

2004年上海大学博士学位论文 ⑩

基于内容多媒体应用的语义视频 对象提取及评价

作 者：杨高波

专 业：通信与信息系统

导 师：张兆扬




上海大学出版社

· 上 海 ·

2004 年上海大学博士学位论文

基于内容多媒体应用的语义视频 对象提取及评价



作 者： 杨高波
专 业： 通信与信息系统
导 师： 张兆扬

上海大学出版社

• 上海 •

Shanghai University Doctoral Dissertation (2004)

**Semantic Video Object Segmentation and Its
Performance Evaluation for
Content-Based Multimedia Applications**

Candidate: Yang Gaobo

Major: Communication and Information System

Supervisor: Prof. Zhang Zhaoyang

Shanghai University Press

• Shanghai •

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，上海交大图象所 200030

委员：戚飞虎 教授，上海交大计算机系 200030

王朔中 教授，上海大学电子信息工程系 200072

王治钢 研究员，上海航天局 809 研究所 200031

张一钧 教授级高工，上海广电(集团)公司 201108

王国忠 教授级高工，SVA 中央研究院 201108

导师：张兆扬 教授，上海大学 200072

评阅人名单:

余松煜	教授, 上海交大图象所	200030
张立明	教授, 复旦大学电子工程系	200433
王朔中	教授, 上海大学电子工程系	200072

评议人名单:

翁默颖	教授, 华东师大电子科学系	200062
王治钢	研究员, 上海航天局 809 研究所	200031
张一钧	教授级高工, 上海广电(集团)公司	201108
方 勇	教授, 上海大学电子信息工程系	200072

答辩委员会对论文的评语

杨高波同学的博士学位论文《基于内容多媒体应用的语义视频对象提取及评价》是国家自然科学基金项目(60172020)和上海市教委重点学科基金(2001-44)的资助项目中的重要内容之一。语义视频对象的提取是当前国际上前沿的研究热点之一。该文从提高视频对象的分割精度和提高分割速度两个角度,展开了深入的研究,在以下几个方面取得创新性研究成果:

(1) 提出了一种有效减少分水岭变换“过分割”的方法。通过小波分解,滤除了图像中的噪声等易于过分割的因素,然后在二级小波子空间进行分水岭变换,可以有效地克服过分割现象,并减少了后续的区域合并的计算负担。算法分割精度比COST211 AM分割精度提高约10%。

(2) 提出了一种适合半自动分割算法初始对象轮廓勾勒的改进智能剪算法。它通过引入边界框、简化代价函数和改进搜索策略等,使分割速度比改进前提高6~8倍。

(3) 针对头肩序列,提出了一种适合于细胞神经网络实现的视频对象自动分割算法。该算法已充分满足实时性要求。

(4) 提出了一种存在参考分割时客观的视频分割算法评价方法。该方法评价结果与主观评价一致,且计算量小。

作者已发表或录用的论文中,属SCI源刊3篇,EI源刊7篇。该论文思路清晰、结构严谨、文字通畅、富于创新。综上所述,杨高波同学掌握了坚实宽广的基础理论和系统深入的专门知

识,独立从事科研创新工作的能力强.在答辩过程中,陈述清晰,回答问题正确.经答辩委员会无记名表决,一致同意通过博士学位文答辩,并建议校学位委员会授予工学博士学位,推荐申报优秀博士学位论文.

答辩委员会表决结果

经答辩委员会表决,全票同意通过杨高波同学的博士学位论文答辩,建议授予工学博士学位.

答辩委员会主席: 余松煜

2004年3月10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-4 引入视频对象的概念,以支持基于对象的交互性和可分层性; MPEG-7 则对各种媒体对象进行统一和规范化的描述.按照 MPEG-4 的校验模型,视频序列必须先分割成具有语义意义的视频对象,然后对其运动、形状和纹理分别进行编码.视频对象的应用价值主要有:对不同的视频对象按其对视觉重要性分配不同的码率,以提高压缩编码效率;支持对象可分级,在较低的网络带宽时获得更好的视觉效果;用视频对象来组织视频内容,实现基于视频内容的存储、交互和查询等功能.

然而, MPEG-4 尽管引入了视频对象的概念,但它并没有指定从视频序列获取视频对象的具体方法.一方面,视频对象的语义一致性难以通过视频的低级物理特性来建模,使得针对各种视频序列的通用视频对象分割算法是一个尚未解决的经典难题;另一方面,针对特定的应用,往往可以利用先验知识设计相应的算法.

