

一种基于时空信息的运动目标提取新算法

朱仲杰^{1),2)} 蒋刚毅^{2),3)} 郁梅^{2),3)} 吴训威^{1),2)}

¹⁾(浙江大学信息与电子工程系,杭州 310027) ²⁾(宁波大学电路与系统研究所,宁波 315211)

³⁾(西安电子科技大学 ISN 国家重点实验室,西安 710071)

摘要 自动、快速的视频目标分割是目标基视频编码中的一项关键技术.为此提出了一种基于时空信息的运动目标提取算法.该算法首先根据多帧运动信息和高阶统计检测方法得到二值运动掩模图象,然后提出一种改进后的分水岭算法对运动区域及其周围部分进行分割,最后将二者所得结果进行投影运算,得到最终运动目标.实验结果说明了该算法的有效性.

关键词 计算机图象处理(520·6040) 视频目标 有限区域分割 改进分水岭算法

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 1006-8961(2003)04-0422-05

New Algorithm for Extracting Moving Object Based on Spatio-temporal Information

ZHU Zhong-jie¹⁾, JIANG Gang-yi^{2),3)}, YU Mei^{2),3)}, WU Xun-wei^{1),2)}

¹⁾(Dept. of Information and Electronics Engineering, Zhejiang University, Hangzhou 310027)

²⁾(Institute of Circus and Systems, Ningbo University, Ningbo 315211)

³⁾(National Key Laboratory of ISN, Xidian University, Xian 710071)

Abstract Fast and automatic segmentation of video object is a key technology in object-based video coding such as MPEG-4. In this paper, a new algorithm for extracting moving object based on spatio-temporal information is proposed. First, a binary motion image is achieved based on high-order statistics detection method and motion information of multiple frames. High-order statistics detection method can get rid of noise and disturbance effectively. Then, an improved watershed algorithm is proposed to segment motion region and its surrounding areas. Compared with traditional watershed algorithm the improved watershed algorithm can reduce the over-segmentation phenomenon as well as the computational load. Finally, a moving object is extracted with projecting operation of spatial and temporal segmented results. In order to test the performance of the proposed algorithm, simulation test is performed with test sequences Sussie and Missa. And good experimental results are acquired which show that the proposed algorithm is efficient.

Keywords Computer image processing, Video object, Limited region segmentation, Improved watershed algorithm

0 引言

第二代视频编码方案 MPEG-4 是以基于内容的方式表示视频数据,它引入视频目标(Video Object)的概念来实现基于内容的表示.视频场景被分割成若干个视频对象,不同视频对象采用不同的

编码策略.与传统基于块运动估计和 DCT 变换的编码方案相比,第二代编码方案具有高效压缩、基于内容交互和基于内容分级扩展等诸多优点,它能消除块效应,改善重建图象的主观质量^[1].

视频目标分割是目标基视频编码中的一项关键技术^[2].视频目标分割的结果不仅要求帧内静止图象分割的合理性,而且还要保持帧间运动物体分割

基金项目:浙江省青年人才基金(RC01057);西安电子科大 ISN 国家重点实验室开放基金(H02025);

浙江省中青年学科带头人基金(2001385);宁波市重点博士基金(01J20300-05);教育部留学归国人员科研基金(2001331).

收稿日期:2002-06-04;改回日期:2002-11-19

的延续性。目标分割算法可以分为两类:自动分割和交互式分割^[3]。自动分割是计算机自动地从视频序列中分割出运动目标,而交互式分割是计算机和人工相结合的方法,需要人工进行干预。要实现基于目标的实时视频编码,视频目标的自动分割是必不可少的。然而,想要实现在任意场景下提取任意形状的视频目标是非常困难的,因为当前的计算机所能识别和利用的信息只能是纹理、边缘、运动等低层次的信息,而不能有效地描述和识别有意义的视频目标和物体^[4,5]。本文研究的重点是运用空域灰度分割和运动检测相结合的方法,提取静止背景中的运动物体。运动检测是基于多帧的运动信息,运用高阶统计检测方法,提取完整连续的运动区域。为了减少计算量和降低算法的复杂度,空域分割仅对运动区域及其附近部分进行,而不是对整帧图象进行分割,其提取的对象是运动目标,这样,对大部分静止区域的分割就毫无意义。为此提出了一种改进后的分水岭算法,它能够减少传统分水岭算法的过分割现象。对二值运动掩模图象和空域分割的结果进行投影运算,最终提取有精确边缘的运动目标。

