

基于模糊量化和 2 bit 深度像素的运动估计算法

宋传鸣^{1,2,3}, 郭延文², 王相海^{1,2,3}, 刘丹¹

(1. 辽宁师范大学 计算机与信息技术学院, 辽宁 大连 116029; 2. 南京大学 计算机软件新技术国家重点实验室, 江苏 南京 210093;
3. 南京邮电大学 江苏省图像处理与图像通信重点实验室, 江苏 南京 210046)

摘要:提出了一种 2 bit 深度像素的运动估计算法。首先, 将像素深度的降采样过程形式化为区间分划和区间映射 2 个步骤, 其中前者为多对一映射, 决定着运动估计性能, 后者为一对一映射; 其次, 提出一种非均匀量化方法求解区间分划的 3 个初始阈值, 并利用隶属度函数对初始阈值细化, 从而克服信号噪声等因素导致的初始阈值周围像素值的误匹配; 再次, 讨论了适用于 2 bit 深度像素运动估计的误差度量准则, 进而提出了基于模糊量化和 2 bit 深度像素的运动估计算法; 最后, 借助信号自相关函数, 建立比特深度转换误差—运动向量精度模型来估计该算法所能达到的预测精度。实验结果证明, 对于多种类型的视频序列, 尤其是场景细节和物体运动比较复杂者, 该算法始终能保持较高的估计精度, 运动补偿的平均峰值信噪比之传统 2 bit 深度像素的运动估计提高 0.27 dB。

关键词: 视频编码; 运动估计; 块匹配; 模糊量化; 低比特分辨率

中图分类号: TP391

文献标识码: A

文章编号: 1000-436X(2013)07-0059-12

Motion estimation algorithm using 2 bit-depth pixel and fuzzy quantization

SONG Chuan-ming^{1,2,3}, GUO Yan-wen², WANG Xiang-hai^{1,2,3}, LIU Dan¹

(1. College of Computer and Information Technology, Liaoning Normal University, Dalian 116029, China;

2. State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China; 3. Key Laboratory for Image Processing & Image Communication of Jiangsu Province, Nanjing University of Posts and Telecommunications, Nanjing 210046, China)

Abstract: A motion estimation algorithm was proposed using 2 bit-depth pixels. The reduction of pixel depth was first formalized by two successive steps, namely interval partitioning and interval mapping. The former is a many-to-one mapping which determines motion estimation performance, while the latter is a one-to-one mapping. A non-uniform quantization method was then presented to compute three initial thresholds of the interval partitioning. These initial thresholds were subsequently refined by using a membership function to solve the mismatch of pixel values near them caused by signal noise and so on. Afterwards, a matching criterion was discussed suitable for the motion estimation using 2 bit-depth pixels. A novel motion estimation algorithm was consequently addressed based on 2 bit-depth pixels and fuzzy quantization. To further predict the precision of the proposed algorithm, a bit resolution reduction error-motion vector precision model was built by exploiting the auto-correlation function. Extensive experimental results show that the proposed algorithm can always achieve high motion estimation precision characteristics, especially for those with detailed scene and complex motion. Compared with traditional 2 bit motion estimation, the proposed algorithm gains 0.27 dB improvement in terms of average peak signal-to-noise ratio of motion compensation.

Key words: video coding; motion estimation; block matching; fuzzy quantization; low bit-resolution

收稿日期: 2013-01-24; 修回日期: 2013-05-31

基金项目: 国家自然科学基金资助项目 (60703084, 41271422, 61073098); 辽宁省自然科学基金资助项目(20102123); 辽宁省博士科研启动基金资助项目(20121076); 辽宁百千万人才工程基金资助项目(2008921036); 计算机软件新技术国家重点实验室(南京大学)开放基金资助项目(KFKT2011B09, KFKT2011B11); 江苏省图像处理与图像通信重点实验室(南京邮电大学)开放基金资助项目(LBEK2010003, LBEK2011001); 辽宁省高等学校科学技术计划基金资助项目(L2011192); 大连市科学技术基金资助项目(2012J21DW008)

Foundation Items: The National Natural Science Foundation of China (60703084, 41271422, 61073098); The Natural Science Foundation of Liaoning Province(20102123); The Scientific Research Foundation for PhD of Liaoning Province (20121076); The BaiQianWan Talents Program of Liaoning Province (2008921036); The Open Foundation of State Key Laboratory for Novel Software Technology of Nanjing University (KFKT2011B09, KFKT2011B11); The Jiangsu Key Laboratory's Open Foundation of Image Processing and Image Communication of Nanjing University of Posts & Telecommunications (LBEK2010003, LBEK2011001); The Foundation of Science and Technology Plan for Higher Education of Liaoning Province (L2011192); The Foundation for Science and Technology of Dalian (2012J21DW008)

1 引言

随着互联网、无线通信、移动计算和普适计算等技术的稳步发展,多种不同类型的视频服务,诸如视频监控、移动视频电话和会议、移动流媒体、无线局域网内的视频访问等^[1],开始向社会经济的各个应用领域不断拓展,为具有不同计算能力、显示分辨率、接入带宽的终端设备提供跨平台、实时、可靠的视频通信。高效率的视频编码器是上述诸多应用能够成功推广的重要保证之一。现有的视频编码器,如 H.261/2/3、MPEG-1/2/4、H.264/AVC 和 AVS 等,均采用了以分块为单位的差分预测+变换的单闭环混合编码结构。其中,运动估计是一项有效的时域预测技术,而且多年来视频编码器效率的大部分提升均是由于其采用了更为优秀的运动估计算法^[2,3]。然而,运动估计环节所消耗的计算资源往往占编码器全部资源的 50% 以上^[4,5]。对于智能手机、掌上电脑等计算能力有限的设备,这样的计算强度并不实用,因此研究一种能够在复杂度和预测精度之间达到较好折衷的运动估计算法是十分必要的。

现有的运动估计算法可被划分为 4 类。1) 基于候选向量下采样的算法^[4~11]。该类算法的基本思想是通过将运动估计限制于搜索窗口中的部分候选向量集合的方式来减少匹配次数。绝大多数快速运动估计算法均属此类,如非对称十字多层次六边形格点搜索^[5]和菱形搜索^[8]等。2) 基于像素下采样的算法^[12~16]。该类算法的基本思想是在统计匹配误差时,只计算宏块中部分像素的匹配误差和,从而减少块匹配过程的运算次数。典型算法包括部分失真搜索^[12]、多分辨率搜索^[13,14]和自适应像素下采样搜索^[15,16]等。3) 基于低复杂度匹配函数的算法^[17]。该类算法的主要思路是在运动估计过程中采用计算复杂度较低的匹配误差函数代替均方误差(MSE, mean squared error)函数。常用的函数包括误差绝对值之和(SAD, sum of absolute difference)、像素误差分类(PDC, pel difference classification)和相异像素数目(DPC, different pixel count)等。4) 基于低比特深度像素的算法^[4,18~33]。这一类算法的主要思路是通过将 8 bit 深度的像素映射为深度较低的像素来降低匹配误差的计算开销。典型算法如 1 bit 深度像素的搜索^[18~27],在字长为 128 bit 的硬件环境下,它能并行地处理 128 像素的匹配误差。由于第 4 类

算法在减少硬件设计、降低能耗方面优势明显,较适用于处理能力有限的设备,而且还可用来降低其他 3 类算法的计算量,本文将工作集中于此。

本文提出一种基于模糊量化和 2 bit 深度像素的运动估计算法,其基本思想是首先将比特深度转换问题转化为均方误差意义下的最优量化进行求解;然后利用隶属度函数来刻画设备噪声、光照等因素引起的像素值变化,从而为每个像素值确定一个最佳的区间,即 2 bit 深度的表示;最后讨论了适用于 2 bit 深度像素运动估计的匹配误差度量准则,进而采用该准则下的全搜索策略进行 2 bit 深度像素的运动估计。另外,本文采用自相关函数来预测不同精度的运动向量所产生的匹配误差,并将该误差与比特深度转换误差建立联系,提出了一种比特深度转换误差—运动向量精度模型。

