

Semi-automatic Semantic Annotation of Images Using Machine Learning Techniques

Oge Marques and Nitish Barman

Department of Computer Science and Engineering
Florida Atlantic University
777 Glades Road, Boca Raton, FL 33431-0991
{omarques,NBAR4294}@fau.edu

Abstract. The success of the Semantic Web hinges on being able to produce semantic markups on Web pages and their components, in a way that is cost-effective and consistent with adopted schemas and ontologies. Since images are an essential component of the Web, this work focuses on an intelligent approach to semantic annotation of images. We propose a three-layer architecture, in which the bottom layer organizes visual information extracted from the raw image contents, which are mapped to semantically meaningful keywords in the middle layer, which are then connected to schemas and ontologies on the top layer. Our key contribution is the use of machine learning algorithms for user-assisted, semi-automatic image annotation, in such a way that the knowledge of previously annotated images – both at metadata and visual levels – is used to speed up the annotation of subsequent images within the same domain (ontology) as well as to improve future query and retrieval of annotated images.

1 Introduction

The success of the Semantic Web hinges on being able to produce semantic markups on Web pages and their components, in a way that is cost-effective and consistent with adopted schemas and ontologies. A significant portion of the Web is comprised of images. And as yet, most of the images on the Web have not been marked-up with respect to their semantic content. Our work is an attempt to address this issue by providing an intelligent approach to expediting and facilitating the semantic annotation of images. We explore the issues of annotating, crawling, querying, and retrieving images in a way that extends our previous work in the related field of content-based image retrieval [16,17], and redirects our efforts toward a ‘Semantic Web-ready’ solution.

Recent developments in Web engineering, such as the popularization of XML and Web services, will allow easy packaging, discovery and communication of structured information such as metadata schemas and indexes for media data such as images. These new advances will not, however, enable the automatic generation of these metadata. In other words, it is still very costly and human intensive to generate powerful metadata and media descriptors to populate such

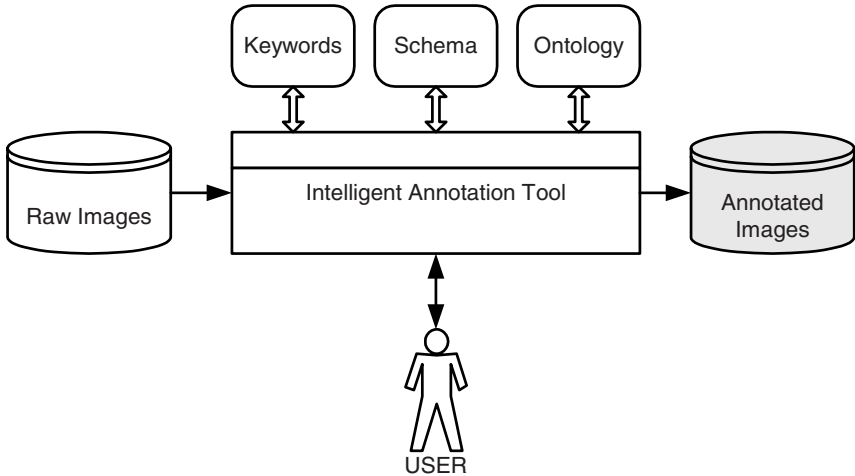


Fig. 1. Intelligent image annotation tool: conceptual diagram

metadata schemas and structures for very large media databases. Our research addresses this topic and suggests an intelligent, user-assisted, semi-automatic solution for annotating (potentially huge) repositories of raw images, and storing the result (e.g., in XML-formatted RDF-schema) in an ‘annotated image’ repository (Figure 1). The resulting annotated image archive would then be available for queries by visual content (giving an image as an example) and/or keywords as well as for data exchange with other Web-based applications.

In addition to its primary goals – consistent, cost-effective, fast, intelligent annotation of visual data – the method proposed in this paper also aims at improving the performance of subsequent image query and retrieval operations on the annotated image repository. The image retrieval problem has not been successfully solved yet, despite almost a decade of work [24] and dozens of research prototypes and commercial products developed as a result. The use of keywords only does not provide satisfactory results (as anyone who has used the ‘Images’ option in GoogleTM [8] knows), mostly because the correlation between the keywords used in the query and the desired image file has to be established by unreliable, vague, not always available metadata, in the form of surrounding text and/or HTML tags, such as the contents of the **alt** attribute of the **img** element in the HTML document. Pure content-based image retrieval (CBIR) also falls short of providing an adequate solution to the problem, mostly due to the limitations of current computer vision techniques in bridging the gap between visual features and their semantic meaning. There is a clear need to cleverly combine the two in order to yield a better solution, and several researchers (e.g., [21]) have been working on techniques that combine textual and visual cues to improve retrieval results.

