

# 基于多帧边缘差异的视频运动对象的分割与跟踪算法<sup>1</sup>

刘 龙 刘贵忠 王占辉 王黎明

(西安交通大学电子与信息工程学院 西安 710049)

**摘 要:** 从视频场景中分割和跟踪感兴趣的视频对象对于 MPEG-4 等基于对象的视频编码来说是关键性的技术之一。针对目前大部分视频对象分割和跟踪算法相当复杂但仍不能有效地去除背景噪声的问题,该文提出用于分割和跟踪视频运动对象的一种基于多帧边缘差异的算法。该算法利用一组帧的边缘差异来提取运动对象区域,通过聚类方法去除背景像素点,利用形态学算子得到对象分割模板,同时通过建立前帧感兴趣对象与当前帧运动对象的帧间向量跟踪当前帧的感兴趣视频对象。不同标准视频测试序列的测试结果表明,该算法能够实现对感兴趣的视频运动对象更为精确、快速和有效地分割和跟踪。

**关键词:** 边缘差异, 视频对象, 分割, 跟踪

**中图分类号:** TN941 **文献标识码:** A **文章编号:** 1009-5896(2004)05-0715-07

## Segmentation and Tracking of Video Object of Interest Based on Change of Multi-frames' Edge

Liu Long Liu Gui-zhong Wang Zhan-hui Wang Li-ming

(School of Electron. & Info. Eng., Xi'an Jiaotong University, Xi'an 710049, China)

**Abstract** It is one of the key technologies to MPEG-4 grade codes based on target to segment and track the concerned video objects from the video scene. Most current segmentation and tracking algorithms are of high complexity but not effective in getting rid of background noise. One algorithm is put forward to segment and trace video objects based on edge difference among multiple frames. According to the proposed algorithm, edge difference between a group of frames is used to draw the area of moving objects; then, background pixels are removed through setting up pixel-measuring window and threshold value; the area of objects is set up by morphology operator; at the same time, vectors between the last concerned objects per frame and moving objects of present frame are established to follow current concerned objects. The result of various standard video test sequences shows that the proposed algorithm offers more accurate, faster and more effective segmentation and tracking of concerned video moving objects.

**Key words** Edge difference, Video object, Segmentation, Tracking

### 1 引言

传统的视频编码标准如 MPEG-1/MPEG-2 以及 H.261/H.263 都没有涉及视频对象的可操作性问题。在新的基于对象的视频编码标准 MPEG-4<sup>[1]</sup> 中,由于需要对感兴趣的区域或对象进行随意的操作,就必须将其与场景中的其它信息分离。在视频场景中人们往往对于运动的视频对象感兴趣,因此,分割和跟踪感兴趣视频运动对象成为基于对象视频编码的关键技术之一。

文献 [2-4] 的分割方案是一种比较流行的视频对象分割方案。这种方法采用形态学和运动预测模型来进行分割,依次完成特征提取、特征匹配和运动分析 3 个步骤。Thomas<sup>[5]</sup> 等提出的分割算法是基于模式识别和对象跟踪技术,这种分割方法在视频对象运动量较大时也能达到

<sup>1</sup> 2002-12-08 收到, 2003-05-19 改回

国家自然科学基金 (No.60272071)、国家教育部博士点基金 (No.2000069828) 资助课题

较好的分割效果。MPEG-4 建议的自动分割方案<sup>[6-9]</sup>是基于变化区域检测的时空法,通过检测序列图象帧的变化区域与不变区域,将运动物体和静止背景进行分割。

仿真试验表明,上述方法虽然对简单背景下的运动对象有较好的分割和提取效果,但是在背景比较复杂的情况下,由于受到背景噪声干扰,背景中的像素点往往不能被很好地剔除。针对这个问题,本文提出基于多帧间边缘差异的分割方法。其做法是把当前帧和前几帧做边缘差异检测,以检测出当前帧中属于运动对象的像素。为去除背景像素点,设定一个检测窗,用指定的阈值进行判断,得到运动对象区域。然后用形态学算子对对象区域内部进行腐蚀,得到分割模板,最终得到分割的视频运动对象。

