Content Meets Semantics: Smarter Exploration of Image Collections
Presentation of Relevant Use Cases

Ilaria Bartolini
DEIS, Università di Bologna, Bologna, Italy

Keywords: Image Databases, Visual Content, Semantics, Browsing.

Abstract: Current techniques for the management of image collections exploit either user-provided annotations or automatically-extracted visual features. Although effective, the approach based on annotations cannot be efficient since the manual process of data tagging prevents its scalability. On the other hand, the organization and search grounded on visual features, such as color and texture, is known to be a powerful (since it can be made fully automatic), yet imprecise, retrieval paradigm, because of the semantic gap problem. This position paper advocates the combination of visual content and semantics as a critical binomial for effectively and efficiently managing and browsing image databases satisfying users' expectations in quickly locating images of interest.

1 INTRODUCTION

The advent of digital photography leads to an increasing production of media data, such as images and videos. Recent years have also witnessed the proliferation of social media and the success of many websites, such as Flickr, Facebook, Youtube, etc., which drastically increase the volume of media resources, available for sharing. Such websites allow users not only to create and share media data, but also to annotate them. Users of above systems are, however, also interested in a number of other challenging applications; among them, image searching and browsing, that avoid users to feel lost when facing large image repositories.

Current solutions for exploring image collections are based on a variety of heterogeneous techniques, like tagging and visual features. Keyword-based retrieval exploits labels (or free tags), used to annotate images, to search for images of interest. Free tags, however, suffer the problem of ambiguity due to the existence of synonymy (a single concept represented by several, different labels) and homonymy/polysemy (a single label representing several, different concepts). For this, it is well known that concept hierarchies represent a simple, yet powerful, way of organizing concepts into meaningful groups (Hearst, 2006). This approach has been used in a variety of contexts, also outside of multimedia. For example, Yahoo uses a topic-based hierarchy to organize web sites according to their topic and allows users to quickly identify web pages of interest, while Wikipedia contents are based on a hierarchy of categories. The biggest drawback of this approach is the fact that, while categorization of items can be performed (semi-)automatically, the hierarchies should be manually built, although studies have also focused on the automatically derivation of hierarchies (Dakka et al., 2005). When the number of categories is large, organizing them into a single taxonomy is detrimental for the usability of the overall structure. To this end, faceted hierarchies (Hearst, 2006) are used in a variety of scenarios as a very flexible, and simple, way to represent complex domains. For example, they are successfully exploited in systems like Catalyst www.gettyimages.nl/Catalyst and Flamenco (Yee et al., 2003).

Although effective, current tags-based approaches cannot be efficient due to the labor-intensive manual process of data annotation that prevents its scalability. We note that, exploiting social community contribution in terms of trustable annotations of images (e.g., Flickr), still represents an unreliable solution due to the very high level of noise in the provided meta-information.

Moreover, it is a fact that above presented techniques are not able alone to reach satisfactory performance levels. Just to give some concrete examples, text-based techniques, as exemplified by the image search extensions of Google, Microsoft Bing, and

Bartolini I..
Content Meets Semantics: Smarter Exploration of Image Collections - Presentation of Relevant Use Cases.
DOI: 10.5220/0004126901860191
In *Proceedings of the International Conference on Signal Processing and Multimedia Applications and Wireless Information Networks and Systems* (SIGMAP-2012), pages 186-191
ISBN: 978-989-8565-25-9

Yahoo!, and by systems like Google Picasa, Apple iPhoto, and Yahoo's Flickr, yield a highly variable retrieval accuracy. This is due to the imprecision and the incompleteness of the manual annotation process, in the case of Picasa, iPhoto, and Flickr, or to the poor correlation that often exists between surrounding text of Web pages and the visual image content, for the case of Google, Bing, and Yahoo!

On the other hand, content-based search, which relies on low-level similarity features, such as color and texture, is known to be a powerful (i.e., full automatic and scalable) retrieval paradigm (Schaefer, 2010; Heesch, 2008; Bartolini et al., 2004; Santini and Jain, 2000), yet it is imprecise because of the semantic gap existing between the user subjective notion of similarity and the one implemented by the system (Smeulders et al., 2000).

