

Interactive search in image retrieval: a survey

Bart Thomee · Michael S. Lew

Received: 7 April 2012 / Revised: 27 April 2012 / Accepted: 28 April 2012 / Published online: 8 June 2012
© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract We are living in an Age of Information where the amount of accessible data from science and culture is almost limitless. However, this also means that finding an item of interest is increasingly difficult, a digital needle in the proverbial haystack. In this article, we focus on the topic of content-based image retrieval using interactive search techniques, i.e., how does one interactively find any kind of imagery from any source, regardless of whether it is photographic, MRI or X-ray? We highlight trends and ideas from over 170 recent research papers aiming to capture the wide spectrum of paradigms and methods in interactive search, including its subarea relevance feedback. Furthermore, we identify promising research directions and several grand challenges for the future.

Keywords Multimedia information retrieval · Content-based image retrieval · Image search · Interactive search · Relevance feedback · Human–computer interaction

1 Introduction

Terabytes of imagery are being accumulated daily from a wide variety of sources such as the Internet, medical centers (MRI, X-ray, CT scans) or digital libraries. It is not uncommon for one's personal computer to contain thousands of photos stored in digital photo albums. At present, billions of

images can even be found on the World Wide Web. But with that many images within our reach, how do we go about finding the ones we want to see at a particular moment in time? Interactive search methods are meant to address the problem of finding the right imagery based on an interactive dialog with the search system. Some recent examples of the interfaces to these interactive image search systems are shown in Fig. 1.

Furthermore, interactive search allows the user to find imagery, even when there is not a word known to the user for the concept he has in mind. Interactive retrieval systems can, for example, assist a virologist in identifying potentially life-threatening bacteria within a databases containing characteristics of tens of thousands of bacteria and viruses, or assist a radiologist in making his diagnosis of the patient by providing the most relevant examples from credible sources.

The areas of interactive search with the greatest societal impact have been in WWW image search engines and recommendation systems. Google, Yahoo! and Microsoft have added interactive visual content-based search methods into their worldwide search engines, which allows search by similar shape and/or color (see Fig. 2) and are used by millions of people each day. The recommendation systems have been implemented by companies such as Amazon, NetFlix and Napster in wide and diverse contexts, from books to clothing, from movies to music. They give recommendations of what the user would be interested in next based on feedback from prior ratings. Furthermore, Internet advertisements are usually driven by relevance feedback strategies where clicked upon products and links are used to show the next set of advertisements to the user in real time. If a user clicks upon some shoes at a major retailer website, he will probably be shown advertisements for shoes at the next websites that he visits. In image retrieval, another good example is Getty Images where the audience is assumed to be knowledgeable

B. Thomee
Yahoo! Research, Avinguda Diagonal 177, 08018 Barcelona, Spain
e-mail: bthomee@yahoo-inc.com

B. Thomee · M. S. Lew (✉)
Leiden University, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands
e-mail: mlew@liacs.nl



Fig. 1 Examples of user interfaces. The ‘tendrill’ interface [1] (left) is specifically designed to support the user in exploring the visual space, where changes to the query result in branching off the initial path. The ‘FreeEye’ interface [2] (right) assists the user in browsing the database, where the selected image is surrounded by similar ones

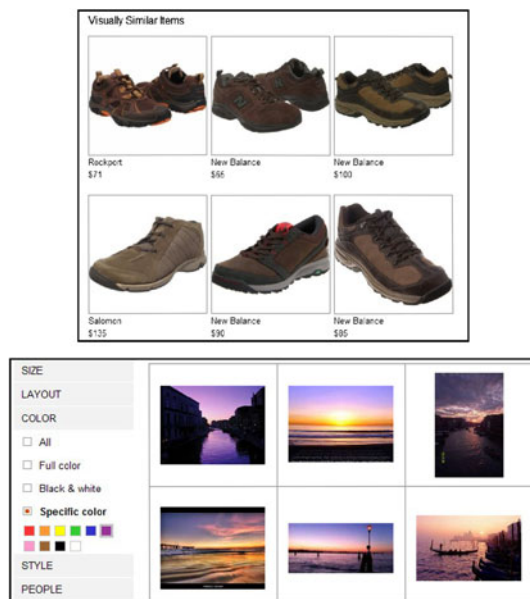


Fig. 2 An example from Google Product Search (top) showing items that are visually similar by shape and color, and from Microsoft Bing image search (bottom) showing the interface and resulting visually similar images by color (purple) (color figure online)

and their image search engine reflects this by having multi-modal interactive image search capabilities by both content, context, style, composition and user feedback. Moreover, interactive image search has become important in medical facilities both in hospitals and in research labs [3]. These systems allow interactive searching on both 2D and 3D imagery from X-ray, MRI, ultrasound and electron microscopy.

Text search relies on annotations that are frequently missing in both personal and public image collections. When annotations are either missing or incomplete, the only alternative is to use methods that analyze the pictorial content of the imagery in order to find the images of interest. This field of research is also known as content-based image retrieval. Since the early 1990s the field has evolved and has made significant breakthroughs. “The early years” of image retrieval were summarized by Smeulders et al. [4], painting a detailed

picture of a field in the process of learning how to successfully harness the enormous potential of computer vision and pattern recognition. The number of publications increased dramatically over the past decade. The comprehensive reviews of Datta et al. [5,6], Lew et al. [7] and Huang et al. [8] provide a good insight into the more recent advances in the entire field of multimedia information retrieval and, in particular, content-based image retrieval.

A particularly well explored subarea of interactive search is called relevance feedback where the search system solicits user feedback on the relevance of results over the course of several rounds of interaction, where after each round the system ideally returns images that better correspond to what the user has in mind. A strength of relevance feedback systems is that the user feedback is simplified to an extreme, typically just a binary “relevant” or “not relevant”. This strength is also a weakness in that the user can often provide richer feedback than relevance. The last review dedicated to relevance feedback in image retrieval was published in 2003 [9], but with the rapid progress of technology, many novel and interesting techniques have been introduced since then. As is covered in this paper, researchers have gone far beyond simple relevance feedback and frequently integrate more diverse information and techniques into the interactive search process.

In this survey, we reviewed all papers in the ACM, IEEE and Springer digital libraries related to interactive search in content-based image retrieval over the period of 2002–2011 and selected a representative set for inclusion in this overview. This survey is aimed at content-based image retrieval researchers and intends to provide insight into the trends and diversity of interactive search techniques in image retrieval from the perspectives of the users and the systems. This paper will not be discussing the simplest uses (i.e. keyword search) of interactive search. We will be covering more sophisticated types of interactive search which delve into deeper levels of interaction such as wider, multimodal queries and answers, and the next generation approaches of using user feedback such as active learning. We try to present the trends, the larger clusters of research, some of the frontier research, and the major challenges.

We have organized our discussion according to the view of interactive image retrieval as a dialog between user and system, looking at both sides of the story. In Sect. 2 we therefore first capture the state of the art by considering how the user interacts with the system and in Sect. 3 we then reverse their roles by considering how the system interacts with the user. Because the majority of research focuses on improving interactive image retrieval from the system’s perspective, we have consequently directed more attention to that side of the discussion. In Sect. 4 we continue by looking at the ways that retrieval systems are presently evaluated and benchmarked. Finally, in Sect. 5 we summarize the promising frontiers and present several grand challenges.

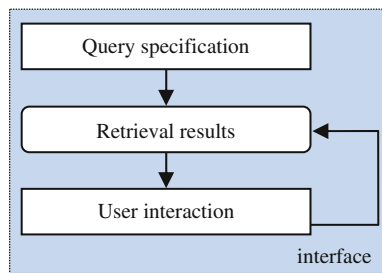


Fig. 3 The interactive search process from the user's point of view

2 Interactive search from the user's point of view

A rough overview of the interactive search process is shown in Fig. 3. Note that real systems typically have significantly greater complexity. In the first step, the user issues a query using the interface of the retrieval system and shortly thereafter is presented with the initial results. The user can then interact with the system in order to obtain improved results. Conceivably, the ideal interaction would be through questions and answers (Q&A), similar to the interaction at a library help desk. Through a series of questions and answers the librarian helps the user find what he is interested in, often with the question “Is this what you are looking for?”. This type of interaction would eventually uncover the images that are relevant to the user and which ones are not. In principle, feedback can be given as many times as the user wants, although generally he will stop giving feedback after a few iterations, either because he is satisfied with the retrieval results or because the results no longer improve.

2.1 Query specification

The most common way for a retrieval session to start is similar to the Q&A interaction one would have with a librarian. One might provide some descriptive text (i.e. keywords) [10], provide an example image [11] or in some situations use the favorites based on the history of the user [2]. The query step can also be skipped directly when the system shows a random selection of images from the database for the user to give feedback on [12]. When image segmentation is involved there are a variety of ways to query the retrieval system, such as selecting one or more pre-segmented regions of interest [13,14] or drawing outlines of objects of interest [15,16]. A novel way to compose the initial query is to let the user first choose keywords from a thesaurus, after which per keyword one of its associated visual regions is selected [17].

