

# 基于 MPEG 的全局运动估计技术\*

詹翊强 戚飞虎 詹劲峰 张文杰

(上海交通大学计算机科学与工程系, 上海, 200030)

**摘要** 提出了一种基于 MPEG (moving picture experts group) 的全局运动估计技术: 利用四参数模型来描述由摄像机运动带来的全局运动, 再利用基于块匹配的算法来求解 4 个参数, 进而作相应的运动补偿。由于 MPEG 视频流中的运动矢量蕴含了最优块匹配的信息, 所以不需要对 MPEG 视频流完全解码, 即可进行参数估计。实验证明该算法速度快, 并能通过运动补偿排除由摄像机运动造成的伪运动区域的干扰。

**关键词** MPEG, 全局运动估计, 运动矢量。

## GLOBAL MOTION ESTIMATION TECHNOLOGY BASED ON MPEG\*

ZHAN Yi-Qiang QI Fei-Hu ZHAN Jing-Feng ZHANG Wen-Jie

(Department of Computer Science and Engineering, Shanghai Jiaotong University, Shanghai 200030, China)

**Abstract** A global motion estimation technology based on MPEG was put forward. First, a four-parameter model, which is used to describe the global motion yielded by camera motion, was proposed. Then these parameters were obtained by using an algorithm which is based on block-matching, and the global motion could be compensated by these four parameters. Since the motion vectors include the information of the optimum match-blocking, these information can be used without completely decoding the MPEG video sequence. The experiment shows that the algorithm is fast, and feint motion field can be excluded by global motion compensation.

**Key words** MPEG, global motion estimation, motion vector.

### 引言

在视频流的记录过程中, 摄像机都会不可避免做出 zoom (缩放)、pan (水平运动)、tilt (竖直运动) 和 roll (旋转运动) 等运动。由摄像机运动产生的全局运动会扩大视频流中的信息量, 给人以愉悦的视觉感受, 但是, 从目标跟踪、视频压缩、场景分析等数字视频技术的角度来看, 这些全局运动将会带来各种负面影响。要解决这个问题, 就必须对摄像机引起的全局运动进行估计并在进行相邻帧匹配时作相应的补偿。目前, 国际上已有相当多的学者在此领域作了大量研究工作<sup>[1~5]</sup>。

### 1 基于 MPEG 的全局运动估计思想及算法框图

摄像机运动在视频流中留下的唯一痕迹就是全局光流场的变化, 因此, 全局运动估计的本质就是如何通过全局光流场的变化来反向推导摄像机运动参数。要达到这个目的, 首先要建立一个摄像机运动模型。一个好的摄像机运动模型需要符合两个条件: (1) 该模型能够反映摄像机的所有运动状态; (2) 该模型参数应尽可能地简化, 并且可以根据视频流中全局光流场的变化进行计算。

建立了摄像机运动参数模型后, 可通过光流方程来求解运动参数。但是这种算法的复杂度太大, 考虑到实际视频流中相邻像素点的运动相似性, Gagan B Rath<sup>[4]</sup>和 R De Min Wang 等人<sup>[6]</sup>等提出了基于相邻帧块匹配的方法。联想到 MPEG 流中的运动矢量包含了最优块匹配信息, 对于 MPEG 流可以不进行完全解码, 即利用运动矢量进行全局运动

\* 国家自然科学基金 (编号 60072029) 资助项目  
稿件收到日期 2001-03-23, 修改稿收到日期 2001-06-25

\* The project supported by the National Natural Science Foundation of China (No. 60072029)  
Received 2001-03-23, revised 2001-06-25

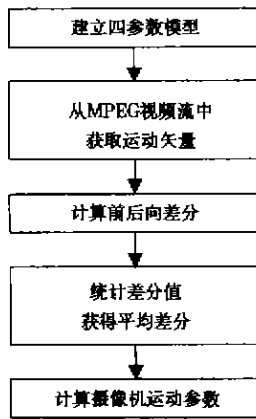


图1 摄像机运动参数估计算法框图  
Fig.1 Diagram of camera motion parameters estimation

估计,算法框图如图1所示.

