

一种基于时空联合的视频对象分割方法

李朝晖^{1),2)} 张艳玲²⁾ 余英林¹⁾

¹⁾(华南理工大学电子与通信工程系, 广州 510641) ²⁾(广州大学信息学院计算机系, 广州 510405)

摘要 视频对象分割在基于内容的视频编码和视频检索中均有重要的应用。为此, 针对视频对象分割, 提出了一种时域和空域信息融合的视频对象分割方案, 该方案首先对时域分割采用基于F-假设检验的方法来得到初始的变化检测模板, 然后通过与基于形态学的空域分割融合来获得最终的运动对象。实验结果表明, 该方案计算比较简单, 能较好地前景运动对象从静止或运动、简单或复杂的背景中分离出来, 且定位精度较好。

关键词 图象处理(510·4050) 视频分割 F-检验 形态学

中图法分类号: TP391.4 TN941.1 **文献标识码:** A **文章编号:** 1006-8961(2003)11-1341-05

A Kind of Video Object Segmentation Based of Spatio-temporal Information

LI Zhao-hui^{1),2)}, ZHANG Yan-ling²⁾, YU Ying-lin¹⁾

¹⁾(Dept. of Comm. and Elec. Engineering, South China University of Technology, Guangzhou 510641)

²⁾(Dept. of computer, Guangzhou Normal University, Guangzhou 510405)

Abstract Video object segmentation has important application in content-based video encoding and video retrieval. In this paper, a new approach based on temporal and spatial union information is proposed. In spatial domain, morphological operators such as morphological filters, morphological gradient operator and watershed algorithm are used, morphological filters are used to simplify the image, morphological gradient is used to mark homogeneous regions, the watershed operator is used to decide spatial boundaries. In temporal segmentation domain, the initial change detection mask is gotten by using F-test, then by combining it with the result of spatial segmentation, the final result is gotten. The procedure toward complete segmentation consists of the following main steps: global motion estimation and compensation (if exist background motion), change detection, marker extraction, watershed transfer, temporal and spatio information fusion. The experimental result shows the effectiveness of the method; it can separate the foreground from still or motion, simple or complex background satisfactorily, preferably locate and segment the video object.

keywords Video segmentation, F-detection, Morphological operator

0 引言

视频对象分割是指从视频中分割出具有一定意义的实体, 这种有意义的实体在数字视频中称为视频对象 VOP (Video Objects Plane)。如今, 视频对象分割在很多领域已有很重要的应用, 如在 MPEG-4 的基于内容的视频编码中, 提取视频对象可极大地提高压缩效率; 又如, 在 MPEG-7 中, 提取视频对象可实现基于内容的视频检索; 此外, 视频分割还在其他方面, 比如模式识别、计算机视觉等领域也得到了

广泛的应用。

然而, 到目前为止, 将视频序列分解成视频对象仍是一个难点, 也是目前的研究热点之一。视频对象分割常用的方法有基于时域的分割方法^[1], 即可单纯根据运动信息的均一性来划分区域和估算图象的运动场, 并将具有相似运动的区域聚类, 以形成视频对象。对运动对象进行分割可采用基于参数模型或非参数模型的统计理论, 如贝叶斯估计方法等。用这种方法来分割, 对运动场的估计要足够的准确, 可是, 运动估计本身就是一个亟待解决的研究课题。由于运动场噪声的影响, 以及遮挡和孔径等问题, 很难

准确逼近运动物体的边缘,为此,在近几年的研究中^[2,3],人们趋向于结合视频对象的一些空间特征信息^[4,5],比如颜色、亮度、边缘、纹理等特征来进行视频对象分割。本文提出一种时域和空域信息的融合方案,即首先时域分割采用基于F-假设检验来得到初始的变化检测模板,然后与基于形态学的空域分割融合来获得最终的运动对象。实验结果表明,该方案计算比较简单,能较好地前景运动对象从背景中分离出来,且定位精度较高。

