

# 像素域运动对象提取算法的研究

周启亚<sup>1</sup>, 杨高波<sup>2</sup>

(1. 湖南铁路科技职业技术学院 信息技术系, 湖南 株洲 412000; 2. 湖南大学 计算机与通信学院, 湖南 长沙 412008)

**摘要:** 针对像素域运动对象的提取, 较系统地分析比较了现有主要算法各自的原理、提取效果及时间特性, 为研究者在各种不同的场合选择不同的提取算法提供了依据。为使算法具有可比性, 实验在同一 Hall\_Monitor 序列上进行。然后, 提出一种改进的、同时进行帧间差分 and 减背景的运动对象提取算法, 其鲁棒性在于能对光照变化、运动物体的暂停, 显露部分的背景区域等复杂情况作出正确判断处理。

**关键词:** 运动对象; 提取; 视频处理; 像素域

中图分类号: TP391.41

文献标识码: A

文章编号: 1673-9833(2008)06-0050-05

## Research on Moving Object Extraction from Video Pixel Domain

Zhou Qiya<sup>1</sup>, Yang Gaobo<sup>2</sup>

(1. Department of Information Technology, Hunan Railway College of Science and Technology, Zhuzhou Hunan 412000, China;  
2. School of Computer and Communication, Hunan University, Changsha 410082, China)

**Abstract:** In view of extraction process, the principles, the extracted effects and extracted time attributions of the main algorithms about the moving object extraction (MOE) from video pixel domain are analyzed and compared. The contributions lie in the selection rules of the appropriation algorithm for own uses according to different conditions. Simulation results are demonstrated, which are conducted on Hall-Monitor sequence. Then, a improved MOE algorithm is presented, which is based on frame difference and background subtraction simultaneously. Its robustness can deal with some complicated situations such as light change, the paused moving object in background and the uncovered background.

**Key words:** moving object; extraction; video processing; pixel domain

视频对象 (VO) 的提取一直是视频信号处理领域具有挑战性的研究内容之一, 在新一代面向对象视频编码、视频检索、智能系统等中有着广泛的应用而倍受关注。运动对象的提取就是要从视频序列中分割出语义上完整的对象, 是视频信号处理领域研究的焦点之一。目前, 大多数运动对象提取算法是基于光流场计算和颜色空间的, 即基于运动信息、颜色信息、纹理信息等特征。其大体分为 2 类: 1) 基于帧间差分的方法, 对连续两帧或多帧进行时域相减来分离前景和背景对象; 2) 基于减背景的方法, 用当前帧减背景模型而分离运动对象。

由于受光照变化、背景复杂度、噪声等多种因素的影响, 上述 2 种方法具有各自的特点, 针对不同应用场合, 其提取效果会有所差异, 执行时间也不尽相同<sup>[1]</sup>。

本文对上述 2 类像素域运动对象提取的基本方法进行了分析、比较, 在此基础上提出一种改进的提取算法。

## 1 像素域运动对象提取的基本方法

### 1.1 基于帧间差分算法

基于帧间差分算法的基本思想是: 对相邻 2 帧或

收稿日期: 2008-10-07

基金项目: 国家自然科学基金资助项目 (60572127), 湖南省自然科学基金资助项目 (05JJ30113)

作者简介: 周启亚 (1980-), 男, 湖南湘乡人, 湖南铁路科技职业技术学院教师, 硕士, 研究方向为视频编码, 视频信号处理;  
杨高波 (1974-), 男, 湖南岳阳人, 湖南大学教授, 博士, 主要研究方向为图像, 视频信号处理。

多帧逐像素计算其差分,  $t$  帧中坐标为  $(x, y)$  像素的像素值表示为  $P(x, y, t)$ ,  $t-1$  帧相应位置的像素值表示为  $P(x, y, t-1)$ 。若差值小于阈值  $\tau$ , 则判定该像素为背景像素, 否则视该像素为运动对象像素<sup>[2,3]</sup>。