Among the recent attempts to match content and semantics, systems exist that profitably use similarity search principles to annotate general purpose images (Pan et al., 2004; Bartolini and Ciaccia, 2008). The basic idea is to exploit automatically extracted low-level features from images that have been manually annotated and use such information as training labels in order to suggest tags for un-labeled visually similar images. As another example, Picasa and iPhoto are tag-based tools that also provide a specific visual facility for automatic face recognition allowing users to propagate manual annotations to the image folder representing the same person; however this is clearly limited to a single content type.

Recently, some efforts have tried to exploit information contained in both visual content and semantics within the same framework, e.g., (Bartolini, 2009). In these approaches, however, the two kinds of information are never fully integrated, rather they are merely combined as a retrieval filter in order to support so-called *mixed* queries (e.g., "images tagged as *bear* that are also visually similar to the one provided as query"). Similarly (but with retrieval finality), Google Images supports keyword-based search, but also provides the user the opportunity to get visual similar images with respect to a particular returned image. A tentative joint use of low-level features and surrounding texts has been pursued by (Gao et al., 2005): here the purpose is create clusters containing images with homogeneous content *and* semantics, so as to display images resulting from a Web search in an opportunely organized way. The focus of the paper is on exploration effectiveness only, while efficiency aspects are not considered due to the bounded dimension of the result set. This clearly prevents the use of such technique on large image collections.

In this position paper we show how visual content and semantic-based paradigms can be seamlessly integrated for effectively and efficiently managing and browsing image databases, satisfying users' expectations in quickly locating images of interest. The basic idea is to use the semantics to overcome visual content limitations and vice versa. We will elaborate our reasoning through a number of relevant use cases on a real browsing system. In particular, we will show how the PIBE system (Bartolini et al., 2004; Bartolini et al., 2006) can be extended to encompass concepts drawn from both semantics and content realms.

1.1 Basics on PIBE

PIBE (Bartolini et al., 2004; Bartolini et al., 2006) is a browsing engine whose principles are rooted on the concept of "adaptable" content-based similarity. PIBE is based on an automatically derived hierarchical tree structure (called Browsing Tree) and provides the user with a set of browsing and personalization actions that enable an effective and efficient exploration of the image collection avoiding global re-organizations.

PIBE provides the user with two main functionalities, browsing and personalization, that are available through an intuitive and user-friendly GUI (see Figure 1).

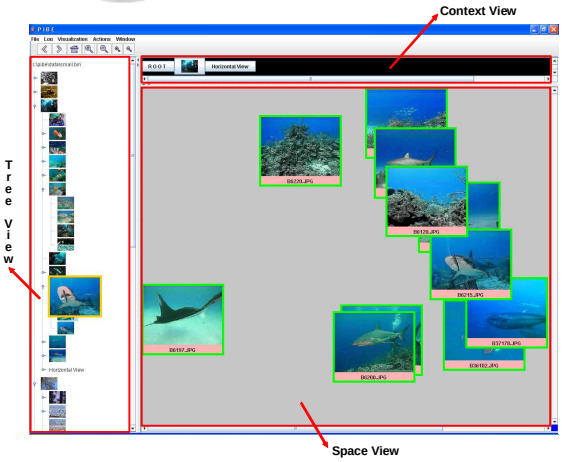

Figure 1: PIBE's interface.

In details, the Browsing Tree (BT) is automatically derived from visual image descriptors (i.e., the color distribution) whose only requirement is to be points (feature vectors) in a N-dimensional space. By recursively applying a clustering (k-means) algorithm on such visual image descriptors, images sharing low level characteristics are grouped together deriving the nodes of the BT. The image comparison criterion is based on a (dis)similarity function between the cor-

responding feature vectors. A key point to highlight with regards to the (dis)similarity function is that each node of the BT uses an adaptable *local* similarity criterion (the weighted Euclidean distance) to compare images. In this way, user preferences are contextualized to the relevant portion of the dataset and the hierarchical organization of the BT avoids costly global reorganizations. In the first step of the clustering process, the *k*-means algorithm is applied to the whole dataset by using default weights; then, it is recursively applied down to the desired granularity level to each of the derived *k* clusters using an updated (dis)similarity function whose weight components correspond to the inverse of the variance of the feature vectors within the cluster along each component. Finally, for each node a *representative* image (the image that is closest to the cluster *centroid*) is selected for visualization purposes.

Unfortunately, due to its fully automatic creation, the initial structure of the BT may not always perfectly satisfy current user preferences. For example, some images may be included in the wrong sub-tree or a node may have too many/few children. To overcome such problems, PIBE provides a set of personalization actions (e.g., move of an image to a different cluster, fusion/split of clusters, etc.) to help the user adapting the automatically built BT to her current preferences.

