

# 一种时域和梯度域相结合的视频对象提取算法

封春升 郝爱民 何兵 王莉莉

(北京航空航天大学虚拟现实新技术教育部重点实验室, 北京 100083)

**摘要** 为了从复杂背景中精确地提取出视频对象,提出了一种融合时域和梯度域信息的视频对象提取算法,可以有效地提取出复杂背景下的视频运动对象,并解决前景与背景一致情况下,背景剔除方法所带来的空洞问题。首先在时域空间中分别采用背景剔除和帧间差分方法生成初步的视频对象,并利用形态学中的二值腐蚀和膨胀方法对视频对象进行处理;然后,在梯度域空间中用 Sobel 算子进行视频对象边缘检测,并结合时域空间中的视频对象,生成精确的视频对象轮廓边缘;最后,采用启发式搜索方法连接视频对象轮廓边缘点,进而提取出视频对象。实验结果表明,该方法能够比较完整精确地从复杂背景中提取出视频对象。

**关键词** 视频对象提取 背景剔除 边缘检测 边缘连接

中图分类号: TP391.9 文献标识码: A 文章编号: 1006-8961(2008)03-0494-05

## A New Algorithm for Video Object Extraction Based on Temporal Domain and Gradient Domain

FENG Chun-sheng, HAO Ai-min, HE Bing, WANG Li-li

(The Key Laboratory of Virtual Reality Technology, BeiHang University, Beijing 100083)

**Abstract** A novel approach for video object extraction fusing temporal and gradient information is proposed. Experiments demonstrate that our method can extract the video object effectively from complex background, and can resolve the vacancy phenomenon when the foreground and background are of the same color using background eliminate method. Firstly, in the temporal domain we introduced background eliminate and adjacent frame difference method to generate elementary video object, and then we used erosion and dilation morphologic method to the video object. Secondly, in the gradient domain we detected video object's edge using Sobel operator. Associating with the video object generated in the temporal domain, we obtained the exact video object contour's edge. Finally, we applied heuristic search method to link video object contour's edge points, sequentially extracted video object.

**Keywords** video object extraction, background eliminate, edge detection, edge connection

## 1 引言

视频对象(video object, VO)是由MPEG-4引入的概念,它将广泛地应用于诸如新一代交互式电视、因特网流视频等多媒体通信系统中。视频对象提取是通过视频对象分割获得的,不仅是实现

MPEG-4<sup>[1]</sup>基于内容的视频编码的前提条件,也是实现基于内容的视频检索和交互式多媒体应用所必需的。视频对象分割技术,按分割过程中所利用信息的不同,主要可分为时域分割、空间分割和时空联合分割。

因为大多数运动对象具有跟背景不同的运动,时域分割通常利用运动信息主要有基于光流法的分

**基金项目:**国家高技术研究发展计划 863 项目(2004AA115130);国家重点基础研究发展计划 973 项目(2002CB312105);国家自然科学基金重点项目(60533070)

**收稿日期:**2006-06-24; **改回日期:**2006-11-06

**第一作者简介:**封春升(1981 ~ ),男。现为北京航空航天大学计算机应用专业硕士研究生。主要研究方向为计算机视觉、虚拟现实、计算机图形学。E-mail: talentfcs@163.com

割和基于变化检测的分割。基于光流法的分割主要是基于光流场的运动参数估计,求出符合运动模型的像素区域,进而合并区域构成运动对象进行视频分割。Yan 等人利用统计学的原理,采用二乘方中值最小法和多项式曲面拟和方法,进行光流估计和视频对象提取<sup>[2]</sup>。基于变化检测的分割使用差分或运动补偿的变化检测算法,对背景静止或者场景只有简单全局运动的视频图像进行分割。Mech 等人利用估计出的变化检测模板来生成视频对象平面<sup>[3]</sup>。

空间分割通常利用图像颜色和纹理特征信息进行分割。主要有基于特性的分割、基于边缘的分割和基于区域的分割。基于特性的分割,由用户按纹理特性选择一些中心像素点,计算机将余下像素按所选中心进行分类,形成不同的对象。基于边缘的分割,是利用边缘检测算子生成边缘图像,再根据边缘图像提取边缘对象。基于区域的分割,则将图像分割为不同的同一性区域,通过组合不同的区域来得到视频对象<sup>[4]</sup>。

时空联合分割是将视频看作像素的时空块,同时考虑时间和空间维进行的分割。Greenspan 等人把每个像素与特征空间的一个点相对应<sup>[5]</sup>,像素点通过 EM (the expectation-maximization algorithm) 框架下的高斯混合模型来聚类,聚类的最优解使用 MDL (minimum description length) 来寻找。在模型学习后,空间和时间维之间的高斯协方差系数就能给出运动的信息。采用时空联合分割方法既利用了