另外,视频场景中往往存在多个运动对象。但是在交互性的场合,我们只是对其中一个或者几个运动对象感兴趣。因此,在前一帧已经确定感兴趣运动对象的情况下,为了准确地辨别出当前帧的感兴趣运动对象,就需要对感兴趣运动对象进行跟踪。传统的视频对象的跟踪方法通常使用 Kalman 滤波器或者条件概率来计算和预测刚体目标的运动<sup>[10]</sup>。在运动对象被遮挡和形状发生变形的情况下,这种跟踪方法不利于视频运动对象的实时跟踪。所以本文还提出了通过建立对象帧间向量来跟踪感兴趣对象,该算法能够有效地从多个运动对象中跟踪感兴趣的视频运动对象。

## 2 感兴趣视频对象的提取

目前,根据帧间变化来检测和提取感兴趣视频对象是一个可行的方案。但是这种方法有一个很大的缺点,就是带来比较大的背景噪声。因此目前的算法主要集中在如何快速有效地去除背景噪声。本文将给出一种新的分割方案来解决这个问题。

### 2.1 多帧间边缘差异检测

分割视频对象最重要的是获取对象的边缘参数,即对象边缘像素点的坐标。由于灰度图本身只保留了背景和运动对象的边缘像素点,这就排除了大量的背景干扰,只是在背景的边缘存在少量的背景干扰。要得到运动对象的边缘参数,只需剔除背景边缘上的干扰像素点。

通过前后帧的边缘图像差异来获取运动对象的边缘参数,首先要提取图像边缘。本文采用 Canny 算子<sup>[11]</sup>对当前帧作边缘检测。假设一组视频序列有  $n$  帧组成,  $I$  表示视频序列的灰度图像,当前帧为  $I_i$ ,前一帧为  $I_{i-1}$ ,我们定义边缘差异为

$$DE_i = |\phi(\nabla G \cdot I_i) - \phi(\nabla G \cdot I_{i-1})| \quad (1)$$

其中  $\nabla G$  定义为二维高斯函数  $G(x, y) = [1/(2\pi\sigma^2)] \exp[-(x^2 + y^2)/(2\sigma^2)]$  的梯度;  $\phi$  代表在梯度方向上强度在其邻域中的最大值对应的像素点。

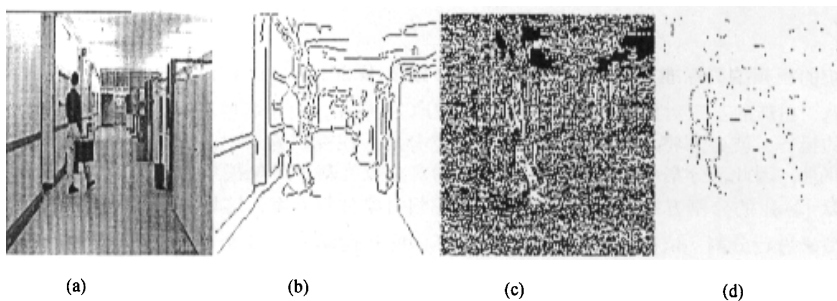


图 1 多帧边缘差异

- (a) “Hall Monitor” 第 13 帧原始图像 (b) 采用 Canny 算子提取“Hall Monitor”第 13 帧的灰度边缘图像  
(c) 第 12 帧与 13 帧的灰度差值图像 (d) “Hall Montior”第 12 帧和第 13 帧间灰度边缘差异图象

图 1 所示表明帧间的边缘差异比帧间差异的噪声点少很多,但同时存在两个问题:一个是对象部分边缘信息丢失,二是背景像素点并没有完全清除。本文采取多帧边缘差异的方法来弥补丢失信息。由于多帧的运动差异比两帧的大,这样就弥补了对象局部运动缓慢而丢失的边缘信息。我们定义多帧差异如下:

$$DE_i^N = \left| \phi(\nabla G * I_i) - \sum_{1 \leq n \leq N} \phi(\nabla G * I_{i-n}) \right| \quad (2)$$

