

上海大学出版社

2005年上海大学博士学位论文 84



基于内容的视频运动对象 分割技术研究

- 作者：包红强
- 专业：通信与信息系统
- 导师：张兆扬



上海交通大学出版社
上海交通大学博士学位论文



基于内容的视频运动对象 分割技术研究

作者：周成德

导师：周圣博、周圣博、周圣博

答辩日期：2008年

上海交通大学
图书馆

上海交通大学图书馆

上海交通大学图书馆

上海大学出版社

2005年上海大学博士学位



基于内容的视频运动对象 分割技术研究

- 作者：包红强
- 专业：通信与信息系统
- 导师：张兆扬



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)

Shanghai University Doctoral Dissertation (2005)

Research on video motion object segmentation for content- based application

Candidate: Bao Hongqiang

Major: Communication and Information System

Supervisor: 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,上海交大电子信息工程学院	200030
委员:	王朔中	教授,上海大学通信学院	200072
	王汝笠	教授,中科院技术物理所	200083
	方 勇	教授,上海大学通信学院	200072
	王浩钢	研究员,上海航天局 809 研究所	200031
导师:	张兆扬	教授,上海大学	200072

评阅人名单:

余松煜	教授,上海交大电子信息工程学院	200030
张立明	教授,复旦大学	200433
王朔中	教授,上海大学通信学院	200072

评议人名单:

王园中	研究员,上广电中央研究院	200054
严壮志	教授,上海交大电子工程系	200030
戚飞虎	教授,上海交通大学计算机学院	200030
翁默颖	教授,华东师大电子科学系	200062

答辩委员会对论文的评语

视频对象分割是目前视频处理方向的前沿研究之一,选题具有重要的理论意义和良好的应用前景.论文在充分研究现有各种算法的基础上,着重解决分割准确性、遮挡处理、多视频对象分割等较难解决的问题,其中主要的创新点如下:

提出一种基于区域多重选择的视频对象分割方法,该方法根据区域的时空特性,结合马尔可夫随机场模型等后处理方法,可提高对象分割的准确性.

提出一种基于背景重建技术的视频对象分割算法,其原理简单,复杂度低,能有效地解决视频分割中对象与背景的遮挡问题.

提出一种基于贝叶斯框架的时空标记场的多视频对象分割算法,该方法结合运动对象的多个时空特性,通过求取不同标记多个对象最大后验概率的方法使得能分别提取场中的多个运动对象.

提出了一种基于时空曲线演化的分割算法,通过运用基于 level set 的时间-空间联合能量模型及演进算法,结合运动矢量有效性判断和贝叶斯最小差错判决法,使得能有效地处理出现的新对象和较好地解决对象间的互遮挡问题.

论文条理清晰、理论分析深入、实验数据丰富可信,表明作者具有坚实、宽广的专业理论基础,独立科研工作能力较强.在答辩过程中,回答问题正确.

答辩委员会表决结果

经答辩委员会表决，全票同意通过包红强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**

2005 年 4 月 28 日

摘 要

随着现代信息社会的不断发展,人们获得的多媒体信息日益增加.因此,对各种多媒体信息的处理技术就变成了相关领域内学者的重要研究任务.在这些多媒体中,视觉信息因其形象、生动和直观的特点而更为人们所关注,但同时视觉信息巨大的数据量给传输、存储、处理带来了挑战.因此对视觉信息尤其是视频信息的各种处理技术的研究就从来没有停止过.

为了更有效地存储、传输和使用视频数据,必须进行视频压缩. MPEG 组织提出了第二代视频压缩标准 MPEG4. 相比与第一压缩标准, MPEG4 最显著的特点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组合,也就是将每一个视频帧分割为一些不同语义对象区域. 同时为视频数据检索服务的 MPEG7 标准则提出了对各种媒体对象进行统一和规范化的描述. 而一些计算机视觉等也需要基于对象的各种功能. 因此,基于对象的视频信息描述方式已成为多媒体信息处理技术中的一种巨大需求.