2.2 Retrieval results

The standard way in which the results are displayed is a ranked list with the images most similar to the query shown

at the top of the list. Because giving feedback on the best matching images does not provide the retrieval system with much additional information other than what it already knows about the user's interest, a second list is also often shown, which contains the images most informative to the system [18]. These are usually the images that the system is most uncertain about, for instance those that are on or near a hyperplane when using SVM-based retrieval. This principle, called *active learning*, is discussed in more detail in Sect. 3.3. Innovative ways of displaying the retrieval results are discussed in Sect. 2.4.

2.3 User interaction

Many of the systems have interaction which is designed to be used by a machine learning algorithm which gives rise naturally to labeling results as either positive and/or negative examples. These examples are given as feedback to the systems to improve the next iteration of results. Researchers have explored using positive feedback only [19], positive and negative feedback [20], positive, neutral and negative feedback [21], and multiple relevance levels: four relevance levels [22,23], five levels [17] or even seven levels [24]. An alternative approach is to let the user indicate by what percentage a sample image meets what he has in mind [25].

While positive/negative examples are important to learning, in many cases it can be advantageous to allow the user to give other kinds of input which may be in other modalities (text, audio, images, etc.), other categories, or personal preferences. Thus, some systems allow the user to input multiple kinds of information in addition to labeled examples [1,2,26–31]. In addition, sketch interfaces allow the user to give a fundamentally different kind of input to the system [32,33], which can potentially give a finer degree of control over the results. In the Q&A paradigm [34,35], results may be dynamically selected to best fit the question, based on deeper analysis of the user query. For example, by detecting verbs in the user query or results, the system can determine that a video showing the actions will provide a better answer than an image or only text.

When the system uses segmented images it is possible to implement more elaborate feedback schemes, for instance allowing the splitting or merging of image regions [36], or supporting drawing a rectangle inside a positive example to select a region of interest [37]. An interesting discussion on the role and impact of negative images and how to interpret their meaning can be found in [38]. Besides giving explicit feedback, it is also possible to consider the user's actions as a form of implicit feedback [39], which may be used to refine the results that are shown to the user in the next result screen. An example of implicit feedback is a click-through action, where the user clicks on an image with the intention to see it in more detail [40]. In contrast with the traditional

query-based retrieval model, the ostensive relevance feedback model [41,42] accommodates for changes in the user's information needs as they evolve over time through exposure to new information over the course of a single search session.

2.4 The interface

The role of the interface in the search process is often limited to displaying a small set of search results that are arranged in a grid, where the user can refine the query by indicating the relevance of each individual image. In recent literature, several interfaces break with this convention, aiming to offer an improved search experience (see Figs. 1, 4). These interfaces mainly focus on one, or a combination, of the following aspects:

Support for easy browsing of the image collection, for instance through an ontological representation of the image collection where the user can zoom in on different concepts of interest [43], by easily shifting the focus of attention from image to image allowing the user to visually explore the local relevant neighborhood surrounding an image [2,44] or by letting users easily navigate to other promising areas in feature space, which is particularly useful when the search no longer improves with the current set of relevant images [12].

Better presentation of the search results, with for instance giving more screen space to images that are likely to be more relevant to the query than to less relevant images [45], dynamically reorganizing the displayed pages into visual islands [46] that enable the user to explore deeper into a particular dimension he is interested in, or visualizing the results where similar images are placed closer together [47,48].

Multiple query modalities, result modalities and ways of giving feedback, for instance by allowing the user to query by grouping and/or moving images [49,50], 'scribbling' on images to make it clear to the retrieval system which parts of an image should be considered foreground and which parts background [51], or providing the user with the best mixture of media for expressing a query or understanding the results.

2.5 Trends and advances

The increasing popularity of higher level image descriptors has expressed itself in approaches that are tailored to support those ways of searching. In particular, we have noticed an increase in research on how to best leverage region-based image retrieval, offering new ways to initiate the search, give feedback and visualize the retrieval results. During the last decade we have seen the interface transition from having only a supportive role to playing a more substantial role in finding images. The interfaces have evolved from simple grids to a wide variety of approaches, which include but are not limited to image clusters, ontologies, image linked representations (e.g. the tendril interface), and 3D visualizations.



Fig. 4 Examples of user interfaces. The 'similarity visualization' interface [47] (left) displays a representative set of images from the entire collection, where similar images are projected close to each other and dissimilar ones far away. The 'visual islands' interface [46] (right) reorganizes search results into colored clusters of related images

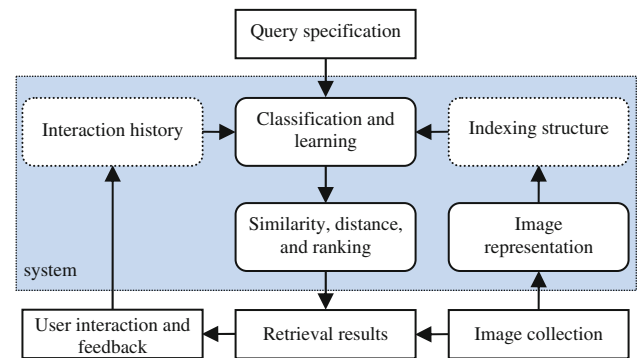


Fig. 5 The interactive search process from the system's point of view

Recent advances have expanded the frontiers in both the user interface and the kinds of interaction the user can have with the system. In particular, these systems allow the user to ask multi-modal queries/questions and also give multi-modal input on the set of results. Furthermore, it is also a growing trend to integrate browsing and search as well as provide varying levels of explanations for why the results were chosen.

3 Interactive search from the system's point of view

A global overview of a retrieval system is shown in Fig. 5. The images in the database are converted into a particular image representation, which can optionally be stored in an indexing structure to speed up the search. Once a query is received, the system applies an algorithm to learn what kind of images the user is interested in, after which the database images are ranked and shown to the user with the best matches first. Any feedback the user gives can optionally be stored in a log for the purpose of discovering search patterns, so learning will improve in the long run. In this section, we cover the recent advances on each of these parts of a retrieval system.

Fig. 6 Images overlaid with detected visual words. Identically colored squares indicate identical visual words, while differently colored squares indicate different visual words (color figure online)



3.1 Image representation

By itself an image is simply a rectangular grid of colored pixels. In the brain of a human observer these pixels form meanings based on the person's memories and experiences, expressing itself in a near-instantaneous recognition of objects, events and locations. However, to a computer an image does not mean anything, unless it is told how to interpret it. Often images are converted into low-level features, which ideally capture the image characteristics in such a way that it is easy for the retrieval system to determine how similar two images are as perceived by the user. In current research, the attention is shifting to mid-level and high-level image representations.

Mid-level representations focus on particular parts of the image that are important, such as sub-images [52], regions [53,54] and salient details [36,55]. After these image elements have been determined, they are often seen as stand-alone entities during the search. However, some approaches represent them in a hierarchical [43,56,57] or graph-based structure and exploit this structure when searching for improved retrieval results. The multiple instance learning and bagging approach [37,58–61] lends itself very well to image retrieval, because an image can be seen as a bag of visual words where these visual words can, for instance, be interest points, regions, patches or objects (see Fig. 6). By incorporating feedback, the idea is that the user can only give feedback on the entire bag (i.e. the image), although he might only be interested in one or more specific instances (i.e. visual words) in that bag. The goal is then for the system to obtain a hypothesis from the feedback images that predicts which visual words the user is looking for. An unconventional way of using bags is presented in [62], where the multiple instance learning technique does not assume that a bag is positive when one or more of its instances are positive.

High-level representations are designed with semantics in mind. The way semantics are expressed is usually in the form of concepts, which are commonly seen as a coherent collection of image patches ('visual concepts') or sometimes as the equivalent of keywords ('textual concepts'). The number of visual concepts present in an image collection can be fixed beforehand [63,64], estimated beforehand [57,65],

or alternatively automatically determined while the system is running using adaptive approaches [66,67]. A thesaurus, such as WordNet [68], is often used to link annotations to image concepts [69,70], for instance by linking them through synonymy, hypernymy, hyponymy, etc. [71] (see Fig. 7). Since manually annotating large collections of images is a tedious task, much research is directed at automatic annotation, mostly offline [72,73], but also driven by relevance feedback [74]. Finding the best balance between using keywords for searching and using visual features for searching is one of the newer topics in image retrieval [75,76]. For instance, in [40] the image ranking presented to the user is composed first using a textual query vector to rank all database images and then using a visual query vector to re-rank them.

3.2 Indexing and filtering

Finding images that have high similarity with a query image often requires the entire database to be traversed for one-on-one comparisons. When dealing with large image collections this becomes prohibitive due to the amount of time the traversal takes. In the last few decades various indexing and filtering schemes have been proposed to reduce the number of database images to look at, thus improving the responsiveness of the system as perceived by the user. A good theoretical overview of indexing structures that can be used to index high-dimensional spaces is given in [77].

The majority of recent research in this direction focuses on the clustering of images, so that a reduction of the number of images to consider is then a matter of finding out which cluster(s) the query image belongs to [14,78,79]. Often the image clusters are stored in a hierarchical indexing structure to allow for a step-wise refinement of the number of images to consider [80,81]. Alternatively, the set of images that are likely relevant to the query can be quickly established by approximating their feature vectors [52,82]. A third way to reduce the number of images to inspect is by partitioning the feature space and only looking at that area of space which the query image belongs to [83,84]. Hashing is a form of space partitioning and is considered to be an efficient approach for indexing [85–87].