帧间时域上的强相关性信息,又利用了帧内空间上的特征,能够取得较好的分割效果。

本文算法融合了时域分析中背景剔除、帧间差分方法和梯度域分析中边缘检测的思想。首先在时域空间中采用中值灰度方法生成背景图像,用背景差分方法提取初步的视频对象并进行二值腐蚀操作,用帧间差分方法生成初步的视频对象轮廓并进行二值膨胀操作;然后,在梯度域空间中用 Sobel 算子进行视频对象边缘检测,并结合时域空间中形态学方法处理后的视频对象,生成精确的视频对象轮廓边缘;最后,采用启发式搜索方法连接视频对象轮廓边缘点,进而提取出视频对象。实验结果表明,本文方法可以有效地提取出复杂背景下的视频运动对象。

## 2 时域分割

### 2.1 背景估计

传统背景估计的方法,采样平均估计器来生成初始化图像。这种算法的原理是取一段时间内像素值的平均值作为背景。 $B(X, Y) = \frac{1}{t} \sum_{i=1}^t C_i(X, Y)$ , 其中  $B(X, Y)$  表示背景图像在  $(X, Y)$  坐标点上的颜色值,  $C_i(X, Y)$  表示第  $i$  帧图像上  $(X, Y)$  坐标点上的颜色值。但这种算法存在明显的缺点,会在运动目标的整个运动轨迹上留下运动的模糊,如图 1 所示。

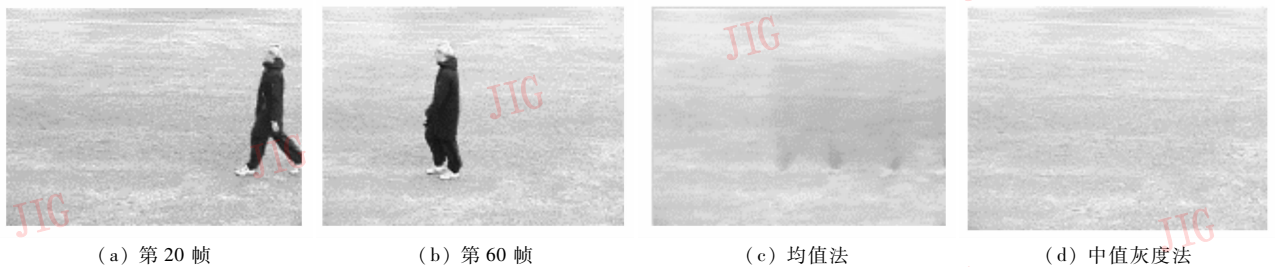


图 1 背景估计方法对比

Fig. 1 Background extraction methods contrast

本文采用中值灰度的方法,来生成背景图像。其主要原理是:计算一段时间内,各像素点的灰度中值,根据灰度中值的像素点生成背景图像。主要步骤如下:

(1) 计算  $t$  帧内,每一帧图像中各像素的灰度值。

$$G_i(X, Y) = 0.299R + 0.587G + 0.114B$$

式中,  $G_i(X, Y)$  表示第  $i$  帧图像中  $(X, Y)$  像素点的灰度值。  $R, G, B$  分别表示像素点的红色、绿色、蓝色分量。

(2) 对于每一个像素点,计算第 1 帧到第  $t$  帧中此像素的灰度中值。

$$G(X, Y) = \text{median} \{ G_1(X, Y), G_2(X, Y), \dots, G_t(X, Y) \}$$

取灰度中值的像素点的颜色值作为背景图像中对应像素点的颜色值,从而获得背景图像  $B(X, Y)$ 。

采用中值灰度的方法,可以很好地解决均值法中因为混合了运动目标而使背景图像模糊的问题,实验结果如图 1(d) 所示。

## 2.2 形态学处理

形态学处理利用了点集的性质、积分几何的结果和拓扑学原理,它的基础是作用于物体形状的非线性算子的代数。本文首先利用形态学中的二值膨胀方法对帧间差分得到的视频对象进行粗画操作,然后利用二值腐蚀方法对背景差分得到的视频对象进行细化操作,从而可以消除视频对象外部和内部的边缘。

二值膨胀(又称填充或生长),是一种各向同性的扩张,即在所有方向上行为相同。其作用是将所有与物体邻近的背景像素变为物体像素。首先,采用时域中的帧间差分思想,用下一帧减去当前帧来获得视频对象  $O_i(X, Y) = C_{i+1}(X, Y) - C_i(X, Y)$ ;然

