IJCSE International Journal of Computer Sciences & Engineering Open Access

Review Paper Volume-4, Special Issue-3, May 2016 E-ISSN: 2347-2693

Active Learning Methods for Interactive Image Retrieval

Balaram Joshi S M[1], Vinayak V Naik[2] and Siddalingappa Kadakol[3]

[1,2,3]*KLE's BCA ,PC Jabin college,Hubli*

Available online at: www.ijcseonline.org

Abstract— Human interactive systems have attracted a lot of research interest in recent years, especially for content- based image retrieval systems. Contrary to the early systems, which focused on fully automatic strategies, recent approaches have introduced human-computer interaction. In this paper, we focus on the retrieval of concepts within a large image collection. We assume that a user is looking for a set of images, the query concept, within a database. The aim is to build a fast and efficient strategy to retrieve the query concept. In content-based image retrieval (CBIR), the search may be initiated using a query as an example. The top rank similar images are then presented to the user. Then, the interactive process allows the user to refine his request as much as necessary in a relevance feedback loop. Many kinds of interaction between the user and the system have been proposed, but most of the time, user information consists of binary labels indicating whether or not the image belongs to the desired concept.

Keywords— Multimedia information retrieval,Content based image retreival,Image search,Interactive search,Relavance feedback.

1. INTRODUCTION

Terabytes of imagery are being accumulated daily from a wide variety of sources such as the Internet, medical centres (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 you go about finding the ones you 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.

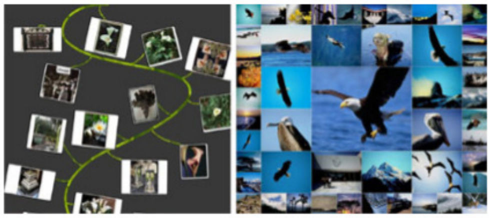
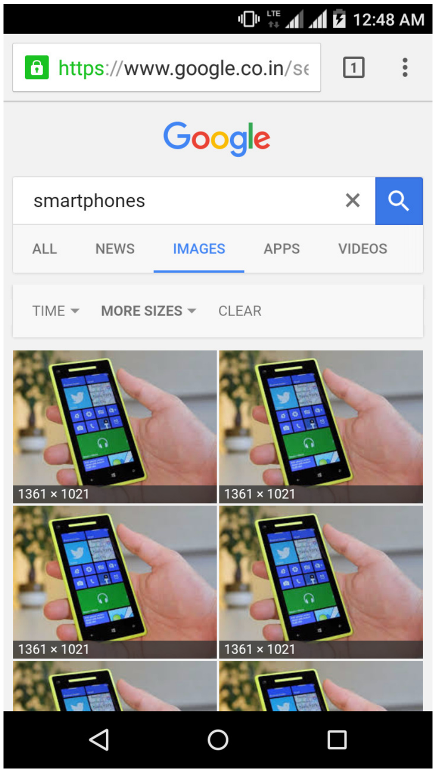


Fig.1. Examples of user interfaces. The 'tendril' interface (left) is specifically designed to support the user in exploring the visual space, where changes to the query result in branching off the initial path. The 'FreeEye' interface (right) assists the user in browsing the database, where the selected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

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 and are used by millions of people each day.

The recommendation systems have been implemented by companies such as Amazon, NetFlix 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.



https://www.google.co.in/se

smartphones

Fig.2. An example from Google Product Search shows items that are visually similar by shape and color.

Text search relies on an notations 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 analyse 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.

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.

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 helpdesk. 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.

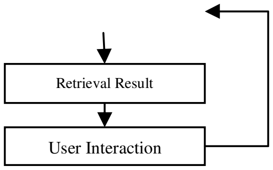


Fig. 3 The interactive search process from the user's point of view

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, provide an example image or in some situations use the favourites based on the history of the user . 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. When image segmentation is involved there area variety of ways to query the retrieval system, such as selecting one or more pre-segmented regions of interest or drawing outlines of objects of interest. 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.

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. These are usually the images that the system is most uncertain about, for instance those that are on or near a hyper plane when using SVM-based retrieval. This principle, called active learning.

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 labelling 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, positive and negative feedback, positive, neutral and negative feedback, and multiple relevance levels: four relevance levels, five levels or even seven levels. An alternative approach is to let the user indicate by what percentage a sample image meets what he has in mind. 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 labelled examples . In addition, sketch interfaces allow the user to give a fundamentally different kind of input to the system, which can potentially give a finer degree of control over the results. In the Q&A paradigm, 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 videos how 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, or supporting drawing a rectangle inside a positive example to select a region of interest . An interesting discussion on the role and impact of negative images and how to interpret their meaning can be found in. Besides giving explicit feedback, it is also possible to consider the user's actions as a form of implicit feedback, 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. In contrast with the traditional query-based retrieval model, the ostensive relevance feedback model 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. 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, by easily shifting the focus of attention from image to image allowing the user to visually explore the local relevant neighbourhood surrounding an image 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. Better presentation of these arch 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, dynamically reorganizing the displayed pages into visual islands 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. Multiple query modalities, result modalities and ways of giving feedback, for instance by allowing the user to query by grouping and/or moving images, ‘scribbling’ on images to make it clear to the retrieval system which parts of an image should be considered foreground and which parts background, or providing the user with the best mixture of media for expressing a query or understanding the results.