2 相关工作

目前,基于低比特深度像素的运动估计算法总体上可分为 3 类:1 bit 搜索^[18~27]、2 bit 搜索^[4,28~30]和更多比特的搜索^[31~33]。

文献研究得较为广泛的是 1 bit 深度像素的搜索。文献[18]利用二维低通滤波器实现比特深度转换,然后采用异或运算度量匹配误差,按照某种搜索策略完成运动估计;文献[19]用移位运算代替了文献[18]滤波过程中的乘法运算;文献[20]引进提前中止策略,改善了算法对低运动复杂度视频的估计效率;文献[21]采用离散余弦变换域的梯度算子进行比特深度转换,以充分利用编解码器的资源;文献[22]将 1 bit 搜索从整像素精度推广到分数像素精度;文献[23]则将分层运动估计策略与 1 bit 搜索相结合,并采用多种类型的低通滤波器进行比特深度转换,但同时增加了其硬件开销。此外,文献[24]改进了比特深度转换的硬阈值方法,提出对于 2 个同时分布在阈值的一定范围内的像素值,不再计算其匹配误差,具有一定的模糊理论思想,提高了运动估计精度。总体来讲,由于 1 bit 的表现能力非常有限,1 bit 深度像素运动估计的精度和补偿质量仍然不高,往往需要在所计算出的运动向量基础上,利用 8 bit 深度像素的运动估计进一步求精,从而得到理想的向量^[25~27]。故而,有必要采用更大的比特深度刻画像素。

在多于 2 bit 的运动估计算法方面,文献[31~33]均通过直接截去每个像素的较低位比特,获得低深

度的表示。然而文献[34]指出，运动补偿误差服从拉普拉斯分布或者高斯分布，这说明匹配误差往往产生在较低的比特位。如果这些比特被截断，将使得位于最佳匹配块周围的一定区域内，所有像素被量化成相同值，以致所有候选向量对应的匹配误差均为一常数，失去对最优向量的判断能力。可见，这种方法对于运动内容不剧烈或者空间复杂度不强的视频，其运动估计性能易受到影响。此外，比特深度越深，算法的计算量和硬件开销都将进一步增加。

文献[28,29]认为 2 bit 深度像素的运动估计可以在性能和复杂度之间达到比较理想的折衷。它们用自适应量化实现像素的 2 bit 深度转换，并利用 DPC 准则完成初始搜索；文献[4]选取像素均值和标准差的线性组合作为阈值进行 2 bit 深度转换，并用线性逼近方法快速求解标准差，减少了比特深度转换过程的开销；文献[30]使用文献[4]进行初始运动估计，得到 2 个向量。当二者的补偿误差均大于一个自适应的阈值时，再借助 8 bit 深度像素的搜索算法确定运动向量，其思路与文献[26]基本相同。虽然文献[4,28~30]表现出了优于 1 bit 深度像素的估计效率，可是与 8 bit 搜索相比，仍有一定差距，其主要瓶颈在于选取的深度转换方法和匹配误差度量准则。首先，在比特深度转换时，文献[4,28~30]为当前帧和参考帧的每个宏块各计算出一组阈值。这不仅使当前帧和参考帧的阈值各不相同，也会使同一帧中的相邻宏块具有不同阈值。而搜索窗口中往往包含多个宏块，当视频的空间复杂度较高时，这些宏块的边界将出现不连续，以致影响搜索精度；其次，比特深度转换均采用了硬阈值法，认为位于阈值两侧、但亮度差异很小的 2 个像素是不匹配的。尽管这符合块匹配运动估计的亮度恒定假设^[30]，可是由于设备精度、光照变化等因素，2 个最佳匹配宏块往往会有一定差异^[31]。因此，硬阈值易导致对最佳匹配像素的误处理，最后，对于 2 个 2 bit 深度的像素而言，其匹配程度共有 4 种情况，而文献[4,28~30]选用相异像素数目作为误差度量准则，只能描述匹配或者不匹配 2 种情况。这显然没有充分利用 2 bit 深度所提供的信息，不利于提高搜索精度。

与前人的工作相比，本文方法有以下 3 方面优点：1) 将比特深度转换形式化为量化问题，并借助最优量化方法求解，有利于获得最小的重构误差，

为提高搜索精度奠定基础；2) 隶属度函数能较为准确地刻画帧间噪声，可以有效地避免硬阈值引起的误匹配，使本文算法对存在光照变化和非平移运动等因素的视频也具有很好的顽健性，尤其是场景细节和物体运动较复杂者，始终能保持较高的估计质量；3) 比特深度转换误差—运动向量精度分析能给出运动向量精度的下界，它与真实下界的平均偏差不超过一个像素。对于给定视频，该模型可预测出本文算法的补偿质量相当于何种精度的 8 bit 深度像素的全搜索所达到的补偿质量，为不同场合的应用提供参考。

3 基于模糊量化的比特深度转换

将 8 bit 深度的像素降采样至 2 bit 深度，势必丢失视频帧的部分信息，这是 2 bit 深度像素运动估计性能下降的重要因素。因此，尽可能多地保留原始视频的信息是比特深度转换有效与否的关键。本节首先将比特深度转换问题转化为量化阈值的选取问题，并用最优量化理论给出解决思路；然后，考虑到帧间噪声等因素对初始阈值的影响，引进隶属度函数确定最终阈值。

3.1 基于最优量化的比特深度转换

令 $G = \{0, 1, 2, \dots, 255\}$ ， $B = \{00, 01, 10, 11\}$ ，映射 $f: G \rightarrow B$ ， $g \rightarrow b$ 表示从 G 到 B 的比特深度转换。再将转换过程分解为区间分划 $t: G \rightarrow G$ ， $g \rightarrow r$ 和区间映射 $m: G \rightarrow B$ ， $r \rightarrow b$ ，则有

$$f(g) = m(t(g)) \quad (1)$$

和

$$t(g) = \begin{cases} r_1, & -1 < g < T_1 \\ r_2, & T_1 < g < T_2 \\ r_3, & T_2 < g < T_3 \\ r_4, & T_3 < g < 255 \end{cases}, \quad m(r) = \begin{cases} 00, & r = r_1 \\ 01, & r = r_2 \\ 10, & r = r_3 \\ 11, & r = r_4 \end{cases} \quad (2)$$

其中，区间分划为多对一映射，区间映射为一对一映射，故区间分划决定了阈值选取和信息失真。为了使经过比特深度转换的像素保留尽可能多的视频信息，就要求解出一种能获得最小重构均方误差的映射 t ，即

$$D = E \left[(g - t(g))^2 \right] = \sum_{g=0}^{255} p(g) (g - t(g))^2 \quad (3)$$

其中， $E[\cdot]$ 和 $p(\cdot)$ 分别表示数学期望和像素值的概率分布函数。根据均方误差意义下的最优量化理

论, 标量量化的目标是为一个概率分布为 p 的随机信号 g , 确定一组量化阈值和重构电平 r , 使其重构误差最小。显然, 区间划分与标量量化是等价的。所以, 可以利用最优标量量化的理论来求解式(2)和式(3)。由文献[35,36]可知, 欲求最优区间分划 t , 就是要确定一组 r_i 和 T_j ($i \in \{1, 2, 3, 4\}$, $j \in \{1, 2, 3\}$)。

Max J 为该非线性问题提出了一种甚为复杂的递归解法^[35], 并且需要给出先验概率分布 p 。后来, Lloyd S P 又设计了一种试错(trial-and-error)法^[36], 通过多次尝试和迭代来计算 r_i 和 T_j , 但该方法的计算代价仍很高。

Max J 和 Lloyd S P 方法的计算开销主要花费在非均匀量化步长的迭代求解过程中。由最优标量量化理论可知, 若已知像素值为均匀分布 U , 则均匀量化就是其最佳的量化策略。这样, 若可使待量化的像素值满足均匀分布, 就可避免迭代求解。基于上述思路, 下文讨论一种快速算法。

