

联合运动估计的多视点视频视差估计新方法^①

胡波, 闫晓勇, 吴建平, 胡杰, 肖志坚

(浙江东方职业技术学院 工程技术系, 温州 325011)

摘要: 针对多视点视频编码中运动估计和视差估计运算量大的特点, 提出了一种联合运动估计的多视点视频视差估计方法. 在时域上对每个视点的宏块的运动状态建立 Kalman 滤波器, 预测当前宏块的运动矢量. 在运动估计基础上结合空域信息, 分析了视差矢量和运动矢量的几何关系, 计算了当前宏块的视差矢量. 实验结果表明, 本文方法和全搜索算法及快速算法相比, 在大幅节省了编码时间基础上又提高了率失真性能.

关键词: 多视点视频编码; 运动估计; 视差估计; Kalman 滤波; 几何关系; 视差矢量

New Disparity Estimation Method for Multiview Video Combined with Motion Estimation

HU Bo, YAN Xiao-Yong, WU Jian-Ping, HU Jie, XIAO Zhi-Jian

(Department of Engineering Technology, Zhejiang Dongfang Vocational and Technical College, Wenzhou 325011, China)

Abstract: The computation complexity of motion estimation and disparity estimation for multiview video encoding was so huge. In order to resolve the problem, a new method for disparity estimation was proposed in the paper, which was combined with motion estimation. In the time domain, Kalman filtering was established for motion state of each view's macroblock, the motion vector of current macroblock was predicted by Kalman filtering. The thesis analyzed the geometry relationship of motion vector and disparity vector in the space domain based on motion estimation, and calculated the present macroblock's parity vector. Experimental results show that the method proposed in this paper not only saves encoding time substantially but also keeps rate distortion, compared to the full search and fast algorithm.

Key words: multiview video coding; motion estimation; disparity estimation; Kalman filtering; geometry relationship; disparity vector

多视点视频编码技术是 3DTV(立体电视)、FTV(自由视点电视)关键技术之一^[1,2], 近年来越来越多受到国内外学者的关注. 最简单的多视点视频编码方法是对每个视点的视频用 H.264/AVC 进行编码^[3], 但该方法没有考虑各视点间包含的大量冗余信息, 因此压缩效率并不高. 由于运动估计和视差估计在整个编码过程中消耗巨大, 考虑在运动估计的基础上结合视点之间的相关性进行视差估计可以提高编码效率^[4,5]. Kim 等^[6]利用视差矢量和运动矢量的几何关系来预测运动矢量, 但该方法需要增加额外的视点信息, 因此限制了其使用范围. Ding 等^[7]根据邻近视点的运动矢量并结合视差矢量预测当前宏块的运动矢量, 但视差估计仍采用全搜索算法, 编码效率还有待提高. Lai 等^[8]在已编码的视差矢量中利用运

动矢量来预测视差矢量, 该算法有效地降低了计算量, 但其运动估计的精度在计算中至关重要. 本文采用 Kalman 滤波器预测运动矢量, 提高了估计的精度, 通过运动矢量和视差矢量的几何关系来计算当前宏块的视差矢量, 使率失真性能和编码效率均获得提高.

1 多视点视频编码结构

目前视频编码结构基本上有 Simulcast^[9]、顺序视点预测、GoGOP 预测^[10]、M 帧预测和分层 B 帧(Hierarchical B Pictures, HBP)^[11]等预测结构, 其中 HHI 提出的 HBP 是现阶段的主流结构. 本文采用的 HBP 结构如图 1 所示, 其中 $S_0 \sim S_7$ 表示视点个数为 8, $T_0 \sim T_7$ 表示 GOP 的 8 个时刻. 图中箭头表示帧间参考方向,

① 收稿时间:2013-01-13;收到修改稿时间:2013-02-25

如 S_1T_1 帧在视点方向上参考了 S_0T_1 帧和 S_2T_1 帧, 在时间方向上参考了 S_1T_0 帧和 S_1T_2 帧.