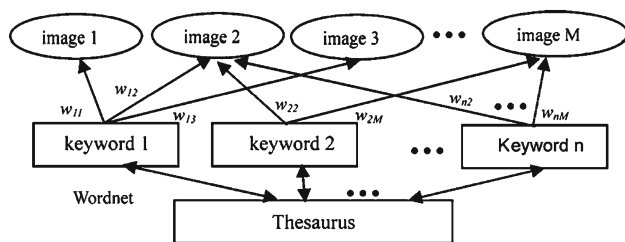


Fig. 7 A thesaurus is used to link keywords to images [74]

3.3 Active learning and classification

The core of the retrieval system is the algorithm that learns which images in the database the user is interested in by analyzing the query image and any implicit or explicit feedback. Typical interactive systems have two categories of images to show the user: (1) *clarification images*, which are images that may not be wanted by the user but that will help the learning algorithm improve its accuracy, and (2) *relevant images*, which are the images wanted by the user. How to decide which imagery to select for the first category is addressed by an area called “active learning”, which we first describe in more detail below.

Active learning Arguably, the most important challenge in interactive search systems is how to reduce the interaction effort from the user while maximizing the accuracy of the results. From a theoretical perspective, how can we measure the information associated with an unlabeled example, so a learner can select the optimal set of unlabeled examples to show to the user that maximizes its information gain and thus minimizes the expected future classification error [88–91]?

This category as pertaining to image search is usually called *active learning* in the research community and is closely related to *relevance feedback*, which many consider to be a special case of active learning. Especially during the last few years researchers are going beyond just selecting the unlabeled examples closest to the decision boundary by also aiming to maximize diversity amongst the chosen images [71, 92–94]. For instance, by trying to avoid selecting examples with certain visual properties that are already overly present in the list of top-ranked images [18] or by clustering the unlabeled candidate images by their similarity, so only a few examples per cluster need to be picked [95–97].

When multiple learners are used, a typical strategy is to select unlabeled examples for which the learners disagree the most in terms of their labeling [98–101]. With large image databases being commonplace, another focus in recent years has been placed on strategies to reduce the computational complexity [102], in particular, by filtering out unlabeled examples that are unlikely to contribute much to the decision boundary, so less examples need to be considered by the

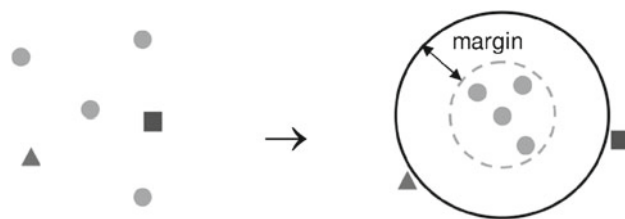


Fig. 8 A manifold is learned by projecting the relevant images close together and the irrelevant ones far away [118]

active learning algorithm [103, 104]. Integrating large external knowledge databases [24, 105, 106] into the search algorithm has seen increasing attention. These systems frequently use the external databases such as the WWW, Wikipedia, or social media networks to provide clarification of the user intent [107] or to form additional links/connections between imagery and multimodal information towards minimizing the number of queries to the user [71].

In the literature we can find diverse and interesting approaches for improving the feature space. Feature selection and manifold learning can reduce the complexity of the feature space and improve the shape of the clusters to make the relevance problem easier to learn by the classifier. The inclusion of synthetic imagery in the feedback process can be especially beneficial towards assisting in active learning. Recent work in each these directions is described below.

Feature selection and weighting One of the ways to discover the hidden information from the user’s feedback is let the search mainly focus on those features that feedback images have in common [108–110]. The feature space can also be transformed to discover hidden properties amongst relevant images, which is often done using principal component analysis [111], discriminant component analysis [112] or linear discriminant analysis [113]. One of the drawbacks of linear discriminant analysis is that negative feedback is treated as belonging to a single class, which is why researchers currently focus on multi-class [114] or biased [115] extensions to improve retrieval performance.

Manifold learning Manifold learning aims to learn the local structure formed by the query and feedback images, by creating a subspace where the relevant images are projected close together while the irrelevant images are projected far away (see Fig. 8). The most promising and popular approaches are currently based on linear extensions of graph embedding [116–120], which mostly differ in their choices of the affinity graph and the constraint graph.

Synthetic and pseudo-imagery An interesting development is the use of synthetic or pseudo-imagery during relevance feedback to improve the search results [11, 121–124]. When the system wants to ask the user about a particular region of



Fig. 9 Example of synthetic imagery such as used in [11], where several images are synthesized containing an object in different arrangements

feature space to clarify the decision boundary, there may not be a suitable image in the database due to the sparsity of images compared to the dimensionality of the feature space. By giving the system the ability to synthesize imagery corresponding to a point in feature space, the system can then clarify the uncertain area, as subsequent feedback on these synthetic images would allow the system to better narrow down what the user is looking for (see Fig. 9).

As the user interacts with the system and gives it positive and/or negative feedback, this feedback can be given to learning algorithms to address the classification of images as relevant images, which can then be cast as a classic machine learning problem:

- Cluster approaches: methods which represent the clusters of the images in feature space, such as query point or nearest neighbor-based learning.
- Decision plane approaches: methods which represent the decision planes between clusters of images, such as artificial neural networks, support vector machines and kernel approaches.
- Combining learners: methods that combine multiple classifiers to improve the overall accuracy.

There is extensive literature describing the theory and motivation for the methods above, which is beyond the scope of this survey. We restrict ourselves to concise descriptions of recent developments in this area.

Artificial neural networks One of the popular approaches is the RBF network [125, 126], which uses radial basis functions as activation functions. These functions have the advantage over sigmoids that generally only one layer of hidden radial units is sufficient to model any function. Another popular approach is the self-organizing map [127, 128], which in contrast with other kinds of neural networks does not need supervision during training. It projects the high-dimensional feature vectors down to only a few dimensions, typically two. Feedback causes the relevance information to spread to the neighboring units, based on the assumption that similar images are located near each other on the map surface. The

spreading of the relevance values happens by convolving the surface with window or kernel functions (see Fig. 10).

Support vector machine The current trend is the development of techniques that aim to overcome the inherent limitations of standard SVMs, such as targeting the imbalanced training set [127, 129, 130], filtering out noisy feedback [131], reducing the amount of computation necessary between rounds of feedback [132] or offering more flexibility in the labeling of examples [133]. For instance, a fuzzy SVM [134] uses the fuzzy class membership values to reduce the effect of less important examples, so that the examples with higher confidence have a larger effect on the decision boundary.

Kernels Many approaches, such as support vector machines, use kernels to convert the feature space to a higher- or lower-dimensional space, where ideally the images of interest can be linearly separated from all other images. We show the popularity of common kernel variations in Table 1. The kernel that is used is generally fixed, i.e. the type of kernel and its parameters are determined beforehand, although particularly in recent work positive and negative feedback is used to guide the design and/or selection of optimal kernels [135–137].

Combining learners Instead of using a single learner to classify an unlabeled image, multiple independent learners can be combined to obtain a better classification, e.g. by combining their individual decision functions into an overall decision function [138, 139], by majority voting [110, 130, 134] or by selecting the most appropriate learner(s) for a particular query [140].

Probabilistic classifiers Mixture models [141, 142] are used to overcome the limitations of using only a single density function to model the relevant class. Mixture models are a combination of multiple probabilistic distributions, where the number of distributions (components) it comprises is ideally identical to the number of classes present in the data. Other approaches in this category aim to learn the probabilistic model and unconditional density of the positive and/or negative classes [143, 144].

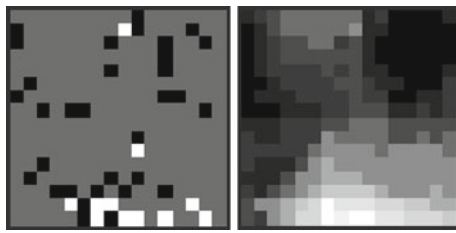


Fig. 10 The positive (*white*) and negative (*black*) map units in a self-organizing map (*left*) are convolved with a low-pass filter mask, leading to the relevance values being spread across the map surface (*right*) [128]

Classification approaches Some methods directly assign relevance scores to each image in the database, whereas other methods attempt to classify the images using a *one-class* approach, where a model is built for only the relevant class [58], or a *two-class* approach, where a model is built that either classifies an image as positive or as negative [145]. Other variations exist that allow for more flexibility, for instance $1 + x$ [92], $x + 1$ [138], $x + y$ [49] and *soft label* [146]. The popularity of the classification approaches as used in the recent literature is shown in Table 2.

3.4 Similarity measures, distance and ranking

What matters the most in image retrieval is the list of results that is shown to the user, with the most relevant images shown first. In general, to obtain this ranking a similarity measure is used that assigns a score to each database image indicating how relevant the system thinks it is to the user's interests. The advantages and disadvantages of using a metric to measure perceptual similarity are discussed in [147], in which the authors argue for incorporating the notion of *betweenness* when ranking images to allow for a better relative ordering between them. Ways of calculating scores include using the relative distance of an image to its nearest relevant and nearest irrelevant neighbors [148, 149] or combining multiple similarity measures to give a single relevance score [59, 150]. Relevance feedback can also be considered to be an ordinal regression problem [23, 151], where users do not give an absolute but rather a relative judgment between images.

