

2001 年上海大学博士学位论文

13

# 视频对象平面分割和视频对象 提取方法的研究

作者：史立

专业：通信与信息系统

导师：张兆扬



上海大学出版社

11.73

2001 年上海大学博士学位论文

# 视频对象平面分割和视频对象 提取方法的研究

作者：史立

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 (2001)

**Study on the Segmentation of Video Object  
Plane and Extraction of Video Object**

**Candidate:** Shi Li

**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：

- |               |                |        |
|---------------|----------------|--------|
| <b>主任：戚飞虎</b> | 教授，上海交大计算机系    | 200030 |
| <b>委员：周源华</b> | 教授，上海交大图象研究所   | 200030 |
| <b>莫玉龙</b>    | 教授，上海大学通信工程系   | 200072 |
| <b>王文骐</b>    | 教授，上海大学电子信息工程系 | 200072 |
| <b>陈泳恩</b>    | 教授，上海同济大学计算机系  | 200092 |
| <b>导师：张兆扬</b> | 教授，上海大学        | 200072 |

### 评阅人名单:

<b>戚飞虎</b>	教授, 上海交大计算机系	200030
<b>余松煜</b>	教授, 上海交大图象研究所	200030
<b>薛向阳</b>	教授, 复旦大学计算机系	200433

### 评议人名单:

<b>朱维乐</b>	教授, 电子科技大学	610054
<b>王朔中</b>	教授, 上海大学电子信息工程系	200072
<b>莫五戈</b>	教授, 上海大学通信工程系	200072
<b>翁默颖</b>	教授, 华东师大电子科学系	200062

## 答辩委员会对论文的评语

基于对象视频分割是当前的研究热点之一,论文以此为选题具有较强的学术意义及很大的应用价值。

该博士论文在自动 VOP 分割技术上作了大量工作,主要创新点有:

(1) 改进了基于形态运动分割和对象跟踪算法,较好地分离出运动不太剧烈的前景对象;

(2) 提出了基于序列中边缘图象帧间差与改进的 Hausdorff 距离跟踪器相结合的前景对象提取方法,提高了分割的精确性;

(3) 以基于迷宫算法的活动轮廓模型来跟踪对象,对运动位移虽不很大但发生了非刚性形变的对象,改善了跟踪效果。

作者收集、消化了大量国内外有关文献,做了大量工作,并取得了创新性成果,表明作者掌握坚实宽广的基础理论知识和系统深入的专门知识,独立从事科学研究的能力强。

在答辩中能正确回答答辩委员提出的问题,经无记名投票,一致同意通过论文答辩。建议授予工学博士学位。

## 答辩委员会表决结果

经答辩委员会表决,全票同意通过史立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: **戚飞虎**

2001年8月7日

## 摘 要

MPEG-1、MPEG-2、H.261 和 H.263 等经典的视频编码标准由于不需要对场景进行分割或分析, 所以从某种意义上说是低级处理技术. 尽管它们可以获得比较高的压缩比且应用非常广泛, 然而随着多媒体和基于内容交互性的应用日益增多, 需有新的视频编码方案.

正在兴起的 MPEG-4 视频编码标准由于引入了视频对象平面 (VOP) 的概念而具备基于内容的交互功能. 由于输入视频序列的每一帧被分割成任意形状的视频对象平面 (VOP), 这样每个 VOP 描述了一个语义意义的对象或所感兴趣的视频内容.

将视频序列分解成 VOP 是一项相当困难的任务. 生成 VOP 的难点是所感兴趣的对象往往难以用诸如彩色、亮度和光流等传统的图象分割特征来表达, 所以经典的分割算法无法获得有语义的划分.

本论文探讨了生成视频对象平面 (VOP) 的技术, 并且提出了从视频序列中自动分割 VOP 和自动提取 VO 的算法, 由于视频对象通常以异于背景的运动来表征, 所以本算法结合运动信息进行处理.

