

# MPEG-4 视频中运动背景下的目标检测算法

田玉敏 万 波 董文涛

(西安电子科技大学计算机外部设备研究所, 陕西 西安 710071)

**摘要** 针对由运动摄像机捕获的 MPEG-4 视频流中的运动目标检测问题, 提出了一种直接利用压缩视频码流进行全局运动估计的新算法。算法从全局运动估计的基础出发, 利用背景宏块运动相似性的特点快速建立背景宏块集合并采用常用的四参数全局运动估计模型估计运动参数。最后, 计算运动矢量残差, 通过对运动矢量残差的筛选检测运动目标。算法利用 MPEG-4 码流中蕴含的运动信息, 不需要对压缩流完全解码, 较大地提高了检测效率; 进一步改善了检测效果。实验验证了提出的全局运动估计算法的检测效率和检测效果。

**关键词** 目标检测; 运动矢量; 直流系数; 全局运动估计; MPEG-4

**中图分类号** TN919.81 **文献标识码** A **doi**: 10.3788/AOS20092905.1227

## Object Detection Algorithm Based on Moving Background in MPEG-4 Video

Tian Yumin Wan Bo Dong Wentao

(Research Institute of Peripherals, Xidian University, Xi'an, Shaanxi 710071, China)

**Abstract** Focusing on the problem of moving object detection in moving camera, a new method for the estimation of global motion from compressed image sequences is proposed. With the global motion estimation as basis, utilizing the similarity of the background micro blocks, the algorithm can build background micro blocks sets speedily. Adopting the common four-parameter global motion estimation model, the motion parameters are estimated. By computing and filtering the motion vectors residual, the moving object is detected. The algorithm can utilize the motion information contained in the MPEG-4 video coded stream without decoding the coded stream completely, so it can improve the detection efficiency and effect dramatically. Testing results validate its advantages in global motion estimation.

**Key words** object detection; motion vector; DC coefficient; global motion estimation; MPEG-4

### 1 引 言

运动目标检测是随着数字视频技术的发展而产生的一个新的研究课题。在描述视频的内容时, 通常把视频中的内容分为背景和前景, 因此, 在视频序列的运动分析中, 也常把运动信息分为背景运动信息和前景运动信息。背景运动是指由摄像机运动造成的镜头内所有点的整体移动, 又称为全局运动或摄像机运动<sup>[1]</sup>; 前景运动是指被拍摄物体在场景中的运动, 又称为局部运动。人们在理解视频中的运动内容时, 通常关心的主要是前景物体的运动, 但在

运动背景下, 物体各点的运动是由全局和局部运动叠加而成的, 因此为了正确的提取局部运动信息, 必须首先估计出全局运动信息, 并在对局部运动进行分析时剔除全局运动的影响<sup>[2]</sup>。由于全局运动和局部运动的性质截然不同, 全局运动的整体性使它可以由一组很少的模型参数来表示, 而局部运动则较为复杂, 只在小范围内表现出一定的一致性, 需要比较精细的方法才能完整的刻画。目前静止背景下的视频运动目标检测算法已经逐渐成熟, 而运动背景下的运动目标检测和跟踪至今仍然存在很多关键

收稿日期: 2008-08-13; 收到修改稿日期: 2008-10-08

基金项目: 陕西省自然科学基金(2006F48)资助项目。

作者简介: 田玉敏(1964—), 女, 教授。主要从事计算机图形图像处理与多媒体信息检索方面的研究。

E-mail: ymtian@mail.xidian.edu.cn

问题尚未解决,因而成为当前运动跟踪领域的一个研究热点<sup>[3~6]</sup>。

针对由运动摄像机捕获的 MPEG-4 视频流中的运动目标检测问题,本文提出了一种直接利用压缩视频码流进行全局运动估计的新算法。算法利用背景宏块运动相似性的特点快速建立背景宏块集合,采用常用的四参数全局运动估计模型估计运动参数。最后,计算运动矢量残差筛选运动目标,改善了检测效果。算法直接利用 MPEG-4 码流中蕴含的运动信息,不需要对压缩流完全解码,较大地提高了检测效率。

