

# 基于 MPEG 压缩域的快速、鲁棒的全局运动信息提取方法

李宏亮, 刘贵忠, 李永利, 张忠伟

(西安交通大学电子与信息工程学院, 信息与通信工程系, 陕西西安 710049)

**摘要:** 全局运动信息在视频分析中起着重要的作用. 本文根据 MPEG 编码特点, 提出了一种从 MPEG 压缩域中快速有效地进行全局运动参数估计的算法. 该算法充分利用了 MPEG 压缩码流中的信息, 通过提取预测残差 DC 图像的运动背景区域, 估计全局运动参数, 从而保证了参数估计的准确性, 有效地克服了已有文献中仅仅采用运动矢量进行全局运动估计的局限性. 根据不同的 MPEG 测试序列的对比分析, 结果表明, 本算法可快速准确地对 MPEG 视频序列进行全局运动信息估计, 同时具有很高的鲁棒性.

**关键词:** 镜头运动; 背景估计; MPEG 视频

**中图分类号:** TN911.73 **文献标识码:** A **文章编号:** 0372-2112 (2004) 01-0121-04

## Efficient Global Motion Estimation Method for MPEG Video

LI Hong-liang, LIU Gui-zhong, LI Yong-li, ZHANG Zhong-wei

(School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China)

**Abstract:** Motion estimation plays an important part in the analysis of video sequences for the purposes of information extraction. In order to process video data effectively, camera motion extraction from compressed video is often required. In this paper, an effective global motion estimation algorithm based on the moving background derived from MPEG compressed data is presented. The algorithm uses the predictive error DC-image to extract the moving background of the current frame, and then exploits a new global motion estimation method based on a 6-parameters motion model to calculate the camera parameters. In the evaluation experiments, five video test sequences with global motions and three video sequences without global motions are used to test the proposed algorithm, and compared with an existing method. The experimental results show our algorithm can estimate global motion parameters more effectively and exhibit more robustness than the existing algorithm.

**Key words:** camera motion; background estimation; MPEG video

## 1 引言

随着计算机和通信技术的发展, 诸如数据的存贮、检索和流媒体网络控制等多媒体信息系统也在不断地更新完善, 被广泛地应用到远程教学、数字图书馆、远程医疗和多媒体在线新闻等可视信息系统中. 大量的存贮和传输的图像视频数据都是采用不同压缩标准完成的, 诸如 MPEG、JPEG、H. 261 和 H. 263 等标准. 如何从图像和视频的压缩数据中提取出关键信息, 提高网络资源的利用率, 快速有效地实现视频管理、数据检索和场景分析等, 已成为视频分析首要解决的问题.

目前, 很多的视频分析方法是通过帧间图像的亮度、色彩、纹理以及物体轮廓形状等特征信息不同, 提取出场景切换帧<sup>[1-3]</sup>. 这些方法对于检测突变 (abrupt) 场景都能起到很好地效果. 然而, 大部分视频序列中存在很多由于镜头运动而产生地渐变 (Gradual) 场景, 对于这种场景的检测至今仍然研究得较少, 而且很难准确地采用突变场景的检测方法进行分析.

近几年, 针对未压缩视频序列中, 提出了一些基于镜头运动模型参数来估计全局运动 (GM) 的方法<sup>[4-6]</sup>. 在压缩域内, Tan 和 Saur 等<sup>[7]</sup>通过抽取 MPEG 码流中宏块的运动矢量, 提出了一种快速的镜头运动参数检测算法. 但该算法很容易造成全局运动的错误估计.

本文从 MPEG 压缩码流中获得的宏块运动矢量以及预测误差 DC 系数, 采用 6 参数的镜头运动模型, 提出了一种快速而有效的全局运动检测方法. 通过对不同视频序列测试, 并与文[7]中的算法进行对比, 实验结果表明, 本文提出的方法具有很高的鲁棒性, 可对全局运动信息进行准确的估计. 本文的组织结构安排如下: 在文章第二部分中, 首先介绍了预测误差 DC 图像以及运动背景的提取. 在此基础上, 第三部分推导出镜头运动模型参数的估计方法. 文章第四部分选取了三组标准视频测试序列, 对本文提出的分类方法进行实验验证, 并与文献[7]的方法进行了对比. 最后, 在第五部分中给出了本文的结论.