理论上可以使用光流或运动场估计进行分割, 但其对噪声相当敏感, 且分割精度也会因遮挡和孔径效应的影响而剧烈降低. 另一方面, 可通过变化检测掩模将遮挡区域标识为变化的, 但除非对象包含足够的纹理信息, 否则就难于精确定位对象位置, 而且还必需由附加装置来填充对象内部的洞隙.



本文提出一套分割算法可有效地分割出视频对象平面 (VOP)。主要有三种方法: 一是基于形态运动滤波和活动轮廓模型的分割方法, 另一是基于运动对象边缘帧差和豪斯道夫对象跟踪的分割方法。此外, 为了解决视频对象仅有局部的偶尔运动, 用形态运动滤波器和边缘帧差图象无法提取出整个初始语义对象的问题, 本文还提出一种基于帧差边缘和灰度边缘之间高度相关性, 通过累积运动信息得到语义对象二值边缘模型而分割前景对象的 VOP 分割算法。

其中第一种方法, 首先根据形态运动滤波器提取出初始对象, 滤波准则为全局支配运动与局部运动的差。标识出初始对象后, 再阈值化和去噪, 就可以提取出初始运动对象和其二值轮廓模型。然后再使用活动轮廓模型在后继帧中跟踪和匹配位置和形状都发生了变化的对象, 最后根据一系列二值轮廓模型, 就可以将序列中的视频对象提取出来。此算法特别适用于前景对象运动不甚剧烈, 静止或运动背景的序列。

第二种方法首先根据边缘帧差图象标识出前景对象, 再使用 Canny 算子得到此前景对象的二值边缘模型, 为了容纳对象在后继帧中的位置变化, 使用经本文改进的豪斯道夫距离跟踪器跟踪此对象, 以修改的二值边缘模型适应对象在后继帧中的形变。最后由一系列二值边缘模型引导提取出视频对象。此算法特别适用于前景对象的运动较激烈和非刚性形变较大而背景是静止的序列。

使用形态运动滤波器或变化检测掩模无法标识轻微运动的语义对象, 例如某些视频会议中的头肩序列的对象。因此我们提出一种基于运动对象边缘和帧差边缘高度相关性的双向运动估计跟踪算法, 且修改运动区域边缘的空间位置, 删除先前帧中帧

差图象的边缘,增加当前帧的帧差图象的边缘,从而形成语义对象的边缘图象,再根据一系列边缘图象提取出运动对象。

本文所提出的算法基于模式识别和对象跟踪原理,因而避免了某些与运动估计相关的问题。实验也证明本文所提出的算法可以处理许多典型的视频序列。最后本文也指出对多视频对象提取和遮挡情况下视频对象提取的问题尚需作更深入探索,才能有效解决。

本文课题是上海市自然科学基金(992D14035)的资助项目。

**关键词** VOP, 形态运动滤波器, 豪斯道夫对象跟踪器, 活动轮廓模型

## Abstract

Traditional video standards such as MPEG-1, MPEG-2, H.261 or H.263 are low-level techniques in the sense that no segmentation or analysis of the scene is required. They can achieve high compression ratios and are suitable for a wide range of applications. However, with the increasing popularity of multimedia applications and content-based interactivity, new video coding schemes are necessary.

The standard MPEG-4, which is currently being developed, enables content-based functionalities by introducing the concept of video object planes (VOP's). Each frame of the input sequence is segmented into arbitrarily shaped image regions (VOP's) such that each VOP describes one semantically meaningful object or video content of interest.

Decomposing a video sequence into VOP's is very difficult in this field. An intrinsic problem of VOP generation is that objects of interest are not homogeneous with respect to low-level features such as color, intensity, or optical flow. Thus, conventional segmentation algorithms will fail to obtain meaningful partitions.

This paper addresses video object plane generation and presents a new algorithm that can automatically extract moving objects from a sequence. Since these objects are characterized by a different motion from that of the background, some type of motion information must be incorporated into the segmentation algorithm.

Optical flow or motion fields could theoretically be used, but they are extremely noise sensitive and their accuracy is limited due to the aperture and occlusion problem. Change detectors or difference images on the other hand mark occlusion areas as changed,

while the objects themselves are unchanged unless they contain sufficient texture. This makes exact boundary location difficult and an additional mechanism is necessary to fill the holes inside objects.

