

基于 H. 264 的多参考帧快速选择算法

张淑芳, 李 华

(天津大学电子信息工程学院, 天津 300072)

摘 要: H. 264 采用了多参考帧技术, 大大提高了预测精度, 改善了视频图像质量, 然而这种特性同时也带来了很高的计算复杂度. 为了降低 H. 264 编码器的计算复杂度, 本文提出了一种多参考帧快速选择算法. 该算法充分考虑多参考帧中运动矢量的相关性以及中心偏置特性, 通过对待选参考帧进行预选择, 有效减少了需要进行全搜索的参考帧个数, 提高了编码速度. 实验结果表明, 该算法在保持图像质量和码率变化很小的前提下, 运动估计模块编码速度提高了 30.91%.

关键词: H. 264; 多参考帧; 运动矢量; 中心偏置特性

中图分类号: TN911.73 **文献标识码:** A **文章编号:** 0372-2112(2009)01-0062-05

A Multi-Reference Frames Fast Selection Algorithm for H. 264

ZHANG Shu-fang, LI Hua

(School of Electronic and Information Engineering, Tianjin University, Tianjin 300072, China)

Abstract: The multi reference frames technique is employed in H. 264, which can improve prediction precision and meliorate video pictures quality greatly. However, the feature calls for higher computational complexity simultaneity. In order to reduce the complexity of H. 264 encoder, a fast multi reference frames selection algorithm is presented in this paper. By considering the correlation and center biased characteristic of motion vectors in multi reference frames adequately, the proposed algorithm makes selection for multiple references in advance, thereby it can effectively decrease the numbers of reference frames making full search and improve the coding efficiency. Experimental results show that the proposed methods can improve the speed of motion estimation module by 32.14% on average with less performance degradation and bit rate increase.

Key words: H. 264; multi reference frames; motion vector; center biased characteristic

1 引言

新一代视频压缩编码标准 H. 264/AVC^[1] 允许采用多个已编码帧作为运动补偿预测的参考, 并且不同的宏块分块可以使用不同的参考帧. 在 JVT 参考模型 JM10.2^[2] 中引入了拉格朗日率失真最优策略, 采用穷尽搜索法对 7 种帧间预测模式 (16×16 , 16×8 , 8×16 , 8×8 , 8×4 , 4×8 , 4×4) 的每一种, 在所有可能的参考帧中选择拉格朗日代价函数最小的参考帧和其相应的运动矢量. 在 16×16 , 16×8 , 8×16 , 8×8 模式中, 每个子块分别在不同参考帧进行运动估计; 而 8×8 块中所有 8×4 , 4×8 , 4×4 模式的子块首先在同一个参考帧进行运动估计并累加 J_{MOTION} , 然后在不同参考帧之间进行比较. 这种对多参考帧的全搜索方法使得帧间预测的可选范围更大, 预测也更为精确, 获得了很好的编码效果. 但同时该技术会显著增加编码器的计算复杂度, 其计算复杂

度随参考帧数目呈线性增加, 参考帧数越多, 需要耗费的编码时间就越长, 给处理器的计算能力和存储空间带来巨大的挑战.

对于多参考帧的快速选择问题, 国内外也有不少研究成果^[3-5], 如 Chung^[3] 对当前帧进行采样得到 $1/4$ 低分辨率图像, 然后参考不同帧对该图像作估计, 并根据估计结果寻找最佳参考帧, 认为该最佳帧就是原始帧的最佳参考帧, 故在此参考帧中对原始帧作运动估计, 获得原始帧的运动估计结果. 因采样图像只有原始帧的 $1/4$, 故选择最佳参考帧的运动估计计算量显著下降, 但它没有考虑视频序列中运动的相关性. Ting^[4] 验证了多参考帧下运动矢量的中心偏置性, 即多参考帧下运动矢量大都位于搜索窗口 $w = \pm 2$ 的范围内 (约 78%).

本文充分考虑多参考帧中运动矢量的相关性以及中心偏置特性, 提出了一种多参考帧快速选择算法.