With respect to browsing functionalities, the user can explore the BT by means of a spatial visualization approach (see the *Space View* in Figure 1), where feature vectors are mapped on the 2-D screen to highlight image similarity (the more images are similar, the closer they are displayed to each other). Among different browsing modalities, PIBE provides within the Space View a traditional top-down navigation, where the user selects an image on the display (by clicking on it) and zooms in the corresponding BT node (i.e., images representing the children nodes are shown). The visualization concerns the *local* content of one node; to distinguish between internal and leaf BT nodes, representative images are framed in yellow for internal nodes and in green for leaf nodes, respectively (see Figure 1).

On the other hand, to offer a *global* visualization modality of the BT, PIBE's GUI provides the user with a sequential visualization tool, named *Tree View* (see the left side of Figure 1), which offers usual navigational facilities (expanding/collapsing nodes, visiting nodes, etc.). When the user clicks on (the representative image of) a BT node in the Tree View, the content of the Space View is replaced with images included in the selected node. Due to space constraints, each image in the Tree View is zoomed in only when the pointer hovers over it (see the "shark" image in Figure 1).

Finally, to help the user in remembering the history of its browsing session, the GUI provides a third visualization tool, named *Context View* (see the top of Figure 1), which takes memory of each vertical/horizontal exploration action by highlighting the sequence of clicked images, for the vertical exploration, and using a grey box to represent a horizontal browsing action. In particular, by clicking on each element of the Context View the user can return back to a previous visualization in the Space View.

2 HOW SEMANTICS CAN HELP CONTENT AND VICE VERSA

Despite the fact that PIBE includes personalization facilities so as to partially fill the semantic gap problem, browsing quality is still far from users' expectations, due to the fact that the hierarchical organization of the BT may prevent semantically related images to appear within the same BT node. PIBE is therefore a good candidate for the integration of semantics into its content-based organization. For this, in this section we show an evolution of the PIBE system, where *semantics* are taken into account, using image labels, to complement the visual similarity computed on low-level features. By means of a set of use cases on this advanced version of PIBE, we demonstrate the key role played by the semantics in the image browsing process. In details, it is possible to see how the basic functionalities of PIBE can be opportunely extended (by exploiting tags associated to images) in order to improve users' satisfaction. Towards this goal, we distinguish two relevant scenarios:

1. use cases where the semantics assist the user during her exploration;
2. use cases where the semantics help in improving the image organization (i.e., the BT).

In the first scenario, existing image tags are used to refine or augment the scope of current exploration. The first use case of this scenario deals with "*hyperlinks to clusters of similar images*". One of the problems when browsing the BT in the original version of PIBE is the fact that semantically correlated images that do not share a similar visual content are likely scattered in different parts of the tree. For example, suppose the user is exploring the cluster represented by the *black bear* in Fig. 2, i.e., the one highlighted in the tree view. During the browsing session, the system, by exploiting existing image tags, may suggest

the user two relevant clusters of semantically correlated images, i.e., *brown bear* and *polar bear* clusters, respectively. She can easily explore such clusters by following provided links within the space view instead of repeating multiple, separate, BT explorations.

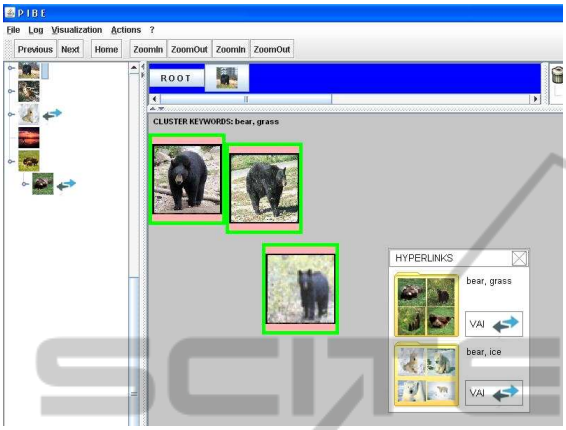

Figure 2: Hyperlinks to similar clusters.

