INTEGRATED REGION-BASED IMAGE RETRIEVAL

James Z. Wang

KLUWER ACADEMIC PUBLISHERS

INTEGRATED REGION-BASED IMAGE RETRIEVAL

by

James Z. Wang
The Pennsylvania State University

KLUWER ACADEMIC PUBLISHERS
Boston / Dordrecht / London

Distributors for North, Central and South America:
Kluwer Academic Publishers
101 Philip Drive
Assinippi Park
Norwell, Massachusetts 02061 USA
Telephone (781) 871-6600
Fax (781) 681-9045
E-Mail <kluwer@wkap.com>

Distributors for all other countries:
Kluwer Academic Publishers Group
Distribution Centre
Post Office Box 322
3300 AH Dordrecht, THE NETHERLANDS
Telephone 31 78 6392 392
Fax 31 78 6392 254
E-Mail <services@wkap.nl>


Electronic Services <http://www.wkap.nl>

Library of Congress Cataloging-in-Publication Data

Wang, James Z., 1972-
Integrated region-based image retrieval / James Z. Wang.
p. cm. -- (The Kluwer international series on information retrieval ; 11)
Includes bibliographical references.
ISBN 0-7923-7350-2 (alk. paper)
1. Optical storage devices. 2. Image processing--Digital techniques. 3. Database management. I. Title. II. Series.
TA1635.W37 2001
006.4'2--dc21 2001020367

Printed on acid-free paper.

Printed in the United States of America

The Publisher offers discounts on this book for course use and bulk purchases. For further information, send email to <scott.delman@wkap.com>.

INTEGRATED REGION-BASED IMAGE RETRIEVAL

THE KLUWER INTERNATIONAL SERIES ON INFORMATION RETRIEVAL

Series Editor

W. Bruce Croft
University of Massachusetts, Amherst

Also in the Series:

MULTIMEDIA INFORMATION RETRIEVAL: ***Content-Based Information Retrieval from Large Text and Audio Databases,*** *by Peter Schäuble;* ISBN: 0-7923-9899-8

INFORMATION RETRIEVAL SYSTEMS: ***Theory and Implementation,*** *by Gerald Kowalski;* ISBN: 0-7923-9926-9

CROSS-LANGUAGE INFORMATION RETRIEVAL, *edited by Gregory Grefenstette;* ISBN: 0-7923-8122-X

TEXT RETRIEVAL AND FILTERING: ***Analytic Models of Performance,*** *by Robert M. Losee;* ISBN: 0-7923-8177-7

INFORMATION RETRIEVAL: UNCERTAINTY AND LOGICS: ***Advanced Models for the Representation and Retrieval of Information,*** *by Fabio Crestani, Mounia Lalmas, and Cornelis Joost van Rijsbergen;* ISBN: 0-7923-8302-8

DOCUMENT COMPUTING: ***Technologies for Managing Electronic Document Collections,*** *by Ross Wilkinson, Timothy Arnold-Moore, Michael Fuller, Ron Sacks-Davis, James Thom, and Justin Zobel;* ISBN: 0-7923-8357-5

AUTOMATIC INDEXING AND ABSTRACTING OF DOCUMENT TEXTS, *by Marie-Francine Moens*; ISBN 0-7923-7793-1

ADVANCES IN INFORMATIONAL RETRIEVAL: ***Recent Research from the Center for Intelligent Information Retrieval,*** *by W. Bruce Croft*; ISBN 0-7923-7812-1

INFORMATION RETRIEVAL SYSTEMS: ***Theory and Implementation*****, Second Edition,** *by Gerald J. Kowalski and Mark T. Maybury;* ISBN: 0-7923-7924-1

PERSPECTIVES ON CONTENT-BASED MULTIMEDIA SYSTEMS, *by Jian Kang Wu; Mohan S. Kankanhalli;Joo-Hwee Lim;Dezhong Hong;* ISBN: 0-7923-7944-6

MINING THE WORLD WIDE WEB: ***An Information Search Approach,*** *by George Chang, Marcus J. Healey, James A. M. McHugh, Jason T. L. Wang;* ISBN: 0-7923-7349-9

INTEGRATED REGION-BASED IMAGE RETRIEVAL, *by James Z. Wang;* ISBN: 0-7923-7350-2

To my parents

Preface

Content-based image retrieval is the set of techniques for retrieving relevant images from an image database on the basis of automatically-derived image features. The need for efficient content-based image retrieval has increased tremendously in many application areas such as biomedicine, the military, commerce, education, and Web image classification and searching. In the biomedical domain, content-based image retrieval can be used in patient digital libraries, clinical diagnosis, searching of 2-D electrophoresis gels, and pathology slides.