We show the popularity of common similarity measures in Table 3. As can be seen the Euclidean (L_2) distance measure is used most frequently, although in a substantial number of papers it was only used in the initial iteration and a more advanced similarity measure was applied once feedback was received. Many similarity measures are tailored to the problem to solve and thus quite specialized, which are therefore not included in the table.

3.5 Long-term learning

In contrast with short-term learning, where the state of the retrieval system is reset after every user session, long-

Table 1 Popularity of kernel variations

Linear	Polynomial	Triangular	Gaussian	Laplacian
1 %	12 %	6 %	73 %	8 %

Table 2 Popularity of classification approaches

One-class	Two-class	$1 + x$	$x + 1$	$x + y$	Soft label
32 %	38 %	18 %	2 %	2 %	8 %

term learning is designed to use the information gathered during previous retrieval sessions to improve the retrieval results in future sessions. Long-term learning is also frequently referred to as collaborative filtering. The most popular approach for long-term learning is to infer relationships between images by analyzing the *feedback log* [52, 79, 152], which contains all feedback given by users over time. From the accumulated feedback logs a semantic space can be learned containing the relationships between the images and one or more classes, typically obtained by applying matrix factorization [153–155] or clustering [156] techniques. Whereas the early long-term learning methods mostly built static relevance models, the recent trend is to continuously update the model after receiving new feedback [157–160].

3.6 Trends and advances

It is generally agreed upon that minimizing the number of questions that need to be asked (small training set problem) is one of the grand challenges. Over the past decade we have seen several different trends that include, but are not limited to, (1) query point movement, (2) query set movement, (3) input near decision borders, and (4) input reflecting additional information sources. By query point movement, we refer to the Rocchio [9] inspired methods where a single query point is shifted towards the positive examples and away from the negative examples. This paradigm has worked surprisingly well when there is little feedback; however, it has a notable problem that it cannot adjust to multiple clusters of relevant results. This led to query set movement approaches, which move multiple query points that ideally end up in each relevant cluster in the database; yet, this method has distinct weaknesses when there are many clusters or when the class separation between positive and negative clusters is small. In reaction, the research community investigated decision border approaches where the user was asked to clarify the ambiguous regions near the borders. In a large image database, however, the number of decision borders can be very large, so that even in the simplest case where the system needs to get feedback for every decision border this can result in

Table 3 Popularity of common similarity measures

Manhattan	Euclidean	Minkowski	Earth Mover's distance	Bhattacharyya	Mahalanobis	Hausdorff	Kullback-Leibler
11 %	49 %	1 %	4 %	1 %	5 %	2 %	4 %
Chi-square	Probability	Cosine	Graph	Unified feature matching	Integrated region matching	dN/dR	
3 %	10 %	2 %	3 %	1 %	1 %	3 %	

Note that (a) Minkowski refers to all similarity measures in its family other than Manhattan and Euclidean, (b) Probability refers to similarity measures that calculate the likelihood of an image belonging to the target category, (c) Graph refers to similarity measures that determine the shortest path between two nodes in a graph, (d) dR/dN refers to similarity measures that compare the relative distance of an image to its nearest relevant and nearest irrelevant neighbors

an overload of questions to the user. This, in turn, has led to methods which attempt to gain clarification by exploiting additional or external sources, such as personal history, the Internet, or Wikipedia. Another challenge has been shown to be the problem of sparsity in the image database which has recently been addressed by using both external sources and synthetic imagery.

From the articles published during the last decade we can see the perception of image retrieval slowly shifting from pixel-based to concept-based, especially because it generally has led to an increase in retrieval performance. This new concept-based view has inspired the development of many new high-level descriptors. The bag-of-words and manifold learning approaches remain popular, and especially the latter has become a particularly active research area, providing a stimulating and competitive research environment. Long-term learning and approaches that combine multiple information sources have also demonstrated steady and significant improvements in retrieval performance over the previous years. Rocchio [9] approaches are currently only used for comparative benchmarks relative to a novel algorithm.

4 Evaluation and benchmarking

Assessing user satisfaction and general evaluation of interactive retrieval systems [7, 161, 162] is well known to be both difficult and challenging. Experiments that are well executed from a statistical point of view require a relatively large number of diverse and independent participants. In our field such studies are rarely performed, although this is understandable due to the difficulty in obtaining cooperation from a large number of users and in the rapidly advancing technological nature of our research. More often than not our experiments limit themselves to a group of (frequently computer science) students [81] or use a computer simulation of user behavior [163]. Simulated users are easy to create, allow for the experiments to be performed quickly and give a rough indication of the performance of the retrieval system. How-

Table 4 User-based evaluation of experiments

Real users	Simulated users	No evaluation
13 %	84 %	3 %

ever, these simulated users are, in general, too perfect in their relevance judgments and do not exhibit the inconsistencies (e.g. mistakenly labeling an image as relevant), individuality (e.g. two users have a different perception of the same image) and laziness (e.g. not wanting to label many images) of real users. By involving simulated users, we can very well end up with skewed results. In Table 4, we show how the experiments are evaluated in current research. As can be seen, the majority of experiments is conducted with simulated users, with only a small number of experiments involving real users. Some works provide no evaluation, because they present a novel idea and only show a proof of concept.

A brief look at current ways of evaluating interactive search systems is covered in [164] and an in-depth review can be found in [165], where guidelines are additionally suggested on how to raise the standard of evaluation. An evaluation benchmarking framework is proposed in [166], so relevance feedback algorithms can be fairly compared with each other.

4.1 Image databases

There is a large variation in the image databases used by the research community that focuses on interactive search. Photographic imagery is the most popular kind of imagery. From our study, the Corel stock photography image set (e.g. [167]) has been used most frequently because it was the first large image set which could be considered representative for real world usage. However, it is also known to have significant and diverse problems [167] and that it is both illegal to distribute and is no longer sold. The copyright situation of the Corel image set motivated the research community to

Table 5 Most popular databases used in image retrieval using interactive search

Rank	Name	Institute	Type	No. of images
1	Corel [167]	Corel	Stock photo	80K
2	MIRFLICKR [168, 169]	Leiden University	General photo	1,000K
3	Brodatz [170]	Brodatz	Texture	1K
4	Ponce [171]	University of Illinois at Urbana-Champaign	Texture	1K
5	VisTex [172]	Massachusetts Institute of Technology	Texture	<1K
6	Caltech-256 [173]	California Institute of Technology	Object	30K
7	PASCAL VOC [174]	VOC Challenge	Object	12K
8	ImageCLEF Medical [175]	ImageCLEF	Medical	231K
9	Columbia-COIL [176]	Columbia University	Object	2K

Table 6 Performance-based evaluation of experiments

Precision-recall	Precision	Recall	Mean average precision	Retrieval time	Other
15%	44%	24%	4%	7%	6%

create large representative image sets which were both legal to redistribute and easily downloadable, such as the MIRFLICKR [168, 169] sets that contain images collected from thousands of users from the photo sharing website Flickr. The list of most popular databases used in image retrieval from our literature search is shown in Table 5 from most frequently to least frequently used. Please note that many of the databases grow over time so the most current version will often be larger than the number listed.

4.2 Performance measures

Recently, several new performance measures have been proposed [177]. A notable measure is *generalized efficiency* [165], which normalizes the performance of a feedback method using the optimal classifier performance. This measure is particularly useful for benchmarking several methods with respect to a baseline method. Table 6 shows the popularity of current methods to evaluate retrieval performance. As can be seen *precision* is the most popular evaluation method, with *recall* second most popular and the combined *precision-recall* as third.

4.3 Trends and advances

Standardization has received significantly greater attention during the past years. We have witnessed several efforts to fulfill this need, ranging from benchmarking frameworks to standard image databases, such as the recent test sets that aim to provide researchers with a large number of images that are well-annotated and free of copyright. Considering that the volume of digital media in the world is rapidly expanding, having access to large image collections for training and testing new algorithms is important because it is not clear

which algorithms scale well to millions. In the recent years, researchers have been moving away from the Corel image database and started creating open access databases for specific areas in image retrieval.

5 Discussion and conclusions

Over the years, we have seen the performance of interactive search systems steadily improve. Nonetheless, much research remains to be done. In this section, we will discuss the most promising research directions and identify several open issues and challenges.

5.1 Promising research directions

Below we outline top research directions that, based on our literature review, are on the frontier of interactive search.

– *Interaction in the question and answer paradigm*

The Q&A paradigm has the strength that it is probably the most natural and intuitive for the user. Recent Q&A research has focused significantly more on multimodal (as opposed to monomodal) approaches for both posing the questions and displaying the answers. These systems can also dynamically select the best types of media for clarifying the answer to a specific question.

– *Interaction on the learned models*

Beyond giving direct feedback on the results, preliminary work was started involving mid-level and high-level representations (see Sect. 3). Multi-scale approaches using

segmented image components are certainly novel and promising.

– *Interaction by explanation: providing reasons along with results*

In the classic relevance feedback model, results are typically given but it is not clear to the user why the results were selected. In future interactive search systems, we expect to see systems which explain to the user why the results were chosen and allow the user to give feedback on the criteria used in the explanations, as opposed to only simply giving feedback on the image results.

