

基于时空信息融合的视频对象分割方法

张 雄, 毕笃彦, 许悦雷, 杨宝强

(空军工程大学 工程学院, 陕西 西安 710038)

摘 要:使用一种改进的分水岭方法对图像进行空域分割,采用抗噪声能力更强的多尺度形态学算子求取图像梯度,并对经典浸没模型的处理单元进行了改进,明显提高了运算效率,改善了过分割现象。最后联合时空域信息确定视频对象的边界位置,进而提取出运动前景对象。

关键词:视频对象;分割;累积帧差;分水岭;形态学

中图分类号: TN919 **文献标识码:** A **文章编号:** 1009-3516(2006)02-0048-03

目前,基于变化区域检测运动对象的时空信息分割算法应用比较广泛,此类方法先通过检测序列中的变化区域,再用空间分割结果来修正变化区域,从而分割出具有精确边界的视频对象,运算相对简单^[1-3],缺点在于门限值必须人为预先设定,而不是在检测过程中自适应确定,而且仅用两帧之差并不能把目标的完整边界体现出来,难以在实际中得到应用。空域分割则常用分水岭方法,但经典的基于浸没模型的分水岭方法存在过分割和计算复杂等缺点,很难满足实时性要求。

1 全局运动补偿

视频序列通常包括两种运动信息:局部运动信息和全局运动信息。局部运动即视频对象的运动,为前景对象相对于背景所做的运动。而全局运动是在视频流的记录过程中,摄像机随目标运动产生的,如高空无人机的图像侦察等,摄像机在很多监控场合需要做旋转和平移等运动,全局运动造成的图像变化是一致的。因此,在视频对象分割中,就必须对摄像机引起的全局运动进行估计并在计算相邻帧的帧差时进行相应的运动补偿。

对全局运动进行估计的方法一般有块匹配运动位移估计法和参数仿射模型,考虑计算精度和运算量,本文采用6参数的仿射运动模型估计全局运动参数: $u(x, y) = a_1x + a_2y + a_3$; $v(x, y) = a_4x + a_5y + a_6$ 这一运动模型足以描述一般视频序列中的摄像机运动,而且在运动模型复杂性和模型参数估计的准确性之间也有较好的折衷。可使用最小二乘法来估计式中的6个参数,即最小化误差函数 $\psi(a_1, a_2, a_3, a_4, a_5, a_6) = \sum_{i=1}^N [(a_1x_i + a_2y_i + a_3 - u_i) + (a_4x_i + a_5y_i + a_6 - v_i)]$ 。式中: N 为帧图像的像素数。采用奇异值分解方法求解6个参数,可以得到可靠的数值解。

2 基于改进分水岭方法的空域分割

由 Vincent 和 Soille 提出来的基于浸没模型的分水岭方法^[4]是常用的图像分割方法之一,但是分水岭方法对图像梯度噪声特别敏感,容易过度分割^[5],且计算耗时非常大^[6]。本文采用一种快速的优化分水岭方法,不仅可以减少过分割,而且计算速度大大加快,易于实时系统的硬件实现。

算法分3个步骤,首先对图像进行形态预处理,以简化图像;然后采用多尺度形态梯度来产生参考图像,

收稿日期:2005-05-09

作者简介:张 雄(1978-),男,湖北钟祥人,博士生,主要从事图像处理,模式识别等研究;

毕笃彦(1962-),男,陕西扶风人,教授,博士生导师,主要从事智能图象处理,模式识别等研究。

以加强目标之间的差异;最后采用改进的分水岭方法进行分割。

2.1 计算形态梯度图像

我们采用一种多尺寸形态学梯度算子,它可以有效地增强模糊边界,减少梯度图像中不相关的局部极小值的个数。多尺寸形态学梯度算子定义为 $MG(f) = \frac{1}{n} \sum_{i=1}^n \{[(f \oplus g_i) - (f \ominus g_i)] \ominus g_{i-1}\}$ 。其中 g_i 是一组方形结构元素,且有 $1 \leq i \leq n$, n 为整数,表示不同大小结构元素的个数, g_i 的大小是 $(2i + 1) \times (2i + 1)$ 。在实际中,由于多尺度的形态学梯度算子使用了取平均运算,所以抗噪声的能力更强。

