PREMIER REFERENCE SOURCE

Semantic Mining Technologies for Multimedia Databases

Dacheng Tao, Dong Xu, & Xuelong Li

Detailed Table of Contents

The authors present a face recognition scheme based on semantic features' extraction from faces and tensor subspace analysis. These semantic features consist of eyes and mouth, plus the region outlined by three weight centres of the edges of these features. The extracted features are compared over images in tensor subspace domain. Singular value decomposition is used to solve the eigenvalue problem and to project the geometrical properties to the face manifold. They also compare the performance of the proposed scheme with that of other established techniques, where the results demonstrate the superiority of the proposed method.

Computer-aided foliage image retrieval systems have the potential to dramatically speed up the process of plant species identification. Despite previous research, this problem remains challenging due to the large intra-class variability and inter-class similarity of leaves. This is particularly true when a large number of species are involved. In this chapter, the authors present a shape-based approach, the inner-distance shape context, as a robust and reliable solution. They show that this approach naturally captures part structures and is appropriate to the shape of leaves. Furthermore, they show that this approach can be easily extended to include texture information arising from the veins of leaves. They also describe a real electronic field guide system that uses our approach. The effectiveness of the proposed method is demonstrated in experiments on two leaf databases involving more than 100 species and 1000 leaves.

Motion estimation necessitates an appropriate choice of similarity function. Because generic similarity functions derived from simple assumptions are insufficient to model complex yet structured appearance variations in motion estimation, the authors propose to learn a discriminative similarity function to match

images under varying appearances by casting image matching into a binary classification problem. They use the LogitBoost algorithm to learn the classifier based on an annotated database that exemplifies the structured appearance variations: An image pair in correspondence is positive and an image pair out of correspondence is negative. To leverage the additional distance structure of negatives, they present a location-sensitive cascade training procedure that bootstraps negatives for later stages of the cascade from the regions closer to the positives, which enables viewing a large number of negatives and steering the training process to yield lower training and test errors. They also apply the learned similarity function to estimating the motion for the endocardial wall of left ventricle in echocardiography and to performing visual tracking. They obtain improved performances when comparing the learned similarity function with conventional ones.

Chapter VI

Relevance feedback is an effective approach to boost the performance of image retrieval. Labeling data is indispensable for relevance feedback, but it is also very tedious and time-consuming. How to alleviate users' burden of labeling has been a crucial problem in relevance feedback. In recent years, active learning approaches have attracted more and more attention, such as query learning, selective sampling, multi-view learning, etc. The well-known examples include Co-training, Co-testing, SVMactive, etc. In this literature, the authors will introduce some representative active learning methods in relevance feedback. Especially they will present a new active learning algorithm based on multi-view learning, named Co-SVM. In Co-SVM algorithm, color and texture are naturally considered as sufficient and uncorrelated views of an image. SVM classifier is learned in color and texture feature subspaces, respectively. Then the two classifiers are used to classify the unlabeled data. These unlabeled samples that disagree in the two classifiers are chose to label. The extensive experiments show that the proposed algorithm is beneficial to image retrieval.

Chapter VII

Visual data mining is a procedure aimed at a selection from a document's repository subsets of documents presenting certain classes of objects; the last may be characterized as classes of objects' similarity or, more generally, as classes of objects satisfying certain relationships. In this chapter attention will be focused on selection of visual documents representing objects belonging to similarity classes.

Section III
Semantic Analysis

The insufficiency of labeled training samples is a major obstacle in automatic semantic analysis of large scale image/video database. Semi-supervised learning, which attempts to learn from both labeled and unlabeled data, is a promising approach to tackle this problem. As a major family of semi-supervised learning, graph-based methods have attracted more and more recent research. In this chapter, a brief introduction is given on popular semi-supervised learning methods, especially the graph-based methods, as well as their applications in the area of image annotation, video annotation, and image retrieval. It is well known that the pair-wise similarity is an essential factor in graph propagation based semi-supervised learning methods. A novel graph-based semi-supervised learning method, named Structure-Sensitive Anisotropic Manifold Ranking (SSAniMR), is derived from a PDE based anisotropic diffusion framework. Instead of using Euclidean distance only, SSAniMR further takes local structural difference into account to more accurately measure pair-wise similarity. Finally some future directions of using semi-supervised learning to analyze the multimedia content are discussed.

With the explosive growth in the amount of video data and rapid advance in computing power, extensive research efforts have been devoted to content-based video analysis. In this chapter, they authors will give a broad discussion on this research area by covering different topics such as video structure analysis, object detection and tracking, event detection, visual attention analysis, etc. In the meantime, different video representation and indexing models are also presented.

Semantic mining is an essential part in knowledgebase and decision support systems where it enables the extraction of useful knowledge form available databases with the ultimate goal of supporting the decision making process. In process systems engineering, decisions are made throughout plant / process / product life cycles. The provision of smart semantic mining techniques will improve the decision making

process for all life cycle activities. In particular, safety and environmental related decisions are highly dependent on process internal and external conditions and dynamics with respect to equipment geometry and plant layout. This chapter discusses practical methods for semantic mining using systematic knowledge representation as integrated with process modeling and domain knowledge. POOM or plant/process object oriented modeling methodology is explained and used as a basis to implement semantic mining as applied on process systems engineering. Case studies are illustrated for biological process engineering, in particular MoFlo systems focusing on process safety and operation design support.

Efficient and effective techniques for managing and browsing large image databases are increasingly sought after. This chapter presents a simple yet efficient and effective approach to navigating image datasets. Based on the concept of a globe as visualisation and navigation medium, thumbnails are projected onto the surface of a sphere based on their colour. Navigation is performed by rotating and tilting the globe as well as zooming into an area of interest. Experiments based on a medium size image database demonstrate the usefulness of the presented approach.

