A Multi-faceted Browsing Interface for Digital Photo Collections*

Ilaria Bartolini

DEIS, University of Bologna, Italy

i.bartolini@unibo.it

Abstract

Managing photos by using visual features (e.g., color and texture) is known to be a powerful, yet imprecise, retrieval paradigm because of the semantic gap problem. The same is true if search relies only on keywords (or tags*), derived from either the image context or user-provided annotations. In this paper we present a new multi-faceted image search and browsing system, named Scenique, that allows the user to manage her photo collections by using* both *visual features and tags, possibly organized into multiple* dimensions *(or* facets*). Each facet can be seen as a coordinate of a multidimensional space describing the image content (for example, the visual appearance, the content type, the geographic location, and so on). We present the basic principles of Scenique and provide evidence of the effectiveness of its visual tools. Feedback supplied by a set of real users indicates that the proposed interface is intuitive, easy to use, and that satisfies users' expectations in managing photo collections and quickly locating images of interest.*

1 Introduction

Thanks to the wide dissemination of digital and mobile phone cameras, it is nowadays extremely easy for any ordinary user to collect photos. Since taking and storing pictures is almost priceless, the size of personal digital photo collections is growing exponentially. Additionally, the sharing of photos on the Web as an easy and global experience helps in enlarging such phenomenon. As a main consequence, the demand for tools for effectively managing personal photo collections has enormously increased. Among such tools, those providing similarity-search functionalities are essential if one wants to provide users with the possibility of looking for images sharing some visual appearances. Even if this content-based approach can be completely automatized, it is known to yield imprecise results because of the semantic gap problem [21].

The alternative to content-based retrieval is to look for images by using text-based techniques. Towards this end, several commercial applications and services have been proposed in recent years: examples include the image search extensions of Google and Yahoo, which consider the original Web context (e.g., file name, title, surrounding text) to infer the relevance of an image, as well as systems like Microsoft's Photo Gallery, Google Picasa, and Yahoo's Flickr, which rely on user-provided *tags* or labels. However, in both cases, the accuracy of the results is highly variable, since it heavily depends on the precision and the completeness of the manual annotation process (in the case of Photo Gallery, Picasa, and Flickr) or it can be completely uncorrelated with respect to the visual image content (for the image search extensions of Google and Yahoo).

Motivated by above observations, in this paper we present Scenique (Semantic and ContENt-based Image QUErying), a novel multi-faceted image search and browsing system.[1] Scenique represents an effective solution in providing users with an *integrated* query/browsing facility that allows photos to be searched and explored using *both* semantic tags and visual features, thus taking the best of the two approaches (refer to Figure 1 for a snapshot of the browsing visual interface). The model of Scenique is
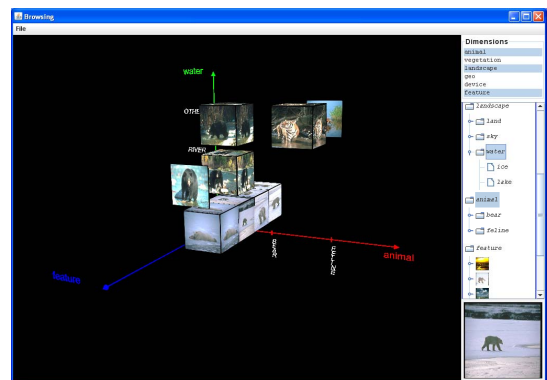

Figure 1. Scenique browsing interface.

[1] A software demo of a preliminary version of the system has been presented in [4].

based on the multi-structural framework proposed in [11]. Each photo is viewed as a set of regions, from which visual features can be automatically extracted, and a set of (semi-)automatically assigned tags. Tags and visual features can be organized in so-called (classification) dimensions, or *facets*, which take the form of *facet trees*. Each facet is thought to be as a particular coordinate describing the content of a picture. For example, in Figure 1, among the many existing dimensions, the semantic facet trees `landscape` and `animal` and the visual facet `feature` (which represent tree hierarchical partitions of the photo collection based on the semantic content and the color similarity of pictures) are used to explore the photos. By means of a multi-faceted interface, Scenique makes available the whole content of the underlying photo collection and help users in quickly locating photos of interest.

