

采用 2 b 深度像素的弹性运动估计算法

宋传鸣^{1,2,4} 闵新^{1,3} 谢维冬^{1,3} 尹宝才² 王相海¹

¹(辽宁师范大学计算机与信息技术学院 辽宁大连 116029)

²(大连理工大学计算机科学与技术学院 辽宁大连 116024)

³(东北大学计算机科学与工程学院 沈阳 110169)

⁴(计算机软件新技术国家重点实验室(南京大学) 南京 210023)

(chmsong@lnnu.edu.cn)

Elastic Motion Estimation Algorithm Using Two-Bit-Depth Pixels

Song Chuanming^{1,2,4}, Min Xin^{1,3}, Xie Weidong^{1,3}, Yin Baocai², and Wang Xianghai¹

¹(School of Computer and Information Technology, Liaoning Normal University, Dalian, Liaoning 116029)

²(School of Computer Science and Technology, Dalian University of Technology, Dalian, Liaoning 116024)

³(School of Computer Science and Engineering, Northeastern University, Shenyang 110169)

⁴(State Key Laboratory for Novel Software Technology (Nanjing University), Nanjing 210023)

Abstract To reduce the computational complexity of traditional elastic motion estimation, this paper proposes a novel elastic motion estimation algorithm using two-bit-depth pixels. First, the Prewitt operator is employed to calculate the gradient of each video frame. The mean and standard deviation of the gradient norm is utilized to down-sample each pixel's depth from 8 b into 2 b. Second, the bitwise operation-based matrix multiplication and the comparison based partial derivative computation are introduced. We subsequently describe an elastic motion model using two-bit-depth pixels, as well as a simplified Gaussian-Newton method which avoids the repetitive computation of the Hessian matrix and its inverse matrix. Meanwhile, we establish the functional relationship of the damping step size versus motion vector increment and motion-compensated errors by the first-order linear approximation, obtaining a method for approximately calculating the initial step size. Furthermore, we address a fast method for solving the elastic motion model with two-bit-depth pixels, using the diamond search algorithm as initial search. Experimental results illustrate that our algorithm obviously outperforms the full search with eight-bit-depth pixels, the full search with two-bit-depth pixels, as well as the conventional elastic motion estimation with eight-bit-depth pixels in terms of the peak signal-to-noise ratio (PSNR) and computational efficiency.

Key words video coding; motion estimation; low bit-depth pixel; elastic motion model; adaptive damping coefficient

摘要 为降低传统弹性运动估计的计算复杂度,提出一种 2 b 深度像素的弹性运动估计方法.首先,利用 Prewitt 算子提取视频帧的梯度,借助梯度模长的均值和标准差将像素深度从 8 b 降采样为 2 b.其次,

收稿日期:2018-10-12;修回日期:2019-04-11

基金项目:国家自然科学基金项目(61402214,41671439,61632006);辽宁省自然科学基金项目(20180550570);辽宁省高等学校创新人才支持计划项目(64);南京大学计算机软件新技术国家重点实验室开放课题项目(KFKT2018B07)

This work was supported by the National Natural Science Foundation of China (61402214, 41671439, 61632006), the Natural Science Foundation of Liaoning Province of China (20180550570), the Program for Liaoning Excellent Talents in University (64), and the Open Foundation of State Key Laboratory for Novel Software Technology (Nanjing University) (KFKT2018B07).

引进基于位操作的矩阵乘法和基于比较操作的偏导运算,提出了 2 b 深度像素的弹性运动模型以及求解该模型的简化高斯-牛顿法,避免了黑塞矩阵及其逆矩阵的重复计算.同时,采用 1 阶线性逼近,得出阻尼步长与运动向量增量、运动补偿误差之间的函数关系以及初始步长的近似计算策略,进而以菱形搜索为初始搜索,给出了 2 b 深度像素的弹性运动模型的快速求解算法.实验表明:该算法的平均峰值信噪比和计算效率明显优于 8 b 全搜索、2 b 全搜索和传统 8 b 弹性运动估计.

关键词 视频编码;运动估计;低位深度像素;弹性运动模型;自适应阻尼系数

中图分类号 TP37;TN941.1

运动估计是一项去除视频冗余的时间域预测技术,为 H.264 和高效视频编码(high efficiency video coding, HEVC)等编码标准贡献了大部分的性能提升^[1-2].提高运动估计的效率,使其搜索过程更加健壮和高效不仅是视频编码领域的热点问题,也是进一步改善视频编码效率的重要途径之一.

多年来,视频编码标准始终采用了仅能刻画水平、竖直等平移运动的块匹配算法,可是块平移模型既无法有效预测由物体旋转、缩放、变形和摄像机运动产生的非刚性复合运动,又不能准确表示具有复杂形状的运动区域,导致在运动物体边缘产生大幅值的预测残差.并且,随着块尺寸的减小和运动向量精度的提高,用于编码运动向量、块划分方式的码流开销以及各种软硬件资源开销的增加幅度甚至超过了率失真性能的提升幅度^[2].故此,研究能有效表示复杂运动场的运动模型及相应的帧间运动估计方法,对下一代视频编码效率的提升具有重要意义^[3-6].

鉴于平移运动模型的不足,研究人员将高阶运动模型引入运动估计,利用高阶函数产生 1 个或多个扭曲的参考帧实现了更高质量的运动补偿.依据模型显式表现形式的不同,本文将这些运动估计/补偿算法划分为 4 类:1) 基于网格模型的运动估计^[7-9];2) 基于多项式模型的运动估计^[10-12];3) 基于缩放模型的运动估计^[13-15];4) 基于弹性模型的运动估计^[16-20].其中,算法 1 对编码率失真的优化较为困难^[21];算法 2 的参数偏多,搜索复杂度高,且对局部运动的刻画能力不足;算法 3 无法有效描述物体的旋转、错切和局部变形运动;算法 4 可刻画物体的平移、错切和扭曲运动,既能表示全局和局部运动,又可通过调整模型参数个数来控制运动估计的计算复杂度.因此,对比各类高阶运动模型并综合现有研究结论,弹性运动模型是一种表示复杂运动场的高效模型.

本文首先简要介绍弹性运动估计的相关工作,然后针对其存在的计算量高的问题,引进一种视频

帧的 2 b 深度像素表示方法,进而提出低位深度的弹性运动模型,并从模型的快速计算和自适应步长选取策略 2 个方面展开论述,提出一种采用 2 b 深度像素加速的弹性运动估计算法.本文的主要贡献有 3 个方面:

1) 利用 Prewitt 梯度算子和梯度模长的均值、标准差提取视频帧的纹理信息,设计了一种将像素从 8 b 深度向 2 b 深度进行变换的方法,可避免视频帧在像素深度降采样后出现大面积的光滑区域;

2) 引进基于位操作的矩阵乘法和基于比较操作的偏导运算,提出了一种基于 2 b 深度像素的弹性运动估计模型,并根据 8 b 乘法运算的符号表,进一步给出求解该模型的简化高斯-牛顿法,其优点在于用异或操作替代乘法运算,且无需反复计算黑塞矩阵及其逆矩阵,计算复杂度明显降低;

3) 根据运动补偿误差的分布特性,利用 1 阶线性逼近策略,推导出高斯-牛顿法的自适应步长与弹性运动向量增量、运动补偿误差之间的函数关系,进而得出其近似计算方法,保证了算法具有较稳定的收敛性.

1 相关工作

弹性运动模型采用离散余弦函数、小波函数和样条函数等调和函数来描述物体的非刚性运动,最初主要应用于医学影像配准、物体跟踪等领域.文献^[16-17]将弹性模型引进视频编码中,提出了基于弹性模型的运动估计,在相同码率下获得了比块平移运动估计高出 0.70 dB 的运动补偿峰值信噪比(peak signal-to-noise ratio, PSNR);文献^[18]的结论显示,弹性运动模型可将 H.265 的输出码率降低 3%~12%;文献^[19-20, 22]则通过优化弹性模型的求解方法取得了更加理想的预测效率,将传统弹性运动估计的平均 PSNR 又提高了 1.42~1.77 dB,比基于块平移模型的全搜索算法高出 1.73~2.54 dB.

然而,上述文献均采用高斯-牛顿法来迭代地求解弹性运动模型,在每次迭代过程中需计算偏导数、黑塞矩阵、逆矩阵和矩阵乘法,其计算复杂度甚至高于块平移模型的全搜索(full search, FS).尽管经过文献[19-20,22]改进后的弹性运动估计仅需 2~3 次迭代即可达到块匹配全搜索的补偿质量,可是仍然无法避免黑塞矩阵及其逆矩阵等的反复求解,计算量约为 FS 的 25.62%,高于菱形搜索(diamond search, DS)和测试域搜索(test zone search, TZSearch)等快速块匹配运动估计.

为了增强高阶运动模型求解的实时性,文献[23]提出一种 1 b 深度像素的高斯-牛顿迭代法,并将其应用到基于 6-参数仿射变换的图像配准中.然而,由于 1 b 匹配误差曲面的梯度较小,该算法至少需迭代 50 次才能缓慢趋于收敛,计算复杂度并未显著降低.文献[24-25]进一步将文献[23]的思路推广到 2 b 深度像素,实现了一种基于仿射模型的 2 b 全局运动估计.可是,为节省逆矩阵的计算量,该方法要求保持迭代步长固定不变,无法实现自适应,且只能朝着单一方向进行迭代.而文献[19-20,22]的研究结论却表明,高斯-牛顿法对初始迭代步长和迭代方向均较为敏感,这就不可避免地导致文献[24-25]在大多数情况下都无法求解出全局最优解.同时,文献[23-25]均利用 8 b 深度表示的图像计算 1 阶偏导,再结合经过位截断后的低位深度图像得到匹配误差和最速下降方向.由于低位深像素的梯度往往不同于 8 b 深度像素,两者混合作用很难保证下降方向的准确性;另一方面,文献[26-27]指出,匹配误差往往产生在较低的位,直接截断低位无法有效保留图像的边缘特征,还将使位于最佳匹配块周围的一定区域内的所有像素被量化成相同值,以致失去对最优向量的判断能力.综合来看,无论是运动估计的精度,抑或收敛速度,文献[23-25]的方法尚不够令人满意,且只能求解仿射模型下的运动向量.因此,研究更加有效的、基于低位深度像素的高斯-牛顿法和弹性运动估计,有助于以弹性模型为代表的高阶运动估计走向实用,对新一代视频编码的发展具有一定意义.从本文作者所掌握的文献来看,目前尚鲜见相关研究.

