

一种快速、鲁棒的压缩视频光流场估计算法

尤隽永^① 刘贵忠^① 李宏亮^②

^①(西安交通大学电子与信息工程学院 西安 710049)

^②(香港中文大学可视信号处理与传输实验室)

摘要: 该文针对视频运动特性的快速分析需求,提出了一种压缩域中的光流场估计算法。首先利用DCT变换后的两个AC系数估计出图像的空间偏导数,在此基础上再利用预测残差以及运动矢量对图像的时间偏导数进行估计。另外对编码过程中未进行前向运动预测的宏块特性进行了详细的分析,给出了这些宏块相对于前向参考帧的运动信息估计方法,并对时域不连续的图像块给出了一种修正的偏导数估计方法,以此解决遮挡、切换等现象。然后通过最小二乘法并结合图像的偏导数进行光流场估计。实验表明该方法在准确度上可以达到或超过像素域中的L-K估计方法,比起现有的压缩域估计方法也有一定提高,而在计算时间上相比像素域估计有了大幅度降低。

关键词: 光流场估计; 视频分析; DCT系数; 偏导数估计

中图分类号: TN919.8

文献标识码: A

文章编号: 1009-5896(2007)09-2154-04

A Fast, Robust Optical Flow Estimation Method for Compressed Video

You Jun-yong^① Liu Gui-zhong^① Li Hong-liang^②

^①(School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

^②(Visual Signal Processing and Communication Laboratory, Chinese University of Hong Kong)

Abstract: To analyze quickly the motion information of videos, a fast optical flow estimation algorithm for compressed domain is proposed. Firstly the spatial partial derivatives of image luminance are estimated by using two AC coefficients, and then the predictive residual errors and motion vectors of blocks are appended to estimate the temporal partial derivatives. In addition, after the detailed analysis of those macro-blocks that had no forward motion estimation, the motion information is given approximatively relative to their forward reference frames. And then, a amendatory partial derivative estimation is given for the inconsecutive image blocks when occlusion and cut had happened. Finally, optical flow estimation is performed based on the least square method and partial derivative estimation. Experiments indicated that this method can more accurately estimate the optical flow than the L-K method in pixel domain and the exist method in compressed domain. Moreover the proposed method can greatly reduce the compute time than the estimation in pixel domain.

Key words: Optical flow estimation; Video analysis; DCT coefficient; Partial derivative estimation

1 引言

运动信息是视频内容变化的一个重要部分,运动场是三维物体的实际运动在二维图像平面上的投影,实际上很难直接观测到,通常能得到的是图像的亮度随时间和空间的变化,即光流场。令 $f(x, y, t)$ 表示视频图像的连续时空亮度分布,通常可以假设沿着运动轨迹上图像亮度保持不变,得到如下的光流方程^[1]:

$$\frac{\partial f}{\partial t} = f_x \cdot v_x + f_y \cdot v_y + f_t = 0 \quad (1)$$

其中 $v_x = \partial x / \partial t$ 和 $v_y = \partial y / \partial t$ 分别表示运动速度矢量的两个分量,光流场估计的目的即是估计这两个速度分量。由于

包含两个未知数,一个方程是无法直接求解的,这种不适定问题常被称为“孔径问题”。为了解决“孔径问题”,研究者提出了很多种方法。Lucas等人^[2]假设图像中一块区域中所有像素点光流一致,从而将光流估计转换为最优块匹配约束问题。Nagel附加了一种光滑性约束,使得光流场在沿着其梯度的垂直方向上变化最小,导出了一种迭代估计方法^[3]。此外,还有很多针对“孔径问题”的约束假设,例如基于块的估计方法^[4],利用颜色信息^[5],以及Snyder提出的针对遮挡等特殊情况的平滑约束条件的通用公式^[6]等。

传统的光流场估计是在像素域中进行的,但目前绝大多数的视频和图像是以压缩码流的形式存储传输的,这就提出了对压缩域图像序列进行光流估计的需要。这方面的研究目前还比较少,Coimbra等人^[7]利用运动矢量和交流系数(AC系数)进行MPEG2码流中的光流场估计,但在对偏导数的估