后对得到的视频对象  $O_i(X, Y)$  进行二值化处理,得到视频对象的二值图像  $\hat{O}_i(X, Y)$ ,如图 2(c) 所示;最后对二值图像  $\hat{O}_i(X, Y)$  采用二值膨胀运算:

$$X \oplus B = \{ p \in \mathcal{E}^2 : p = x + b, x \in X \text{ 且 } b \in B \}$$

式中,  $X, B$  表示两幅图像,  $p$  表示像素,  $\mathcal{E}^2$  表示像素空间,  $x$  表示图像  $X$  的像素,  $b$  表示图像  $B$  中的像素。获得的粗绘的视频对象  $O_i^{(2)}(X, Y)$  如图 2(d) 所示。

二值腐蚀(又称收缩或缩小),是一种各向同性的收缩,其作用是通过减少物体轮廓的像素来缩小图像。利用上文中值灰度法计算得到的背景图像  $B(X, Y)$ ,采用背景差分思想,用当前帧减去背景图像来获得视频对象  $O_i(X, Y) = C_i(X, Y) - B(X, Y)$ ;然后对得到的视频对象  $O_i(X, Y)$  进行二值化处理,得到视频对象的二值图像  $\hat{O}_i(X, Y)$ ,如图 2(e) 所示;最后对图像  $\hat{O}_i(X, Y)$  采用二值腐蚀运算:

$$X \ominus B = \{ p \in \mathcal{E}^2 : p + b \in X, \text{ 对每一个 } b \in B \}$$

从而获得细画的视频对象  $O_i^A(X, Y)$ ,如图 2(f) 所示。

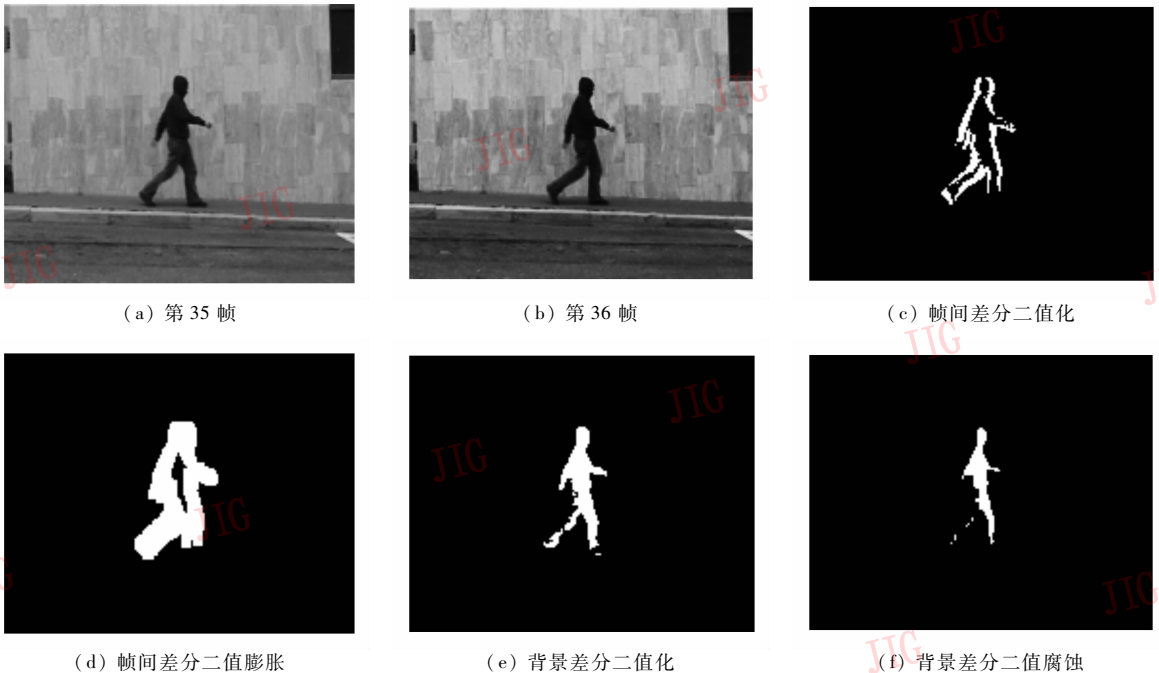


图 2 形态学二值处理结果

Fig. 2 Images dealing with morphology

## 3 梯度域分析

### 3.1 边缘检测

如果一个像素落在视频图像中某一个物体的边

界上,那么它的邻域将成为一个灰度级变化的带。利用这种梯度域信息,采用 Sobel 边缘检测算子<sup>[6]</sup>,获得边缘图像  $E_i(X, Y)$ ,然后进行二值化处理得到  $\hat{E}_i(X, Y)$ 。由于 Sobel 边缘检测算子是两个  $3 \times 3$  的卷积核,因而能够更好的处理灰度渐变和噪声较多