本论文重点研究 MPEG-4 框架下的从视频序列中分割出视频对象的方法和技术,以及其在基于内容多媒体中的应用.研究目标是:对特定类型的序列如头肩序列,算法满足实时性要求;

对背景静止的序列,全自动分割算法取得较好的分割效果;对复杂背景和前景运动视频序列,采用半自动分割算法,要求得到较好的分割质量,而且人机交互简单.具体地,本文研究的主要内容和贡献包括:

提出了两种全自动的视频对象分割算法.第一种采用背景记录和变化检测,主要由预处理、背景记录、背景缓冲、变化检测和后处理等几部分组成.它不需要诸如运动估计、特征空间分析等计算量大的操作,并能有效去除阴影和光照变化造成的影响.它能够生成背景信息,支持 MPEG-4 的精灵编码.第二种是一种基于时空分割融合的视频对象提取改进算法.时间分割基于变化检测,其关键的阈值选取是通过直方图分析得到的.空间分割是本算法的核心,采用基于小波变换的分水岭变换算法.

提出了一种半自动的视频对象分割算法.为方便用户定义初始对象轮廓,提出了一种修改的智能剪.它通过引入边界框、简化代价函数和改进搜索策略等,可提高优化路径搜索速度约 6~8 倍,而几乎不损失分割精度,完全满足半自动分割算法对初始对象轮廓勾勒的要求.为克服对象跟踪过程中的误差积累,按视频对象的刚性、非刚性以及全局、局部直方图比较进行视频分解得到后续帧的视频对象.由于视频分解以及人工参与,它可以在很大程度上解决遮挡问题,取得了比 COST211 AM 更好的分割效果.

针对目前的视频分割算法大多数难以满足实时性的要求,采用了一种新的计算体系结构,即将细胞神经网络引入视频对象分割.细胞神经网络(CNN)是一种非线性模拟电路,由大量胞元组成,且只允许最邻近的胞元间直接通信.由于它具有的高度并行的实时处理能力和机理类似于人类视觉系统,特别适合于图像处

理等领域. 然而, 与传统的 CISC 处理器相比, CNN 只能利用一些简单的基于像素的函数, 有相对狭窄的指令集——尽管有很高的速度. 因此, 基于 CNN 体系结构的视频对象分割算法的关键是充分考虑到细胞神经网络的特点, 将复杂的视频分割算法分解为一些 CNN 胞元能够完成的低级操作. 论文提出基于彩色边缘变化检测的视频分割算法. 所有的模板都是 3×3 的线性模板, 并能在 CNN 的模板库中得到, 因此, 其易于 CNN 实现.

本文提出了一种客观的存在参考分割时分割算法评价方法. 视频分割算法往往只适合特定的应用, 其性能依赖于具体的序列. 目前, 视频分割算法的性能评价以对已知序列的分割结果的主观评价为主, 尚没有一种广泛接受的客观评价方法. 视频分割算法的评价是重要的, 它有助于针对具体的应用选取合适的算法并设置恰当的参数, 以及有利于通过融合各种算法的优点发展新的算法. 而且, 自动分割算法采用性能评价作反馈可改进分割性能. 空间精确度通过相对前景面积、位置、边界像素距离以及像素分类来进行, 并将其按对人类视觉系统的重要性线性加权. 而时间一致性反映了分割算法分割各帧时的稳定性, 它通过空间准确度的变化来刻画. 实验证明, 其评价结果与主观评价结果一致, 而且计算量小.

综上所述, 本论文系统地研究了 MPEG-4 框架下的语义视频对象分割问题, 根据具体的问题提出了满足实际需要的全自动、半自动分割算法, 并探讨了在存在参考分割的情况下客观地评价视频分割算法性能的方法.

关键词 视频对象分割, 细胞神经网络, 分割评价, MPEG-4

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 applications. With the proliferation of multimedia information, people are no more satisfied with simple navigation of video contents, but require object-based functionalities. Therefore, MPEG-4 introduces the concept of video object to support content-based functionalities. MPEG-7 defines a universal and normalized description of various multimedia objects. According to the MPEG-4 verification model, video sequence must be segmented into semantic video objects. Their motion, shape and texture information are coded respectively. The main values are: improved coding efficiency by allocating different bit rate to different video object in accordance with their importance to human visual system; object-based scalability so as to obtain better visual effect at low bit rate applications; content-based storage, interactivity and retrieval by organizing video content according to video object.

Though MPEG-4 introduces the concept of video object, it does not specify any concrete techniques for obtaining video objects from video sequence. On one hand, the semantic homogeneity of video object is hard to be modeled by any low level features, which makes a generic segmentation algorithm for various video sequences still a

classical problem to be resolved; On the other hand, priori knowledge can often be utilized for specific applications.