2006-02-20收到,2006-07-31改回

国家自然科学基金(60272072),教育部跨世纪优秀人才基金和国家教育部“十五”“211”工程西安交通大学重点科研项目资助课题

计时存在一定的误差, 同时没有详细分析不同类型宏块的处理方法, 因此估计结果不是非常精确, 尤其在对视频中出现遮挡、切换等情况时该方法会出现较大误差。另外, 由于光流场估计需要利用图像亮度的偏导数信息, 因此如何从压缩视频流中快速提取相应的数据信息也尤为重要, Yeo 等人提出了一种从 MPEG 序列中快速计算 DC 图像的方法^[8]。

本文在分析 DCT 变换和预测编码信息的基础上, 提出了图像亮度偏导数的估计方法, 并对不同类型的宏块进行了详细的分析和处理, 使得光流方程对所有的宏块均可应用, 最后结合块运动估计的思想以及最小二乘法进行光流场的估计。实验表明该方法可以准确、快速地估计出压缩域中光流场信息, 并且对遮挡、切变等现象也具有一定的鲁棒性。

2 基于压缩域的光流场估计

光流场估计的一个关键问题是计算 f_x , f_y 和 f_t , 由于在压缩视频流中存储的是经过预测编码后的残差数据, 因此如何快速计算图像亮度的偏导数是我们要解决的一个首要问题。

2.1 压缩域数据变换

(1) 子块 DCT 系数的快速计算 为了获取图像 DCT 系数等信息, 同时避免复杂的解码过程, Yeo 等人提出了一种从 MPEG 码流中直接计算 DC 系数的方法^[8], 我们将其推广到计算 AC 系数上, 如图 1 所示, 参考块的 AC 系数可由下式近似计算得到

$$A(B_r) = \sum_{i=1}^4 \frac{w_i h_i}{64} A(B_i) \quad (2)$$

其中 $A(B_r)$ 和 $A(B_i)$ 分别表示参考块 B_r 和原始块 B_i 的 AC 系数。

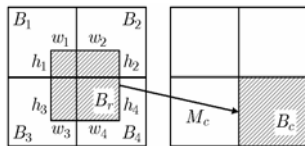


图 1 参考块 (B_r), 当前块 (B_c), 原始块 ($B_i, i = 1, 2, 3, 4$) 和运动矢量 (M_c) 示意图

另外, 如果当前块在 B 帧中并采用双向预测模式, 将其前后两个参考块的 AC 系数分别计算出来, 然后取中值作为计算结果。记当前块和参考块之间的残差图像矩阵为 B_d , 它的 DCT 系数从压缩码流中可以直接得到。由于 DCT 是一种线性变换, 则当前块的近似 AC 系数可由下式计算:

$$A(B_c) = A(B_r) + A(B_d) \quad (3)$$

其中 $A(B_d)$ 表示残差块 B_d 的 AC 系数。

(2) 压缩域中的偏导数估计 图像函数 $f(x, y)$ 的空间偏导数 f_x 和 f_y 表示的是图像在 x 和 y 方向上的变化情况, 而图像的 AC 系数也可以在一定程度上表达图像像素值的变化水平, 因此我们考虑利用 AC 系数来估计 f_x 和 f_y 。根据子块的 DCT 变换定义, 第(1,0)个 AC 系数的形式如下:

$$\begin{aligned} F_{1,0} &= \frac{1}{4\sqrt{2}} \sum_{x=0}^7 \sum_{y=0}^7 \cos \frac{(2x+1)\pi}{16} f(x, y) \\ &= \frac{1}{4\sqrt{2}} \sum_{x=0}^7 \cos \frac{(2x+1)\pi}{16} \sum_{y=0}^7 f(x, y) \end{aligned} \quad (4)$$