In previous work [17] we assumed a purely content-based scenario and dismissed the use of keywords because of the tedious, time-consuming, and subjective nature of human annotation and the ambiguous results it often produces. However, recent developments in RDF schemas, ontologies, and other Semantic Web-related efforts, have led us to look at the problem from a different perspective, one in which the use of metadata is a required part of the solution. We decided to concentrate our efforts in improving the efficiency of the annotation process, ensuring the meaningfulness of the chosen keywords, and preparing the framework for exchange and reuse of annotated visual data.

The content of this paper is structured as follows. Section 2 summarizes related work, particularly in the fields of image annotation and combination of textual and visual features for image retrieval. Section 3 explains the proposed architecture and its main components. Section 4 focuses on the techniques and algorithms for image annotation, and subsequent query and retrieval. Finally, Section 5 presents some concluding remarks.

2 Related Work

In this Section we discuss works related to semi-automatic annotation of textual data, image annotation, combination of keywords and visual features in image retrieval, and use of metadata for description of multimedia content, particularly images.

2.1 Semi-automatic Semantic Annotation of Text

There have been a number of papers on semi-automatic semantic annotation of natural language text. In [6], Erdmann et al. describe an ontology-based annotation tool developed under the KA^2 -initiative. Since the raw data is in text format, their system uses a tokenizer, a lexical analysis component, and a chunk parser, in such a way as to associate single words or complex expressions with a concept from the ontology, connected through the domain lexicon linkage. Recognized concepts and dependency relations between concepts are highlighted as suggested annotations, so that all relevant information in the document with regard to the ontology is proposed to the annotator, whereas the actual process of annotation is delegated to the annotation tool. This work also addresses the need of treating document annotation and ontology maintenance in a cyclic fashion. Even though it focuses on textual information, it provided us some useful guidelines for the development of our work, such as the suggestion to keep the ontology simple and the need to support annotators with interactive GUIs. In subsequent papers [9,10], this work is extended to an annotation environment that allows construction of relational metadata, while maintaining the same strict guidelines of annotation for the sake of generating machine-understandable data, viz. to allow relational metadata, proper reference and consistency.

2.2 Image Annotation

Work on semantic image annotation is still in its infancy, and we found only two examples that relate more closely to the work proposed in this paper. One is Photo-RDF [15], a project for describing and retrieving JPEG pictures with (RDF) metadata. Photo-RDF describes the RDF schemas, offers a data-entry program (RDFPic) for quickly entering metadata for large numbers of photos, proposes a way to serve the photos and the metadata over HTTP, and suggests search methods to retrieve photos based on their descriptions. It uses the Dublin Core (DC) schema as well as additional optional schemas for technical data. Its data-entry tool (RDFPic) has a GUI that allows the user to reuse the annotation values that were entered for previous photos in a nice and friendly way, saving annotation time.

The other is CBSA [4], a content-based soft annotation tool that allows adding semantic labels to a small set of images. Machine learning algorithms extend the labels to larger image sets, based on the visual similarity between the new (unlabeled) images and previously labeled images. This work is one of the most recent relevant results in combination of visual and textual features in image retrieval, but it does not rely on ontologies and/or schemas.

2.3 Combining Visual and Textual Features in Image Retrieval

Researchers in the field of visual information retrieval have recognized the many difficulties in retrieving images from large repositories using only visual contents. It is also widely accepted that the retrieval of images annotated with keywords may provide potentially better results, which unfortunately depend on the amount, quality, and consistency of the metadata associated with each image. In very well-controlled annotation environments, the use of keywords, even when hidden from the end user (see [5] for an example), is likely to provide better results than the sole reliance on visual properties extracted using current computer vision techniques.

The main drawbacks of image query and retrieval by keywords can be classified in three levels, in increasing degree of complexity. First, on a practical level, most of the images available over the Web – or the HTML pages in which they are embedded – are not annotated in a consistent, meaningful way. Secondly, when keywords are used, they are often isolated from any framework – such as schema or ontology – that could put them in context and allow further inference of relevant knowledge. Third, on a deeper level, the contents, meaning, and emotions associated with or evoked by a picture can hardly be adequately expressed in words [24].

One of the best-known efforts in combining visual and textual features in image retrieval is the work by Sclaroff et al. [21], in which textual and visual statistics from HTML documents are combined into a single index vector. Their system, ImageRover [13], enables the user to provide relevance feedback which is used to set the weights of the distance metrics used to compare the images. It relies on whatever textual information is available next to the image within the

HTML page in which it is embedded, using the **Latent Semantic Indexing (LSI)** method, which assigns different weights to words, depending on their importance; for instance, words found inside the **alt** attribute of the **img** element are assigned a higher weight than those found within a pair of **h2** tags. But this work also does not refer to schemas or ontologies.

2.4 Use of Metadata for Description of Multimedia Content

The standardization and widespread adoption of metadata for describing multimedia content has been a goal of many researchers and standard bodies in the field. It is widely recognized that the use of standardized descriptors, description schemes, etc., such as proposed in the MPEG-7 standard [12], will promote an easier integration of tools and applications without limiting the creative efforts of researchers involved in open topics such as efficient multidimensional indexing structures or faster search algorithms for multimedia data.