本文提出的 2 b 深度像素的弹性运动估计,利用梯度均值和标准差提取图像的边缘和纹理特征,将像素从 8 b 深度向 2 b 深度进行转换,并实现了基于位操作的快速高斯-牛顿优化求解.不仅解决了初始迭代步长的自适应选取和迭代方向的多样性,也

使得 1 阶偏导与最速下降方向均在 2 b 深度像素域进行统一计算,从而保证基于低位深像素的弹性运动估计精度和收敛速度.

2 弹性运动模型及其高斯-牛顿求解方法

为便于下文工作的论述,本节首先介绍弹性运动模型的定义,再给出求解弹性运动向量的高斯-牛顿方法.

2.1 弹性运动模型的基本定义

视频运动估计的目标是在参考帧的某个搜索窗口内,为当前待预测块 I (尺寸为 $B \times B$ 像素) 搜索到 1 个运动向量,使得 I 与其最佳匹配块 R 之间的误差平方和(sum of squared difference, SSD) 最小,即:

$$D = \arg \min_m \left\{ \sum_{i=1}^B \sum_{j=1}^B [R(\omega(x_{ij}, \mathbf{m}), \omega(y_{ij}, \mathbf{m})) - I(x_{ij}, y_{ij})]^2 \right\}, \quad (1)$$

其中, x_{ij} 和 y_{ij} 分别表示当前待预测块中 i 行 j 列像素的 x 坐标和 y 坐标, \mathbf{m} 表示某种运动模型 ω 下的运动向量.对于弹性运动模型,采用弹性基函数 φ_k 建立待预测块中每个像素的坐标及其匹配像素的坐标映射,其定义为

$$x'_{ij} = \omega(x_{ij}; \mathbf{m}) = x_{ij} + \sum_{k=1}^{p/2} m_k \varphi_k(i, j), \quad (2)$$

$$y'_{ij} = \omega(y_{ij}; \mathbf{m}) = y_{ij} + \sum_{k=p/2+1}^p m_k \varphi_k(i, j), \quad (3)$$

其中, p 表示运动矢量的分量数目, m_k 表示 \mathbf{m} 的第 k 个分量,且 φ_k 为离散余弦函数:

$$\varphi_k(i, j) = \varphi_{k+p/2}(i, j) = \cos \left[\frac{(2i+1)\pi u}{2B} \right] \cos \left[\frac{(2j+1)\pi v}{2B} \right], \quad (4)$$

其中, $p=8, k=qu+v+1, u, v=0, 1, \dots, q-1, q=\sqrt{p/2}, 0 \leq i, j < B$.

2.2 弹性运动模型的高斯-牛顿解法

为了获得弹性运动向量 \mathbf{m} , 文献[16-20]均采用了高斯-牛顿法进行求解,主要思想是对匹配误差函数进行线性逼近,再通过反复迭代求出极小值.具体地,假设当前迭代点为 \mathbf{m} , 并令 (x_{ij}, y_{ij}) 处的预测误差为 $e_{ij}(\mathbf{m}) = R(\omega(x_{ij}, \mathbf{m}), \omega(y_{ij}, \mathbf{m})) - I(x_{ij}, y_{ij})$, 则 $e_{ij}(\mathbf{m})$ 可用其在 \mathbf{m} 处的 1 阶泰勒展开式近似表示,即:

$$e_{ij}(\mathbf{m} + \Delta \mathbf{m}) \approx e_{ij}(\mathbf{m}) + \mathbf{e}'_{ij}(\mathbf{m}) \Delta \mathbf{m}, \quad (5)$$

其中, $e'_{ij}(m)$ 表示 $e_{ij}(m)$ 在 m 处的 1 阶偏导数 $\frac{\partial e_{ij}}{\partial m}$.

将式(5)代入式(1), 则有

$$D = \arg \min_m \left\{ \sum_{i=1}^B \sum_{j=1}^B [R(w(x_{ij}, m), w(y_{ij}, m)) - I(x_{ij}, y_{ij})]^2 \right\} \approx \arg \min_{\Delta m} \sum_{i=1}^B \sum_{j=1}^B [e_{ij}(m) + e'_{ij}(m) \Delta m]^2. \quad (6)$$

为取得式(6)的最小值, 将其对 Δm 取导并令导数为 0, 整理后得到:

$$\sum_{i=1}^B \sum_{j=1}^B [e'_{ij}(m)]^T e'_{ij}(m) \Delta m = - \sum_{i=1}^B \sum_{j=1}^B [e'_{ij}(m)]^T e_{ij}(m), \quad (7)$$

$$\text{令 } \mathbf{b} = - \sum_{i=1}^B \sum_{j=1}^B [e'_{ij}(m)]^T e_{ij}(m), \quad (8)$$

$$\mathbf{H} = \sum_{i=1}^B \sum_{j=1}^B [e'_{ij}(m)]^T e'_{ij}(m), \quad (9)$$

显然, \mathbf{H} 是一个 $p \times p$ 阶的黑塞矩阵, 并且根据求导的链式法则和 $e_{ij}(m)$ 、式(2)、式(3)的定义, 有:

$$\mathbf{H} = \sum_{i=1}^B \sum_{j=1}^B \left[\frac{\partial \mathbf{R}}{\partial x'_{ij}} \frac{\partial w}{\partial m} + \frac{\partial \mathbf{R}}{\partial y'_{ij}} \frac{\partial w}{\partial m} \right]^T \times \left[\frac{\partial \mathbf{R}}{\partial x'_{ij}} \frac{\partial w}{\partial m} + \frac{\partial \mathbf{R}}{\partial y'_{ij}} \frac{\partial w}{\partial m} \right], \quad (10)$$

其中, $\frac{\partial \mathbf{R}}{\partial x'_{ij}}$ 和 $\frac{\partial \mathbf{R}}{\partial y'_{ij}}$ 分别表示匹配块沿着水平和垂直

方向的梯度分量, $\frac{\partial w}{\partial m}$ 是雅克比矩阵, 并且 $\frac{\partial w}{\partial m_k} =$

$\frac{\partial w}{\partial m_{p/2+k}} = \varphi_k(i, j), k \in [1, p/2]$. 于是, 有 $\Delta m = \mathbf{H}^{-1} \mathbf{b}$,

进而得到更新后的弹性运动向量 $m: m \leftarrow m + \Delta m$.

图 1 给出了基于高斯-牛顿法的弹性运动估计过程, 其中 T 表示迭代次数, T_{\min} 表示迭代的条件阈值, $\|\cdot\|$ 表示向量的模长.

3 基于 Prewitt 梯度算子的 2b 变换

由 2.2 节可见, 弹性运动模型的高斯-牛顿解法需反复计算偏导数、黑塞矩阵、逆矩阵和矩阵乘法, 计算量明显高于块平移模型的全搜索. 为此, 我们引进基于低位深度像素的策略, 将像素从 8b 深度降采样为 2b 深度, 以达到降低计算开销的目的.

3.1 本文的 2b 变换算法

将像素从 8b 深度转换至 2b 深度, 势必会丢失视频帧的部分信息, 尤其是细腻的纹理细节, 产生大面积的平滑区域, 这是导致基于低位截断的低位深

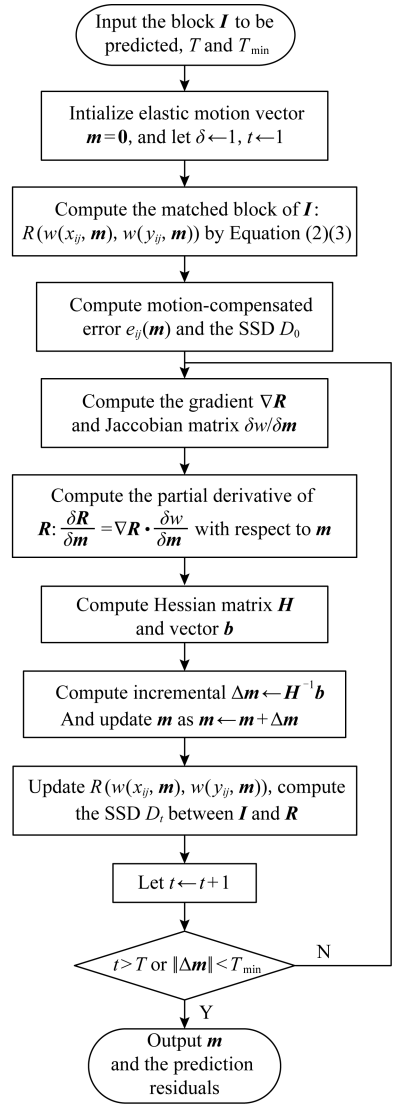


Fig. 1 Flowchart of elastic motion estimation using the Gauss-Newton Method

图 1 基于高斯-牛顿法的弹性运动估计流程图

度像素运动估计出现性能下降的重要原因^[26-27]. 故此, 在像素深度变换过程中尽可能保持视频的纹理特征, 对于衡量像素深度降采样的有效性、保证运动估计精度具有重要意义. 为了达到这一目的, 我们首先通过 Prewitt 算子计算待预测帧的梯度; 其次, 假定像素梯度的模长服从正态分布, 若用均值 μ_c 和标准差 σ_c 将概率密度划分为 4 个区间, 则知梯度模长位于区间 $(-\infty, \mu_c - 0.68\sigma_c)$, $[\mu_c - 0.68\sigma_c, \mu_c)$, $[\mu_c, \mu_c + 0.68\sigma_c)$, $[\mu_c + 0.68\sigma_c, +\infty)$ 的概率均近似等于 0.25, 根据这一规律, 本文以像素块为单位, 将每个像素处的边缘和纹理强度平均分为弱、较弱、较强、强 4 个等级, 从而实现将像素 (x, y) 的深度从 8b 变换为 2b:

$$C'(x,y) = \begin{cases} 11, & \|\nabla G\|_{C(x,y)} \geq \mu_c + 0.68\sigma_c, \\ 10, & \mu_c + 0.68\sigma_c > \|\nabla G\|_{C(x,y)} \geq \mu_c, \\ 00, & \mu_c - 0.68\sigma_c \leq \|\nabla G\|_{C(x,y)} < \mu_c, \\ 01, & \|\nabla G\|_{C(x,y)} < \mu_c - 0.68\sigma_c, \end{cases} \quad (11)$$

其中, $\|\nabla G\|_{C(x,y)}$ 表示块 C 中 (x,y) 处的梯度模长, $C'(x,y)$ 表示像素 (x,y) 经过变换后的 2 b 深度表示.