I started my work on content-based image retrieval in 1995 when I was with Stanford University. The project was initiated by the Stanford University Libraries and later funded by a research grant from the National Science Foundation. The goal was to design and implement a computer system capable of indexing and retrieving large collections of digitized multimedia data available in the libraries based on the media contents. At the time, it seemed reasonable to me that I should discover the solution to the image retrieval problem during the project. Experience has certainly demonstrated how far we are as yet from solving this basic problem.

CBIR for general-purpose image databases is a highly challenging problem because of the large size of the database, the difficulty of understanding images, both by people and computers, the difficulty of formulating a query, and the problem of evaluating the results. The objectives of this book are to introduce the fundamental problems, to review a collection of selected and well-tested methods, and to introduce our work in this rapidly developing research field.

We designed a content-based image retrieval system with wavelet-based feature extraction, semantics classification, and integrated region matching (IRM). An image in the database, or a portion of an image, is represented by a set of regions, roughly corresponding to ob-

jects, which are characterized by color, texture, shape, and location. The system classifies images into semantic categories, such as textured-nontextured, objectionable-benign, or graph-photograph. The categorization enhances retrieval by permitting semantically-adaptive searching methods and narrowing down the searching range in a database. A measure for the overall similarity between images is developed as a region-matching scheme that integrates properties of all the regions in the images. Compared with retrieval based on individual regions, the overall similarity approach reduces the adverse effect of inaccurate segmentation, helps to clarify the semantics of a particular region, and enables a *simple* querying interface for region-based image retrieval systems.

We built an experimental image retrieval system, the SIMPLIcity (Semantics-sensitive Integrated Matching for Picture LIbraries) system, to validate these methods on various image databases, including a database of about 200,000 general-purpose images and a database of more than 70,000 pathology images. We have shown that our methods perform much better and much faster than existing methods. The system is exceptionally robust to image alterations such as intensity variation, sharpness variation, intentional distortions, cropping, shifting, and rotation. These features are important to biomedical image databases because visual features in the query image are not exactly the same as the visual features in the images in the database. The work has also been applied to the classification of on-line images and web sites.

JAMES Z. WANG

Acknowledgments

This work would not have been possible without the guidance and advice of my dissertation advisor Gio Wiederhold. He has led me to new areas of research, pointed me to interesting research problems, and offered me substantial encouragement. Gio has cultivated a creative atmosphere and provided me with unconditional support.

I would like to thank Dennis A. Hejhal for introducing me to the excitement of conducting scientific research, and for being everlastingly supportive during the past nine years. I would like to thank Martin A. Fischler and Oscar Firschein for inspiring me with the fascinating field of image understanding, and encouraging me. Discussions with Desmond Chan, Shih-Fu Chang, Eldar Giladi, Robert M. Gray, Yoshi Hara, Kyoji Hirata, Xiaoming Huo, Yvan Leclerc, Quang-Tuan Luong, Thomas P. Minka, Wayne Niblack, Richard Olshen, Dragutin Petkovic, Donald Regula, Xin Wei Sha, Michael Walker, and Tong Zhang have been very helpful in different stages of my research. Special thanks goes to Russ B. Altman, W. Bruce Croft, Oscar Firschein, Hector Garcia-Molina, Rosalind W. Picard, Mu-Tao Wang, Stephen T.C. Wong, and anonymous reviewers, who provided numerous constructive comments to the manuscript and its related publications.

I would also like to thank my friends in the Stanford Database Group, the Stanford Biomedical Informatics Group, the Stanford Mathematics Department, the Perception Research Group at SRI International, the QBIC Group at the IBM Almaden Research Center, and the School of Information Sciences and Technology and the Department of Computer Science and Engineering at the Pennsylvania State University for their generous help.

My wife Jia Li is the most essential contributor to my success and my well-being. Her talents and professional expertise in statistics, information theory, and image processing have enlightened me numerous times

throughout my research. We have coauthored several publications and experimental systems.

My work was funded primarily by a research grant from the National Science Foundation's Digital Libraries initiative and a research fund from the Stanford University Libraries. I have also received support from IBM Almaden Research Center, NEC Research Lab, SRI International, Stanford Computer Science Department, Stanford Mathematics Department, Stanford Biomedical Informatics, The Pennsylvania State University, and the PNC Foundation. I am truly grateful for the support.

Finally, I acknowledge the Institute for Electrical and Electronic Engineers (IEEE) for their generous permission to use material published in their *Transactions* and conference proceedings in this book as detailed in specific citations in the text. I would like to thank Scott E. Delman and Melissa Fearon, the editor and editorial assistant at Kluwer Academic Publishers, for making the publication of this book go smoothly.