3 INTERACTIVE SEARCH FROM THE SYSTEM’S POINT OF VIEW

A global overview of a retrieval system is shown inFig.4.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. This section covers the recent advances on each of these parts of a retrieval system.

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, regions and salient details. After these image elements have been determined, they are often seen as standalone entities during the search. However, some approaches represent them in a hierarchical or graph-based structure and exploit this structure when searching for improved retrieval results. The multiple instance learning and bagging approach 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. 5). 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 , 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 , estimated beforehand , or alternatively automatically determined while the system is running using adaptive approaches. A thesaurus, such as Word Net, is often used to link annotations to image concepts, for instance by linking them through synonymy, hyponymy, hyponymy, etc. (See fig 6). Since manually annotating large collections of images is a tedious task, much research is directed at automatic annotation, mostly offline, but also driven by relevance feedback. Finding the best balance between using keywords for searching and using visual features for searching is one of the newer topics in image retrieval. For instance, in 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.

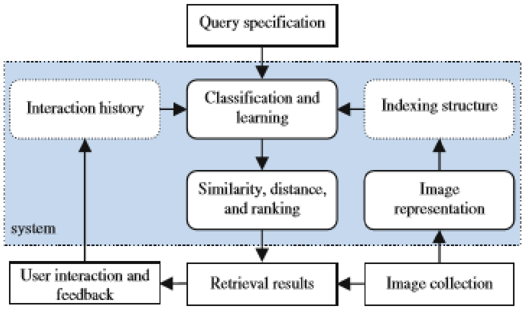


Fig.4. The interactive search process from the system’s point of view

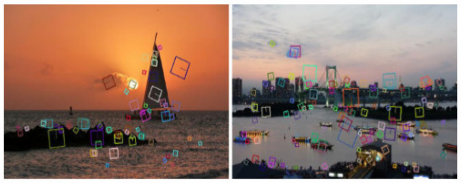


Fig.5. Images overlaid with detected visual words. Identically colored squares indicate identical visual words, while differently colored squares indicate different visual words (color figure online).

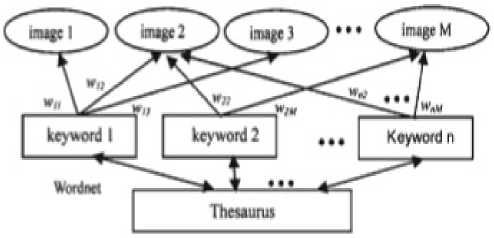


Fig.6. A thesaurus is used to link keywords to images

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 comparison. 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. 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. 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. Alternatively, the set of images that are likely relevant to the query can be quickly established by approximating their feature vectors. 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. Hashing is a form of space partitioning and is considered to be an efficient approach for indexing.

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 analysing the query image and any implicit or explicit feedback. Typical interactive system shave 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”.
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 one can measure the information associated with an unlabelled example, so a learner can select the optimal set of unlabelled examples to show to the user that maximizes its information gain and thus minimizes the expected future classification error? 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.

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, in which the authors argue for incorporating the notion of between ness 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 neighbours or combining multiple similarity measures to give a single relevance score.

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 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 analysing the feedback log, 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 or clustering 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.

4. DISCUSSION AND CONCLUSIONS

Over the years, the performance of interactive search systems steadily improved. Nonetheless, much research remains to be done. This section provides the most promising research directions.

4.1. PROMISING RESEARCH DIRECTIONS

Some top research directions that are based on this article are outlined below.

- **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 mono modal) 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. 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 a n 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. 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 labelled 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 smartphones 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 sometime reaching a large number of users. Now that the Web 2.0, the social internet, is also becoming more and more prevalent, techniques that analyse 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-labelled 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, 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 bench marks 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, there search 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 criteria on 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 behavioural 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 behaviour may not mirror human user behaviour. 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 behaviour of simulated users after their real counterparts. It is noteworthy that the user behaviour information largely exists in the logs of the WWW search engines. Thus, on the one hand, a research community 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.

6. 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

[2] Aggarwal G, Ashwin TV, Ghosal S (2002) An image retrieval system with automatic query modification. IEEE Trans. on Multimedia 4(2):201–214

[3] Bian W, Tao D (2010) Biased discriminant Euclidean embedding for content-based image retrieval. IEEE Trans Image Process 19(2):545–554

[4] 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

[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] 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

[7] Ren K, Sarvas R, Calic J (2010) Interactive search and browsing interface for large-scale visual repositories. Multimedia Tools Appl 49:513–528

[8] 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