1 运动检测

设 $f_k(x, y)$ 和 $f_{k+1}(x, y)$ 为视频序列中连续的两帧图象,二者可表示为

$$f_k(x, y) = s(x, y) + m_k(x, y) + n_k(x, y) \quad (1)$$

$$f_{k+1}(x, y) = s(x, y) + m_{k+1}(x, y) + n_{k+1}(x, y) \quad (2)$$

其中, $s(x, y)$ 表示静止背景部分, $m_k(x, y)$ 表示运动区域(包括物体运动、暴露和遮挡引起的), $n_k(x, y)$ 为图象噪声, k 为帧序号。定义连续两帧的帧差图象为:

$$d_{k,k+1}(x, y) = m_{k+1}(x, y) - m_k(x, y) + n_{k+1}(x, y) - n_k(x, y) \quad (3)$$

理论上, $d_{k,k+1}(x, y)$ 为零的点代表静止区域, $d_{k,k+1}(x, y)$ 不为零的点代表运动区域。但是由于存在观察噪声,使得实际上静止区域的 $d_{k,k+1}(x, y)$ 并不等于零。为了抑止噪声,精确提取出属于运动区域的像素点,采用了高阶统计的检测方法。运动区域的帧差可看作是非高斯信号,而噪声可看作是均值为零的高斯信号。这样,运动区域检测过程就可看作是从高斯信号中提取随机非高斯信号的过程^[6],采用四阶矩检测器进行检测。为了判断像素 (x, y) 是否运动,先在帧差图象上计算以 (x, y) 为中心的矩形窗口 $w(x, y)$ (窗口尺寸: $N_w \times N_w$) 的四阶矩

$$D_{k,k+1}^{(4)}(x, y) = \frac{1}{N_w^2} \sum_{(p, q) \in w(x, y)} (d_{k,k+1}(p, q) - \bar{d}_{w(k,k+1)}(x, y))^4 \quad (4)$$

其中, $\bar{d}_{w(k,k+1)}(x, y)$ 是窗口 $w(x, y)$ 内帧差信号均值,即

$$\bar{d}_{w(k,k+1)}(x, y) = \frac{1}{N_w^2} \sum_{(p, q) \in w(x, y)} d_{k,k+1}(p, q) \quad (5)$$

给定阈值 $u(\bar{\sigma}_d^2)^2$, 可得到一个二值图象

$$B_{k,k+1}(x, y) = \begin{cases} 1 & D_{k,k+1}^{(4)}(x, y) \geq u(\bar{\sigma}_d^2)^2 \\ 0 & D_{k,k+1}^{(4)}(x, y) < u(\bar{\sigma}_d^2)^2 \end{cases} \quad (6)$$

$B_{k,k+1}(x, y) = 1$ 表示像素点 (x, y) 属于运动区域, $B_{k,k+1}(x, y) = 0$ 表示像素点 (x, y) 属于静止区域。式中, u 是由实验确定的常数, $\bar{\sigma}_d^2$ 是噪声方差,由下式给出

$$\bar{\sigma}_d^2 = \frac{1}{N_s^2} \sum_{(p, q) \in S(x, y)} (d_{k,k+1}(p, q) - \bar{d}_{s(k,k+1)}(x, y))^2 \quad (7)$$

式中, $S(x, y)$ 是属于图象静止背景中的区域,因为图象的运动区域一般位于图象的中间,所以 $S(x, y)$ 可以在图象的边界选取, N_s 为选取区域的尺寸大小。