其中  $N$  代表帧数,本文采用三帧边缘差异。

### 2.2 采用聚类方法去除背景像素点

在多帧边缘差异图中,噪声像素点比较离散,而运动像素点相对集中。根据这个特点本文采用聚类的思想方法来去除背景像素点,其具体做法是:多帧边缘差异像素点的集合为  $DE_i = \{e_1, e_2, e_3, \dots\}$ ,  $DE_i = DE_i^{\text{change}} \cup DE_i^{\text{still}}$ , 其中  $DE_i^{\text{change}}$  代表视频运动对象的边缘运动像素点集合,  $DE_i^{\text{still}}$  代表背景的边缘噪声像素点集合。根据背景像素和运动像素的不同特点,采用像素点离散度  $\xi$  作为距离。先在多帧边缘差异图上以  $(x, y)$  为中心的矩形窗口  $\eta(x, y)$  (窗口大小为  $N_\eta \times N_\eta$ ) 中定义点  $(x, y)$  的离散度:

$$\xi = \frac{1}{N_s^2} \sum_{s(x,y) \in \eta(x,y)} \sqrt{(x_s - x)^2 + (y_s - y)^2} \quad (3)$$

其中  $s(x, y)$  代表灰度值不为 0 的像素坐标,  $N_s$  为矩形窗口  $\eta(x, y)$  中灰度值不为 0 的像素点数目。将  $\xi$  与阈值  $\xi_T$  进行比较,  $\xi_T$  是由试验确定的常数,如果  $\xi \leq \xi_T$ , 则此  $(x, y) \in DE_i^{\text{change}}$ , 若  $\xi > \xi_T$ , 则  $(x, y) \in DE_i^{\text{still}}$ 。从图 2 中能看出采用这种方法去除像素的效果很好。

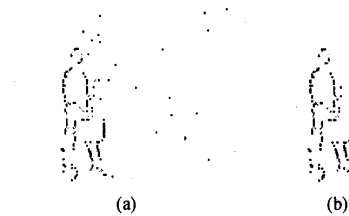


图 2 去除背景像素点  
(a) “Hall Montior” 第 11 帧, 第 12 帧和第 13 帧间灰度边缘差异图像  
(b) 采用聚类方法去除三帧边缘差异图像背景像素点

### 2.3 采用形态学提取视频对象

通过去除多帧边缘差异灰度图像中的背景像素点,可以得到运动对象的粗略轮廓。根据运动对象的边缘轮廓参数,本文采用形态学的腐蚀过程对边缘内部进行腐蚀从而得到准确的对象模板,如图 3(a) 所示。根据对象模板,清除帧内不在模板内的像素点就可以得到运动对象的分割结果,如图 3(b) 所示。



图 3 提取运动对象  
(a) 腐蚀后的结果 (b) 根据对象模板提取的运动对象

### 3 多个运动对象场景下感兴趣的视频对象的跟踪

在视频场景中往往有几个不同的视频运动对象，人们常常对某一运动对象感兴趣，这就需要在多个运动对象中跟踪感兴趣的运动对象。传统的跟踪方法<sup>[9]</sup>需要大量的计算，不利于实时跟踪处理。本文建立了一种前后帧间的向量，比较向量的模值，从而确定当前帧的感兴趣的运动对象。该算法具体做法如下：

首先确定感兴趣的运动对象 (Video Object Plane, VOP) 的位置

$$A(\text{VOP}) = \sum_{(x,y) \in \text{VOP}} 1, \quad \text{且} \quad \bar{x} = \frac{1}{A} \sum_{x \in \text{VOP}} x, \quad \bar{y} = \frac{1}{A} \sum_{y \in \text{VOP}} y \quad (4)$$