– *Interaction with external or synthesized knowledge sources*

In the prior work in this area, most of the systems limited themselves only to the imagery in the local collection. However, it has been found that utilizing additional image collections and knowledge sources can significantly improve the quality of results. Currently, using very large multimedia databases such as Wikipedia as external knowledge sources is an active and fertile direction.

– *Social interaction: recommendation systems and collaborative filtering*

The small training set problem is of particular concern because humans do not want to label thousands of images. An interesting approach is to examine potential benefits from using algorithms from the area of collaborative filtering and recommendation systems. These systems have remarkably high performance in deciding which media items (often video) will be of interest to the user based on a social database of ranked items.

5.2 Grand challenges

The past decade has brought many scientific advances in interactive image search theory and techniques. Moreover, there has been significant societal impact through the adoption of interactive image search in the largest WWW image search engines (Google, Bing, and Yahoo!), as well as in numerous systems in application areas such as medical image retrieval, professional stock photography databases, and cultural heritage preservation. Arguably, interactive search is the most important paradigm, because in a human sense it is the most effective method for us, while in a theoretical sense it allows the system to minimize the information required for answering a query by making careful choices about the questions to pose to the user. In conclusion, the grand challenges can be summarized as follows:

1. *What is the optimal user interface and information transfer for queries and results?*

Our current systems usually seek to minimize the number of user labeled examples or the search time on the assumption that it will improve the user satisfaction or experience. A fundamentally different perspective is to focus on the user experience. This means that other aspects than accuracy may be considered important, such as the user's satisfaction/enjoyment or the user's feeling of understanding why the results were given. A longer search time might be preferable if the overall user experience is better. Recent developments in the industry have led to new interfaces that may be more intuitive. For example, touch-based technology has become intuitive and user-friendly through the popularity of smart phones and tablets. These developments open up new interaction possibilities between the search engine and the user. Novel interfaces can be potentially created that deliver a better search experience to such devices, while at the same time reaching a large number of users. Now that the Web 2.0, the social internet, is also becoming more and more prevalent, techniques that analyze the content produced by users all over the world show great promise to further the state of the art. The millions of photos that are commented on and tagged on a daily basis can provide invaluable knowledge to better understand the relations between images and their content.

2. *How can we achieve good accuracy with the least number of training examples?*

The most commonly cited challenge in the research literature is the small training set problem, which means that, in general, the user does not want to manually label a large number of images. Developing new learning algorithms and/or integrating knowledge databases that can give good accuracy using only a small set of user-labeled images is perhaps the most important grand challenge of our field. Other promising techniques include manifold learning, multimodal fusion and utilizing implicit feedback. Novel learning algorithms are being regularly developed in the machine learning and the neuroscience fields. A particularly interesting direction comes from spiking networks and BCM theory [178], which conceivably is the most accurate model of learning in the visual cortex. Another recent novel direction is that of synthetic imagery.

3. *How should we evaluate and improve our interactive systems?*

Evaluation projects in interactive search systems are in their infancy. There are several major issues to address in how to create or obtain high-quality ground truth for real image search contexts. One major issue is the way in which evaluation benchmarks are constructed. The current ones typically focus on the overall performance/accuracy of a search engine. However, it would

be of significantly greater value if they could focus on benchmarks which give insight into each system's weaknesses and strengths.

Another issue is to determine what kinds of results are satisfactory to a user. For assessing the performance of a system, precision- and recall-based performance measures are the most popular choices at the moment. However, the research literature has shown that these measures are unable to provide a complete assessment of the system under study and argues that the notion of *generality*, i.e. the fraction of relevant items in the database, should be an important criterion when evaluating and comparing the performance of systems.

A third issue is that currently researchers are largely guessing what kinds of imagery users are interested in, the kinds of queries and also the amount of effort (and other behavioral aspects) the user is willing to expend on a search. Currently, most researchers attempt to use simulated users to test their algorithms, while knowing that the simulated behavior may not mirror human user behavior. While simulations are very useful to get an initial impression on the performance of a new algorithm, they cannot replace actual user experiments since retrieval systems are specifically designed for users. One valuable direction for further study would thus be to properly model the behavior of simulated users after their real counterparts. It is noteworthy that the user behavior information largely exists in the logs of the WWW search engines. Thus, on the one hand, as a research community, we would like to have the user history from large search engines such as Yahoo! and Google. On the other hand, we realize that there are many legal concerns (e.g. user privacy) that prevent this information from being distributed. Finding a solution to this impasse could result in major improvements in interactive image search engines.

Acknowledgments Leiden University and NWO BSIK/BRICKS supported this research under Grant #642.066.603.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. Andre P, Cutrell E, Tan D, Smith G (2009) Designing novel image search interfaces by understanding unique characteristics and usage. In: Proceedings of international conference on human-computer interaction, vol 2, pp 340–353
2. Ren K, Sarvas R, Calic J (2010) Interactive search and browsing interface for large-scale visual repositories. *Multimedia Tools Appl* 49:513–528
3. Zhou X, Zillner S, Moeller M, Sintek M, Zhan Y, Krishnan A, Gupta A (2008) Semantics and CBIR: a medical imaging perspective. In: Proceedings of ACM international conference on image and video retrieval, pp 571–580
4. 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 Machine Intell* 22(12):1349–1380
5. Datta R, Li J, Wang JZ (2005) Content-based image retrieval: approaches and trends of the new age. In: Proceedings of ACM international workshop on multimedia, information retrieval, pp 253–262
6. Datta R, Joshi D, Li J, Wang JZ (2008) Image retrieval: ideas, influences, and trends of the new age. *ACM Comput Surv* 40(2): 1–60
7. Lew MS, Sebe N, Djeraba C, Jain R (2006) Content-based multimedia information retrieval: state of the art and challenges. *ACM Trans Multimedia Comput Commun Appl* 2(1):1–19
8. Huang TS, Dagli CK, Rajaram S, Chang EY, Mandel MI, Poliner GE, Ellis DPW (2008) Active learning for interactive multimedia retrieval. *Proc IEEE* 96(4):648–667
9. Zhou XS, Huang TS (2003) Relevance feedback in image retrieval: a comprehensive review. *ACM Multimedia Syst* 8(6):536–544
10. Kherfi ML, Brahmi D, Ziou D (2004) Combining visual features with semantics for a more effective image retrieval. In: Proceedings of IEEE international conference on pattern recognition, vol 2, pp 961–964
11. Aggarwal G, Ashwin TV, Ghosal S (2002) An image retrieval system with automatic query modification. *IEEE Trans. on Multimedia* 4(2):201–214
12. Thomee B, Huiskes MJ, Bakker EM, Lew MS (2009) Deep exploration for experiential image retrieval. In: Proceedings of ACM International Conference on Multimedia, pp 673–676
13. Kutics A, Nakagawa A, Tanaka K, Yamada M, Sanbe Y, Ohtsuka S (2003) Linking images and keywords for semantics-based image retrieval. In: Proceedings of IEEE international conference on multimedia and expo, vol 1, pp 777–780
14. Chiang C-C, Hsieh M-H, Hung Y-P, Lee GC (2005) Region filtering using color and texture features for image retrieval. In: Proceedings of ACM conference on image and video retrieval, pp 487–496
15. Amores J, Sebe N, Redeva P, Gevers T, Smeulders A (2004) Boosting contextual information in content-based image retrieval. In: Proceedings of ACM international workshop on multimedia, information retrieval, pp 31–38
16. Ko BC, Byun H (2002) Integrated region-based image retrieval using region's spatial relationships. In: Proceedings of IEEE international conference on pattern recognition, vol 1, pp 196–199
17. Torres JM, Hutchison D, Reis LP (2007) Semantic image retrieval using region-based relevance feedback. In: Proceedings of international workshop on adaptive multimedia retrieval: user, context, and, feedback, pp 192–206
18. Huiskes MJ (2006) Image searching and browsing by active aspect-based relevance learning. In: Proceedings of international conference on image and video retrieval, pp 211–220
19. Jin X, French JC (2003) Improving image retrieval effectiveness via multiple queries. In: Proceedings of ACM international workshop on multimedia databases, pp 86–94
20. Zhang C, Chen X (2005) Region-based image clustering and retrieval using multiple instance learning. In: Proceedings of international conference on image and video retrieval, pp 194–204
21. Yang J, Li Q, Zhuang Y (2002) Image retrieval and relevance feedback using peer indexing. In: Proceedings of IEEE international conference on multimedia and expo, vol 2, pp 409–412
22. Ko BC, Byun H (2002) Probabilistic neural networks supporting multi-class relevance feedback in region-based image retrieval. In: Proceedings of IEEE international conference on pattern recognition, vol 4, pp 138–141