### 1.1.1 diff2

连续 2 帧时域相减的情况记为 diff2, 即:

$$|P(x, y, t) - P(x, y, t-1)| < \tau \Leftrightarrow (x, y) \in B, \quad (1)$$

式 (1) 中  $B$  表示背景。diff2 的实验结果见图 1。

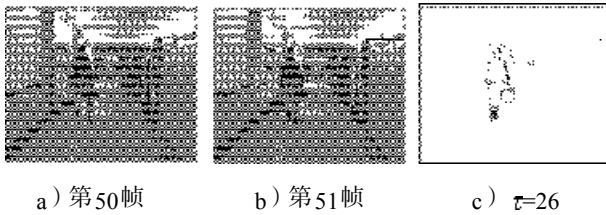


图 1 Monitor 序列第 51 帧 diff2 的实验结果

Fig. 1 The experimental result of diff2 in 51 frame of monitor sequence

### 1.1.2 diff3

通常为了减少错误的判断, 可同时对连续多帧进行比较, 连续 3 帧时域相减的情况记为 diff3:

$$\begin{cases} |P(x, y, t) - P(x, y, t-1)| < \tau, & (a) \\ |P(x, y, t) - P(x, y, t-2)| < \tau, & (b) \end{cases} \Leftrightarrow (x, y) \in B_c \quad (2)$$

对 Monitor 序列第 50 帧、51 帧和 52 帧进行实验, diff3 的实验结果见图 2。



由(a)式提取的结果 由(b)式提取的结果 diff3 的实验结果

图 2 Monitor 序列第 52 帧用 diff3 提取的实验结果 ( $\tau=26$ )

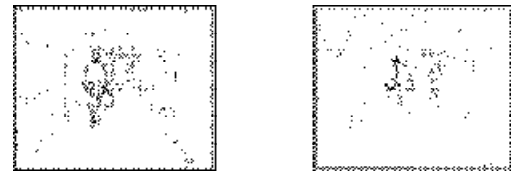
Fig. 2 The experimental result of diff3 in 52 frame on monitor sequence ( $\tau=26$ )

显然, 理论上 diff3 比 diff2 效果要好。

为了减少光照条件变化的影响, 可以先进行预处理再执行 diff2 操作。根据所进行的预处理方式不同, 帧间差分算法又有几种。

### 1.1.3 diff2trans

可对每帧的像素值进行线性变换再进行时域相减。研究表明: 有些颜色空间对光照变化不是很敏感, 这样有利于提高算法的稳定性。Elgammal 等把  $\{R, G, B\}$  转化为  $(G/(R+G+B), B/(R+G+B), (R+G+B))$  再进行时域相减, 把这种情况记为 diff2trans<sup>[4]</sup>。图 3 为对 Monitor 序列第 51 帧和第 95 帧进行 diff2trans 的实验结果。虽然 Monitor 序列背景的亮度较大, 但序列光照没有发生明显变化, 所以用此方法效果也不是很好。



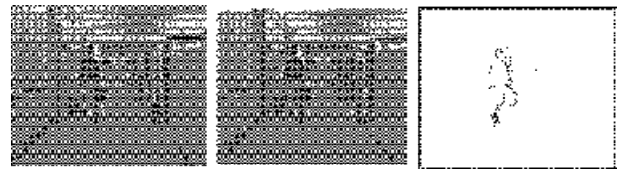
a) 第 51 帧 b) 第 95 帧

图 3 Monitor 序列 diff2trans 的实验结果

Fig. 3 The experimental result of diff2 trans on monitor sequence

### 1.1.4 diff2homo

先对执行帧差分的图像进行滤波处理 (记为 diff2homo), 使帧图像经过高通滤波器, 滤除高频分量  $r(x, y)$ , 保留低频分量  $i(x, y)$ 。因为原图像可视为由低频分量和高频分量构成的 ( $I(x, y) = i(x, y) \cdot r(x, y)$ ), diff2homo 的实验结果见图 4。



a) 第 50 帧高通滤波 b) 51 帧高通滤波 c) diff2homo

图 4 Monitor 序列用 diff2homo 提取的实验结果 ( $\tau=26$ )