1 算法原理

1.1 时域分割

时域分割的基本思想是先对连续两帧图象进行变化检测,然后通过比较前景和背景是否属于同一分布来得到运动对象的模板。

对于背景静止的视频序列,通常说来,由于对象的运动使连续帧之间的亮度发生变化,因而,亮度变化是定位时间和空间运动物体的重要线索,通常用变化检测模板 CDM(Change Detection Mask)来表示,然而,由于成像设备的热噪声和量化噪声也可能引起亮度改变,即背景噪声,因此分割的任务就是要从背景噪声中检测出运动对象,即从背景噪声中区分出由对象运动所引起亮度改变的像素,也就是说,运动检测就是将序列帧相对前一帧标注为“变化的”和“未变化”的区域;对于非静止背景的视频序列,由于前景和背景都存在运动,因此,可首先对背景的运动进行估计和补偿,即进行全局运动估计,并经过补偿背景的移动,使连续两帧视频图象背景基本相同;然后用如下的步骤来检测运动对象:

首先计算两帧图象灰度之差,假定观察噪声服从均值为零的高斯正态分布,而由运动物体引起的变化与噪声独立同分布,即服从均值未知的高斯分布。由统计学原理可知,帧差图象服从两个正态总体的混合分布。若以大小为 n 的观察窗口对帧差图象进行取样,则可得到 n 个随机样本 $\{x_1, x_2, \dots, x_n\}$ 。由先验知识可知,运动物体引起的方差变化要显著大于由噪声引起的背景的方差变化。为此,取假设检验为

$$\begin{aligned} H_0: \sigma_1^2 &= \sigma_2^2 \\ H_1: \sigma_1^2 &< \sigma_2^2 \end{aligned} \quad (1)$$

零假设 H_0 是指若检测区域为“未变化”区域,则其背景方差 σ_1^2 应与观察窗内方差 σ_2^2 相等。备择假设 H_1 表示检测区域是“变化了”的区域,即如果观

察窗滑过运动对象区域,那么观察窗内的方差 σ_2^2 应大于背景方差 σ_1^2 。

设 $\hat{\sigma}_1^2$ 和 $\hat{\sigma}_2^2$ 分别是 σ_1^2 和 σ_2^2 的估计值, n_1 和 n_2 为样本数,则随机变量 $\hat{\sigma}_1^2/\sigma_1^2$ 和 $\hat{\sigma}_2^2/\sigma_2^2$ 服从自由度分别为 $n_1 - 1$ 和 $n_2 - 1$ 的 χ^2 分布,因而统计测试

$$V = \frac{\hat{\sigma}_2^2/\sigma_2^2}{\hat{\sigma}_1^2/\sigma_1^2} \quad (2)$$

服从自由度为 $n_1 - 1$ 和 $n_2 - 1$ 的 F 分布。

在零假设条件下,由于 $\sigma_1^2 = \sigma_2^2$,式(2)就变成 $V = \hat{\sigma}_2^2/\hat{\sigma}_1^2$,说明不必用真实的方差作为先验知识来计算 V ,因此,可设置一阈值 $V_{\text{threshold}}$ 来决定是否接受 H_0 判决,若 $V < V_{\text{threshold}}$,则接受 H_0 ;若 $V > V_{\text{threshold}}$,则 H_0 被拒绝;

根据上述分析,整个时域分割算法的步骤如下:首先计算得到连续两帧图象的帧间差图象,并取其绝对值;同时依次扫描每个像素点,并以其为中心点,构造边长为 $m \times m$ 的观察窗口;然后,分别计算观察窗口和选择的背景区域各自的方差;最后再与一个预设的阈值进行比较,如果大于该阈值,则将该像素点标记为运动点,否则为背景点。采用上述方法对每个像素进行同样处理,直到所有点都扫描完为止。处理时,通常取图象上下左右离边框 10 个像素左右的区域作为背景区域,因为该区域出现运动物体的概率较低。经过假设检验后,就可以得到初始的时域分割模板了。

用 F 检验得到初始模板后,就可根据变化检测的一些先验知识来进行滤波。因为一般说来,属于运动物体的像素都是 8-连通的,而由噪声引起的变化一般都是随机孤立的点或小区域,所以可以对结果中面积大于一定门限的变化区域进行标识,以消除噪声对结果的影响和保留大的运动区域。