In this thesis we propose a set of algorithms which can efficiently segment sequences into video object planes. The proposed algorithms have three versions: one is based on morphological motion filter and active contour model, and another is based on difference of edges of moving objects and modified Hausdorff distance object tracker. Furthermore, if parts of an object only move occasionally, there is no sufficient motion information using morphological motion filter or difference of edges to identify the whole semantic meaningful object. We propose a new technique which is based on high correlation between edge of frame difference and edge of gray-level image. From collecting enough motion information it can form the binary edge of the whole object and can extract the VOP of the corresponded foreground object.

As for the first VOP segmentation algorithm, the initial object is first extracted using morphological motion filter. The filter criterion is the difference between the global dominant motion and local motion. The initial moving object and its binary contour model is obtained after thresholding initial object and removing the noise. Then the binary contour model is tracked and matched to accommodate the change of position and shape of object using active contour model in the following frames. From a series of binary contour models we can extract the video object from the sequences. This algorithm is much suitable for sequences with little motion and stationary or moving background.

For the second version VOP segmentation algorithm the moving foreground is identified using absolute difference between edge images first, and its binary edge model is obtained by canny

operator . In order to accommodate the change of position of object the translation the object has undergone is indicated by the modified Hausdorff object tracker and a two-stage model update method accommodate changes in shape of the object in the following frames. Finally from a series of binary edge models we can extract the video object. This algorithm is much suitable for sequences with fast motion and stationary background.

It is difficult to identify the whole semantic meaningful object which suffers from very little motion, eg. head-shoulder sequence in video phone system. So we proposed another algorithm to solve it. The proposed algorithm is based on the high level correlation between the edge of moving object and the edge of motion area in frame difference(FD) image. It begins with tracking object by double-direction motion estimation , updating the spatial position of moving object's edge, erasing the FD image correspondence edge of previous frame and adding the FD image correspondence edge of current frame, then forming the binary edge model of the whole semantic meaningful foreground object after several frames. In the end, from a series of binary edge models we can extract the semantic meaningful video object.

Our proposed algorithms are based on pattern recognition and object tracking principles and thereby avoid many of the problems associated with motion estimation. Experiments also demonstrate our proposed algorithm can handle many traditional types of sequences. Finally we also give some ideas on how to extract multi objects and extract whole semantic meaningful occluded object.

This thesis is supported by the National Natural Science Foundation of Shanghai (992D14035).

**Key words** VOP, morphological motion filter, Hausdorff object tracker, active contour model

## 目 录

第一章 绪论 .....	1
1.1 背景和贡献 .....	1
1.2 论文结构 .....	5
第二章 图象的边缘检测和形态分割 .....	7
2.1 引言 .....	7
2.2 边缘检测 .....	8
2.3 基于形态学的静止图象分割 .....	15
第三章 视频对象提取和基于运动信息分割的基础 .....	25
3.1 引言 .....	25
3.2 运 动 .....	29
3.3 二维运动估计 .....	38
3.4 运动分割和视频对象平面提取现状和发展 .....	45
第四章 基于形态运动滤波器和活动轮廓模型 的 VOP 分割算法 .....	62
4.1 基于形态运动滤波器和 Snake 跟踪的 VOP 分 割算法流程图 .....	65
4.2 运动模型的初始化 .....	81
4.3 基于活动轮廓模型(Snake)的对象跟踪匹配器 .....	82
4.4 VOP 提取 .....	100
4.5 实验结果 .....	100
第五章 基于修改的豪斯道夫对象跟踪器的 VOP 分割技术 .....	105
5.1 初始运动对象的检测和二值模型的建立 .....	106