The contributions of this papers are: (i) the description of the integrated software architecture implementing the Scenique engine; (ii) a multi-faceted model for the definition and the construction of the dimensions; (iii) effective and easy-to-use visual tools able to support facet construction, photos annotations, and search/browsing facilities. The rest of the paper is organized as follows. Section 2 overviews related work. Section 3 presents the model. The architecture of the system is described in Section 4. Section 5 introduces the Scenique interface. Section 6 gives details on the system implementation and provides evidence of the effectiveness and the usability of the user interface. Section 7 concludes the paper.

2 Related Work

Concept hierarchies are commonly used to support browsing task on top of large set of objects. 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. Most of the existing systems use a single hierarchy to present the content of an image collection. However, academically-developed systems like Flamenco [23] demonstrated that *faceted hierarchies*, which allow users to explore a data collection across multiple, orthogonal classification schemas, are more effective than a single monolithic hierarchy. In the particular, context of personal photo collections (where the cardinality of the databases is usually not too high) such faceted hierarchies can be identified and manually constructed by the user (provided that the GUI contains appropriate tools). On the other hand, when the cardinality of text annotated object databases is large, techniques for automatically deriving multi-faceted hierarchies are required [9].

The annotation process can be completely manual or take the advantage from some background knowledge, including the file name, the title, and the surrounding text. Thanks to

improvements in GPS and mobile phone technology, recent solutions automatically incorporate contextual metadata to photos, such as the time events, the location (for example, through GPS coordinates) and, more in general, all information that can help in knowing when and where the photos are taken in order to discover what they represent [13, 20, 1].

To overcome possible inaccuracy and incompleteness of image annotations, several techniques have been proposed in recent years for (semi-)automatically assigning meaningful terms to photos trying to capture the correspondences between low-level features and semantic concepts [16, 10, 14, 18, 15, 3]. In this context, Scenique represents a new interaction paradigm between the user and the system, so that users can exploit multi-faceted interfaces to concurrently explore both semantic concepts and visual features. Advanced multimodal searching and browsing facilities based on user-defined facets allow the effective management of personal photo collections through high-level, concept-based interactions. Systems which are able to deal with multi-criteria queries involving both metadata and content-based representations of images have been investigated both in the context of relational [17] and medical [8] databases. With respect to such systems, the novelty of Scenique consists in the ability of organizing pictures through multiple semantic and feature-based hierarchies by automatically associating photos to nodes of such dimensions.

Two systems which are in some ways related to our contribution have been presented in [22, 19]. FacetMap [22] is an interactive driven-based visualization for personal image collection which relies on metadata facets as filter only. The Cortina system [19] retrieves images using visual features and text annotations. However, the correspondence between low-level features and semantic descriptors is manually done by the user, who first enters a keyword-based search, then provides relevance feedback on the returned images to capture the content [6].

3 The Model

Scenique is based on the *multi-structural* framework [11], that consists of a set of objects (in our case, photos), a *schema* that defines a classification of the objects according to multiple distinct criteria (i.e., the *facets*), and a membership relation specifying the elements of each facet to which an object belongs. In such a way, the user can construct several dimensions, with the aim to organize images from different points of view, and, at query time, browse images through the facets or by formulating complex queries. Formally, our multi-faceted DB is a triple $(\mathcal{P}, \mathcal{F}, \mathcal{R})$, where $\mathcal{P} = \{P_1, \ldots, P_N\}$ is the photo DB, $\mathcal{F} = \{F_1, \ldots, F_M\}$ is the set of facets, and $\mathcal{R}$ is the membership relation between objects in $\mathcal{P}$ and elements in $\mathcal{F}$. Two relevant cases