Therefore, the dissertation focuses on the methodology and techniques for video object segmentation under the framework of MPEG-4 and its application in content-based multimedia systems. The main objectives are as follows. For some specific video sequences such as head-shoulder sequence, video object segmentation should meet the real-time performance. Automatic video segmentation can achieve better results for video sequences with simple or still background. For sequences with complex background, semi-automatic segmentation can achieve satisfactory results, and the human intervention should be simple. Major works of this dissertation are as follows:

First, two automatic video object segmentation schemes are proposed. The first one is based on background registration and change detection. It consists of preprocessing, background registration, background buffering, change detection and post-processing. It doesn't need computation-intensive operations such as motion estimation, and it can overcome the influence of shadow and illumination variance. It can produce background information, which makes it support MPEG-4 sprite coding. The second one is an improved spatio-temporal segmentation. Temporal segmentation is based on change detection, and its key is the selection of threshold, which is obtained by threshold analysis. Spatial segmentation is the core of the whole algorithm, which is a wavelet based watershed scheme.

Second, a semi-automatic video object segmentation algorithm is proposed. To facilitate users defining the initial object contour, a modified intelligent scissors is proposed on the basis of original intelligent scissors. By introducing bounding box, simplified image features and improved searching strategies, it can improve about 6~8 times the processing speed with just slight sacrifice of segmentation accuracy, which fully meets the requirements for initial object extraction in semi-automatic segmentation. To avoid errors accumulating and propagating during object tracking, video decomposing is conducted based on the rigidity of video object and global/local histogram comparisons. Then, region-based backward projection is utilized to interpolate the VOPs of successive frames. Because of video decomposing and human intervention, it can solve the occlusion problem to most extent. Experimental results demonstrate that it can achieve better segmentation results than COST211 AM.

Third, video object segmentation in the cellular neural networks is proposed. Since most of the current algorithms are hard to meet the real-time performance, a new architecture, i.e. cellular neural networks (CNN), is introduced into video object segmentation. CNN is a non-linear circuit made up of many cells. Only direct communication between adjacent cells is allowed. Because of its high parallel mechanism and similarities with human visual system, it is very suitable for image/video processing. Contrary to classical SISC processor, only simple pixel-based functions are defined with a relative narrow instruction set. Therefore, video object segmentation

in the CNN architecture should take the characteristics of CNN into consideration, and decompose the complex algorithm into low-level operations that CNN can perform. A video object segmentation based on color edge based change is proposed. All the templates are 3×3 linear ones, which can be found in the CNN template library. So it can be realized in the CNN architecture.

Fourth, a methodology for objective evaluation of video segmentation algorithms with ground-truth is proposed. Most of video segmentation algorithms are only suitable for specific applications and their performance depends on the video sequences. Up to now, performance evaluation of video segmentation algorithms is mainly subjective evaluation by human observers, and a widely accepted objective evaluation is in absence. However, performance evaluation of video segmentation algorithms is important. It can help users' selection of appropriate algorithms and their parameters according to specific applications. It can also do benefit to developing new algorithm by fusion of the advantages of various algorithms. Moreover, feedback based on performance evaluation can improve the segmentation accuracy. An objective evaluation methodology with ground truth is proposed. Four metrics based on relative foreground area, position, distance between edge pixels and pixel classification are weighted to address the spatial accuracy. Temporal coherency reflects the stability when segmenting every frame, which is defined as the differences of spatial accuracy between adjacent frames.

In summary, the dissertation systematically discusses video

object segmentation under the framework of MPEG-4. Fully automatic and semi-automatic algorithms are proposed for specific applications. Objective performance evaluation of video segmentation algorithms with ground truth is also discussed.

Key words video object segmentation, cellular neural networks, performance evaluation, MPEG-4

目 录

第一章 绪 论	1
1.1 问题背景	1
1.2 视频对象分割的综述	4
1.3 论文研究的内容	17
1.4 论文的主要贡献及结构	19
1.5 本章小结	22
第二章 视频对象分割基础	23
2.1 数学形态学	23
2.2 图象分割	31
2.3 运动估计	39
2.4 对象跟踪	46
2.5 本章小结	54
第三章 全自动视频对象分割	55
3.1 基于背景记录 and 变化检测的全自动分割算法	56
3.2 基于小波分解和分水岭变换的分割算法	64
3.3 本章小结	79
第四章 半自动视频对象分割算法	80
4.1 半自动分割算法概述	81
4.2 半自动视频分割算法	84
4.3 本章小结	113
第五章 基于细胞神经网络的视频对象分割算法	114
5.1 细胞神经网络	115

5.2	针对头肩序列的视频对象分割算法治	127
5.3	CNN 模板的设计方法简介	144
5.4	本章小结	145
第六章	视频分割算法的性能评价	146
6.1	视频分割算法的性能评价概述	146
6.2	分割评价现状	148
6.3	客观性能评价方法	150
6.4	本章小结	162
第七章	结论及进一步工作	163
7.1	论文工作总结	163
7.2	进一步的工作	165
参考文献		167
致 谢		179