Fig. 4 The experimental result of diff2homo on monitor sequence ( $\tau=26$ )

### 1.1.5 diff2grad

可以对相邻帧的梯度图进行时域相减来提取运动对象, 记为 diff2grad。先对帧检测对象的边缘, 然后执行 diff2 相同的操作, 从而提取出运动对象的轮廓。实验中先用 Roberts 算子对 Monitor 帧进行对象边缘的检测, 再执行 diff2 操作, 实验结果见图 5。



a) 第 50 帧边缘检测 b) 第 51 帧边缘检测 c) diff2grad

图 5 Monitor 序列 diff2grad 的实验结果

Fig. 5 The experimental result of diff2grad on monitor sequence

## 1.2 基于减背景的算法

基于减背景算法的基本思想是: 用当前帧减背景模型而分离运动对象, 可表示为:

$$|P(x, y, t) - B(x, y, t)| < \tau \Leftrightarrow (x, y) \in B, \quad (3)$$

式中  $B(x, y, t)$  为背景模型。

该类算法的关键在于背景模型的建立, 由于受多种因素的影响, 背景实际上是不不断变化更新的, 鲁棒的背景模型应能自适应地得到更新<sup>[5]</sup>。

### 1.2.1 backadapt

简单的自适应背景更新模型如式(4),称为backadapt:

$$B(x, y, t) = \alpha P(x, y, t) + (1 - \alpha) B(x, y, t-1), \quad (4)$$

即用当前帧的信息反馈给背景模型,使背景模型得到及时更新。式(4)中 $\alpha$  ( $0 < \alpha < 1$ )为背景学习速率,值越大,背景更新得越快。

### 1.2.2 backgauss

一般由噪声引起的同一位置像素值变化可以视为满足高斯概率分布,而由物体运动引起的像素值变化会不满足高斯分布。通过统计方法求得像素的概率密度函数,即建立高斯背景模型  $B(x, y, t) \sim N(\mu, \sigma^2)$ ,称为backgauss<sup>[6]</sup>:

$$|r(x, y) - \mu_r(x, y)| < \tau; \quad (5)$$

$$|g(x, y) - \mu_g(x, y)| < \tau; \quad (6)$$

$$|b(x, y) - \mu_b(x, y)| < \tau. \quad (7)$$

式中: $r, g, b$ 分别表示坐标为 $(x, y)$ 像素的3个颜色分量值;

$\mu_r, \mu_g, \mu_b$ 分别为3个颜色分量值的均值;

$\sigma_r(x, y), \sigma_g(x, y), \sigma_b(x, y)$ 分别为3个颜色分量值的方差;

$\tau$ 为阈值。

式(5)~(7)同时满足,则认为该像素为背景像素。backgauss的实验结果见图6。

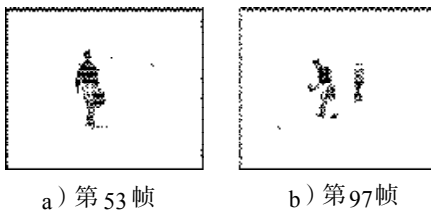


图6 Monitor序列backgauss的实验结果

Fig. 6 The experimental result of backgauss on monitor sequence

实验中先对序列的前13帧进行统计学习,得到背景的高斯模型, $\tau$ 取值为32。

### 1.2.3 backW4

若通过统计学习先得到 $M(x, y)$ (图像序列中坐标为 $(x, y)$ 处像素值的最大值), $m(x, y)$ (图像序列中坐标为 $(x, y)$ 处像素值的最小值), $D(x, y)$ (图像序列中坐标为 $(x, y)$ 处帧差最大值),这种方式建立的背景模型为backW4<sup>[7]</sup>。把当前帧的像素经过以下判断,若两式同时满足,则认为该像素为背景像素。