以上得到的 $\{B_{k,k+1}(x, y)\}$ 仅仅是根据相邻两帧的帧差图象得到的二值运动掩模图象,当图象中运动物体的运动幅度很小时,仅根据相邻两帧帧差值得到的二值图象很不完整。在后续的空域分割中,由于不是对整帧图象进行分割,而是仅仅对运动区域及其附近部分进行分割,这就要求在前期的运动检测中,分割出的运动区域应当完整。因此,运动区域的提取是基于连续多帧的帧差信息,设 $\{B(x, y)\}$ 表示提取的二值运动掩模图象,由下式给出

$$B(x, y) = \max\{B_{k,k+1}(x, y) | k=0, 1, \dots, M-1\} \quad (8)$$

式中, M 表示参与运动区域提取的图象总共有 $M+1$ 帧,然后,对 $B(x, y)$ 进行形态开运算和形态闭运算,以消除凸凹区,得到最终的二值运动掩模图象。

2 空域分割

分水岭算法是一种经典的图象分割工具,它能够将图象分割成许多纹理和灰度一致的区域,分割后的图象能很好地反映图象轮廓和边界^[7]。但传统的分水岭算法容易产生过分割现象,分割后图象的区域过多、过细,因此提出了一种改进后的分水岭算法,以减少过分割现象。

在分割前需要对图象进行形态预处理,以化简图象。形态开运算起到消峰作用,而形态闭运算主要

起到填谷作用.为了同时起到消峰填谷的作用,通常采用形态开、闭级联组合形式.对于灰度图象 $f(x)$,形态膨胀 $f \oplus \beta$ 和形态腐蚀 $f \ominus \beta$ 定义如下

$$f \oplus \beta = \max\{f(x-a), (x-a) \in D_r, a \in \beta\} \quad (9)$$

$$f \ominus \beta = \min\{f(x+a), (x+a) \in D_r, a \in \beta\} \quad (10)$$

式中, β 为结构元素, D_r 为图象 $f(x)$ 的定义域. 设 f_a 表示图象 f 的形态梯度图象, 定义形态梯度运算如下

$$f_a = f \oplus \beta - f \ominus \beta \quad (11)$$

得到形态梯度图象 f_a 后, 对其进行形态闭滤波和形态开滤波, 以消除图象中的凸区和凹区, 得到最终化简后的形态梯度图象 \tilde{f}_a , 空域分割针对 \tilde{f}_a 进行, 采用改进后的分水岭算法. 改进后的分水岭算法实际上是一种迭代分割算法, 其步骤如下:

(1) 对于形态梯度图象 \tilde{f}_a , 根据图象中像素灰度最大与最小值, 确定一个灰度尺度 h . 从最小灰度的像素开始, 像素差值小于等于 h 的所有像素划归同一个等级. 整帧图象划分为一定数量的等级, 每个像素都属于某一个等级. 灰度值最低的像素所在的等级定义为第 1 个等级, 以后随着像素值的增大等级依次增高.

(2) 取出第 1 个等级中的所有像素, 按下述规则给每个像素分配标号: 相邻的像素点分配相同的标号, 孤立的像素点分配一个独立的标号.

(3) 取出下一个等级中的所有像素点, 按以下步骤 4~步骤 6 给每个像素分配标号.

(4) 在像素的邻域内, 除水线标号外, 如果存在已经分配了标号的像素点, 且标号相同, 则该像素赋予相同的标号.

(5) 在像素的邻域内, 除水线标号外, 如果存在已经分配了标号的像素点, 但标号不同, 则该像素赋予一个特殊的标号, 称为水线.

(6) 在像素的邻域内, 除水线标号外, 不存在已经分配了标号的像素点, 则该像素赋予一个新的标号, 表示是一个新的区域.

(7) 重复第 3 步~第 6 步, 直到所有等级中所有像素都被分配了标号.

(8) 分配相同标号的所有像素属于同一个区域, 检查图象分割后每个区域的大小, 如果某个区域的像素个数少于 N , 则撤销该区域内所有像素已分配的标号, 重复第 2 步~第 7 步, 重新给每个像素分配标号. 其中, N 表示区域的临界尺寸, 它由图象的尺寸大小来确定.

