

Interaction Harvesting for Document Retrieval

Proposal for degree of Master of Science – Fall 2004

Noah S. Fields

Advisor:

John Maeda
E. Rudge and Nancy Allen Professor of Media Arts and Sciences
MIT Media Laboratory

Reader:

Deb Roy
AT&T Career Development Professor of Media Arts and Sciences
MIT Media Laboratory

Reader:

Walter Bender
Senior Principal Research Scientist
MIT Media Laboratory

Interaction Harvesting for Document Retrieval

Table of Contents

1.0 Abstract.....	3
2.0 Introduction.....	4
3.0 Background.....	5
4.0 Proposed Approach.....	6
4.1 Assessment and Evaluation.....	7
5.0 Summary.....	8
6.0 Deliverables.....	8
7.0 Schedule.....	8
8.0 Bibliography.....	9

1.0 Abstract

Despite advances in search technology, few software systems have been developed which accurately categorize multimedia files. The most successful systems for searching images, sounds, or movies rely on keyword annotation to provide meaningful search terms for non-text documents. Unfortunately, such systems usually require the author to enter the keywords manually, a task which is commonly neglected, or is executed poorly. This thesis proposes an approach to document categorization called Interaction Harvesting, wherein systems establish document relationships based on organizational and curatorial cues, harvested from the mouse and click gestures of an online community. Specifically, the spatial and temporal proximity and placement of documents are taken as indicators of document similarity. We propose an expansion technique whereby such proximal documents exert weighted keyword influences on each other. We hypothesize that these approaches will form a document classification framework that relieves some of the difficulty of the annotation process, while providing keyword-equivalent retrieval performance.

2.0 Introduction

Professional stock photographers and photo banks have elaborate organizational procedures that help them categorize their photographic assets. These organizational structures exist because the photographers expend time categorizing and tagging their photographs. The better the organization of the imagery, the easier it is to find photographs for clients. Similarly, digital asset-management software relies heavily on the user to provide text annotation for each asset. Although professionals know the value of keyword tagging, it can be a time-consuming task. Non-professionals rarely take this time to annotate their images, and as a result, the knowledge represented by those images is difficult to recover. What is needed is a system that facilitates or partially automates the categorization process, making the media more retrievable, while reducing the cost of annotation for users.

This thesis proposes a new categorization technique called “Interaction Harvesting,” a system in which documents are characterized in part by how the users interact with their documents through an application interface. In the simplest case, the system observes users' actions as they manipulate media files. These user actions are recorded in a database, where they are combined with actions of other users of the system, and analyzed by software agents. A specific application is presented which demonstrates the Interaction Harvesting technique. In this application, users create a spatial narrative with images from their photo library. By observing the curatorial decision-making process of the user as she positions documents in the landscape, the system is able to enhance its own representation how these documents are related. A detailed description of how these observations are used to establish similarities between documents is given in the proposed approach.

3.0 Background

Previous research in document classification and search has principally been focused on three major areas: computational analysis of the document contents, classification based on document context, and classification based on user annotation. Research focusing on document classification by observing users' interactions with the documents is less common.

In the digital image domain, visual content analysis has been the most widely studied approach to document search and classification. Many systems attempt to index image files by attribute histograms, such as color, or texture distributions. Images can then be categorized and grouped based on the similarities of these attribute histograms. Searches can be performed to find the closest matching histogram (Pentland 1994). Although perceptually similar images can be grouped using such systems, the resultant groups rarely share subject or content items. Additional research has been addressing regional pattern matching and contour matching approaches to visual content analysis (Jitendra 2001, Rubner 1998).

Completely automatic annotation methods have been obtained from approaches that focus on document context. In these approaches, a document is characterized by factors such as the document creation time, the document author, and documents that appear in the same location. An example of such a system is Google Image Search, which uses the text of the web page in which an image file was located, as a source for keyword search terms. More subtle context data is used by AutoAlbum, software that attempts to auto-correlate document creation time with image analysis signatures (Platt 2002).

Interaction Harvesting for Document Retrieval

Systems have been designed that attempt to reduce the effort required for manual image annotation. Examples range from systems that simplify the annotation process by enabling drag and drop keyword editing to systems that use user queries and user-relevance feedback to assist in the keyword generation process (Shneiderman 2000, Wenyin 2001). Aria is a system that extracts possible keyword annotations both from the context surrounding an image and from monitoring the keystrokes of the user (Lieberman 2001). Additionally image retrieval systems have been developed which seek to broaden the keyword terms used in annotation. In one such extension, WordNet and OpenMind databases are used to expand keyword set membership (Fellbaum 1998, Singh 2002, Dai 2003).

4.0 Proposed Approach

Treehouse Studio is an online community with a suite of design tools and a database for storage of media files and metadata. This system is well-suited for research involving document retrieval, because it uses a central, relational database to store file contents and file attributes, and therefore can quickly read and write attributes such as creation time, keywords, and location. Additionally, Treehouse Studio is a multi-user system already equipped with a document-retrieval framework. Within this framework, applications can be developed to explore document retrieval using Interaction Harvesting document classification.