1.2 空域分割

经过以上分割得到的时域分割模板,虽能够标记出运动对象的大概位置,但是由于边缘定位不精确,不能得到精确的分割对象,因此还需与空间分割结果相结合,才能得到比较精确的分割结果。本文的空间分割是采用数学形态学的方法,它包括以下几个部分:

(1) 图象简化

对图象进行简化的目的是为了去掉小的噪声干扰,并去除掉对感知不重要的细节,即对图象进行平滑处理,以便于分割。分割可采用按面积大小分割和按对比度分割两种准则。其中,按面积大小分割是指

分割出所有面积均大于给定门限的区域;按对比度分割则是指提取对比度强的区域. 图象简化常用的形态滤波器有开、闭重构滤波器、区间开、闭滤波器以及交替顺序的滤波器等. 它们可通过去除面积小的区域或对比度弱的区域来简化信号,同时使轮廓信息保持.

实验首先使用高斯滤波器来对图象进行平滑处理,然后采用开闭重构滤波器来对平滑处理后的图象进一步平滑. 设 f 表示一个 N 维信号, γ 表示形态开, φ 表示形态闭,则形态开、闭重构的定义如下:

形态开重构

$$\gamma^{(\text{rec})}(\gamma_n(f), f) = \dots \gamma^{(1)}(\dots \gamma^{(1)}(\gamma_n(\gamma f), f) \dots f) \quad (3)$$

形态闭重构:

$$\varphi^{(\text{rec})}(\varphi_n(f), f) = \dots \varphi^{(1)}(\dots \varphi^{(1)}(\varphi_n(\gamma f), f) \dots f) \quad (4)$$

其中, (rec) 代表重构.

(2) 标记提取

标记提取的目的是为了对图象内的平坦或均匀区域作标记. 该区域可用形态梯度来标识,由于均匀区域内的内部梯度较低,而区域边缘的梯度值则较大,因此形态梯度可表达为

$$g = \delta_1(f) - \varepsilon_1(f) \quad (5)$$

其中, δ_1 表示对图象进行膨胀运算, ε_1 表示腐蚀运算. 对于人类视觉系统来说,对高对比度的小区域的敏感性要高于对低对比度的大区域的敏感性,由于基于面积的形态滤波器有可能去掉了前者,而保留了后者,所以可采用基于对比度的方法来进行标记提取.

(3) 分水岭变换

经标记提取后,虽然已知道要分割区域的数目和区域内部,然而,大量的像素还没有归到某一给定区域,即所谓的不确定区. 这些像素主要集中在区域边界附近. 其判决过程就是将这些像素归到某一给定区域,以便准确定位区域之间的边界,而经典的形态判决工具则是分水岭法(Watershed). 分水岭变换的过程实际上就是一个在梯度图象上不断“灌水”的过程,即首先对最低集水盆灌水,设定灌水高度为 H ,然后对每一个相应极小值小于或等于 H 的集水盆分配一个唯一的标记,再检测梯度值为 $H+1$ 的像素,如果该像素周围有已经标记过的像素,则给它分配相同标记;如果该像素周围没有一个标记过的像素,就给它分配一个新的标记,并作为一个新的集水盆,直到每一个像素都分配到一个标记为止,就完成了分水岭的变换.

1.3 时空融合

经过以上的步骤,即已得到空域分割和时域分割模板,由此可见,空间分割可准确地定位物体的边

沿,而时间分割则可以大致地确定前景和背景的分割模板,如果将时空分割结果通过空间投影融合起来,那么就得到视频对象,但在做融合之前,还需对时域分割得到的初始模板进行形态变换处理,才能用于融合过程. 因为变化检测模板,只能将其中变化大的区域以及边缘和纹理区域标记出来,而不一定能将整个对象标记出来,即对象的边界不一定是连续的,而对象内部也有待于填充,所以可首先通过形态膨胀来实现区域闭合,然后进行区域填充,最后将处理后的时域分割模板与空间分割模板进行融合. 该融合过程是以当前帧空间分割得到的区域为处理对象来将当前帧的时域分割结果投影到空间区域. 如果时域分割得到的模板中像素的投影落在该区域内的比率超过一定门限,则将该区域作为运动物体的一部分. 该比率为

$$P = \frac{A_{\text{space}}(i) \cap A_{\text{time}}}{A_{\text{space}}(i)} \begin{cases} \geq \tau & G_F \\ < \tau & G_{\text{Back}} \end{cases} \quad (6)$$