The second use case consists in "*suggesting good browsing directions*". When looking for a particular image, the user usually starts her exploration from the root of the BT and navigates down the tree until she reaches a leaf node. During this browsing session, she has to select, at each step, which is the "best" subtree to visit next, according to visual characteristics. When siblings nodes are visually similar, such selection becomes a critical task, because a choice made in the upper levels of the tree may lead the user to the "wrong" part of the tree. The integration of semantics with low-level features may alleviate the problem by opportunely highlighting portions of the BT that are not relevant to the specific browsing task, i.e., subtrees containing images whose annotations are not correlated to the tags specified by the user.

In the example depicted in Fig. 3, the user is looking for an image representing a "bird in the sky" (input tags *bird* and *sky*, respectively). The system interacts with the user by suggesting three different types of browsing directions, each one distinguished by a specific color (i.e., green for "visit", red for "avoid to visit", and yellow for "visit, but only after exploring green clusters"). This suggestion could be extremely helpful for the user that, in order to quickly reach images of interest, should probably visit the green node first (the sub-tree represented by the *sub* picture); in such node, the presence of at least one image semantically relevant for the search is guaranteed. Then, the user could possibly follow the yellow branch (represented by the *brown bear* image), where she would likely find untagged images that may be interesting

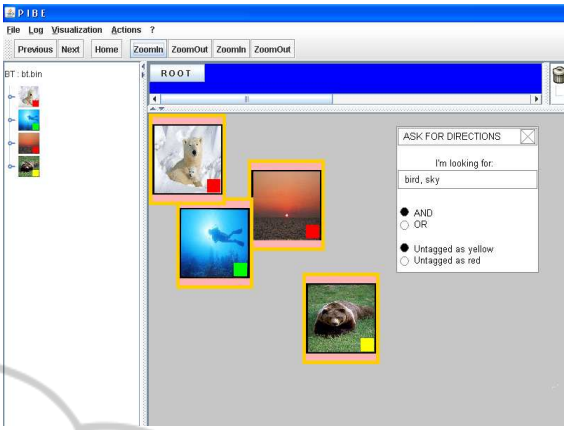

Figure 3: Suggesting good browsing directions.

for her. Finally, she should avoid exploring red clusters, that contain no images with tags semantically correlated with the input ones.

Our third use case deals with a facility which is typically included in several multimedia applications, i.e., "*mixed queries*". In this case, the user asks for images related to a specific concept (e.g., tag *sub* in the example of Fig. 4) and that are also visually similar to a pre-selected image (the *fish* image in our example). The system returns the set of images belonging the whole BT that match the input tag, providing them to the user in descending order of visual similarity with respect to the selected image.

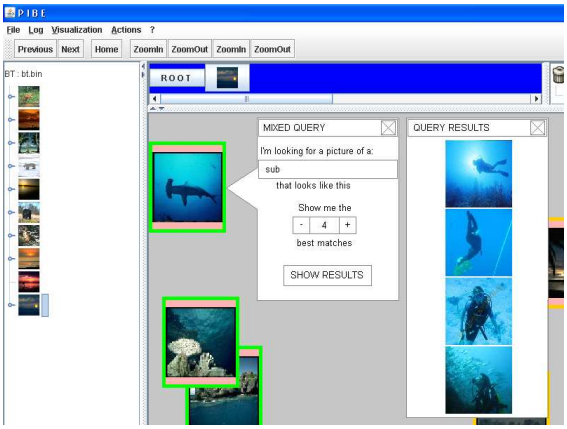

Figure 4: Supporting mixed queries.

The final use case for the first scenario (how semantics could assist the user during her exploration) covers the "*tag report*" functionality, providing the user with a general idea of the content of the image collection. Given a user provided concept (e.g., the tag *bear* in the example of Fig. 5), the system computes a statistical report containing the set of tags that are correlated with the input one (i.e., tags that co-

Figure 5: Supporting tag report.

occur with *bear* in the images of the collection). In the example, such tags are *ice*, *grass*, *sky*, and *tree*. For each correlated tag, the number of relevant images for that tag is also provided in the report together with a link to the image set containing both tags, that the user is allowed to explore. In our example, the collection contains 11 images tagged with *bear* and *ice*, 10 pictures labeled with *bear* and *grass*, etc.

We now present relevant use cases belonging to the second scenario, where semantics could help in improving the BT. The basic idea is to highlight possible discrepancies between tags associated to images belonging a same cluster, and to suggest viable actions to solve the problem (by consistently updating the browsing structure) in order to improve exploration effectiveness for future browsing sessions. Regarding efficiency, we note that in all cases, reorganizations of the BT are always performed using local visual features, thus scalability of this process is inherited from the original version of PIBE.