本文提出的基于梯度均值和标准差的 2 b 变换过程如图 2 所示:

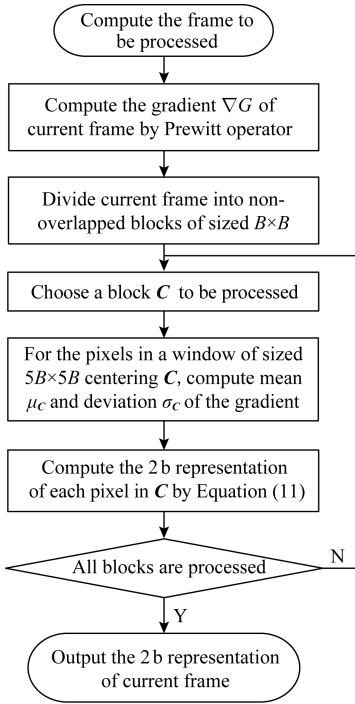


Fig. 2 Flowchart of the proposed 2 b transform

图 2 本文提出的 2 b 变换流程图

图 3 给出了分别采用低位截断方法^[28]、基于亮度均值和方差的 2 b 变换方法^[29], 以及本文方法处理 Football 序列和 Foreman 序列第 2 帧的结果. 从图 3 不难发现, 基于低位截断的变换方法损失了 Football 序列包含的复杂背景纹理(图 3(b)); 虽然基于亮度均值和方差的 2 b 变换方法^[29]在一定程度上避免了位截断方法的不足, 可是未充分考虑正态分布的特点, 选用亮度均值 μ 、标准差 σ 及其线性组合 $\mu - \sigma$ 和 $\mu + \sigma$ 作为变换阈值, 这将导致约有 68% 的像素会从 8 b 深度被降采样为 00 和 01, 产生较大面积的平滑区, 如 Football 的运动员身体和 Foreman 的脸部区域中大量像素被转换为相同值, 易使运动估计陷入局部最优, 甚至失去对最优向量的判断能力; 而本文方法无论在复杂的背景纹理区,

还是在平滑的渐变区(如 Foreman 的脸部), 均能保持适当的图像细节, 进而形成能有效区分错误运动向量的特征.

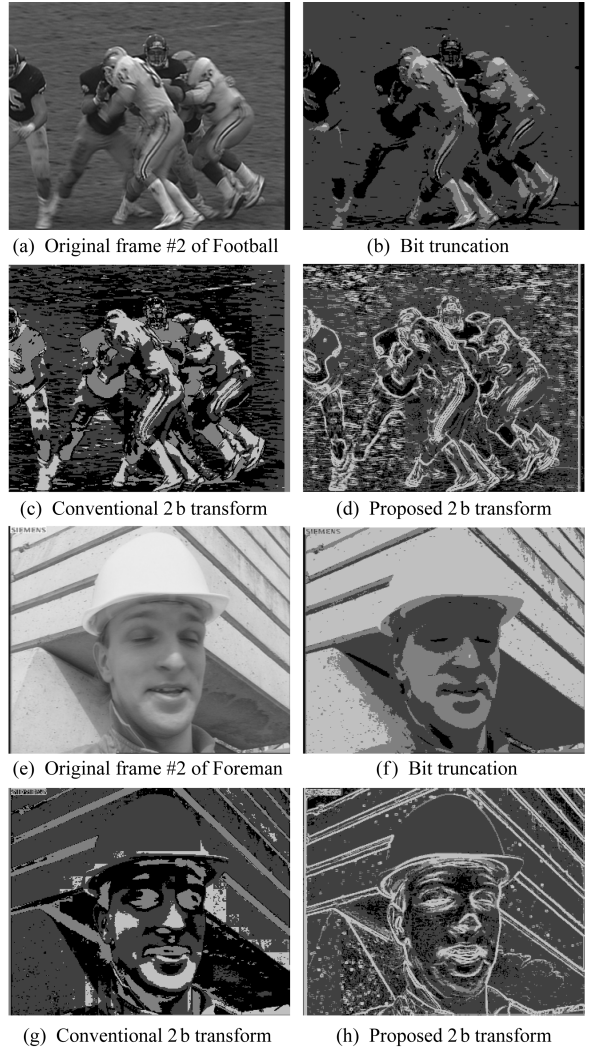


Fig. 3 Comparison among different 2 b transform results

图 3 不同的 2 b 变换方法的效果对比

3.2 2 b 深度像素的匹配误差曲面的特点

由于位深度变换过程难免丢失少量纹理细节, 会出现 2 种情况: 1) 当一定区域内的像素均被转换为相同值时, 像素值的变化率降低、自相关系数增大. 这样, 当待预测块在该区域内搜寻最佳匹配块时, 匹配误差的变化趋于平缓. 我们称这样的区域为“曲面平滑区域”. 2) 在 2 个曲面平滑区的交界线两侧, 像素值被转换为不同值, 像素值的变化率提高, 而自相关系数减小. 对 2 b 深度的像素而言, 由于匹配误差的值域显著小于 8 b 深度像素, 即使匹配误差增加 1, 也意味着 25% 的增幅, 这相当于 8 b 误差增幅的 32 倍. 因此, 在交界区域中不同候选向量的

运动补偿误差易产生较大变化.2 b 匹配误差曲面的这种“区域内平滑、区域间锐化”的分布特点,会使其呈现更加明显的单调性和梯度变化.

图 4 给出了 Stefan 序列第 2 帧内左上角坐标为(144,112)的块和 Football 序列第 2 帧内左上角坐标为(176,176)的块,分别进行基于 8 b 深度像素和基于本文 2 b 深度像素的整数精度全搜索后得到的匹配误差曲面,搜索窗口为 33×33 .对比可见:首先,8 b 深度像素的匹配误差曲面(图 4(a)(c))较之 2 b 深度像素的匹配误差曲面(图 4(b)(d))具有更强的波动性,局部极值点也更多,易导致典型的快速

搜索陷入局部最优.其次,在曲面平滑区域,8 b 深度像素的匹配误差的变化幅度和频率均大于 2 b 深度像素的匹配误差;而在极值点的一定邻域以及 2 个局部极值区域的交界,8 b 深度像素匹配误差的变化趋势却较平滑,其数值梯度和衰减速度明显小于 2 b 深度像素的匹配误差曲面.这不仅印证了上一段的理论分析,也充分说明 2 b 深度像素的运动估计陷入局部最优的概率更小,且能通过梯度下降策略以更快的速度向全局最优点收敛,尤其是当搜索初始点较为准确、位于全局极值点的单调区间时,从而有利于提高运动补偿精度和运动估计效率.

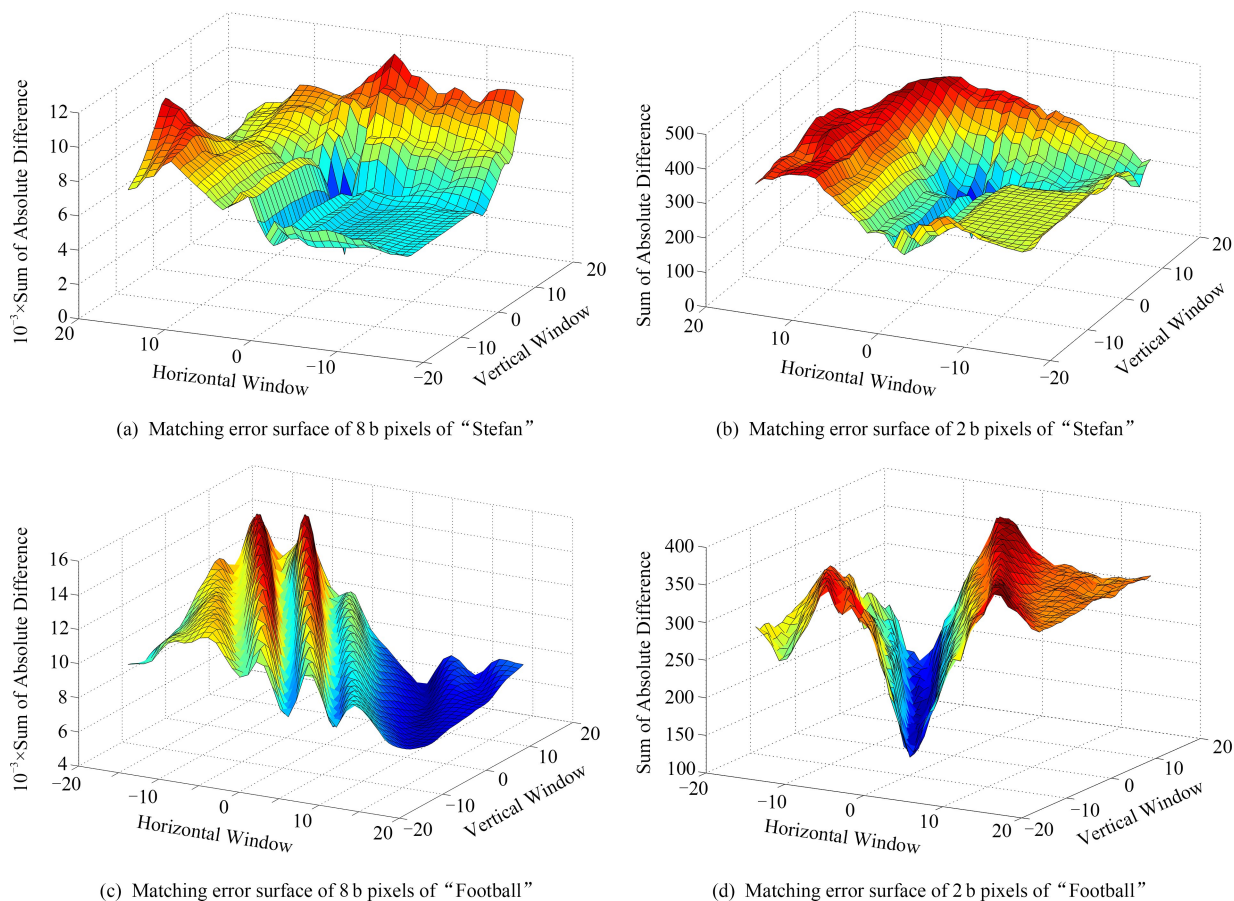


Fig. 4 Comparison between matching error surfaces of 8 b pixel and 2 b pixel under full search

图 4 8 b 全搜索与 2 b 全搜索的匹配误差曲面比较

4 低位深度的弹性运动模型

传统高斯-牛顿法的每次迭代都需要计算待预测块的偏导数、黑塞矩阵、逆矩阵、矩阵乘法和运动补偿误差等,其计算复杂度达到了 $\max\{O(p^3), O(p^2 B^2)\}$.若能降低反复执行上述操作的运算量,

则可进一步改善弹性运动估计/补偿的效率.同时,经过第 3 节的处理,视频帧的全部像素已由 8 b 深度降采样到 2 b 深度,若继续沿用传统高斯-牛顿法中基于 8 b 整数的偏导、矩阵乘法等运算,则达不到通过精度下采样来提高计算效率的目的.故此,本文对匹配误差 $e_{ij}(m)$ 及其 1 阶偏导数 $e'_{ij}(m)$ 也进行了位深度变换,利用异或运算取代乘法运算,提出了