(9) 重复第 2 步~第 8 步, 直到图象中的所有像素都分配了标号, 且不存在尺寸小于 N 的区域.

3 运动目标提取算法步骤

运动目标提取新算法的基本步骤如下:

(1) 根据多帧的帧差信息, 基于高阶统计的检测方法, 提取完整连续的运动区域, 得到最终的二值运动掩模图象 $B(x, y)$.

(2) 由 $B(x, y)$ 确定空域分割的图象区域大小. 由于运动区域的提取是基于多帧的帧差信息, 提取的运动区域较为完整和连续, 运动目标位于运动区域内部, 因此空域分割只需要对运动区域及其边界部分进行处理, 而非整帧图象.

(3) 对待分割的图象区域计算形态梯度图象并化简, 利用改进分水岭算法进行空域分割.

(4) 对分割后的图象与二值运动掩模图象 $B(x, y)$ 进行投影运算, 得到最终的运动目标二值掩模图象 $B_o(x, y)$. 设 $A_i (i=0, 1, \dots, N-1)$ 为空域分割后得到的各个区域, N_i 为 A_i 的大小, 给定阈值 T , 由如下投影运算得到最终的 $B_o(x, y)$

$$B_o(x, y) = \begin{cases} 1 & (x, y) \in A_i \& \frac{\sum_{(x, y) \in A_i} m(x, y)}{N_i} \geq T \\ 0 & (x, y) \in A_i \& \frac{\sum_{(x, y) \in A_i} m(x, y)}{N_i} < T \end{cases} \quad (12)$$

式中, $i=0, 1, \dots, N-1$.

$$B_o(x, y) = \bigcap \{B_o(x, y)\} \quad (i=0, 1, \dots, N-1) \quad (13)$$

(5) 对 $B_o(x, y)$ 进行形态后处理, 结合原图象提取最终运动目标.

需要说明的是, 在基于多帧帧差信息提取运动区域的步骤中, 存在一个最佳帧数的选择问题. 对于不同的视频序列, 为了提取完整的运动区域, 所需的帧数不太一样, 这和视频序列的运动特征有关. 在空域分割时, 周围区域不需太大, 也不宜太小. 一般选取运动区域周围 50~100 个像素区域.

4 实验结果

为了检验新算法的有效性, 对测试序列 Susie 进行了测试, 序列具有简单静止背景, 图象大小为 720×480 pixels, 提取对象是序列第 30 帧中的运动

目标。首先,根据多帧帧差信息,运用高阶统计检测方法,得到二值运动掩模图象,结果如图 1(b)所示;然后,根据二值运动掩模图象确定空域分割的图象区域大小,本文中分割的图象区域是运动区域及其周围 50 个像素点,如图 1(c)所示。在基于多帧帧差信息提取运动区域的步骤中,图 1 中所采用的帧数

为 9 帧,图 2 为采用 Missa 图象序列得到的实验结果。由上述实验结果可以看出,应用本文算法能有效地提取运动目标,并具有较为精确的边缘。此外,由于改进分水岭算法减少或消除了过分割现象,同时也降低了计算复杂度,因此与原有分水岭算法相比,在性能上,它综合地提高了处理速度和分割效果。



图 1 Susie 图象序列实验结果(第 30 帧)



图 2 Missa 图象序列实验结果

5 结 论

基于连续多帧的帧差信息和高阶统计检测方法,能有效去除噪声,提取完整的二值运动掩模图象。改进分水岭算法在保持原算法能提取精确物体边缘和轮廓的前提下,可以有效地减轻过分割现象。空域分割仅对运动区域及其周围部分进行,可以减少计算量和降低算法的复杂度。将空域分割和运动分割的结果进行投影运算可得到最终有精确边缘的运动目标,实验结果表明,对于背景简单且是静止的目标视频序列,本文所提算法是有效的,而对于具有复杂背景和复杂运动目标的视频序列,还有待作进一步的研究。