其中  $(\bar{x}, \bar{y})$  用来表述运动对象的中心坐标。假设当前帧为第  $i$  帧，第  $i$  帧内的运动对象有  $N$  个，则每个运动对象的中心坐标表示为  $(\bar{x}_i^n, \bar{y}_i^n)$ ， $1 \leq n \leq N$ ；假设第  $i-1$  帧内的感兴趣运动对象只有一个并且已经确定，第  $i-1$  帧内感兴趣的运动对象表示为  $\text{VOP}_{i-1}^G$ ，其中心坐标为  $(\bar{x}_{i-1}^G, \bar{y}_{i-1}^G)$ ，用  $\text{MV}$  (Move Vector) 表示帧间运动向量，定义运动对象的帧间向量为

$$\text{MV}_i^{Gn} = (\bar{x}_i^n - \bar{x}_{i-1}^G, \bar{y}_i^n - \bar{y}_{i-1}^G) \quad (5)$$

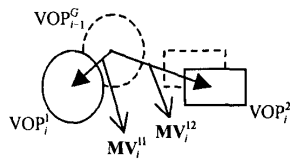


图 4 帧间向量

由于第  $i$  帧内的运动对象有  $N$  个，则有  $\{0 \leq n \leq N | \text{MV}_i^{Gn}\}$ 。以图 4 所示为例，虚线代表第  $i-1$  帧中两个运动对象，实线代表第  $i$  帧中两个运动对象，即  $\text{VOP}_i^1$  和  $\text{VOP}_i^2$ ， $\text{VOP}_{i-1}^G$  为第  $i-1$  帧中感兴趣的运动对象； $\{\text{MV}_i^{11}, \text{MV}_i^{12}\}$  为以第  $i-1$  帧内  $\text{VOP}_{i-1}^G$  为基准与第  $i-1$  帧内的  $\text{VOP}_i^1$  和  $\text{VOP}_i^2$  所形成的帧间向量集合。

我们可以计算出每个向量的模值，即

$$|\text{MV}_i^{Gn}| = \sqrt{(\bar{x}_i^n - \bar{x}_{i-1}^G)^2 + (\bar{y}_i^n - \bar{y}_{i-1}^G)^2} \quad (6)$$

计算出向量模值后， $\text{Min}\{0 \leq n \leq N | \text{MV}_i^{Gn}\}$  所对应的运动对象即是要跟踪的运动对象。

在多个运动对象的场景中，多个对象可能分离互不关联，也有可能相互遮掩或者重叠；运动对象还有可能突然出现或者消失。在出现上述两种情况时，本文的跟踪算法有相应的处理原则。以图 5 为例，场景中两个运动对象分别表示为  $\text{VOP}_1$  和  $\text{VOP}_2$ ，假定  $\text{VOP}_1$  为感兴趣的视频运动对象。如图 5 所示， $i-1$  帧， $i-2$  帧和  $i-3$  帧属于感兴趣的对象相互遮掩或者重叠变化。如果出现遮掩或重叠，则把它们作为一个整体判断。