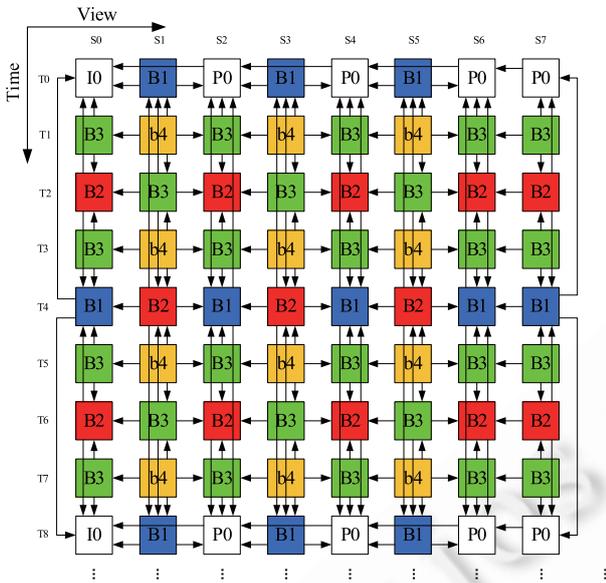


图 1 基于 HBP 的多视点编码结构

2 Kalman滤波预测运动矢量

目前运动估计快速搜索算法^[3]有 3 步搜索法 (TSA), 新 3 步搜索法 (NTSA), 二维对数搜索 (2DLOG), 菱形搜索 (DS) 等, 但是不可避免会面临局部最优问题, 对视频的质量影响较大. 本文对各视点宏块分别建立 Kalman 滤波器来预测其运动矢量. Kalman 滤波器采用线性最小均方误差作为最优准则, 根据时变随机信号的统计特性递推滤波计算未来值, 使其达到最接近真值的估计, 其估计过程通过状态方程和观测方程来描述:

$$X(k) = AX(k-1) + w(k-1) \quad (1)$$

$$Z(k) = HX(k) + v(k) \quad (2)$$

其中, A 、 H 分别表示状态转移矩阵和测量矩阵, $w(k-1)$ 、 $v(k)$ 表示过程噪声和测量噪声, 均为高斯白噪声. 这里将宏块在间隔 T_0 内的运动近似分解为受随机加速扰动的直线运动^[12], 建立的状态向量 $[x \ v_x \ y \ v_y]$ 表示宏块在图像中的位移及移动速度, 则宏块的状态方程和观测方程为:

$$\begin{bmatrix} x(k) \\ v_x(k) \\ y(k) \\ v_y(k) \end{bmatrix} = \begin{bmatrix} 1 & T_0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x(k-1) \\ v_x(k-1) \\ y(k-1) \\ v_y(k-1) \end{bmatrix} + \begin{bmatrix} T_0^2/2 \\ T_0 \\ T_0^2/2 \\ T_0 \end{bmatrix} \quad (3)$$

$$\begin{bmatrix} x(k) \\ y(k) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x(k) \\ v_x(k) \\ y(k) \\ v_y(k) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} v(k) \quad (4)$$

通过上述方程就可以预测得到宏块的位移即运动矢量.

3 运动矢量和视差矢量相关性分析

预测视差矢量主要有帧间预测矢量、帧内相邻块预测矢量以及上层模式的预测矢量. 经过统计发现帧间预测视差矢量相关性最强^[13], 因此有必要对其进行分析. 如图 2 所示, 按从左往右从上往下顺序依次为 t 时刻 S_i 和 S_{i+1} 视点的图像(参考帧), $t+T$ 时刻 S_i 和 S_{i+1} 视点的图像, $P_{i,t}$ 表示 t 时刻 S_i 视点中宏块所处的位置. $DV(t)$ 、 $DV(t+T)$ 表示 t 时刻和 $t+T$ 时刻宏块的视差矢量, $MV(i)$ 、 $MV(i+1)$ 表示 i 视点和 $i+1$ 视点宏块的运动矢量. 通过几何关系可以得到:

$$DV(t) = P_{i+1,t} - P_{i,t}, \quad DV(t+T) = P_{i+1,t+T} - P_{i,t+T} \quad (5)$$

$$MV(i) = P_{i,t+1} - P_{i,t}, \quad MV(i+1) = P_{i+1,t+1} - P_{i+1,t} \quad (6)$$

从上述两式可以得到:

$$DV(t+T) - DV(t) = MV(i+1) - MV(i) = P_{i+1,t+T} - P_{i,t+T} - P_{i+1,t} + P_{i,t} \quad (7)$$

即

$$DV(t+T) = DV(t) + MV(i+1) - MV(i) \quad (8)$$

通过上式分析发现当前帧的视差矢量由参考帧的视差矢量以及两者的运动矢量组成. 通过对两个视点分别建立 Kalman 滤波器就可以得到其运动矢量 $MV(i+1)$ 和 $MV(i)$, 同时结合 t 时刻的视差矢量 $DV(t)$ 就可以得到视差矢量 $DV(t+T)$, 将该值作为当前帧宏块的视差矢量的初始值进行下一步匹配, 可以大大减少搜索点数.