参 考 文 献

- 1 Zhu Z, Jiang G, Yu M. New approach to reducing blocking effects in stereo video coding[A]. In: Proc. of IEEE Int. Conf. on Signal Processing[C], Beijing, 2002;1027~1030.
- 2 Salembier P, Marques F. Region-based representations of image and video; Segmentation tools for multimedia services[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1999, 9(8): 1147~1169.
- 3 Kim M, Choi J, Kim D *et al.* A VOP generation tool; Automatic segmentation of moving objects in image sequences based on spatial-temporal information [J]. IEEE Trans. On Circuits and Systems for Video Technology, 1999, 9(8): 1216~1226.
- 4 Sobottka K, Pitas I. A novel method for automatic face segmentation, facial feature extraction and tracking[J]. Signal Processing; Image Communication, 1998,12(3): 263~281.
- 5 Castagno R, Ebrahimi T, Kunt M. Video segmentation based on multiple features for interactive multimedia applications [J]. IEEE Transactions on Circuits and Systems for Video Technology, 1998,8(5): 562~571.
- 6 Neri A, Colonnese S, Russo G. Automatic moving objects and background segmentation by means of higher order statistics [A]. In: Proc. of SPIE[C]. San Jose, California, USA, 1997, 3024;8~14.
- 7 Vincent L, Soille P. Watersheds in digital spaces; An efficient algorithm based on immersion simulation [J]. IEEE Trans PAMI, 1991, 13(6):583~598.



朱仲杰 1976年生,浙江大学信电系博士研究生.主要研究领域为多视域视频图象编码、无线视频编码与通信、视频水印与信息安全等.



郁梅 1968年生,博士,副教授.研究方向包括计算机视觉及其应用、多媒体信号处理、编码和水印技术.



蒋刚毅 1964年生,博士,教授.主要研究领域为多媒体信息传输与信息安全、图象处理与视频信号编码、基于视觉的智能控制.

吴训威 1940年生,教授,博士生导师,宁波大学电路与系统研究所所长.主要研究方向为集成电路设计.

第12届中国多媒体学术会议(NCMT2003)

征 文 通 知

由中国计算机学会和中国图象图形学会联合召开的全国多媒体学术会议从1992年在北京第一次召开以来,已成功地举行了11届.第12届全国多媒体学术会议(NCMT2003)将于2003年9月18日~20日在风景秀丽的贵州省贵阳市召开.在会议期间将组织著名学者就宽带和无线网络、新型计算模式、流媒体等热点领域做大会特邀报告和进行广泛的学术讨论.

欢迎各位同行踊跃投稿,这次大会的内容将覆盖以下广泛的领域,但并不局限于这些内容.

多媒体信息处理和编码:多媒体信息处理和压缩、内容分析、基于内容的检索、数字水印、和信息安全

多媒体系统支持和网络技术:网络协议、无线网络、操作系统、中间件、服务器、服务质量保证(QoS)、数据库、传感器和执行元件、客户终端、流媒体技术

多媒体工具、应用系统:超媒体系统、用户接口、著作工具、多媒体人机交互和集成、多媒体教育系统、分布式多媒体系统和应用、虚拟空间的设计和应用、系统集成、计算机支持的协同工作

关于投稿的重要日期:投稿截至日期(以邮件寄出日期):2003年7月13日;录取通知:2003年7月27日前;正式稿件发回:2003年8月10日前

NCMT2003上的优秀论文将推荐到国内著名学术刊物发表,其中包括中国图象图形学报的多媒体技术专辑.

大会程序委员会主席:徐光祐(清华大学计算机系)

组织委员会主席:陈孝威(贵州大学计算机系)

投稿请寄:清华大学计算机系,北京,100084(投稿请同时寄打印稿和通过电子邮件发电子版文件)

联系人:孙兴华,清华大学计算机系,北京,100084

电话:010-62784141,010-62786910 **传真:**010-62771138 **电子邮件:**ncmt03@media.cs.tsinghua.edu.cn

有关会议的更详细的信息请访问:<http://media.cs.tsinghua.edu.cn/~ncmt03>