## 2 预测误差 DC 图像以及运动背景的提取

### 2.1 预测误差 DC 图像

MPEG<sup>[8]</sup> 编码引入了三种主要的图像类型. 帧内编码帧 (I 帧), 这种类型帧采用帧内空域压缩方法, 提供了对编码图像数据序列的访问点. 预测编码帧 (P 帧) 通过过去的 I 帧或 P 帧进行补偿预测, 再对预测误差图像进行空域压缩. 双向预测编码帧 (B) 需要过去和未来的参考帧来进行运动补偿, 而本身不能作为预测参考帧. 这些不同类型的帧按一定方式排列就构成了 MPEG 视频序列组 (GOP), 一个 GOP 必须以 I 帧开始, 后面可跟任意个 I 和 P 帧. 每帧都是由三个整数矩阵组成: 一个亮度矩阵 (Y) 和两个色度矩阵 (Cb 和 Cr). 对于 MPEG 2 编码, 帧是一个顶场和一个底场的合并. 对于 I 帧中的每个

宏块, 均采用了帧内编码. 而对于 P 和 B 帧内的宏块, 或者采用帧内编码, 或者采用帧间编码. 对于帧间编码的宏块, 通过该宏块的运动补偿, 找到参考帧中相应的预测宏块, 对预测残差进行 DCT 变换编码.

P 帧内的每个宏块均采用了前向预测, 如果宏块与前向参考帧的预测宏块的残差大于设定的阈值, 该宏块采用帧内 DCT 变换编码. 如果小于给定的阈值, 此宏块采用预测残差的 DCT 变换编码. 同样, B 帧内的每个宏块与 P 帧类似, 只是多采用了后向预测, 可提高压缩效率. 由此可知, 通过对 P 和 B 帧的位流解码, 可获得 DCT 系数的预测残差 DC 图像, 而该图像正是当前帧与参考帧经过运动补偿 (MC) 后所得到的差值图, 大小是原始图像的六十四分之一.



图 1 Foreman0 和 Foreman3 的 DC 差分对比分析

### 2.2 运动背景的提取

一般情况下, 称不存在局部运动的区域为背景, 前景代表局部运动的区域. 所以全局运动参数应该从背景中估计, 要尽量避免采用前景的运动信息. 目前, 对于静态背景的估计算法<sup>[9~10]</sup>主要应用在静态背景中运动物体的分割及识别等. 对于存在全局运动的背景的提取, HU 和 Uchimura<sup>[11]</sup>提出了基于镜头 3-D 运动分析的运动背景检测方法. 而在 MPEG 编码序列中, 为了获得高的压缩比, P 帧和 B 帧都采用了预测方式进行编码. 帧内宏块通过运动估计在预测参考帧中找到相匹配的宏块, 对该块预测后的残差进行编码. 所以预测帧的 DC 残差图是运动补偿后获得的, 这样通过采用一般的背景提取方法就可把运动背景区域估计出来.

以 Foreman 测试序列的第 0 帧到第 3 帧为例. 伴随镜头 Pan Left 运动同时, 还存在着人物头部的局部运动. 图 1 (a) ~ (d) 分别给出了 Foreman 测试序列的第 0 帧 (I 帧) 和第 3 帧 (P 帧) 的 DC 图. 图 1 (c) 是未经过运动补偿 DC 图像残差图. 从图中可以看出, 一些运动背景信息仍然保留得很清楚. 而图 1 (d) 是通过 MPEG 编码位流中得到的 DC 残差系数图, 图中的背景信息通过运动补偿后, 幅值已经很小, 保留下来的基本上是前景中头部的局部运动信息. 这样采用一般的背景提取方法, 通过设定适当的阈值, 就可很好的提取出运动背景区域点, 获得该帧的背景区域. 如图 1 (e) 所示, 黑色区域表示具有全局运动的背景, 白色区域表示局部运动的前景.

## 3 镜头运动模型参数的估计方法

### 3.1 摄像机的运动<sup>[12]</sup>

如图 2 所示, 摄像机是把 3-D 空间点 (X, Y, Z) 映射到 2-D 空间平面点 (x, y) 上. 图像平面与 Z-轴垂直, 中心坐标为 (0, 0, f), 这里 f 表示摄像机的焦距. 映射关系可表示为:

$$x = \frac{fX}{Z}, \quad y = \frac{fY}{Z}$$

摄像机的运动主要有: Pan (摄像机绕 Y 轴的角度水平旋转运动), Track (水平轴的运动导致了图像的水平移动), Tilt (摄像机绕 X 轴的角度垂直旋转运动), Boom

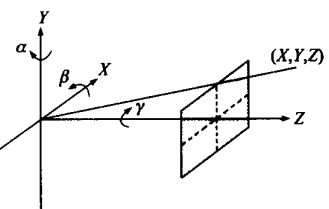


图 2 摄像机运动成像模型

(沿着垂直轴的运动导致了图像的垂直移动)、Zoom (由摄像机焦距改变而产生的 zoom out 和 zoom in 的图像中心的仿射运动) 和 Dolly (水平后向移动).