首先, 利用直方图均衡变换 e 将像素值 g 映射为 g' , 从而使 g' 满足均匀分布 $U: g': U(0, 255)$ 。

$$e(g) = \left\lfloor \sum_{k=0}^g 255 p(k) \right\rfloor \quad (4)$$

其次, 均匀量化 g' 。2 bit 深度转换需要 3 个量化阶, 则阈值为 $T'_j = 64j - 1$, $j \in \{1, 2, 3\}$ 。

最后, 利用逆变换 $T_j = e^{-1}(T'_j)$ 计算 g 的初始区间分划阈值。由于直方图均衡变换不是一一映射, 由反函数的性质, e 不存在逆变换, 故而本文定义 e 的逆变换为

$$e^{-1}(T'_j) = \min \{k\} \quad \text{s.t.} \quad \{e(k) \geq T'_j\} \quad (5)$$

其中, $k \in \{0, 1, 2, \dots, 255\}$ 。再令 $T_0 = -1$, $T_4 = 255$, 进而有

$$r_i = \lfloor (T_i - T_{i-1})/2 \rfloor \quad (6)$$

上述方法能以很低的计算量处理服从任意概率分布的视频序列。但是, 上述方法只能获得非均匀量化的次最优解, 其原因在于直方图均衡变换往往不能将离散的像素值映射为真正的均匀分布。另外需指出, 除了直方图均衡变换以外, 也可采用具有相同映射作用的变换。

3.2 基于隶属度函数的初始阈值细化

经过基于最优量化的比特深度转换后, 获得了当前帧和参考帧的 2 bit 深度的表示, 进一步可利用

它们进行块匹配运动估计。按照块匹配运动估计的亮度恒定假设, 即当前宏块及其最佳匹配宏块中的对应像素具有相同的像素值, 可知当前宏块及其最佳匹配宏块在比特深度转换后也应相同。但是, 实际应用中设备噪声和光照条件等因素的存在却导致 2 个最佳匹配块之间存在一定差异^[37], 即

$$\hat{M}_c(x, y) = M_r(x + u, y + v) = M_c(x, y) + n(x, y) \quad (7)$$

其中, (u, v) 表示运动向量, 它建立了当前帧宏块 $M_c(x, y)$ 与参考帧宏块 $M_r(x + u, y + v)$ 之间的预测关系; \hat{M}_c 表示运动估计获得的当前宏块的预测; $n(\cdot, \cdot)$ 表示随机噪声。同时, 为了刻画视频帧的像素值变化, 本文将视频看做一个随机场, 将每个像素视为一个随机变量, 并用方差 s_g^2 衡量每个像素值的变化。此时, 若令当前宏块的最佳预测为 \hat{M}_c^* , 则 $n(\cdot, \cdot)$ 和像素值的变化 s_g^2 将导致存在某些 (x, y) , 有 $\hat{M}_c^*(x, y) \neq M_c(x, y)$ 。当 $\hat{M}_c^*(x, y) > T_j$ 且 $M_c(x, y) < T_j$ 时, $\hat{M}_c^*(x, y)$ 和 $M_c(x, y)$ 将被划分到不同区间, 继而映射到不同的值, 反之亦然, 此时 2 个像素的匹配误差至少为 1。然而, 对于 2 bit 深度的像素来说, 2 个像素的最大匹配误差仅为 4。一旦比特深度转换不准确, 就会由于这至少 25% 的匹配偏差导致 $\hat{M}_c^* \neq \hat{M}_c$ 。这正是基于低比特深度像素的运动估计方法, 尤其是 1 bit、2 bit 深度像素的运动估计, 对阈值选取非常敏感的关键所在。综上所述, 忽略噪声和像素变化而进行比特深度转换, 不利于搜索到最佳的运动向量。为了减小比特深度转换导致的偏差, 就需要对分布在初始阈值两侧一定范围内的像素值进行细化。如果某像素值属于当前区间的概率不高, 则修改初始阈值, 使该像素值分布在最可能的区间。为实现这一目的, 本文定义了一种隶属度函数来度量某像素值属于每个区间的概率。

3.2.1 隶属度函数的定义

初始阈值 (T_1, T_2, T_3) 给出了 G 的 4 个区间分划 Z_0 、 Z_1 、 Z_2 和 Z_3 , 其中, T_j 是 Z_{j-1} 和 Z_j 的分界点。由于存在设备噪声、编码失真和光照变化等, 分布于 T_j 一定范围内的像素值 g , 既可能原本属于 Z_{j-1} , 又可能原本属于 Z_j , 本文将 g 属于 Z_j 的概率称为隶属度, 并利用一个隶属度函数来度量。隶属度函数的定义为

$$M_1(g) = \int_g^{+\infty} p_1(s) ds \quad (8)$$

$$M_2(g) = 1 - \int_g^{+\infty} p_1(s) ds - \int_{-\infty}^g p_2(s) ds \quad (9)$$

$$M_3(g) = 1 - \int_g^{+\infty} p_2(s) ds - \int_{-\infty}^g p_3(s) ds \quad (10)$$

$$M_4(g) = \int_{-\infty}^g p_3(s) ds \quad (11)$$

其中, $M_i(g)$ 表示 g 属于区间 Z_i 的隶属度; p_j 表示 T_j 的概率密度, 它与噪声 (本文将设备噪声和编码失真统称为噪声) 和视频信号本身均有关。假设噪声和视频信号相互独立, 且噪声服从正态分布 $n: N(0, s_n^2)$, 则 T_j 的概率密度函数 $p_j(s)$ 为

$$p_j(s) = \frac{1}{\sqrt{2p(s_n^2 + s_g^2)}} \exp\left[-\frac{(s - T_j)^2}{2(s_n^2 + s_g^2)}\right] \quad (12)$$

上述隶属度函数的示意如图 1 所示。根据具体应用的不同, 也可为不同的 T_j 选择不同的概率密度函数。

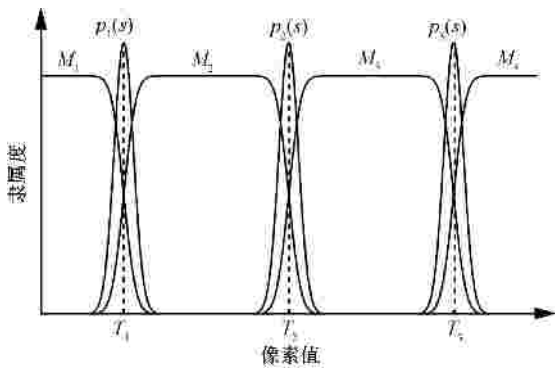


图 1 隶属度函数

3.2.2 初始阈值细化

在定义了隶属度函数之后, 像素值的最优区间分划就转化为对最大隶属度的求解, 也就是确定一组更加准确的区间分划 (T_1'', T_2'', T_3'') , 使得

$$\begin{aligned} & \sum_{g=0}^{T_1''} M_1(g)p(g) + \sum_{g=T_1''+1}^{T_2''} M_2(g)p(g) + \\ & \sum_{g=T_2''+1}^{T_3''} M_3(g)p(g) + \sum_{g=T_3''+1}^{255} M_4(g)p(g) \\ \text{s.t. } & |T_j - T_j''| \leq \sqrt{s_n^2 + s_g^2} \end{aligned} \quad (13)$$

取得最大值。根据正态分布的特性, 噪声和像素值变化的幅度有近 70% 的概率会在一倍方差以内。所以, 本文利用式(13)后部的约束条件限制 (T_1'', T_2'', T_3'') 的选取范围, 使之在 T_j 的 1 倍方差范围内寻找 T_j'' 。若超出 1 倍方差范围, 得出的新阈值很可能不是均