---

5.2 模型初始化 .....	108
5.3 基于经典的豪斯道夫距离的模型跟踪及其改进 .....	109
5.4 模型修改 .....	116
5.5 背景滤波 .....	118
5.6 VOP 提取 .....	119
5.7 实验结果 .....	120
<b>第六章 对仅含局部运动区域的头-肩序列中运动</b>	
<b>对象的分割提取 .....</b>	<b>123</b>
6.1 头-肩序列 VOP 分割算法 .....	127
6.2 填入技术 .....	129
6.3 实验结果及分析 .....	139
<b>第七章 结论和进一步的工作 .....</b>	<b>142</b>
<b>参考文献 .....</b>	<b>144</b>
<b>致 谢 .....</b>	<b>159</b>

# 第一章 绪 论

## 1.1 背景和贡献

数字图象处理, 在过去 20 年中是一个相当活跃且发展十分迅速的领域. 现在数字图象处理技术在商业、工业和医学等应用领域正发挥着日益重要的作用.

分割历来是数字图象处理中的难题之一. 传统的分割是指把图象分成具有相似属性的区域, 基本属性包括亮度、彩色和光流等. 这样的基本分割区域往往和物理对象没有一一对应关系, 所以一般把这种根据基本属性进行分割称为基于区域相关或基于特征的分割, 在某种意义上也可归为“低级”分割.

当图象必须被分割成若干个有语义意义的实体(物理对象)时, 就需要上下文信息或某些人工智能形式的信息, 因此也被称之为“高级”分割. 相比于“低级”分割, 更增加了分割难度, 研究得也相对要少.

应根据具体应用决定是否进行“低级”或“高级”分割. 这些应用包括: 医学图象处理中的血管检测、机器人学中的对象识别、图象与视频编码、图象恢复、工业产品探伤、地质学中的制图、资源勘探、市区管理和军事遥感成像等等.

目前两种最流行的低级分割是基于数学形态学<sup>[1-3]</sup>和贝叶斯决策<sup>[4-6]</sup>的分割, 形态学技术是研究动植物形式和结构的生态学



的一个分支. 在图象处理和计算机视觉中, 使用形态学技术来研究图象的拓扑结构. 这就是众所周知的面向形状的图象处理方法, 与之相反的, 即为面向频谱的方法.

Serra<sup>[7]</sup>对数学形态学的研究做出了许多贡献, 这包括早期一些公式的定义. 形态分割技术主要包括三个主要的步骤: 图象简化、标志提取和分水岭算法, 在初始阶段可获得良好的分割效果. 其缺陷是分割时在加强空域连续性上缺乏约束力.

贝叶斯分割对未知分割进行最大后验概率估计(MAP). 基于这个目的, 将分割标签场和图象假设为二维随机场的样本值. 标签场通常建模为马尔可夫随机场(MRF). 尽管在物理系统中使用 MRF 描绘空间交互可以溯及到 20 世纪 20 年代<sup>[8]</sup>, 但直到 1974 年它才投入实际应用<sup>[9]</sup>. 自提出 Hammersley-Cliffors 定理(阐述了 MRF 和吉布斯随机场的对偶原理)后, 人们才有可能根据派势函数而确定 MRF. 随着计算机运算能力的增加, 贝叶斯分割技术在 80 年代得到迅速的发展.

贝叶斯分割技术由于合并了空域连续性约束而优于形态技术. 但需要初始估计且对分割结果起决定作用的输入参数  $K$ (指定所使用的标签数目)有强烈依赖性. 是贝叶斯分割算法的缺点. 从整体说, 形态分割和贝叶斯分割间互补性很强.

图象和视频分割在高性能的图象和视频压缩中的作用越来越重要. 众所周知, 图象和视频编码的目的是通过删除冗余数据来达到以最少的比特数来表征图象或序列. 离散余弦变换 (DCT)<sup>[10]</sup>通过丢弃被认为不甚重要的高频系数而达到相当高的压缩率. 现行的图象视频标准例如 JPEG<sup>[11,12]</sup>、MPEG-1<sup>[13]</sup>、MPEG-2<sup>[14]</sup>、H.261<sup>[15]</sup>和 H.263<sup>[16]</sup>都使用基于 DCT 的技术. 其主要缺点是在甚低比特率下沿着块边界引入了可见的不连续性, 通