分类号_____ 密级_____

UDC _____ 编号_____

中国科学院研究生院

硕士学位论文

基于图像局部不变特征的图像场景匹配和视频运动分析

马述高

指导教师 王伟强 副教授

中国科学院研究生院信息科学与工程学院

申请学位级别 工科硕士 学科专业名称 计算机应用技术

论文提交日期 2009 年 4 月 论文答辩日期 2009 年 5 月

培养单位 中国科学院研究生院（本部）

学位授予单位 中国科学院研究生院

答辩委员会主席 卢汉清

摘　要

近十几年来，随着数字图像及视频获取设备在人们生产及生活中的日益普及、计算机存储介质成本的不断降低、互联网技术的迅速发展，同时随着人们对视觉媒体的日益关注，许多企业、机构和个人积累了大量的数字图像及视频数据，而互联网上数字图像和视频的数量更是以前所未有的速度增长，互联网正变成一个无比巨大的数字图像及视频库。因此，如何有效地对数字图像和视频进行分析和处理以满足人们生产和生活的需要，成为一个具有意义和挑战性的课题，而图像场景匹配和视频运动分析正是其中重要的研究内容。图像局部不变特征能够鲁棒的描述图像的局部结构，具有良好的视角和光照不变性，近年来引起许多研究者的关注，并被应用于图像检索、图像匹配、图像分类、物体识别和视频数据挖掘等诸多领域。本文重点研究了如何将图像局部不变特征应用于图像场景匹配及视频运动分析问题以获得高效的算法，做出的主要贡献如下：

第一，提出了一个新的图像场景匹配方法。图像场景匹配能够度量图像中场景的相似度，对于按照图像场景管理大规模数字图像库具有重要应用价值。对于一副图像，新方法首先提取其中的图像局部不变特征，然后基于亲和度聚类算法，将特征按照远近距离和相似度聚类，最终将图像表示为带权重的局部不变特征聚类中心的集合（称为特征签名）。两幅图像的场景相似度通过计算它们的特征签名间的改进的推土机距离（Earth Mover Distance -- EMD）获得。与近年国际上发表的相关算法的比较实验证明了该方法的有效性。

第二，提出了一个新的视频运动分析方法。视频中的运动信息包括摄像机的运动信息和前景物体的运动信息。在视频中摄像机或前景物体剧烈运动的情况下（例如动作片），有效的分析视频运动信息是一个十分困难、具有挑战性的问题。本文基于可靠的图像局部不变特征点的运动信息分析视频中的运动信息。该方法首先在相邻采样帧间鲁棒匹配局部不变特征点以获得可靠的特征点运动信息，进而基于这些特征点的运动信息分析并获得视频中摄像机的运动信息和前景物体的运动信息。对于摄像机的运动，该方法可以定性分析摄像机运动类型（平移或缩放），并能够估算摄像机运动参数模型；对于前景物体的运动，该方法可以提取一个描述前景物体运动特征的前景

运动特征向量。实验证明该方法可以有效分析视频（包括前背景剧烈运动的视频）中摄像机的运动类型。我们基于前景运动特征向量提出了一个电影中打斗镜头判别方法，并用实验证明了该方法的有效性，这也证明了前景运动特征向量能够有效的描述视频中前景物体的运动。

关键词：图像场景匹配、视频运动分析、图像局部不变特征、摄像机运动特征、前景物体运动特征、打斗镜头判别。

Research on Image Scene Matching and Video Motion Analysis

Based on Local Invariant Image Features

Ma Shugao(Computer Application)
Directed by: **Wang Weiqiang** (Associate Professor)

In the last decade, with increasing popularity of digital image or video capture devices, decreasing prices of storage devices, advances of web technologies and increasing interest in visual media from users, lots of organizations and individuals are accumulating large amounts of digital images and videos, and the number of digital images and videos on the Internet is also growing explosively, which makes the Internet an unprecedented large digital image and video library. It is a meaningful and challenging problem how to analyze and process so many images and videos, and image scene matching and video motion analysis are important research areas for this problem. Local invariant image features are robust descriptions of local image structures, which have good view point and lighting invariance. In recent years, local invariant image features attract many researchers and are explored in many applications such as image retrieval, image matching, image classification, object recognition and video data mining, etc. This thesis aims at finding effective methods for image scene matching and video motion analysis by utilizing local invariant image features, and the main contributions are summarized as follows:

First, this thesis presents a new image scene matching method. Image scene matching measures the scene similarity for two images, and is of central importance for the applications where we have to properly organize large amount of digital images by scene categories. For a given image, this method compactly represents its scene as a set of weighted cluster centers, called local feature signature, where the clusters are obtained using the affinity propagation algorithm to aggregate local features according to their spatial closeness and appearance similarity. The similarity of scenes in two images is then measured by a modified Earth Mover Distance (EMD) between their corresponding local feature signatures. Empirical experiments on real world images show that our method is comparable to the state-of-the-arts methods.

Second, this thesis presents a new video motion analysis method. Motion information

in a video includes both camera motion information and foreground object motion information. It is quite challenging to analyze the video motion when the camera or the foreground object moves violently. This thesis utilizes the reliable motion information of local invariant image feature points to analyze the video motion. First, a robust method is devised for the matching of feature points between two consecutively sampled frames to extract feature points' motion information, and then the camera motion and foreground object motion are analyzed based on these feature points' motion information. For camera motion, this method can both qualitatively decide the camera motion type and quantitatively estimate a camera motion parametric model; for foreground object motion, this method can extract a motion feature vector to describe the foreground motion. Experiments show that this method can effectively analyze the camera motion type in videos (including video with violent camera or object motion). The effectiveness of the foreground motion feature vector is proved by using it for discriminating the movie shots which contain fighting events.

Keywords: image scene matching, video motion analysis, local invariant image features, camera motion, object motion, fighting shot discrimination.