The first example, of an Interaction Harvesting application called “Narrative Landscape”, has already been started. This application is a presentation application similar to a slide-show editor. Users assemble groups of media files and arrange them in specific order, and on specific locations of the screen as they create their presentation. As the author composes and arranges these media files, the user interface communicates these actions to

Interaction Harvesting for Document Retrieval

the Treehouse database, where the interactions are recorded in real time. Information about when documents were added, where they were placed, etc., is stored. Document clustering happens as a result of correlating one or more of these gesture attributes amongst the documents. Spatial clusters, for example, are groups of documents that have been placed near each other.

Document Clusters are assumed to have similar content, or to be semantically relevant to each other. These clusters exert influence on each other in the form of fuzzy tags. Fuzzy tags are weighted keywords similar to vector-space keywords (Salton 1983). As documents are positioned near each other, they add diluted versions of their keyword tags to their neighbors. The dilution of the significance of the tag is proportional to the distance of the neighbor. For example assume a photograph of a child blowing out candles on a cake has amongst its many tags the following: ('birthday', 1). Any picture which is placed close to this one would inherit a tag ('birthday', W) where W is the weight of the keyword, computed as the weight of the source keyword divided by the square of the distance of the two documents (W/d^2). This is one possible approach to observing the curatorial process of software users to build document knowledge. Time permitting Interaction Harvesting techniques will also be applied to an existing drawing application.

4.1 Assessment and Evaluation

It will be difficult to accurately assess the success of Interaction Harvesting techniques quantifiably. Because the research is being conducted in a closed system, it will be impossible to compare results with that of Google Image search, for example. It may be useful to compare the tool with existing search facilities inside the system. Treehouse studio has an active user community; a critique of the document retrieval processes by the community can be used as part of an evaluation. Additional Evaluation methods will be

Interaction Harvesting for Document Retrieval

sought during the course of research; all results will be reported as part of the thesis.

5.0 Summary

Multimedia document retrieval remains a difficult problem. So far the best approach to organizing multimedia documents has been keyword annotation. Because of the general reluctance by authors to provide keyword annotations to their materials, a system is required to help facilitate and streamline the annotation process. Our proposed method of Interaction Harvesting involves loose categorization of images based on how users work with documents. Once collections of documents have been harvested, similar documents benefit from each others' annotations.

6.0 Deliverables

1. A summary of document retrieval techniques, and a presentation of current state-of-the-art systems.
2. A software application that implements Interaction Harvesting and fuzzy tagging techniques to help individuals organize their digital image libraries.
3. An analysis and critique of the Interaction Harvesting tools based in part on feedback from the Treehouse studio community.

7.0 Schedule

- January:** Collecting references to findings, practices, and papers concerning image organization, classification, and searching techniques.
- February:** Write software. write summary of previous research and background.
- March:** Begin assessment and evaluation process, write up critique of tools from online user community. Continue writing.
- April & May:** Finish writing.

8.0 Bibliography

- Dai, James J. *Visual Intelligence for Online Communities, Common-Sense Image Retrieval by Query Expansion* Thesis (S.M.), Massachusetts Institute of Technology, 2003
- Fellbaum, Christiane. *WordNet An Electronic Lexical Database* Cambridge, Ma: MIT Press, 1999
- Jitendra Malik, Serge Belongie, Thomas Leung, Jianbo Shi, "Contour and Texture Analysis for Image Segmentation", *International Journal of Computer Vision*, Volume 43, Issue 1, Pages 7 – 27, June 2001
- Lieberman, H., and Rosenzweig E. And Singh P., "Aria: An Agent for Annotating and Retrieving Images" *IEEE Computer*, pages 57 – 61, July 2001
- Lienhart R., and Hartmann, A. "Classifying images on the web automatically," *Journal of Electronic Imaging*, vol. 4, no. 11, pp. 445-454, October 2002
- Pentland, Alex. P., and Picard, Ros, W. and Scarloff, S., "Photobook: tools for content-based manipulation of image databases," *Technical Report*, Volume 255, MIT, Media Lab., 1993
- Platt, John C. and Czerwinski, Mary and Field, Brent A., "PhotoTOC: Automatic Clustering for Browsing Personal Photographs", *In Proc. of Fourth IEEE Pacific Rim Conference on Multimedia*, 2003
- Rubner, Yossi and Tomasi, Carlo and Guibas, Leonidas J., "The Earth Mover's Distance as a Metric for Image Retrieval" *Stanford University*, 1998
<http://vision.stanford.edu/public/publication/rubner/rubnerTr98.pdf>
- Salton, G. and McGill, M.J . *Introduction to Modern Information Retrieval* NewYork, NY: McGraw-Hill Publishing Company, 1983
- Shneiderman, Ben. "Direct Annotation: A Drag-and-Drop Strategy for Labeling Photos", *Proceedings of the International Conference on Information Visualization*, p.88, July 19-21, 2000
- Singh, P. "The Public Acquisition of Common-Sense Knowledge." *In Proc of AAAI Spring Symposium: Acquiring (and Using) Linguistic (and World) Knowledge for Information Access*. Palo Alto, CA, AAAI, 2002
- Wenyin, S. Dumais, Y. Sun, and H. Zhang, M. Czerwinski, and B. Field. "Semi-automatic image annotation." *In Proc. of Interact: Conference on HCI*, pages 326--333, July 2001
<http://citeseer.ist.psu.edu/wenyin01semiautomatic.html>