

上海大学出版社

2005年上海大学博士学位论文 81



视频对象自动分割技术 及其细胞神经网络 实现方法的研究

- 作者：张庆利
- 专业：通信与信息系统
- 导师：张兆扬 莫玉龙



2005年上海大学博士学位论文 81



视频对象自动分割技术 及其细胞神经网络 实现方法的研究

- 作者：张庆利
- 专业：通信与信息系统
- 导师：张兆扬 莫玉龙

上海大学出版社

Shanghai University Doctoral Dissertation (2005)

Study on Video Object Segmentation and Its Realization Method on CNUM

Candidate: Zhang Qingli

Major: Communication and Information System

Supervisor: Zhang Zhaoyang

Mo Yulong

Shanghai University Press

• Shanghai •

图书在版编目(CIP)数据

2005 年上海大学博士学位论文. 第 2 辑/博士论文编辑部编. —上海: 上海大学出版社, 2009. 6

ISBN 978-7-81118-367-2

I. 2… II. 博… III. 博士—学位论文—汇编—上海市—2005 IV. G643.8

中国版本图书馆 CIP 数据核字(2008)第 180878 号

2005 年上海大学博士学位论文

——第 2 辑

上海大学出版社出版发行

(上海市上大路 99 号 邮政编码 200444)

(<http://www.shangdapress.com> 发行热线 66135110)

出版人: 姚铁军

*

南京展望文化发展有限公司排版

上海华业装潢印刷厂印刷 各地新华书店经销

开本 890×1240 1/32 印张 274.25 字数 7641 千

2009 年 6 月第 1 版 2009 年 6 月第 1 次印刷

印数: 1~400

ISBN 978-7-81118-367-2/G·490 定价: 980.00 元(49 册)

上海大学

本论文经答辩委员会全体委员审查,确认符合上海大学博士学位论文质量要求。

答辩委员会名单:

主任:	余松煜	教授,上海交通大学	200030
委员:	王汝笠	研究员,中科院技物所	200433
	王朔中	教授,上海大学	200072
	王国中	教授, SVA 中央研究院	200039
	方 勇	教授,上海大学	200072
导师:	张兆扬	教授,上海大学	200072
	莫玉龙	教授,上海大学	200072

评阅人名单:

余松煜	教授, 上海交通大学	200030
张立明	教授, 复旦大学	200433
王朔中	教授, 上海大学	200072

评议人名单:

王浩钢	教授, 航天局 802 所	200050
张一均	教授, 索广公司	200240
翁默颖	教授, 华东师范大学	200062
严壮志	教授, 上海大学	200072

答辩委员会对论文的评语

该博士学位论文深入研究了基于 MPEG-4 的视频对象分割方法且应用了细胞神经网络(CNN)进行仿真,论文选题有前沿性和应用前景。论文取得的主要成果如下:

一、算法改进方面,提出以下几种改进的算法:1) 基于运动投影的头肩序列人脸提取算法,能较有效地从头肩序列中分割出人脸;2) 提出了两种头肩序列的视频对象分割方法:一是时空结合的视频对象分割,适合任意头肩序列,另一是基于最大梯度淹没区域边缘聚类的快速分割算法,在简单背景下有比较高的分割精度;3) 提出一种光流和改进分水岭相结合的视频对象分割方法,在复杂背景下,有较好的分割性能。

二、算法实时性方面。为提高分割的实时性,对上述1)、3)两种分割方法设计了由 CNN 实现的方案,该方案可大幅度缩减所需要的运算时间。

论文理论正确,条理清晰,结构合理,文笔流畅,实验结果可靠。反映出作者具有扎实的理论基础和系统的专业知识,具有较强的数学分析能力和独立从事科研工作的能力。在答辩中表达清楚,回答问题正确。

答辩委员会表决结果

经答辩委员会表决,全票同意通过张庆利同学的博士学位论文答辩,建议授予工学博士学位。

答辩委员会主任: **余松煜**

2005 年 4 月 28 日

摘 要

传统的视频压缩编码标准 MPEG 1/2 和 H. 26x 都采用基于帧的技术,不要求对场景进行分割。它们能获得较高的压缩比,并在许多领域得到了广泛的应用。随着多媒体信息的日益丰富,人们不再满足于对视频信息的简单浏览,而要求提供基于对象的操纵、交互等功能。

为此, MPEG 组织提出了第二代视频压缩标准 MPEG-4, MPEG-4 最显著的特征之一就是基于对象的编码方式。要实现基于对象的视频编码,首先要进行视频对象分割。视频对象分割是指在时空域上将视频分割为一些视频语义对象的集合,也就是将每一个视频帧分割为一些不同语义对象区域。事实上,视频对象分割技术不仅仅可以为 MPEG-4 服务,在许多计算机视觉领域都能发挥视频对象分割技术的作用。视频对象分割技术有很广阔的应用前景,其中的一些典型例子包括视频编码、视频制作和编辑、视频检索、视频监控等。

