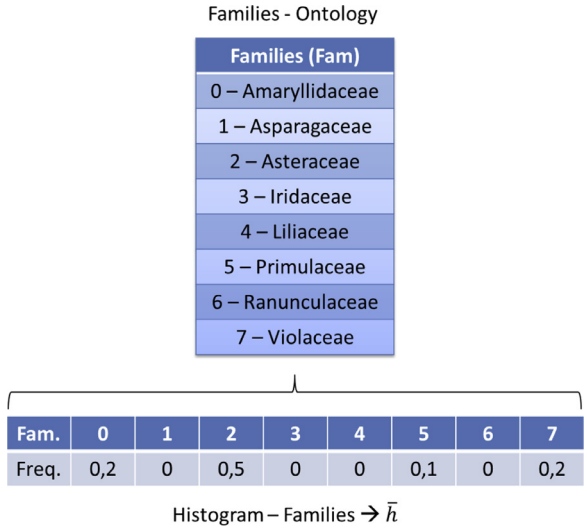


Fig. 5. Histogram for a hypothetical ranked list.

An $|F| \times |A_t|$ ontology matrix **O** is used as an adjacency matrix for defining the set of edges E_o of the graph. The matrix **O** is computed as:

$$O_{ij} = f_o(f_i, a_j), \tag{6}$$

where f_o is a binary function computed based on the ontology information which returns 1, if the family f_i owns the attribute a_j, and 0 otherwise.

An edge e_{ij} between a family f_i and an attribute a_j indicates the presence of such attribute for the family. Technically, the edge e_{ij} is computed through the OWL API, responsible for querying the ontology.

Fig. 4 shows an example on how the matrix **O** is computed based on the information collected in the modeled ontology. The value "1" for the element $o_{2,\,1}$ indicates the presence of a *Capitulum Inflorescence* on the *Asteraceae* family.

6.2. Family rank-based histogram

While the ontology graph represents the semantic information, a structure for modeling the low-level visual information is also required. The ranked lists computed by the unsupervised distance learning step are exploited with this objective.

An analysis is conducted considering the top-k positions of ranked lists computed for each query image img_q. The objective is to compute the frequency of each family in the retrieved results. A histogram is computed with this purpose, as illustrated in Fig. 3B. The number of bins is defined according to the number of families analyzed, as $|F|$.

The histogram is formally defined as follows: let $\mathcal{N}(q,k)$ be a neighborhood set which retrieves the k-nearest neighbors of image img_q, such that $|\mathcal{N}(q,k)| = k$. Let $h(i)$ be the number of images from the family f_i in the neighborhood set $\mathcal{N}(q,k)$. Let $\bar{h}(i)$ be the normalized frequency of family f_i, which can be defined as $\bar{h}(i) = h(i)/k$. We can also say that:

$$\sum_{i=1}^{r} \bar{h}(i) = 1, \tag{7}$$

where r represents the number of families, such that $r = |F|$. In this way, the histogram $\bar{h}$ is proportionally defined according to the frequency of each family at top positions, providing a summarized information extracted from visual characteristics.

Fig. 5 presents the histogram of the families on the top-10 positions of a ranked list. The *Asteraceae* family has 5 images on the first 10 positions of the rank, resulting in a $\bar{h}[2] = 0.5$.

6.3. Discriminative Attribute Graph

Each plant attribute has a distinct potential for identifying plant families. For example, an attribute which is present in many fam-

Matrix **O**

F\A	0	1	2	3
0	0	0	0	1
1	1	0	0	1
2	1	1	1	1
3	0	0	0	1
4	0	0	0	1
5	1	0	1	0
6	1	0	0	1
7	1	0	0	1

Histogram – Families → $\bar{h}$

Fam.	0	1	2	3	4	5	6	7
Freq.	0,2	0	0,5	0	0	0,1	0	0,2

$$d_{ij} = \sum_{c=0}^{7} |o_{ij} - o_{cj}| \times (\bar{h_i} \times \bar{h_c})$$

Example Calculus – Matrix **D**

$$d_{0,0} = 1 \times (0{,}2 \times 0) + 1 \times (0{,}2 \times 0{,}5) + 0 + 0 + 1 \times (0{,}2 \times 0{,}1) + 1 \times (0{,}2 \times 0) + 1 \times (0{,}2 \times 0{,}2)$$

$$d_{0,0} = 0{,}16$$

Fig. 6. Computation of adjacency matrix **D** – Discriminative Attribute Graph.

F\A	0	1	2	3
0	0,16	0,1	0,12	0,02
1	0	0	0	0
2	0,1	0,25	0,2	0,05
3	0	0	0	0
4	0	0	0	0
5	0,02	0,05	0,04	0,09
6	0	0	0	0
7	0,04	0,1	0,12	0,02

Matrix **D**

Fig. 7. Matrix **D** fully computed.

ilies can be useless for identification purposes. Additionally, such discriminative potential can vary according to the most frequent families at the top retrieved images.

In this way, the Discriminative Attribute Graph (DAG) aims at providing a graph structure for determining the most discriminative attributes for identification purposes. The main idea consists in combining both semantic and low-level visual information into a single graph. In fact, such step consists in one challenging task of the proposed approach, where the semantic gap problem is addressed.

The Discriminative Attribute Graph (DAG) combines information from the Bipartite Ontology Graph (BOG – Fig. 3A) and the Family Rank-Based Histogram (Fig. 3B). The proposed approach is illustrated in Fig. 3C. While the histogram identifies the most frequent families at top retrieved images, the BOG graph is analyzed for discovering the most appropriated attributes for discriminating such families.

Formally, the DAG graph can be defined as an undirected graph $G_d = (V_d, E_d)$. The set of nodes V_d is defined in the same way of the BOG graph, as a union of sets of families and attributes, such $V_d = F \cup A_t$. The set of edges E_d is defined by an adjacency matrix **D**.

An edge e_{ij} between a family f_i and an attribute a_j indicates the capacity of such attribute for discriminating the family f_i from the other families. Additionally, the value of the edge is weighted by the frequency of families in top retrieved images, given by the histogram $\bar{h}$. In this way, adjacency matrix **D** is computed as follows:

$$D_{ij} = \sum_{c=1}^{r} |O_{ij} - O_{cj}| \times (\bar{h}(i) \times \bar{h}(c)). \quad (8)$$

Fig. 6 shows how to calculate, based on the equation above, the element $d_{0,\,0}$ from the examples in Figs. 4 and 5. After computing all elements, matrix **D** is illustrated in Fig. 7. It is important to

Matrix **D**

F\A	0	1	2	3
0	0,16	0,1	0,12	0,02
1	0	0	0	0
2	0,1	0,25	0,2	0,05
3	0	0	0	0
4	0	0	0	0
5	0,02	0,05	0,04	0,09
6	0	0	0	0
7	0,04	0,1	0,12	0,02

$$s(a_j) = \sum_{i=0}^{7} d_{ij}$$

Atrib.	0	1	2	3
Value	0,32	0,50	0,48	0,18

Semantic Attribute Selection - s

Fig. 8. Computation of the accumulated adjacency.

notice that elements with value "0" refer to those families that are not presented on the top-10 positions of the ranked list (Fig. 5).

6.4. Semantic Attribute Selection

The DAG graph combines semantic and low-level information for identifying the most discriminative attribute for each family (Fig. 3C). In this way, the adjacency information can be used for identifying the most appropriated attribute to be used in the interactive image retrieval step (Fig. 3D). The most discriminative attribute is given by the node a_j which presents the greater accumulated adjacency (i.e.: the attribute which can be exploited for differentiating the most frequent families).