The first use case deals with "*alerting for misplaced images and suggesting destination clusters*". Due to the semantic gap problem suffered by the original version of PIBE, it may happen that an image included in a cluster is semantically different with respect to all the other images in the same cluster. In such situation, by looking at image semantics (i.e., tags) the system may alert the user of these misplaced images and consistently suggest possible relevant destination clusters where those images could be moved.

In the example of Fig. 6, the alert concerns the image representing a *fish* (such image is labeled as *fish* and *sea*, while all the other 4 images are tagged with tags *sub* and *sea*). The system concludes that, even if visually similar to the others, the *fish* image is misplaced in the current cluster and automatically provides the user with a more suitable destination for it. In order to derive a suitable destination cluster, the system looks for cluster of images in the BT sharing the same tags (o similar ones, exploiting a lexical ontology like WordNet (Miller, 1995)) as the misplaced image. The user is then offered the choice to persistently move that image to one of suggested clusters, depending on her personal preferences. In alternative, a re-tagging action is also contemplated by the

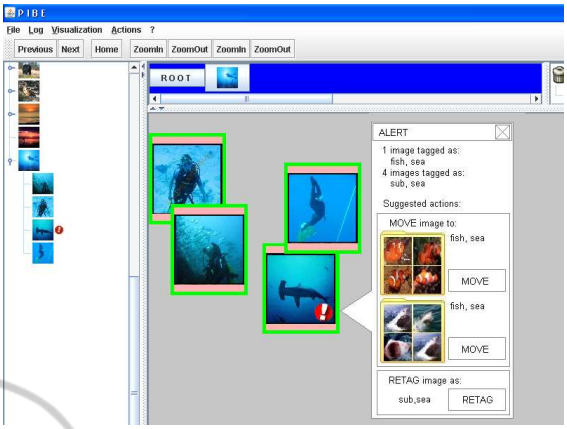

Figure 6: Suggesting clusters for misplaced images.

system; this is important because the misplacement alert may depend on a wrong annotation of the image. In this way, the user has the possibility to re-tag (at run time) the supposedly misplaced picture with the same tags associated to the majority of other components (the *first-class citizens*) of the same cluster (in the example, *sub* and *sea*).

Our second example deals with "*suggesting cluster split*". This is a particular case of the previous task. The envisioned scenario is as follows: some clusters may exist containing images that, according to their tags, should be semantically divided into two or more (sub-)groups.

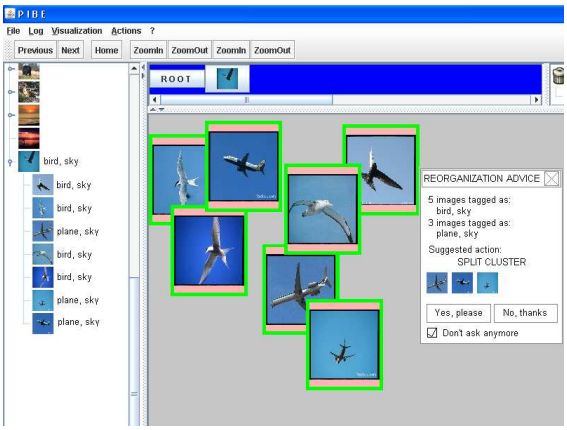

Figure 7: Suggesting cluster splitting.

In the example depicted in Fig. 7, the cluster is composed of 5 images of *bird*s and 3 images of *plane*s (again, this might be the result of a visual similarity existing between images). The system notices a questionable situation and suggests the user a viable solution to solve the problem, that is, to divide (split) the current cluster into two separate sub-clusters. Again, the user may decide to persistently apply such modi-

fications to the BT or to ignore the suggestion.

The last use case we envision, "*suggesting cluster merge*", is complementary to the previous one. This time, the user is offered the possibility of merging clusters containing images that are semantically similar but visually different (and are therefore scattered in different points of the original BT).

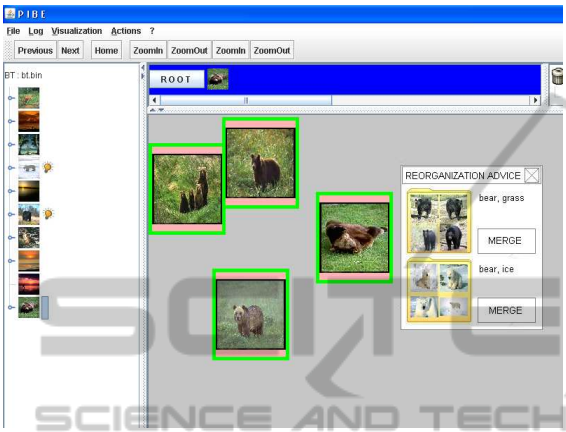

Figure 8: Suggesting cluster merging.