For the specific case of images, two recent developments toward metadata adoption and standardization are worth mentioning: the “Data Dictionary for Images” [19], a comprehensive list of technical data elements required to manage digital image collections, and the Adobe eXtensible Metadata Platform XMP [1], an RDF-based framework for adding machine-readable labels, or semantic content to application files, databases and content repositories.

The adoption of RDF, XML, and other Semantic Web-oriented technologies and concepts in visual information retrieval can also be found in the description of European projects ARTISTE (An integrated Art Analysis and Navigation Environment) [3] and SCULPTEUR (Semantic and Content-based mULTimedia exPloiTation for EURopean benefit) [22].

3 Proposed Architecture

In this Section, we outline our proposed architecture. First, we present our holistic vision, which should put in perspective where our work fits in the grand scheme. Next, we describe our proposed three-layer model for annotation, query, and retrieval in more detail.

3.1 The Big Picture

The work described in this paper is part of a bigger vision, in which our intelligent image indexing and annotation solution is exposed as an Intelligent Image Indexing Web Service (I3WS) to potential clients. This Web service will take a raw image repository (plus a few optional constraints and parameters – schemas, keywords, ontologies, etc.) as an input and return its annotated version as a result. Clients of the service would save annotation time and manpower and get a free semantic mark-up, which they could use in any way (e.g., to improve the visibility and rank of their pages to a search engine).

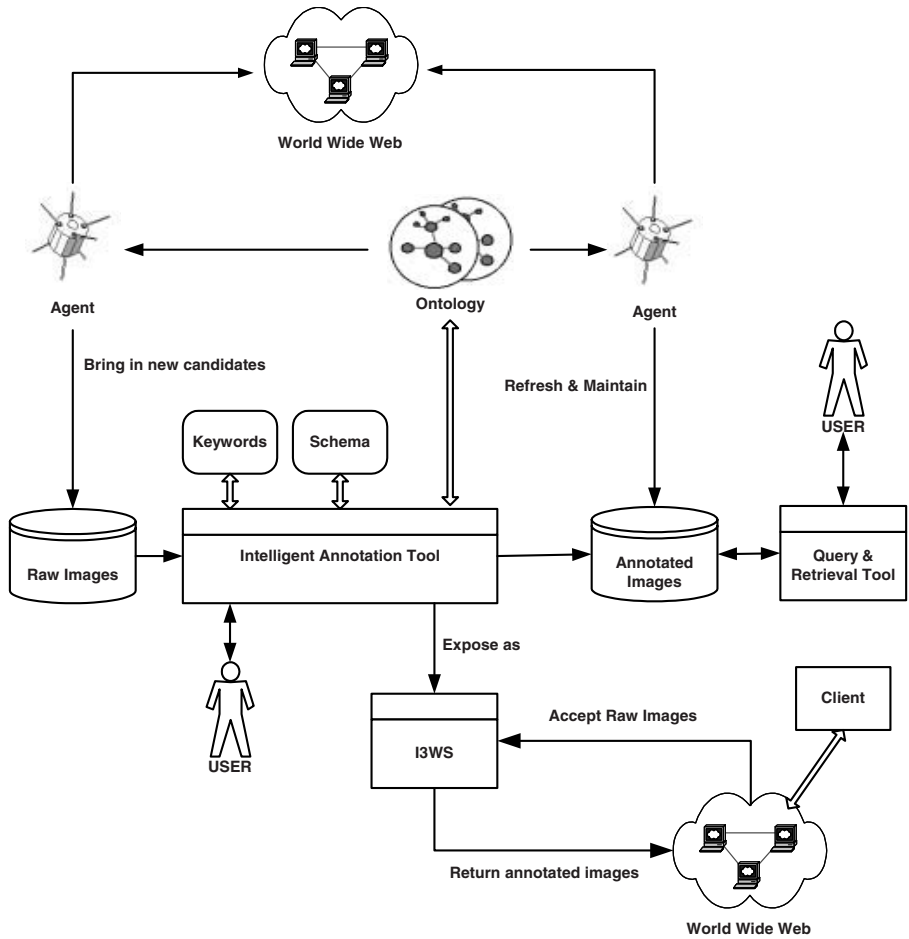


Fig. 2. The big picture.

In addition to annotating third-party repositories, our system keeps its own image repositories (with locally stored thumbnails and links to the actual images' locations on the Web) and a Web-based GUI for performing query and retrieval on them (see Section 4.3 for more details). In a future stage, we also plan to use ontology-focused agents [11] to crawl the Web and bring new candidate images that fit under a given ontology to the 'raw image repository' for subsequent annotation. The agents will also periodically update the annotated repository by checking the integrity of the links between our local thumbnails and the full-size images to which they correspond. Figure 2 depicts our overarching vision. The focus of our current work, however, is the development of the Intelligent Annotation Tool (IAT) and the Query and Retrieval Tool (QRT).