方误差意义下的最优。下面分 2 种情况讨论 (T_1'', T_2'', T_3'') 的求解方法。

首先, 若概率分布 p 为均匀分布, 则 $p(\cdot)$ 可作为公因子从式(13)的各项中提取出。

$$\begin{aligned} & \sum_{g=0}^{T_1''} M_1(g) + \sum_{g=T_1''+1}^{T_2''} M_2(g) + \\ & \sum_{g=T_2''+1}^{T_3''} M_3(g) + \sum_{g=T_3''+1}^{255} M_4(g) \\ \text{s.t. } & |T_j - T_j''| \leq \sqrt{s_n^2 + s_g^2} \end{aligned} \quad (14)$$

这样式(14)取得最大值与式(13)取得最大值等价。从图 1 可见, 此时的 T_1'' 、 T_2'' 和 T_3'' 分别应为 M_1 与 M_2 、 M_2 与 M_3 、 M_3 与 M_4 的交点, 即 T_1 、 T_2 和 T_3 。这一结果也与最优标量量化的结论一致。

其次, 概率分布 p 为不均匀分布, 由于几乎所有图像的像素值分布都是不均匀的, 视频亦属于此类情况。这时, 可以利用某种最优化方法来求解式(13)的最大值。但是, 若为每个视频帧都计算一次最优区间分划, 则难免带来很高的计算开销。于是, 近似求解是一种折衷的途径。

因为 T_j 是经过直方图均衡得到的, 所以由式(4)可推知每个区间 Z_i 内的像素数目 (或像素累积分布) 基本相同。若令 $|Z_i| = T_{i+1} - T_i$ 表示区间长度, 则 $|Z_i|$ 越小, 区间 Z_i 内像素值的平均概率分布 \bar{p} 越大。由于均匀分布对应的区间长度为 64, 本文将 $64/|Z_i|$ 作为权重来粗略度量 p , 则式(13)转化为

$$\begin{aligned} & \frac{64}{T_1''+1} \sum_{g=0}^{T_1''} M_1(g) + \frac{64}{T_2''-T_1''} \sum_{g=T_1''+1}^{T_2''} M_2(g) + \\ & \frac{64}{T_3''-T_2''} \sum_{g=T_2''+1}^{T_3''} M_3(g) + \frac{64}{255-T_3''} \sum_{g=T_3''+1}^{255} M_4(g) \\ \text{s.t. } & |T_j - T_j''| \leq \sqrt{s_n^2 + s_g^2} \end{aligned} \quad (15)$$

此时, T_j'' 的选择与其所在区间的权重直接相关。 $|Z_{j-1}|$ 和 $|Z_j|$ 的差越大 (假设 $|Z_{j-1}| > |Z_j|$), 说明原始视频帧中分布在 Z_j 内和 Z_{j-1} 内的像素值数目差异越大, 位于 T_j 两侧一定范围的像素值原本属于 Z_j 的概率则较高, 因为这些像素值很有可能是受到噪声影响才被划分到 Z_{j-1} 内的, 这样 T_j'' 应远离 T_j , 向 T_{j-1} 移动。

基于上述思路, 本文给出一种计算近似解的快速方法, 其主要思想是根据各个区间的长度和视频

信号的方差,自适应地扩展权值较大的区间,间接缩小权值较小的区间,进而实现调整区间分划阈值的目的。具体步骤如下:

Step1 初始化:令 $j=0$ 。

Step2 若 $|Z_j| \geq 64$, 则令 $s_n = \min\{s_n, 16\}$, $s_g = \min\{s_g, 32\}$, $D = (64/|Z_j|) \left[\sqrt{s_n^2 + s_g^2} / 2 \right]$, $|Z'_j| = \min\{|Z_j| + 2D, 64 + s_n\}$; 否则, 令 $|Z'_j| = |Z_j|$, 转入 Step4。

Step3 $j = j + 1$ 。如果 $j = 4$, 则转入 Step4; 否则, 转入 Step2。

Step4 令 $L = \sum_j |Z'_j|$, 再更新各区间的长度和阈值, 即令 $|Z'_j| = 256 \times |Z'_j| / L$ ($j \in \{0, 1, 2, 3\}$), $T_0'' = -1$, $T_4'' = 255$, $T_i'' = T_{i-1}'' + |Z'_{i-1}| - 1$ ($i \in \{1, 2, 3\}$)。算法结束。

将计算得到的 $\{T_j''\}$ 代入式(6)和式(2), 即可求出每个视频帧的比特深度转换结果 B 。

需要指出, 上述 Step2 中参数 ρ 的作用是减少区间调整的次数, ρ 越小, 需要的计算次数就越少, 从而控制了计算量。

4 2 bit 深度像素的运动估计算法

视频帧经过比特深度转换后, 就可以用某种匹配误差度量准则和搜索策略进行运动估计。本节首先讨论适用于 2 bit 深度像素运动估计的匹配误差度量准则, 然后提出一种 2 bit 运动估计算法。

4.1 2 bit 深度像素的匹配误差度量准则

匹配误差度量准则是高效率运动估计的一项重要保证。大多数现有的 2 bit 深度像素的运动估计算法均采用了基于异或运算的相异像素数目准则, 其定义为^[4]

$$NNMP(u, v) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \left\{ \left[B_c^l(x, y) \oplus B_r^l(x+u, y+v) \right] \oplus \left[B_c^h(x, y) \oplus B_r^h(x+u, y+v) \right] \right\} \quad (16)$$

其中, $B_c^l(\cdot, \cdot)$ 和 $B_c^h(\cdot, \cdot)$ 分别为当前帧进行比特深度转换后的低位平面和高位平面; $B_r^l(\cdot, \cdot)$ 和 $B_r^h(\cdot, \cdot)$ 分别为参考帧进行比特深度转换后的低位平面和高位平面; \oplus 表示异或运算; P 表示或运算(OR)。

相异像素数目准则计算简单, 尤其便于硬件实现, 而且 2 bit 的匹配误差能够并行计算。可是, 该

准则只能描述匹配或者不匹配 2 种情况。而对于深度为 2 bit 的像素, 其值域为 $\{00, 01, 10, 11\}$, 这 4 个值之间的匹配误差绝对值共有 4 种取值, 即 $\{00, 01, 10, 11\}$ 。用只能度量 2 种匹配误差的准则评价存在 4 种取值的误差, 这显然是不准确的, 没有充分利用 2 bit 深度所提供的信息, 不利于提高运动估计的预测精度。

因此, 本文采用能够度量 4 种取值的基于位截断(BT, bit truncation)的 SAD^[29]作为运动估计阶段的匹配误差度量准则, 其定义为

$$BTSAD(u, v) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \left| I_c^{q:q+1}(x, y) - I_r^{q:q+1}(x+u, y+v) \right| \quad (17)$$

其中, c 和 r 用于标识当前宏块和参考宏块, 上标 $q:q+1$ 表示匹配运算仅在第 $q \sim (q+1)$ bit 上进行。这样可以将多个像素的 2 bit 表示存储在 1 byte 中, 既节省缓存空间, 又可以并行计算多个像素的匹配误差。此外, 由于匹配误差只有 4 种情况, 在具体实现时可首先在内存中建立一个表格用于存储 4 种误差值, 再以 $I_c^{q:q+1}(x, y)$ 和 $I_r^{q:q+1}(x+u, y+v)$ 为偏移地址、用汇编语言的查表指令完成匹配, 其需要的机器时钟周期数目要少于先做减法、再取绝对值所需的机器时钟周期数目。

4.2 2 bit 深度像素的运动估计

在上文阐述的基础上, 本文提出一种基于模糊量化和 2 bit 深度像素的运动估计算法, 其基本思想是首先利用模糊量化确定区间分划的阈值, 从而将当前帧和参考帧转换为 2 bit 深度像素的表示, 然后采用 BTSAD 准则在参考帧的一定范围内, 为当前帧的每个宏块搜索运动向量。算法步骤如下。