Formally, a function $s(a_j)$ is computed defining the sum of the adjacencies of a given attribute a_j. The accumulated adjacency $s(a_j)$ is computed based on the adjacency matrix **D**, as follows:

$$s(a_j) = \sum_{i=1}^{m} d_{ij} \quad (9)$$

Fig. 8 illustrates the computation of the accumulated adjacency in order to select the most discriminative attribute.

After this step, the attributes are sorted in decreasing order in relation to their accumulated adjacency value. The attribute which presents the greater accumulated adjacency is selected for com-

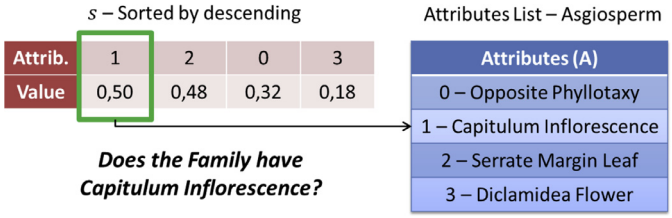


Fig. 9. Choosing a question for the user.

posing an interactive retrieval step. A question is created for asking the user with regard to the presence of the selected attribute in the query image, as illustrated in Fig. 9.

One advantage of the proposed method consists in the capacity of selecting the most discriminative attributes in order to reduce the number of questions and interactions required. In this way, the proposed approach can achieve more effective results requiring fewer user efforts than traditional biological approaches, as a dichotomous key.

6.5. Interactive image retrieval

In the interactive image retrieval step (Fig. 3E), given a selected attribute a_j (Fig. 3D), the system composes a question for the user. The user answers "yes", "no" or "do not know" to questions, triggering a different system feedback for each situation. The answer given by the user is used for updating and improving the retrieval results.

For example, let a_j be the selected attribute for a particular experiment which indicates the presence of *Capitulum Inflorescence*. The user would be asked:

"*Does the family have a capitulum inflorescence?*"

Suppose the queried image was referring to the *Asteraceae* family and the answer would be "yes". Thus, all images from families that have a *Capitulum Inflorescence* would receive a distance updating. The distances to such images would be decreased and they would be moved to top positions of ranked lists, improving the quality of results retrieved.

The approach used for the distance updating is based on a multiplication by a constant $\alpha < 1$. Let img_q be a query image. Let img_j denotes an image which has a certain attribute a_j. The answer "yes" for the presence of the attribute a_j implicates the decrease of the distance between img_q and img_j. Therefore the distance matrix A is updated as follows:

$$A_{qj} = A_{qj} \times \alpha. \tag{10}$$

If the answer regarding of the presence of a certain attribute is "no", the distance updating follows the same principle. Let img_n be an image which does not have the attribute, the matrix A is updated as:

$$A_{qn} = A_{qn} \times (1 + \alpha). \tag{11}$$

However, the answer "no" is inconclusive, since a family may present more than one feature for the same category. This may occur, if the queried image is, for example, a *Primulaceae* plant. This family may have either a *Raceme, Cymose* or *Panicle Inflorescence*. If the chosen attribute a_j is a *Raceme Inflorescence* and the image does not have such feature (i.e.: has another inflorescence type, such as *Cymose* or *Panicle*), the distance should not affect the ranked lists in a strong way. Therefore, the value used for the constant α is very small.

The final situation occurs when the user "do not know" the answer. In this case, the next attribute with the greater accumulated adjacency is used for composing a new question.

7. Experimental evaluation

An experimental evaluation was conducted aiming at assessing the effectiveness of the presented approach. Section 7.1 discusses the experimental protocol and Section 7.2 describes the datasets considered. Section 7.3 shows a visual evaluation of the proposed approach. Section 7.4 discusses the evaluation of visual features and unsupervised learning while Section 7.5 describes the experimental results of the proposed interactive approach.

7.1. Experimental protocol

The evaluation considers an experimental protocol mainly based on retrieval tasks, in which all dataset images are considered as query images. Various effectiveness measures are reported: the precision at different depths (P@5, P@10), the Mean Average Precision (MAP) and the *Precision* × *Recall* curve (PR curve) before and after the use of the proposed approach. In order to allow a deeper experimental analysis and comparisons with other methods, classifications tasks are also considered. A *k*NN classifier built upon the retrieval results is evaluated by the accuracy of the recognition rate obtained.

For both retrieval and classification tasks, the first steps involved in the evaluation are the same. Initial experiments aim at evaluating the effectiveness of visual features and the impact of unsupervised learning step. The visual features are extracted, the distances among images are computed and ranked lists are obtained. Subsequently, the unsupervised learning step is performed by the RL-Sim algorithm, considering isolated features and aggregation of different features. In order to evaluate the impact of parameters of RL-Sim algorithm, an analysis is performed varying the number of iterations in the range of 1–3 and the neighborhood size k in the range 5–35 (in intervals of 5). The retrieval results obtained before and after the RL-Sim algorithm are evaluated by effectiveness measures, as precision, recall, and MAP. After the parameters definition, the execution of the unsupervised learning algorithm is performed once for the whole dataset and used by the next steps. Since no label information is used, the retrieval results obtained at this stage can be shared by all query images.

Next, various experiments were conducted to assess the effectiveness of Semantic Interactive Image Retrieval (Section 6). For this stage, the interactive retrieval process is evaluated independently for each query image. Ground-truth information used to simulate the user's responses is based on the Bipartite Ontology Graph (Section 6.1), which encodes information about the presence or absence of attributes modeled for the plants. Such information is available for the whole dataset, excluding the query image. The number of users interactions ranged from 1 to 10 questions and the evolution of results are evaluated for each iteration. The reported results represent the average of the measures obtained for all query images, constituting a leave-one-out cross validation. Both retrieval and classification tasks are considered. For retrieval, precision, MAP and PR curves are reported as effectiveness measures. For classification, the accuracy of *k*NN classifier is considered.

7.2. Datasets

In order to evaluate the proposed approach, two popular flowers datasets were considered. Firstly, the Oxford Flowers 17 Classes dataset (Nilsback & Zisserman, 2006), which contains 17 classes from different Angiosperm species. Each class has 80 images, totalizing 1360 images in the dataset.

We also used the Oxford Flowers 102 Classes dataset (Nilsback & Zisserman, 2006). This dataset contains 102 classes from different Angiosperm species and each class has a different number of

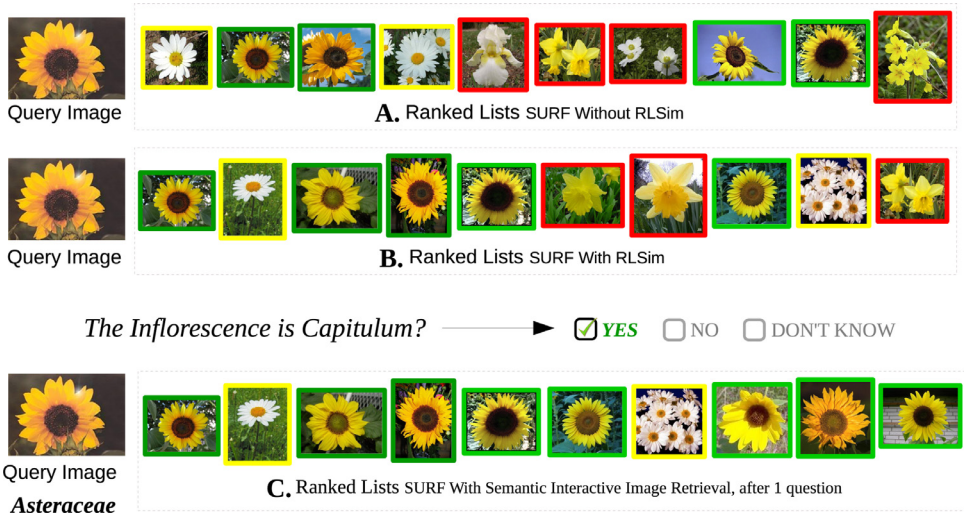


Fig. 10. Ranked lists behavior after the use of the proposed approach (SIIR).

images, varying from 40 to 251 images per class. In this way, the 102 classes dataset presents a total of 8189 images.