## 2 全局运动估计方法

### 2.1 全局运动估计模型

摄像机运动时所拍摄的视频序列中背景也是运动的,但这种运动是一种被动的运动,并不是由背景自主产生的,这种由于摄像机运动而造成的背景的相对运动通常被称之为全局运动,全局运动可以采用参数模型来进行描述。普遍应用的模型有平移模型、旋转缩放模型、仿射模型、平面透视模型等<sup>[7]</sup>。这些模型根据所含参数多少,又分为六参数、四参数及两参数模型<sup>[8,9]</sup>。目前,人们提出了多种全局运动估计算法<sup>[2,10,11]</sup>,这里以四参数模型为例,分析线性参数模型的迭代算法。常用的四参数模型是摄像机运动模型,可表示为

$$\begin{bmatrix} V_X \\ V_Y \end{bmatrix} = \begin{bmatrix} a_1 & -a_2 \\ a_2 & a_1 \end{bmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{bmatrix} a_3 \\ a_4 \end{bmatrix}, \quad (1)$$

(1)式表明,摄像机的四参数运动模型的参数可以由图像的运动矢量场来估计。理想的情况下,只需两个像素点的运动矢量即可求出(1)式中的四个参数(即  $a_1, a_2, a_3, a_4$ )。但是块匹配方法无法得到子像素级的运动矢量,因此需要更多像素的运动矢量参与估计。也就是说,参数估计是一个解矛盾方程组的问题,一般常用最小二乘法求解四个参数的近似值。

### 2.2 参数估计的迭代算法

设:所有参与估计的像素集合为  $A$ ,则  $A$  中任一点  $(X, Y)$  的模型估计误差可以表示为

$$\begin{aligned} E_1 &= \sum_{(X,Y) \in A} (V_X - a_1 X + a_2 Y - a_3)^2, \\ E_2 &= \sum_{(X,Y) \in A} (V_Y - a_1 X + a_2 Y - a_4)^2. \end{aligned} \quad (2)$$

为得到  $E_1, E_2$  的最小值,由

$$\frac{\partial E_1}{\partial a_1} = 0, \quad \frac{\partial E_1}{\partial a_2} = 0, \quad \frac{\partial E_1}{\partial a_3} = 0, \quad (3)$$

$$\frac{\partial E_2}{\partial a_1} = 0, \quad \frac{\partial E_2}{\partial a_2} = 0, \quad \frac{\partial E_2}{\partial a_4} = 0. \quad (4)$$

可得到

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} \sum_{(X,Y) \in A} X^2 & - \sum_{(X,Y) \in A} XY & \sum_{(X,Y) \in A} X \\ \sum_{(X,Y) \in A} XY & - \sum_{(X,Y) \in A} Y^2 & \sum_{(X,Y) \in A} Y \\ \sum_{(X,Y) \in A} X & - \sum_{(X,Y) \in A} Y & \sum_{(X,Y) \in A} 1 \end{bmatrix}^{-1} \times \begin{bmatrix} \sum_{(X,Y) \in A} V_X X \\ \sum_{(X,Y) \in A} V_X Y \\ \sum_{(X,Y) \in A} V_X \end{bmatrix}, \quad (5)$$

$$\begin{bmatrix} a_1 \\ a_2 \\ a_4 \end{bmatrix} = \begin{bmatrix} \sum_{(X,Y) \in A} Y^2 & \sum_{(X,Y) \in A} XY & \sum_{(X,Y) \in A} Y \\ \sum_{(X,Y) \in A} XY & \sum_{(X,Y) \in A} X^2 & \sum_{(X,Y) \in A} X \\ \sum_{(X,Y) \in A} Y & \sum_{(X,Y) \in A} X & \sum_{(X,Y) \in A} 1 \end{bmatrix}^{-1} \times \begin{bmatrix} \sum_{(X,Y) \in A} V_Y Y \\ \sum_{(X,Y) \in A} V_Y X \\ \sum_{(X,Y) \in A} V_Y \end{bmatrix}. \quad (6)$$