$$|P(x, y, t) - M(x, y)| < D(x, y), \quad (8)$$

$$|P(x, y, t) - m(x, y)| < D(x, y). \quad (9)$$

backW4的实验结果见图7。

实验中,把24位图像每个像素的 $r, g, b$ 颜色分量通过公式: $gray = 0.299r + 0.587g + 0.114b$ 转换成灰度值

来进行比较,从而避免对每个分量进行处理,以减少计算工作量,并对序列的前10帧进行统计,得到了背景模型。

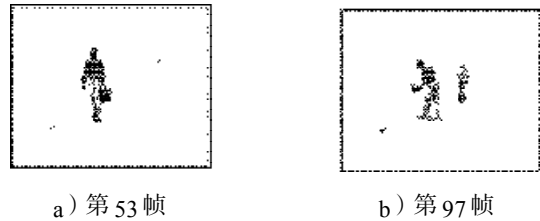


图7 Monitor序列backW4的实验结果

Fig. 7 The experimental result of backW4 on monitor sequence

## 2 一种鲁棒的运动对象提取算法

上述像素域基于帧间差分的算法实现起来简单,利用前1帧或2帧作为参考帧,较少受光照慢变化的影响。但当视频对象的运动不显著或物体部分运动时,帧间变化检测无法分割出整个视频对象,得到的只是对象的一部分;对运动对象出现暂停时,也无法检测。同时,由于视频对象的运动而暴露出的背景区域会标记为变化,被误当作视频对象的一部分。这些是进行简单的帧间差分无法解决的问题。

基于减背景算法能较好地解决以上问题,对暂停的物体和因对象的运动而显露的背景区域作出正确的判断。然而,基于减背景算法中,背景通常需要一段学习统计阶段才能得到,时间复杂度较高;背景模型也只能在序列背景相对稳定的情况才能建立。例如,对变化丰富的水纹区域、树叶摆动区域,用此方法是不能实现的<sup>[8]</sup>。

本文提出一种改进的同时进行帧间差分 and 减背景像素域运动对象提取算法,算法框图见图8。

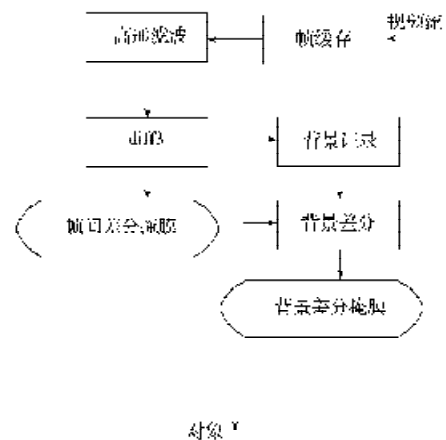


图8 算法框图

Fig. 8 The proposed algorithm framework

算法首先对缓存中的帧进行高通滤波处理,这样就能消除光照变化的影响,然后执行相邻多帧的帧间差分(diif3),得到帧间差分掩膜,记为FDM。在进行

帧间差分的过程中, 同时建立序列的背景模型, 其中背景采用文献[9]中背景记录的方法得到, 实质上是一种自适应背景更新模型。

根据 FDM 的值, 较长时间内没有运动的像素才视为背景像素, 即:

$$SI(x, y, t) = \begin{cases} SI(x, y, t-1), & \text{if } FDM = 0, \\ 0, & \text{if } FDM = 1. \end{cases} \quad (10)$$

$$BG(x, y, t) = \begin{cases} I(x, y, t), & \text{if } SI(x, y, t) = F_{th}, \\ BG(x, y, t-1), & \text{else.} \end{cases} \quad (11)$$

式中  $SI(x, y, t)$  的值为某像素属于背景的概率值。若连续多帧其 FDM 均为 0, 表明该像素属于背景的概率越大; 当  $SI(x, y, t)$  达到阈值  $F_{th}$  时, 则视该像素为背景

素。显然, 这种背景记录过程具有背景更新的能力。

得到背景模型后, 再用当前帧与背景模型相减, 求得背景差分掩膜, 记为 BDM。结合帧间差分掩膜和背景差分掩膜, 就能提取序列中各种情况下的视频对象, 初始对象掩膜记为 IOM, 从而提取的过程可以表

$$IOM(x, y, t) = \begin{cases} BDM(x, y, t), & \text{if } BI(x, y, t) = 1, \\ FDM(x, y, t), & \text{else.} \end{cases} \quad (12)$$