The 17 classes dataset represents 8 Angiosperm families. Each family was modeled on the developed ontology as well as their specific characteristics that allow their identification in a total of 132 restrictions. The 102 classes dataset represents 47 Angiosperm families. All of these families were modeled on the ontology, so as their specific morphological characteristics, which presents a total of 350 restrictions.

7.3. Visual evaluation

The capacity of the proposed approach in improving the plant identification tasks is illustrated in this section. Fig. 10 shows a real case which highlights the effectiveness improvements obtained by the proposed approach for a given ranked list. Borders in red represent incorrect results considering the images classes (species), while borders in green show the correct retrieved results for each step. Borders in yellow represents an incorrect retrieved results considering species, but a correct result considering the information from the families.

The query image had its features extracted by SURF, and the *Rank A* presents the top-10 similar images according to this descriptor. According to Fig. 10A, P@10 accounts for only 40% of the correct classification, since it has four images that belong to the class of the query image. The precision of the families in the top-10 (PF@10) represents a value of 60%

Fig. 10B shows the results of the ranked list for the query image after the execution of the RL-Sim algorithm (k=20, t=2). It can be observed that P@10 increased after this step, when compared with the case in Fig. 10A, since its value represents now 50% at the top-10 first positions in this rank.

Fig. 10C shows the results of the query image after 1 question and suggests to the user that the family is *Asteraceae*. It can be noticed that the proposed approach reaches P@10 with 80% of class precision and PF@10 with 100% of family precision. Notice that with only 1 simple question, our approach improved the precision with a huge gain when compared with the initial ranks in Fig. 10A and B.

The question "*The inflorescence is Capitulum?*" was chosen by the Semantic Attribute Selection once the question is the most informative for distinguishing the images in *Rank B*. Additionally, it is an easy question to solve, since its concepts are known by biologists and botany enthusiasts, and represents a concept that can be visually observed in the image.

7.4. Visual features and unsupervised learning

This section presents the experimental results obtained by: (*i*) visual features; (*ii*) visual features + unsupervised learning; (*iii*) fusion of visual features through rank aggregation.

The experimental results demonstrate the importance of combining different features through unsupervised learning. Combining different approaches of visual features achieved the higher effective results in retrieval tasks. We experimentally evaluated more than 19 features (color, texture, and local descriptors) and 1 feature based on Convolutional Neural Networks (CNN). By combining a color, a local and a CNN feature, we achieved the highest effective results. Such positive results are mainly due to the complementarity among diverse features and the capacity of the rank aggregation based on unsupervised learning of combining in an effective manner. It is worth mentioning the importance of improving the initial retrieved results, since the ranks will be used for the next steps of the proposed approach.

After features extraction evaluation, three descriptors (ACC, BIC and SURF) were selected, as well as the Caffe framework (CNN) to proceed with the image analysis. This criterion was set because these tools presented the best results for the effectiveness metrics. Since we applied the proposed approach in two datasets, the results of each one are shown in two distinct Sections 7.4.1 and 7.4.2.

7.4.1. Oxford Flowers – 17 Classes

In this section, we present the results for different features with and without the unsupervised learning (UL) step on the Oxford 17 classes dataset. Table 1 summarizes the initial retrieval results (without UL) and the considering the unsupervised learning with the best parameters settings for each feature.

It can be noticed in Table 1 that Caffe framework (CNN), when isolated (without Rank Aggregation), shows the best results among the three metrics for visual features extraction method and for the Unsupervised Learning. These results reach 87.71% on ranking precision for the five first positions after the execution of the RL-Sim algorithm.

Considering the other three image descriptors (ACC, BIC and SURF), it can be noticed that SURF presents the best results. For

Table 1
Effectiveness results for various features and unsupervised learning (UL) on the Oxford 17 Classes dataset.

Effectiveness metrics	ACC	BIC	SURF	ACC+SURF	CNN	CNN+SURF
P@5 without UL	0.5310	0.6001	0.5204	0.6085	0.8569	0.8859
P@5 + UL	0.5394	0.6001	0.5518	0.6278	0.8771	**0.9138**
P@10 without UL	0.4215	0.5015	0.4184	0.5071	0.7959	0.8207
P@10 + UL	0.4440	0.5104	0.4530	0.5439	0.8411	**0.8921**
MAP without UL	0.1928	0.2625	0.2155	0.2466	0.5025	0.4804
MAP + UL	0.2410	0.3097	0.2391	0.3216	0.7023	**0.7485**

Table 2
Effectiveness results for various features and unsupervised learning (UL) on the Oxford 102 Classes dataset.

Effectiveness metrics	BIC	SURF	CNN	CNN+SURF+BIC
P@5 without UL	0.5399	0.3661	0.5751	0.7455
P@5 + UL	0.5432	0.4201	0.6009	**0.8020**
P@10 without UL	0.4269	0.2624	0.4816	0.6491
P@10 + UL	0.4353	0.3173	0.5292	**0.7431**
MAP without UL	0.1766	0.0979	0.1871	0.2872
MAP + UL	0.1905	0.1219	0.2645	**0.4326**

a combination of two features, a significant improvement can be observed. The fusion of ACC+SURF shows an increase of approximately 16.39% for its best P@5 (0.6278; $k = 35,\ t = 1$), when compared to the best result presented by ACC in this metric (0.5394; $k = 10,\ t = 1$). Considering the aggregation of features that presented the best retrieval results (CNN+SURF), the gains in the effectiveness measures analyzed are even more significant.

The use of unsupervised learning through the RL-Sim algorithm achieved a major advancement for the MAP gain for CNN+SURF in 55.8%, when comparing the value of the best MAP of the union CNN+SURF (0.7485; $k = 35,\ t = 2$) with the map without the application of RL-Sim to the same union (0.4804).

7.4.2. Oxford Flowers – 102 Classes

This section presents the results of the visual features extraction methods and the unsupervised learning for 102 classes. Table 2 summarizes the initial retrieval results (without UL) and the unsupervised learning (UL) with the best parameters settings for the Oxford Flowers 102 Classes dataset.

As we can see in Table 2, the results are lower than those presented by the 17 classes. It occurs since the 102 classes dataset presents a large number of classes, compared with 17 classes, which difficult the retrieval of the correct results. Table 2 also presents the aggregation results of three extraction methods (CNN+SURF+BIC), since this combination shows the best retrieved results for the 102 classes dataset. We also combined CNN+SURF, CNN+BIC, BIC+SURF and also ACC with these selected methods, but the results were lower than those presented by the union CNN+SURF+BIC.

The results of aggregation of the three methods show how the unsupervised learning assists the improvement of the proposed approach. The initial MAP results of CNN+SURF+BIC is 0.2872, while the results obtained after the unsupervised learning is 0.4326, representing a gain of 50.63%. Although SURF presented lowest results compared to the others methods chosen for the 102 classes dataset, when aggregated with CNN and BIC it presented an excellent gain.