of facets that arise in practice are *numerical* and *hierarchical* dimensions; the former corresponding to points on the real line and the latter to trees. Each facet is modelled as a *lattice*, whose components are *intervals* of the real line, for the numerical case (e.g., time), and nodes of the tree representing concepts (i.e., tags), for the hierarchical scenario (e.g., geographical location). We refer to hierarchy trees as *semantic facet trees* (or *tag trees*). We then enrich the model by means of a set of operations for efficiently updating and exploring the structures (e.g., add/delete concepts, add/delete branches of concepts, get dimension).

In the following, we focus on hierarchical facets. We define the universe of tags $T = \{t_1, \ldots, t_l\}$ (also referred as tag DB) as the union of tags in all facets $F_i \in \mathcal{F}$. The membership relation $\mathcal{R} \subseteq \mathcal{P} \times T$ indicates that an object belongs to a facet element (similarly, that a photo has assigned a tag of a semantic facet tree).

Each photo P_j is viewed as a set of regions from which features, such as the color and the texture, can be automatically extracted, and a set T_j of (semi-)automatically assigned tags. The bridge between the tags of tag trees and the photo information (i.e., visual descriptors and tags) is obtained by means of a (semi-)automatic image annotator that, starting from a training set of pre-annotated photos, is able to predict sets of "good" tags for other pictures. This annotation process is modelled as a nearest neighbors (NN) problem on image regions.[2] The set S containing the k-NN regions of each region of a new photo P_j is first determined. The initial set T_j of tags for the new image consists of the tags included in images containing regions in S; each tag in T_j is also given a frequency score f. However, tags in T_j might include unrelated, or even contradictory, terms. To overcome such limit, we exploit the pairwise term correlation by associating to each couple of tags a correlation score c. In particular, we reduce the cardinality of T_j by combining the f and c scores. To this end, we build an undirected and weighted graph G whose nodes correspond to tags in T with highest values of f, while weights correspond to f values. An edge between two nodes is added if their correlation score c exceeds a fixed threshold value. Starting from the graph G, we derive the set of final tags that are both related to the new photo and that share a semantic correlation among themselves, by determining the maximum subset of fully connected nodes [7].

Any tag occurrence is actually a specific node in a *tag tree*, each tag tree representing the organization of tags for a specific dimension. As an example, the tag `bear` is a node in the tag tree of the `animal` dimension of Figure 1. Note that, in principle, the same tag can appear in different tag trees, which allows to discriminate between the different usages and/or meanings different tag occurrences can have.

For instance, the tag `Italy` might appear as a node for the `geographic` dimension (used to organize photos according to the place they have been shot) as well as a node in the `sport` dimension (which only applies to photos related to sport events).

By default, each tag is initially a node of a generic, unstructured, `default` dimension. User-defined dimensions can be built to fit specific needs. For instance, in order to organize photos according to their main subject, the dimensions `person` and `animal` can be defined. The facet `animal` is then structured by creating, for example, the three nodes `mammal`, `bird`, and `fish`. The node `mammal` can be further specialized into nodes `bear`, `horse`, etc. The same rationale can be applied to `person`.

Within the reference model, a set of operations, such as *meet* (logical AND) and *join* (logical OR) are defined. In this way, the user formulates compound queries over different dimensions by means of logical expressions (e.g., `(south america AND bear)` or `(person OR (animal AND europe))`.

