

摘 要

近几年来,随着计算机和网络技术的发展,数字化视频与图像信息越来越多的涌现,基于多媒体信息服务的信息时代正在向我们走来。人们对视频和图像等视觉媒体内容的需求也越来越多,越来越广泛。这就需要行之有效的技术手段来满足用户的各种需求。而“语义鸿沟”是横在人与计算机和谐交互中的一个重要障碍,这是由于人的大脑对视觉媒体的评判标准和计算机系统对视觉媒体的评判标准存在着很大差异。虽然目前针对视觉媒体的语义分析和理解有了很多研究,但这一倍受关注的技术还远远不能满足用户的普遍需求。他们需要利用更多自动提取的语义信息。

本文对视觉媒体语义自动提取中的几项关键技术进行了研究,提出了语义提取的四层技术框架,即对象语义层、场景语义层、知识及情感语义层和语义应用层,并分别研究了对象检测、场景分类、高级语义概念提取和基于本体的语义应用等多项关键技术。由于想找到一条普遍通用的语义提取技术是非常困难的,因此往往针对给定应用和利用专业领域知识对特定的视觉媒体内容采取各个击破的策略来分析和自动理解。体育视频的分析 and 理解由于具有广泛的用户群和巨大的市场潜力而成为近几年来的一个热门研究方向,而随着北京奥运会的临近,体育视频的语义分析和理解对中国具有更强的现实意义。另一方面,通过计算机技术对数字化艺术图像进行分析,并提取它们类别、风格、以及包含的内容等语义信息是一个非常重要而且迫切的问题,正逐渐获得越来越多的关注,国画是中华艺术的瑰宝,对国画等数字化艺术图像的研究也是一个重要的问题。因此本文针对视频和图像这两种视觉媒体,分别研究了体育视频和艺术图像中的语义提取技术。最后还给出了夜景图像的场景分类方法,该技术也具有重要的应用价值。具体来说,论文主要的研究成果包括:

- 1) 首先对视觉媒体的语义自动提取的系统框架进行了宏观分析,这是必要的,一方面可以对整个问题有个全局的认识,另一方面可以指导我们实现具体的语义提取技术。给出其中所包含的各个层次的语义信息;并对视觉媒体语义提取的应用框架和解决方案分别进行了系统分析。
- 2) 针对体育视频提出了一个鲁棒的球场对象分割检测方法。在很多种体育视频的自动分析中,球场区域起着至关重要的基础性作用,许多语义线索可以在球场分割结果的基础上获取。采用高斯混合模型(GMMs)为球场区域建立颜色模型,这是由于 GMMs 可以对复杂的,非线性的颜色分布进行建模,从而在进行球场区域的像素检测时具有足够的通用性。经过高斯混合模型的像素检测过程之后,采用区域分析方法把检测的像素连成区域,区域分析主要包括形态学的方法和区域增长的方法,这样得到最终的分割结果。实验证明,本文提出的方法对于不同的体育视频均能有效地实现球场区域的检测。论文还研究了体育视频场景

语义分类的技术。针对足球体育视频提出了三层分类框架，共九种场景画面。并利用球场分割的结果所得到的颜色特征，以及形状和边缘等特征，从图像底层语义的角度分析各个场景画面之间的不同之处。由于可利用的训练数据相对较少，利用支持向量机(SVM)作为分类器，它具有较好的推广能力。本文提出的分类方法可以直接用在语义标注，也可被用来进行更高级的语义分析。比赛形势的分析对于体育专业人士和长期的体育爱好者来说是十分有帮助的，这是一个较新的方向，少有人涉及。对于给定的视频序列，将利用球场分割的结果进行球员分析以及利用摄像机运动估计进行球场变化分析。并利用这两方面的信息为比赛形式建模。从而判定哪个球队在这段时间内更占优势一些。这样就可以对镜头进行自动语义标注，从而利于自动的语义视频检索，也可以用来分析整个比赛。