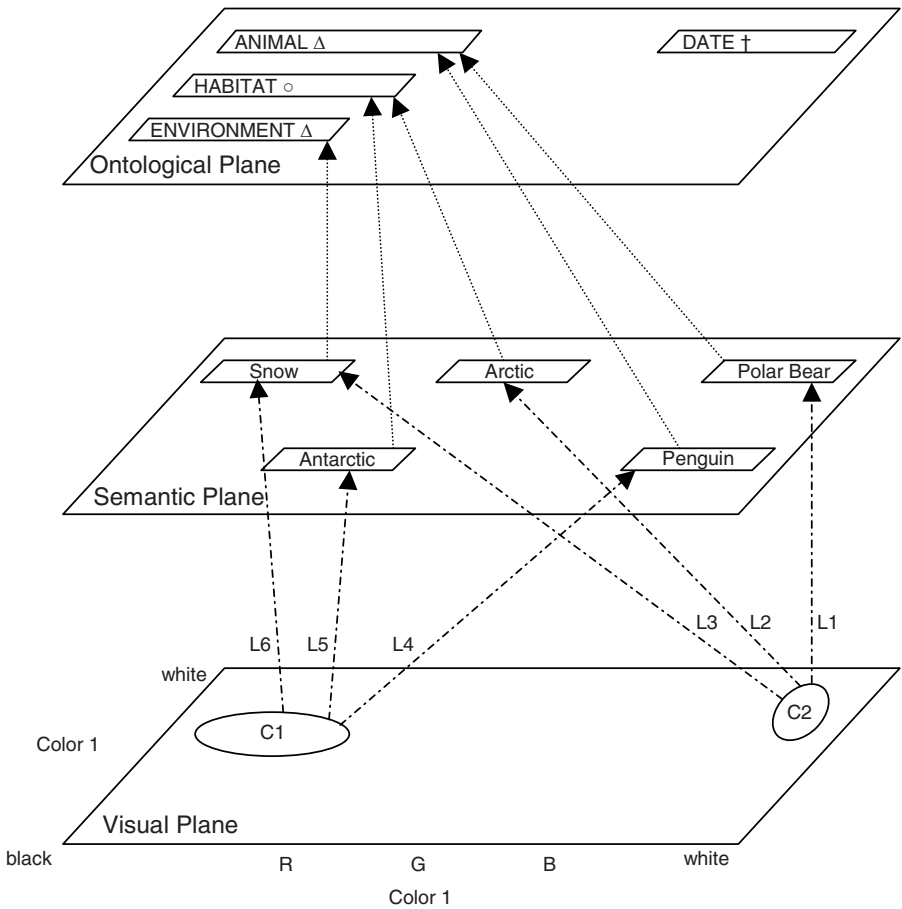


Fig. 3. The proposed three layer architecture, where: \circ = content-dependent metadata, Δ = content-descriptive metadata, \dagger = content-independent metadata.

3.2 The Three-Layer Model for Annotation, Query, and Retrieval

We propose a three-layer model, in which the inter-planar relationships are used as a basis for making inferences about the semantic content of not yet annotated images. The visual plane organizes visual information extracted from the raw image contents. The information in the visual plane is mapped to semantically meaningful keywords in the semantic plane. The keywords in the semantic plane, in turn, map to specific slots in the domain-specific schema(s) contained in the ontological plane.

The bottom layer (*visual plane*) is a K -D space containing the visual feature vectors corresponding to each image in the annotated database. The size and nature of these feature vectors is purely dependent on the visual feature

extraction algorithms that one wishes to implement, from simple color-based measures such as color histograms, to complex and task-specific features, such as eigenfaces. Ideally, the extracted feature vectors should be compact (for the sake of both storage and processing) and discriminative, i.e., the distribution of the vectors in the K - D space should resemble the relative visual similarity among the images to which they map, in such a way that visually similar images should appear clustered together.

The middle layer (*semantic plane*) contains all the keywords that have been used to annotate the images.

The top layer (*ontological plane*) contains the schemas and ontologies to which the keywords belong. It contains information about the relationships (i.e. dependencies) between the various tags in the RDF schema. Storing this information in a separate plane allows us to easily extend the RDF schema and automatically re-annotate the images in the annotated images database with the new keyword-tag combinations used in the extended schema.

Figure 3 shows a simplified version of the model in which the visual plane is a 2 - D plane whose axes are the two predominant colors in the image, containing two clusters (C1 and C2). Each cluster maps to the keywords that have been used to annotate at least one image in that cluster. The weight of each link (L1, L2, ..., L6) is proportional to the number of images in that cluster which have been annotated with that keyword. The ontological plane depicts a very simple 4-slot schema, with different types of metadata: *content-dependent metadata* (e.g., *environment*, which may relate strongly to the predominant color in a picture), *content-descriptive metadata* (e.g., name of an *animal*), and *content-independent metadata* (e.g., the *date* when the photo was taken).