With respect to *visual dimensions*, hierarchical structures are automatically derived over visual photo descriptors by means of recursive clustering [5]. First, we apply the partitioning algorithm *k-means* to the whole photos DB. Each of the so-obtained clusters, containing photos sharing similar visual features, is then recursively partitioned into k subclusters, and so on until less than k pictures are left in a cluster. Each node of the so derived *visual facet tree* is represented by means of a representative image, named *centroid*, which corresponds to the photo of the cluster that is closer to the point obtained by averaging all the feature vectors within the cluster (see Figure 1 for an example). The so obtained hierarchical views of images can be customized according to user preferences [5]. Note that visual facet trees are similar to the structure of semantic facet trees, with the only difference that each node of a visual tree is not a tag but a cluster of photos represented by its centroid. Thus, from a functional point of view, they can be considered completely analogous.

4 System Overview

Figure 2 illustrates the Scenique architecture. It is composed of a *Feature DB*, storing the feature vectors that are automatically extracted from photos, and a *Tag DB* which maintains the current tags in each tag tree. The *Data Manger* is in charge of the construction and maintenance of facets.

At population time (Figure 2 (a)), visual facet trees are automatically derived over the visual photo descriptors by the *Feature Organizer*. The *Facet Processor* is in charge of the definition of semantic facets and the construction of their tag trees. As described in Section 5, the user interacts

---

[2] Here we summarize the main idea of the photo annotation process (for a complete description, please refer to [3]).

(a) (b)

Figure 2. The system architecture: population (a) and searching/browsing (b) time.

with the Tag DB through *drag&drop* facilities to easily derive her tag trees. The *Annotation Processor* populates the tag trees.

At query time (Figure 2 (b)), the *Search&Browse Processor* manages querying and browsing users requests by interacting with visual and tag trees. The Search Processor supports three main query modalities: *tag*-based (T), *visual*-based (V), and *tag&visual*-based (TV).

Queries of type T are formulated using the available facets. In the simplest case, the retrieval is based on the resolution of user-provided logical expressions, that relies upon the exact match between selected tags and image associated tags. More interesting queries are derived when the *parent-child* relations between nodes of the tag trees (e.g., "the bear is a mammal") are exploited by the Search Processor to improve the quality of the results. Suppose that the user is looking for mammal images, the result provided by the Search Processor might include not only images with the tag mammal, but also images annotated with the tag bear, because, in this case, the Search Processor takes advantage of the relation "the bear is a mammal". With the same aim, lexical ontologies, such as WordNet[3], can be exploited instead of user-defined facets. This allows to deal with the case when provided search keywords do not belong to any dimension. Another interesting type of queries supported by the Search Processor are those involving the so-called *semantic relaxation* on concepts. Let us consider an user interested in localizing photos of "asiatic_black_bear" and that, for such selected tree node of the animal dimension, no photo is available. To avoid an empty result, the system allows the user to specify at query time the level of relaxation on the specified concept she can tolerate (e.g., "*if there is no Asiatic black bear, I will be happy even with photo of bears*"). The level corresponds to the number of *child-parent* relations the Search Processor can exploit to compute the final result. Continuing our example, if the user admits, as worst case, photos of mammals (i.e., the two levels of relaxation "asiatic_black_bear is a bear" and "bear is a mammal" can be used), the Search Processor

will compute the semantic similarity between the node "asiatic_black_bear" and all nodes bounded by the node "mammal" in order to return photos in descendent order of similarity. As a measure of the semantic similarity we follow the well-known metric for hierarchical domain structure proposed in [12]. The similarity between two nodes of the tree is then defined as the ratio of twice the depth of their lowest common ancestor to the sum of the depths of the two involved nodes.

With queries of type V, the user is looking for images that are similar, from the point of view of visual features, to a specific query image. In details, the Search Processor supports k nearest neighbors (k-NN) queries. Given a query image, it ranks photos according to a specific similarity criterion and returns the k images with highest similarity score.

Queries of type TV combine T and V modalities. Visual descriptors of query images are used as a filter for the tag-based searches. Returned images contains photos in the intersection of both the T and V results first, followed by pictures in the T list only and, finally, by images in the V result only.