23. Wu H, Lu H, Ma S (2004) WillHunter: interactive image retrieval with multilevel relevance measurement. In: Proceedings of IEEE international conference on pattern recognition, vol 2, pp 1009–1012
24. Haas M, Oerlemans A, Lew MS (2005) Relevance feedback methods in content based retrieval and video summarization. In: Proceedings of IEEE international conference on multimedia and expo, pp 1038–1041
25. Huang X, Chen S-C, Shyu M-L (2003) Incorporating real-valued multiple instance learning into relevance feedback for image retrieval. In: Proceedings of IEEE international conference on multimedia and expo, vol 2, pp 321–324
26. Li G, Ming Z, Li H, Chua T (2009) Video reference: question answering on YouTube. In: Proceedings of ACM international conference on multimedia, pp 773–776
27. von Ahn L, Dabbish L (2004) Labeling images with a computer game. In: Proceedings of ACM conference on human factors in computing systems, pp 319–326
28. Andrenucci A, Sneiders E (2005) Automated question answering: review of the main approaches. In: Proceedings of IEEE international conference on information technology and applications, vol 1, pp 514–519
29. Sahbi H, Etyngier P, Audibert J, Keriven R (2008) Manifold learning using robust graph Laplacian for interactive image search. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 1–8
30. Xu H, Wang J, Hua X (2010) Interactive image search by 2D semantic map. In: Proceedings of ACM international conference on World Wide Web, pp 1321–1324
31. Meng J, Yuan J, Jiang Y, Narashimhan N, Vasudevan V, Wu Y (2010) Interactive visual object search through mutual information maximization. In: Proceedings of ACM international conference on multimedia, pp 1147–1150
32. Wang C, Li Z, Zhang L (2010) MindFinder: image search by interactive sketching and tagging. Proc. ACM International Conference on, World Wide Web, pp 1309–1312
33. Cao Y, Wang C, Zhang L (2011) Edgel index for large-scale sketch-based image search. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 761–768
34. Nie L, Wang M, Zha Z, Li G, Chua T (2011) Multimedia answering: enriching text QA with media information. In: Proceedings of ACM conference on research and development in information retrieval, pp 695–704
35. Yeh T, Lee J, Darrell T (2008) Photo-based question answering. In: Proceedings of ACM international conference on multimedia, pp 389–398
36. Nguyen GP, Worring M (2005) Relevance feedback based saliency adaptation in CBIR. ACM Multimedia Syst 10(6):499–512
37. Tran DA, Pamidimukkala SR, Nguyen P (2008) Relevance-feedback image retrieval based on multiple-instance learning. In: Proceedings of IEEE international conference on computer and information science, pp 597–602
38. Kherfi ML, Ziou D, Bernardi A (2002) Learning from negative example in relevance feedback for content-based image retrieval. In: Proceedings of IEEE international conference on pattern recognition, vol 2, pp 933–936
39. Liu J, Li Z, Li M, Lu H, Ma S (2007) Human behaviour consistent relevance feedback model for image retrieval. In: Proceedings of ACM international conference on multimedia, pp 269–272
40. Cheng E, Jing F, Zhang L (2009) A unified relevance feedback framework for web image retrieval. IEEE Trans Image Process 18(6):1350–1357
41. Campbell I (2000) Interactive evaluation of the ostensive model using a new test collection of images with multiple relevance assessments. J Inf Retrieval 2:87–114
42. Urban J, Jose JM, Rijsbergen CJ (2006) An adaptive technique for content-based image retrieval. Multimedia Tools Appl 31(1):1–28
43. Fan J, Gao Y, Luo H, Jain R (2008) Mining multilevel image semantics via hierarchical classification. IEEE Trans Multimedia 10(2):167–181
44. Thomee B, Huiskes MJ, Bakker EM, Lew MS (2009) An exploration-based interface for interactive image retrieval. In: Proceedings of IEEE international symposium on image and signal processing, pp 192–197
45. Mavandadi S, Aarabi P, Khaleghi A, Appel R (2006) Predictive dynamic user interfaces for interactive visual search. In: Proceedings of IEEE international conference on multimedia and expo, pp 381–384
46. Zavesky E, Chang S-F, Yang C-C (2008) Visual islands: intuitive browsing of visual search results. In: Proceedings of ACM international conference on image and video retrieval, pp 617–626
47. Nguyen GP, Worring M (2008) Optimization of interactive visual-similarity-based search. ACM Trans Multimedia Comput Commun Appl 4(1):499–512
48. Wang X, McKenna SJ, Han J (2009) High-entropy layouts for content-based browsing and retrieval. In: Proceedings of ACM international conference on image and video retrieval, article 16
49. Nakazato M, Huang TS (2002) Extending image retrieval with group-oriented interface. In: Proceedings of IEEE international conference on multimedia and expo, vol 1, pp 201–204
50. Urban J, Jose JM (2007) Evaluating a workspace’s usefulness for image retrieval. ACM Multimedia Syst 12(4–5):355–373
51. Guan J, Qiu G (2007) Learning user intention in relevance feedback using optimization. In: Proceedings of ACM international workshop on multimedia, information retrieval, pp 41–50
52. Shyu M, Chen S-C, Chen M, Zhang C, Sarinnapakorn K (2003) Image database retrieval utilizing affinity relationships. In: Proceedings of ACM international workshop on multimedia databases, pp 78–85
53. Sun Y, Ozawa S (2005) HIRBIR: a hierarchical approach to region-based image retrieval. ACM Multimedia Syst 10(6):559–569
54. Chen Y, Wang JZ (2002) A region-based fuzzy feature matching approach to content-based image retrieval. IEEE Trans Pattern Anal Mach Intell 24(9):1252–1267
55. Ko BC, Kwak SY, Byun H (2004) SVM-based salient region(s) extraction method for image retrieval. In: Proceedings of IEEE international conference on pattern recognition, vol 2, pp 977–980
56. Luo J, Nascimento MA (2004) Content-based sub-image retrieval using relevance feedback. In: Proceedings of ACM international workshop on multimedia databases, pp 2–9
57. Zhang R, Zhang Z (2004) Hidden semantic concept discovery in region based image retrieval. Proc. IEEE Conference on Computer Vision and Pattern Recognition 2:996–1001
58. Chen X, Zhang C, Chen S-C, Chen M (2005) A latent semantic indexing based method for solving multiple instance learning problem in region-based image retrieval. In: Proceedings of IEEE international symposium on multimedia, pp 37–45
59. Rahmani R, Goldman SA, Zhang H, Cholleti SR, Fritts JE (2008) Localized content-based image retrieval. IEEE Trans Pattern Anal Mach Intell 30(11):1902–1912
60. Fu Z, Robles-Kelly A (2009) An instance selection approach to multiple instance learning. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 911–918
61. Sivic J, Russell BC, Efros AA, Zisserman A, Freeman WT (2005) Discovering objects and their location in images. In: Proceedings of IEEE international conference on computer vision, vol 1, pp 370–377