## 2 四参数全局运动模型

### 2.1 摄像机运动类型及其数学表述

假设摄像机本身位置固定不动,则其运动可分为4种: zoom、pan、tilt 和 roll(如图2所示).以摄像机镜头为坐标轴的原点,以摄像机的视角的方向为  $z$  轴建立一个三维的笛卡尔坐标系,在该坐标系中可以对这4种动作作描述为:(1) Zoom 是摄像机焦距改变所引起的全局运动,(2) Pan 是摄像机关于摄像机坐标系  $x$  轴的转动,(3) Tilt 是摄像机关于摄像机坐标系  $y$  轴的转动;(4) Roll 是摄像机关于摄像机坐标系  $z$  轴的转动.摄像机运动可以表示为

$$u' = SRu + T \quad (1)$$

式中  $u$  和  $u'$  分别为摄像机坐标系中的一个静止点在摄像机运动前后不同的坐标下的空间位置, $S$  为尺度因子, $R$  为摄像机三维旋转矩阵, $T$  为摄像机运动相对于坐标原点的偏移.

在笛卡尔坐标中的三维旋转可以用相对于3个

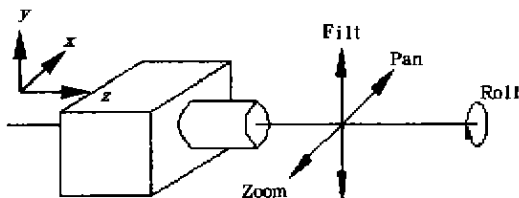


图2 摄像机运动  
Fig.2 Camera motions

坐标轴旋转的欧拉角来表示,即

$$R = R_{\theta_z} \cdot R_{\theta_y} \cdot R_{\theta_x} \quad (2)$$

其中

$$R_{\theta_x} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{bmatrix},$$

$$R_{\theta_y} = \begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y \\ 0 & 1 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y \end{bmatrix},$$

$$R_{\theta_z} = \begin{bmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

### 2.2 相关坐标系

为了更明确地表示摄像机的运动与视频流中全局运动的关系,在讨论摄像机各种运动模型建立之前,需要引入两个坐标系:(1) 摄像机坐标系,该系统是一个三维的笛卡尔坐标系,摄像机镜头为坐标轴的原点,其中  $z$  轴为摄像机的视角的方向;(2) 图像空间坐标系,该系统是一个二维坐标系,其坐标平面与摄像机坐标系的  $z$  轴垂直,其原点与摄像机坐标模型原点的距离  $F$  为摄像机焦距.坐标系关系如图3所示.

其坐标关系为:  $X/F = x/z, Y/F = y/z$ .其中  $F$  是摄像机的焦距.

分析了摄像机运动种类并建立相应数学表述后可以发现摄像机的运动依照摄像机坐标系是否变动,可以分为两类:(1) pan、tilt 和 roll 这3种运动中摄像机关于摄像机坐标轴  $x, y$  和  $z$  的旋转运动,实际上,摄像机坐标系的  $z$  轴已经发生了变化.(2) 在 zoom 运动中,摄像机本身不存在物理运动,只是改变焦距,因此摄像机坐标系不变,改变的不仅仅是图像空间坐标.

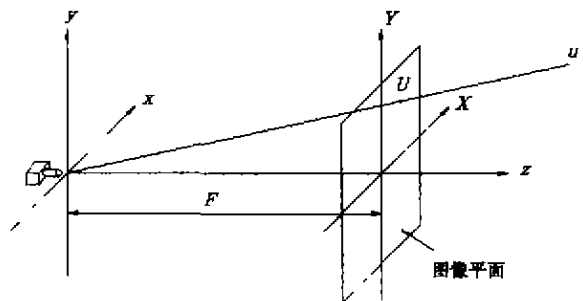


图3 摄像机坐标系和图像空间坐标  
Fig.3 The camera coordinate and the image-space coordinate

2.3 Pan、Tilt、Roll 运动模型

pan, tilt 和 roll 是摄像机关于摄像机坐标轴  $x$ 、 $y$  和  $z$  的旋转运动, 刚体在笛卡尔坐标中的三维旋转可以用相对于 3 个坐标轴旋转的欧拉角来表示, 以 tilt 运动为例, 有

$$u' = R_y \cdot u. \tag{3}$$

由于在实际视频流中, 帧采样频率通常可以达到 30 帧/s, 因此摄像机在两帧图像之间时间内的 Tilt 旋转角度  $\theta_t$  都是很小的, 对于小角度  $\theta_t$ ,  $\cos\theta_t \approx 1$ ,  $\sin\theta_t \approx \theta_t$ , 则根据式(1), 有

$$\begin{cases} x' = x - \theta_t \cdot z, \\ y = y', \\ z' = -\theta_t \cdot x + z. \end{cases} \tag{4}$$

