Semantic-Based Image Retrieval

基于语义的图像检索

Ying Liu （刘 颖）

科 学 出 版 社

Semantic-Based Image Retrieval

基于语义的图像检索

Ying Liu (刘颖)

Science Press
Beijing

Responsible Editor: Xiangping Yang and Wuhan Song

Published by Science Press
16 Donghuangchenggen North Street
Beijing 100717, P. R. China

Printed in Beijing

ISBN 978-7-03-049490-0

Preface

In an attempt to improve the performance of traditional content-based image retrieval (CBIR) systems, research effort is needed not only to find salient low-level image features to represent images, but also to reduce the 'semantic gap' between low-level image features and high-level human semantics. Such systems with high level semantics not only improve the accuracy in image retrieval but also support query-by-keywords which is more convenient for users.

The aim of this book is to discuss CBIR with high-level semantics focusing on low-level feature extraction and image semantic learning. A general introduction to this area is given in Chapter 1. Then, Chapter 2 discusses the key techniques in semantic-based image retrieval including image segmentation, low-level image feature extraction, image similarity measure, and high-level semantic learning. Recent and traditional technical achievements are reviewed and six categories of the existing image semantic learning techniques are identified. This form the second chapter of the book.

Then, Chapter 3～5 describe in details the key components in semantic-based image retrieval. A proposed region-based image retrieval algorithm with high level semantics is used for illustration purpose with experimental results provided. The algorithm first segments an image into regions of interest using image segmentation technique, and then extracts low-level features of each region including color and texture. In Chapter 3, a 'dominant color' is defined as the color feature of a region, which is converted to semantic color name using a color naming method. In Chapter 4, an effective method for texture feature extraction from arbitrary-shaped regions is presented. This method, named 'POCS-ER', utilizes projection-onto-convex-sets (POCS) theory to obtain a set of coefficients that best describe a region, from which the texture feature of the region is calculated. From low-level region features, high-level concepts are obtained through a semantic learning method DT-ST which combines decision tree (DT) and image semantic templates (ST) as explained in Chapter 5.

In Chapter 6, the proposed semantic learning algorithm DT-ST is applied to Web image search to filter out those images which are semantically irrelevant to the query from the set of images returned by existing text-based search engines, thus to improve web image retrieval performance. At last, Chapter 7 concludes the book and suggests a few future research directions in semantic-based image retrieval.

The content of this book is suitable for postgraduate students in literature survey as well as in understanding how to design and implement image retrieval system. In addition, the book is useful for researchers in the field of image retrieval as reference reading.

List of Abbreviations

CBIR: Content-Based Image Retrieval

CSMask: Coefficient Selection Mask

DCT: Discrete Cosine Transform

D-EM: Discriminant-EM

Dm: Dominant Color

Dm+: Multiple Dominant Colors

DT: Decision Tree

DT-ST: Integrating Decision Tree with Semantic Templates for image concepts learning

EM: Expected Maximization

EMD: Earth Mover's Distance

ER: Extended Rectangular area

FF: False Filtering

Gabor-ER: Gabor texture featuring from Extended Rectangle of a region

Gabor-IR: Gabor texture featuring from Inner Rectangle of a region

GRBF: Gaussian Radial Basis Function

HSL: Hue, Saturation, Lightness (Luminance)

HSV: Hue, Saturation, Value

HSV-ave: Average color of a region in HSV space

IR: Inner Rectangle

KMCC: K-means with Connectivity Constraints

LPC: Locality Preserving Clustering

MAP: Maximum-a-posteriori

ML: Maximum-Likelihood

MP: Mean-intensity Padding

MR: Mirroring

NCut: Normalized Cut
PCK-means: Pair-wise Constraints K-means
POCS: Projection Onto Convex Sets
POCS-ER: Texture feature extraction from ER of a region based on POCS theory
Pr: Precision
QPM: Query-Point-Movement
RBF: Radial Basis Function
RBIR: Region-Based Image Retrieval
Re: Recall
RGB: Red, Green, Blue color space
RGB-ave: Average color of a region in RGB space
ST: Semantic Template
SVT: Semantic Visual Template
ZR: Zero Padding
SPM: Spatial Pyramid Matching