2.2 用改进分水岭方法对区域进行标识

梯度图像建立后就可进行分水岭变换,分水岭变换的过程实际上就是一个在梯度图像上不断“灌水”的过程。关于分水岭方法的严格数学描述请参见文献[7-8]。在 Vincent 的经典分水岭方法中,构造了一个处理单元(PE),用来标识分水线。比如基于单像素的分割中,像素的标识由其 8 邻域的像素 $N_1 - N_8$ 来决定,此时需 3×3 的像素阵作为 PE 的输入,如图 1 所示,这样计算时间比较慢,很难用于实时系统中。为此我们提出用增加的 16 个支撑像素 $S_1 - S_{16}$ 来决定 3×3 像素的标识,PE 的输入为 5×5 像素阵,如图 1。具体实现过程如下:

首先根据 8 像素邻域的标识来决定像素 C 的标识,该过程同于 Vincent 的方法,具体细节见文献[9];然后对像素 C 的 8 个邻域像素进行标识,每个像素 N_i 的标识仅由像素 C 和与其相对的 3 个支撑像素决定,如标识 N_1 时,只考虑像素 C 、 S_1 、 S_2 和 S_{16} 的值,同样的, N_2 的标识要由 C 、 S_2 、 S_3 和 S_4 的像素值决

S_1	S_2	S_3	S_4	S_5
S_{16}	N_1	N_2	N_3	S_6
S_{15}	N_8	C	N_4	S_7
S_{14}	N_7	N_6	N_5	S_8
S_{13}	S_{12}	S_{11}	S_{10}	S_9

图 1 5×5 输入像素矩阵

定。但对 $N_i(i=1, \dots, 8)$ 的分水线判断则要由低-高-低(LHL)原则决定,以 N_1 为例,当 N_1 的像素值大于其邻域像素 C 、 S_1 、 S_2 和 S_{16} 的值时, N_1 被标识为分水线。

仅用 C 、 S_1 、 S_2 和 S_{16} 判断像素 N_1 是否为分水线的原因(同样适用于 $N_2 - N_8$):由于 N_1 在像素 C 标识之后仍未标识,则像素 C 的值在 N_1 的邻域中是很小的,因此 $C - N_1 - S_1$ 不可能成为分水线。而且,当处理的图像分辨率较大时,分水岭在局部可认为是一条直线,而通过 C 、 N_1 、 S_1 、 S_2 和 S_{16} 即可决定 $S_3 - N_1 - S_{15}$ 、 $S_{16} - N_1 - N_2$ 或 $S_2 - N_1 - N_8$ 三种可能的分水线,因此用 C 、 S_1 、 S_2 和 S_{16} 的像素值即可判断 N_1 是否为分水线。

所提出的这种增强型分水岭方法处理能力从 $1/9$ 提高到 $9/25$,由于每次处理像素较多,相对减少了对外部存储器的访问,速度明显提高。对两种方法分割 Lena 图像计算复杂度进行对比,明显看到本文方法要比经典 Vincent 方法效率高,明显减少了过分割,总体性能要比 Vincent 方法好。

分水岭分割能够比较准确地得到运动物体的边缘,特别是细节部分不会丢失,但对于背景或运动物体本身比较复杂的情况,分水岭分割并不能得到具有语义的运动物体,必须联合时域方法,对分割后的小块与时域信息进行融合,从而提取出运动对象。

3 时空信息融合分割视频对象

由时域分割得到了视频对象的大致区域,空域分割也已得到了视频对象的一致区域的准确边界,因此只需要融合时域分割和空域分割的结果即可。我们将空域分割后得到的区域向时域分割得到的视频对象区域投影,如果前者区域的像素属于后者区域的比率高于一预先设置的阈值 TH,则可认为此区域属于视频对象。合并所有的此类区域,即可得到最终的视频对象。阈值 TH 的选取与具体的视频序列有关,如阴影、噪声等。在本算法中,由于空域分割前进行了多尺度形态学滤波,有效地去除了阴影和噪声的影响,故 TH 的选取不需要太大,一般可取 $TH = 75\%$ 左右。

4 实验结果及分析

我们对多个具有不同特点的视频序列进行了实验,图 2 是对 Coastguard 序列进行视频对象分割的结果。Coastguard 序列帧图像分辨率为 352×288 。由于该序列中目标较大,移动也比较快,加上水波的影响,对时