2 多参考帧分析

为了了解不同参考帧在实际序列中被选用的情况, 在参考软件 JM10.2 上对不同的视频序列进行了测试, 表 1 为参考帧个数为 5、QP 为 32 时各序列中参考帧被选中的概率。其中 $REF(0)$ 表示与当前待编码帧最相邻的前一参考帧, $REF(1)$ 表示前一参考帧的前一帧, 以

表 1 参考帧概率统计结果

格式	序列	$REF(0)$ 概率	$REF(1)$ 概率	$REF(2)$ 概率	$REF(3)$ 概率	$REF(4)$ 概率
QCIF	Bus	86.72%	6.30%	3.98%	1.60%	1.40%
	Foreman	88.59%	5.57%	3.84%	1.04%	0.97%
	Silent	95.80%	1.75%	1.32%	0.46%	0.67%
	M&D	94.00%	2.91%	1.98%	0.47%	0.63%
CIF	Foreman	90.44%	4.65%	2.90%	0.90%	1.10%
	Container	93.07%	3.04%	1.79%	1.17%	0.93%
	Mobile	74.08%	8.48%	8.87%	4.61%	3.96%
平均值		88.96%	4.67%	3.53%	1.46%	1.38%

上述结果是比较合理的, 因为通常情况下距离当前帧越近的参考帧其相关性越高, 因此它被选择为最优参考帧的概率就越大。但在一些特殊情况下, 距离较远的参考帧被选择为最优参考帧的概率比较大, 例如在对象边缘遮挡和显露时, 物体做周期性运动, 或者在场景切换时, 待编码宏块在第一参考帧中可能无法找到最匹配的块, 多参考帧中的某一帧可能更符合当前宏块的需要。

由于从参考帧 $REF(0)$ 中获得的运动搜索结果对整个编码性能有着重要影响, 在进行多参考帧选择时, 对该参考帧应该进行大范围的搜索, 得到最优的运动矢量。相比较, 由于其余的参考帧被选中的概率很低, 而且每增加一个参考帧就会增加很大一部分计算量, 所以应利用一定的判定准则, 尽量减少对它们的搜索, 降低视频编码计算复杂度。

3 多参考帧运动矢量特性

3.1 多参考帧运动矢量相关性

在视频序列图像中, 物体运动是连续的, 所以描述物体运动的宏块的运动矢量在空间和时间上具有相关性。物体运动的连续性同样也导致多参考帧中运动矢量域的相关性。如图 1 所示, 沿着运动轨迹进行追踪, 就能得到相邻帧运动矢量之间的关联性^[5], 即

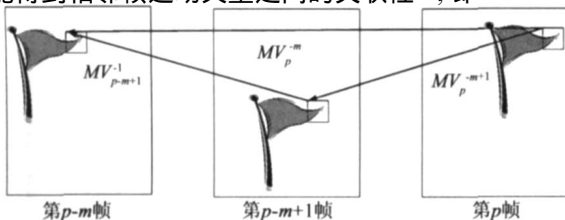


图1 运动矢量相关性示意图

此类推。

从表 1 可看出, 一般情况下, 在所有参考帧中距离当前帧越近的参考帧被选择为最优参考帧的概率就越高, 尤其是第一参考帧 $REF(0)$ 。 $REF(0)$ 成为最优参考帧的概率平均为 88.96%, 最高的序列达到了 95.80%, 而其余参考帧被选中的概率远远低于 $REF(0)$, 平均概率仅为 11.04%。

$$MV_p^{-m} = MV_p^{-m+1} + MV_{p-m+1}^{-1} \quad (1)$$

其中, MV_p^{-m} 表示从第 p 帧指向 $p-m$ 帧的运动矢量。采用运动矢量相关性进行多参考帧预测, 需将每帧相对于前一帧的运动矢量都保存下来, 每个宏块需要保存 7 种块类型的运动矢量, 这对于多参考帧情形下的快速搜索很有帮助, 可有效定位编码块在每一帧的预测起点, 从而减少搜索点数节约运动估计时间。