面向低位深度像素的弹性运动模型,从而加快弹性运动估计的速度。

根据 1 阶偏导 $e'_{ij}(\mathbf{m})$ 的符号,文献[23]将其直接 2 值化为

$$\bar{e}'_{ij}(m_k) = \begin{cases} 1, & e'_{ij}(m_k) > 0, \\ 0, & e'_{ij}(m_k) \leq 0, \end{cases} \quad (12)$$

其中,用 bit “1”表示正数,用 bit “0”表示非正数, $e'_{ij}(m_k)$ 表示 $e'_{ij}(\mathbf{m})$ 的第 k 个分量。同理,采用同样的规则将匹配误差 $e_{ij}(\mathbf{m})$ 进行位深度的变换,即:

$$\bar{e}_{ij}(\mathbf{m}) = \begin{cases} 1, & e_{ij}(\mathbf{m}) > 0, \\ 0, & e_{ij}(\mathbf{m}) \leq 0. \end{cases} \quad (13)$$

此时,将经过 2 值化后的 $e_{ij}(\mathbf{m})$ 和 $e'_{ij}(\mathbf{m})$ 代入式(8),并将式(8)中的乘法运算替换为异或运算来计算向量 \mathbf{b} 。为了验证上述规则的有效性,表 1 给出了基于 8 b 运算的 $[e'_{ij}(\mathbf{m})]^T \times e_{ij}(\mathbf{m})$ 的乘法符号表,而表 2 所示则为低位深度的 $e'_{ij}(\mathbf{m})$ 和 $\bar{e}_{ij}(\mathbf{m})$ 经过 1 b 异或运算“ \oplus ”的真值表。

Table 1 Sign of 8 b Multiplication Operation

表 1 8 b 乘法运算的符号表

$e_{ij}(\mathbf{m})$	$e'_{ij}(m_k)$	$e'_{ij}(m_k) \times e_{ij}(\mathbf{m})$
-	-	+
-	+	-
+	-	-
+	+	+

Table 2 Truth Table of 1 Bit-Wise XOR Operation

表 2 1 b 异或运算的真值表

$\bar{e}_{ij}(\mathbf{m})$	$\bar{e}'_{ij}(m_k)$	$\bar{e}'_{ij}(m_k) \oplus \bar{e}_{ij}(\mathbf{m})$
0	0	0
0	1	1
1	0	1
1	1	0

对比表 1 和表 2 可以发现,8 b 乘法运算的符号与 1 b 异或运算的真值并不一致。例如正数与负数

的乘积为负数,也就是“1”与“0”的乘积应为“0”,而异或的结果却等于“1”。另一方面,高斯-牛顿法通过 $[e'_{ij}(\mathbf{m})]^T \times e_{ij}(\mathbf{m})$ 控制每次迭代的方向,若 $[e'_{ij}(\mathbf{m})]^T \oplus \bar{e}_{ij}(\mathbf{m}) = \mathbf{0}$,则由 $\Delta \mathbf{m} = \mathbf{H}^{-1} \mathbf{b}$ 可知,将导致迭代步长 $\|\Delta \mathbf{m}\| = 0$,失去向全局最优解收敛的驱动力。这表明,采用文献[23]的方法将弹性模型进行 2 值化,仅能产生单一方向的迭代动力(也就是当 $[e'_{ij}(\mathbf{m})]^T \times e_{ij}(\mathbf{m}) = \mathbf{1}$ 时的情形),不利于保证解的精度和收敛速度。

为了解决这个问题,本文通过将匹配误差进行 2 次互为相反数的 2 值化转换,提出了基于 2 b 深度的弹性运动模型,即:

$$\bar{e}_{ij}^+(\mathbf{m}) = \begin{cases} 1, & I'_{ij} < R'_{ij}, \\ 0, & I'_{ij} \geq R'_{ij}, \end{cases} \quad (14)$$

$$\bar{e}_{ij}^-(\mathbf{m}) = \begin{cases} 0, & I'_{ij} < R'_{ij}, \\ 1, & I'_{ij} \geq R'_{ij}, \end{cases} \quad (15)$$

其中, I'_{ij} 和 R'_{ij} 表示当前帧和参考帧的 2 b 深度表示, $\bar{e}_{ij}^+(\mathbf{m})$ 和 $\bar{e}_{ij}^-(\mathbf{m})$ 分别表示在弹性运动参数 \mathbf{m} 下像素 (i, j) 处匹配误差的高位和低位。在此基础上,将高、低位分别与 $e'_{ij}(\mathbf{m})$ 进行异或运算,再将所得结果相减:

$$(\bar{e}_{ij}^+(\mathbf{m}) \oplus e'_{ij}(\mathbf{m})) - (\bar{e}_{ij}^-(\mathbf{m}) \oplus e'_{ij}(\mathbf{m})). \quad (16)$$

表 3 给出了经过上述改进后,1 b 异或运算和 8 b 乘法运算的对比结果。可见, $[e'_{ij}(\mathbf{m})]^T \times e_{ij}(\mathbf{m})$ 与 $(\bar{e}_{ij}^+(\mathbf{m}) \oplus e'_{ij}(\mathbf{m})) - (\bar{e}_{ij}^-(\mathbf{m}) \oplus e'_{ij}(\mathbf{m}))$ 表现出一致的极性,并且 $(\bar{e}_{ij}^+(\mathbf{m}) \oplus e'_{ij}(\mathbf{m})) - (\bar{e}_{ij}^-(\mathbf{m}) \oplus e'_{ij}(\mathbf{m})) = \pm 1$ 表明,上述基于 2 b 深度的弹性运动模型不仅避免了 $[e'_{ij}(\mathbf{m})]^T \oplus \bar{e}_{ij}(\mathbf{m}) = \mathbf{0}$ 时可能出现的不收敛现象,也使得求解过程能够朝着 2 个不同方向进行迭代,从而提高取得全局最优解的概率。鉴于这一原因,本文向量 \mathbf{b} 的计算过程为

$$\mathbf{b} = \sum_{i=1}^B \sum_{j=1}^B \bar{e}_{ij}^+(\mathbf{m}) \oplus e'_{ij}(\mathbf{m}) - \sum_{i=1}^B \sum_{j=1}^B \bar{e}_{ij}^-(\mathbf{m}) \oplus e'_{ij}(\mathbf{m}). \quad (17)$$

Table 3 Truth Table of Improved XOR Operation

表 3 改进的异或运算真值表

$e_{ij}(\mathbf{m})$	$e'_{ij}(m_k)$	$e'_{ij}(m_k) \times e_{ij}(\mathbf{m})$	$\bar{e}_{ij}^+(\mathbf{m})$	$\bar{e}_{ij}^-(\mathbf{m})$	$\bar{e}'_{ij}(m_k)$	$(\bar{e}_{ij}^+(\mathbf{m}) \oplus e'_{ij}(m_k)) - (\bar{e}_{ij}^-(\mathbf{m}) \oplus e'_{ij}(m_k))$
-	-	+	1	0	0	$(1 \oplus 0) - (0 \oplus 0) = +1$
-	+	-	1	0	1	$(1 \oplus 1) - (0 \oplus 1) = -1$
+	-	-	0	1	0	$(0 \oplus 0) - (1 \oplus 0) = -1$
+	+	+	0	1	1	$(0 \oplus 1) - (1 \oplus 1) = +1$

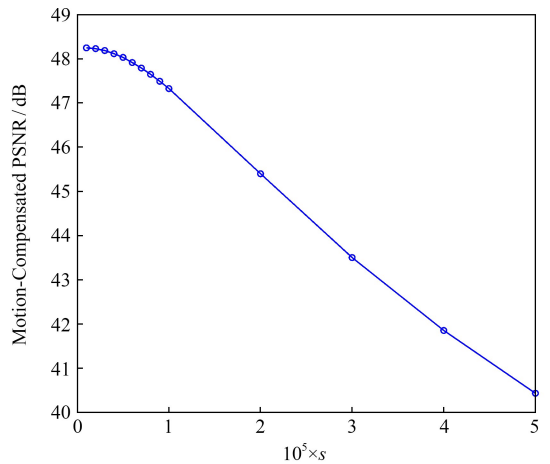
5 迭代步长的自适应计算

在高斯-牛顿方法中, \mathbf{b} 用于确定每次迭代的方向, 黑塞矩阵 \mathbf{H} 的逆 \mathbf{H}^{-1} 则用于确定每次迭代的步长. 然而, 计算 \mathbf{H}^{-1} 需要大量的乘法运算, 其时间复杂度为 $O(p^3)$; 并且, 1 阶偏导 $e'_{ij}(\mathbf{m})$ 经过第 4 节的 2b 转换后, \mathbf{H}^{-1} 所给出的步长亦不够精确, 没有必要再通过 \mathbf{H}^{-1} 计算迭代步长. 所以, 文献[25]采用固定步长 s 来近似 \mathbf{H}^{-1} , 于是弹性运动向量的增量表示为

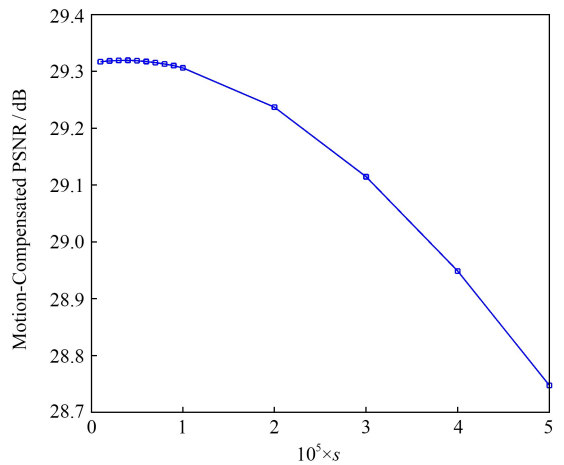
$$\Delta \mathbf{m} = -s \times \text{sgn}(\mathbf{b}), \quad (18)$$

其中, $\text{sgn}(\cdot)$ 表示符号函数. 但是, 这种做法无异于放弃了黑塞矩阵对高斯-牛顿迭代过程的控制. 而文献[19-20, 22]的研究结论却表明, 高斯-牛顿法对初始迭代步长较为敏感, 文献[25]的这种固定初始步长的方式显然不利于迭代过程收敛到全局最优解. 为了验证这一推论, 我们选取 Akiyo, Tempete, Football 这 3 个运动特性和纹理复杂度各异的视频序列的前 10 帧, 应用固定步长的 2b 深度弹性模型进行了运动估计/补偿实验. 图 5 给出了在不同的初始步长 s 下, 3 个序列经过 15 次迭代后所获得的平均运动补偿 PSNR 曲线. 从图 5 可见, Akiyo 序列的运动幅度很小、画面纹理细节简单, 随着初始步长的缩小, 该序列的运动补偿 PSNR 明显升高(图 5(a)), 但是即使初始步长缩小至 1×10^{-6} , 仍未取得最高的 PSNR; Tempete 序列的运动幅度中等, 可是纹理细节复杂, 且包含摄像机的拉摄运动, 其运动补偿的平均 PSNR 在初始步长为 4×10^{-6} 时取得极值, 然后随着初始步长的增大迅速下降(图 5(b)); Football 序列既包含快速的前景运动和中等复杂度的纹理, 又存在物体的局部形变, 它在初始步长等于 3×10^{-5} 时取得了运动补偿 PSNR 的极值(图 5(c)), 而此时, Akiyo 的平均 PSNR 已比其最高值下降了 4.74 dB.