Contents

Chapter 1
Introduction

1.1 Background

With the development of the Internet and the availability of image capturing devices such as digital cameras and image scanners, the size of digital image collection is increasing rapidly. Efficient image searching, browsing and retrieval tools are required by users from various domains including: remote sensing, fashion, crime prevention, publishing, medicine, and architecture, etc. For this purpose, many general-purpose image retrieval systems have been developed under two broad frameworks: text-based and content-based.

The text-based approach can be tracked back to 1970's. In such systems, the images are manually annotated by text descriptors, which are then used by a Database Management System (DBMS) to perform image retrieval. There are two disadvantages of this approach. The first is that a considerable level of human labour is required for manual annotation. The second is the impreciseness and incompleteness of annotation due to the subjectivity of human perception [1, 2].

To overcome these problems in text-based retrieval, content-based image retrieval (CBIR) was introduced in the early 1980's. In CBIR, images are indexed by their own visual contents such as color, texture and shape. A pioneering work was published by Chang in 1984 [3], in which the author presented a picture indexing and abstraction approach for pictorial database retrieval. In the past decade, a number of commercial products and experimental prototype systems have been developed, such as QBIC [4], Photobook [5], Virage [6], VisualSEEK [7], Netra [8] and Simplicity [9]. Comprehensive surveys in text-based image retrieval and content-based image retrieval can be found in [10, 11].

1.1.1 The 'Semantic Gap'

The fundamental difference between content-based and text-based retrieval systems is that the human interaction is an indispensable part of the latter system. Humans tend to use high-level features (concepts), such as keywords and text descriptors, to interpret images and measure their similarity, whereas the features extracted automatically using computer vision techniques are mostly low level features (color, texture, shape, spatial layout, etc.) [2].

Though many sophisticated algorithms have been designed to describe color, shape and texture features; these algorithms can't adequately model image semantics and have many limitations when dealing with broad content image databases [12]. Extensive experiments on CBIR systems show that low-level contents often fail to describe the high level semantic concepts in user's mind [13]. Therefore, the performance of CBIR is still far from user's expectations.

In [1], John Eakins mentioned three levels of queries in CBIR:

Level 1: Retrieval by primitive features such as color, texture, shape or the spatial location of image elements. Typical query is query by example such as: 'find pictures like this'.

Level 2: Retrieval of objects of given type identified by derived features, with some degree of logical inference. For example, 'find a picture of a flower'.

Level 3: Retrieval by abstract attributes, involving a significant amount of high-level reasoning about the purpose of the objects or scenes depicted. This includes retrieval of named events, and pictures with emotional or religious significance, etc. A typical query example is: 'find pictures of a joyful crowd'.

Level 2 and Level 3 together are referred to as semantic image retrieval, and the gap between level 1 and 2 as the semantic gap [1]. More specifically, the discrepancy between the limited descriptive power of low-level image features and the richness of user semantics, is referred to as the 'semantic gap' [14, 15].

Users in Level 1 retrieval are usually required to submit an example image or sketch as query. But what if the user does not have an example image at hand? Thus, semantic image retrieval is more convenient for users as it supports query by keywords or textual descriptions.

Therefore, to further improve retrieval accuracy and to support query by high-level concepts, a CBIR system should bridge the 'semantic gap' between low-level image features and human semantics [13, 15].

1.1.2 Query by Keywords

An important advantage of semantic-based image retrieval is that it allows users to perform query by keywords which is more convenient to users. This book describes a special application scenario: Web image search.

Due to the explosive growth of the World-Wide Web (the Web), nowadays we can easily access a huge amount of images from the Web. The Web can be viewed as a large, unstructured image database and Web image search has been actively explored and developed in academic as well as commercial areas for over a decade [16-18]. The advantage lies in that Web images are usually accompanied by rich textual descriptions included in the HTML file. Most current Web image systems are text-based that try to explore textual information associated with an image, such as filename, caption and surrounding text to represent images [16,19-21]. To measure the similarity between an image and user query, these systems estimate the probability that the textual information corresponding to the image is relevant to the query.