因为 $\cos(\pi - \alpha) = -\cos(\alpha)$, 从而上式可转化为

$$\begin{aligned} F_{1,0} &= \frac{1}{4\sqrt{2}} \left\{ \cos \frac{\pi}{16} \sum_{y=0}^7 [f(0, y) - f(7, y)] \right. \\ &\quad + \cos \frac{3\pi}{16} \sum_{y=0}^7 [f(1, y) - f(6, y)] + \cos \frac{5\pi}{16} \sum_{y=0}^7 [f(2, y) - f(5, y)] \\ &\quad \left. + \cos \frac{7\pi}{16} \sum_{y=0}^7 [f(3, y) - f(4, y)] \right\} \end{aligned} \quad (5)$$

则 $F_{1,0}$ 能够描述图像在垂直方向上的一种均匀变化情况, 对于垂直方向上的平均偏导数 \bar{f}_x , 有如下的估计式:

$$\bar{f}_x = -F_{1,0} \sqrt{2} \left[7 \cos \frac{\pi}{16} + 5 \cos \frac{3\pi}{16} + 3 \cos \frac{5\pi}{16} + \cos \frac{7\pi}{16} \right] \quad (6)$$

类似地, 对图像在水平方向上的平均偏导数, 可按如下式计算:

$$\bar{f}_y = -F_{0,1} \sqrt{2} \left[7 \cos \frac{\pi}{16} + 5 \cos \frac{3\pi}{16} + 3 \cos \frac{5\pi}{16} + \cos \frac{7\pi}{16} \right] \quad (7)$$

对时间轴偏导数 f_t 进行估计时, 需要利用块的运动估计信息。图 2 表示的是当前块 (B_c) 和参考块 (B_r) 之间的关系, 其中 ΔE 表示二者之间的预测残差, 而块 B_o 与 B_c 在空间位置上是相同的, 它们之间的差值矩阵即可作为 f_t 的估计值, B_o 与 B_r 之间的差值可以利用空间偏导数以及当前块的运动矢量 (M) 进行计算, 因此可以按照下式估计 f_t :

$$f_t = \Delta E - (\bar{f}'_x \cdot M_x + \bar{f}'_y \cdot M_y) \quad (8)$$

其中 (M_x, M_y) 分别表示运动矢量 M 在 x 和 y 方向上的分量, \bar{f}'_x 和 \bar{f}'_y 表示参考块中在两个方向上的空间平均偏导数。

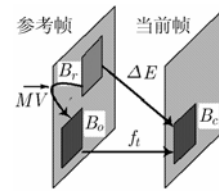


图 2 当前帧和参考帧中子块的位置及差值示意图

2.2 不同类型宏块的处理

首先需要说明的是, I-帧在编码时没有运动估计信息, 因此对其进行光流估计时需要首先解码, 然后利用像素域中的方法进行估计。而 P-帧采用了前向预测编码技术, B-帧还另外采用了后向及双向预测方式。针对它们的前向预测型和跳过型宏块, 其亮度偏导数可以直接按照前面所述方法进行估计。而对帧内编码型和具有后向运动预测的宏块, 其相对前向参考帧的时域偏导数无法直接得到, 下面我们说明对这些宏块相应的处理方法。

(a) P-帧和 B-帧中帧内编码型宏块 帧内编码型宏块完全采用空域编码的方式, 其 DCT 系数可从码流中直接计算得到, 因此对 \bar{f}_x 和 \bar{f}_y 的估计仍可按式(6)和式(7)进行。由

于不存在运动矢量,无法利用式(8)进行 f_t 的估计,而且可能出现了遮挡、切换等变化,使得图像的时域连续性遭到破坏,即 f_t 可能并不存在,此时引入一种修正估计方案,将光流方程修正为如下形式:

$$f_x v_x + f_y v_y + \tilde{f}_t = 0 \quad (9)$$

其中 $\tilde{f}_t = \lambda f_t + \varepsilon$, ε 是一个误差估计量。

根据 DCT 变换的定义,首先利用当前子块与参考帧中相同位置子块 DC 系数的差值估计 \tilde{f}_t :