It is worth mentioning that the higher gains obtained with CNN+ SURF+ BIC aggregation, shown in Table 2, were due to the fact that those extraction methods complemented each other. In general, it means that those methods had same hits and different misses for the same query. With a higher precision of the ranked lists, more effective will be semantic image retrieval and therefore, fewer user efforts will be required.

7.5. Semantic Interactive Image Retrieval

The results presented in this section are related to the experiments using the Semantic Interactive Image Retrieval (SIIR). As the configurations of the optimal points achieved by each metric were different to the extraction methods (Tables 1 and 2), we used a standardized parameters settings for all the experiments involving the SIIR evaluation ($k = 20,\ t = 2$). The value of the constant $\alpha = 0.01$ was set through an empirical analysis. For the experiments, 10 interactive sessions were considered, since we simulated the user's answer. But in practical applications the user decides when to stop the retrieval process. We also computed the Confidence Interval (CI) with a 0.95 confidence value.

7.5.1. Oxford Flowers – 17 Classes

Fig. 11(a) shows the improvements achieved by the proposed approach in *Precision* × *Recall* (PR) curves considering ACC and SURF descriptors. Fig. 11(a) presents three PR curves: the ACC and SURF in isolation (without UL) and the ACC+SURF combined through rank aggregation in the interactive image retrieval approach after 10 questions.

It can be observed that the gain – represented by the distance between the curves of the two descriptors (ACC and SURF) and the ACC+SURF aggregation with SIIR, reaches a very high value demonstrating the effectiveness of the proposed interactive approach.

Fig. 11(b) illustrates analogous results considering the CNN-Caffe feature. We can observe that higher effectiveness results were achieved, demonstrated by *Precision* × *Recall* curves. The CNN+SURF with SIIR curve shows the best results presented for this analysis.

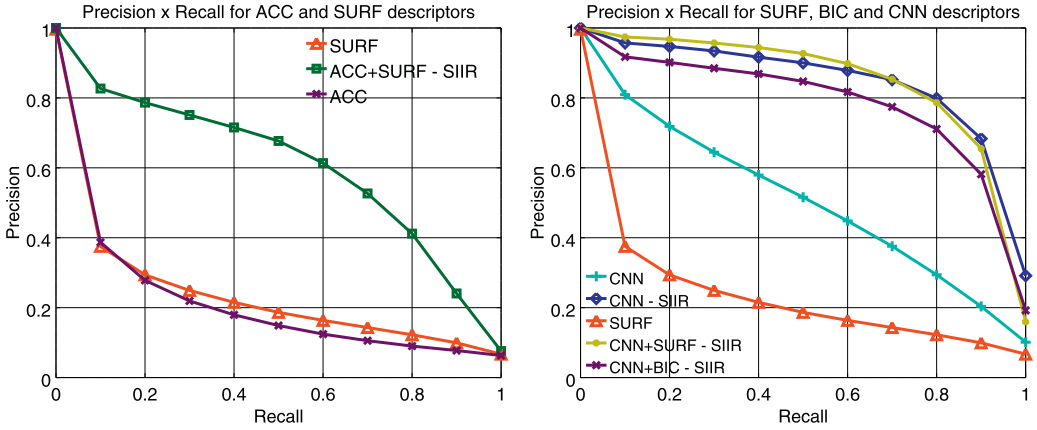
Significant improvements were also obtained considering other effectiveness metrics. Fig. 12 (a) presents the results of precision in the five first positions (P@5) of the ranks for the extraction methods analyzed.

It is observed that the P@5 value of CNN+SURF aggregation with SIIR is the highest achieved in this experiment. The value of this metric increases every question answered by the user, reaching its maximum value within 10 questions answered (0.9707). This value represents a gain of approximately 6.97% in relation to the initial value of P@5 (0.9074) presented in Fig. 12(a).

If compared the maximum value of P@5 for CNN+SURF (0.9707) after evaluation of the Semantics Interactive Image Retrieval, with the P@5 of feature's extractions for the same method without the Unsupervised Learning (0.8859 – Table 1), it is observed a gain of approximately 9.57%.

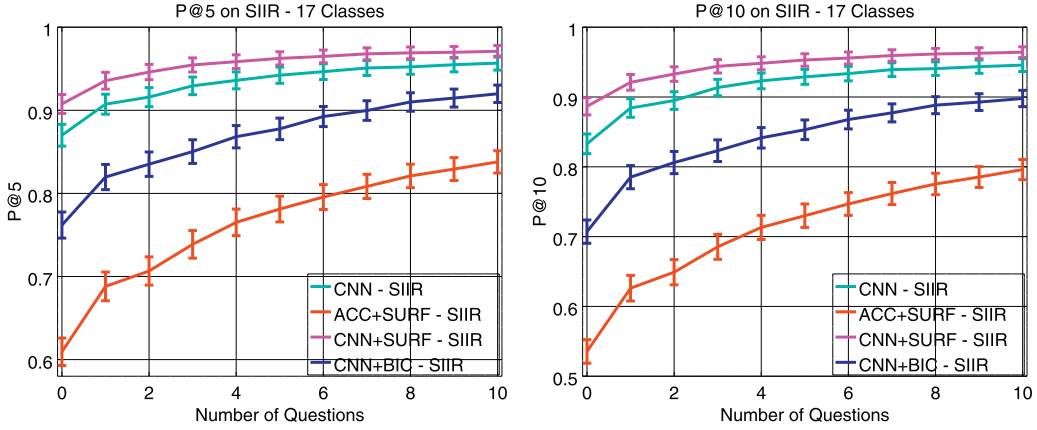
The analysis of the precision in the top-10 positions (P@10) also shows an increase in the three extraction methods. Fig. 12(b) illustrates the obtained results for P@10, with similar gains to P@5. In addition, for image family retrieval, the results, after 10 questions, are even better: 99.17% on PF@5 and 99.11% on PF@10. It can be observed in Fig. 12(a) and (b) that several points do not present intersection of their respective error bars, thus demonstrating statistical differences between some points analyzed.

We can observe that the first iterations are responsible for the highest effectiveness obtained in the interactive retrieval pro-



(a) ACC and SURF. (b) CNN, SURF and BIC.

Fig. 11. *Precision* × *Recall* for extraction methods with Semantic Interactive Image Retrieval after 10 questions, on Oxford 17 Classes.



(a) P@5. (b) P@10.

Fig. 12. Evolution of *Precision* for Semantic Interactive Image Retrieval (SIIR) along with questions, on Oxford 17 Classes.

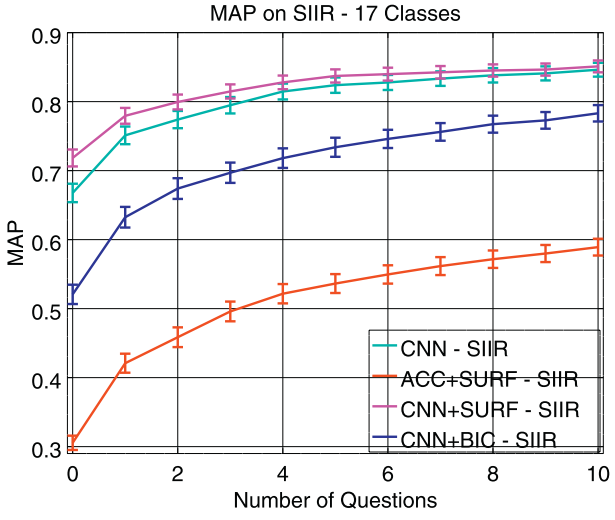


Fig. 13. Evolution of MAP for Semantic Interactive Image Retrieval (SIIR) along with questions, on Oxford 17 Classes.

cess. It is worth mentioning the fact that answering only one question increases the effectiveness of the system in a positive way. For example, Fig. 13 illustrates the significant increase of the MAP measure, demonstrating the evolution of effectiveness of the retrieval results. This result presents a major advancement for Angiosperm families identification, since traditional identification techniques based on a dichotomous key for Flowers with Perianth and Polypetalous Corolla reaches 188 leads (Souza & Lorenzi, 2007).