With respect to browsing facilities, the Browse Processor supports a user-friendly modality to explore tag trees, visual trees, or their combination. By means of a mouse-driven selection of facet tree nodes, the user can browse her photo collection from different points of view by obtaining a display of the result in a 3-D *viewing room*, whose coordinates correspond to selected facets. Multiple selections of tree nodes equal to formulate a logical AND expression on selected tags/representative photos (see Figure 1).

The user interacts with the Scenique *Engine* through the Graphic User Interface (GUI), which provides advanced visual tools for facets construction, photo annotation, and search and browse visualization, as detailed in Section 5.

5 Scenique Interface

In this section we present the visual tools for facets construction, image annotation, and search/browsing facilities.

---

[3] WordNet: http://wordnet.princeton.edu/.

Facets construction: Facets construction is supported by an intuitive interface that requires the user to set the name of the dimension (e.g., `geographic`), so as to obtain the corresponding tree root construction. At this point, through *drag&drop* actions, the user can add tags from a provided set (the tag DB), or add new tags. When done, the user can add the so-constructed dimension to the list of available facets and permanently save it within the reference schema.

Photo annotation: For annotating an image, the user selects a photo together with a dimension of interest (the unstructured `default` dimension or, alternatively, a user-defined one). The system predicts a set of tags which are bound to the specified dimension. Then, the user provides feedback on the suggested labels by confirming correct tags, deleting wrong ones and/or adding missing keywords. Finally, she permanently saves the annotation within the set of membership relations (i.e., assigning the tags of the selected dimension to the photo).

Search facilities: To search the photo collection, the user can follow one of the three query modalities *T*, *V*, and *TV* described in Section 4. Here, we propose the more interesting example, where the user formulates a query of type *TV*. By selecting the facets of interest (i.e., `animal` and `geographic`) and by using provided buttons, the user composes the logical expression (`black_bear` AND `asia`). The result of the expression is then filtered by means of the selected image, which reflects user preferences from a visual similarity point of view.

3-D browsing: Photo collections can be explored by the user through an intuitive browsing interface, as shown by Figure 1. The interface is mainly composed by a 3-D *viewing room* (the black window in Figure 1) and by a 2-D *facet panel* (the white panel on the right side in Figure 1).

The user starts her browsing session by selecting some facets from the available list of semantic and visual dimensions (in our example, `animal`, `landscape`, and `feature`). Then, the corresponding facet trees are displayed within the 2-D facet panel. At this point, the user starts her exploration through the trees by *clicking* with the mouse on nodes of interest. Each time a node is selected, the corresponding facet is highlighted in the 3-D room, as an axis of the space being browsed, and the corresponding photos are shown. Image cubes in the 3-D view are clusters of images that can be further explored, whereas standard images correspond to leaf nodes. We define the 3-D view to be *active*, in the sense that the user can explore the photo collection even by clicking on images. When more than one facet tree is involved in the browsing process (this is the case of our example, where tags `water` and `animal` and the representative image "bear" of a visual cluster are involved), each axis in the 3-D room corresponds to a different dimension.

The 3-D viewing room and the 2-D facet panel are *syn-chronized* at each time, representing the former the visualization of the browsing result and the latter the user request. For a better view of the photo collection content (from different angle-shots), multiple browsing visualization can take advantage of the rotation functionality of the 3-D space.

6   System Implementation

We implemented Scenique in Java JDK 5.0 and tested it on a laptop, equipped with 512MB of memory running Microsoft Windows XP, using a database of 5,000 annotated images extracted from the Corel collection. The whole data (photo DB, feature DB, tag DB, tag trees, and visual trees) is persistently maintained in secondary memory under the open source MySQL DBMS and accessed through JDBC. Visual tools are based on Java3D APIs.

Each photo is automatically segmented into a set of homogeneous regions which convey information about color and texture features. Each region corresponds to a cluster of pixels and is represented through a 37-dimensional feature vector. With respect to regions comparison the Bhattacharyya metric is used. For more details, see [2].

