1995-05-27

World Wide Web Image Search Engines

Boston University Computer Science Dept. TR95-016
Position paper for *NSF Workshop on Visual Information Management*, **Cambridge, MA, June 1995.**

World Wide Web Image Search Engines

Stan Sclaroff
Image and Video Computing Group
Computer Science Department
Boston University
Boston, MA 02215

Abstract

We propose the development of a world wide web image search engine that crawls the web collecting information about the images it finds, computes the appropriate image decompositions and indices, and stores this extracted information for searches based on image content. Indexing and searching images need not require solving the image understanding problem. Instead, the general approach should be to provide an arsenal of image decompositions and discriminants that can be precomputed for images. At search time, users can select a weighted subset of these decompositions to be used for computing image similarity measurements. While this approach avoids the search-time-dependent problem of labeling what is important in images, it still holds several important problems that require further research in the area of query by image content. We briefly explore some of these problems as they pertain to shape.

1 Introduction

For over a year now there have been "worms" and "spiders" crawling the World Wide Web, collecting index information about the text documents they find. These search engines extract text indexing information that is later used to guide interactive searches without having to retrieve the actual web documents. The scale of these databases is impressive. For instance, Lycos, a web search engine at CMU, supports interactive keyword-based searches for an index of over four million web documents. To remain up to date, the engine retrieves and analyzes thousands of documents daily.

What is needed is an equivalent web image search engine that crawls the web collecting information about the images it finds, computes the appropriate image decompositions and indices, and stores this extracted information for Lycos-style searches based on image content.

Web image search engines could be applied profitably in many areas; for instance, in searching on-line catalogs of consumer goods and services, or for enforcing image copyrights by sniffing out unauthorized copies on the web. Such a web crawler would also be useful to researchers studying image databases, serving as a very large testbed for image database indexing methods. Given the number of unsolved problems in image understanding, however, building a web image worm may seem overly ambitious.

Fortunately, just as searching text need not require understanding the text's meaning (Lycos simply extracts keywords using an algorithm that considers characteristics like word placement, word frequencies, etc.), indexing and searching images need not require solving the image understanding problem. Instead, the general approach is to provide an arsenal of image decompositions and discriminants that can be precomputed for images: color histograms, edge orientation histograms, texture measures, shape invariants, eigendecompositions, etc. At search time, users can select a weighted subset of these decompositions to be used for computing image similarity measurements.

2 Shape-based Search

While in some ways, this approach avoids the search-time-dependent problem of labeling what is important in images, it still holds several important problems that require further research in the area of query by image content. We will now briefly explore some of these problems as they pertain to shape. Much of this is relevant to searches on other image properties (e.g., texture, color, grayscale appearance).

First, there is the problem of developing appropriate image decompositions and invariants. For shape-based search, a whole cadre of shape discriminants has been employed: area, axes of inertia, Hausdorff distance, higher-order algebraic moment invariants, etc. Discriminants can be useful for constraining a search space, an important tool in searching an index of possibly millions of web images. Unfortunately, these discriminants discard significant semantic information about the shape. In other words none of these discriminants allows for reconstruction of the original shape, and thus it is impossible to perform closer comparison of shapes once a search is narrowed to a small subspace. Since what is important in images can be search-dependent, decompositions should preserve as much perceptually important information about the image as possible, while providing a compact (lossy) encoding of the signal [1].

Next, there is the problem of determining the appropriate subset of similarity measures for a particular search, and then how the various measures should be weighted and combined. Directly prompting users for weightings and thresholds is problematic, since it may require that users grasp the technical details of the underlying representation (e.g., higher-order moment invariants). One way around this is to allow the users to train the system for a search by providing positive and negative examples, and or to allow users to iteratively refine a search. The search engine can then automatically select the appropriate weighting and thresholds for guiding the search [2].

Another problem associated with selecting subsets of shape measures is one of providing selective invariance. For instance, in human shape similarity judgments, sometimes scale and rotation invariance are important, other times not [3]; it is therefore desirable to duplicate this performance in our image database search algorithms. Many of the shape discriminants mentioned are not scaling, translation, or rotation invariant.

Invariance to shape deformation is also important: humans will report that shapes are "similar" even when the two shapes are actually deformed versions of each other. Nearly none of the currently employed shape discriminants performs consistently well when there is nonrigid deformation or significant change in 3-D viewing angle. Furthermore, it is not enough to provide robustness to deformation. To measure the shape similarities between two objects, a shape decomposition must be able to encode the types of deformations that relate them. Shape decompositions should provide deformation "control knobs" that roughly correspond to a human user's notions of perceptual similarity, and it should be possible to selectively turn these control knobs on and off.

Modal matching is a physically-motivated technique for deformable shape description that was developed with many of these issues in mind [4]. Shape is described in terms of an ordered set of deformations from an initial shape: starting with bends, tapers, shears, and moving up towards higher-frequency wiggles. In contrast to the shape discriminants mentioned, the finite element integrals used in the modal model formulation provide greater robustness to sampling, outliers, and missing data. Furthermore, modal models provide quasi-invariance to different types of nonrigid deformation, while also providing an ordered, orthogonal, *encoding* of the nonrigid deformation that relates a candidate shape to a shape prototype or shape category.

3 Conclusion

In building a web image search engine, there are some problems in other research areas that need to be addressed. In particular, there are the logistical problems of a user interface that gives feedback for iteratively searching web images. There are also problems in partitioning such a large index to allow interactive search; work in categories and prototype-based descriptions may help in this area. Lastly, spatial relationships may be important in image search, and so support for this must also be included.

Whether or not all of these problems are solved, the idea of using the web as a large distributed image database testbed is a powerful one. It may even be possible to provide simple but useful web image search engines given current technology.

References

[1] A. Pentland, R. Picard, and S. Sclaroff. Photobook: Tools for content-based manipulation of image databases. *International Journal of Computer Vision*, to appear 1995.

[2] R. Picard and T. Minka. Vision texture for annotation. *Multimedia Systems*, 3(3):3–14, 1995.

[3] D. Mumford. Mathematical theories of shape: Do they model perception? In *Proc. SPIE Conf. on Geometric Methods in Computer Vision*, volume 1570, 1991.

[4] S. Sclaroff and A. Pentland. Modal Matching for Correspondence and Recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(6):545–561, 1995.