Although textual information is still the prevailing choice to index Web images, it can be incomplete or ambiguous in describing the actual image content. For example, filenames may be misleading or surrounding text might not describe the content of an image due to page layout considerations [18]. To address this issue, some systems have been designed to boost the performance of Web image search by utilizing other information sources such as visual image features, link structure and user feedbacks [22-27].

1.2 Objectives

From the above discussion, we realize that in order to further improve the performance of content-based image retrieval, effort has to be made not only in extracting salient low-level image features, but also in reducing the 'semantic gap' between low-level image features and high-level concepts.

For CBIR, image features can be extracted either from the entire image or from image regions. It has been found that representation of images at region level is more close to human perception system [28] and users are usually more interested in specific regions rather than the entire image. Figure 1.1 gives two example images with different interested regions. In the left image, the region 'flower' represents the semantic meaning of the image. In the right image, the 'ants' region might be what users are interested in.

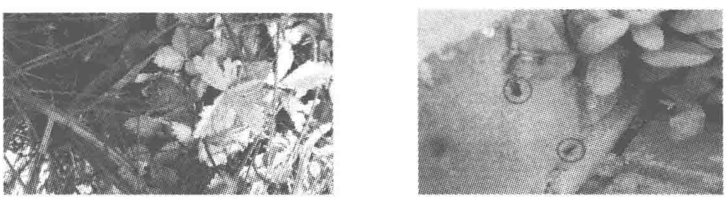

Figure 1.1 Region of Interest in Images

For the reasons described above, this book studies on region-based image retrieval (RBIR) focusing on the following two aspects: extracting salient low-level region features and deriving high-level semantic concepts from such region features. A semantic-based RBIR system is described with the intention to improve the performance of image search.

Web image search is an important application scenario for image retrieval techniques, with the advantage that additional textual information is available to facilitate image search. However, such textual description of Web images could be incorrect or misleading, resulting in irrelevant images being returned for the query. Therefore, the proposed RBIR system is used to improve the performance of pure text-based Web image search.

In summary, the main objectives of this book are:

To review region-based image retrieval with high-level semantics;

To describe salient low-level region feature extraction;

To study how to derive high-level concepts from low-level region features;

To apply RBIR to Web image search.

1.3 Contributions of this Book

This book focuses on region-based image retrieval (RBIR) with high-level semantics. Our key contributions are summarized in the following.

1.3.1 Identifying Existing Semantic Learning Techniques

Through an extensive survey of many works in this area, we identify six major categories of the existing techniques that attempt to narrow down the 'semantic gap' between low-level image features and high-level semantics. These are: ① Using object ontology to define high-level concepts; ② Using supervised or unsupervised machine learning tools to associate low-level features with query concepts; ③ Introducing relevance feedback into retrieval loop for continuous learning of users' intention; ④ Generating semantic template based on low-level image features to support high-level image retrieval; ⑤ Making use of multiple information sources (such as the visual content of images and the textual information on the Web) for Web image retrieval; ⑥ Deep learning for classification of large scale image database with large vocabulary size.

1.3.2 Designing Effective Feature Extraction Methods for Arbitrary-Shaped Regions

Although many algorithms are available for low-level feature extraction, comparatively less work has been done for RBIR in terms of feature extraction from arbitrary-shaped regions.

The color features used in most existing systems such as color-covariance matrix, color histogram and color moments [28-30], though efficient in describing colors, are not directly related to high-level semantics. We study various methods to define color features for arbitrary shaped regions and propose to use the HSV space 'dominant color' as region color feature. In addition, a color naming method is designed to convert 'dominant color' of a region into semantic color name.