运动矢量 MV_p^{-m+1} 在 $p-m+1$ 帧指向的位置不总是和 p 帧中被匹配块大小相等块的起始位置, 这时需要在 $p-m+1$ 帧中和匹配块相邻的四个块中找一个主导块, 即重叠面积最大的块, 作为匹配块^[6]。如图 2 所示, p 帧中 8×8 的块 B_p 在 $p-m+1$ 帧中搜索到的匹配块为 B_{p-m+1} , 但 B_{p-m+1} 所在位置不是 $p-m+1$ 帧某个 8×8 的块的起始位置, 需要把 B_{p-m+1} 相邻的四个 8×8 块中和 B_{p-m+1} 重叠面积最大的块 B'_{p-m+1} 作为 B_p 的匹配块。

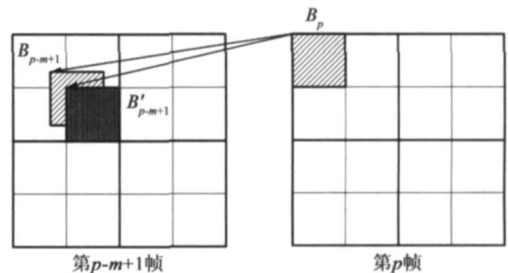


图2 主导块选择示意图

为了证实运动矢量间具有这种很强的相关性, 用合成的运动矢量 MV_p^{-m} 和利用参考软件进行搜索得到的真实运动矢量 MV_p^{-m} 相比, 并用 MCE 来表示运动矢量合成误差, 即

$$MCE = |MV_p^{-m} - MV_p^{-m}|$$

$$= |MV_p^m(x) - MV_p^{m-1}(x)| + |MV_p^m(y) - MV_p^{m-1}(y)| \quad (2)$$

文献[5]证明了90%左右合成运动矢量的误差都在1个像素之内。

由于参考帧 $REF(0)$ 的搜索结果对编码性能影响极大,在进行多参考帧选择时,需要对 $REF(0)$ 的所有参考位置进行搜索,得到该参考帧中最优的运动矢量 MV_p^0 , 并求得其相应的拉格朗日代价 $J_{MOTION}^{REF(0)}$. 令多参考帧最优运动矢量的代价函数 $J_{MOTION}^{REF(best)} = J_{MOTION}^{REF(0)}$, 最优参考帧 $REF(best) = REF(0)$. 然后根据运动矢量合成公式求得待编码帧指向参考帧 $REF(m)$ ($m = 1, 2, \dots, N-1$) 的运动矢量 MV_p^m , N 为参考帧的个数, 由于该运动矢量 MV_p^m 是 $REF(m-1)$ 中最优运动矢量的近似值, 还需要进行进一步修正, 具体修正方法见 3.2, 以得到最优运动矢量 $(MV_p^m)''$, 相应的拉格朗日代价 $J_{MOTION}^{REF(m)}$ 的计算如式(3)所示:

$$J_{MOTION}^{REF(m)} = J_{MOTION}(m, rcf | \lambda_{MOTION}) = SAD(s, r(rcf, m)) + \lambda_{MOTION}(R(m-p) + R(rcf)) \quad (3)$$

其中, s 为当前块的像素值, rcf 为要选择的参考帧, $m = (m_x, m_y)^T$ 为参考帧 rcf 的运动矢量, $p = (p_x, p_y)^T$ 为利用相邻的运动矢量得到的预测运动矢量, $r(rcf, m)$ 为参考块的像素值, $SAD(s, r(rcf, m))$ 为当前块与参考块像素差值的绝对值和, 通过式(4)进行计算, 码率 R 包括 $R(m-p)$ 和 $R(rcf)$ 两部分, $R(m-p)$ 为运动矢量与其预测值的差值编码所需比特, $R(rcf)$ 为编码参考帧的比特数。

$$SAD(s, r(rcf, m)) = \sum_{x=1}^{n_1} \sum_{y=1}^{n_2} |s(x, y) - r(x - m_x, y - m_y)| \quad (4)$$