由于  $X = Fx/z, Y = Fy/z$ , 所以  $X' = Fx'/z', Y' = Fy'/z'$ , 又因为  $\theta_t X/F \leq 1$ , 可进一步简化为  $X' = X + F\theta_t, Y' = Y$ . 其矩阵表示为

$$\begin{pmatrix} X' \\ Y' \end{pmatrix} = \begin{pmatrix} 1 & -\theta_t \\ \theta_t & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}. \tag{5}$$

同理摄像机的 Pan 运动和 Roll 运动可以表示为

$$\begin{pmatrix} X' \\ Y' \end{pmatrix} = \begin{pmatrix} 1 & -\theta_r \\ \theta_r & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}. \tag{6}$$

摄像机的 Roll 运动可以得到

$$\begin{pmatrix} X' \\ Y' \end{pmatrix} = \begin{pmatrix} 1 & -\theta_z \\ \theta_z & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}. \tag{7}$$

2.4 Zoom 运动模型

Zoom 运动中, 摄像机本身位置不变, 只是通过焦距长度变化引起的图像空间的改变, 因此, 它可以表示为摄像机坐标系统不变, 图像空间坐标沿摄像机坐标系统的  $z$  轴移动, 如图 6 所示. Zoom 运动的矩阵表示为

$$\begin{pmatrix} X' \\ Y' \end{pmatrix} = \begin{pmatrix} f_z & 0 \\ 0 & f_z \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}. \tag{8}$$

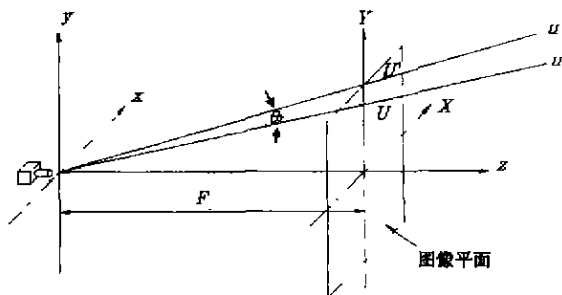


图 4 摄像机 tilt 运动示意图  
Fig. 4 Sketch map of tilt motion of camera

其中, Zoom 因子  $f_z = F'/F$ ,  $F$  和  $F'$  是摄像机 Zoom 前后的焦距.

2.5 摄像机运动四参数模型

综合 tilt, pan, roll 和 zoom 运动的数学表述(5)~(8), 可以将摄像机运动表述为

$$\begin{pmatrix} X' \\ Y' \end{pmatrix} = \begin{pmatrix} f_z & -f_z\theta_t \\ f_z\theta_t & f_z \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} F\theta_t \\ F\theta_r \end{pmatrix}, \tag{9}$$

令  $a_1 = f_z, a_2 = f_z\theta_t, a_3 = F\theta_t, a_4 = F\theta_r$ , 得

$$\begin{pmatrix} X' \\ Y' \end{pmatrix} = \begin{pmatrix} a_1 & -a_2 \\ a_2 & a_1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} a_3 \\ a_4 \end{pmatrix}. \tag{10}$$

$a_1, a_2, a_3, a_4$  4 个参数的物理意义分别为:  $a_1$  代表摄像机 zoom 运动,  $a_2$  代表摄像机 zoom 和 roll 运动,  $a_3$  代表摄像机 tilt 运动,  $a_4$  代表摄像机 pan 运动.