For example, consider again Fig. 2, where the system suggested the user alternative browsing directions: the system could also offer the user the option to persistently merge suggested clusters with the current one. Clearly, such re-organization would move images in suggested clusters into a single cluster, that may be subsequently divided into several subclusters. In the specific case of Fig. 8, the system allows the user to merge the currently opened cluster (*brown bear*) with clusters containing *polar bear* and/or *black bear* images, respectively.

3 CONCLUSIONS

In this position paper we advocate the combined use of visual content and semantics as a critical binomial for effective and efficient browsing of large image collections, so as to satisfy users' expectations in quickly locating images of interest. We have elaborated our reasoning through a set of relevant use cases on a real browsing system, namely PIBE. Such use cases testify how semantics can help visual content and vice versa, both in assisting the user during her exploration sessions and in improving image organization. We finally note that, although use cases presented in this paper all used simple labels (free tags), the underlying model exploited in the improved version of PIBE allows the use of *semantic tags*, i.e., paths extracted from a variety of existing taxonomies (semantic dimensions); this is known to solve problems of ambiguity, polysemy, etc. that plague solutions based on free tags (Hearst, 2006; Bartolini, 2009).

In the future, we plan to provide a thorough experimental analysis and comparison evaluation of PIBE on real users with large image benchmarks.

REFERENCES

Bartolini, I. (2009). Multi-faceted Browsing Interface for Digital Photo Collections. In *Proc. of CBMI 2009*, pages 65–72, Chania, Crete.

Bartolini, I. and Ciaccia, P. (2008). Imagination: Exploiting Link Analysis for Accurate Image Annotation. *Adaptive Multimedia Retrieval: Retrieval, User, and Semantics (LNCS)*, 4918:322–44.

Bartolini, I., Ciaccia, P., and Patella, M. (2004). The PIBE Personalizable image Browsing Engine. In *Proc. of CVDB 2004*, pages 43–50, Paris, France.

Bartolini, I., Ciaccia, P., and Patella, M. (2006). Adaptively Browsing Image Databases with PIBE. *Multimedia Tools Appl.*, 31(3):269–286.

Dakka, W., Ipeirotis, P. G., and Wood, K. R. (2005). Automatic Construction of Multifaceted Browsing Interfaces. In *Proc. of CIKM 2005*, pages 768–775, Bremen, Germany.

Gao, B., Liu, T.-Y., Qin, T., Zheng, X., Cheng, Q., and Ma, W.-Y. (2005). Web Image Clustering by Consistent Utilization of Visual Features and Surrounding Texts. In *Proc. of ACM MM 2005*, pages 112–121, New York, USA.

Hearst, M. A. (2006). Clustering Versus Faceted Categories for Information Exploration. *Commun. ACM*, 49(4):59–61.

Heesch, D. (2008). A Survey of browsing Models for Content-based Image Retrieval. *Multimedia Tools Appl.*, 40(2):261–284.

Miller, G. A. (1995). WordNet: A Lexical Database for English. *Commun. ACM*, 38(11):39–41.

Pan, J.-Y., Yang, H.-J., Faloutsos, C., and Duygulu, P. (2004). Automatic Multimedia Cross-modal Correlation Discovery. In *Proc. of SIGKDD 2004*, pages 653–658, Seattle, USA.

Santini, S. and Jain, R. (2000). Integrated Browsing and Querying for Image Databases. *IEEE MultiMedia*, 7(3):26–39.

Schaefer, G. (2010). A Next Generation Browsing Environment for Large Image Repositories. *Multimedia Tools Appl.*, 47(1):105–120.

Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., and Jain, R. (2000). Content-based Image Retrieval at the End of the Early Years. *IEEE TPAMI*, 22(12):1349–1380.

Yee, K.-P., Swearingen, K., Li, K., and Hearst, M. A. (2003). Faceted Metadata for Image Search and Browsing. In *Proc. of CHI 2003*, pages 401–408, Ft. Lauderdale, Florida, USA.