Step1 计算当前帧和参考帧的像素均值 $m_c = E[I_c]$ 、 $m_r = E[I_r]$ 和方差 $s_c^2 = E[I_c^2] - E^2[I_c]$ 、 $s_r^2 = E[I_r^2] - E^2[I_r]$ 。

Step2 将当前帧进行直方图均衡变换得到 g' 。

Step3 对 g' 施以均匀量化, 从而确定 $\{T_j''\}$ 。

Step4 采用式(5)计算初始阈值 $\{T_j''\}$ 。

Step5 计算像素方差 s_g^2 和噪声方差 s_n^2 。本文假设噪声和视频信号相互独立, 故有 $s_n^2 = |s_c^2 - s_r^2|$, $s_g^2 = z s_c^2$, z 表示比例常数。

Step6 利用 2.2.2 节的快速算法计算区间分划的阈值 T_j'' , 并用该组阈值对当前帧和参考帧进行比

特深度转换。

Step7 对当前帧进行基于块的全搜索运动估计(FS, full search)(或者采用某种基于块的快速搜索策略), 匹配误差度量准则为 BTSAD。算法结束。

5 2 bit 深度像素运动估计的性能分析

虽然 2 bit 深度像素的运动估计可有效降低传统运动估计的计算量, 可是, 原始视频信息的损失势必降低其运动向量的精度。至于精度究竟会降低多少, 从作者所掌握的文献来看, 报道中并没有进行过分析和讨论。本节首先分析比特深度转换所导致的误差, 进而讨论运动向量的精度, 最后分析算法的计算量。

5.1 比特深度转换误差——运动向量的精度分析

由 2.1 节可知, 本文的比特深度转换误差源于映射 t , 并且等于其逆量化的重构误差, 再由式(6)可得, 区间 Z_{i-1} ($i \in \{1, 2, 3, 4\}$) 内各个像素值的重构值是 r_i , 而 $r_i = \lfloor (T_i'' + T_{i-1}'') / 2 \rfloor$, 则利用式(3)就可计算出平均比特深度转换误差的数学期望为

$$D = \sum_{i=1}^4 \sum_{g=T_{i-1}''+1}^{T_i''} p(g) \left[g - \lfloor (T_i'' + T_{i-1}'') / 2 \rfloor \right]^2 \quad (18)$$

比特深度转换误差有可能将位于最佳匹配宏块 \hat{M}_c^* 周围、半径为 d 像素的区域内所有像素量化成相同值, 导致这个区域内所有候选向量产生相同的匹配误差, 无法判断最优向量。本文称此时运动向量可达到 d 像素精度。故此, 部分细节信息丢失是运动向量不准确的关键成因, 而比特深度转换误差则决定着 d 的高低。

若假设利用 8 bit 深度像素的全搜索计算出的向量为理想运动向量, 则图 2 可很直观地说明不同精度运动向量之间的关系。在图 2 的参考帧中, 深色宏块表示最佳匹配宏块 \hat{M}_c^* , 浅色宏块是 2 bit 深度像素运动估计确定的匹配宏块 \hat{M}_c , 后者相当于是前者经空间平移 (d, d) 得到的。此时, 宏块 \hat{M}_c^* 与 \hat{M}_c 的平均像素差异约为

$$D' = E \left\{ I_r(x, y) - I_r(x+d, y+d) \right\}^2 = E \left\{ I_r(x, y) \right\}^2 + E \left\{ I_r(x+d, y+d) \right\}^2 - 2E \left\{ I_r(x, y) I_r(x+d, y+d) \right\} \quad (19)$$

假设搜索窗口内的像素处于一致分布的随机

场, 则式(19)可转化为

$$D' = 2(m_r^2 + s_r^2) - 2r(d, d) \quad (20)$$

其中, $r(\cdot, \cdot)$ 表示随机场的自相关函数。

$$r(d, d) = E \left[I_r(x, y) I_r(x+d, y+d) \right] \quad (21)$$

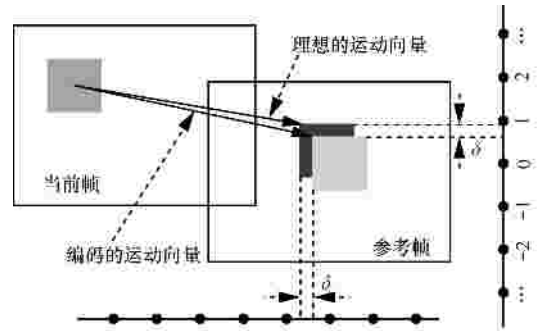


图 2 不同精度运动向量间的关系示意

由于当 $d \neq 0$ 时噪声的自相关函数值往往等于 0, 此时可简单认为 \hat{M}_c^* 与 \hat{M}_c 的差异完全是由比特深度转换所导致的, 即

$$D \approx D' \quad (22)$$

这样, 利用比特深度转换误差和参考帧的低阶统计参数就可以考察 2 bit 深度像素的运动估计精度。具体步骤如下。

Step1 利用式(18)计算比特深度转换误差 D 。

Step2 计算参考帧的像素均值 m_r 和方差 s_r^2 。

令 $d=1$ 。

Step3 计算参考帧的自相关函数值 $r(d, d)$, 再将结果代入式(20)计算匹配误差的预测 D' 。

Step4 如果 $D > D'$, 说明比特深度转换误差将掩盖错误运动向量所导致的匹配误差, 无法在 d 像素精度下区别出 \hat{M}_c^* , 那么令 $d=d+1$, 转入 Step3; 否则, 输出 d , 算法结束。

上述算法的输出值 d 就是 2 bit 深度像素运动估计搜索到的运动向量的精度, 它可以衡量 \hat{M}_c 与 \hat{M}_c^* 的平均距离。

5.2 计算量分析

计算时间复杂度是衡量运动估计算法性能的重要指标之一。下文将加、减法通称加法, 乘、除法通称乘法, 并设视频帧的宽度为 w , 高度为 h , 则可从 2 个方面来统计本文算法的计算量。

1) 比特深度转换: 该过程包括直方图均衡变换和基于隶属度函数的阈值细化 2 个步骤。直方图

均衡变换：统计当前帧的概率分布 p 需 wh 次加法，直方图均衡 e 需 256 次乘法和 255 次加法；计算 e^{-1} 至多需 255 次比较运算。基于隶属度函数的阈值细化。首先，计算当前帧和参考帧的像素均值需 $2(wh-1)$ 次加法，计算两帧的像素方差需 $2wh$ 次乘法和 $2wh$ 次加法；其次，计算 s_n^2 和 s_g^2 需 1 次加法、1 次乘法和 1 次取绝对值运算；再次，2.2.2 节的阈值细化算法中 Step2 至多需 10 次加法、6 次乘法、1 次开平方和 4 次比较运算，Step4 又需 9 次加法和 6 次乘法运算；最后，将每个像素划分到不同区间最多需 3 次比较运算。

综合和可得，对于 $352 \times 288 \times 2$ 像素的视频帧（包括当前帧和参考帧），每个像素为完成比特深度转换需做 1.250 7 次加法、0.500 7 次乘法和 3.001 3 次比较。此外，两帧还需 1 次取绝对值和 1 次开平方运算。

2) 基于块匹配的运动估计。以全搜索为例，由式(17)可知，对于 $W \times W$ 的搜索窗口，执行全搜索共需 $W \times W \times N \times N$ 次 2 bit 加法和 $W \times W \times 4 \times N \times N$ 次位运算。特别地，当 $W = 32$ 、 $N = 16$ 时，平均每个像素需进行 1 024 次 2 bit 加法和 4 096 次位运算。

6 实验结果与分析

为了验证本文所提出算法的性能，分别以 CIF (common intermediate format) @ 30 帧/s 的 Akiyo、Bus、Football、Flower、Husky、Mobile & Calendar、Paris、Tennis、Silent、Stefan、Soccer、Sign Irene、