3 基于 MPEG 的模型参数估计快速算法

我们根据模型式(9)研究全局光流场与摄像机运动模型参数的定量关系. 对式(9)作移项, 令  $a_1 = f_z - 1, a_2 = f_z\theta_t, a_3 = F\theta_t, a_4 = F\theta_r$ , 得

$$\begin{pmatrix} V_x \\ V_y \end{pmatrix} = \begin{pmatrix} a_1 & -a_2 \\ a_2 & a_1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} a_3 \\ a_4 \end{pmatrix}. \tag{11}$$

$V_x, V_y$  表示像素的运动矢量. 对式(11)两边取偏导, 得

$$\begin{cases} \frac{\partial}{\partial X} V_x = \frac{\partial}{\partial Y} V_y = a_1, \\ \frac{\partial}{\partial X} V_y = -\frac{\partial}{\partial Y} V_x = a_2. \end{cases} \tag{12}$$

式(12)成立的条件是视频流中光流场完全由摄像机的运动产生, 而没有独立运动的物体. 若存在由独立运动的物体产生的局部光流场, 那么等式所基于的前提将被破坏, 进而导致计算结果的误差. 但是, 在很多应用场合视频流具有两个特点: (1) 运动物体占图像整体比例很小, (2) 物体的运动近似于刚体

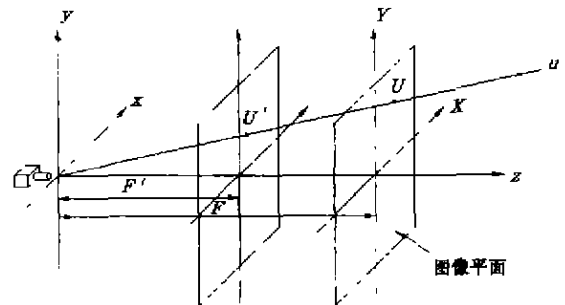


图 5 摄像机 zoom 运动示意图  
Fig. 5 Sketch map of camera zoom

运动. 基于这两个假设, 重新来分析局部运动对式(12)的影响.

运动矢量可以被分解为两部分, 一部分由全局运动产生, 另一部分由物体的局部运动产生, 故

$$\begin{bmatrix} V_x \\ V_y \end{bmatrix} \text{ 可以表示为 } \begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} V_{xv} \\ V_{yv} \end{bmatrix} + \begin{bmatrix} V_{xl} \\ V_{yl} \end{bmatrix}. \quad (13)$$

其中,  $\begin{bmatrix} V_{xv} \\ V_{yv} \end{bmatrix}$  为全局运动矢量,  $\begin{bmatrix} V_{xl} \\ V_{yl} \end{bmatrix}$  为局部运动矢量.

根据刚体运动的特点, 在刚体内部的局部运动矢量都是一致的, 因此, 在刚体内部有

$$\begin{aligned} \frac{\partial}{\partial X} V_x &= V_x(X+1, Y) - V_x(X, Y) \\ &= V_{xv}(X+1, Y) - V_{xv}(X, Y) + \\ &\quad V_{xl}(X+1, Y) - V_{xl}(X, Y) \\ &= V_{xv}(X+1, Y) - V_{xv}(X, Y). \end{aligned} \quad (14)$$

同理,

$$\frac{\partial}{\partial Y} V_y = V_{yv}(X, Y+1) - V_{yv}(X, Y). \quad (15)$$

即在刚体内部, 局部运动对运动矢量的偏导没有影响, 表达式(12)依然成立. 由于局部运动光流只在运动目标的边缘影响运动矢量的偏导, 而运动目标本身占图像整体比例又很小, 所以那些不同于全局运动矢量偏导的运动矢量数量很少, 使用统计的方法可以排除这些运动矢量对摄像机运动参数估计的影响.

在以上的讨论中,  $V_x, V_y$  均表示像素的运动矢量. 考虑到空间邻接区域具有运动状态相似性, 完全可以利用块(块大小取  $16 \times 16$ )的运动矢量来代表像素的运动矢量. 根据 MPEG 标准<sup>[7]</sup>, 可以在对 MPEG 视频流中  $P$  帧和  $B$  帧中宏块进行运动矢量解码时获得  $V_x, V_y$ . 这里尽管忽略了  $I$  帧, 但由于  $I$  帧的出现频率不高(一般一个图像组 15 帧图像中, 有一帧  $I$  帧), 对算法结果影响不大.