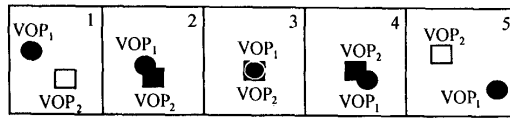


图 5 场景中的运动对象的遮掩与重叠

表 1 场景中运动对象变化情况

	前一帧的对象数目	当前的对象数目
情况 1	2	2
情况 2	2	1
情况 3	1	1
情况 4	1	2

如果出现遮掩或重叠，则把它们作为一个整体判断。另外，在两个对象场景变化中，对象的突然消失或出现有 4 种不同的情况，如表 1 所列。因此我们在判断感兴趣的视频运动对象时，要根据不同的情况来做不同的处理。在情况 1 和情况 4 的情况下，仍然取最小  $|MV_i^{G^n}|$  所对应的运动对象为感兴趣的运动对象；情况 2 的情况下，我们需要判断感兴趣对象是否在当前帧消失。我们设定阈值， $|MV_i^{G^n}|$  与  $T$  进行比较，如果  $|MV_i^{G^n}| \leq T$  认为当前对象是感兴趣对象， $|MV_i^{G^n}| > T$  判断感兴趣对象消失，这时候无任何输出；在情况 3 情况下，认为当前对象即为感兴趣的运动对象。图 6 是感兴趣视频对象的跟踪算法流程，Per\_vop 表示前一帧的对象数目，Cur\_vop 表示当前帧的对象数目。

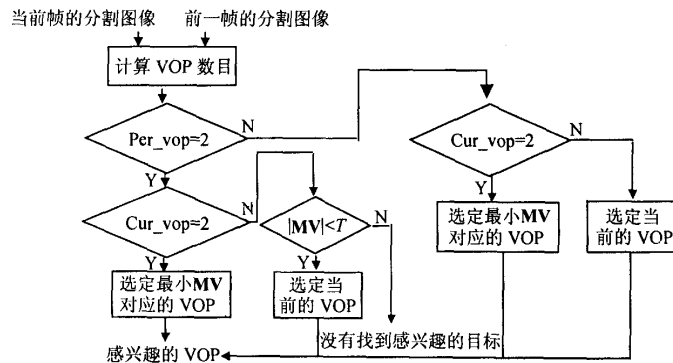


图 6 感兴趣视频对象跟踪算法流程框图

### 4 实验结果

选取两组测试序列来验证本文算法的正确性。Miss America 属于较缓慢变化的头肩序列，而 Hall Monitor 属于多运动对象序列。试验条件是 PIV1615-MHz(256MB RAM)。

#### 4.1 视频运动对象的分割算法测试

图 7 是采用本文的算法分别对 Miss America 第 2 帧和 Hall Monitor 的第 20 帧进行运动对象的分割结果。测试时，采用三帧边缘差异，离散度阈值  $\xi_T = 0.8$ ，检测窗  $N_T = 8$ 。表 2 比较了在对序列 Miss America 分割时，传统的帧间差异法和本文提出的多帧间边缘差异法所消耗的时间。其中传统的帧间差异法采用四阶矩的检测器去除噪声。



图 7 Miss America 和 Hall Monitor 的分割结果

(a) 为 Miss America 第 2 帧的原始灰度图像

(b) 为 Miss America 第 2 帧的分割结果

(c) 为 Hall Monitor 第 20 帧的原始灰度图像

(d) 为 Hall Monitor 第 20 帧的分割结果

表 2 分割算法各部分占用时间

多帧边缘差异法	计算多帧边缘变异	去除背景噪声	提取边缘轮廓	提取视频对象	总计
损耗时间 (ms)	185	70	22	123	400
帧间差异法	计算帧间变异	去除背景噪声	提取边缘轮廓	提取视频对象	总计
损耗时间 (ms)	166	104	20	124	414

从上面的试验可以看出, 本文提出的分割算法从主观上看有较好的分割效果, 完全剔除了背景的噪声干扰。在消耗时间方面, 计算多帧边缘差异要比计算帧间差异耗费时间多。然而由于多帧边缘差异所带来的噪声很少, 所以去除噪声时花费的时间少很多。从总体上看, 本文提出的多帧边缘差异法比帧间差异法节省时间, 更适用于实时处理应用。

#### 4.2 视频运动对象的跟踪算法测试

图 8 是采用本文的追踪算法对 Hall Monitor 序列视频对象进行跟踪的结果, 跟踪的感兴趣对象是第 13 帧开始的“提箱子的人”。跟踪判断参数  $T = 1$ 。