的图像。

### 3.2 结合时域分割的边缘确定

为了更精确的得到视频对象轮廓的边缘,结合时域空间中获得的图像,对梯度域空间中获得的边缘检测后的二值图像  $\hat{E}_i(X, Y)$  进行处理。首先用边缘检测图像减去时域空间中得到的细绘的视频对象  $O_i^{(1)}(X, Y)$ , 获得消除了视频对象内部边缘的图像  $E_i^1(X, Y) = \hat{E}_i(X, Y) - O_i^{(1)}(X, Y)$ 。

然后用消除了视频对象内部边缘的图像  $E_i^1(X, Y)$  乘以时域空间中得到的粗绘的视频对象边缘图像  $O_i^{(2)}(X, Y)$  来除去视频对象外部的边缘,从而获得精确的视频对象轮廓的边缘图像  $E_i^0(X, Y) = E_i^1(X, Y) \times O_i^{(2)}(X, Y)$ 。

### 3.3 边缘连接

由于边缘检测得到的视频对象轮廓的边缘并不是闭合连通的,所以需要进行边缘点连接,将邻近的边缘点连接起来,从而产生一条闭合连通的视频对象轮廓,进而获得视频对象。

本文采用启发式搜索方法连接边缘点,从一个边缘点  $A$  开始,搜索以点  $A$  为中心的  $7 \times 7$  邻域,在邻域像素点中根据离点  $A$  距离由近至远,依次判断像素点是否存在边缘点。如果是,则连接点  $A$  和此像素点  $B$ ,填充  $AB$  直线上的像素点为边缘像素。然后,根据  $AB$  直线的方向作为启发信息,优先从此方向上判断是否存在边缘像素,来连接其他的边缘点。如图3所示,点  $C$  和点  $D$  都是边缘点,则根据启发式搜索点  $B$  连接的下一个边缘点是点  $C$ ,而不是点  $D$ 。

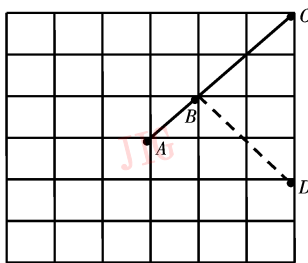


图3 启发式搜索边缘连接

Fig. 3 Edge connection using heuristic search method

## 4 视频对象提取

通过边缘连接得到的视频对象轮廓是一个闭合曲线(记为  $C$ ),通过以下方法来获得视频对象内部区域:

(1) 获得闭合曲线  $C$  的拐点集合  $G$ 。记  $C_{i-1}$ ,  $C_i$ ,  $C_{i+1}$  是曲线  $C$  上相邻的3个像素点,  $Y(C_i)$  为  $C_i$  点在图像纵轴上的坐标值。如果  $Y(C_i) < Y(C_{i+1})$  且  $Y(C_i) < Y(C_{i-1})$  或者  $Y(C_i) > Y(C_{i+1})$  且  $Y(C_i) > Y(C_{i-1})$ , 那么  $C_i$  属于闭合曲线  $C$  上的拐点集合  $G$ 。

(2) 获得闭合曲线  $C$  的拐点水平直线段和非拐点水平直线段。记  $C_i, C_{i+1}, \dots, C_{i+j-1}, C_{i+j}$  是曲线  $C$  上相邻的像素点。如果  $Y(C_{i+1}) = Y(C_{i+2}) = \dots = Y(C_{i+j-1}) < Y(C_{i+j})$  且  $Y(C_{i+1}) = Y(C_{i+2}) = \dots = Y(C_{i+j-1}) < Y(C_i)$ , 或者  $Y(C_{i+1}) = Y(C_{i+2}) = \dots = Y(C_{i+j-1}) > Y(C_{i+j})$  且  $Y(C_{i+1}) = Y(C_{i+2}) = \dots = Y(C_{i+j-1}) > Y(C_i)$ , 则直线段  $C_i C_{i+j}$  为拐点水平直线段。否则直线段  $C_i C_{i+j}$  为非拐点水平直线段。