从(5)式或者(6)式可以计算出参数  $(a_1, a_2)$ ,为 提高准确性可取其均值。

为得到最优的像素集合  $A$ ,一般在迭代估计参数中引入运动矢量残差

$$\delta(X, Y) = \sqrt{(V_X - V'_X)^2 + (V_Y - V'_Y)^2}, \quad (7)$$

其中  $V_X$  和  $V_Y$  为  $(X, Y)$  点原始解码运动矢量,  $V'_X$  和  $V'_Y$  为全局运动估计后  $(X, Y)$  点的估计运动矢量,可表示为

$$\begin{bmatrix} V'_X \\ V'_Y \end{bmatrix} = \begin{bmatrix} a_1 & -a_2 \\ a_2 & a_1 \end{bmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{bmatrix} a_3 \\ a_4 \end{bmatrix}. \quad (8)$$

具体实现时,首先对初始集合  $A$ ,利用(5)式和(6)式估计模型参数  $(a_1, a_2, a_3, a_4)$ ,再用该模型参数计算运动矢量残差图,将  $A$  更新为运动矢量残差小于阈值  $T$  的所有像素的集合;在下一轮迭代中,应用更新后的集合  $A$  重新估计模型参数,再利用更新后的模型参数计算运动矢量残差图,进一步更新  $A$ ;依此迭代,收敛后得到一个稳定的集合  $A$  和最终的模型参数。

### 3 基于背景块搜索的全局运动参数估计

由于摄像机的运动,导致视频中不仅存在运动物体的自主运动,而且还存在背景的不自主运动,同时运动物体的自主运动中也叠加了摄像机的运动。可以看出,背景的运动完全是由摄像机运动引起的,他们并不自发运动。文献[12]对背景运动信息进行了详细的分析,摄像机各种运动对背景运动矢量的影响如图 1 所示。

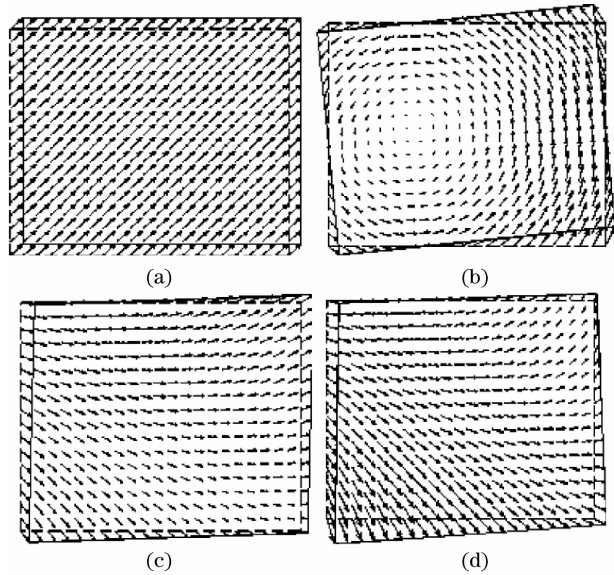


图 1 摄像机各种运动对背景运动矢量的影响  
Fig.1 Influence of various camera motion on background motion vector

由图 1 可以看出对于不同的摄像机运动方式,虽然背景的运动矢量有差异,但总体上都具有一定的相似性。根据这种相似性,提出了一种建立基于背景集合的运动目标提取算法。算法如下:

- 1) 从 MPEG-4 视频里获取运动矢量信息;
- 2) 建立背景集合: 这是一个搜索的过程,算法假定左上角宏块为背景宏块,接着从左上角宏块(种子)开始,以  $3 \times 3$  宏块集合作为滑动窗口,搜索背景宏块,直到构成一片连通区域。搜索过程如下:

(i) 归并: 对于  $3 \times 3$  窗口中的八个比较块,按照右、下、左、上、右上、右下、左下、左上的顺序依次与目标块(中心块)的运动矢量比较其相似性,包括方向和大小,当其差异小于指定阈值时,将其归并到目标块集合中;