视频对象分割是计算机视觉领域中的一个难点,这主要是由于现实世界中场景的复杂性和多样性,很难建立一个统一的先验模型,这就决定了很难找到一种单一的方法来完成这项任务。研究通用视频对象分割方法的困难还来自对“语义对象”的定义:视频对象的分割不仅仅是低层次的视觉分割,还涉及高层次的语义定义和特征提取。所以,尽管视频对象分割算法的研究得到了广泛的重视,并且已经提出了众多的算法,但是,

视频对象分割算法与 MPEG-4 标准的要求相比还远未成熟。概括起来,目前的视频对象算法主要存在以下一些问题:首先,目前还不存在一种适合于所有序列的通用算法;其次,由于分割过程中计算量大,尚难满足实时性要求。

本论文重点研究 MPEG-4 框架下从视频序列中分割出视频对象的方法和技术,针对目前的视频分割算法大多数难以满足实时性要求的缺点,引入细胞神经网络来解决。

论文主要工作包括以下几个方面:

1) 提出了一种基于边缘投影的头肩序列人脸提取算法。通过对头肩序列进行分析,可知面部的运动细节比较丰富,也比较集中。因此可以通过对帧间差分图像做水平和垂直投影,确定人脸的大致坐标,搜寻范围变为一个矩形小区域。之后对这小块区域使用游程编码进行区域划分,结合前面得到的帧差运动信息,提取出人脸区域。这种算法综合利用了空间灰度信息和运动信息,因而能够比较准确地分割出人脸。

2) 提出了两种头肩序列中视频对象的分割算法。其中一种是时空结合的视频对象分割方法,首先使用帧间差分得到运动信息,然后使用灰度连通区域标记法对图像进行空间划分,最后结合两种信息得到最终的对象。另外一种是基于最大梯度淹没区域聚类的快速分割算法。由于头肩序列中通常都有大片相同颜色的背景区域,这样在经过梯度淹没后,就会在背景区域出现大片的集水盆,而最大集水盆对应的边缘就会近似头肩序列的对象边缘,然后对边缘附近像素从灰度、欧拉距离等特征出发进行多尺度聚类分析,达到精确分割的目的。

3) 提出了一种适合于运动和静止背景的基于光流和改进

的分水岭的视频对象分割方法。这种方法也是一种时空联合的方法,首先对待处理图像进行形态学重建滤波预处理,然后使用多尺度形态学算法求取梯度,对梯度淹没形成集水盆后使用改进的分水岭算法进行区域划分;利用 Horn-Schunck 光流法求取运动信息,其中使用了一种新的有限差分方法来估算光流法中的偏微分;最后结合时空信息得到最终的视频对象。

4) 使用 CNN 实现了上述的 1) 和 3) 两种算法,使之能实时分割。一种是头肩序列中人脸提取算法,另一种是全自动的视频对象分割方法。给出了 13 种模板结构,其中构建了逻辑“与”模板、+1 模板和有限差分模板等模板。

本文对提出的所有方法都进行了实验验证。

关键词 视频对象分割, MPEG-4, 时空联合, 数学形态学, 细胞神经网络, 光流, 分水岭

Abstract

Classical video coding standards such as H.26x and MPEG-1/2 are frame-based techniques, and no segmentation of video scenarios is required. Their high compression performance makes them widely used in video application. With the proliferation of multimedia information, people are no more satisfied with simple navigation of video contents, but require object-based functionalities.

Therefore, Motion Picture Experts Group (MPEG) published the secondary generation video coding standard, MPEG-4. Compared with the first generation standard, a significant character of MPEG-4 is object-based coding which means to code videos as a set of semantic video objects. Video object segmentation is necessary to get each video object. Video object segmentation has been promoted greatly by the occurrence of MPEG-4 but is not limited to MPEG-4. Video object segmentation can serve for many application in computer vision area. Some typical applications of video object segmentation are video coding, video authoring and edit, video retrieval, video-based monitoring, etc.

Video object segmentation is a nut in video processing and computer vision. The difficulty of video object

segmentation lies on two aspects. One is the extreme complexity of video scenes, which means no uniform model for all video objects. Another one is the definition and description of semantic video object. Video object segmentation is carried out on low vision level while semantic video objects are defined on high vision level. It is difficult to get high-level objects by low-level segmentation. Generally, there are two problems in the current segmentation methods: one is no universal algorithm suited for all the scenes, another one is most of the current algorithms are hard to meet the real-time performance.

This dissertation focuses on the methodology and techniques for video object segmentation under the framework of MPEG - 4, cellular neural network is introduced to conquer the problem of real-time performance.