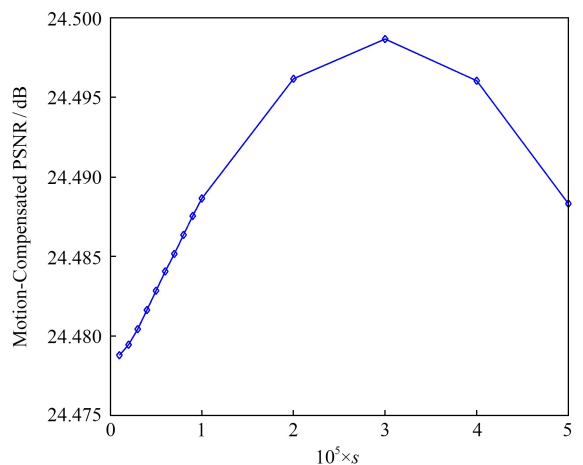
对比这 3 个视频序列的实验结果可知, 不同序列的最优初始步长相差几十倍. 对于运动幅度较小、画面细节较多的视频, 应为其选取一个较小的迭代步长; 反之, 对于运动速度较快、局部形变较多的视频, 为其设定一个较大的迭代步长则更加合适. 可见, 文献[25]固定初始步长的方式不仅无法保证运动补偿的质量, 而且在现实中亦很难做到折中选择. 为此, 本节采用一种自适应步长 s 来近似 \mathbf{H}^{-1} , 下面讨论其计算方法.



(a) Curve of step size s vs PSNR for "Akiyo" sequence



(b) Curve of step size s vs PSNR for "Tempete" sequence



(c) Curve of step size s vs PSNR for "Football" sequence

Fig. 5 The motion-compensated PSNR comparison among with different initial step size s

图 5 不同初始步长 s 下的运动补偿 PSNR 对比

由式(6)可知, 高斯-牛顿法的基本思路是在每次迭代中利用 2 次函数来逼近匹配误差 $D(\mathbf{m})$, 并

求解近似函数的极小点作为下一个迭代点,则第 $k+1$ 次迭代的弹性运动向量 \mathbf{m}^{k+1} 有:

$$\mathbf{m}^{k+1} = \mathbf{m}^k + \mathbf{H}^{-1} \mathbf{b}, \quad (19)$$

其中, $D(\mathbf{m}) = \sum_{i=1}^B \sum_{j=1}^B [e_{ij}(\mathbf{m})]^2$. 那么:

$$\frac{\partial D(\mathbf{m})}{\partial \mathbf{m}} = \sum_{i=1}^B \sum_{j=1}^B 2[e'_{ij}(\mathbf{m})]^T e_{ij}(\mathbf{m}) = -2\mathbf{b}, \quad (20)$$

$$\frac{\partial^2 D(\mathbf{m})}{\partial \mathbf{m}^2} = 2 \sum_{i=1}^B \sum_{j=1}^B \left\{ [e'_{ij}(\mathbf{m})]^T e'_{ij}(\mathbf{m}) + [e''_{ij}(\mathbf{m})]^T e_{ij}(\mathbf{m}) \right\}, \quad (21)$$

由于匹配误差 $e_{ij}(\mathbf{m})$ 在全局最优解附近“几乎”是白噪声^[30], 式(21)第 2 项 $[e''_{ij}(\mathbf{m})]^T e_{ij}(\mathbf{m})$ 的数学期望近似为 $\mathbf{0}$, 就有:

$$\frac{\partial^2 E(\mathbf{m})}{\partial \mathbf{m}^2} \approx 2\mathbf{H}. \quad (22)$$

而由微分的 1 阶近似性质 $\Delta f(x) \approx f'(x) \cdot \Delta x$, 可知:

$$\frac{\partial^2 D(\mathbf{m})}{\partial \mathbf{m}^2} \approx \frac{\frac{\partial D(\mathbf{m}^{k+1})}{\partial \mathbf{m}^{k+1}} - \frac{\partial D(\mathbf{m}^k)}{\partial \mathbf{m}^k}}{\mathbf{m}^{k+1} - \mathbf{m}^k}, \quad (23)$$

进一步, 将式(20)代入式(23), 得到:

$$\frac{\partial^2 D(\mathbf{m})}{\partial \mathbf{m}^2} \approx \frac{2\mathbf{b}^k - 2\mathbf{b}^{k+1}}{\mathbf{m}^{k+1} - \mathbf{m}^k} = -2\Delta\mathbf{b}/\Delta\mathbf{m}, \quad (24)$$

再由式(22)易知 $\mathbf{H} \approx -\Delta\mathbf{b}/\Delta\mathbf{m}$, 整理后即可得到自适应步长 s :

$$s = \|\mathbf{H}^{-1}\| = -\|\Delta\mathbf{m}\| \times \|\Delta\mathbf{b}\|^{-1}. \quad (25)$$

式(25)表明: 自适应步长 s 与运动向量的变化量 $\Delta\mathbf{m}$ 成正比, 而与匹配误差的变化量及其关于 \mathbf{m} 的导数呈反比. 对比图 5 的实验结果, Akiyo 序列的运动幅度很小, 其 $\Delta\mathbf{m}$ 几乎为 $\mathbf{0}$, 而由于播音员的头部和嘴部存在局部运动, 第 1 次迭代后产生的匹配误差必然大于 0, 这就使得该序列的最优初始步长趋近于 0; 尽管 Tempete 序列存在中等幅度的运动 (即 $\Delta\mathbf{m} > \mathbf{0}$), 可是复杂纹理细节势必产生较大的匹配误差, 故此, 由式(25)可推知 Tempete 序列的最优初始步长也相对较小; 考虑到 Football 序列包含快速运动, 其 $\Delta\mathbf{m}$ 将明显超过 Akiyo 和 Tempete, 且纹理细节的复杂度一般, 匹配误差变化量在预期上不会显著大于 Tempete 序列, 于是 Football 序列的最优初始步长较大. 综合上述分析, 式(25)的结论与图 5 的实验结果一致.

为了进一步降低式(25)所需要的计算量, 本文

采用取绝对值运算代替乘法和开平方运算来近似取模长运算:

$$\|\Delta\mathbf{m}\| \approx \sum_{k=1}^p |\Delta m_k|, \quad (26)$$

$$\|\Delta\mathbf{b}\| \approx \sum_{k=1}^p |\Delta b_k|, \quad (27)$$

其中, $\Delta m_k, \Delta b_k$ 分别表示 $\Delta\mathbf{m}$ 和 $\Delta\mathbf{b}$ 的第 k 个分量. 同时, 本文仅利用式(25)~(27)确定初始迭代步长, 而对于后续的迭代, 为保证算法具有较稳定的收敛性, 本文在步长值满足以下条件时将其减半^[25]: \mathbf{b} 的符号在连续 3 次迭代中改变 3 次, 即在某个解的局部邻域内发生往返式的迭代. 如第 4 节所述, 本文的低位深度的弹性运动模型能朝着 2 个不同方向进行迭代, 若发生折返式迭代, 则表明算法很可能处于某个全局/局部最优解的单调区间内, 且当前步长较大, 导致每次迭代均越过该最优解, 故应选取更小的步长向这个解进行逼近.

6 2 b 深度像素的弹性运动估计步骤

在第 5 节算法的基础上, 本节给出一种基于 2 b 深度像素的弹性运动估计方法, 具体步骤如图 6 所示, 其中, x_i, y_i 表示 \mathbf{I} 中某个像素的横、纵坐标, $0 \leq x_i, y_i \leq B-1$; \mathbf{m}^0 表示只包含平移分量的初始运动向量, 且 $\mathbf{m}^0 = (m_1, 0, \dots, m_{p/2+1}, 0, \dots)$; Th 表示预设的迭代次数阈值; $R'(\cdot)$ 表示参考帧中位于坐标“ \cdot ”处的 2 b 深度的像素值; 8 b 深度像素的菱形搜索和弹性运动估计的搜索窗口尺寸为 $W \times W$ 像素.

7 实验结果与分析

为了验证本文算法的有效性, 以通用中间格式 (common intermediate format, CIF)、4CIF 和高清格式的 34 个标准视频序列 (见表 4) 的 1~90f 为例进行实验, 并将结果与基于 8 b 深度像素的菱形运动估计^[31] (简称 8 b-DS)、基于 8 b 深度像素的块平移全搜索 (简称 8 b-FS)、基于 8 b 深度像素的弹性运动估计^[16] (简称 8 b-Elastic)、基于传统 2 b 深度像素的块平移全搜索算法^[29] (简称 2 b-FS) 以及未引进自适应初始步长的本文算法 (简称 2 b-Elastic-NAStep) 进行比较.