$BI(x, y, t)$  指示  $t$  时刻背景模型是否存在, 可通过下式求

$$BI(x, y, t) = \begin{cases} 1, & \text{if } SI(x, y, t) = F_{th}, \\ BI(x, y, t-1), & \text{else.} \end{cases} \quad (13)$$

算法在 Monitor 序列上的实验结果如图 9 所示。

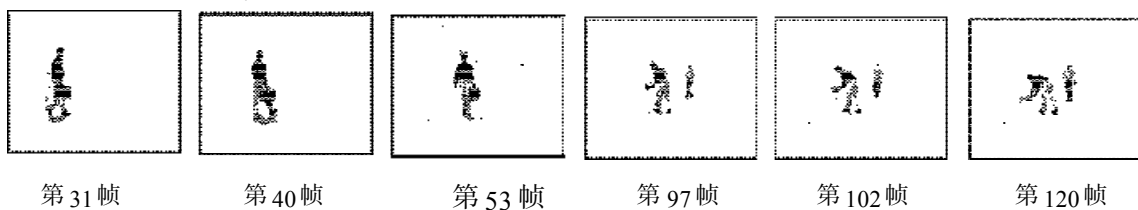


图9 本文提出的算法对 Monitor 序列的实验结果

Fig. 9 The experimental result of our proposed algorithm conducted on monitor sequence

该算法的优点是: 1) 经滤波预处理, 算法能适应对光照变化的影响。2) 鲁棒性高, 具有帧间差分和减背景算法各自的优点, 对运动物体、运动对象的暂停情况和因对象的运动而显露的背景区域均能作出正确的判断。缺点是: 1) 算法的时间复杂度较高, 基于像素的处理使得算法效率低, 这是像素域视频处理的瓶颈。2) 阈值的选择对求得的掩膜影响是很关键的, 若阈值太低, 掩膜中会包括一些伪运动像素; 若阈值太高, 一些真正的运动像素又会被忽略掉。本文实验中采用的是统计值, 针对不同序列自适应地选择阈值是下一步要解决的问题。3) 算法只是在低级特征空间进行的, 对于复杂场景的提取, 必须进一步利用语义上的知识, 即场景上下文知识和经验知识。

### 3 像素域提取算法比较

为便于比较, 本文把各种典型的像素域运动对象提取算法的执行时间与主观质量评价列入表 1。

表 1 典型算法间的比较

Table 1 Comparison of extraction algorithms

Algorithm	执行时间 /ms	质量评价
diff2	3.575	一般
diff3	5.251	较好
diff2trans	5.451	较差
diff2homo	$9.414 \times 2 + 3.535$	较好
diff2grad	$8.424 \times 2 + 3.048$	一般
backgauss	9.560	较好
backW4	8.027	较好
proposed	40.361	好

为使算法具有可比性, 实验中对同一序列进行提取, 序列为 Hall\_Monitor, 格式为 qcif (176 × 144), 实验平台: Intel Pentium 4 3.00 GHz CPU, 512 MB DDR 内存, VC++ 6.0。表 1 中 diff2homo 和 diff2grad 的执行时间包括对 2 帧分别进行高通滤波和边缘检测的时间, 之后再 diff2 操作。

综上所述, 各算法在不同场合所提取的效果不尽相同, 时间复杂度也有所差异, 实际应用中可根据需要, 尽量选择合适的算法实现运动对象的提取。

### 4 结语

本文主要对像素域运动对象提取的基本方法进行了归纳和总结, 为研究者在不同的情况下选择适合自己需求的算法提供依据; 在提出一种鲁棒的同时进行帧间差分和减背景的像素域运动对象提取算法。视频对象的提取是一个复杂的问题, 在由像机造成的全局运动估计与补偿、对象阴影的消除等方面都是值得进一步深入研究的。随着交叉学科的发展, 可以引入人工智能、图像理解、运动分析等领域的方法来解决视频对象提取问题。