For example, analyzing the Fig. 13, from the total gain represented by the user's interaction for ACC+SURF MAP on 17 Classes, 40.69% of this gain is reached after the first user's interaction. It also can be noticed in Fig. 13 that the MAP for CNN+SURF also achieves a higher gain score on the first user's interaction, 45.94% from the total gain represented for this joined method, after SIIR execution.

When compared with the initial values displayed in the Fig. 13, the best MAP score achieved gains of 18.52% for CNN+SURF, 26.81% for CNN, 50.49% for CNN+BIC, and 92.89% for ACC+SURF. These gains are related only to the user's interaction process.

These results demonstrate that the proposed approach showed effectiveness to several cases of classification analysis and images retrieval, for both metrics with low income and the best cases.

It is interesting to note that, in curves with low values of P@5, P@10 and MAP, the distances between the confidence intervals (Figs. 12(a), (b) and 13) are greater than the curves that present higher values (CNN and CNN+SURF). This demonstrates the importance of asking more questions for low accuracy methods, as well as corroborating the effectiveness of the proposed approach.

Table 3
Accuracy of 20-NN family classification of SIIR on 17 Oxford flower dataset.

Methods	Family recognition rate
SIIR –CNN+SURF 1 question	**93.97%**
SIIR – CNN+SURF 5 questions	**97.20%**
SIIR – CNN+SURF 10 questions	**98.97%**

Table 4
Accuracy of 20-NN species classification of SIIR, in comparison with state-of-the-art methods on 17 Oxford flower dataset.

Methods	Class recognition rate
Visual Vocabulary (Nilsback & Zisserman, 2006)	71.76%
Discrim. Power-Invar. (Varma & Ray, 2007)	82.55%
Auto. Flower Classif. (Nilsback & Zisserman, 2008)	88.33%
Top-down color attention (Khan, van de Weijer, & Vanrell, 2009)	89%
Bin-ratio information (Xie, Ling, Hu, & Zhang, 2010)	89.02%
BiCoS (Chai, Lempitsky, & Zisserman, 2011)	90.04%
RL-Sim – CNN+SURF	**90.44%**
Multi-scale fusion (Hu, Hu, Xie, Ling, & Maybank, 2014)	91.39%
SIIR – CNN+SURF 1 question	**92.06%**
SIIR – CNN+SURF 5 questions	**95.22%**
SIIR – CNN+SURF 10 questions	**96.84%**

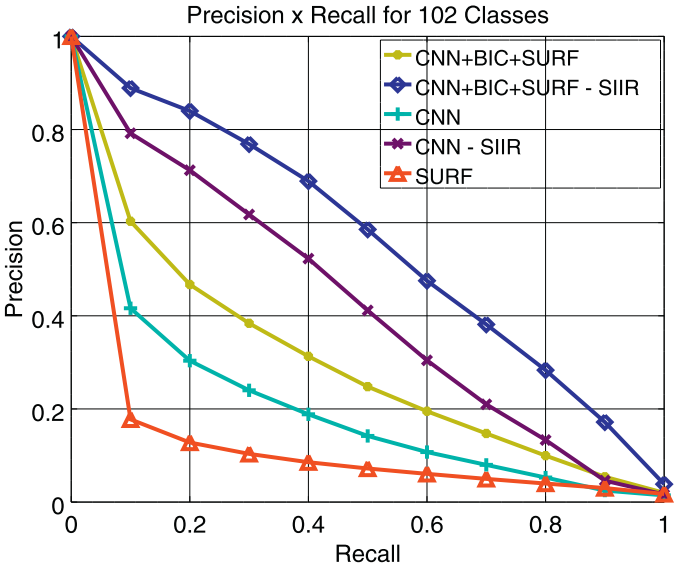


Fig. 14. *Precision* × *Recall* for CNN, SURF and BIC with Semantic Interactive Image Retrieval (SIIR) after 10 questions, on Oxford 102 Classes.

Table 3 presents the accuracy of the 20-NN classification of Angiosperm families. It can be seen that after 10 questions, the family recognition rate reaches 98.97%. Since the results summarized in Table 3 cannot be compared with other studies that use the same dataset due to the lack of literature in similar works to the proposed approach, we also present a class comparison (instead of family comparison) with other state-of-the-art approaches.

Despite the fact that the experimental protocol of our method differs from the others, a brief comparison is presented. Table 4 summarizes the recognition accuracies published for several methods from the literature, along with the accuracy of our proposed approach (SIIR). The Semantic Interactive Image Retrieval obtained the highest accuracy result, when compared to other approaches, reaching **96.84%** of accuracy for the 20-NN classification. Even when the user answers only 1 question, the Semantic Interactive Image Retrieval demonstrates its effectiveness illustrating a higher value of accuracy than other approaches.

7.5.2. Oxford Flowers – 102 Classes

This section introduces the results of the Semantic Interactive Image Retrieval for the Oxford Flowers 102 Classes dataset. The configuration of the execution is the same as those presented by the 17 classes.

Fig. 14 shows the improvement of the proposed approach for the 102 classes dataset after 10 questions. It can be seen once again the effectiveness of the SIIR by the distances of the curves. The highest curves present the results from SIIR, while the low ones show the results from the extraction methods only (without UL). The aggregation of CNN+BIC+SURF with SIIR achieved the highest precision scores for this dataset.

When compared the curves CNN+BIC+SURF after SIIR application with the CNN+BIC+SURF without UL at the point of 20% of the images classes recalled, it can be observed that the precision reaches more than 80% for the curve with SIIR, while the curve without the proposed approach shows a precision of less than 50%.

Fig. 15(a) shows the gain in precision on the top-5 positions in the ranked lists for the 102 classes. It can be seen that CNN+BIC+SURF after 10 questions presents the best P@5. The gain of this aggregation when compared with the initial value on this curve represents 14.92%. The best gain was of SURF (64.04%).

Fig. 15(b) also presents high values on P@10 after applying the questions selected from the Semantic Attribute Selection (Section 6.4).

The precision on the first ten positions with SURF shows the highest gain with 96.36%. When analyzed BIC only, it can be seen that its gain is 67.31%, showing again that the proposed approach is very effective in low retrieval values also.

It can be observed in Fig. 15(a) and (b) the non-intersection between the confidence intervals of some points, thus demonstrating statistical differences between such analyzed points.

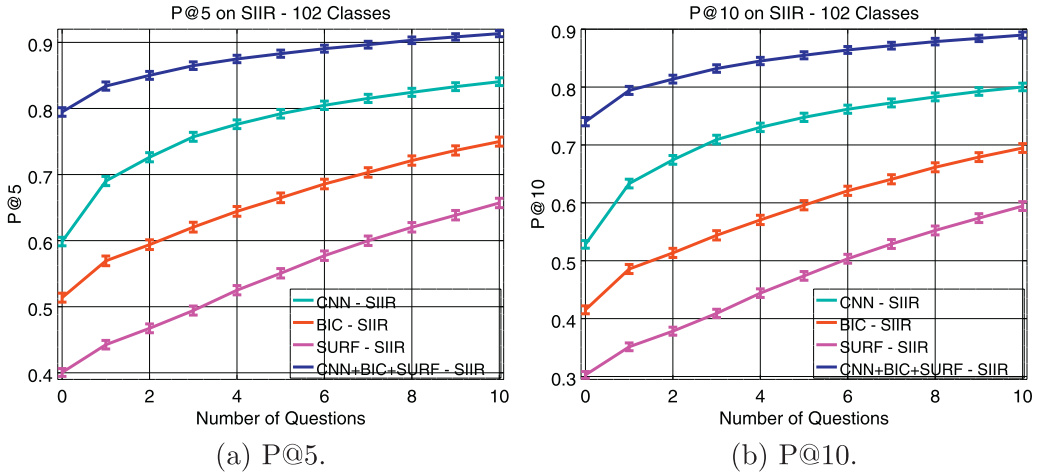
The precisions for image family retrieval for the results presented in 102 classes also demonstrates high scores after 10 questions: 93.20% on PF@5 and 91.59% on PF@10.