Major work of this dissertation is as follows:

First, a face extraction algorithm based on edge projection is proposed. An important character is discovered via the analysis of head-shoulder sequences, the motion details are rich in the face region, so the approximate coordinates of the face can be determined by the vertical and horizontal projection of DFD image. Then this small rectangle region is divided into several small regions using run length coding algorithm, judged by the motion informations in DFD, the face can be extracted.

Second, two automatic video object segmentation algorithms of head-shoulder sequence are proposed. The first

one is a spatio-temporal algorithm, a symmetrical inter-frame difference is performed on a group of gray image initially, then the image is divided into some flat zones with uninterrupted grey scale information. Finally some regions are merged and forming the object according to a rule, others are discarded. The second one is a fast segmentation method, this method is based on the maximal gradient region edge clustering.

Third, an automatic video object segmentation algorithm based on optical flow and improved watershed is proposed. In this method, a new finite difference is used in the optical flow computation of the proposed algorithm at first; then, the original image is labeled into different regions by improved watershed algorithm; finally, some regions are merged into video object according to a certain criterion, others are discarded.

Fourth, the first and the third algorithms are realized by CNN to reach the target of real-time performance. 13 templates are used in this task include three templates made by myself: “And” template, + 1 template and limited difference template.

All the proposed algorithms are validated through experiments.

Key words video object segmentation, MPEG - 4, spatio-temporal, mathematic morphology, cellular neural networks, optical flow, watershed

目 录

第一章 绪论	1
1.1 引 言	1
1.2 问题背景	2
1.3 视频对象分割综述	4
1.4 发展趋势及存在问题	19
1.5 细胞神经网络的引入	20
1.6 论文的主要内容与结构	22
1.7 本章小结	24
第二章 视频分割基础	25
2.1 图像分割	25
2.2 运动估计	36
2.3 对象跟踪	44
2.4 数学形态学	48
2.5 本章小节	52
第三章 头肩序列的分割算法	53
3.1 引言	53
3.2 头肩视频序列分析	53
3.3 基于时空信息的头肩序列分割算法	56
3.4 基于运动投影的人脸提取算法	60
3.5 基于梯度最大淹没区域的头肩序列快速分割算法	66
3.6 小结	73

第四章 基于光流和改进分水岭的视频对象分割方法	74
4.1 引言	74
4.2 算法详述	74
4.3 实验结果与分析	91
4.4 小结	92
第五章 细胞神经网络分析及其用于分割算法的实现	93
5.1 引言	93
5.2 CNN 的基本模型	93
5.3 CNN 网络特性分析	98
5.4 CNN 的输入输出量化及用于图像处理的基本思想	103
5.5 CNN 网络权的设计	105
5.6 人脸提取算法的 CNN 实现	107
5.7 基于光流和改进分水岭分割算法的 CNN 实现	121
5.8 本章小结	132
第六章 结论与展望	133
6.1 论文工作总结	133
6.2 展望	134
参考文献	136
致谢	147

第一章 绪 论

1.1 引言

当今时代,信息技术和计算机互联网飞速发展,在此背景下,多媒体信息已成为人类获取信息的最主要载体,同时也成为电子信息领域技术开发和研究的热点^[1]。多媒体信息经数字化处理后具有易于加密、抗干扰能力强、可再生中继等优点,但同时也伴随海量数据的产生,这对信息存储设备及通信网络均提出了很高要求,从而成为影响人们有效获取和使用信息的瓶颈。

因此研究高效的多媒体数据压缩编码方法,以压缩形式存储和传输数字化的多媒体信息具有重要意义。作为多媒体技术的核心及关键,多媒体数据压缩编码近年来在技术及应用方面都取得了长足的进展,它的进步和完善正深刻影响着现代社会的方方面面。

人类获取的信息中 70% 来自视觉,视频信息在多媒体信息中占有重要地位;同时视频数据冗余度最大,经压缩处理后的视频质量高低是决定多媒体服务质量的关键因素。因此数字视频技术是多媒体应用的核心技术,对视频编码的研究已成为信息技术领域的热门话题。

伴随着视频编码相关学科及新兴学科的迅速发展,新一代数据压缩技术不断诞生并日益成熟,其编码思想由基于像素和像素块转为基于内容(content-based)。它突破了香农信息论框架的束缚,充分考虑了人眼视觉特性及信源特性,通过去除内容冗余来实现数据压缩^[3]。

与此同时,视频编码相关标准的制定也日臻完善。视频编码标准主要由 ITU-T 和 ISO/IEC 开发。ITU-T 发布的视频标准有 H. 261、H. 262、H. 263、H. 263+、H. 263++、ISO/IEC 公布的 MPEG