Section IV
Multimedia Resource Annotation

Although important in practice, manual image annotation and retrieval has rarely been studied by means of formal modeling methods. In this paper, we propose a set of formal models to characterize the annotation times for two commonly-used manual annotation approaches, i.e., tagging and browsing. Based on the complementary properties of these models, we design new hybrid approaches, called frequency-based annotation and learning-based annotation, to improve the efficiency of manual image annotation as well as retrieval. Both our simulation and experimental results show that the proposed algorithms can achieve up to a 50% reduction in annotation time over baseline methods for manual image annotation, and produce significantly better annotation and retrieval results in the same amount of time.

This chapter introduces the application of active learning in video annotation. The insufficiency of training data is a major obstacle in learning-based video annotation. Active learning is a promising approach to dealing with this difficulty. It iteratively annotates a selected set of most informative samples, such that the obtained training set is more effective than that gathered randomly. We present a brief review of the typical active learning approaches. We categorize the sample selection strategies in these methods into five criteria, i.e., *risk reduction*, *uncertainty*, *positivity*, *density*, and *diversity*. In particular, we introduce the Support Vector Machine (SVM)-based active learning scheme which has been widely applied. Afterwards, we analyze the deficiency of the existing active learning methods for video annotation, i.e., in most of these methods the to-be-annotated concepts are treated equally without preference and only one modality is applied. To address these two issues, we introduce a multi-concept multi-modality active learning scheme. This scheme is able to better explore human labeling effort by considering both the learnabilities of different concepts and the potential of different modalities.

Chapter XIV

Although it has been studied for years by computer vision and machine learning communities, image annotation is still far from practical. In this paper, we propose a novel attempt of modeless image annotation, which investigates how effective a data-driven approach can be, and suggest annotating an uncaptioned image by mining its search results. We collected 2.4 million images with their surrounding texts from a few photo forum websites as our database to support this data-driven approach. The entire process contains three steps: 1) the search process to discover visually and semantically similar search results; 2) the mining process to discover salient terms from textual descriptions of the search results; and 3) the annotation rejection process to filter noisy terms yielded by step 2). To ensure real time annotation, two key techniques are leveraged – one is to map the high dimensional image visual features into hash codes, the other is to implement it as a distributed system, of which the search and mining processes are provided as Web services. As a typical result, the entire process finishes in less than 1 second. Since no training dataset is required, our proposed approach enables annotating with unlimited vocabulary, and is highly scalable and robust to outliers. Experimental results on real web images show the effectiveness and efficiency of the proposed algorithm.

Chapter XV

With the rapid growth of image collections, content-based image retrieval (CBIR) has been an active area of research with notable recent progress. However, automatic image retrieval by semantics still remains a challenging problem. In this chapter, we will describe two promising techniques towards semantic

image retrieval — semantic image classification and automatic image annotation. For each technique, four aspects are presented: task definition, image representation, computational models, and evaluation. Finally, we will give a brief discussion of their application in image retrieval.

Section V
Other Topics Related to Semantic Mining

With advances in computer technology and the World Wide Web there has been an explosion in the amount and complexity of multimedia data that are generated, stored, transmitted, analyzed, and accessed. In order to extract useful information from this huge amount of data, many content-based image retrieval (CBIR) systems have been developed in the last decade. A Typical CBIR system captures image features that represent image properties such as color, texture, or shape of objects in the query image and try to retrieve images from the database with similar features. Recent advances in CBIR systems include relevance feedback based interactive systems. The main advantage of CBIR systems with relevance feedback is that these systems take into account the gap between the high-level concepts and low-level features and subjectivity of human perception of visual content. CBIR systems with relevance feedback are more efficient than conventional CBIR systems; however, these systems depend on human interaction. In this chapter, we describe a new approach for image storage and retrieval called association-based image retrieval (ABIR). We try to mimic human memory. The human brain stores and retrieves images by association. We use a generalized bi-directional associative memory (GBAM) to store associations between feature vectors that represent images stored in the database. Section I introduces the reader to the CBIR system. In Section II, we present architecture for the ABIR system, Section III deals with preprocessing and feature extraction techniques, and Section IV presents various models of GBAM. In Section V, we present case studies.

Image retrieval and image compression have been typically pursued separately. Only little research has been done on a synthesis of the two by allowing image retrieval to be performed directly in the compressed domain of images without the need to uncompress them first. In this chapter we show that such compressed domain image retrieval can indeed be done and lead to effective and efficient retrieval performance. We introduce a novel compression algorithm – colour visual pattern image coding (CVPIC) – and present several retrieval algorithms that operate directly on compressed CVPIC data. Our experiments demonstrate that it is not only possible to realise such midstream content access, but also that the presented techniques outperform standard retrieval techniques such as colour histograms and colour correlograms.

Discovery of the multimedia resources on network is the focus of the many researches in post semantic web. The task of resources discovery can be automated by using agent. This chapter reviews the current most used technologies that facilitate the resource discovery process. The chapter also the presents the case study to present a fully functioning resource discovery system using mobile agents.

The rapid advances in multimedia capture, storage and communication technologies and capabilities have ushered an era of unprecedented growth of digital media content, in audio, visual, and synthetic forms, and both personal and commercially produced. How to manage these data to make them more accessible and searchable to users is a key challenge in current multimedia computing research. In this chapter, we discuss the problems and challenges in multimedia data management, and review the state of the art in data structures and algorithms for multimedia indexing, media feature space management and organization, and applications of these techniques in multimedia data management.