域信息提取会有一些的干扰。(a)是 Coastguard 的第 170 帧,(b)是第 165 到 170 帧的累积帧差,(c)是自适应二值化结果,(d)为形态学处理后得到的模板,(e)为改进分水岭空域分割结果,(f)图为时空融合提取得到的第 170 帧的视频对象。

该序列摄像机的运动也基本上是平移的,最终得到的结果不完整并且没有连通在一起,一个原因是船体的很多部分纹理与水面非常相似,另一个原因是摄像机和船基本都做水平移动,而船的底部中间纹理在水平方向基本相似,致使帧差累积不明显,因此造成船体底部中间出现中断。

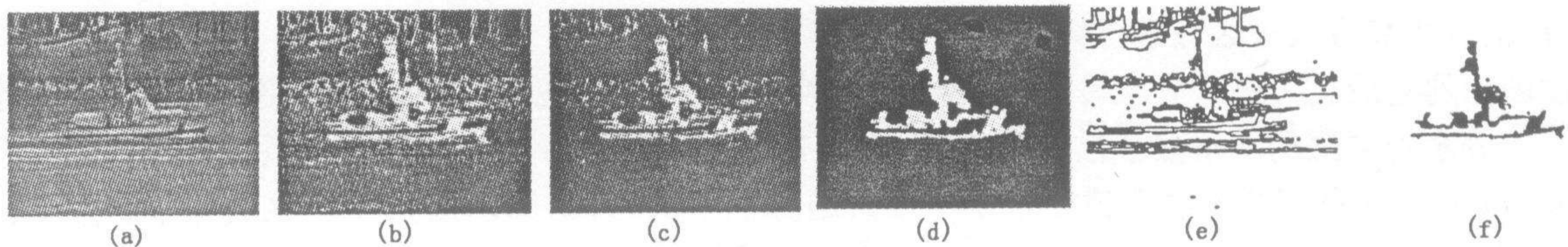


图2 Coastguard 序列时空联合分割结果

参考文献:

- [1] Chang M M, Tekalp A M, Sezan M I. Motion - field Segmentation Using an Adaptive MAP Criterion[A]. Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing, Minneapolis[C]. MN, USA, 1993, 5: 33 - 36.
- [2] Thomas Meier, Ngan K N. Automatic Segmentation of Moving Objects for Video Object Plane Generation[J]. IEEE Transactions on Circuits and Systems for Video Technology, September 1998, 8(5): 525 - 538.
- [3] 许悦雷,左继章,张 雄,等. 一种累积帧差视频对象分割算法[J]. 光电工程, 2004, (7): 69 - 72.
- [4] Vincent L, Soille P. Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 13: 583 - 597.
- [5] Haris K. Hybrid Image Segmentation Using Watersheds and Fast Region Merging[J]. IEEE Transactions on Image Processing, 1998, 7(12): 1684 - 1699.
- [6] Smet P D. Performance and Scalability of High Optimized Rainfalling Watersheds Algorithm[A]. CISST98.
- [7] Bleau A, Leon L J. Watershed - based segmentation and region merging[J]. Computer Vision and Image Understanding, 2000, 77(3): 317 - 370.
- [8] Najm an L, Schm itt M. Geodesic Saliency of Watershed Contours and Hierarchical Segmentation. IEEE Transaction on Pattern Analysis and Machine Intelligence 1998, 18(12): 1163 - 1173.
- [9] Salembier P, Pardhs M. Hierarchical Morphological Segmentation for Image Sequence Coding[J]. IEEE Transactions on Image Processing. 1994, 3: 639 - 651.

(编辑:姚树峰)

An Algorithm of Video Object Segmentation Based on Spatial - temporal Information Fusion

ZHANG Xiong, BI Du - yan, XU Yue - lei, YANG Bao - qiang

(The Engineering Institute, Air Force Engineering University, Xi'an, Shaanxi 710038, China)

Abstract: An improved watershed segmentation method in spatial segmentation is used. In the method the multi - scale morphological operator is used, which has very strong anti - noise capability to obtain the image grads and modify the processing element in conventional immersing model. As a result, the calculation efficiency is greatly increased and the over - segmentation is also restrained. By fusing spatial and temporal results the borderline is located and the moving foreground object is segmented.

Key words: video object; segmentation; sum of absolute frame differences; watershed; morphology