The reference data model is enriched by a SQL-like definition language, based on a EBNF representation grammar, which allows the user to create and manipulate her facets at runtime (e.g., `insert into facet` *subject* `concept` *animal/mammal/dog*). Persistency of user-defined facets is guaranteed over several working sessions by an *XML wrapper*, which is in charge of saving and loading the so-defined structures.

6.1   User Study

We conducted a user study over a set of ten participants (half male and half female) to evaluate the effectiveness of Scenique and the usability of its visual tools. Participants ranged from 25 to 50 years (average 35.8) and were all expert users of the Web and multimedia search engines. After a demonstration of the Scenique functionalities, the users were given a fixed time of 20 minutes to freely play with the system. Finally, they were requested to fill a questionnaire, using 0 as "strongly disagree" and 4 as "strongly agree" for each question (see Table 1).

The system setup for the user study was as follows. Several facets (among which `default`, `animal`, `vegetation`, `landscape`, `geographic`, and `device`) were available to the users for searching and exploring the photo collection. They were encouraged to also define new, personal, dimensions. In details, the unstructured `default` facet contains all tag in the tag DB, whereas `animal`, `vegetation`, and `landscape` dimensions classify the photos based on their main subject.

	Question	Agreement 0 1 2 3 4	Average Rating
1)	*"I like the Scenique system"*	0 0 0 2 8	3.80
2)	*"The multi-faceted paradigm helps me in search and browse the photo collection"*	0 0 0 7 3	3.30
3)	*"I found the integration of semantic tags and visual descriptors helpful"*	0 0 1 1 8	3.70
4)	*"The power expression in formulating requests satisfies me"*	0 1 0 2 7	3.50
5)	*"I found visual tools intuitive and easy-to-use"*	0 1 0 1 8	3.60
	Total average		3.58

Table 1. Mean user satisfaction ratings.

Facets `geographic` and `device` are used to organize pictures according to the place they have been shot and the electronic device that has been used to take them (e.g., Sony Ericsson mobile phone vs. Kodac Easyshare digital camera), respectively.

Table 1 shows that there is an overall positive agreement from all participants (3.58/4 on average) on all the questionnaire statements. In particular, all users liked Scenique (see question (1)) and, more important, believed that an integrated use of semantic annotations and visual descriptors is vital to get the retrieval process truly effective (statement (3)). Most of the participants found the multi-faceted paradigm helpful for their tasks (question (2)) and they judged the expressive power in formulating requests to the system more than sufficient (question (4)). With respect to the Scenique user interface, most of the participants found it very intuitive and easy-to-use; only one user rated 1 and commented: "*I found the user interface intuitive in general; however, I would prefer the possibility to exchange the labels of the axes space in the 3-D viewing room than to rotate the entire space*".

Such results are extremely encouraging, especially for the fact that the participants were not familiar with the image collection. We believe that results would be even better when using personal photo collections.

7 Conclusions

In this paper we introduced Scenique, a novel multi-faceted image search and browse system for effectively managing personal photo collections. Within Scenique, the user can define multiple facets with the aim to organize her photos under different points of view. In order to quickly locate image of interest, at querying time the user can formulate complex queries though logical expressions. She can also exploits provided visual tools for tag and visual trees exploration by taking benefit of a synchonized and active 3-D viewing room showing the current selected portion of photo collection. Feedback provided by a set of real users testifies that Scenique is an effective system and that its GUI is intuitive and easy-to-use.

References

[1] S. Ahern, M. Naaman, R. Nair, and J. H.-I. Yang. World Explorer: Visualizing Aggregate Data from Unstructured Text in Geo-referenced Collections. *JCDL*, pages 1–10, Vancouver, BC, Canada, 2007.

[2] S. Ardizzoni, I. Bartolini, and M. Patella. Windsurf: Region-Based Image Retrieval Using Wavelets. *IWOSS*, pages 167–173, Florence, Italy, 1999.