在运动矢量的基础上做前向差分, 并以差分的均值作为偏微分, 这样做的好处是可以克服噪声的影响, 即

$$\begin{aligned} \frac{\partial}{\partial X} V_x &= \frac{1}{2} \{ \hat{V}_x(X+1, Y) - \hat{V}_x(X, Y) + \\ &\quad \hat{V}_x(X, Y) - \hat{V}_x(X-1, Y) \}, \\ \frac{\partial}{\partial Y} V_y &= \frac{1}{2} \{ \hat{V}_y(X, Y+1) - \hat{V}_y(X, Y) + \\ &\quad \hat{V}_y(X, Y) - \hat{V}_y(X, Y-1) \}. \end{aligned}$$

$$\begin{aligned} &\hat{V}_x(X, Y) - \hat{V}_x(X, Y-1) \}, \\ \frac{\partial}{\partial X} V_y &= \frac{1}{2} \{ \hat{V}_y(X-1, Y) - \hat{V}_y(X, Y) + \\ &\quad \hat{V}_y(X, Y) - \hat{V}_y(X-1, Y) \}, \\ \frac{\partial}{\partial Y} V_x &= \frac{1}{2} \{ \hat{V}_x(X, Y-1) - \hat{V}_x(X, Y) + \\ &\quad \hat{V}_x(X, Y) - \hat{V}_x(X, Y-1) \}. \end{aligned} \quad (16)$$

然后, 利用统计的方法计算平均偏微分, 并利用其结果结合式(12)估计摄像机运动参数  $a_1$ , 有

$$a_1 = \frac{1}{i-j} \left[ \sum_{|\frac{\partial}{\partial X} V_x - \bar{m}| < T-1} \frac{\partial}{\partial X} V_x + \sum_{|\frac{\partial}{\partial Y} V_y - \bar{m}| < T-1} \frac{\partial}{\partial Y} V_y \right]. \quad (17)$$

其中  $T$  为统计均方差

$$m = \frac{1}{2K} \sum_{i,x,y} \left( \frac{\partial}{\partial X} V_x + \frac{\partial}{\partial Y} V_y \right). \quad (18)$$

$\Omega$  为图像块中心点领域,  $K$  为块的个数. 同理, 可以估计参数  $a_2$ .

在  $a_1, a_2$  基础上可以按照式(19)计算参数  $a_3, a_4$ , 根据式(15)计算 zoom 因子和绕  $z$  轴旋转角度, 即

$$\begin{aligned} a_1 &= V_x - a_3 X + a_4 Y, \\ a_2 &= V_y - a_3 - a_4 Y. \end{aligned} \quad (19)$$

## 4 实验结果

我们选取了两段长度为 1min 的 CIF (352 \* 288) 格式的 MPEG-1 视频流作为实验样本, 第一段视频流的没有运动物体, 只有摄像机作左右 pan 运动; 第二段视频流中除了有摄像机作左右 pan 运动外, 还有两个人, 其中一个人基本静止不动, 另一个则作独立于摄像机运动的局部运动.

利用上述算法对两段 MPEG 视频流进行全局运动补偿实验, 在 P II-400, 128M RAM 的 PC 上进行.

第一段视频流 1520 帧, 耗时 208s, 平均处理速度 7.3 帧/s, 实验结果如图 6 所示. 其中, 113 帧和 296 帧时, 摄像机作自左向右的 Pan 运动, 615 帧时, 摄像机作自右向左的 pan 运动. 从实验结果中可以看到, 由摄像机运动引起的伪运动目标经过运动补偿后大大减少.