实验参数选取为: 搜索窗口为 33×33 像素, 宏块尺寸为 $B=16$ 像素, 这与文献^[16, 29, 31] 的参数

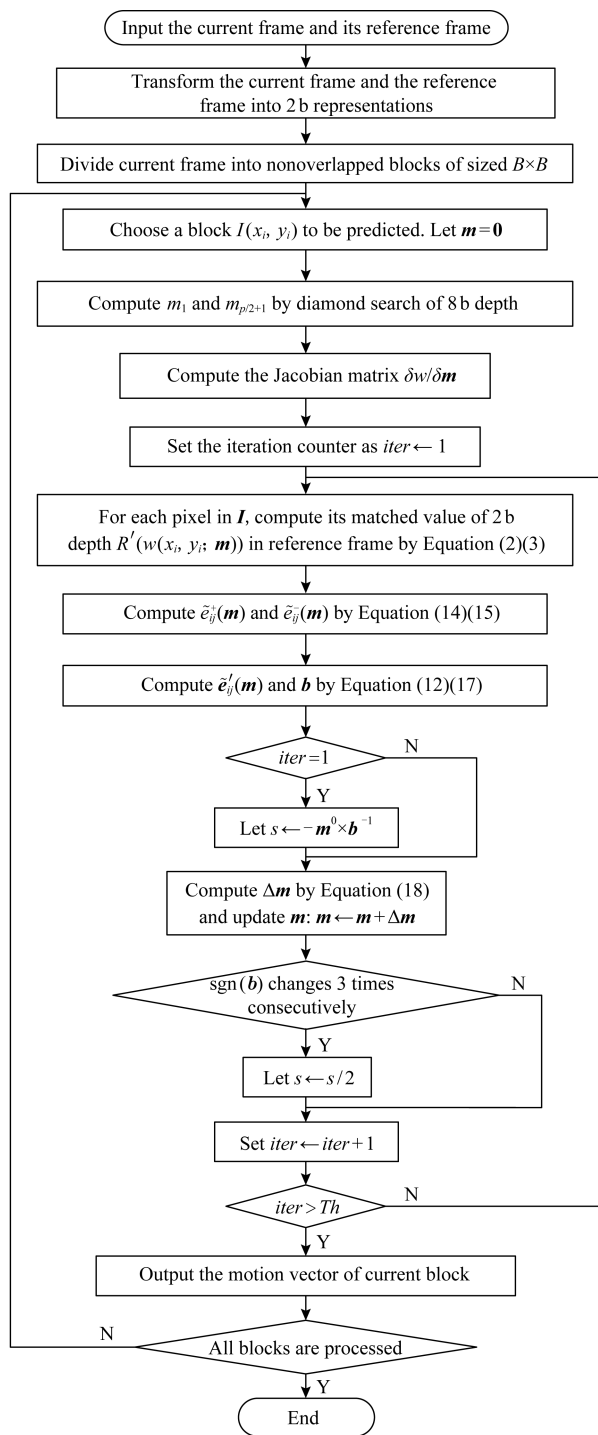


Fig. 6 Flowchart of the proposed 2 b elastic motion estimation

图6 本文提出的2 b弹性运动估计流程图

取值相同,也是视频编码器的典型设置.为了在预测精度和收敛速度之间实现折中,令迭代次数阈值 $Th = 5$. 7.2节还将对此进行深入讨论.同时,为公平起见,将传统8 b-Elastic的迭代次数也设置为5次.运动补偿质量采用PSNR进行评价.

Table 4 Test Video Sequences' Names and Their Formats

表4 测试视频序列名称及其格式

Resolution	Sequence
352×288	Akiyo, Bowling, Coastguard, Football, Foreman, Hall, Highway, Husky, Ice, Intros_cif, Mobile, Mother_daug, Pamphlet, Paris, Sign_irene, Soccer, Stefan, Students, Tempete, Tennis, Vtc1nw, Waterfall
704×576	Crew, City, Harbour
832×480	BQmall, Flowervase
1280×720	sc_Robot, FourPeople, Johnny, KristenAndSara
960×540	BQTerrace, Cactus, ParkScene

7.1 运动估计/补偿质量的比较

表5罗列了各个测试序列的亮度分量采用不同运动估计算法所得到的平均PSNR.从表5可见,相对于传统的2 b深度像素的全搜索算法2 b-FS^[29],未引进自适应初始步长的本文算法2 b-Elastic-NAStep将PSNR平均提高了1.15 dB,表明基于Prewitt算子的2 b变换和2 b深度像素的弹性运动模型能够进行更有效的运动估计/补偿.2 b-Elastic-NAStep算法的平均PSNR比传统的8 b深度弹性运动估计8 b-Elastic^[16]提高了0.46 dB,其根本原因在于8 b像素的匹配误差曲面富含局部极值点,在缺少初始搜索的情况下易陷入局部最优,并且8 b匹配误差曲面的下降梯度为高斯-牛顿法提供的迭代驱动力不足,收敛速度较慢;而2 b-Elastic-NAStep算法则利用8 b-DS计算弹性运动向量的水平和垂直分量,将起始搜索点置入全局最优点的单调区间,再沿着梯度下降方向进行迭代,这表明起始搜索点对于提高弹性运动估计的精度有重要作用,该结论与文献[19-20, 22]的研究结果一致.同时,2 b-Elastic-NAStep算法的平均PSNR较之8 b深度像素的菱形运动估计8 b-DS高出0.90 dB,并且比8 b深度像素的块平移全搜索8 b-FS提高0.23 dB,说明本文提出的低位深度弹性运动模型在初始搜索的基础上取得了进一步的性能增益.在引进了自适应初始步长之后,本文算法的运动补偿PSNR比2 b-Elastic-NAStep又提高了0.92 dB,这表明通过自适应步长实现黑塞矩阵对迭代过程的控制,有助于弹性运动估计收敛到全局最优解,进而表现出与8 b深度像素的阻尼高斯-牛顿法相近的准确程度.

图7给出了Akiyo, Husky, City, Flowervase序列的PSNR逐帧比较情况.其中, Akiyo序列仅含有局部慢速运动,且细节简单; Husky序列则与之相

反,前景具有较大幅度的平移运动且纹理复杂性高; City 序列含有摄像机的中速摇摄运动和丰富的细节信息;而 Flowervase 序列则包含摄像机的慢速拉摄运动和中等复杂性的纹理.如图 7 所示,尽管 4 个

序列的运动类型、速度和纹理复杂度各有不同,本文算法的运动补偿质量均较为稳定,在绝大部分情况下都取得了最高的 PSNR,而其他算法的运动估计精度却会随着视频特点的不同出现一定波动.

Table 5 Motion-Compensated PSNR Comparison

表 5 运动补偿的 PSNR 比较

Sequence	8 b-FS	8 b-DS	8 b-Elastic	2 b-FS	2 b-Elastic-NASep	Proposed
Akiyo	42.67	42.65	39.99	42.23	41.91	44.41
Bowing	42.39	40.78	38.57	39.65	40.40	42.88
Coastguard	29.71	29.19	30.93	29.30	30.64	30.97
Football	27.37	25.52	25.37	25.76	27.09	27.10
Foreman	33.52	32.47	30.25	31.97	33.43	34.26
Hall	34.48	34.37	34.92	33.37	34.74	36.14
Highway	36.08	34.63	30.63	33.97	35.88	36.42
Husky	19.70	19.17	19.83	19.46	20.93	20.95
Ice	30.99	30.11	29.47	29.10	31.84	32.08
Intros_cif	37.30	36.36	32.71	36.29	37.87	38.48
Mobile	23.96	23.58	27.02	23.61	25.11	25.34
Mother_daug	40.26	39.95	42.08	39.45	40.78	41.76
Pamphlet	43.12	41.87	41.85	40.88	38.83	43.86
Paris	30.70	30.50	32.38	30.17	29.47	32.27
Sign_irene	33.56	32.95	34.59	32.59	34.65	34.70
Soccer	29.61	27.79	28.42	28.32	29.67	29.81
Stefan	25.90	24.45	24.90	25.44	25.95	26.22
Students	39.49	39.40	41.09	38.79	38.93	41.16
Tempete	27.75	27.37	29.32	27.29	28.87	29.21
Tennis	28.91	27.16	28.21	27.95	28.74	29.05
Vtc1nw	44.70	44.70	42.08	44.01	43.77	46.46
Waterfall	35.53	35.52	39.43	35.41	36.64	37.28
Crew	32.36	31.64	32.17	30.93	33.48	33.53
City	30.86	29.46	28.77	30.29	31.06	31.24
Harbour	28.10	27.79	30.23	27.45	29.35	29.55
BQmall	29.63	28.22	28.59	28.71	29.56	30.14
Flowervase	37.42	37.34	36.24	35.71	37.66	39.10
sc_Robot	32.20	32.22	32.75	32.89	33.93	34.36
FourPeople	38.43	38.17	39.86	37.82	38.92	39.99
Johnny	38.83	38.61	39.71	38.08	39.71	40.37
KristenAndSara	39.38	39.16	41.01	38.71	39.47	40.93
BQTerrace	25.51	25.31	26.60	24.96	26.59	27.03
Cactus	29.53	28.95	30.98	28.57	30.52	30.84
ParkScene	29.98	29.57	31.14	29.43	31.08	31.18
Average	33.23	32.56	33.00	32.31	33.46	34.38

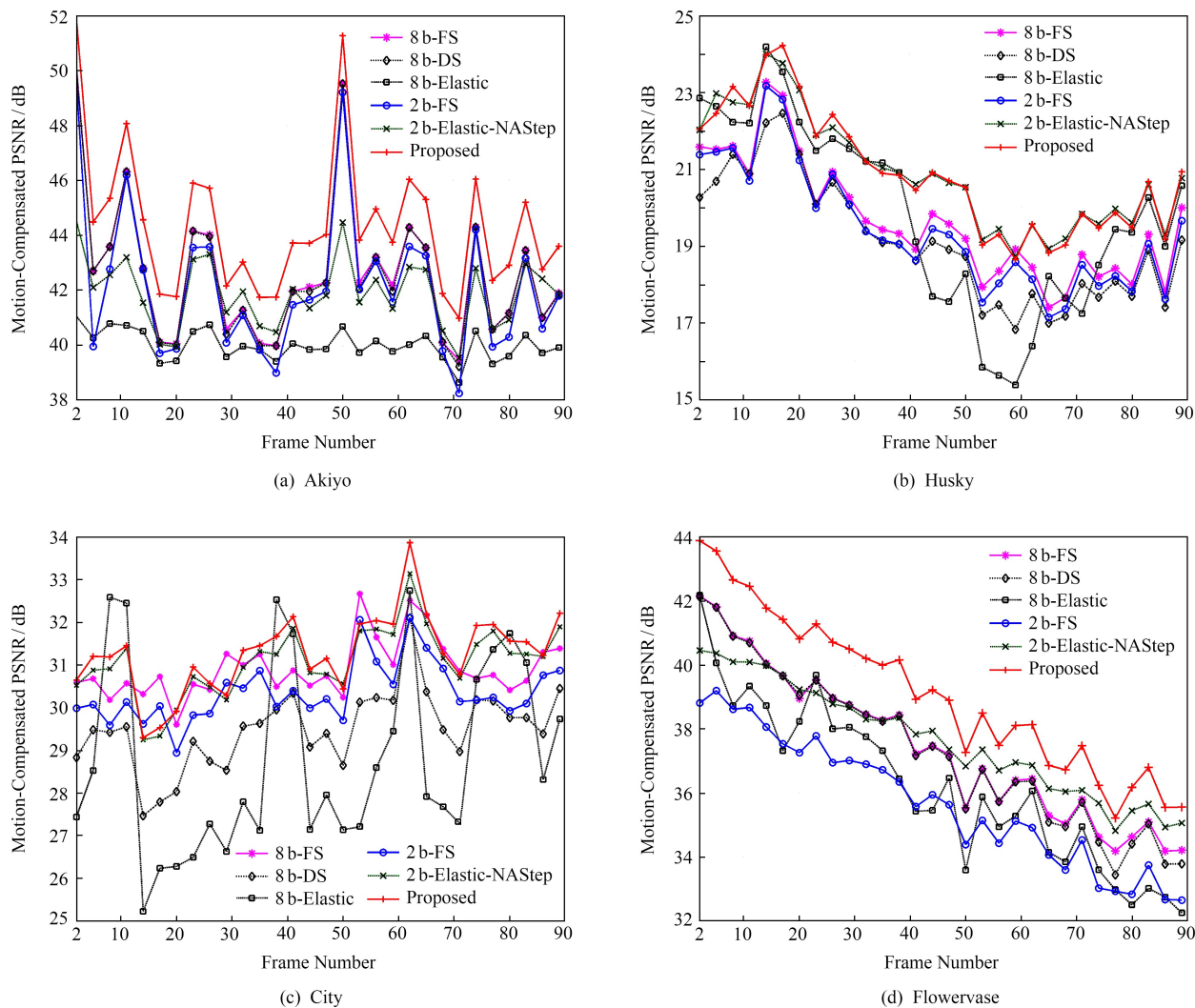


Fig. 7 PSNR comparison among different algorithms on the first 90 frames of Akiyo, Husky, City, and Flowervase