Tempete, QCIF (Quarter CIF) @ 30 帧/s 的 Salesman 和 4CIF @ 30 帧/s 的 City 和 Harbour 等序列的 1~90 帧为例，进行了大量实验，并将结果与 8 bit 深度像素的全搜索(下文简称 8b-FS)、1 bit 深度像素的 FS^[18] (1b-FS)、2 bit 深度像素的 FS^[4] (2b-FS)、基于位截断的 FS^[32] (T-FS) 和基于自适应量化的 2 bit 深度像素的 FS^[29] (AQ-FS) 等 5 种算法相比较。实验参数设置如下：搜索窗口为 32×32 像素，宏块尺寸为 16×16 像素，搜索精度为整数像素。本文算法(FQ-FS)的参数值分别设置成 $W = 32$ ， $N = 16$ ， $z = 0.2$ 和 $\rho = 0.625$ 。重构帧的质量采用峰值信噪比(PSNR, peak signal-to-noise)进行评价。注意，原始的 AQ-FS 算法^[29] 包括 2 个步骤。为公平比较 FQ-FS 和 AQ-FS 的 2 bit 深度像素运动估计的有效性，本文只与 AQ-FS 算法第 1 步的预测进行对比，而忽略其后续的 8 bit 深度像素的搜索结果。

6.1 运动估计/补偿质量的比较

在人眼主观质量方面，分别利用 6 种算法对 Football 序列的第 2 帧进行运动估计和补偿，得到的补偿帧如图 3 所示。对比发现，图 3(c)~图 3(f) 中运动员的腿部、肘部、小腿至脚踝处等运动幅度比较大的部位均有明显的影子(ghosting artifact)或块效应(blocking artifact)，而图 3(g) 中除了 1 处影子外，其他部分与 8b-FS 的补偿质量(图 3(b)) 几乎相同。可见，FQ-FS 能获得最接近理想结果的运动估计性能，尤其是在物体运动较为剧烈的区域。

在客观质量方面，表 1 罗列了各个测试序列的

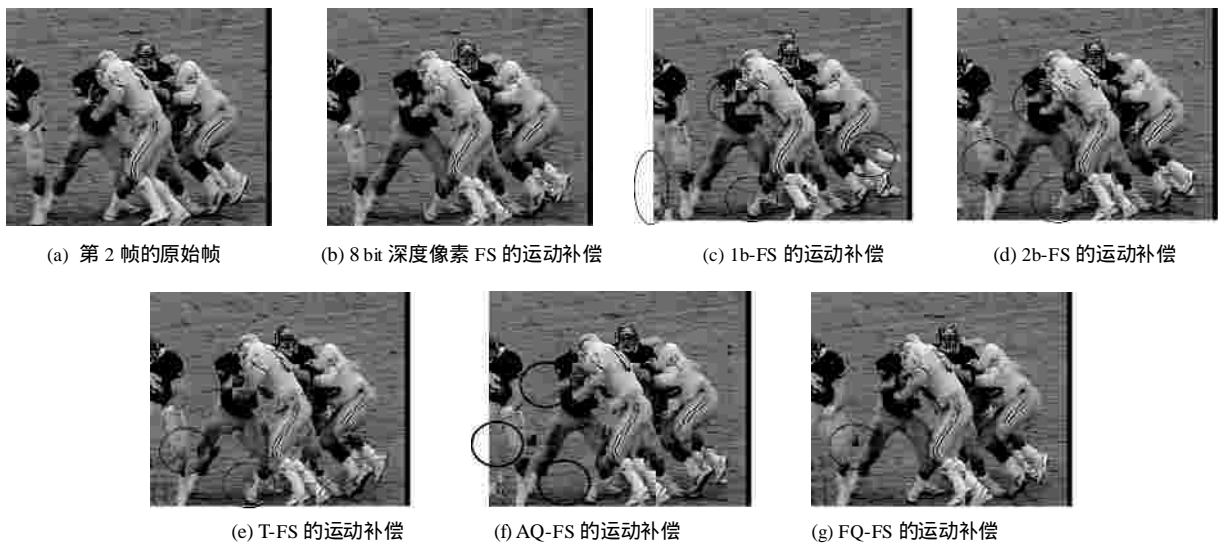
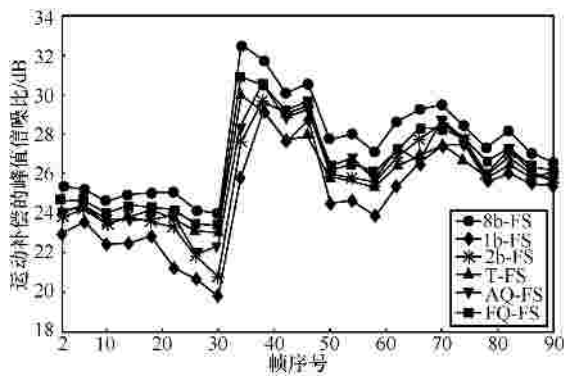


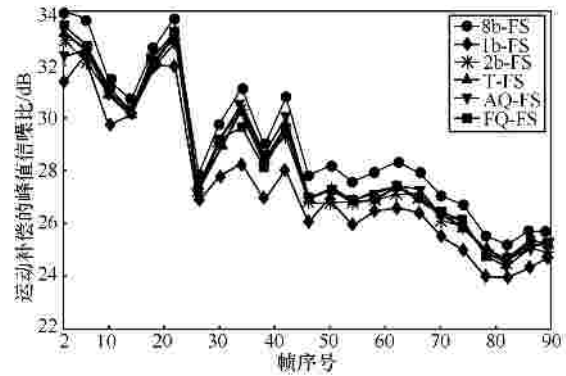
图 3 Football 序列的第 2 帧分别经过 8b-FS、1b-FS、2b-FS、T-FS、AQ-FS 和本文算法 FQ-FS 的运动补偿帧。(b)~(h)的 PSNR 分别为 25.37 dB、22.97 dB、23.81 dB、24.11 dB、24.08 dB 和 24.64 dB

亮度分量采用不同算法所得到的平均 PSNR。图 4(a)~图 4(d)所示分别为 Football、Tennis、Silent 和 Paris 序列的 PSNR 逐帧比较情况。由表 1 可知, 6 种

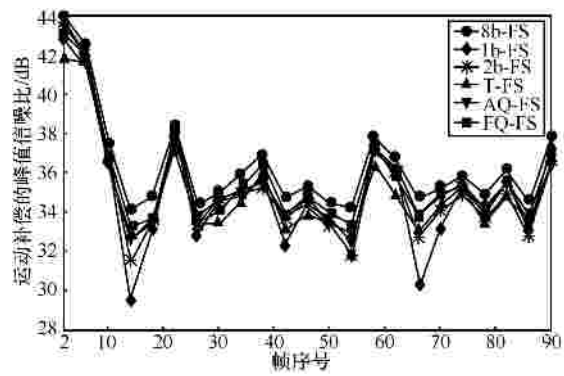
算法中, 1b-FS 的运动估计性能最差, 平均 PSNR 较 8b-FS 低 1.01 dB ;其次是 T-FS 算法, 平均 PSNR 比 8b-FS 降低 0.75 dB ;再次为 2b-FS 算法, 平均