第二段视频流 1745 帧, 耗时 237s, 平均处理速度 7.4 帧/s, 实验结果如图 7 所示. 其中, 109 帧时,

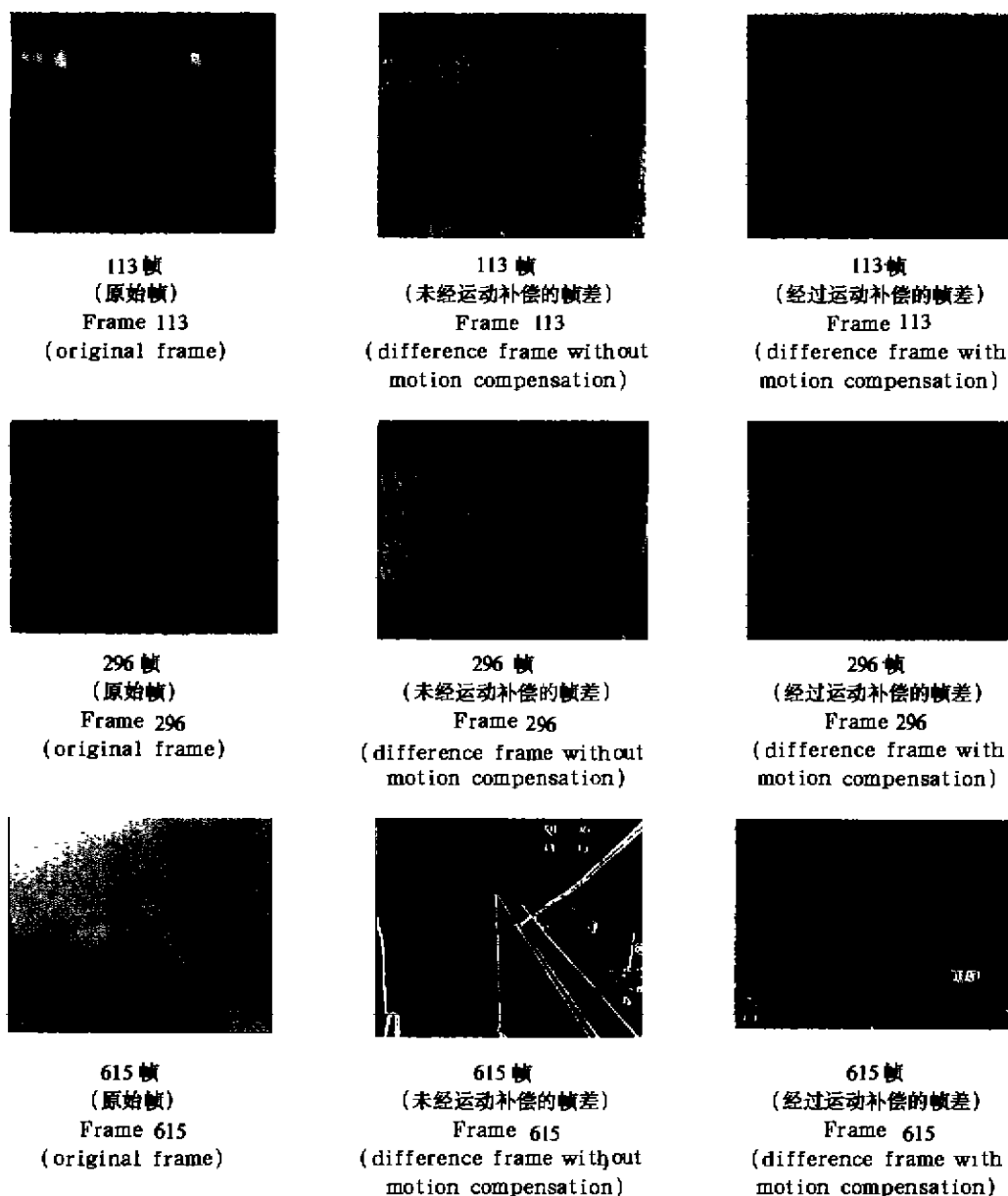


图 6 全局运动估计和补偿实验结果(视频流 1)

Fig. 6 Result of global motion estimation and compensation (video sequence 1)

摄像机作自右向左的 pan 运动, 视频流中的运动目标作与摄像机的同向运动。331 帧时, 摄像机作自右向左的 pan 运动, 视频流中的运动目标作与摄像机的反向运动, 从实验结果中可以看到, 由摄像机运动引起的伪运动目标在经过运动补偿后大大减少, 而对视频流中实际运动的物体影响不大。

## 5 小结

全局运动估计是数字视领域中的一个关键的技术, 本文通过分析摄像机运动和全局光流场的关系,

构造一个摄像机的四参数全局运动模型, 并根据 MPEG 视频流的特点, 提出了一种基于 MPEG 的参数估计快速算法, 实验证明, 该算法获得了比较理想的结果, 可以使目标跟踪<sup>[7]</sup>、目标分析等后继处理获得更好的效果。