### 3.2 摄像机运动模型参数

本文采用 6 个参数的摄像机运动模型<sup>[4-7]</sup>, 如式 (1) 所示:

$$\begin{cases} x_i = a_0 + a_1 x_i + a_2 y_i + & (a) \\ y_i = a_3 + a_4 x_i + a_5 y_i + & (b) \end{cases} \quad (1)$$

这里点 (x<sub>i</sub>, y<sub>i</sub>) 表示当前帧像素点的坐标, 点 (x<sub>i</sub>, y<sub>i</sub>) 表示参考帧像素点的坐标. 其中, 参考帧内像素点的坐标位置可由当前帧像素坐标和宏块运动矢量求得. 坐标原点为图像中心. 当 a<sub>1</sub> = a<sub>5</sub> = 1 时, 无 Zoom 发生; 当 a<sub>1</sub> = a<sub>5</sub> < 1 时, 为 Zoom out; 当 a<sub>1</sub> = a<sub>5</sub> > 1 时, 为 Zoom in. a<sub>2</sub>/a<sub>1</sub> 和 a<sub>4</sub>/a<sub>5</sub> 分别表示两帧之间镜头的旋转的变化. a<sub>0</sub>/a<sub>1</sub> 和 a<sub>3</sub>/a<sub>5</sub> 分别表示两帧之间镜头的 Pan 和 Tilt 旋转. 当 a<sub>2</sub> = a<sub>4</sub> = 0, a<sub>1</sub> = a<sub>5</sub> 时, 即转化为 3 参数的运动估计模型.

对于式 (1) 中的参数 (a<sub>0</sub>, a<sub>1</sub>, a<sub>2</sub>, a<sub>3</sub>, a<sub>4</sub>, a<sub>5</sub>) 的估计, 可运用二元线性回归的方法求解. 首先可采用最小二乘法求 a<sub>0</sub>, a<sub>1</sub>, a<sub>2</sub> 的估计量. 作离差平方和

$$Q = \sum_{i=1}^n (x_i - a_0 - a_1 x_i - a_2 y_i)^2$$

这里  $n$  表示背景像素点的数目.

选择  $a_0, a_1, a_2$  使  $Q$  达到最小,即  $Q = \min$ . 通过求偏导,

即:

$$\begin{cases} \frac{\partial Q}{\partial a_0} = -2 \sum_{i=1}^n (x_i - a_0 - a_1 x_i - a_2 y_i) = 0 \\ \frac{\partial Q}{\partial a_1} = -2 \sum_{i=1}^n (x_i - a_0 - a_1 x_i - a_2 y_i) x_i = 0 \\ \frac{\partial Q}{\partial a_2} = -2 \sum_{i=1}^n (x_i - a_0 - a_1 x_i - a_2 y_i) y_i = 0 \end{cases}$$

上式可转化为  $\hat{a} = A^+ B$  的方式求解. 这里系数矩阵  $A^+$  是矩阵  $A$  的 Moore-Penrose 广义逆或着称作矩阵  $A$  的极小最小二乘  $g$  逆. 对式(1)中的式(b)也可采用同样的方法估计出参数  $a_3, a_4, a_5$ . 由于本文提出的方法是采用背景中的像素点的运动矢量进行全局运动估计,不考虑前景中局部运动信息,因此具有很高的鲁棒性.

### 4 实验研究

为了在实际 MPEG 压缩流中测试上面提出的镜头运动参数估计算法,本文把该方法应用到三组标准测试序列中,如表 1 所示. 我们首先得到不同序列的原始  $Y, U, V$  图像,选用 25/s 的帧率和 5Mb/s 的码率,采用 MPEG2 对测试序列编码, GOP 格式为 IBBPBBPBBPBB,长为 12. 选用基于 P 帧中的信息进行全局运动参数估计. 对于压缩码流,首先对得到的 P 帧中的预测误差 DC 系数,采用第二部分的背景估计算法,提取出属于

运动背景区域的象素点及其运动矢量. 然后利用背景区内象素点,根据第三部分提出的 6 参数镜头运动模型进行运动参数估计. 实验中我们选取了一组仅存在局部运动信息的测试序列 Container 与文献[7]中的算法进行对比.