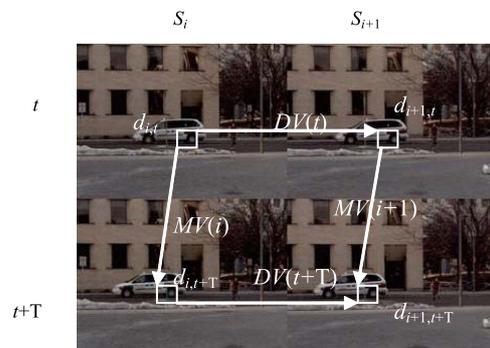


图 2 视差估计及运动估计关系图

4 实验结果及分析

本文在多视点视频编码参考模型 JMVM7.0 平台上进行测试研究, 硬件环境为 Intel(R) Core(TM) 2 Duo 2.39 GHz CPU, 1G 内存的 PC 机. 多视点测试序列由 MERL 的 Ballroom 和 Exit 序列, KDDI 的 Race1 序列提供. 通过和 JMVM7.0 全搜索算法以及 JMVM7.0 快速算法比较来验证本文方法的有效性, 评价指标有峰值信噪比(PSNR)和码率(Bitrate), 编码时间. 为了突出

比较效果, 通过定义 ρ 、 δ 和 γ 来进一步表示:

$$\rho = 1 - \frac{T_1}{T_2}, \delta = PSNR_1 - PSNR_2, \gamma = \frac{BR_1}{BR_2} - 1 \quad (9)$$

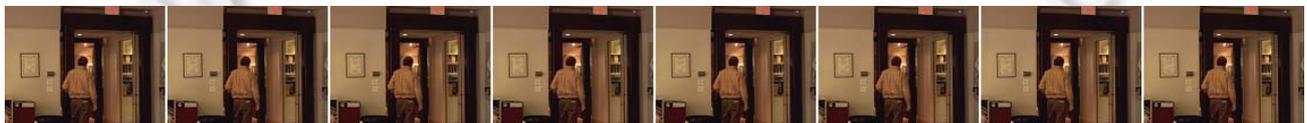
其中 T_1 和 T_2 表示本文方法和 JMVM7.0 全搜索算法或 JMVM7.0 快速算法的编码时间, $PSNR$ 和 BR 定义类似. 显然 δ 为正数或 γ 为负数表示率失真性能更好. 测试环境如表 1 所示, 测试序列如图 3 所示, 测试结果如表 2 所示.

表 1 测试环境列表

| 编解码器 | MV 搜索范围 | Basis QP | DeltaLayerXQuant | | | | | | | 编码结构 | GOP 长度 | 编码帧数 |
|---------|----------|----------|------------------|---|---|---|---|---|--------|---------|--------|-------|
| | | | 0 | 1 | 2 | 3 | 4 | 5 | | | | |
| JMVM7.0 | ±96 | 22,27, | 0 | 1 | 2 | 3 | 4 | 5 | HBP | 8 | 41 | |
| | | 32,37 | 0 | 3 | 4 | 5 | 6 | 7 | | | | |
| 测试序列 | 分辨率 | 序列属性 | | | | | | | 间距(cm) | 帧率(fps) | 阵列属性 | |
| MERL | Ballroom | 640×480 | 大视差 | | | | | | | 19.5 | 25 | 1D/平行 |
| MERL | Exit | 640×480 | 静止区域较多 | | | | | | | 19.5 | 25 | 1D/平行 |
| KDDI | Race1 | 320×240 | 运动剧烈 | | | | | | | 20 | 30 | 1D/平行 |



(a) Ballroom 序列



(b) Exit 序列



(c) Race1 序列

图 3 多视点视频测试序列

表 2 本文算法与 JMVM7.0 全搜索算法、JMVM7.0 快速算法比较

| 测试序列 | QP 值 | 本文算法和 JMVM7.0 全搜索算法比较 | | | 本文算法和 JMVM7.0 快速算法比较 | | |
|----------|------|-----------------------|---------------|--------------|----------------------|---------------|--------------|
| | | ρ (%) | δ (db) | γ (%) | ρ (%) | δ (db) | γ (%) |
| Ballroom | 22 | 95.02 | 0.01 | -0.04 | 8.5 | 0.06 | -0.25 |
| | 27 | 95.30 | 0.02 | -0.39 | 9.1 | 0.08 | -0.82 |
| | 32 | 95.53 | 0.03 | -0.48 | 8.8 | 0.12 | -1.12 |
| | 37 | 95.76 | 0.05 | -0.67 | 10.2 | 0.13 | -1.36 |
| Exit | 22 | 95.26 | 0.02 | -0.32 | 8.5 | 0.08 | -0.55 |
| | 27 | 95.62 | 0.04 | -0.63 | 8.3 | 0.10 | -0.98 |
| | 32 | 95.76 | 0.05 | -0.81 | 8.6 | 0.12 | -1.45 |