(ii) 增长: 对于目标块,如果未找到下一个合适的种子,则转步骤(iii)。对于某个比较块,如果其与中心块属于同一集合,那么指定该比较块为新的

目标块(种子),转步骤(i);

(iii) 在已生成的背景宏块集合中,对所有未作过种子的宏块,再次实行归并和增长;

3) 在业已建立的背景集合上,应用 2.2 节所述的迭代算法,使用四参数模型进行全局运动参数估计,得到最终的模型参数( $a_1, a_2, a_3, a_4$ );

4) 运动目标提取。

对帧中每一个宏块( $X, Y$ ),按照(7)式计算运动矢量残差  $\delta(X, Y)$ ,残差小于阈值的宏块被认为是背景,否则是前景即运动目标。

### 4 实验结果

为测试算法的效果,采用 Microsoft 的 MPEG-4 参考编解码软件对经常用于全局运动估计研究的 Foreman 和 Coastguard 序列进行编码。Foreman 为  $352 \text{ pixel} \times 288 \text{ pixel/frame}$ ,序列长度为 58 帧; Coastguard 序列格式为  $352 \text{ pixel} \times 288 \text{ pixel/frame}$ ,序列长度为 52 帧。视频流帧结构序列均为“IPBBPBB……”。

图 2、图 3 所示为本文算法与文献[9]算法对视频序列 Foreman 第 32 帧、第 46 帧的检测结果对比,其中绿色部分代表检测出的运动宏块。从中可以看出文献[9]的检测结果不完整,而本文算法

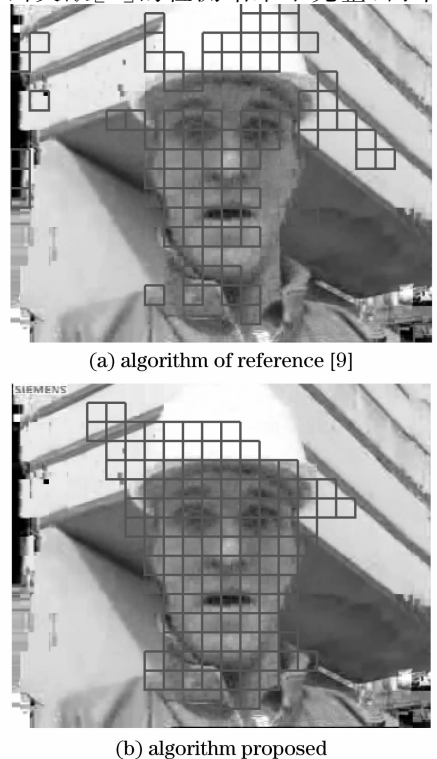


图 2 Foreman 第 32 帧的检测结果对比  
Fig.2 Detection results of the 32<sup>th</sup> frame of Foreman

得到的检测结果较前者有很大改善,所得到的运动目标的整个包络更加完整,并且不会出现不连续的现象。

在对检测效果进行比较的基础上,对本文的算

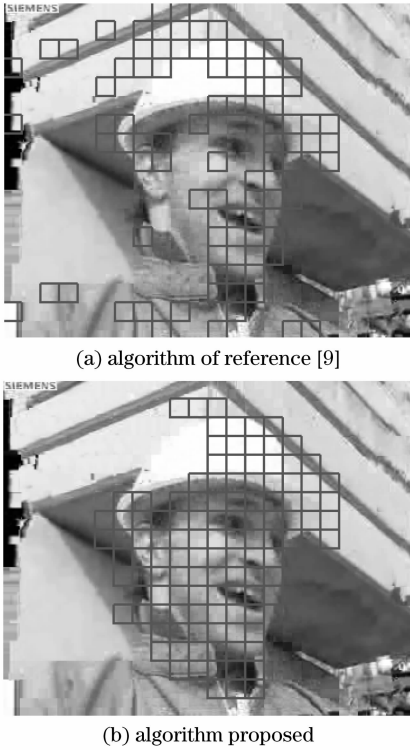
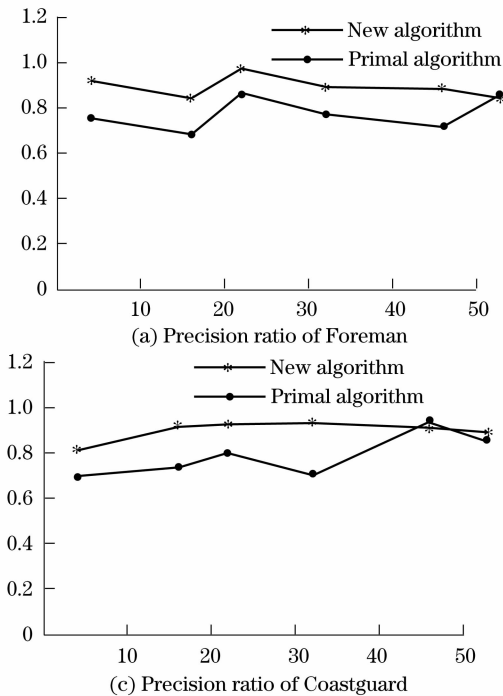