Note, also, that each keyword in the semantic plane maps to a slot in the schema contained in the ontological plane. The knowledge that visually similar images may map onto independent – maybe even mutually exclusive – keywords is modeled using a Bayesian network, whose details are explained in Section 4.1.

4 Detailed Description of Our Approach

Broadly, our application makes use of knowledge (semantic and visual) derived from previously annotated images to expedite and facilitate the annotation of subsequent images. The knowledge derived is also used to improve query and retrieval of annotated images.

In this Section, we describe our application in more detail. We begin by noting that the proposed approach assumes that meaningful visual features are extracted from the raw images, and that images that are sufficiently similar are grouped together in clusters. It is beyond the scope of this paper to discuss the implications behind the choice and design of feature extraction and clustering strategies, issues that have been addressed by [2,7,14,16,20,23], among many others. Next, we describe what we believe to be the most important and innovative aspect of our contribution – the Intelligent Inference Engine (IIE). Finally, we provide a more detailed procedural description of each phase of our application.

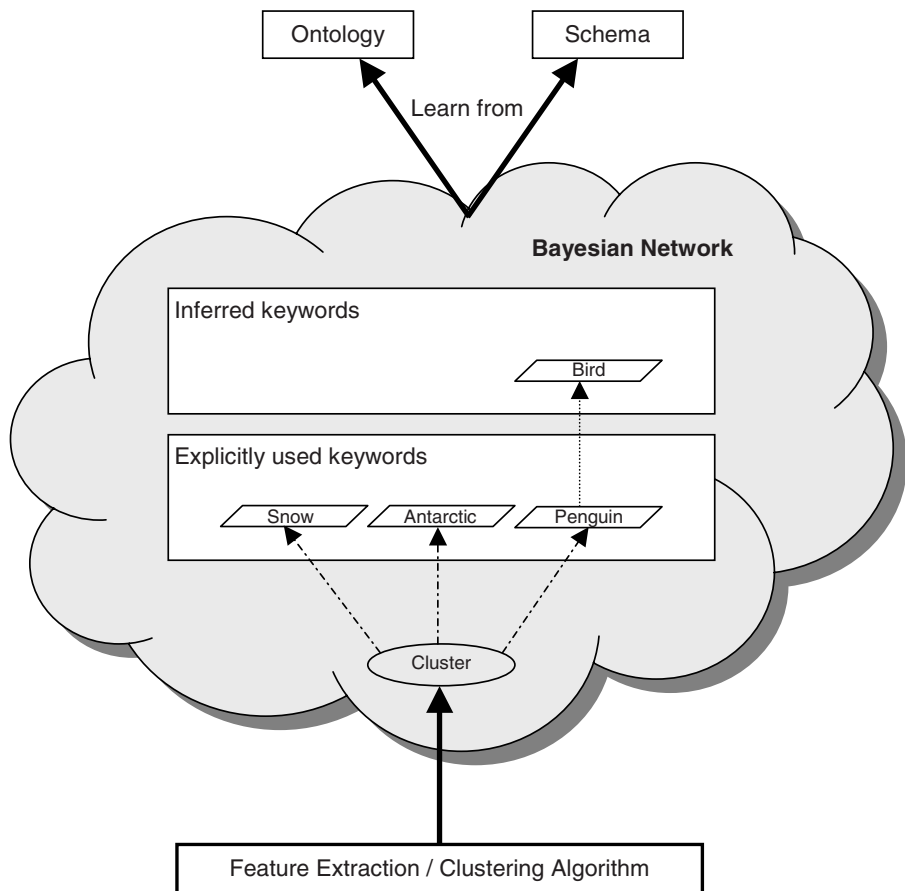


Fig. 4. The Bayesian network used to model relationships between low-level features and keywords – simplified view.

4.1 The Intelligent Inference Engine

We model the relationship between visual features and the keywords using a Bayesian network [18] (Figure 4), whose root node represents cluster membership and whose first level of nodes includes all the keywords explicitly used in the annotation process. The network can include additional levels which correspond to keywords that are not explicitly used in the annotation process.

The prior probabilities are derived from the clusters' sizes as the ratio of the size of cluster C_i and the total number of images in the repository (N). The mapping between clusters and explicitly used keywords is performed at the first level of the Bayesian network. The conditional probability of occurrence of any keyword given that the image belongs to a certain cluster in the visual plane can

be easily computed and updated every time an individual image is annotated as can the conditional probabilities of keywords in other levels of the Bayesian network.

What distinguishes our approach from classical learning Bayesian networks is the fact that our system learns both from the ontologies and schemas from the top layer as well from the joint occurrence of visual features and keywords in annotated images. What makes our approach innovative is that we use not only the joint probabilities of various keywords in order to make inferences but we also use knowledge derived from the domain specific schema (e.g., which keywords are valid, how a keyword may be a subcategory of another, mutually exclusive keywords, etc.) to “validate” those inferences and derive additional inferences which could not be otherwise made.