然而,由于视频对象分割是相当困难的问题, MPEG4 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. 视频对象分割的困难性主要体现在两方面: (1) 现实世界中视频场景极度复杂和多样,很难用一种方法进行统一的分割; (2) 视频对象的定义是一种基于高级语义的描

述,很难用低层次的视觉描述方式如边缘、颜色、运动等特征来描述,而目前的图像分析技术又远远不足以将高层次语义对象用机器语言来准确的定义和描述。

本论文重点研究 MPEG4 框架下的从视频序列中分割出视频对象的方法和技术。研究目标包括视频序列中运动对象在各种复杂情况下的分割方法,既有单视频对象的分割方法,又有多视频对象的分割方法,对视频对象分割中的一些难点技术如遮挡、对象的分裂合并、时空融合等进行了重点研究。具体地,本文研究的主要内容和贡献包括:

针对视频对象分割精度不高的问题,提出了一种基于区域多重选择的视频运动对象分割提取方法:首先利用一种有效的方法在空间域(帧内)进行区域划分,得到基于区域表示的空间图;然后利用改进的帧差法计算序列时间域(帧间)信息,对所有区域进行分类选择,区域内运动像素点较多的确定为对象区域,运动点非常少的为背景区域,而介于两者中间的区域为候选区域;为了对候选区域进行分类选择,建立时空区域能量模型,进一步选择;最后利用马尔可夫模型和标记的方法对所得到的对象区域进行后处理,得到完整的对象掩模,通过映射,即可获得运动的视频对象。

针对视频对象分割中存在的背景与对象之间的遮挡问题,提出了一种利用背景重建技术进行视频对象提取的方法。首先提出一种新的噪声参数估计方法,运用变化检测技术得到表示视频对象运动区域的二进制变化检测掩模(CDM),针对 CDM 纹理性差的特点,以邻域最大值法来修补边缘并填补空洞,经数学形态学滤波后得到完整的二进制掩模;然后利用二进制掩模映射获得本帧背景图像,通过多帧背景图像相互补充可得到

较为完整的重构背景图像;最后用当前帧和重构的背景图像相减的办法消除背景遮挡的影响,得到初始的视频对象,经数学形态学边缘修正后,获得准确的视频对象。

针对多视频对象的分割问题,提出了一种基于贝叶斯框架的时空标记场最大后验概率的多视频对象分割算法,根据视频序列帧间(时间域)和帧内(空间域)信息的不同特点,建立基于多个对象分割标记场的最大后验概率公式,并导出其最小能量函数,通过求解最小能量使其分割标记的后验概率达到最大。最小能量的求解过程用优化方法——迭代条件模式(ICM)方法,初始分割标记场用矢量直方图法得到。提出的算法是不仅对单独运动的多对象分割是有效的,而还能处理对象间的部分遮挡问题。

针对具有复杂运动的多视频对象,提出了一种基于时空曲线演化的多视频对象自动分割方法,首先根据视频序列帧间(时间域)和帧内(空间域)信息的不同特点,建立基于全局和局部特征的能量模型,并由此导出基于 level sets 方法的曲线演化方程;然后用视频序列的连续两帧帧差得到初始的视频对象,分别进行时间和空间曲线演化跟踪,提取多个视频对象;当对象因运动而发生相互遮挡现象时,利用基于 Bayes 最小错误概率决策法则的判断方法,分割遮挡对象和显露对象。

综上所述,本论文系统地研究了 MPEG4 框架下的语义视频对象分割问题,根据具体的问题不仅提出了满足实际需要的自动分割算法,而且对其中的一些困难和关键的技术进行重点研究,提出了相应的解决方法。

关键词 图像处理,视频对象分割,时空联合,MPEG4

Abstract

With the development of the modern information society, more and more multimedia information is available. So the technology of multimedia processing is becoming the important task for the irrelevant area of scientist. Among of the multimedia, the visual information is more attractive due to its direct, vivid characteristic, but at the same time the huge amount of video data causes many challenges if the video storage, processing and transmission.

Video compression is necessary for efficient storage, processing and transmission. Motion picture experts group (MPEG) published the secondary generation video coding standard, MPEG4. Compared with the first generation standard, a significant character of MPEG4 is object-based coding which means to code videos as a set of semantic video objects. Segmentation is necessary to get each video object. MPEG7, which is based on video database querying, also defines a universal and normalized description of various multimedia objects, and the computer vision needs function based on object. Therefore the method of description based on object is becoming a huge requirement of multimedia information processing.