Fig. 16 shows that the improvement of the retrieved results just not occur only in the first positions, but also above deeper positions over the rank. This analysis can be done because when the user answers a question that corresponds to some attribute, even the deepest plant images that contain that attribute can be moved to the rank initial positions. Fig. 16 shows that SURF descriptor had a gain of 112.76% on this metric, after the SIIR approach been applied. It can be seen that BIC descriptor had a 101.45% gain in MAP, while the union CNN+BIC+SURF reaches a gain of 32.88%.

When comparing the graphics of the confidence interval of the 102 classes dataset (Figs. 15(a), (b) and 16) with the graphics of the 17 classes dataset (Figs. 12(a), (b) and 13), it can be observed that the confidence intervals of the analyzed points for the 102 classes are smaller than those presented for the 17 classes. This is due to the fact that the number of comparisons between the ranks and the number of analyzed images is bigger in the set of 102 classes. This way, even curves with higher values for the metrics tested demonstrates the need to apply the proposed approach.

When analyzing P@5, P@10 and MAP metrics for 102 classes dataset, the gains on the first user's interaction are lower than those presented by 17 classes, but also illustrates that the biggest part of the gain (Figs. 15(a), (b) and 16) is related to the first interaction and consequently, demonstrates the reduction on the effort in classifying plants.

Finally, we present the accuracy of the 20-NN images of Angiosperm families (Table 5). It can be seen that after 10 questions, the family recognition rate reaches 85.39%. We also present Table 6



(a) P@5. (b) P@10.

Fig. 15. Evolution of *Precision* for Semantic Interactive Image Retrieval (SIIR) along with questions, on Oxford 102 Classes.

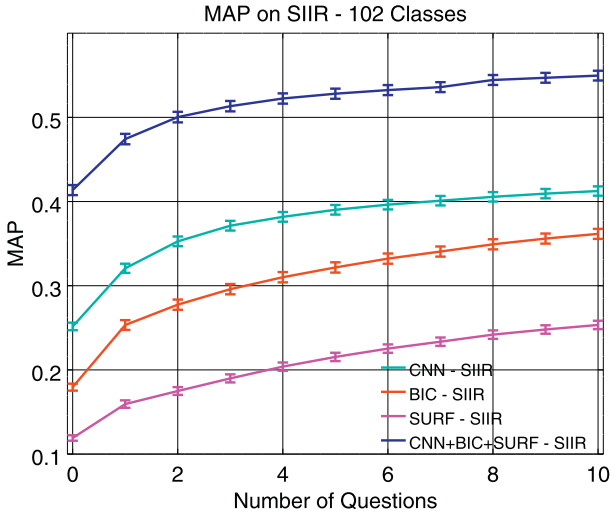


Fig. 16. Evolution of MAP for Semantic Interactive Image Retrieval (SIIR) along with questions, on Oxford 102 Classes.

Table 5
Accuracy of 20-NN family classification of SIIR on 102 Oxford flower dataset.

Methods	Family recognition rate
SIIR – CNN+BIC+SURF 1 question	**74.94%**
SIIR – CNN+BIC+SURF 5 questions	**81.13%**
SIIR – CNN+BIC+SURF 10 questions	**85.39%**

Table 6
Accuracy of 20-NN species classification of SIIR in comparison with state-of-the-art methods on 102 Oxford flower dataset.

Methods	Class recognition rate
Ito and Kubota (2010)	53.9%
Nilsback and Zisserman (2008)	72.8%
Khan, van de Weijer, Bagdanov, and Vanrell (2011)	73.3%
Kanan and Cottrell (2010)	75.2%
Nilsback (2009)	76.3%
Angelova, Wong, Zhu, Specht, and Lin (2012)	76.7%
SIIR – CNN+BIC+SURF 1 question	**79.15%**
Chai et al. (2011)	80.0%
Angelova, Zhu, and Lin (2013)	80.6%
Mattos, Herrmann, Shigeno, and Feris (2014)	80.8%
SIIR – CNN+BIC+SURF 5 questions	**84.89%**
SIIR – CNN+BIC+SURF 10 questions	**88.88%**

with a comparison of SIIR with others methods presented in the literature.

It is worth mentioning that the accuracy cannot be compared precisely with others approaches presented in Table 6 since the experimental protocol of our method differs from the others. However, it can be seen that the accuracy of the top-20 images of SIIR achieves high-accuracy results when compared to the others state-of-the-art methods.

The class accuracy of SIIR at 1-NN classification on 102 classes using CNN+BIC+SURF after 10 questions is equal to 90.36%. The class accuracy at the 5-NN classification of the same joined configuration is 90.31%, while the class accuracy at 10-NN is equal to 89.79%.

8. Discussion

The difficulties associated with identification of Angiosperm families is addressed by SIIR through a combination of low-level features, ontology information and interactive retrieval. The major advantage of the proposed approach consists in the capacity of reducing the user's efforts by suggesting a question that can potentially differentiate the highest number of families presented on the top retrieved images. As demonstrated by experimental evaluation, a single user response can improve significantly the retrieval results. Since a traditional dichotomous key has nearly by 188 leads to identify an Angiosperm family, it represents a relevant contribution.

The SIIR approach also yields very high effective retrieval and classification results in comparison with others approaches. For example, our best results in accuracy reach 96.84% and 88.88% for the Oxford 17 classes and 102 classes, respectively; when others state-of-the-art approaches reach, respectively, 91.39% and 80.8% in accuracy results for the same datasets.

Another strong point of our proposed method relies on the fact that if the user has any doubts about what the plant structure or the property means (i.e.: question with a selected attribute) it is possible to consult the ontology in order to understand and clarify the meaning of that question. Since the object properties relate the classes of the ontology, the user has a full high-level concepts and information about the Systematic Botany domain by analyzing the ontology graph and/or reading the ontology annotations about that concept in doubt.

On the other hand, identifying a plant is not an easy task and it is possible that the system will suggest a hard question to the user. Despite the fact that the user can consult the ontology to clarify the doubts about the question and sometimes that is the

only question that needs to be solved, it would be necessary to extend the approach to consider profiles of different users, asking easy question for those inexperienced ones.

Another weakness of the system in its current formulation occurs when the user answers the question in a wrong way. The system will probably reduce the retrieval effectiveness, since it moves the top ranked images that present the attribute answered incorrectly.

The SIIR approach also requires an ontology in some domain. If there is no ontology developed for the studied domain of the images, some features and functionality of the proposed approach would not work properly. As the tendency is to provide information in the patterns of a new Semantic Web, it can be said that the development of specific vocabularies must grow in the coming decades. When this development occurs jointly by a determined community of researchers from different areas, following established standards, the knowledge ends up interconnecting in several areas and can be shared and reused. This way, it is possible to affirm that the SIIR can be applied in others areas of knowledge.