其中, $A_{\text{space}}(i)$ 表示第 i 个空间分割区域的面积, A_{time} 为时域模板面积, G_F 为前景, G_{Back} 为背景. 实验所取门限为 0.6 左右. 如果上述比率 P 小于门限值,则将该区域与前一帧的分割结果相比较,如果前一帧运动物体在某一区域的投影占该区域中像素的比例超过 0.9,则将该区域当作当前帧运动物体的一部分. 这样就可以追踪运动物体在帧间的变化,以弥补因时域分割噪声估计不准而导致的区域缺失.

用上述的方法来检测运动物体,虽然具有较好的抗噪鲁棒性,但在实际检测中,由于总会有误警的现象发生,从而会导致一些小的背景区域被当作前景区域误检出来. 为此,可采用形态连接算子来滤除小的干扰区域. 这样整个分割过程结束后,就得到最终的分割结果.

2 实验结果及分析

为了验证本文算法的效果,选择 MPEG-4 的几组 QCIF 格式、大小为 176×144 的测试图象进行了实验. 图 1~图 3 是 Akiyo、Hall objects、Mother and daughter 序列的部分实验结果. 图 1(a)~图 1(d)、图 2(a)~图 2(d)、图 3(a)~图 3(d) 分别为 Akiyo 序列第 60 帧、Hall objects 序列第 60 帧、Mother and daughter 序列第 64 帧的原图、时域分割初始模板、空域分割模板和最终的分割结果图. 从实验结果来看,时域分割能大致地定位运动对象,但存在一定

的噪声和孔洞;空域分割能比较精确地定位对象边界,但存在“过分割”的问题.若将两者结合起来,并进行一定的形态运算,则能得到比较好的分割结果.

图 4 是跟踪 Akiyo 序列其中第 80、第 97、第 100 帧的分割结果.由图 4 可以看到,前景运动对象能较好地 从背景中分割出来,且定位精度较好.