Video object segmentation belongs to the most difficult

and important problem. Though MPEG4 introduces the concept of video object, it does not specify any concrete techniques for obtaining video objects from video sequence. 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. The semantic homogeneity of video object is hard to be modeled by any low level features, such as edge, texture, color and motion.

Therefore, the dissertation focuses on the methodology and techniques for video object segmentation under the framework of MPEG4. The main objective is the objects segmentation with the various of complexity motions, which not only include the single object segmentation, but also multiple objects. The key techniques were the emphases such as occlusion, objects combining and separating, and spatio-temporal segmentation. Major works of this dissertation are as follows:

To solve the problem of inaccurate segmentation, an automatic video segmentation is presented. The algorithm is designed based on regions classification. Regions partition is simple and fast in respect to others methods, which just meets the requirements for automatic video object segmentation. To overcome the errors of regions classification between moving object and background, the candidate regions is introduced as

one of three classes. Energy model based on the motion and the spatial constraints is build to perform the candidate regions selecting. A post processing is utilized to get the accurate object.

To solve the occlusion problem between background and motion object, a novel video object segmentation algorithm based on background rebuilding is proposed. Firstly, the change detection is used to achieve the mask representing moving regions with a new estimation noise parameter. Due to the shortage of the moving video object textures, the methods of maximum in neighborhood regions is present, which is used to smooth the mask boundary and fill the interior hole, then a morphological filter is applied to refine the moving mask. Secondly, the background image is available by mapping the mask to the correspondence frame of sequences, and a complete background image is reconstructed from multiple frames. Finally, the initial video object is derived in each frame by subtracting the background from this image, and the occlusion regions (covered/uncovered regions) are eliminated, then an accuracy video object was extracted after mathematic morphology post-processing.

To solve the problem of multiple objects segmentation, a novel multiple objects segmentation algorithm is proposed based on a Bayesian framework. According to the characteristic of the intra-frame and inter-frame (spatial and temporal) information, a representation of Maximization of the A posteriori Probability (MAP) of spatio-temporal label

field is proposed, and related energy function is minimized to obtain multiple objects with different label. The optimization of solution is carried out by Iterated Conditional Mode (ICM) method. The initial segmentation label fields are available with vector classification.

To solve the problem of multiple objects segmentation under complexity motion, a novel multiple objects segmentation algorithm is proposed based on spatial-temporal curve evolution. First, according to the dissimilar characteristic of the intra-frame and inter-frame (Spatial and Temporal) information, a joint energy model is proposed with global and local features, thus, a curve evolution equation could be achieved based on the method of level sets. Then, after an initial object model is achieved with a simplified method of the difference between two successive frames, multiple objects are tracked and extracted with spatio-temporal curve evolution. Finally, while the occlusion is emerged due to multiple objects overlapping motion, the objects could be segmented using Bayes classification for minimum error.

In summary, the dissertation systematically analysis video object segmentation under the framework of MPEG4. Some innovated algorithms are proposed for the several key and difficult techniques. The experiments show that these methods are effective.

Key words image processing, video object segmentation, spatio-temporal framework, MPEG4

目 录

第一章 绪论	1
1.1 引言	1
1.2 视频对象分割的基本概念	2
1.2.1 对象的定义	2
1.2.2 视频分割与图像分割	3
1.2.3 研究视频对象分割的意义	4
1.3 视频对象分割的现状与存在问题	6
1.3.1 研究现状	6
1.3.2 算法基本分类	7
1.3.3 复杂运动的视频对象提取问题	9
1.4 本论文的研究背景及主要内容	9
1.5 论文结构	11
1.6 本章小结	12
第二章 视频对象分割的基本理论	13
2.1 引言	13
2.2 数学形态学预处理及后处理技术	15
2.2.1 数学形态学基本概念	15
2.2.2 形态学图像处理	15
2.3 空间域分割技术	18
2.3.1 阈值法	19
2.3.2 聚类法	20
2.3.3 分水岭变换	21
2.4 时间域分割技术	23