Contents

Chapter 1

INTRODUCTION

Make everything as simple as possible, but not simpler.
—— Albert Einstein (1879-1955)

The need for efficient content-based image retrieval has increased tremendously in many application areas such as biomedicine, crime prevention, the military, commerce, culture, education, and entertainment. Content-based image retrieval is also crucial to Web image classification and searching.

With the steady growth of computer power, rapidly declining cost of storage, and ever-increasing access to the Internet, digital acquisition of information has become increasingly popular in recent years. Digital information is preferable to analog formats because of convenient sharing and distribution properties. This trend has motivated research in image databases, which were nearly ignored by traditional computer systems because of the large amount of data required to represent images and the difficulty of automatically analyzing images. Currently, storage is less of an issue since huge storage capacities are available at low cost. However, effective indexing[1] and searching of large-scale image databases remain as challenges for computer systems.

The automatic derivation of semantically-meaningful information from the content of an image is the focus of interest for most research on image databases. The image "semantics", i.e., the meanings of an

[1] Here, the term *indexing* means the combination of both feature extraction and feature space indexing.

image, has several levels. From the lowest to the highest, these levels can be roughly categorized as follows:

1 Semantic types (e.g., MRI, X-ray, landscape photograph, clip art)

2 Object composition (e.g., a lesion in the left brain, a bike and a car parked on a beach, a sunset scene)

3 Abstract semantics (e.g., people fighting, happy person, objectionable photograph)

4 Detailed semantics (e.g., a detailed description of a given picture)

Image retrieval is defined as the retrieval of semantically-relevant images from a database of images. In the following sections (Section 1 and Section 2), we discuss text-based image retrieval and content-based image retrieval.

1. TEXT-BASED IMAGE RETRIEVAL

In current commercial image databases, the prevalent retrieval techniques involve human-supplied text annotations to describe image semantics. These text annotations are then used as the basis for searching, using mature text search algorithms developed in the database management and information retrieval communities [11, 42, 112]. It is often easier to design and implement an image search engine based on keywords (e.g., classification codes) or full-text descriptions (e.g., surrounding text) than on the image content. The query processing of such search engines is typically very fast due to the available efficient database management technology. The text-based image retrieval approach is accepted for high-value pictures such as museum pictures.

Recently, researchers have proposed community-wide *social* entry of descriptive text to facilitate subsequent retrieval. This approach is feasible with the widely-available Internet. However, it is limited to image sets that are of wide interest and stable.

There are many problems in using text-based approach alone. For example, different people may supply different textual annotations for the same image. This makes it extremely difficult to answer user queries reliably. Furthermore, entering textual annotations manually is excessively expensive for large-scale image databases (e.g., space-based observations).

2. CONTENT-BASED IMAGE RETRIEVAL

Content-based image retrieval (CBIR) is the set of techniques for retrieving relevant images from an image database on the basis of automatically-derived image features.

CBIR functions differently from text-based image retrieval. Features describing image content, such as color histogram, color layout, texture [72], shape and object composition, are computed for both images in the database and query images. These features are then used to select the images that are most similar to the query. High-level semantic features, such as the types of objects in the images and the purpose of the images are extremely difficult to extract. Derivation of semantically-meaningful features remains a great challenge.

CBIR is also important for video indexing and retrieval. In a typical video retrieval system, long video sequences are broken up into separate clips and key frames are extracted to represent the content of the clips. Searching of relevant clips is done by combining CBIR, speech recognition, and searching for specific movements of the objects in the shots [122, 17]. In this book, we focus on content-based *image* retrieval.

3. APPLICATIONS OF CBIR

Content-based image retrieval (CBIR) has applications in various domains in many areas of our society.

3.1. BIOMEDICAL APPLICATIONS

CBIR is critical in developing patient care digital libraries. McKeown, Chang, Cimino, and Hripcsak [50] of Columbia University plan[2] to develop a personalized search and summarization system over multimedia information within a healthcare setting. Both patients and healthcare providers are targeted users of the system. Efficient CBIR is the most important core technology within such systems. With the help of such a mediator [152, 153], healthcare consumers and providers can quickly access a wide range of online resources: patients and their families can find information about their personal situation, and clinicians can find clinically relevant information for individual patients. A similar research effort is the Stanford SHINE project [58].

CBIR can be applied to clinical diagnosis and decision making. Currently, more and more hospitals and radiology departments are equipped with Picture Archive and Communications Systems (PACS) [154]. Be-

[2]Recently funded by a joint National Science Foundation and National Institute of Health grant.