62. Chen Y, Bi J, Wang JZ (2006) Miles: multiple-instance learning via embedded instance selection. *IEEE Trans Pattern Anal Mach Intell* 28(12):1–17
63. Chatzis S, Doulamis A, Varvarigou T (2007) A content-based image retrieval scheme allowing for robust automatic personalization. In: *Proceedings of ACM international conference on image and video retrieval*, pp 1–8
64. Lim J-H, Jin JS (2005) A structured learning framework for content-based image indexing and visual query. *ACM Multimedia Syst* 10(4):317–331
65. Dong A, Bhanu B (2003) A new semi-supervised EM algorithm for image retrieval. In: *Proceedings of IEEE conference on computer vision and pattern recognition*, vol 2, pp 662–667
66. Dong A, Bhanu B (2003) Active concept learning for image retrieval in dynamic databases. In: *Proceedings of IEEE international conference on computer vision*, pp 90–95
67. Fung CC, Chung K-P (2007) Establishing semantic relationship in inter-query learning for content-based image retrieval systems. In: *Proceedings of Pacific-Asia conference on knowledge discovery and data mining*, pp 498–506
68. Fellbaum C (1998) *WordNet: an electronic lexical database*. MIT Press, Cambridge
69. Lu Y, Zhang H-J, Wenyin L, Hu C (2003) Joint semantics and feature based image retrieval using relevance feedback. *IEEE Trans Multimedia* 5(3):339–347
70. Yang C, Dong M, Fotouhi F (2005) Semantic feedback for interactive image retrieval. In: *Proceedings of ACM international conference on multimedia*, pp 415–418
71. Ferecatu M, Boujema N, Crucianu M (2008) Semantic interactive image retrieval combining visual and conceptual content description. *ACM Multimedia Syst* 13(5–6):309–322
72. Liu X, Cheng B, Yan S, Tang J, Chua TS, Jin H (2009) Label to region by bi-layer sparsity priors. In: *Proceedings of ACM international conference on multimedia*, pp 115–124
73. Lu Z, Ip HHS, He Q (2009) Context-based multi-label image annotation. In: *Proceedings of ACM international conference on image and video retrieval*, article 30
74. Zhang H-J, Chen Z, Li M, Su Z (2003) Relevance feedback and learning in content-based image search. *J World Wide Web* 6(2):131–155
75. Urban J, Jose JM (2006) Adaptive image retrieval using a graph model for semantic feature integration. In: *Proceedings of ACM international workshop on multimedia, information retrieval*, pp 117–126
76. Wang X-J, Ma W-Y, Zhang L, Li X (2005) Multi-graph enabled active learning for multimodal web image retrieval. In: *Proceedings of ACM international workshop on multimedia, information retrieval*, pp 65–72
77. Böhm C, Berchtold S, Keim DA (2001) Searching in high-dimensional spaces: index structures for improving the performance of multimedia databases. *ACM Comput Surv* 33(3):322–373
78. Goh K-S, Li B, Chang EY (2002) DynDex: a dynamic and non-metric space indexer. In: *Proceedings of ACM international conference on multimedia*, pp 466–475
79. Zhou X, Zhang Q, Lin L, Deng A, Wu G (2003) Image retrieval by fuzzy clustering of relevance feedback records. In: *Proceedings of IEEE international conference on multimedia and expo*, vol 2, pp 305–308
80. Wang T, Rui Y, Hu S-M, Sun J-G (2003) Adaptive tree similarity learning for image retrieval. *ACM Multimedia Syst* 9(2):131–143
81. Zhang R, Zhang Z (2005) FAST: toward more effective and efficient image retrieval. *ACM Multimedia Syst* 10(6):529–543
82. Heisterkamp DR, Peng J (2005) Kernel vector approximation files for relevance feedback retrieval in large image databases. *Multimedia Tools Appl* 26(2):175–189
83. Tandon P, Nigam P, Pudi V, Jawahar CV (2008) FISH: a practical system for fast interactive image search in huge databases. In: *Proceedings of ACM international conference on image and video retrieval*, pp 369–378
84. Yu N, Vu K, Hua KA (2007) An in-memory relevance feedback technique for high-performance image retrieval systems. In: *Proceedings of ACM international conference on image and video retrieval*, pp 9–16
85. Indyk P, Motwani R (1998) Approximate nearest neighbors: towards removing the curse of dimensionality. In: *Proceedings of ACM symposium on theory of computing*, pp 604–613
86. Kuo Y-H, Chen K-T, Chiang C-H, Hsu WH (2009) Query expansion for hash-based image object retrieval. In: *Proceedings of ACM international conference on multimedia*, pp 65–74
87. Yang H, Wang Q, He Z (2008) Randomized sub-vectors hashing for high-dimensional image feature matching. In: *Proceedings of ACM international conference on multimedia*, pp 705–708
88. Jing F, Li M, Zhang H-J, Zhang B (2004) Entropy-based active learning with support vector machines for content-based image retrieval. In: *Proceedings of IEEE international conference on multimedia and expo*, vol 1, pp 85–88
89. Hoi SCH, Jin R, Zu J, Lyu MR (2009) Semisupervised SVM batch mode active learning with applications to image retrieval. *IEEE Trans Inf Syst* 27(3) (article 16)
90. He X (2010) Laplacian regularized D-Optimal Design for active learning and its application to image retrieval. *IEEE Trans Image Process* 19(1):254–263
91. Peng X, King I (2006) Biased minimax probability machine active learning for relevance feedback in content-based image retrieval. In: *Proceedings of intelligent data engineering and automated learning*, pp 953–960
92. Dagli CK, Rajaram S, Huang TS (2006) Leveraging active learning for relevance feedback using an information theoretic diversity measure. In: *Proceedings of international conference on image and video retrieval*, pp 123–132
93. Chang EY, Lai W-C (2004) Active learning and its scalability for image retrieval. In: *Proceedings of IEEE international conference on multimedia and expo*, vol 1, pp 73–76
94. Goh K-S, Chang EY, Lai W-C (2004) Multimodal concept-dependent active learning for image retrieval. In: *Proceedings of ACM international conference on multimedia*, pp 564–571
95. Yang J, Li Y, Tian Y, Duan L, Gao W (2009) Multiple kernel active learning for image classification. In: *Proceedings of IEEE international conference on multimedia and expo*, pp 550–553
96. Liu R, Wang Y, Baba T, Masumoto D, Nagata S (2008) SVM-based active feedback in image retrieval using clustering and unlabeled data. *Pattern Recogn* 41(8):2645–2655
97. Cord M, Gosselin PH, Philipp-Foliguet S (2007) Stochastic exploration and active learning for image retrieval. *Image Vis Comput* 25(1):14–23
98. Singh R, Kothari R (2003) Relevance feedback algorithm based on learning from labeled and unlabeled data. In: *Proceedings of IEEE international conference on multimedia and expo*, pp 433–436
99. Zhou Z-H, Chen K-J, Dai H-B (2006) Enhancing relevance feedback in image retrieval using unlabeled data. *ACM Trans Inf Syst* 24(2):219–244
100. Cheng J, Wang K Multi-view sampling for relevance feedback in image retrieval. In: *Proceedings of IEEE international conference on pattern recognition*, pp 881–884
101. Zhang X, Cheng J, Xu C, Lu H, Ma S (2009) Multi-view multi-label active learning for image classification. In: *Proceedings of*

- IEEE international conference on multimedia and expo, pp 258–261
102. Zhang X, Cheng J, Lu H, Ma S (2008) Selective sampling based on dynamic certainty propagation for image retrieval. In: Proceedings of international multimedia modeling conference, pp 425–435
 103. He X, Min W, Cai D, Zhou K (2007) Laplacian optimal design for image retrieval. In: Proceedings of ACM conference on research and development in information retrieval, pp 119–126
 104. Hörster E, Lienhart R, Slaney M (2007) Image retrieval on large-scale image databases. In: Proceedings of ACM conference on image and video retrieval, pp 17–24
 105. Popescu A, Grefenstette G (2011) Social media driven image retrieval. In: Proceedings of ACM international conference on multimedia retrieval (article 33)
 106. Rawashdeh M, Kim H, El Saddik A (2011) Folksonomy-boosted social media search and ranking. In: Proceedings of ACM international conference on multimedia retrieval (article 27)
 107. Hu J, Wang G, Lochovsky F, Sun J, Chen Z (2009) Understanding user's query intent with wikipedia. In: Proceedings of international conference on WWW, pp 471–480
 108. Das G, Ray S, Wilson C (2006) Feature re-weighting in content-based image retrieval. In: Proceedings of international conference on image and video retrieval, pp 193–200
 109. Grigorova A, De Natale FGB, Dagli CK, Huang TS (2007) Content-based image retrieval by feature adaptation and relevance feedback. *IEEE Trans Multimedia* 9(6):1183–1192
 110. Wu Y, Zhang A (2004) Interactive pattern analysis for relevance feedback in multimedia information retrieval. *ACM Multimedia Syst* 10(1):41–55
 111. Franco A, Lumini A, Maio D (2004) A new approach for relevance feedback through positive and negative samples. In: Proceedings of IEEE international conference on pattern recognition, vol 4, pp 905–908
 112. Hoi SCH, Liu W, Lyu MR, Ma W-Y (2006) Learning distance metrics with contextual constraints for image retrieval. In: Proceedings of IEEE conference on computer vision and pattern recognition, vol 2, pp 2072–2078
 113. Huang R, Liu Q, Lu H, Ma S (2002) Solving the small sample size problem of LDA. In: Proceedings of IEEE international conference on pattern recognition, vol 3, pp 29–32
 114. Yoshizawa T, Schweitzer H (2004) Long-term learning of semantic grouping from relevance-feedback. In: Proceedings of ACM international workshop on multimedia, information retrieval, pp 165–172
 115. Tao D, Tang X, Li X, Rui Y (2006) Direct kernel biased discriminant analysis: a new content-based image retrieval relevance feedback algorithm. *IEEE Trans Multimedia* 8(4):716–727
 116. Lin Y-Y, Liu T-L, Chen H-T (2005) Semantic manifold learning for image retrieval. In: Proceedings of ACM international conference on multimedia, pp 249–258
 117. He X, Niyogi P (2003) Locality preserving projections. *Advances in neural information processing systems*, vol 16. MIT Press, Cambridge
 118. He X, Cai D, Han J (2008) Learning a maximum margin subspace for image retrieval. *IEEE Trans Knowl Data Eng* 20(2):189–201
 119. Yu J, Tian Q (2006) Learning image manifolds by semantic subspace projection. In: Proceedings of ACM international conference on multimedia, pp 297–306
 120. Bian W, Tao D (2010) Biased discriminant Euclidean embedding for content-based image retrieval. *IEEE Trans Image Process* 19(2):545–554
 121. Hoiem D, Sukthankar R, Schneiderman H, Huston L (2004) Object-based image retrieval using the statistical structure of images. In: Proceedings of IEEE conference on computer vision and pattern recognition, vol 2, pp 490–497
 122. Thomee B, Huiskes MJ, Bakker EM, Lew MS (2008) Using an artificial imagination for texture retrieval. In: Proceedings of IEEE international conference on pattern recognition, pp 1–4
 123. Jing F, Li M, Zhang L, Zhang H-J, Zhang B (2003) Learning in region-based image retrieval. In: Proceedings of ACM conference on image and video retrieval, pp 199–204
 124. Karthik S, Jawahar CV (2006) Efficient region based indexing and retrieval for images with elastic bucket tries. In: Proceedings of IEEE international conference on pattern recognition, vol 4, pp 169–172
 125. Wu K, Yap K-H, Chau L-P (2006) Region-based image retrieval using radial basis function network. In: Proceedings of IEEE international conference on multimedia and expo, pp 1777–1780
 126. Muneesawang P, Guan L (2004) An interactive approach for CBIR using a network of radial basis functions. *IEEE Trans Multimedia* 6(5):703–716
 127. Chan C-H, King I (2004) Using biased support vector machine to improve retrieval result in image retrieval with self-organizing map. In: Proceedings of international conference on neural information processing, pp 714–719
 128. Koskela M, Laaksonen J, Oja E (2002) Implementing relevance feedback as convolutions of local neighborhoods on self-organizing maps. In: Proceedings of international conference on artificial, neural networks, pp 137–142
 129. Hoi C-H, Chan C, Huang K, Lyu MR, King I (2004) Biased support vector machine for relevance feedback in image retrieval. In: Proceedings of IEEE international joint conference on neural networks, vol 4, pp 3189–3194
 130. Tao D, Tang X, Li X, Wu X (2006) Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval. *IEEE Trans Pattern Anal Mach Intell* 28(7):1088–1099
 131. Zhang J, Ye L (2009) Content based image retrieval using unclean positive examples. *IEEE Trans Image Process* 18(10):2370–2375
 132. Wang L, Li X, Xue P, Chan KL (2005) A novel framework for SVM-based image retrieval on large databases. In: Proceedings of ACM international conference on multimedia, pp 487–490
 133. Hoi SCH, Lyu MR, Jin R (2006) A unified log-based relevance feedback scheme for image retrieval. *IEEE Trans Knowl Data Eng* 18(4):509–524
 134. Rao Y, Mundur P, Yesha Y (2006) Fuzzy SVM ensembles for relevance feedback in image retrieval. In: Proceedings of international conference on image and video retrieval, pp 350–359
 135. Zhou XS, Garg A, Huang TS (2004) A discussion of nonlinear variants of biased discriminants for interactive image retrieval. In: Proceedings of international conference on image and video retrieval, pp 1948–1959
 136. Wang L, Gao Y, Chan KL, Xue P, Yau W-Y (2005) Retrieval with knowledge-driven kernel design: an approach to improving SVM-based CBIR with relevance feedback. In: Proceedings of IEEE international conference on computer vision, vol 2, pp 1355–1362
 137. Xie H, Andreu V, Ortega A (2006) Quantization-based probabilistic feature modeling for kernel design in content-based image retrieval. In: Proceedings of ACM international workshop on multimedia, information retrieval, pp 23–32
 138. Hoi C-H, Lyu MR (2004) Group-based relevance feedback with support vector machine ensembles. In: Proceedings of IEEE international conference on pattern recognition, vol 3, pp 874–877
 139. Tieu K, Viola P (2004) Boosting image retrieval. *Int J Comput Vis* 56(1–2):17–36
 140. Yin P-Y, Bhanu B, Chang K-C, Dong A (2005) Integrating relevance feedback techniques for image retrieval using