We offer a few simple examples (see Figures 3 and 4) to help clarify what we mean.

Example 1: Using joint probabilities of keywords to make inferences.

If several images have already been labeled with the keywords ‘antarctic’ and ‘penguin’, and the user labels the current image with the keyword ‘penguin’, the IAT can suggest that perhaps it should also be labeled with ‘antarctic’.

Example 2: Using joint probabilities of keywords to prevent inconsistent or contradictory annotation.

If the IAT has accumulated knowledge that keywords ‘arctic’ and ‘penguin’ never occur together, and the user tries to label the current image with both these keywords, the IAT can issue a warning against it.

Example 3: Validating an inference based on the knowledge of schemas. If the user attempts to label an image with both ‘arctic’ and ‘antarctic’, the IAT prevents him from doing so, based on the knowledge that both these keywords map onto the same slot in the schema (i.e., they are mutually exclusive).

Example 4: Extending keywords based on the knowledge of schemas.

An image annotated with the keyword ‘penguin’ could also be associated with the keyword ‘bird’, without requiring explicit user intervention.

4.2 Annotation

The annotation phase is responsible for adding metadata to a collection of raw images. It takes a raw image and a set of keywords and returns the RDF-formatted description of the image. The annotation process is divided into two stages: initialization and cataloguing.

Initialization Stage. The initialization stage establishes and populates the Bayesian network. **At the beginning of this stage the annotated image repository is empty. The purpose of this stage is to populate it with enough images to allow the IAT to make intelligent inferences. This is a human-intensive stage in that the user is required to annotate raw images without receiving any suggestions from the IAT.** We do, however, attempt to reduce the burden on the user by offering him a pre-selected subset of allowed keywords that can be used during annotation, in such a way as to save annotation time, minimize ambiguities, and ensure compliance with the selected schema/ontology.

Once a raw image has been annotated, it is saved in the annotated image repository. At the end of this stage, we subject the annotated images to feature extraction and clustering algorithms. We then map these images and their attributes into the three-layer model described earlier. This process establishes our Bayesian network.

Cataloguing Stage. The cataloguing stage is similar to the initialization stage in that raw images are annotated with keywords and saved to the annotated image repository. However, this stage is not as human-intensive because the IAT offers suggestions about which keywords should be used to annotate the current (raw) image. **The process of generating and offering suggestions is as follows. First, visual information (feature vectors) is extracted from the current image, using a combination of color-, texture, shape-based computer vision techniques.** Next, the extracted feature vector is mapped onto the visual plane. We select the cluster (if any) that is closest to the feature vector of the current image and analyze the weights of the links from the cluster to the keyword vectors in the semantic plane. Based on that analysis (including calculation of relevant probabilities and knowledge inferred from the schema), we infer the most probable semantic content of the raw image and suggest the appropriate keywords to the user. As part of the ongoing learning process, at the end of the cataloguing stage (or after a certain percentage increase in the number of annotated images), the annotated images are re-clustered, the weights in the Bayesian network are updated, thus increasing the amount of empirical data available to us in order to make inferences.

4.3 Query and Retrieval

The other distinct phase in our application is the Query and Retrieval Phase. The application allows the user to search for an image or set of images that match the parameters specified in the user's query. The query could be either a textual description (i.e. a set of keywords) of the target image(s), an example image, or both. The query information (keywords, visual features extracted on-the-fly from the query image, or both) would then be compared with the data associated with the images in the annotated image repository. In all cases, a set of images that most closely match the parameters defined in the query would be returned (as thumbnails) to the user. If the desired image is among the set of

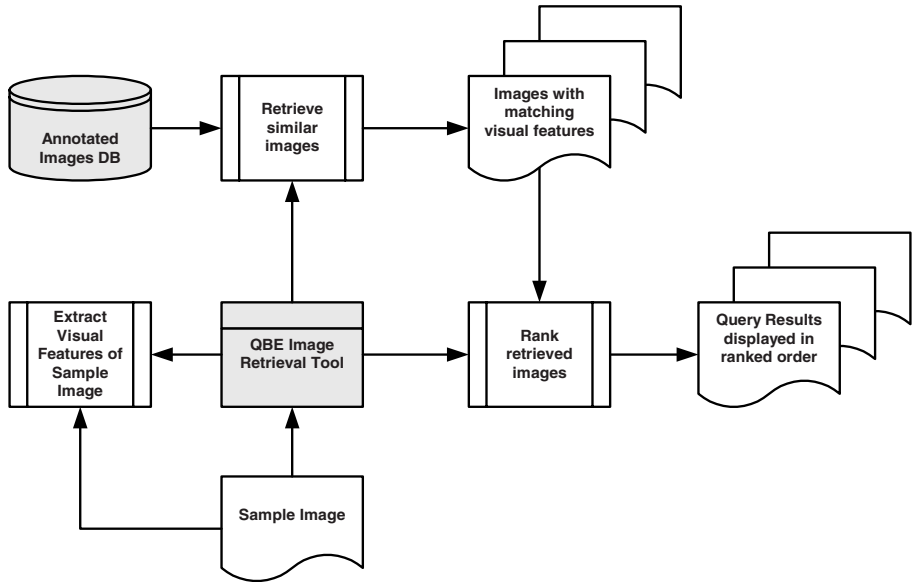


Fig. 5. Query by example

images returned, the search is completed. At any rate, the user can interactively browse and refine the results of the query.