$$\tilde{f}_t = \frac{1}{8}(D_B - D_{B'}) \quad (10)$$

其中 D_B 和 $D_{B'}$ 分别表示当前子块 B 以及参考帧中与 B 位置相同块 B' 的 DC 系数。

另外,利用当前宏块周围存在前向运动预测的宏块的时域偏导数对 ε 进行估计,假设当前宏块周围所有存在前向运动估计宏块的时域偏导数为 $\{f_t(k), k=1,2,\dots,L\}$, 定义 ε 为

$$\varepsilon = \frac{1}{2L} \sum_{k=1}^L f_t(k) \quad (11)$$

取 $\lambda = 1/2$, 即 $\tilde{f}_t = \bar{f}_t/2 + \varepsilon$, 从而可以利用 \tilde{f}_t 作为时域偏导数进行下一步估计。

(b) B-帧中含有后向运动矢量的宏块 B-帧中的双向及后向预测宏块均含有后向运动矢量,而光流场估计是对两帧图像间的运动进行估计,因此需要将后向运动矢量以及预测残差转化为当前 B-帧相对前向参考帧之间的运动矢量和残差值。假设如图 1 所示,当前子块 B_c 的后向运动矢量及残差为 \mathbf{M}'_B 和 $\Delta E'_B$, 后向参考块为 B_i 并且位于 4 个原始块 $B_i (i=1,2,3,4)$ 的中间,另外假设后向参考帧是 P-帧,而且 $B_i (i=1,2,3,4)$ 的运动矢量以及预测残差分别为 \mathbf{M}_i 和 $\Delta E_i (i=1,2,3,4)$ 。则如果 B_c 是双向预测模式,其前向运动矢量及残差为 \mathbf{M}''_B 和 $\Delta E''_B$, 那么我们按如下方法估计 B_c 相对于前向参考帧的运动矢量和残差,其中 $\omega_i, h_i (i=1,2,3,4)$ 如图 1 中所示。

$$\left. \begin{aligned} \mathbf{M}_B &= \left(\mathbf{M}'_B + \sum_{i=1}^4 \frac{\omega_i h_i}{64} \mathbf{M}_i + \mathbf{M}''_B \right) / 2 \\ \Delta E_B &= \left(\Delta E'_B + \sum_{i=1}^4 \frac{\omega_i h_i}{64} \Delta E_i + \Delta E''_B \right) / 2 \end{aligned} \right\} \quad (12)$$

如果 B_c 仅仅是后向预测模式,无前向运动估计的话,则如下估计 B_c 相对于前向参考帧的运动矢量和残差。

$$\left. \begin{aligned} \mathbf{M}_B &= \sum_{i=1}^4 \frac{\omega_i h_i}{64} \mathbf{M}_i + \mathbf{M}''_B \\ \Delta E_B &= \sum_{i=1}^4 \frac{\omega_i h_i}{64} \Delta E_i + \Delta E''_B \end{aligned} \right\} \quad (13)$$

另外,如果后向参考帧是 I-帧,那么此时无法直接从码流中估计出所需的前向运动矢量和残差,可首先计算出 DC 和 AC 系数,再进行图像亮度的时空域偏导数估计。

2.3 光流场估计算法

由于 MPEG 编码中采用的是块运动估计方法,为了解

决“孔径问题”,我们可以自然地引入一个假设:光流函数在子块中是一致的。由于视频编码的 DCT 变换单元是子块,而且一般的视频图像内容具有很强的空间相关性,相邻像素点的光流运动也具有较强的相关性,因此我们选择以子块而作为光流估计单元就可以较为准确地反映出视频序列的运动特征。

假设在当前块 B 中光流函数为 (v_x, v_y) , 估计出的空间平均偏导数为 (\bar{f}_x, \bar{f}_y) , 另外参考块中的偏导数为 (\bar{f}'_x, \bar{f}'_y) , 则结合式(8)可以将光流方程转化为

$$\bar{f}_x v_x + \bar{f}_y v_y + \Delta E - (\bar{f}'_x M_x + \bar{f}'_y M_y) = 0 \quad (14)$$