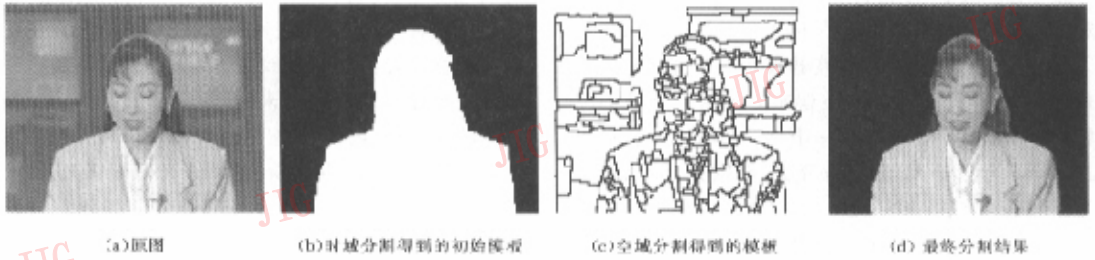


图 1 Akiyo 第 60 帧分割结果

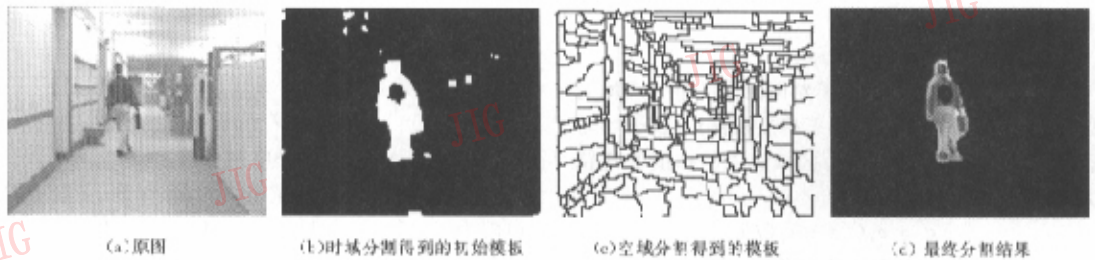


图 2 Hall objects 第 60 帧分割结果

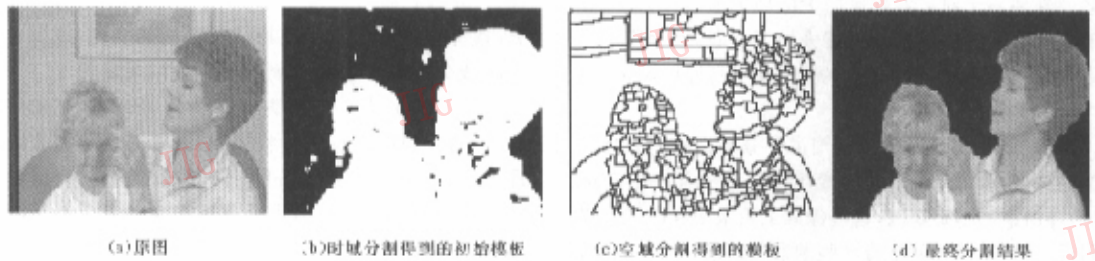


图 3 Mother and daughter 第 64 帧分割结果



图 4 Akiyo 图象第 80 帧、97 帧、100 帧分割结果(上图是原始图象,下图是分割结果图象)

3 结 语

本文针对视频序列对象分割,提出了一种时空相结合的分割方案,其中,时域分割采用基于 F-假设检验来检测运动对象,空域分割则利用形态算子和分水岭算法进行分割,然后将时空分割结果融合起来,并通过进行适当后处理来获得最终的分割对象.模拟实验结果表明,该方案能较好地分割运动对象,并具有计算简单、分割精度较高等特点,可适应于背景静止或有全局运动的视频对象分割,但实验中发现一个值得注意的问题是:背景噪声方差估计的精确程度对分割结果有一定的影响,就目前来说,噪声的估计还没有什么很好的算法,一般只能是根据先验知识,将某一些背景常出现的区域作为噪声的样本进行估计.另外,由于形态算子的大小和形状对实验结果也有一定影响,考虑到边缘一般为曲线,因此取形态结构元素的形状为圆形,其大小一般取 3×3 或 5×5 .此外对不同视频序列的时域分割门限值还需进行一定的调整.这说明此算法还有一定的局限性,为此对于如何自适应地调整参数,以增强算法的通用性,将是下一步需研究的问题.

参 考 文 献

- 1 Neri A, Colonnese S. Automatic moving object and background separation [J]. *Signal Processing*, 1998, 66(2):203~217.
- 2 Kim Munchurl, Choi Jae Gark. A VOP generation tool: Automatic Segmentatic of Moving Objects in ImageSequences Based on Spatio-Temporal Information [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 1999, 9(8): 1216~1226.
- 3 Wang Demin. Unsupervised video segmentation based on watersheds and temporal tracking [J]. *IEEE Transactions on Circuits and video technology*, 1998, 8(4):339~546.
- 4 Philippe Salembier. Morphological multiscale segmentation for image coding[J]. *Signal Processing*, 1994,38:359~386.
- 5 Vincent Luc, Soille Pierre. Watershed in digital space; an implementation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):582~598.



李朝晖 1966年生,1991年获南京航空航天大学硕士学位,现为华南理工大学在职博士生,广州大学讲师.主要研究方向为图象和视频分析、模式识别.



张艳玲 1971年生,1996年获南京理工大学硕士学位,现为广州大学讲师.主要研究方向为信息处理、计算机网络.



余英林 1932年生,华南理工大学教授,博士生导师.主要研究方向为信息处理、图象处理、模式识别、神经网络.多次获省部级以上科研成果奖,发表学术论文60余篇.