Our proposed system supports four types of queries:

Query by example (QBE). In a QBE (Figure 5) the user provides the system with an example image and expects the system to retrieve similar images. The user may or may not have a target image in mind. The query is considered successful if the user selects one or more images returned as a result of the query. The QBE works as follows:

1. Extract visual information from the example image.
2. Map the extracted feature vector onto the visual plane.
3. Determine if there are any images (*not* clusters) within range (experimentally determined threshold) of the example image's feature vector.
4. Retrieve up to M images that are closest to the feature vector of the example image using an appropriate metric.
5. Rank the images based on their proximity to the example image's feature vector.
6. Display retrieved images in rank order.

Query by example – extended (QBEE). A QBEE is very similar to a QBE except that similarity of (inferred) semantic content, instead of visual similarity,

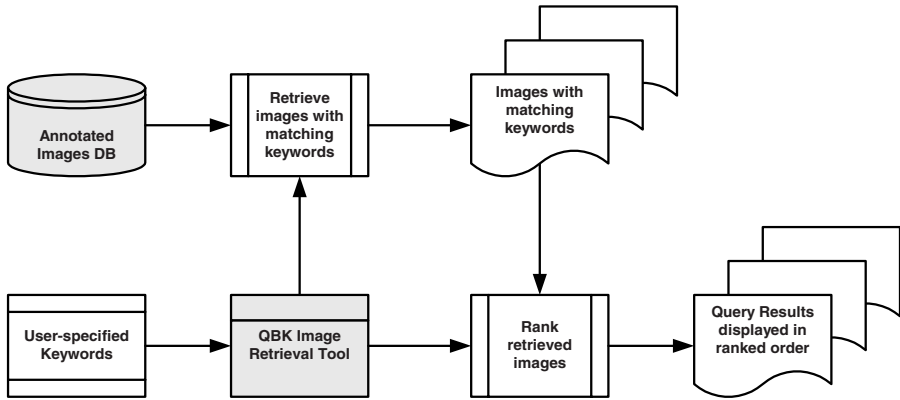


Fig. 6. Query by keyword

is the basis for measuring similarity to the example image. The process is outlined below:

1. Extract visual information from the example image.
2. Map the extracted feature vector onto the visual plane.
3. Determine if there are any image clusters within range (experimentally determined threshold) of the example image's feature vector.
4. Select the cluster (if any) that is closest to the feature vector of the example image.
5. Analyze the weights of the links from the images in the selected cluster to the keyword vectors in the semantic plane.
6. Based on the above analysis (including calculation of joint probabilities and probabilities of co-occurrence), infer the most probable semantic content of the example image.
7. Retrieve up to M images that are most similar to the inferred semantic content of the example image and rank accordingly.
8. Display retrieved images in rank order.

Query by keyword (QBK). In a QBK (Figure 6) the user types in one or more keywords, which are mapped to the schema(s) that characterize(s) the domain(s) to which the annotated images belong. The system is expected to retrieve those images that have at least one selected keyword associated with them. The process by which the system retrieves images based on a QBK is as follows:

1. Retrieve all images that have at least one of the specified keywords associated with them.
2. Rank the images based on the number of specified keywords associated with them as well as knowledge inferred from our Bayesian network.
3. Display retrieved images in rank order.

Query by both example and keyword (QBEEK). Finally, in a QBEEK the user provides a sample image as well as keywords in order to retrieve matching images. The procedure for retrieving images in this case is essentially analogous to a QBK followed by a QBE. The procedure is outlined below:

1. Extract visual information from the example image.
2. Map the extracted feature vector into the visual plane.
3. Select a subset of images that have at least one of the specified keywords associated with them.
4. From the subset of images selected in the step above, select all images that are within range of the example image's feature vector.
5. Rank the selected images based on both visual similarity to the example image and the semantic similarity to the set of query keywords.
6. Retrieve up to M images with the highest rank (score).
7. Display retrieved images in rank order.

5 Concluding Remarks

We have proposed a three-layer architecture for semi-automatic semantic annotation of images using machine learning techniques. This work extends and redirects our previous work in the related field of content-based image retrieval toward a Semantic Web-oriented solution.