- 3) 国画图像是中华艺术的瑰宝。本文研究了国画图像的检测算法。使用了三个低级特征来实现这个高级的语义提取问题，分别为：颜色直方图、颜色一致性向量和自相关纹理特征。检测采用决策树与支持向量机相结合的方法来实现，并采用支持向量机作为主分类器。在一个中等规模的数据集上的正确检测率为94.85%。国画基本上可以分为工笔、写意两大类。为了区分这两种国画，提出一种新的图像特征：边缘大小直方图。这个特征反映了图像边缘的稀疏程度。使用支持向量机作为国画图像检测和分类的主要分类器，并采用颜色、纹理和新提出的边缘特征，最终得到了较好的分类结果。
- 4) 利用本体来进行多媒体的语义理解受到了越来越多的关注。文本针对艺术图像建立了视觉本体；还针对艺术图像提出了图像的非写实语义的概念。建立的本体包括艺术图像各个方面的语义概念，从而可使用户从各个角度查找需要的视觉信息。本体中的语义概念可以自动提取。最终目标是使得用户方便的根据语义查找图像，从而缩小“语义鸿沟”。
- 5) 夜景图像在数字图像尤其是家庭照片或旅游图像中占有相当的比例。夜景图像一般由比较黑暗的背景区域和非常明亮的前景区域组成。另一方面，由于夜景图像在不同的地点不同的光照环境下拍摄，也往往呈现不同的外观。本文针对夜景图像的这些特点提出了一个基于高斯混合模型(GMMs)的图像分类检测算法；在实验数据集上的分类结果为89.79%。

总之，本文的研究工作基于用户迫切的应用需求和广泛的应用前景而展开的，研究了图像和视频等视觉媒体中的不同层次的语义提取技术，重点为体育、艺术等多种视觉媒体形式的语义理解提供技术方法，从而为帮助用户更好地获取并使用他们感兴趣的数字化多媒体信息提供解决方案。

关键词：语义提取，视觉媒体，体育视频分析，图像分类，支持向量机，本体，艺术图像，高斯混合模型

Study on Key Techniques in Automatically Extracting Semantic Information from Visual Media

Jiang Shuqiang (Computer Application)

Directed By Professor Gao Wen

In the past few years, techniques of computers and Internet are improving very fast. This causes the amount of image and video content increasing drastically, and more and more people could conveniently access various equipments to obtain the desired visual information. Information era as the core of multimedia information services is coming to us. Techniques to process and analyze visual information need to be constructed to meet various application demands of users on images and video clips. However, “Semantic Gap” is a great challenge for human and computer harmonious interaction; this is because low-level features used by computers could not be always interpreted to high-level concepts that are commonly used by human. Although there exist many research works on semantic analysis and understanding of visual media, this important research area is far from satisfactory, as users need more automatically extracted semantic information.

In this thesis, we make a study on some key techniques in automatically extracting semantic information from visual media. A four-level technical framework for semantic extraction is proposed, including object semantic layer, scene semantic layer, knowledge and emotion semantic layer, and semantic application layer. Four kinds of key techniques are investigated respectively: object detection, scene classification, high-level concept extraction and ontology-based semantic application. It is hard to provide a general solution to extract all the semantic concepts from visual media, and is best approached by a divide-and-conquer strategy. Sports video always appeals to large audiences, automatically extracting useful semantic information from sports video to facilitate retrieval and organization is an important problem; and this has emerged as a hot research area recently. With the Beijing 2008 Olympic Games being near; research on semantic understanding of sports video has a special meaning for China. On the other hand, automatically analyzing and understanding digitized art images and extracting their type, style and other semantic information is an important and imperative problem that needs to be addressed. Traditional Chinese Painting is the gem of of Chinese traditional arts; research on this kind of art images is also an important problem. This thesis investigates on extracting semantics from visual media including video and image content; particularly, we concentrate on sports video and art images. At the end, we propose a technique to classify night scene images, which is also an important problem in semantic processing of images. The contributions of the thesis are as follows:

- 1) Firstly, we perform a global analysis on system framework of automatically extracting semantic information from visual media. This is a necessary work, as on

one kind, we shall have a general image on the whole problem, and on the other hand, we can benefit from implementing specified semantic extraction techniques. This paper discusses various levels of semantic information existing in visual media; and further performs systematically analysis on application framework and implementation solutions of semantic information extraction from visual media.