| | | | | | | | |
|-------|----|-------|------|-------|-----|------|-------|
| | 37 | 96.05 | 0.08 | -1.12 | 8.7 | 0.13 | -1.56 |
| | 22 | 96.02 | 0.01 | -0.21 | 7.6 | 0.05 | -0.43 |
| Race1 | 27 | 96.12 | 0.01 | -0.32 | 7.5 | 0.05 | -0.66 |
| | 32 | 96.25 | 0.02 | -0.56 | 7.8 | 0.06 | -0.98 |
| | 37 | 96.33 | 0.03 | -0.73 | 7.6 | 0.08 | -1.34 |

从表2可以发现本文算法和JMVM7.0全搜索算法相比,编码速度大幅提高,信噪比提高,码率下降,说明其率失真性能有所提高.而与JMVM7.0快速算法相比,编码速度略有提高,信噪比获得较大的提高,码率也下降较多,说明其率失真性能提升较多.这是因为本文方法获得的运动矢量能够有效减少搜索的扩散,大大降低了编码时间.在运动估计中使用了Kalman滤波器,使运动矢量的预测更加准确,从而获得了较高的信噪比.

5 结语

多视点视频需要处理的数据量非常庞大,本文从多视点视频编码中资源消耗最多的运动估计和视差估计入手,提出了一种结合运动估计的多视点视频视差估计方法.针对传统的运动估计及快速算法在搜索宏块时容易陷入局部最优的情况,采用Kalman滤波器预测宏块的运动矢量.分析了多视点视频视差矢量和运动矢量的几何关系,使当前宏块的视差估计更加准确.通过实验发现本文方法和JMVM全搜索法相比在保证率失真性能基础上大幅提高了编码速度,和快速算法相比在编码速度一步提高的基础上改善了率失真性能.

参考文献

- 1 Kalva H, Christodoulou L, Mayorn L, et al. Challenges and opportunities in video coding for 3DTV. Proc. of IEEE International Conference on Multimedia and Expo.2006: 1689-1692.
- 2 Kimata H, Kitahara M, Kamikura K, et al. Low-delay multiview video coding for free-viewpoint video communication. Systems and Computers in Japan,2007,38(5):14-28.
- 3 Wiegand T, Sullivan G, Bjontegaard G, et al. Overview of the H.264/AVC video coding standard.IEEE Transactions on Circuits and Systems for Video Technology, 2003,13(7): 560-576.
- 4 Li XM, Zhao DB, Ma SW, et al. Fast disparity and motion estimation based on correlations for multiview video coding. IEEE Transactions on Consumer Electronics, 2008, 54(4): 2037-2044.
- 5 Shen LQ, Liu Z, Yan T, et al. View-adaptive motion estimation and disparity estimation for low complexity Multi-view Video Coding. IEEE Transactions on Circuits and Systems for Video Technology,2010,20(6):925-930.
- 6 Kim Y, Kim J, Sohn K. Fast disparity and motion estimation for multi-view video coding. IEEE Transactions on Consumer Electronics, 2007,53(2):712-719.
- 7 Ding LF, Tsung PK, Chien SY, et al. Content-aware prediction algorithm with inter-view mode decision for multi-view video coding. IEEE Transactions on Multimedia,2008, 10(8):1553-1564.
- 8 Lai P, Ortega A. Predictive fast motion/disparity search for multiview video coding. Proc. of SPIE Conference Visual Communications and Image Processing,2006:1-11.
- 9 ISO/IEC JTC1/SC29/WG11, Survey of algorithms used for multi-view video coding(MVC).Hong Kong:ISO/IEC,2005.
- 10 Kimata H, Kitahara M, Kamikura K. System design of free viewpoint video communication.International Conference on Computer and Information Technology,2004:52- 59.
- 11 Mueller K, Merkle P, Schwarz H, et al. Multi-view video coding based on H.264/MPEG4-AVC using hierarchical B pictures. Proc. of the Picture Coding Symposium.2006: 385-390.
- 12 胡波.一种时空域联合的机动视频目标精确跟踪方法.计算机应用,2011,31(4):1047-1049.
- 13 朱仲杰.立体视频编码中的关键技术研究[学位论文].杭州:浙江大学,2003.