图7 Akiyo, Husky, City, Flowervase 的前 90 帧采用不同运动估计算法所得到的 PSNR 比较

需要指出,作为逆黑塞矩阵的 1 阶近似,本文的自适应步长对于前景运动慢、纹理细节简单或者含有慢速摇摄、拉摄运动的视频(如 Akiyo, Johnny, Flowervase)尤为有效,而对于含有快速运动和复杂纹理的视频(如 Husky 和 City),其性能增益则不够明显.究其原因在于,后者的匹配误差曲面非常复杂,导致每次迭代的最优步长之间存在较大差异,尤其当迭代过程从搜索起始点所在的单调区间进入其他单调区间时,初始步长的普适性难免会受到影响.这也是除了像素深度降采样这个原因以外,本文算法在 Mobile, Waterfall, Harbour 等序列上的运动补偿 PSNR 低于传统 8 b-Elastic 运动估计的根本缘故.当然,作为一种基于低位深度像素的快速运动估计算法,本文工作的基本目的是在计算量和运动估计精度之间尽量实现折中,而不是通过精确估计阻尼步长来取得最理想的补偿质量.从这个意义上讲,

本文提出的低位深度的弹性模型及其初始步长的自适应计算能够满足较低计算负载下的弹性运动估计需要.

7.2 收敛速度与迭代次数的讨论

收敛速度是衡量弹性运动估计效率的指标之一.图 8 所示为 Akiyo, Husky, City 序列第 2 帧采用本文算法的前 15 次迭代结果.从图 8 中能够发现,对于运动幅度小、细节简单的 Akiyo 序列,本文算法在第 1 次迭代后就趋于收敛;对于运动幅度较大、纹理复杂的 Husky 序列,运动补偿的 PSNR 随着迭代次数的增加而不断升高,但是经过 5 次迭代就已达到最高 PSNR 值的 99.97%;而对于含有摇摄运动且纹理丰富的 City 序列,本文算法在第 8 次迭代时取得了最高的运动补偿质量,之后出现 PSNR 的下降,不过在第 5 次迭代后已达到 PSNR 峰值的 99.99%;由于 Flowervase 序列的情况与 City 类似,

这里不再赘述.图 8 结果表明,一方面,本文算法在视频具有不同运动模式和空间分布特点的情况下,均能取得较快的收敛速度.另一方面,视频的运动幅度和模式,以及空间纹理复杂性对本文算法的收敛速度存在一定影响:当匹配误差曲面在全局/局部最优解附近呈现单峰分布时(如 Akiyo),本文算法将快速收敛,这与 3.2 节所讨论的 2 b 匹配误差曲面具有“区域内平滑、区域间锐化”的分布特点一致,下降梯度能提供足够的迭代驱动力;反之,当迭代过程需越过不同的单峰区间时,可能引起运动补偿质量的波动甚至小幅下降.这种情况与 7.1 节末尾的讨论一致,源于本文为了节省计算量,仅求解 1 次自适应步长并且未对相邻 2 次迭代的均方误差进行比较(若设 1 个视频帧包含 N 个像素,则计算均方误差的时间复杂度为 $O(Th \times N)$).

考虑到本文算法经过 5 次迭代即可获得超过最优补偿质量 99% 的 PSNR,并且此时迭代过程在很大概率下仍然处于起始搜索点的单调区间内,自适应步长对 $\|\mathbf{H}^{-1}\|$ 具有较高的逼近程度,我们在实验中将设置迭代次数阈值 $Th = 5$.同时,经过 2~3 次迭代后,本文算法就能够达到与 8 b-FS 相当的 PSNR,这也表明本文基于 Prewitt 梯度算子的像素深度转换方法保持了视频帧的纹理特征,并未由于深度降采样导致运动估计精度的明显下降,有效避免了传统像素深度转换方法(如文献[28-29])产生大面积平滑区时陷入局部最优的现象.

7.3 计算复杂度分析

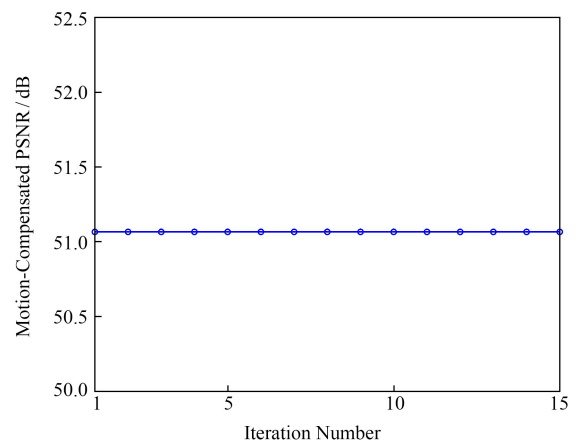
设宏块尺寸为 $B \times B$ 像素,搜索窗口为 $W \times W$ 像素,则不同运动估计算法处理 1 个宏块的计算量为:

1) 8 b-FS 和 2 b-FS 的时间复杂度均为 $O(B^2W^2)$.另外,2 b-FS 在像素深度变换阶段还需 $O(B^2)$ 的复杂度.

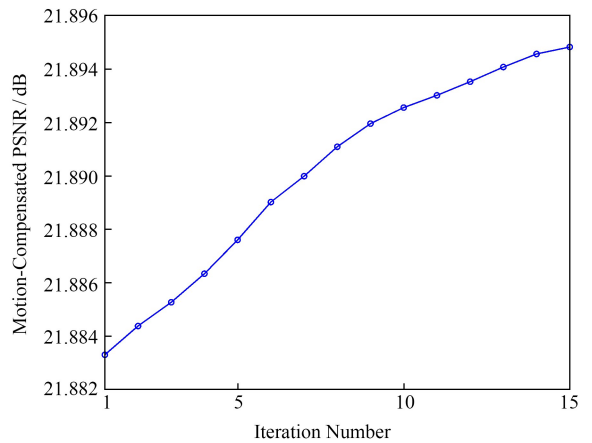
2) 由文献[31-32]可知,8 b-DS 的计算复杂度约为 8 b-FS 的 7%,即 $O(0.07B^2W^2)$.

3) 用高斯-牛顿法求解 8 b 深度弹性运动模型的每次迭代需计算偏导数、黑塞矩阵、逆矩阵、矩阵乘法、双线性插值和当前的运动补偿误差,其渐近时间复杂度分别为 $O(pB^2)$, $O(p^2B^2)$, $O(p^3)$, $O(pB^2)$, $O(B^2)$, $O(B^2)$.

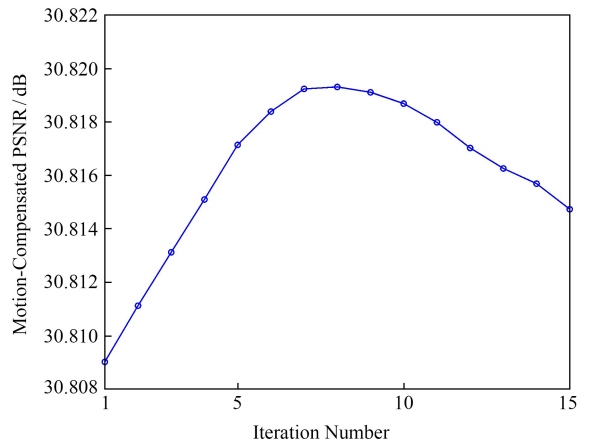
4) 由于省去了黑塞矩阵及其逆矩阵的计算,本文算法的每次迭代过程只需求解偏导数、向量乘法、双线性插值和运动补偿误差,其渐近时间复杂度分别为 $O(pB^2)$, $O(pB^2)$, $O(B^2)$, $O(B^2)$, 基于菱形



(a) Akiyo



(b) Husky



(c) City

Fig. 8 Result of the second frame of Akiyo, Husky, and City during the first 15 iterations

图 8 Akiyo, Husky, City 的第 2 帧的前 15 次迭代结果

搜索的初始点预测还需约为 $O(0.07B^2W^2)$ 的计算复杂度.

当搜索窗口 $W = 33$ 、块尺寸 $B = 16$ 、迭代次数阈值 $Th = 5$ 时,本文算法的时间复杂度约为 8 b-FS 算法和 2 b-FS 算法的 15.26%、8 b-Elastic 算法的

39.58%,但是高于 8 b-DS.若进一步将迭代次数 Th 减少到 2~3 次,本文算法则可在 8 b-FS 算法的 10.31%~11.60%的时间复杂度内,获得与之相当的运动估计精度.

具体地,表 6 给出了各种算法求解 1 个块的运动向量所需的平均操作次数,其中运动估计阶段的“Addition”操作包括 8 b 定点加法和浮点加法.从表 6 可见,本文方法的操作次数显著少于 8 b-FS, 2 b-FS 和 8 b-Elastic.尽管 8 b 加法、2 b 加法、乘法、

取绝对值、比较、位运算等各类操作所需的 CPU 周期数和计算代价存在明显差异,可是这些指标依赖于具体的硬件设计和指令集,已超出本文的讨论范围,所以在此不予进一步的定量比较,详细了解可参见文献[28].不过,考虑到本文算法采用位操作和比较操作来求解偏导数、向量乘法和运动补偿误差,其计算开销将低于 8 b 像素的加法和乘法,且有利于实现多个像素的并行计算,其实际开销比渐近时间复杂度更低.