9. Conclusions

In this paper, we have presented a novel approach for Interactive Image Retrieval. The proposed approach reduces the semantic gap between low-level features of the images and the high-level semantic concepts by introducing a Semantic Guided Interactive Image Retrieval.

The main idea consists in retrieving images based on their visual features, relating such images with their concepts, defined by the developed ontology, supported by user interactions. A set of experiments was conducted for assessing the effectiveness of the proposed approach. The results demonstrated that high effectiveness can be obtained in various scenarios. Experiments also showed the effectiveness of SIIR in two different datasets, reaching high values of the metrics analyzed in both. Additionally, showed that SIIR can be very effective for both low and high-effective input retrieved results.

The proposed approach can be also applied for educational purposes since the information defined by the ontology represents clear assertions for both humans and machines. In this way, every restriction can be consulted by the user in order to clarify any doubt about the concepts. Furthermore, the relationships between each ontology entity show what structure or property belongs to a particular class, thus facilitating the teaching through the developed ontology.

Future work focuses on the investigation of novel formulations for distance updating in the Semantic Attribute Selection stage. Although effective, the current formulation is extremely simple and can be improved. We also intend to develop a species ontology, instead of a family ontology used in this work. Since species presents specific attributes (more restrictions when compared to families), we believe that Semantic Attribute Selection will be more effective, impacting positively the retrieval and classification process. With a more specific attribute, the user's answer will be more effective regarding the interactive image retrieval process. More specifically, the user answer "no" will be even more effective, since images that does not present that value for the selected attribute also does not belong to the specie of the query image and therefore should be moved to lower positions in the rank.

Future work also focuses on an adaptive interaction process. The main idea is to formulate the questions according to the user profile, asking easier questions for those inexperienced ones. In order to fulfill this feature, more information must be included in the ontology. For example, the structures and properties easily viewable in images can be annotated as an "*easy*" mark, while internal attribute as a "*hard*" one. With this feature, the Semantic Interactive Image Retrieval can reach different people and adapt itself for many kinds of users.

The proposed approach is completely unsupervised until the start of user interactions. Other line of investigation consists in the use of semi-supervised learning, by exploiting training data relating low-level features and morphological structures in order to distinguish families or species. It is possible to use the feature vector of each image and the attributes modeled in the ontology as input to train a machine in order to group image plants into their families.

Other possibility of future work is ranking images not only based on their low-level features, but also taking into account their attributes modeled in the ontology. The idea consists in sorting the ranked lists considering the plant images that also have the biggest similarity about their attributes. In addition, this idea simulates a phylogenetic tree, which relates plants that present similar properties and structures.

Another promising line of investigation focuses on the use of SIIR for different areas of knowledge, such as archaeology (to identify different kinds of remains), geology (to identify minerals and rocks), ichthyology (in order to identify fish by their scales, for example) and buildings (to identify the period of some construction, for example, baroque). In fact, for applying SIIR in other domains it is only required the creation of an ontology for the studied domain.

As the idea of a new Semantic Web is to provide, share and reuse information in standards, it can be said that SIIR can be fit in those patterns, since uses an ontology to represent concepts and information about Angiosperm families. In this sense, the development of new vocabularies must grow in the next decades, enabling the relationship to several different domains and interconnecting the knowledge in several areas, which also propitiate the expansion and application of SIIR in other areas of knowledge.

Thus, the present work presented a novel approach with high efficacy in image retrieval that unites the knowledge of the studied domain (Systematic Botany) with the visuals images features.

Acknowledgments

The authors are grateful to CAPES (Coordination for Higher Education Staff Development), and São Paulo Research Foundation – FAPESP (Grant 2013/08645-0).

References

Angelova, A., Wong, J., Zhu, S., Specht, C., & Lin, Y. (December, 2012). Development and Deployment of a Large-Scale Flower Recognition Mobile App, Technical Report, NEC Laboratories America.

Angelova, A., Zhu, S., & Lin, Y. (2013). Image segmentation for large-scale subcategory flower recognition. In *Proceedings of IEEE workshop on applications of computer vision (WACV)* (pp. 39–45).

Arevalillo-Herraez, M., & Ferri, F. J. (2013). An improved distance-based relevance feedback strategy for image retrieval. *Image and Vision Computing, 31*(10), 704–713.

Arvor, D., Durieux, L., Andres, S., & Laporte, M.-A. (2013). Advances in geographic object-based image analysis with ontologies: A review of main contributions and limitations from a remote sensing perspective. *ISPRS Journal of Photogrammetry and Remote Sensing, 82*, 125–137.

Bai, S., & Bai, X. (2016). Sparse contextual activation for efficient visual re-ranking. *IEEE Transactions on Image Processing, 25*(3), 1056–1069.

Bai, X., Bai, S., & Wang, X. (2015). Beyond diffusion process: Neighbor set similarity for fast re-ranking. *Information Sciences, 325*, 342–354.

Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (surf). *Computer Vision and Image Understanding, 110*(3), 346–359.

Caballero, C., & Aranda, M. C. (2010). Plant species identification using leaf image retrieval. In *Proceedings of the ACM international conference on image and video retrieval, CIVR'10* (pp. 327–334). ACM.

Chai, Y., Lempitsky, V., & Zisserman, A. (2011). BiCoS: A bi-level co-segmentation method for image classification. In *Proceedings of international conference on computer vision, ICCV'2011* (pp. 2579–2586).

Cheng, E., Jing, F., & Zhang, L. (2009). A unified relevance feedback framework for web image retrieval. *IEEE Transactions on Image Processing, 18*(6), 1350–1357.

Coto, A. L. (2008). *The use of ontologies for improving image retrieval and annotation* p. 79. Knowledge Media Institute (KMI).

Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys, 40*(2), 5:1–5:60.

Feng, H., & Chua, T.-S. (2003). A bootstrapping approach to annotating large image collection. In *Proceedings of the 5th ACM SIGMM international workshop on multimedia information retrieval, MIR'03* (pp. 55–62). ACM.

Fernández-López, M., Gómez-Pérez, A., & Juristo, N. (1997). Methontology: From ontological art towards ontological engineering. In *Proceedings of the ontological engineering AAAI-97 spring symposium series*. American Association for Artificial Intelligence.

Giacinto, G. (2007). A nearest-neighbor approach to relevance feedback in content based image retrieval. In *Proceedings of the international conference on image and video retrieval, CIVR'07* (pp. 456–463). ACM.

Goëau, H., Bonnet, P., Joly, A., Bakić, V., Barbe, J., Yahiaoui, I., … Péronnet, A. (2013). Pl@ntnet mobile app. In *Proceedings of the 21st ACM international conference on multimedia, MM'13* (pp. 423–424). ACM.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition, 5*(2), 199–220.

Guan, J., & Qiu, G. (2007). Learning user intention in relevance feedback using optimization. In *Proceedings of the international workshop on workshop on multimedia information retrieval, ACM MIR'07* (pp. 41–50). New York, NY, USA: ACM.

Guarino, N. (1998). *Formal ontology in information systems: Proceedings of the 1st international conference June 6–8, 1998, Trento, Italy* (1st). Amsterdam, The Netherlands: IOS Press.

Halaschek-Wiener, C., Schain, A., Grove, M., Parsia, B., & Hendler, J. (2005). Management of digital images on the semantic web. In *Proceedings of the international semantic web conference (ISWC2005), Galway, Ireland*.

Hoi, S. C., Liu, W., & Chang, S.-F. (2010). Semi-supervised distance metric learning for collaborative image retrieval and clustering. *ACM Transactions on Multimedia Computing and Communication Applications, 6*(3), 18:1–18:26.