Although much research has been done in texture feature extraction from rectangular images, feature extraction from arbitrary-shaped regions has not been well addressed. In this book, we describe an effective algorithm for texture feature extraction from arbitrary-shaped regions. This algorithm first extends an arbitrary-shaped region to a rectangular area. Then, based on the projection-onto-convex-sets (POCS) theory, it finds a set of transform coefficients that best describes the region, from which texture feature of the region can be extracted.

Experimental results on real-world images are provided to verify the effectiveness of the proposed color naming and texture feature extraction methods.

1.3.3 High-Level Concept Learning Using Decision Tree

Decision tree has great potential in image semantic learning due to its simplicity in implementation and its robustness to incomplete and noisy input data. Decision tree learning naturally requires the input attributes to be nominal (discrete). However, proper discretization of image features is a difficult task. Some generally designed algorithms have been proposed to discretize continuous attributes, but they usually do not provide meaningful quantization of image feature space.

In this book, we integrate semantic templates (STs) with decision tree (DT) induction in order to avoid the difficult image feature discretization problem. The method, named DT-ST, first generates a set of semantic templates from the low-level features of a collection of sample regions. These STs are then used to convert continuous-valued color and texture features to discrete color and texture labels. Moreover, a new discrete attribute is formed by combining color and texture labels, in order to classify those concepts which require both color and texture for their accurate representation. These discrete features are used as the input attributes for decision tree induction. In addition, we propose a hybrid of pre-pruning and post-pruning methods to simplify the tree in order to prevent tree fragmentation and minimize the risk of misclassifications due to noisy data samples. Using DT-ST, a set of decision rules are derived to associate low-level region features with high-level image concepts. Experimental results confirm the effectiveness of DT-ST in image semantic learning and improved classification performance as compared to other tree based algorithms.

1.3.4 Applying RBIR with Semantics to Web Image Search

Web image search is a special application scenario for image retrieval techniques. The additional textual information on the Web facilitates effective image retrieval. However, this information might be ambiguous and incomplete, resulting in many irrelevant images returned by text-based search engines. In this work, we present an algorithm named False Filtering (FF) that applies the RBIR system with semantics we developed to Web image search. Based on the high-level concepts obtained using semantic learning, FF filters out those images which are semantically irrelevant to the query from the result list returned by text-based search engines. A few practical ways are suggested to integrate FF with existing text-based Web image search engines. Experimental results on Google image search show that the precision of Web image search is significantly improved by using FF.

1.4 Organization of the Book

The rest of this book is organized as follows:

Chapter 2 provides a comprehensive survey of the technical achievements in semantic-based image retrieval. Many recent techniques are included in this survey covering different aspects of the research in this area that include: low-level feature extraction, similarity measure, high-level semantic concept learning. Some other related issues such as image test-bed and retrieval performance evaluation are also discussed. We analyze the problems in low-level feature extraction from arbitrary-shaped regions in RBIR systems. In addition, we identify five major categories of the state-of-the-art techniques in reducing the 'semantic gap'.

In Chapter 3, an RBIR system based on semantic color names is proposed. We propose to use the 'dominant color' in HSV space to describe region perceptual color. In order to support keyword (color name) based queries, a color naming method is devised to convert the dominant color into semantic color names.

The results in Chapter 3 show that color alone can not well represent images. Therefore, texture feature has to be included for more accurate image description. In Chapter 4, based on the theory of projection onto convex sets (POCS), we propose an effective texture feature extraction method for arbitrary-shaped regions. Experimental results are provided to demonstrate the superior performance of the proposed method over other contemporary texture feature extraction algorithms.

In Chapter 5, we describe a decision tree based semantic learning algorithm. To avert the difficult image feature discretization problem in decision tree learning, this algorithm makes use of the semantic templates defined for the concepts in our database to convert low-level color(texture) features to color(texture) labels. The discrete color(texture) labels are used as input attributes to induce a decision tree from which decision rules are formulated to classify query regions. The tree induction technique combines pre-pruning and post-pruning methods for tree simplification in order to solve the noise and tree fragmentation problems.