为了估计 (v_x, v_y) , 利用最小二乘法,令 $Q = \sum_{x,y \in B} [\bar{f}_x v_x + \bar{f}_y v_y + \Delta E - (\bar{f}'_x M_x + \bar{f}'_y M_y)]^2$, 将 Q 对 v_x 和 v_y 求导并令其分别为 0, 即可得到 $\mathbf{A} \cdot \mathbf{v} = \mathbf{b}$, 其中

$$\left. \begin{aligned} \mathbf{A} &= \begin{bmatrix} \sum_{x,y \in B} \bar{f}_x^2 & \sum_{x,y \in B} \bar{f}_x \bar{f}_y \\ \sum_{x,y \in B} \bar{f}_x \bar{f}_y & \sum_{x,y \in B} \bar{f}_y^2 \end{bmatrix}, & \mathbf{v} &= \begin{bmatrix} v_x \\ v_y \end{bmatrix} \\ \mathbf{b} &= \begin{bmatrix} -\sum_{x,y \in B} \bar{f}_x \cdot (\Delta E - \bar{f}'_x M_x + \bar{f}'_y M_y) \\ -\sum_{x,y \in B} \bar{f}_y \cdot (\Delta E - \bar{f}'_x M_x + \bar{f}'_y M_y) \end{bmatrix} \end{aligned} \right\} \quad (15)$$

令 \mathbf{A}^+ 是矩阵 \mathbf{A} 的最小二乘逆,则光流函数 \mathbf{v} 可由下式估计得到

$$\mathbf{v} = \mathbf{A}^+ \cdot \mathbf{b} \quad (16)$$

3 实验结果

为了验证前面提出的压缩域中光流场快速估计算法,本文首先对一些典型的标准测试视频流(经过 MPEG2 编码)进行了实验,图 3 中给出了 Tabletennis 序列其中两帧的估计结果,从估计出的光流场图中可以看出,根据本文提出的压缩域光流场估计算法可以准确地描绘出视频的运动场变化。

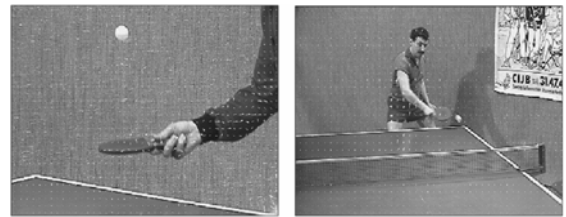


图 3 Tabletennis 序列其中两帧图像的光流场估计结果

另外,为了更充分地验证光流场估计结果的准确性,本文利用两帧图像之间经过光流场补偿后的峰值信噪比 (PSNR) 作为判断标准,假设 $f(x,y)$ 表示当前帧图像, $f'(x,y)$ 表示参考帧图像,并且 $d_1(x,y)$ 和 $d_2(x,y)$ 分别表示估计出的光流分量,则 PSNR 的定义如下:

$$\text{PSNR} = 10 \cdot \lg \frac{255^2 \cdot M \cdot N}{\sum_{x=1}^M \sum_{y=1}^N [f'(x,y) - f(x+d_1(x,y), y+d_2(x,y))]^2} \quad (17)$$

其中 M, N 表示视频帧的高度和宽度。

表1 几种不同方法估计出的 PSNR 均值(dB)

估计方法	Football	Coastguard	Tabletennis	News1	Animals	Donqui
本文方法	30.59	29.82	31.36	29.07	28.85	29.97
L-K 法	30.67	30.19	31.24	28.96	29.33	30.56
文献[7]方法	30.24	29.90	30.67	25.43	27.14	28.05

表2 几种不同方法进行一帧视频的光流场估计平均时间(ms)

估计方法	Football	Coastguard	Tabletennis	News1	Animals	Donqui
本文方法	54.8	34.1	39.6	60.4	59.5	61.6
L-K 法	858.9	305.2	784.3	1016.4	1167.5	984.6
文献[7]方法	51.5	33.7	37.5	56.8	59.8	57.3