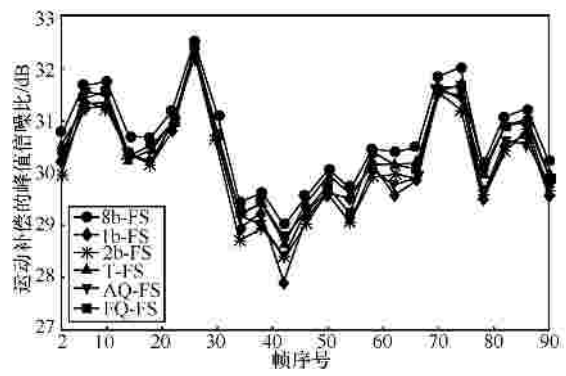
(a) Football 序列的重构 PSNR 对比



(b) Tennis 序列的重构 PSNR 对比



(c) Silent 序列的重构 PSNR 对比



(d) Paris 序列的重构 PSNR 对比

图 4 Football、Tennis、Silent 和 Paris 序列的逐帧 PSNR 比较

表 1 不同视频序列 Y 分量采用 6 种运动估计方法所得到的平均 PSNR 比较

序列名称	8b-FS	1b-FS	2b-FS	T-FS	AQ-FS	FQ-FS
Football	27.46	24.79	25.85	25.75	26.14	26.58
Flower	26.03	25.67	25.84	25.83	25.86	25.93
Mobile	23.96	23.52	23.61	23.78	23.62	23.80
Paris	30.70	30.17	30.17	30.45	30.27	30.49
Tennis	28.91	27.28	27.95	28.19	28.10	28.25
Silent	36.27	34.80	35.12	34.98	35.40	35.59
Stefan	25.90	25.27	25.44	25.70	25.49	25.70
Soccer	29.61	28.22	28.32	28.01	28.57	28.84
Tempete	27.75	27.09	27.29	27.47	27.36	27.58
Akiyo	42.67	41.93	42.23	41.84	42.33	42.11
Bus	24.87	23.80	24.14	24.49	24.31	24.50
Sign Irene	33.56	32.35	32.59	32.57	32.84	33.03
Husky	19.70	19.43	19.46	19.61	19.47	19.61
Salesman	39.41	39.19	39.25	38.78	39.18	39.27
City	30.86	30.07	30.29	28.66	30.40	30.41
Harbour	28.10	26.77	27.45	27.00	27.48	27.70

PSNR 较 8b-FS 下降 0.69 dB；然后是 AQ-FS，平均 PSNR 比 8b-FS 低 0.58 dB；最后，FQ-FS 的平均 PSNR 比 8b-FS 低 0.41 dB。而且，从图 4 能发现，本文算法在每帧上表现出的性能也较其他 4 种快速算法更加稳定。

6.2 运动估计精度

为了验证 4.1 节提出的比特深度转换误差—运动向量精度预测算法，本文比较了 FQ-FS、 d 像素精度的 8b-FS 和 $(d-1)$ 像素精度的 8b-FS 的运动估计结果，并用整像素精度的 8b-FS 的运动估计结果作为参考值。图 5 给出了上述 4 种算法在 Soccer 和 Tennis 序列上进行实验的 PSNR。由图 5 可见，本文算法的 PSNR 几乎处处高于 d 像素精度的 8b-FS，并且在绝大多数情况下，低于 $(d-1)$ 像素精度的 8b-FS。这说明本文预测出的向量精度下界 d 较为准确，它与真实下界的平均误差不超过 1 个像素。所以，可以将下界 d

和计算复杂度作为评价 2 bit 深度像素运动估计算法效率的指标，为用户或者编码器提供性能参考。

6.3 计算量的比较

6 种算法的计算量可分为 2 个阶段来统计，即比特深度转换阶段和运动估计阶段。表 2 以 $W = 32$ 、 $N = 16$ 为例，给出了各种算法在不同阶段所需的计算量。从中可见，本文算法在比特深度转换阶段的计算量明显低于除位截断方法以外的 3 种低比特深度像素的运动估计算法；在运动估计阶段，本文采用了 BTSAD 匹配误差度量准则，由于它可用查表指令完成，它的计算量不会高于 1b-FS、2b-FS 和 AQ-FS 所采用准则的计算量。所以总体来说，本文方法的计算量低于 1 bit 及现有 2 bit 深度像素的搜索，并且显著低于 8 bit 深度像素的运动估计算法。因为这些指标依赖于具体的硬件设计和所采用的指令集，已超出本文的讨论范围，所以

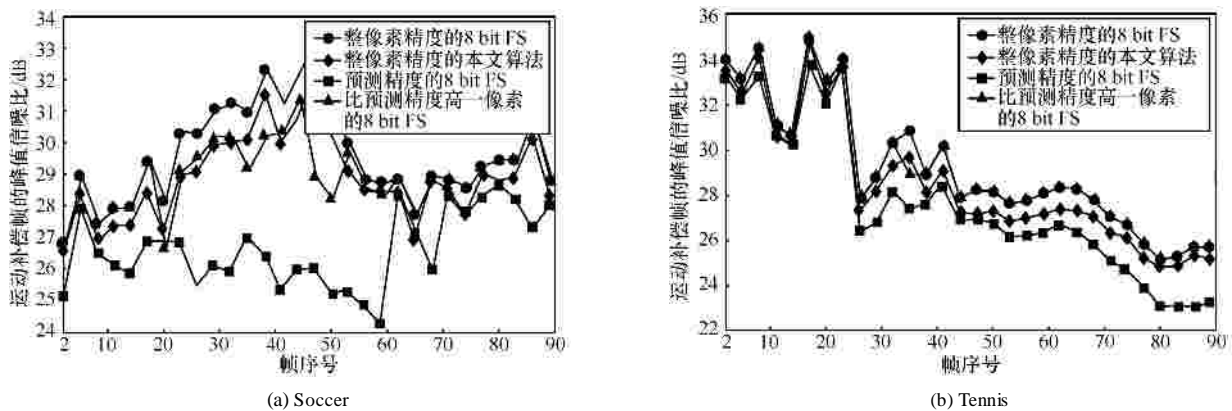


图 5 FQ-FS 的运动向量精度预测结果

表 2 5 种快速运动估计方法对每个像素的平均计算次数

阶段	操作	8b-FS	1b-FS	2b-FS	T-FS	AQ-FS	FQ-FS
比特深度转换阶段	加法	0	24	2.843 8	0	1.5	1.250 7
	乘法	0	1	1.062 5	0	$5.859 4 \times 10^{-3}$	0.500 7
	比较	0	1	3	0	7.5	3.001 3
	取绝对值	0	0	0	0	0.5	$4.932 2 \times 10^{-6}$
	开平方	0	0	0	0	0	$4.932 2 \times 10^{-6}$
运动估计阶段	位运算	0	0	0	2	0	0
	8 bit 定点整数加法	2 048	0	0	0	0	0
	2 bit 整数加法	0	0	1 024	1 024	1 024	1 024
	1 bit 整数加法	0	1 024	0	0	0	0
	取绝对值	1 024	0	0	0	0	0
	位运算	0	1 024	3 072	0	3 072	0
查表运算	0	0	0	1 024	0	1 024	

在此不予定量比较,进一步了解相关介绍可参考文献[29,32]。

6.4 讨论和进一步的工作

上述实验已验证了本文算法的有效性。事实上,该算法还有进一步扩展的可能。

首先,4.1节所提出的比特深度转换误差—运动向量精度分析不仅可用来评价本文算法的预测精度,也能用于预测其他 2 bit 深度像素运动估计方法的向量精度,只要给出对应方法的重构误差即可。在实际应用中,可认为对于统计特性相似的视频,给定的运动估计算法均能达到近似的精度,从而不必对每个序列的每一帧都计算其向量精度。此外,为了克服该方法对于比特深度转换方法的依赖,可假设比特深度转换采用的是均匀量化,其平均误差为 32。这样,当对某一视频序列进行 2 bit 深度像素运动估计时,就能粗略估计所能达到的向量精度,进而根据具体应用需求,决定是否选用低比特深度像素的估计算法。

其次,本文第 5 节的实验采用了全搜索的运动估计策略,其目的是排除不同搜索策略对各个算法的影响,以致不利于性能对比。实际上,本文算法既可单独使用,也可作为初始估计,应用于其他快速算法或复杂度可分级的运动估计中。

最后,特殊的直方图分布对本文算法会造成一定干扰。可根据直方图特点,自适应地调整参数 α 和 β ,进一步提高算法对各种类型视频的估计性能,这也是作者下一步的工作所在。