参考文献:

- [1] Benezeth Y, Emile B, Rosenberger C. Comparative Study on Foreground Detection Algorithm for Human Detection[C]// Proceedings of the Fourth International Conference on Image and Graphics. America: IEEE Computer Society, 2007:

- 661-666.
- [2] 刘李杰, 蔡德钧. 一种面向运动的视频对象分割算法[J]. 计算机学报, 2000, 23(12): 1326-1331.
- [3] 沈未名, 江 柳, 种衍文. 视频对象分割及跟踪方法研究[J]. 武汉大学学报: 信息科学版, 2004, 29(3): 274-277.
- [4] Elgammal A, Duraiswami R, Harwood D, et al. Background and foreground modeling using non-parametric kernel density estimation for visual surveillance[C]// Proceedings of IEEE. America: IEEE Computer Society, 2002: 1151-1163.
- [5] Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking: Proceedings of CVPR[C]// Proceedings of Computer Vision and Pattern Recognition. America: IEEE Computer Society, 1999: 246-252.
- [6] Ju Han Bhanu B. Detecting moving humans using color and infrared video[C]// Proceedings of IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems. America: IEEE Computer Society, 2003: 228-233.
- [7] Haritaoglu I, Harwood D, Davis L. W 4: real-time surveillance of people and their activities[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 28(8): 809-830.
- [8] Caplier A, Bonnaud L, Chassery J M. Robust Fast Extraction of Video Objects Combining Frame Differences and Adaptive Reference Image[J]. Image Processing, 2001(2): 785-788.
- [9] Chien ShaoYi, Huang YuWen, Hsieh BingYu, et al. Fast Video Segmentation Algorithm with Shadow Cancellation, Global Motion Compensation, and Adaptive Threshold Techniques[J]. IEEE Transactions on Multimedia, 2004, 6(5): 732-748.

(责任编辑: 张亦静)

(上接第 49 页)

## 5 结语

本文针对常见板材表面缺陷的特征, 提出了一种基于视觉的表面检测方法。首先针对图像中存在的噪声利用中值滤波器消除了椒盐噪声; 再用边缘检测将缺陷边缘提取出来。然后用形态学处理进行孔洞填充、将细小微粒连接成完整的缺陷信息。最后对提取的形态特征进行分析, 确定缺陷分类标准, 根据标准判别缺陷类型。结果表明, 这种方法能够有效地判别缺陷类型, 具有速度快、检测结果客观等特性, 但仍有不足, 即在背景光照不均的条件下, 正确率明显降低。在今后的研究中加强光照补偿算法的研究, 以进一步降低缺陷检测的误判率。

### 参考文献:

- [1] 廖水碧, 肖明富. 金属制品表面质量缺陷无损检测的研究现状与展望[J]. 中国冶金, 2007, 17(3): 129-133.
- [2] Elias N Malamasa, Euripides G M Petrakisa, Michalis Zervakisa, et al. A survey on industrial vision systems, applications and tools[J]. Image and Vision Computing, 2003(21): 171-188.
- [3] Hou T H, Pern M D. A computer vision-based shape-classification system using image projection and a neural network[J]. International Journal of Advanced Manufacturing Technology, 1999(15): 843-850.
- [4] 刘庆民. 基于计算机视觉的小尺寸零件精密测量技术研究[D]. 长春: 吉林大学, 2006.
- [5] 韩 斌, 刘以安, 王士同. 基于图像处理的印刷缺陷计算机自动检测[J]. 自动化技术与应用, 2002, 21(3): 37-38.
- [6] 张 洪, 段发阶, 丁克勤, 等. 带钢表面缺陷视觉检测系统关键技术研究[J]. 计量学报, 2007(7): 129-132.
- [7] 王吉晖, 伟 其, 王 霞, 等. 基于数学形态学的像增强器缺陷的图像检测方法[J]. 光学技术报, 2005, 31(3): 129-131.

(责任编辑: 廖友媛)