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

[4] I. Bartolini and P. Ciaccia. Scenique: A Multimodal Image Retrieval Interface. *AVI*, pages 476–477, Napoli, Italy, 2008.

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

[6] I. Bartolini, P. Ciaccia, and F. Waas. FeedbackBypass: A New Approach to Interactive Similarity Query Processing. *VLDB*, pages 201–210, Rome, Italy, 2001.

[7] I. Bomze, M. Budinich, P. Pardalos, and M. Pelillo. *The Maximum Clique Problem*. Kluwer Academic Publishers, Boston, MA, 1999.

[8] R. Chbeir, S. Atnafu, and L. Brunie. Image Data Model for an Efficient Multi-Criteria Query: A Case in Medical Databases. *SSDBM*, pages 165–174, Edinburgh, Scotland, UK, 2002.

[9] W. Dakka, P. G. Ipeirotis, and K. R. Wood. Automatic Construction of Multi-faceted Browsing Interfaces. *CIKM*, pages 768–775, Bremen, Germany, 2005.

[10] P. Duygulu, K. Barnard, J. F. G. de Freitas, and D. A. Forsyth. Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary. *ECCV*, pages 97–123, Copenhagen, Denmark, 2002.

[11] R. Fagin, R. V. Guha, R. Kumar, J. Novak, D. Sivakumar, and A. Tomkins. Multi-structural Databases. *PODS*, pages 184–195, Baltimore, Maryland, USA, 2005.

[12] P. Ganesan, H. Garcia-Molina, and J. Widom. Exploiting Hierarchical Domain Structure to Compute Similarity. *ACM Trans. Inf. Syst.*, 21(1):64–93, 2003.

[13] A. Graham, H. Garcia-Molina, A. Paepcke, and T. Winograd. Time as Essence for Photo Browsing through Personal Digital Libraries. *JCDL*, pages 326–335, Portland, Oregon, USA, 2002.

[14] J. Jeon, V. Lavrenko, and R. Manmatha. Automatic Image Annotation and Retrieval Using Cross-media Relevance Models. *ACM SIGIR*, pages 119–126, Toronto, Canada, 2003.

[15] J. Li and J. Z. Wang. Real-time Computerized Annotation of Pictures. *ACM MM*, pages 911–920, Santa Barbara, CA, USA, 2006.

[16] O. Maron and A. L. Ratan. Multiple-instance Learning for Natural Scene Classification. *ICML*, pages 341–349, San Francisco, CA, USA, 1998.

[17] V. Oria and M. T. Özsu. Views or Points of View on Images. *Int. J. Image Graphics*, 3(1):55–80, 2003.

[18] J.-Y. Pan, H.-J. Yang, C. Faloutsos, and P. Duygulu. Automatic Multimedia Cross-modal Correlation Discovery. *ACM SIGKDD*, pages 653–658, Seattle, USA, 2004.

[19] T. Quack, U. Mönich, L. Thiele, and B. S. Manjunath. Cortina: A System for Large-scale, Content-based Web Image Retrieval. *ACM MM*, pages 508–511, New York, NY, USA, 2004.

[20] L. A. Rowe, H. M. Vin, T. Plagemann, P. J. Shenoy, and J. R. Smith. Geographic Location Tags on Digital Images. *ACM MM*, pages 156–166, Berkeley, CA, USA, 2003.

[21] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-Based Image Retrieval at the End of the Early Years. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(12):1349–1380, 2000.

[22] G. Smith, M. Czerwinski, B. Meyers, D. Robbins, G. Robertson, and D. S. Tan. FacetMap: A Scalable Search and Browse Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):797–804, 2006.

[23] K.-P. Yee, K. Swearingen, K. Li, and M. A. Hearst. Faceted Metadata for Image Search and Browsing. *CHI*, pages 401–408, Ft. Lauderdale, Florida, USA, 2003.