The proposed architecture allows a good degree of independence among efforts in improving its supporting technologies and algorithms, such as:

- Development of better algorithms for extraction of visual features.
- Implementation of better clustering algorithms.
- Investigation of improved learning strategies.
- Deployment of ontology-focused intelligent agents.
- Enlargement and maintenance of ontologies and their mapping to keywords.

It is also acknowledged that the quality of the individual algorithms – particularly those used for visual feature extraction, clustering, similarity calculations, and mapping between the bottom and middle layer of our model – is a crucial factor to the success of the proposed methodology. A more systematic way of addressing these potential problems should be devised in future stages of this research.

In addition to its primary goals – consistent, cost-effective, fast, intelligent annotation of visual data – the method proposed in this paper also aims at improving the performance of subsequent image query and retrieval operations on the annotated image repository.

When fully implemented, the work described in this paper will provide an intelligent image indexing and annotation solution to Web-based clients, who will be able to save precious and costly annotation time and manpower and get a free semantic mark-up on their raw images.

Acknowledgments. The authors want to thank Dragutin Petkovic for his excellent comments and suggestions.

References

1. Adobe XMP. <http://www.adobe.com/products/xmp/main.html>.
2. S. Aksoy and R.M. Haralick. Graph-theoretic clustering for image grouping and retrieval. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 23–25, Fort Collins, CO, USA, 1999.
3. ARTISTE (An integrated Art Analysis and Navigation Environment). <http://www.ecs.soton.ac.uk/~km/projs/artiste/>.
4. E. Chang, K. Goh, G. Sychay, and G. Wu. CBSA: content-based soft annotation for multimodal image retrieval using bayes point machines. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(1):26–38, January 2003.
5. I. J. Cox, M. L. Miller, T. Papathomas, J. Ghosn, and P.N. Yianilos. Hidden annotation in content based image retrieval. In *Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries*, pages 76–81, San Juan, Puerto Rico, 1997.
6. M. Erdmann, A. Maedche, H. Schnurr, and S. Staab. From manual to semi-automatic semantic annotation: About ontology-based text annotation tools. In *Proceedings of the COLING 2000 Workshop on Semantic Annotation and Intelligent Content*, Luxembourg, 2000.
7. Y. Gong, G. Proietti, and C. Faloutsos. Image indexing and retrieval based on human perceptual color clustering. In *Proceedings in IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 578–583, 1998.
8. GoogleTM image search. <http://www.google.com/imghp>.
9. S. Handschuh, S. Staab, and F. Ciravegna. S-CREAM – Semi-automatic CREAtion of Metadata. In *Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02)*, October 2002.
10. S. Handschuh, S. Staab, and A. Maedche. CREAM – creating relational metadata with a component-based, ontology-driven annotation framework. In *Proceedings of the International Semantic Web Working Symposium*, Stanford University, California, USA, 2001.
11. J. Hendler. Agents and the semantic web. *IEEE Intelligent Systems*, March/April 2001.
12. J. Hunter. Adding Multimedia to the Semantic Web – Building an MPEG-7 Ontology. In *Proceedings of the International Semantic Web Working Symposium*, Stanford University, California, USA, 2001.
13. ImageRover. <http://www.cs.bu.edu/faculty/sclaroff/ivc/ImageRover/Home.html>.
14. S. Krishnamachari and M. Abdel-Mottaleb. Image browsing using hierarchical clustering. In *Proceedings of the IEEE International Symposium on Computers and Communications*, pages 301–307, 1999.
15. Y. Lafon and B. Bos. Describing and retrieving photos using rdf and http. Technical report, W3C, April 2002. <http://www.w3.org/TR/photo-rdf>.
16. O. Marques and B. Furht. *Content-Based Image and Video Retrieval*. Kluwer Academic Publishers, Norwell, MA, 2002.
17. O. Marques and B. Furht. MUSE: A content-based image search and retrieval system using relevance feedback. *Multimedia Tools and Applications*, May 2002.

18. R.E. Neapolitan. *Learning Bayesian Networks*. Prentice-Hall, Upper Saddle River, NJ, 2004.
19. NISO Z39.87-2002 AIIM 20-2002 data dictionary – technical metadata for digital still images. <http://www.niso.org/committees/index.html>.
20. Y. Rui, T. S. Huang, and S.-F. Chang. Image retrieval: Current techniques, promising directions and open issues. *Journal of Visual Communication and Image Representation*, 10(1):1–23, March 1999.
21. S. Sclaroff, M. La Cascia, and S. Sethi. Unifying textual and visual cues for content-based image retrieval on the world wide web. *Computer Vision and Image Understanding*, 75(1/2):86–98, July/August 1999.
22. SCULPTEUR (Semantic and Content-based mULTimedia exPloiTation for EUROpean benefit). <http://www.sculpteurweb.org/>.
23. G. Sheikholeslami, W. Chang, and A. Zhang. Semantic clustering and querying on heterogeneous features for visual data. In *Proceedings of the 6th ACM International Multimedia Conference*, pages 3–12, Bristol, England, 1998.
24. 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 Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, December 2000.