- reinforcement learning. *IEEE Trans Pattern Anal Mach Intell* 27(10):1536–1551
141. Amin T, Zeytinoglu M, Guan L (2007) Application of Laplacian mixture model to image and video retrieval. *IEEE Trans Multimedia* 9(7):1416–1429
 142. Qian F, Li M, Zhang L, Zhang H-J, Zhang B (2002) Gaussian mixture model for relevance feedback in image retrieval. In: *Proceedings of IEEE international conference on multimedia and expo*, vol 1, pp 229–232
 143. Zhang R, Zhang Z (2004) Stretching Bayesian learning in the relevance feedback of image retrieval. In: *Proceedings of European conference on computer vision*, vol 3, pp 996–1001
 144. Wu H, Lu H, Ma S (2002) The role of sample distribution in relevance feedback for content based image retrieval. In: *Proceedings of IEEE international conference on multimedia and expo*, pp 225–228
 145. Gondra I, Heisterkamp DR (2004) Learning in region-based image retrieval with generalized support vector machines. In: *Proceedings of IEEE conference on computer vision and pattern recognition, workshop*, pp 149–156
 146. Chen Y-S, Shahabi C (2003) Yoda, an adaptive soft classification model: content-based similarity queries and beyond. *ACM Multimedia Syst* 8(6):523–535
 147. ten Brinke W, Squire DMcG, Bigelow J (2004) Similarity: measurement, ordering and betweenness. In: *Proceedings of international conference on knowledge-based intelligent information and engineering systems*, pp 169–184
 148. Giacinto G, Roli F (2004) Nearest-prototype relevance feedback for content based image retrieval. In: *Proceedings of IEEE international conference on pattern recognition*, vol 2, pp 989–992
 149. Royal M, Chang R, Qi X (2007) Learning from relevance feedback sessions using a k-nearest-neighbor-based semantic repository. In: *Proceedings of IEEE international conference on multimedia and expo*, pp 1994–1997
 150. Zhang J, Ye L (2007) An unified framework based on p-norm for feature aggregation in content-based image retrieval. In: *Proceedings of IEEE international symposium on multimedia*, pp 195–201
 151. Wu H, Lu H, Ma S (2003) A practical SVM-based algorithm for ordinal regression in image retrieval. In: *Proceedings of ACM international conference on multimedia*, pp 612–621
 152. Müller H, Pun T (2004) Learning from user behavior in image retrieval: application of market basket analysis. *Int J Comput Vis* 56(1):65–77
 153. He X, Ma W-Y, King O, Li M, Zhang H-J (2002) Learning and inferring a semantic space from user's relevance feedback for image retrieval. In: *Proceedings of ACM international conference on multimedia*, pp 343–346
 154. Shah-hosseini A, Knapp GM (2006) Semantic image retrieval based on probabilistic latent semantic analysis. In: *Proceedings of ACM international conference on multimedia*, pp 703–706
 155. Chen Y, Rege M, Dong M, Fotouhi F (2007) Deriving semantics for image clustering from accumulated user feedbacks. In: *Proceedings of ACM international conference on multimedia*, pp 313–316
 156. Cheng H, Hua KA, Vu K (2008) Leveraging user query log: toward improving image data clustering. In: *Proceedings of ACM conference on image and video retrieval*, pp 27–36
 157. Yin P-Y, Bhanu B, Chang K-C, Dong A (2008) Long-term cross-session relevance feedback using virtual features. *IEEE Trans Knowl Data Eng* 20(3):352–368
 158. Barrett S, Chang R, Qi X (2009) A fuzzy combined learning approach to content-based image retrieval. In: *Proceedings of IEEE international conference on multimedia and expo*, pp 838–841
 159. Oh S, Chung MG, Sull S (2004) Relevance feedback reinforced with semantics accumulation. In: *Proceedings of conference on image and video retrieval*, pp 448–454
 160. Rege M, Dong M, Fotouhi F (2007) Building a user-centered semantic hierarchy in image databases. *ACM Multimedia Syst* 12(4):325–338
 161. Huijsmans DP, Sebe N (2005) How to complete performance graphs in content-based image retrieval: add generality and normalize scope. *IEEE Trans Pattern Anal Machine Intell* 27(2):245–251
 162. Tronci R, Falqui L, Piras L, Giacinto G (2011) A study on the evaluation of relevance feedback in multi-tagged image datasets. In: *Proceedings of IEEE symposium on multimedia*, pp 452–457
 163. Li C-J, Hsu C-T (2008) Image retrieval with relevance feedback based on graph-theoretic region correspondence estimation. *IEEE Trans Multimedia* 10(3):447–456
 164. Marchand-Maillet S, Worring M (2006) Benchmarking image and video retrieval: an overview. In: *Proceedings of ACM international workshop on multimedia, information retrieval*, pp 297–300
 165. Huiskes MJ, Lew MS (2008) Performance evaluation of relevance feedback methods. In: *Proceedings of ACM international conference on image and video retrieval*, pp 239–248
 166. Jin X, French JC, Michel J (2006) Toward consistent evaluation of relevance feedback approaches in multimedia retrieval. In: *Proceedings of international workshop on adaptive multimedia retrieval: user, context, and feedback*, pp 191–206
 167. Müller H, Marchand-Maillet S, Pun T (2002) The truth about Corel—evaluation in image retrieval. In: *Proceedings of international conference on image and video retrieval*, pp 38–49
 168. Huiskes MJ, Lew MS (2008) The MIR Flickr retrieval evaluation. In: *Proceedings of ACM international conference on multimedia, information retrieval*, pp 39–43
 169. Huiskes MJ, Thomee B, Lew MS (2010) New trends and ideas in visual concept detection. In: *Proceedings of ACM international conference on multimedia, information retrieval*, pp 527–536
 170. Brodatz P (1966) *Textures: a photographic album for artists and designers*. Dover, NY
 171. Lazebnik S, Schmid C, Ponce J (2005) A sparse texture representation using local affine regions. *IEEE Trans Pattern Anal Mach Intell* 27(8):1265–1278
 172. Pickard R, Graszyc C, Mann S, Wachman J, Pickard L, Campbell L (1995) *VisTex databases*. Technical report, MIT Media Laboratory
 173. Griffin G, Holub A, Perona P (2007) *Caltech-256 object category dataset*. Technical report, California Institute of Technology
 174. Everingham H, Winn J (2007) *The Pascal VOC challenge 2007 development kit*. University of Leeds, Technical report
 175. Kalpathy-Cramer J, Müller H, Bedrick S, Eggel I, Garcia Seco de Herrera A, Tsirikia T (2011) Overview of the CLEF 2011 medical image classification and retrieval tasks. In: *Proceedings of Cross-Language Evaluation Forum*
 176. Nene SA, Nayar SK, Murase H (1996) *Columbia Object Image Library (COIL-100)*, Technical Report CUCS-006-96. Columbia University
 177. Chang H, Yeung D-Y (2007) Locally smooth metric learning with application to image retrieval. In: *Proceedings of IEEE international conference on computer vision*, pp 1–7
 178. Baras D, Meir R (2007) Reinforcement learning, spike time dependent plasticity, and the BCM rule. *Neural Comput* 19(8):2245–2279