实验中,我们对一些测试视频流的其中 100 帧进行了光流场估计并计算出相应的 PSNR 值,表 1 给出了部分视频流这 100 帧 PSNR 值的均值。另外,我们还利用文献[2]中的 L-K 法和文献[7]中的方法对相同的码流进行了估计,并计算出了相应的 PSNR 均值。在计算时间上,由于基于像素域操作的 L-K 法需要进行复杂的偏导数估计,因此这种方法不如压缩域中利用 DCT 系数进行偏导数估计快捷,L-K 法的估计时间也是最长的。相对来说,虽然本文提出的方法需要进行矩阵求逆等运算,但比起像素域中的偏导数估计还是可以节省大量计算时间。表 2 给出了在相同的操作环境下,这几种方法对视频流中一帧图像进行光流场估计所需的平均时间。从统计结果可以看出,本文提出的估计方法在计算时间上较 L-K 法有非常大的提高;与文献[7]中的估计方法相比,由于本文对偏导数的估计更加准确,同时引入了修正的光流方程,在准确性上可以提高将近 2dB。

4 结束语

光流法是分析视频运动信息的一种重要方法,以往的光流场估计多数是在像素域进行,但目前绝大多数的视频是以压缩码流的形式存储传输的,为了对压缩域视频进行快速分析,本文提出了一种针对压缩视频码流的快速光流场估计算法。首先利用图像的两个 AC 系数进行亮度的空间偏导数估计,在此基础上再通过 MPEG 编码过程中的预测残差和运动矢量估计出时间偏导数,然后对未进行前向预测编码的宏块给出了详细的分析以及相应的偏导数估计方法,并对域上不连续的图像块估计出的偏导数进行修正。最后利用块运动假设解决“孔径问题”,并结合估计出的时空域偏导数以及最小二乘法进行光流场估计。实验结果表明,本文方法可以利用较少的压缩域数据准确地估计出光流场信息,并且计算复杂度较小,同时可以较为有效地解决图像出现遮挡、切换时的光流估计问题,因此能够满足视频运动特性的快速鲁棒的分析需求,在此基础上,可以很方便地进行基于视频运动特性的实时对象检测、跟踪监视以及其他方面的应用。

参考文献

- [1] Horn B and Schunck G. Determining optical flow[J]. *Artificial Intelligence*, 1981, 17(1-3): 185-203.
- [2] Lucas B, and Kanade T. An iterative image registration technique with an application to stereo vision[A]. *Proceeding the 7th International Joint Conference on Artificial Intelligence*[C], Vancouver, 1981: 674-679.
- [3] Nagel H H. Constrains for the estimation of displacement vector fields from image sequences [A]. *Proceeding International Joint Conference on Artificial Intelligence*[C], Karlsruhe, Germany, 1983: 945-951.
- [4] Seferidis V. General approach to block-matching motion estimation[J]. *Optical Engineering*, 1993, 32(1): 79-91.
- [5] Lai J, Gauch J, and Crisman J. Computing optical flow in color image sequences[J]. *Innovation and Technology in Biology and Medicine*, 1994, 15(3): 76-87.
- [6] Snyder M A. On the mathematical foundations of smoothness constraints for the determination of optical flow and for surface reconstruction[J]. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1991, 13(11): 1105-1114.
- [7] Coimbra M T and Davies M. Approximating optical flow within the MPEG-2 compressed domain[J]. *IEEE Trans. on Circuits and Systems for Video Technology*, 2005, 15(1): 103-107.
- [8] Yeo B and Liu B. On the extraction of DC sequence from MPEG compressed video[A]. *Proceeding ICIP*[C], Washington, DC, USA, IEEE, 1995, 2: 260-263.

尤隽永: 男, 1977年生, 博士生, 目前研究方向为视频运动对象提取及分析、场景分析、视频注意力区域分析、视频摘要等。

刘贵忠: 男, 1962年生, 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。