(3) 对于图像中不属于闭合曲线  $C$  上的任意一点  $A$ , 过点  $A$  做  $X$  轴水平线  $L$ , 记录所有与闭合曲线  $C$  相交的点的集合  $J$ 。如果水平线  $L$  穿过拐点水平直线段, 则从集合  $J$  中删除拐点水平直线段上的点; 如果水平线  $L$  穿过非拐点水平直线段, 则集合  $J$  中只保留此非拐点水平直线段中的任意一点, 删除其他的点。最后, 再从集合  $J$  中, 删除属于拐点集合  $G$  中的点, 得到集合  $\hat{J}: \hat{J} = J - J \cap G$ 。

(4) 如果集合  $\hat{J}$  中有奇数个点的横坐标大于点  $A$  的横坐标, 并且有奇数个点的横坐标小于点  $A$  的横坐标, 则点  $A$  属于闭合曲线  $C$  的内部区域, 即视频对象区域。否则点  $A$  属于闭合曲线  $C$  的外部区域。

## 5 实验结果及分析

为了验证本文算法的有效性, 对一个741帧的人体行走图像序列进行分析得到图4所示的结果。从实验结果可以看出, 本文算法可以很好的获得视频对象轮廓, 从而提取出视频对象。此方法与传统的背景差分方法相比, 能够更好的处理因视频对象局部区域与背景图像一致而产生的空洞问题, 如图4(b)中手和头部的空洞。

## 6 结论

本文提出了一种时域和梯度域相结合的视频对象提取方法, 综合利用了多帧图像时间域中的像素运动信息, 和空间域中的梯度信息, 能够从存在运动的复杂背景中提取视频对象。这种方法在视频对象



(a) 第 81 帧图像

(b) 中值灰度法背景差分获得的  
前景图像

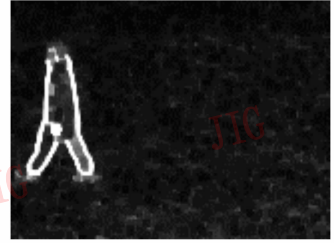
(c) 对图(b)进行二值腐蚀后的图像



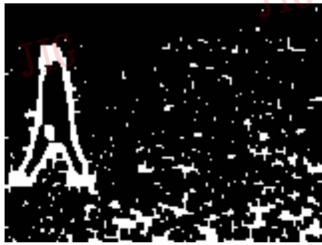
(d) 帧间差分法获得的前景图像



(e) 对图(d)进行二值膨胀后的图像



(f) Sobel 边缘检测后的图像

(g) 图(f)二值化减去图(d)后的图像  
(除去了视频对象内部的边缘)(h) 图(g)乘以图(e)后的图像  
(除去了视频对象外部的边缘)

(i) 对图(h)进行边缘连接的结果

图 4 本文算法实验结果

Fig. 4 The experiment result of this paper's algorithm

提取过程中不需要人机交互,能够自动完成。另外,由于利用了图像边缘梯度信息,能较好地解决视频对象局部区域和背景一致时,采用背景差分方法产生的空洞现象。并能解决视频对象运动速度较快的情况下,帧间差分方法分离出来的视频对象沿着运动方向拉伸过长的的问题,如图 4(d)中白色区域会变大。实验结果表明,本文提出的方法能够比较完整精确的从复杂背景中提取出视频对象。

在本文算法的基础上,可以进一步采用阴影的亮度和色度信息特征来处理带有阴影的视频对象提取,使用全局运动估计方法来处理背景运动的视频序列中的视频对象提取,从而使本算法有更好的通用性。

### 参考文献 (References)

1 MPEG Video Group. Overview of the MPEG-4 standard [S]. ISO/IEC JTC1/SC29/WG11 N 4668, 2002.

2 Yan Hong-shi, Tjahjadi T. Optical flow estimation and segmentation through surface fitting and robust statistics [A]. In: Proceedings of IEEE International Conference on Systems, Man and Cybernetics [C], Washington, DC, USA, 2003, (2): 1390 ~ 1395.

3 Roland Mech, Michael Wollborn. A noise robust method for 2D shape estimation of moving objects in video sequences considering a moving camera [J]. Signal Processing, 1998, 66(2): 203 ~ 217.

4 Tsaig Y, Averbuch A. Automatic segmentation of moving objects in video sequences: a region labeling approach [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2002, 12(7): 597 ~ 612.

5 Greenspan H. A probabilistic framework for spatio-temporal video representation and indexing [A]. In: Proceedings of European Conference on Computer Vision [C], Berlin, Germany, 2002: 461 ~ 475.

6 Zhu Zhi-gang. Digital Image Processing [M]. Beijing: Publishing House of Electronics Industry, 2004: 389 ~ 391. [朱志刚. 数字图像处理 [M]. 北京: 电子工业出版社, 2004: 389 ~ 391.]