Table 6 Average Operation Number Comparison Among Various Motion Estimation Algorithms on Each Block

表 6 不同运动估计算法对每个块的平均操作次数比较

Stage	Operation	8 b-FS	8 b-DS	8 b-Elastic	2 b-FS	Proposed
Bit-depth Conversion Stage	Fix-Point Addition	0	0	0	1 076	1 715
	Multiplication	0	0	0	264	263
	Comparison	0	0	0	768	640
	Absolute	0	0	0	0	512
	Square	0	0	0	0	1
Motion Estimation Stage	Addition	557 568	41 093	119 890	0	41 269
	2 b Addition	0	0	0	278 784	32 512
	Absolute	278 784	20 546	0	0	20 562
	Comparison	0	0	0	0	11 680
	Multiplication	0	0	155 840	0	60
	Bitwise Operation	0	0	0	836 352	20 480

8 结 论

为进一步改善弹性运动估计的实时性,本文引进了基于 Prewitt 梯度算子和正态分布的像素深度变换,以及基于位操作和比较操作的矩阵乘法和偏导运算,推导出高斯-牛顿法的自适应初始步长的 1 阶逼近方法,从而提出了一种基于 2 b 深度像素的弹性运动估计.其优点在于尽可能地简化求解弹性运动模型的高斯-牛顿法,用异或运算替代乘法运算,且无需反复计算黑塞矩阵及其逆矩阵.实验结果表明,该算法的运动补偿质量和计算效率优于 8 b 深度像素的块平移全搜索算法、菱形搜索算法、传统弹性搜索算法,以及传统 2 b 深度像素的块平移全搜索算法.

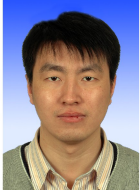
另外,本文算法仍存在若干问题可臻完善,如快速预测起始搜索点、根据匹配误差曲面的曲率变化来自适应地更新初始步长,以及设计能够高效执行本文算法的硬件电路等,我们将在今后的工作中进一步深入研究这些问题的解决思路.

参 考 文 献

- [1] Ma Siwei. History and recent development of AVS video coding standards [J]. Journal of Computer Research and Development, 2015, 52(1): 27-37 (in Chinese)
(马思伟. AVS 视频编码标准技术回顾及最新进展[J]. 计算机研究与发展, 2015, 52(1): 27-37)
- [2] Sinangil M E, Sze V, Zhou Minhua, et al. Cost and coding efficient motion estimation design considerations for high efficiency video coding (HEVC) standard [J]. IEEE Journal of Selected Topics in Signal Processing, 2013, 7(6): 1017-1028
- [3] Narroschke M, Swoboda R. Extending HEVC by an affine motion model [C] //Proc of the 30th PCS. Piscataway, NJ: IEEE, 2013: 321-324
- [4] Ahmmed A, Hannuksela M M, Gabbouj M. Fisheye video coding using elastic motion compensated reference frames [C] //Proc of the 23rd IEEE ICIP. Piscataway, NJ: IEEE, 2016: 2027-2031
- [5] Liu Dong, Wu Feng. Advances and trends of video coding technologies [J]. Communications of the CCF, 2016, 12(6): 20-23 (in Chinese)
(刘东, 吴枫. 视频编码技术发展及趋势[J]. 中国计算机学会通讯, 2016, 12(6): 20-23)

- [6] Ahmmed A, Xu Rui, Naman A T, et al. Motion segmentation initialization strategies for bi-directional inter-frame prediction [C] //Proc of the 15th IEEE MMSP. Piscataway, NJ: IEEE, 2013; 58-63
- [7] Wu Feng, Gao Peng, Gao Wen. Motion estimation technologies based on mesh model [J]. Acta Electronica Sinica, 2000, 28(5): 47-51 (in Chinese)
(吴枫, 高鹏, 高文. 基于网格模型的运动估计技术[J]. 电子学报, 2000, 28(5): 47-51)
- [8] Al-Regib G, Altunbasak Y, Mersereau R M. Hierarchical motion estimation with content-based meshes [J]. IEEE Transactions on Circuits and System for Video Technology, 2003, 13(10): 1000-1005
- [9] Cui Suxia, Wang Yonghui, Fowler J E. Motion estimation and compensation in the redundant-wavelet domain using triangle meshes [J]. Signal Processing: Image Communication, 2006, 21(7): 586-598
- [10] Huang Han, Woods J W, Zhao Yao, et al. Control-point representation and differential coding affine-motion compensation [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 23(10): 1651-1660
- [11] Li Li, Li Houqiang, Liu Dong, et al. An efficient four-parameter affine motion model for video coding [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 28(8): 1934-1947
- [12] Zheng Jiali, Qin Tuanfa, Ni Guangnan. Adaptive global motion estimation method based on rate distortion optimization [J]. Journal of Image and Graphics, 2011, 16(8): 1346-1352 (in Chinese)
(郑嘉利, 覃团发, 倪光南. 结合率失真优化的自适应全局运动估计方法[J]. 中国图象图形学报, 2011, 16(8): 1346-1352)
- [13] Po L M, Wong K M, Cheung K W, et al. Subsampled block-matching for zoom motion compensated prediction [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2010, 20(11): 1625-1637
- [14] Yuan Hui, Liu Ju, Sun Jiande, et al. Affine model based motion compensation prediction for zoom [J]. IEEE Transactions on Multimedia, 2012, 14(4): 1370-1375
- [15] Kim H S, Lee J H, Kim C K, et al. Zoom motion estimation using block-based fast local area scaling [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22(9): 1280-1291
- [16] Muhit A A, Pickering M R, Frater M R, et al. Video coding using elastic motion model and larger blocks [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2010, 20(5): 661-672
- [17] Muhit A A, Pickering M R, Frater M R, et al. Video coding using fast geometry-adaptive partitioning and an elastic motion model [J]. Journal of Visual Communication and Image Representation, 2012, 23(1): 31-41
- [18] Deng Binyou. Study on video compression based on elastic motion model [D]. Xiamen: Xiamen University, 2014 (in Chinese)
(邓斌攸. 基于弹性运动模型的视频压缩算法研究[D]. 厦门: 厦门大学, 2014)
- [19] Song Chuanming, Zhao Changwei, Liu Dan, et al. Elastic motion estimation of video using improved Gauss-Newton method [J]. Journal of Software, 2016, 27(11): 2946-2960 (in Chinese)
(宋传鸣, 赵长伟, 刘丹, 等. 采用改进高斯-牛顿法的视频弹性运动估计[J]. 软件学报, 2016, 27(11): 2946-2960)
- [20] Zhao Changwei. Research on motion estimation algorithm using elastic model [D]. Dalian: Liaoning Normal University, 2015 (in Chinese)
(赵长伟. 基于弹性模型的运动估计算法研究[D]. 大连: 辽宁师范大学, 2015)
- [21] Huang Han. Research on inter/intra prediction and optimization techniques in HEVC [D]. Beijing: Beijing Jiaotong University, 2013 (in Chinese)
(黄晗. HEVC 帧间/帧内预测及优化技术研究[D]. 北京: 北京交通大学, 2013)
- [22] Song Chuanming, Min Xin, Yan Xiaohong, et al. Fast elastic motion estimation using improved Levenberg-Marquardt method [J/OL]. Journal of Software, 2019. [2019-02-09]. <http://www.jos.org.cn/1000-9825/5487.htm> (in Chinese)
(宋传鸣, 闵新, 闫小红, 等. 采用改进 Levenberg-Marquardt 法的快速弹性运动估计[J/OL]. 软件学报, 2019. [2019-02-09]. <http://www.jos.org.cn/1000-9825/5487.htm>)
- [23] Haque M N, Biswas M, Pickering M R, et al. A low complexity algorithm for global motion parameter estimation targeting hardware implementation [C] //Proc of the 11th DICTA. Piscataway, NJ: IEEE, 2009; 1-5
- [24] Haque M N, Biswas M, Pickering M R, et al. An adaptive low-complexity global motion estimation algorithm [C] //Proc of the 27th PCS. Piscataway, NJ: IEEE, 2010; 598-601
- [25] Haque M N, Biswas M, Pickering M R, et al. A low-complexity image registration algorithm for global motion estimation [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22(3): 426-433
- [26] Song Chuanming, Guo Yanwen, Wang Xianghai, et al. Fuzzy quantization based bit transform for low bit-resolution motion estimation [J]. Signal Processing: Image Communication, 2013, 28(10): 1435-1447
- [27] Song Chuanming, Guo Yanwen, Wang Xianghai, et al. Motion estimation algorithm using 2 bit-depth pixel and fuzzy quantization [J]. Journal on Communications, 2013, 34(7): 59-70 (in Chinese)
(宋传鸣, 郭延文, 王相海, 等. 基于模糊量化和 2 bit 深度像素的运动估计算法[J]. 通信学报, 2013, 34(7): 59-70)
- [28] He Zhongli, Tsui C Y, Chan K K, et al. Low-power VLSI design for motion estimation using adaptive pixel truncation [J]. IEEE Transactions on Circuits Systems for Video Technology, 2000, 10(8): 669-678
- [29] Ertürk A, Ertürk S. Two-bit transform for binary block motion estimation [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2005, 15(7): 938-946

- [30] Diniz P S R. Adaptive Filtering: Algorithm and Practical Implementation [M]. New York: Springer US, 2013
- [31] Zhu Shan, Ma K K. A new diamond search algorithm for fast block-matching motion estimation [J]. IEEE Transactions on Image Processing, 2000, 9(2): 287-290
- [32] Li Renxiang, Zeng Bing, Liou M L. A new three-step search algorithm for block motion estimation [J]. IEEE Transactions on Circuits and Systems for Video Technology, 1994, 4(4): 438-442



Song Chuanming, born in 1980. Associate professor of the School of Computer and Information Technology of Liaoning Normal University. Received his PhD degree at the Department of Computer Science & Technology of Nanjing University. Member of CCF. His main research interests include image and video coding, and digital watermarking of multimedia.



Min Xin, born in 1992. Received his master degree in computer science and technology at the School of Computer and Information Technology of Liaoning Normal University. His main research interests include video coding. (hy_minxin@126.com)



Xie Weidong, born 1992. PhD candidate. Received his master degree in computer science and technology from the School of Computer and Information Technology of Liaoning Normal University. His main research interests include video and image coding. (xwdshiwo@163.com)



Yin Baocai, born in 1963. Professor and PhD supervisor of the School of Computer Science and Technology of Dalian University of Technology. Senior member of CCF. His main research interests include digital multimedia, multi-functional perception, virtual reality, as well as computer graphics. (ybc@dut.edu.cn)



Wang Xianghai, born in 1965. Professor and PhD supervisor of the School of Computer and Information Technology of Liaoning Normal University. Senior member of CCF. His main research interests include computer graphics and multimedia information processing. (xhwang@lnnu.edu.cn)