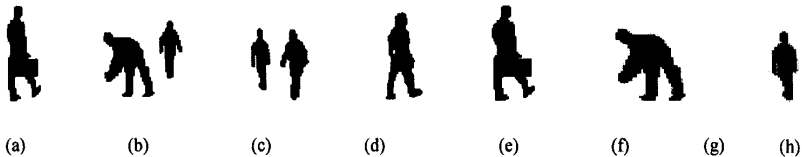


图 8 Hall Monitor 序列的运动对象跟踪结果

(a)-(d) 为第 13, 20, 39 和 96 帧视频对象分割模板

(e)-(g) 为第 13, 20, 和 39 帧跟踪结果

(h) 表示 96 帧中未找到跟踪目标

试验表明, 比较帧间向量的模值可以准确地跟踪感兴趣的运动对象。同时本文的算法比采用 Kalman 滤波器或者条件概率来跟踪平均节省时间 30-40ms。因此该算法是一种快速有效的跟踪算法。

## 5 结论

本文提出了一种自动分割和跟踪方案。引入了多帧间边缘差异的概念,提出了利用聚类思想定义像素离散度来剔除残留的背景像素点,建立帧间向量来跟踪感兴趣的运动对象。试验证实本文提出的分割和跟踪算法是有效的。该算法进一步提高了分割对象的轮廓精确度,并在多对象场景中准确地跟踪了感兴趣的运动对象。

## 参 考 文 献

- [1] Sikora T. The MPEG-4 video standard verification model. *IEEE Trans. on Circuits and Systems for Video Technology*, 1997, 7(1): 19-30.
- [2] Marques F, Molina C. Object tracking for content-based functionalities. in SPIE Vis. Commun. Image Processing, 1997, vol. 3024: 190-199.
- [3] C. Gu, Lee M C. Semantic segmentation and tracking of semantic video objects. *IEEE Trans. on Circuits and Systems for Video Technology*, 1998, 8(5): 572-584.
- [4] C. Gu. Tracking of multiple semantic video objects for internet applications. *SPIE*, 1999, vol.3653: 806-820,.
- [5] Thomas M, King N. Automatic segmentation of moving objects for video object plane generation. *IEEE Trans. on Circuits and Systems for Video Technology*, 1998, 8(5): 525-538.
- [6] Mech R, Wollborn M. Automatic segmentation of moving objects (partial results of core experiment N2), ISO/ IEC JTC1 /SC2 9/ WG11 MPEG98/ m3187, 1998.
- [7] Mech R, Wollborn M. A noise robust method for 2D shape estimation of moving objects in video sequences considering a moving camera. *Signal Processing*, 1998, 66(2): 203-217.
- [8] Kim M, Choi J G, Lee M H. Performance analysis of an ETRI's global motion compensation and scene cut detection algorithms for automatic segmentation. ISO/ IEC JTC1/SC29/WG11 MPEG97/m2387, 1997.
- [9] Bouthemy P, Francois E. Motion segmentation and qualitative dynamic scene analysis from an image sequence. *Int. Journal of Computer Vision*, 1993, 10(2): 157-182.
- [10] Meyer F, Bouthemy P. Region-based tracking in an image sequence, in Proc. European Conf. Computer Vision (ECCV), Rennes, May 1992: 476-484.
- [11] Canny J F. A computational approach to edge detection. *IEEE Trans. on PAMI*, 1986, 6(3): 679-698.

- 刘 龙: 男, 1976 年生, 博士生, 主要研究兴趣包括基于对象的视频编码、无线视频传输技术、小波分析、视频信号分析等。
- 刘贵忠: 男, 1962 年生, 1989 年 6 月获荷兰 Eindhoven 大学博士学位, 教授, 博士生导师, 主要从事非平稳信号的分析与处理、音视频数据压缩、模式识别、反演等方面的理论、方法与应用研究。
- 王占辉: 男, 1978 年生, 硕士生, 主要研究兴趣包括图像处理、视频分析和多媒体技术、小波分析等。
- 王黎明: 男, 1978 年生, 硕士生, 主要研究兴趣包括多媒体技术、小波理论、时频分析等。