图 3 Foreman 第 46 帧的检测结果对比  
Fig. 3 Detection results of the 46<sup>th</sup> frame of Foreman



法效率进行了对比。在 Pentium 4, 2.80 GHz, 512 MB RAM 的计算机上对同样的视频流进行处理,其估计时间如表 1 所示。

表 1 文献[9]算法与本文算法检测时间对比结果

Table 1 Detection time of the algorithms of reference[9] and the proposed

Video sequence	Algorithm of reference[9] /s	Algorithm proposed /s
Foreman	10.997	10.793
Coastguard	18.597	17.347

由表 1 可以看出,本文算法的速度优于文献[9],能够快速得出运动参数,但对 Coastguard 序列的检测性能提升较 Foreman 序列小,是因为 Coastguard 的背景变化较 Foreman 大,因而其运动矢量的大小、方向相似性较低。

为了更进一步评估算法的性能,还引入了精确率和漏检率等评估准则。精确率和漏检率的定义如下:

$$R_{PS} = \frac{n_H}{n_H + n_F}, \quad (9)$$

$$R_{RC} = \frac{n_M}{n_H + n_M}, \quad (10)$$

其中  $R_{PS}$ 、 $R_{RC}$ 、 $n_H$ 、 $n_F$  和  $n_M$  分别表示一帧中检测结果的精确率、漏检率、运动块个数、伪运动块个数和伪静止块个数。Foreman 和 Coastguard 视频序列的精确率及漏检率如图 4 所示。

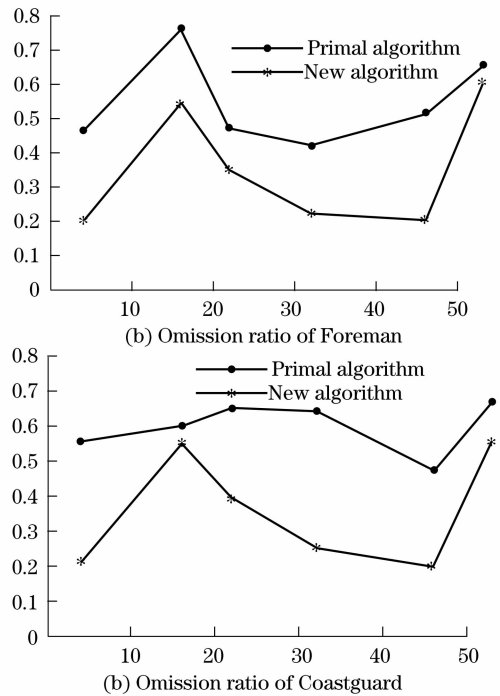


图 4 视频序列 foreman,coastguard 的精确率和漏检率  
Fig. 4 Precision and omission ratio of foreman and coastguard