根据不同的预测块模式, n_1, n_2 取 16, 8 和 4。

比较 $J_{MOTION}^{REF(m)}$ 和 $J_{MOTION}^{REF(best)}$, 如果 $J_{MOTION}^{REF(m)} < J_{MOTION}^{REF(best)}$, 则不需要对该参考帧进行进一步搜索, 令 $m = m + 1$; 否则以 $(MV_p^m)''$ 作为初始运动矢量对 $REF(m)$ 的所有参考位置进行搜索, 并计算出拉格朗日代价 $J_{MOTION}^{REF(m)}$, 令 $J_{MOTION}^{REF(best)} = J_{MOTION}^{REF(m)}$. 以此类推直到所有参考帧都完成为止。

3.2 多参考帧运动矢量中心偏置特性

所谓中心偏置特性, 是指最佳匹配点在搜索区域开始中心点附近分布的特性. 通常视频序列图像中物体运动范围和运动速度都不大, 所以宏块的运动矢量分布在中心位置上的概率较大^[7], 图像序列“Miss America”和“Table Tennis”分别有 83% 和 81% 的运动位移分布在 5×5 的中心区域窗内, 分别有 52% 和 61% 的运动位移分布在 3×3 的中心区域窗内. 快速块匹配算法

(BMA) 如: 新三步搜索法(N3SS)、四步搜索法(4SS)等都采用了运动矢量的中心偏置特性。

在多参考帧模式下, 运动矢量同样也具有中心偏置性, 即多参考帧下运动矢量大都位于搜索区开始中心点的附近. 文献[4]验证了多参考帧下运动矢量的中心偏置性, 在多参考帧下运动矢量大都位于搜索窗口 $w = \pm 2$ 的范围内(约 78%)。

本文利用运动矢量中心偏置特性对采用运动矢量合成法得到的 $REF(m)$ 参考帧的近似运动矢量 MV_p^m 进行修正. 由于运动矢量大都位于搜索窗口 $w = \pm 3$ 的范围内, 因此本文把运动矢量 MV_p^m 作为初始预测运动矢量, 采用快速运动估计算法, 对 $w = \pm 3$ 的小区域进行整像素搜索来得到参考帧 $REF(m)$ 修正后的运动矢量 $(MV_p^m)''$ 。

4 算法具体实现

对从 16×16 到 4×4 的每一种帧间编码模式, 都可采用多参考帧快速选择算法来进行参考帧的选择, 从而大大减少多参考帧选择的计算量。

算法的具体实现步骤为:

步骤 1 以预测运动矢量(MVP)指向的位置作为搜索点的中心, 对第一参考帧 $REF(0)$ 的所有参考位置进行搜索, 得到 $REF(0)$ 中最优的运动矢量 MV_p^0 及其相应的拉格朗日代价 $J_{MOTION}^{REF(0)}$, 令:

$$J_{MOTION}^{REF(best)} = J_{MOTION}^{REF(0)}, \quad REF(best) = REF(0)$$

步骤 2 根据运动矢量合成式(1)求得待编码帧指向参考帧 $REF(m)$ ($m = 1, 2, \dots, N-1$) 的运动矢量 MV_p^m 。

步骤 3 利用运动矢量中心偏置特性, 把运动矢量 MV_p^m 作为初始预测运动矢量, 采用快速运动估计算法, 对 $w = \pm 3$ 的小区域进行搜索, 得到参考帧 $REF(m)$ 修正后的运动矢量 $(MV_p^m)''$, 并利用式(3)求得相应的拉格朗日代价 $J_{MOTION}^{REF(m)}$ 。

步骤 4 比较 $J_{MOTION}^{REF(best)}$ 和 $J_{MOTION}^{REF(m)}$, 如果 $J_{MOTION}^{REF(m)} < J_{MOTION}^{REF(best)}$, 进行步骤 5, 否则进行步骤 6。

步骤 5 把修正过的运动矢量 $(MV_p^m)''$ 作为初始运动矢量对 $REF(m)$ 的所有参考位置进行搜索, 得到 $J_{MOTION}^{REF(m)}$, 并令 $J_{MOTION}^{REF(best)} = J_{MOTION}^{REF(m)}$, $REF(best) = REF(m)$ 。

步骤 6 令 $m = m + 1$, 如果 $m < N$, 跳到步骤 4, 否则该算法结束。

5 实验结果

采用 JVT 的 JM10.2 作为平台, 实验条件为:

(1) MV 的搜索范围: QCIF 为 ± 16 个像素, CIF 为 ± 32 个像素;

- (2) 采用了 RD 优化和滤波;
- (3) 参考图像为 5 帧;
- (4) 编码结构为 IPPP;
- (5) 对每一个序列编码帧数为 100 帧;
- (6) 运动估计采用 FME 算法.