- 2) A novel playfield segmentation method based on Gaussian mixture models (GMMs) is proposed. Playfield plays a fundamental role in analyzing many sports video. Many semantic clues could be inferred from the result of playfield segmentation. The GMMs method is sufficiently general to model highly complex, non-linear distributions. Firstly, training pixels are automatically sampled from frames. Then, by supposing that field pixels are the dominant components in most of the video frames, we build the GMMs of the field pixels and use these models to detect playfield pixels. Finally region analysis operations are employed to segment the playfield regions from the background. Experimental results show that the proposed method is robust to various sports videos even for very poor grass field conditions. Techniques of semantic scene classification for sports video are investigated. Particularly, we propose a hierarchical view classification framework for soccer video. There are three levels in the framework including nine kinds of scenes, and for each of the level we extract different features for classification. Color feature from results of playfield segmentation, shape features and edge features are employed to discriminate various levels of scenes. SVM is selected as the view classifier based on the extracted features since it offers a possibility to train generalizable, nonlinear classifiers using small training set. Satisfactory results are obtained and could be directly used in semantic annotation as well as high-level semantic analysis. Match situation analysis is very useful for professional sports person and most of the long time sports fans, and this is a new research direction that has rarely been explored. In this thesis, a match analysis model is presented based on player analysis and playfield transition analysis. Player analysis is based on results of playfield segmentation and playfield transition analysis is based on global motion analysis. Thus which team is superior in this period of time can be estimated. Through an experiment on seven shots in a recent match between China and Iran, promising results are derived. This could not only facilitate semantic video retrieval, but also help users analysing the whole match.
- 3) Traditional Chinese Painting is the gem of Chinese traditional arts. This thesis proposes a scheme to classify traditional Chinese paintings. The algorithm uses three low-level features to achieve such a high-level classification: Ohta histogram, color

coherence vector and auto-correlation. By using a combined classifier of decision tree and support vector machine, a classification accuracy of 94.85% is achieved on the medium-size image database. Traditional Chinese Paintings could be generally classified into two styles: Xieyi (freehand strokes) and Gongbi ("skilled brush"). We propose a method to categorize Traditional Chinese Paintings into these two schools. A new low-level feature measuring the sparseness and granularity of edges in an image called edge-size histogram is proposed and used to achieve such a high level classification. Autocorrelation texture feature is also used. Our method based on SVM classifier achieves good classification accuracy.

- 4) Using ontology techniques to better understand semantic information of multimedia content are receiving more and more attention. In this work, a scheme for constructing visual ontology oriented to retrieve art images is proposed. The proposed ontology describes images in various aspects of global perceptual effects. Concepts in the ontology could be automatically derived. Nonobjective semantics are introduced, and how to express these semantics is provided. The goal of our method is to make users more naturally find visual information and thus narrows the "semantic gap".
- 5) Night scene images take considerable proportion in digital images especially in sightseeing and family photos. Night images normally contain dark regions and bright regions, thus showing a striking contrast. On the other hand, night scene images take various appearances due to different places and different lighting conditions. In this thesis, a novel method is proposed to detect night scene images for the first time by a method of Gaussian mixture color models because GMMs have the ability to form smooth approximations to arbitrary distributions. Features used in our method are mainly derived from the Gaussian mixture night color models and bright color models. The final detection accuracy is 89.79%.

Above all, based on users' urgent application requirement and extensive application expectation, this thesis investigates on various levels of semantic information extracting techniques for visual media, especially for sports and arts media types, thus providing adaptive solutions to help users better acquire their interested digital media information.

Keywords: Semantic extraction, visual media, sports video analysis, image classification, support vector machine, ontology, art image, and Gaussian mixture models