7 结束语

本文利用量化理论和模糊隶属度函数方法,提出了一种基于模糊量化和 2 bit 深度像素的运动估计方法。将 8 bit 深度像素的降采样问题转变为均方误差意义下的最优量化,并用直方图均衡和均匀量化给出一种快速求解方法;隶属度函数可以有效减少设备噪声等诸多因素造成的区间分划错误,改善比特深度转换质量,提高搜索效率。为预测本文算法的向量精度,提出了基于自相关函数的比特深度转换误差—运动向量精度分析方法,能较为准确地给出向量精度的一个下界,并且该下界与真实下界的平均误差不超过 1 个像素。大量实验表明,本文算法性能稳定,对于多种类型的视频序列,尤其是场景细节和物体运动比较复杂者,本文算法始终能保持较高的估计质量,运动补偿的平均 PSNR 较之传

统 2 bit 深度像素的运动估计提高 0.27 dB。

另外,本文算法还能应用于其他快速算法或复杂度可分级的运动估计中。模糊机制的采用和运动向量精度分析方法,对于其他 8 bit 深度像素的搜索算法亦具有一定的理论意义。

参考文献:

- [1] ISO/IEC JTC1/SC29/WG11. SVC Requirements Specified by MPEG: JVT-N026[S]. 2005.
- [2] SULLIVAN G J, OHM J R, ORTEGA A, *et al.* Future of video coding and transmission[J]. IEEE Signal Processing Magazine, 2003(6): 76-82.
- [3] 高文,赵德斌,马思伟.数字视频编码技术原理[M].北京:科学出版社,2010.
GAO W, ZHAO D B, MA S W. Principles of Digital Video Coding Techniques[M]. Beijing: Science Press, 2010.
- [4] 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.
- [5] CHEN Z, XU J, HE Y, *et al.* Fast integer-pel and fractional-pel motion estimation for H.264/AVC[J]. Journal of Visual Communication and Image Representation, 2006, 17(2):264-290.
- [6] PO L M, MA W C. A novel four-step search algorithm for fast block motion estimation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1996, 6(3):313-317.
- [7] LIU L K, FEIG E. A block-based gradient descent search algorithm for block motion estimation in video coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1996, 6(4):419-423.
- [8] ZHU S, 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.
- [9] CHEUNG C H, PO L M. Novel cross-diamond-hexagonal search algorithms for fast block motion estimation[J]. IEEE Transactions on Multimedia, 2005, 7(1):16-22.
- [10] LOU C C, LEE S W, JAY KUO C C. Adaptive motion search range prediction for video encoding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2010, 20(12):1903-1908.
- [11] LIN W, PANUSOPONE K, BAYLON D M, *et al.* A fast sub-pixel motion estimation algorithm for H.264/AVC video coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2011, 21(2):237-243.
- [12] CHEUNG C K, PO L M. Normalized partial distortion search algorithm for block motion estimation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2000, 10(3):417-422.
- [13] 宋传鸣,王相海.小波域视频运动估计研究进展[J].计算机学报, 2005, 28(10):1716-1727.
SONG C M, WANG X H. Research progress in video motion estimation techniques in the wavelet domain[J]. Chinese Journal of Computers, 2005, 28(10):1716-1727.
- [14] ZHANG R, COMER M L. Rate distortion performance of pyramid and subband motion compensation based on quantization theory[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2010, 20(12):1876-1881.
- [15] WANG Y, WANG Y, KURODA H. A globally adaptive pix-

- el-decimation algorithm for block-motion estimation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2000, 10(6):1006-1011.
- [16] KIM J, CHOI T. A fast full-search motion-estimation algorithm using representative pixels and adaptive matching scan[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2000, 10(7):1040-1048.
- [17] CHEN M J, CHEN L G, CHIUH T D, *et al.* A new block-matching criterion for motion estimation and its implementation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1995,5(3): 231-236.
- [18] NATARAJAN B, BHASKARAN V, KONSTANTINIDES K. Low-complexity block-based motion estimation via one-bit transforms[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1997, 7(4):702-706.
- [19] ERTÜRK S. Multiplication-free one-bit transform for low-complexity block-based motion estimation[J]. IEEE Signal Processing Letters, 2007, 14(2):109-112.
- [20] LEE H, JEONG J. Early termination scheme for binary motion estimation[J]. IEEE Transactions on Consumer Electronics, 2007, 53(4): 1682-1686.
- [21] WU C B, YAO C Y, LIU B D, *et al.* DCT-based adaptive thresholding algorithm for binary motion estimation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2005, 15(5):694-703.
- [22] CELEBI A, AKBULUT O, URHAN O, *et al.* An all binary sub-pixel motion estimation approach and its hardware architecture[J]. IEEE Transactions on Consumer Electronics, 2008, 54(4):1928-1937.
- [23] LUO J H, WANG C N, CHIANG T. A novel all-binary motion estimation (ABME) with optimized hardware architectures[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2002, 12(8):700-712.
- [24] URHAN O, ERTÜRK S. Constrained one-bit transform for low complexity block motion estimation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2007, 17(4):478-482.
- [25] LEE X, ZHANG Y Q. A fast hierarchical motion-compensation scheme for video coding using block feature matching[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1996,6(6):627-635.
- [26] WONG P H W, AU O C. Modified one-bit transform for motion estimation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1999, 9(7):1020-1024.
- [27] SONG X, CHIANG T, LEE X, *et al.* New fast binary pyramid motion estimation for MPEG2 and HDTV encoding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2000, 10(7):1015-1028.
- [28] LEE S, CHAE S I. Motion estimation algorithm using low resolution quantisation[J]. Electronics Letters, 1996, 32(7):647-648.
- [29] LEE S, KIM J M, CHAE S I. New motion estimation algorithm using adaptively quantized low bit-resolution image and its VLSI architecture for MPEG2 video encoding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1998, 8(6):734-744.
- [30] DEMIR B, ERTÜRK S. Block motion estimation using adaptive modified two-bit transform[J]. IET Image Processing, 2007, 1(2):215-222.
- [31] BAEK Y, OH H S, LEE H K. An efficient block-matching criterion for motion estimation and its VLSI implementation[J]. IEEE Transactions on Consumer Electronics, 1996, 42(4):885-892.
- [32] HE Z L, TSUI C Y, CHAN K K, *et al.* Low-power VLSI design for motion estimation using adaptive pixel truncation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2000, 10(5):669-678.
- [33] 张武健, 邱晓海, 周润德等. 一种新的使用两种比特分辨率图像的块匹配运动估计算法及其低功耗 VLSI 结构[J]. 电子学报, 2001, 29(6):860-864.
- ZHANG W J, QIU X H, ZHOU R D, *et al.* New block-matching motion estimation algorithm using dual-bit-resolution image and its low-power VLSI architecture[J]. Acta Electronica Sinica, 2001, 29(6): 860-864.
- [34] PATRAS I, HENDRIKS E A, LAGENDIJK R L. Probabilistic confidence measures for block matching motion estimation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2007,17(8): 988-995.
- [35] MAX J. Quantizing for minimum distortion[J]. IRE Transactions on Information Theory, 1960, 6(1):7-12.
- [36] LLOYD S P. Least squares quantization in PCM[J]. IEEE Transactions on Information Theory, 1982, 28(2):129-137.
- [37] RIBAS-CORBERA J, NEUHOFF D L. Optimizing motion-vector accuracy in block-based video coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2001, 11(4):497-511.

作者简介:



宋传鸣 (1980-), 男, 辽宁沈阳人, 博士, 辽宁师范大学讲师, 主要研究方向为图像及视频编码。



郭延文 (1980-), 男, 山东济宁人, 博士, 南京大学副教授, 主要研究方向为计算机图形学、图像与视频处理。



王相海 (1964-), 男, 吉林汪清人, 博士, 辽宁师范大学教授、博士生导师, 主要研究方向为计算机图形学及多媒体信息处理。



刘丹 (1981-), 女, 辽宁沈阳人, 硕士, 辽宁师范大学讲师, 主要研究方向为图像及视频编码。