表 2 为量化参数 QP 为 28、32、36 和 40 的情况下, 本文快速算法与 JM10.2 参考算法的性能比较结果, 其中负值表示减少, 正值表示增加. 从表 2 可以看出, 与原始的采用穷尽搜索的拉格朗日率失真最优模式选择方

表 2 本文快速算法与 JM10.2 参考算法性能比较

格式	序列	ISNR			比特率			编码速度提高	
		参考算法 (dB)	本文算法 (dB)	变化值 (dB)	参考算法 (kbit/s)	本文算法 (kbit/s)	变化率 (%)	运动估计模块 (%)	整个编码器 (%)
QCIF	Bus	30.80	30.82	0.02	201.32	209.27	3.95	41.94	10.81
	Foreman	32.97	33.03	0.06	59.94	60.28	0.57	35.72	8.68
	Silent	32.94	32.93	-0.01	49.38	49.99	1.23	18.41	5.60
	M&D	33.46	33.51	0.05	44.38	44.82	0.99	16.25	5.16
CIF	Foreman	34.39	34.47	0.08	113.03	115.44	2.14	42.11	11.55
	Container	33.52	33.52	0	39.88	40.25	0.93	18.21	4.79
	Mobile	30.84	30.79	-0.05	403.08	416.02	3.21	43.72	13.50
平均值		—	—	0.02	—	—	1.86	30.91	8.58

对于 Mobile, Bus 和 Foreman 这类运动比较剧烈的序列, 码率升高较多, 编码速度提高幅度相对也较大; 而对于 Silent, M&D 和 Container 运动量较小的序列, 编码速度提高的相对较少. 这是由于运动量较小的序列, 物体位置在相继的几个参考帧中相差不大, 因此运用本文算法得到的相继几个参考帧的修正运动矢量 (MV_p^m) 以及相应的拉格朗日代价 $J_{MOTION}^{REF(m)}$ 都比较相似. 在 H. 264 中, 为了使得编解码匹配, 在编码端采用重构图像代替原始图像作为参考帧, 在参考帧中就可能存在噪声, 并且由于摄像机和其它因素对原始图像引起的噪声也会被重构图像复制, 这些噪声就会对参考帧的选择产生很大的影响, 即使在图像的静止区域, 由于噪声的影响可能使得离待编码图像比较远的参考帧的 $J_{MOTION}^{REF(m)}$ 反而小, 从而在比较以前的参考帧中找到最匹配的块. 在进行编码时就搜索了较多的参考帧, 从而影响了编码速度, 使得速度提升较小. 而对于运动比较剧烈的序列, 物体位置在相继的几个参考帧中相差较大, 计算出的各参考帧的拉格朗日代价 $J_{MOTION}^{REF(m)}$ 比较

法相比, 采用本文所提快速多参考帧选择算法可使编码时间大幅度减少, 其中运动估计模块编码速度提高了 30.91%, 编码器总的编码速度提高了 8.58%, 同时平均 PSNR 上升了 0.02dB, 码率仅仅增长了 1.86%. 原因在于, 该算法充分利用了运动矢量相关性和运动矢量分布的中心偏置特性, 对待选参考帧进行预选择, 有效减少了需要进行全搜索的参考帧个数, 在图像质量和比特率变化很小的情况下, 可以较大幅度提高编码速度.

分散, 这时噪声对其产生的影响可以忽略不计, 利用本文算法可以很好地预测出所选用的参考帧, 所以编码速度提高较多.