表 1 各种测试视频序列的规格及全局运动信息

序列	Size	Num	Rate	GOP	Zoom	Pan	Tilt
Container	176 × 144	280	25	12	No	No	No
Coastguard	176 × 144	300	25	12	in	Left, Right	Up
Football	360 × 240	210	25	12	No	Left, Right	Up, Down

Container 序列中存在货船的局部向右平移的局部运动(图 3). 全局运动参数的估计结果如图 4 所示,图中带 \* 的虚线表示采用文献[7]算法的估计结果,而实线表示采用本文方法的参数计算结果. 从图中可清楚的看到文献[7]算法估计的全局运动参数,造成了很大的偏差,而采用本文提出的算法却仍可很好地进行参数估计.

为了验证本文提出的全局运动信息提取算法,我们选取了 2 组具有全局运动信息的标准测试序列 Coast guard, 和 Football, 表 1 给出了相应的全局运动信息. 如图 3 所示, Coast guard 测试序列中存在着镜头的 Pan Left, Pan Right, Tilt up 和 Zoom in 运动,在 football 中存在着摄像机的 Pan 以及 Tilt 运动. 采用本文提出的估计 6 参数估计方法,并与文献[7]中的算法对比. 图 5~6 分别给出了 2 组测试序列的全局运动参数估计结果. 实验结果表明,本文提出的方法可有效地提取出全局运动信息,更好地估计镜头运动参数. 同时可很好地排除在参数估计中局部运动的信息干扰,具有高的鲁棒性.



图 3 Container, Coastguard, 和 Football 部分帧图

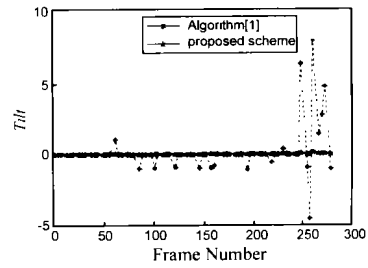
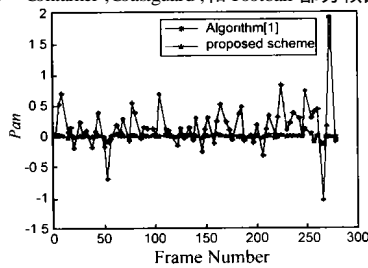
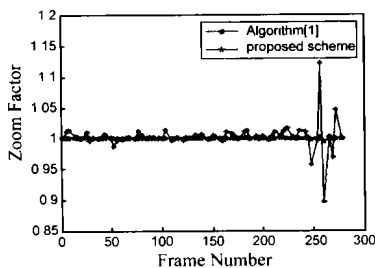


图 4 Container 测试序列的对比结果

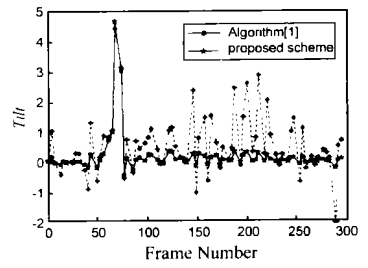
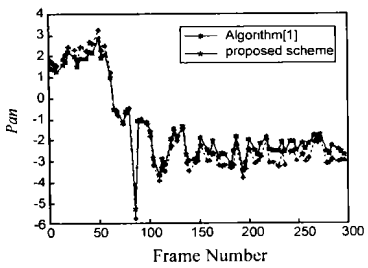
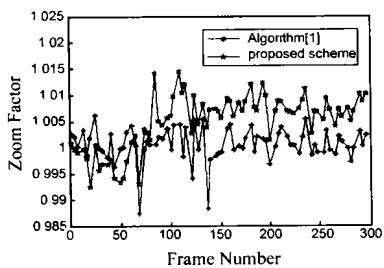