## REFERENCES

- [1] Nicolas H, Labit C. Global motion identification for image sequence analysis and coding. *Proc. of ICASSP* 91, 1991: 2825-2828
- [2] Chiou-Shann Fuh, Petros Maragos. Affine models for

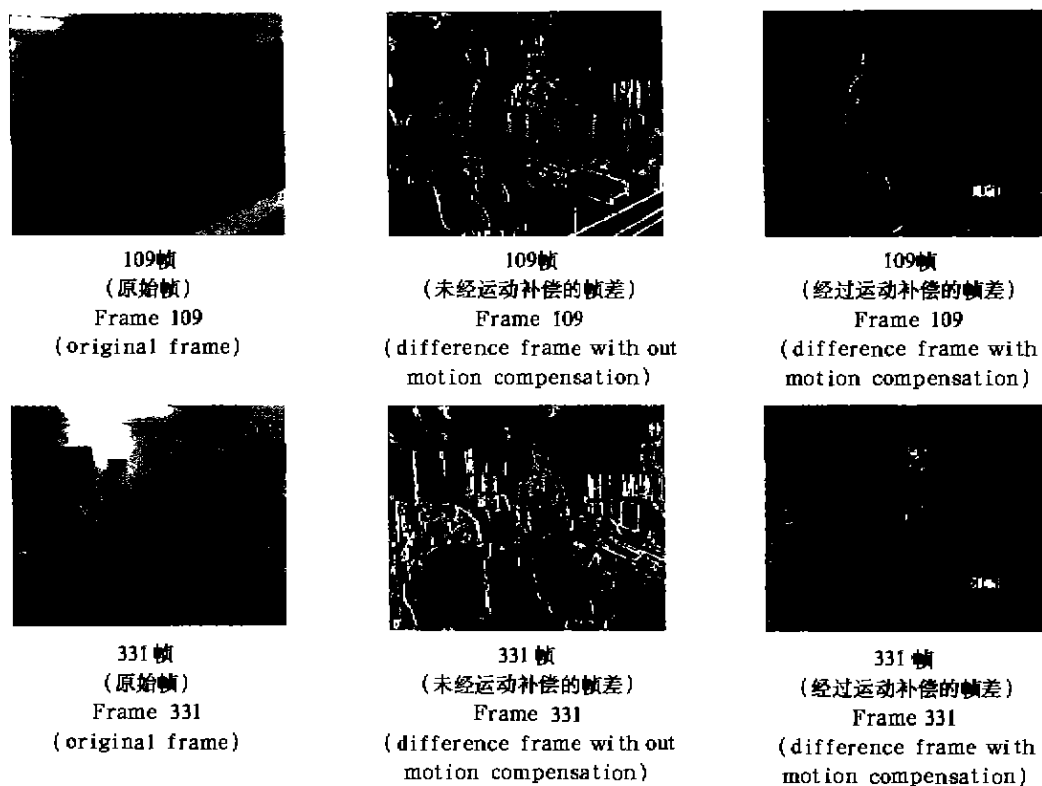


图7 全局运动估计和补偿实验结果(视频流2)

Fig. 7 Result of global motion estimation and compensation (video sequence 2)

- image matching and motion detection. *Proc. of ICASSP'91*, 1991, 2409—2412
- [3] Srinivasan M V, Venkatesh S, Hoste R. Qualitative estimation of camera motion parameters for video sequences. *Pattern Recognition*, 1997, **30**: 593—606
- [4] Rath Gagan B, Makur Anamitra. Iterative least squares and compression based estimations for a four-parameter linear global motion model and global motion compensation. *IEEE Trans. on Circuits and Systems for Video Technology*, 1999, **7**: 1075—1099
- [5] Michael Hoetter. Differential estimation of the global motion parameters zoom and pan. *Signal Processing*, 1989, **16**: 249—265
- [6] Demin Wang, Claude Labit. Morphological spatio-temporal simplification for video image segmentation. *Signal Processing: Image Communication*, 1997: 161—170
- [7] Lorenzo Favalli, Alessandro Mecocci, Fulvio Moschetti. Object tracking for retrieval application in MPEG-2. *IEEE Transactions on Circuits and Systems for Video Technology*, 2000, **10**(3): 427—432