实验结果表明,本文算法估计镜头运动参数的速度更快、精度更高,从而提高了检测精确率,降低了漏检率,而且该算法能够很好地排除局部运动信息的干扰,具有较高的稳健性。

## 5 结 论

全局运动参数估计在运动目标提取过程中极其重要。根据 MPEG 压缩码流的特点和全局运动模型的适用范围,提出了一种改进的全局运动估计迭代模型。该模型通过搜索背景块,建立合理的背景宏块集合,迭代估计全局运动参数。算法利用 MPEG-4 压缩域中的信息进行全局运动估计,不需要解码整个压缩码流,提高了算法效率;采用基于背景搜索的方法仅对运动背景区域进行全局运动参数估计,有效地避免前景中的局部运动带来的误差,具有很高的稳健性,得到了较高的检测精确率,而漏检率较低。

## 参 考 文 献

- 1 Peter M. Kuhn, SONY Corporation. Camera motion estimation using feature points in mpeg compressed domain[C]. *IEEE, Internat. Conf. on Image Processing*, 2000, 596~599
- 2 RATH G B. MAKUR A. Iterative least squares and compression based estimation for a four-parameter linear global motion model and global motion compensation[J]. *IEEE Trans Circuits and System for Video Technology*, 1999, **9**(7):1075~1099
- 3 Xu Gongyi, Cai Jingju, Su Ding. Detection method for moving target in complex environment [J]. *Infrared and Laser Engineering*, 2007, **36**(6): 992~995  
徐功益,蔡敬菊,宿 丁. 复杂背景下运动目标检测方法[J]. *红外与激光工程*, 2007, **36**(6): 992~995
- 4 Ren Chen, Zhang Tanping. New method for detecting of moving targets based on Kalman filter theory [J]. *Opto-Electronic Engineering*, 2007, **34**(4): 7~11  
任 臣,张覃平. 基于 Kalman 滤波理论的运动目标检测新方法[J]. *光电工程*, 2007, **34**(4): 7~11
- 5 Zha Chengdong, Wang changsong, Kong Xianfeng *et al.*. Moving target detection based on adaptive background model[J]. *Opto-Electronic Engineering*, 2008, **35**(1): 26~30  
查成东,王长松,巩宪锋 等. 基于自适应背景模型的运动目标检测[J]. *光电工程*, 2008, **35**(1): 26~30
- 6 Wang Kunpeng, Zhang Xiaohu, Li Lichun *et al.*. New method for moving target detection based on positive and negative difference images[J]. *J. Appl. Opt.*, 2007, **28**(5): 521~525  
王鲲鹏,张小虎,李立春 等. 一种基于正负差图像的运动目标检测新方法[J]. *应用光学*, 2007, **28**(5): 521~525
- 7 Jiang huajun. Visual Motion Analysis and Feature Extraction [D]. Changchun: Master's thesis of Jilin University, 2004  
江华俊,视频图像运动特征的分析与提取[D]. 长春: 吉林大学硕士论文, 2004
- 8 Chiawen Lin, Zhihong Ling, Yuancheng Chang. Compressed domain fall incident detection for intelligent home surveillance [J]. *IEEE, Circuits and Systems*, 2005, **4**: 3781~3784
- 9 Chen Hanfen, Qi Feihu. Double-iteration algorithm for motion field based global motion estimation [J]. *J. China Institute of Communications*, 2004, **25**(6):126~131  
陈韩锋,戚飞虎. 基于运动矢量场的双迭代全局运动估计方法[J]. *通信学报*, 2004, **25**(6): 126~131
- 10 Zhibo Chen, Zhengang Nie, Xiaodong Gu *et al.*. Fast global motion estimation based on iteration least-square estimation with sustained symmetrical structure[C]. Proc. IEEE International Symposium on Circuits and Systems, 2006. 4695~4698
- 11 Guo Li, Gong Shengrong, Cui Zhiming. A fast global motion estimation algorithm[J]. *J. Image and Graphics*, 2007, **12**(12): 2119~2126  
郭 丽,龚声蓉,崔志明. 一种快速全局运动估计算法[J]. *计算机图形图像学报*, 2007, **12**(12): 2119~2126
- 12 Yeping Su, Mingting Sun, Vincent Hsu. Global motion estimation from coarsely sampled motion vector field and the applications [J]. *IEEE, Circuits and Systems for Video Technology*, 2005, **15**: 232~242