Hu, W., Hu, R., Xie, N., Ling, H., & Maybank, S. (2014). Image classification using multiscale information fusion based on saliency driven nonlinear diffusion filtering. *IEEE Transactions on Image Processing, 23*(4), 1513–1526.

Huang, J., Kumar, S. R., Mitra, M., Zhu, W.-J., & Zabih, R. (1997). Image indexing using color correlograms. In *Proceedings of IEEE conference on computer vision and pattern recognition, CVPR* (pp. 762–768).

Hui, H. W., Mohamad, D., & Ismail, N. (2010). Semantic gap in CBIR: Automatic objects spatial relationships semantic extraction and representation. *International Journal Of Image Processing (IJIP), 4*(3), 192–204.

Ito, S., & Kubota, S. (2010). Object classification using heterogeneous co-occurrence features. In *Proceedings of the 11th European conference on computer vision: Part II, ECCV'10* (pp. 209–222).

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., … Darrell, T. (2014). Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on multimedia*.

Kanan, C., & Cottrell, G. (2010). Robust classification of objects, faces, and flowers using natural image statistics. In *Proceedings of IEEE conference on computer vision and pattern recognition, CVPR* (pp. 2472–2479).

Kebapci, H., Yanikoglu, B., & Unal, G. (2009). Plant image retrieval using color and texture features. In *Proceedings of the 24th international symposium on Computer and information sciences, ISCIS 2009* (pp. 82–87).

Khan, F. S., van de Weijer, J., Bagdanov, A. D., & Vanrell, M. (2011). Portmanteau vocabularies for multi-cue image representations. In *Proceedings of the international conference on neural information processing systems, NIPS 2011*.

Khan, F. S., van de Weijer, J., & Vanrell, M. (2009). Top-down color attention for object recognition. In *Proceedings of international conference on computer vision, ICCV'2012* (pp. 979–986).

Knublauch, H., Fergerson, R. W., Noy, N. F., & Musen, M. A. (2004). *The protege owl plugin: An open development environment for semantic web applications* (pp. 229–243). Springer.

Kundu, M. K., Chowdhury, M., & Bulo, S. R. (2015). A graph-based relevance feedback mechanism in content-based image retrieval. *Knowledge-Based Systems, 73*, 254–264.

Kurtz, C., Depeursinge, A., Napel, S., Beaulieu, C., & Rubin, D. (2014). On combining image-based and ontological semantic dissimilarities for medical image retrieval applications. *Medical Image Analysis, 18*(7), 1082–1100.

Kwan, P. W., Welch, M. C., & Foley, J. J. (2015). A knowledge-based decision support system for adaptive fingerprint identification that uses relevance feedback. *Knowledge-Based Systems, 73*, 236–253.

Lacy, L. W. (2005). *OWL: Representing information using the Web Ontology Language*. Victoria BC, Canada: Trafford Publishing.

Lew, M. S., Sebe, N., Djeraba, C., & Jain, R. (2006). Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications, 2*(1), 1–19.

Liu, Y., Zhang, D., Lu, G., & Ma, W.-Y. (2007a). A survey of content-based image retrieval with high-level semantics. *Pattern Recognition, 40*(1), 262–282.

Liu, Y.-T., Liu, T.-Y., Qin, T., Ma, Z.-M., & Li, H. (2007b). Supervised rank aggregation. In *Proceedings of international conference on world wide web (WWW'2007)* (pp. 481–490).

Lux, M. (2013). Lire: Open source image retrieval in java. In *Proceedings of the 21st ACM international conference on multimedia, MM'13* (pp. 843–846). ACM.

Lux, M., & Chatzichristofis, S. A. (2008). Lire: Lucene image retrieval: An extensible java CBIR library. In *Proceedings of the 16th ACM international conference on multimedia* (pp. 1085–1088). ACM.

Manzoor, U., Usman, M., Balubaid, M. A., & Mueen, A. (2015). Ontology-based clinical decision support system for predicting high-risk pregnant woman. *system, 6*(12).

Mattos, A. B., Herrmann, R. G., Shigeno, K. K., & Feris, R. S. (2014). A mission-oriented citizen science platform for efficient flower classification based on combination of feature descriptors. In *Proceedings of international workshop on environmental multimedia retrieval*.

Nilsback, M.-E. (2009). *An automatic visual flora – Segmentation and classification of flowers images*. Ph.D. thesis. University of Oxford.

Nilsback, M.-E., & Zisserman, A. (2006). A visual vocabulary for flower classification. In *Proceedings of IEEE conference on computer vision and pattern recognition, CVPR'2006: 2* (pp. 1447–1454).

Nilsback, M.-E., & Zisserman, A. (2008). Automated flower classification over a large number of classes. In *Proceedings of Indian conference on computer vision, graphics and image processing*.

Pandey, S., Khanna, P., & Yokota, H. (2015). An effective use of adaptive combination of visual features to retrieve image semantics from a hierarchical image database. *Journal of Visual Communication and Image Representation, 30*, 136–152.

Pedronette, D. C. G., Gonçalves, F. M. F., & Guilherme, I. R. (2017). Unsupervised manifold learning through reciprocal kNN graph and connected components for image retrieval tasks. *Pattern Recognition*.

Pedronette, D. C. G., & da S. Torres, R. (2013). Image re-ranking and rank aggregation based on similarity of ranked lists. *Pattern Recognition, 46*(8), 2350–2360.

Pedronette, D. C. G., & da S. Torres, R. (2014). Unsupervised manifold learning using reciprocal kNN graphs in image re-ranking and rank aggregation tasks. *Image and Vision Computing, 32*(2), 120–130.

Razavian, A. S., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN features off-the-shelf: an astounding baseline for recognition. In *Proceedings of IEEE conference on computer vision and pattern recognition workshops (CVPRW'14)* (pp. 512–519).

Reddy, V. R. K., & Bandikolla, P. (2008). Image retrieval using a combination of keywords and image features.

Smeulders, A. W., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*, 1349–1380.

Souza, V. C., & Lorenzi, H. (2005). *Botanica Sistematica: Guia ilustrado para identificacão das familias de Angiospermas da flora brasileira, baseado em APG*. Instituto Plantarum.

Souza, V. C., & Lorenzi, H. (2007). *Chave de Identificacão: Para as principais familias de Angiospermas nativas e cultivadas do Brasil*. Instituto Plantarum.

Stehling, R. O., Nascimento, M. A., & Falcão, A. X. (2002). A compact and efficient image retrieval approach based on border/interior pixel classification. In *Proceedings of the eleventh international conference on information and knowledge management* (pp. 102–109). New York, NY, USA: ACM.

Thomee, B., & Lew, M. (2012). Interactive search in image retrieval: A survey. *International Journal of Multimedia Information Retrieval, 1*(2), 71–86.

Varma, M., & Ray, D. (2007). Learning the discriminative power-invariance trade-off. In *Proceedings of international conference on computer vision, ICCV'2007* (pp. 1–8).

Vogel, J., & Schiele, B. (2007). Semantic modeling of natural scenes for content-based image retrieval. *International Journal of Computer Vision, 72*(2), 133–157.

Walls, R. L., Athreya, B., Cooper, L., Elser, J., Gandolfo, M. A., Jaiswal, P., … Stevenson, D. W. (2012). Ontologies as integrative tools for plant science. *American Journal of Botany, 99*, 1263–1275.

Xie, N., Ling, H., Hu, W., & Zhang, X. (2010). Use bin-ratio information for category and scene classification. In *Proceedings of IEEE conference on computer vision and pattern recognition, CVPR'2010* (pp. 2313–2319).