图 3 为本文快速算法和 JM10.2 参考算法部分视频序列的 R-D 性能比较图. 从图中可以更直观的看出, 本文算法和参考算法的率失真性能相差很小, 对于运动不剧烈的视频序列 (如 M&D 序列), 本文算法的率失真曲线和参考算法的率失真曲线几乎重合, 对于运动较复杂的视频序列 (如 Mobile 序列), 本文算法的率失真性能略低于参考算法, 但性能损失很小. 原因在于本文算法利用了运动矢量的中心偏置特性, 在对第 $REF(0)$ 参考帧以外的其它参考帧进行搜索时, 用修正过的最优运动矢量 (MV_p^m) 作为初始运动矢量, 而不是采用待编码帧中相邻块的运动矢量预测得到的预测运动矢量 MVP , 这样在参考帧中可搜索到更匹配的块, 使得预测更准确. 因此, 对视频图像的失真和比特率产生的影响很小.

6 结论

本文考虑到多参考帧中运动矢量的相关性和中心偏置特性, 提出了一种多参考帧快速选择算法. 该算法能有效减少需要进行全搜索的参考帧个数, 大大降低编码器的计算复杂度. 在对图像质量损失很小, 比特率变化不大的条件下, 采用该算法, 运动估计模块编码速度提高了 30.91%, 编码器总的编码速度提高了 8.58%. 并且该算法的计算量小, 简单可行, 适合于实际

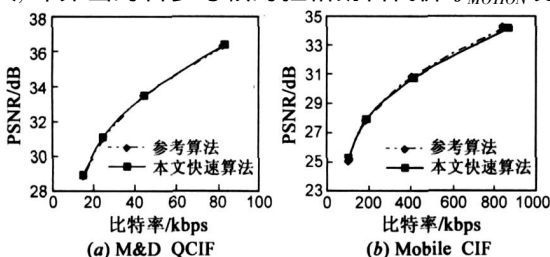


图 3 视频序列 R-D 性能比较

应用.

参考文献:

- [1] Joint Video Team(JVT) of ISO/IEC MPEG & ITU-T VCEG. Draft ITU-T Recommendation and Final Draft international Standard of Joint Video Specification [S]. (ITU-T Rec. H. 264//IEC 14496-10 AVC), Doc. G050r1, Mar. 2003
- [2] Joint Video Team(JVT) Reference Software[OL]. <http://bs.hhi.de/~suehring/tml/download/>
- [3] H Chung, D Romacho, A Ortega. Fast long term motion estimation for H. 264 using multiresolution search[A]. IEEE International Conference on Image Processing[C]. Barcelona: Institute of Electrical and Electronics Engineers Computer Society, 2003. 905- 908.
- [4] C W Ting, L M Po, C H Cheung. Center biased frame selection algorithms for fast multiframe motion estimation in H. 264 [A]. International Conference on Neural Networks and Signal Processing[C]. Nanjing: CA SS, vol. 2, 2003. 1258- 1261.
- [5] Y Su, M T Sun. Fast multiple reference frame motion estimation for H. 264/AVC[J]. IEEE Transaction on Circuits and System for Video Technology, 2006, 16(3): 447- 452
- [6] J Youn, M T Sun. A fast motion vector composition method for temporal transcoding [A]. IEEE International Symposium on Circuits and Systems[C]. Orlando, FL USA: IEEE. 1999. IV- 243 IV- 246.

- [7] 何振亚, 邹采荣, 骆立俊. 一种用于视频编码的块运动估计算法——块特征匹配预测搜索算法[J]. 电子学报, 1999, 27(1): 32- 36.

He Zhenya, Zou Cairong, Luo Lijun. A new block motion estimation scheme for video coding—block feature matching algorithm using predictive search[J]. Acta Electronica Sinica, 1999, 27(1): 32- 36. (in Chinese)

作者简介:



张淑芳 女, 1979 年 2 月出生于山西省襄汾县. 博士, 天津大学电子信息工程学院讲师. 主要研究领域为视频编解码、计算机视觉、图像处理.

E-mail: shufangzhang@tju.edu.cn



李 华 女, 1942 年 5 月出生于天津市. 天津大学电子信息工程学院教授、博士生导师. 主要研究领域为数字图像处理、数字高清晰度电视、视频编解码.