图 5 Coastguard 测试序列对比结果

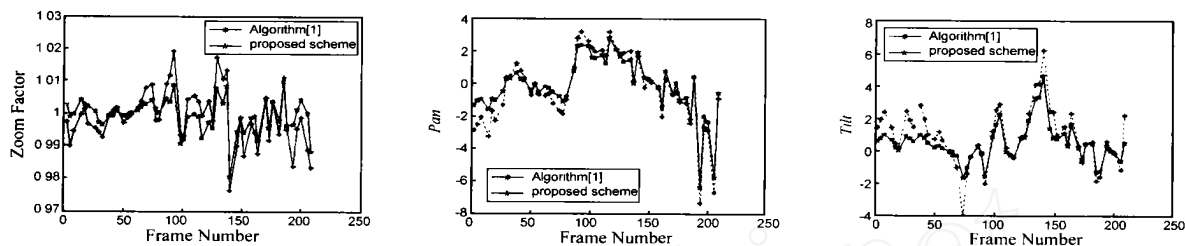


图6 Football 测试序列对比结果

## 5 结论

全局运动参数估计在因镜头运动产生的缓变场景的检测中具有重要的作用. 本文根据 MPEG 压缩码流特点, 提出了在压缩域内采用预测帧的残差 DC 系数, 进行运动背景提取. 根据背景区域内宏块运动矢量, 采用 6 参数的镜头运动模型, 估计全局运动参数. 该算法主要的优点表现在:

- \* 采用 MPEG 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, 再在整个像素域中进行运动分析, 只需部分解压缩, 因此可节省大量运算时间, 实现快速的运动参数估计.

- \* 采用运动背景区域进行全局运动参数估计, 可有效地避免前景中的局部运动带来的误差, 准确地进行参数估计, 具有很高的鲁棒性.

- \* 根据本文提出的镜头运动模型参数估计方法的实验测试结果表明, 该算法可快速有效地对全局运动信息进行提取, 提高了基于镜头运动的缓变场景检测效果.

本文提出的方法, 就是针对于场景分析的一个重要特征-全局运动, 从压缩域中进行分析 and 检测. 在下面的工作中, 可把得到的估计算法应用到视频数据后续处理中, 比如, 根据估计的全局运动信息, 提取出缓变场景关键帧. 为基于内容的视频检索以及基于内容的动态资源分配奠定很好的前期处理基础.

### 参考文献:

- [1] J Meng, Y Juan, S F Chang. Scene change detection in a MPEG compressed video sequence [J]. Digital Video Compression: Algorithms and Technol., Feb. 1995, SPIE 2419: 14 - 25.
- [2] B Yeo, B Liu. Rapid scene analysis on compressed video [J]. IEEE Trans. Circuits Systems Video Technol., 1995, 5(6): 533 - 544.
- [3] H J Zhang, C Y Low, S W Smiliar. Video parsing and browsing using compressed data [J]. Multimedia Tools and Applicat., Mar. 1995, 1(1): 89 - 111.
- [4] F Dufaux, J Konrad. Efficient, robust, and fast global motion estimation for video coding [J]. IEEE Trans. Image Process, Mar. 2000, 9(3): 497 - 501.
- [5] G Gunta, U Mascia. Estimation of global motion parameters by complex linear regression [J]. IEEE Trans. Image Process, Nov. 1999, 8(11): 1652 - 1657.

- [6] K Y Yoo, J K Kim. A new fast local motion estimation algorithm using global motion [J]. Signal Processing Jul. 1998, 68(2): 219 - 224.
- [7] Yap-Peng Tan, D D Saur, S R Kulkarni, P J Ramadge. Rapid estimation of camera motion from compressed video with application to video annotation [J]. IEEE Trans. Circuits Syst. Video Technol., 2000, 10(1): 133 - 145.
- [8] ISO/IEC 13813: 'Information technology — Generic coding of moving pictures and associated audio (MPEG-2)' [S]. 1995.
- [9] E Durucan, T Ebrahim. Change detection and background extraction by linear algebra [J]. Proceedings of the IEEE, Oct. 2001, 89(10): 1368 - 1381.
- [10] K E Matthews, N M Namazi. A Bayes decision test for detecting uncovered background and moving pixels in images sequences [J]. IEEE Trans. Image Process, May 1998, 7(5): 720 - 728.
- [11] Z Hu, K Uchimura. Moving objects detection from time-varied background: an application of camera 3D motion analysis [A]. IEEE Second International Conference on 3-D Digital Imaging and Modeling [C]. Ottawa, Canada: IEEE, 1999, 59 - 67.
- [12] F M Idris, S Panchanathan. Detection of camera operation in compressed video sequences [A]. Storage and Retrieval for Image and Video Databases [C]. San Jose, California, Proc. SPIE 3022, 1997, 493 - 505.

### 作者简介:



李宏亮 男, 1970 年 8 月出生于河南省, 西安交通大学电子与信息工程学院在读博士生, 主要研究兴趣包括: 多媒体通信、小波分析、时频分析、非平稳信号处理、视频图像编码.



刘贵忠 男, 1962 年 6 月出生于陕西省, 1989 年 6 月获荷兰 Eindhoven 大学博士学位, 现为西安交通大学电子与信息学院教授、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、反演等方面的理论与应用研究.