

# 基于时空相关的 H. 264 多参考帧快速选择算法

仲伟波 孟艳茹

(江苏科技大学电子信息学院 镇江 212003)

**摘要** H. 264/AVC 视频编码标准采用的多参考帧技术改善了视频质量,但同时也增加了编码复杂度。为避免搜索多余的参考帧,提高 H. 264 编码器的编码效率,基于视频序列帧间具有较强时间相关性、相邻宏块具有较强时空相关性,提出了一种多参考帧选择算法。该算法根据视频序列的帧间时间相关程度确定参考帧个数;由相邻宏块的时空相关性自适应确定当前宏块的编码模式及其最佳参考帧。实验结果表明,与校验模型 JM16.2 相比,该算法在编码质量几乎不变的情况下,平均可节约 60% 的编码时间。

**关键词** H. 264/AVC, 多参考帧选择, 帧间时间相关, 宏块时空相关

**中图分类号** TP919.81 **文献标识码** A

## Fast Multiple Reference Frame Selection Algorithm for H. 264 Based on Temporal and Spatial Correlation

ZHONG Wei-bo MENG Yan-ru

(School of Electronics and Information, Jiangsu University of Science and Technology, Zhenjiang 212003, China)

**Abstract** The multi-reference frames technique is employed in H. 264/AVC, which ameliorates video quality greatly, but also increases the complexity of coding. In order to avoid searching redundant reference frame and increase the efficiency of the H. 264 encoder, a fast multi-reference frames selection algorithm was proposed according to the temporal correlation between inter frames, and the temporal and spatial correlation among the adjacent macro blocks. The number of reference frames was determined according to the degree of the temporal correlation between inter frames, and the coding mode and the best reference frame of the current macro block were determined by the degree of temporal and spatial correlation between its adjacent macro blocks adaptively. Experimental results show that the proposed algorithm can save about 60% of encoding time with the encoding quality almost unchanged in comparison with JM16.2.

**Keywords** H. 264/AVC, Multiple reference frame selection, Interframe temporal correlation, Macro block temporal and spatial correlation

### 1 引言

H. 264 作为新一代视频编码标准,采用了许多新的视频编码技术,与 MPEG-4 相比,压缩效率提高约 50%,但其计算复杂度也成倍增加<sup>[1-3]</sup>。为提高视频编码质量,H. 264 在帧间预测中采用了多参考帧技术,每个宏块在多个参考帧中进行运动估计与补偿,通过比较率失真决定最佳宏块模式和最佳参考帧,参考帧选择占用了大量编码时间。

目前已有多种多参考帧快速选择算法,如利用视频序列的空间相关性自适应确定运动估计的参考帧数目<sup>[4]</sup>;利用相邻宏块间的相关性和已搜索参考帧的运动预测信息排除不必要的参考帧<sup>[5]</sup>;根据多参考帧中运动矢量的相关性及中心偏置特性减少参考帧数<sup>[6]</sup>;把当前宏块分为 4 个 8×8 块,并在第一个参考帧内进行运动估计,若不满足第一帧提前判决,则对 4 个 8×8 块在其余参考帧进行运动估计,代价函数最小的参考帧作为最佳参考帧<sup>[7]</sup>;根据当前宏块及前一帧对应位置

宏块差值的大小来确定参考帧的数目<sup>[8]</sup>;根据序列的空间特性,利用已编码 2 个 p 帧中 Skip 编码模式的宏块数目来确定不同模式下的宏块参考帧数目<sup>[9]</sup>等。上述算法大多依据已编码宏块模式、运动矢量特性进行多参考帧快速选择,计算复杂度虽有降低,但仍存在改进的空间。本文根据视频序列固有的帧间时间强相关,以及相邻宏块间时空强相关的特性,提出一种多参考帧快速选择算法。根据视频序列的帧间时间相关程度,由已编码 5 帧中最佳参考帧的选择情况确定当前宏块的参考帧数目;利用相邻宏块的时空相关程度,根据相邻宏块的编码模式自适应地确定当前宏块编码模式及其最佳参考帧。实验结果表明,本算法与校验模型 JM16.2 相比,在视频质量和码率基本保持不变的情况下,可节省 60% 的编码时间。

### 2 多参考帧快速选择算法

#### 2.1 参考帧数的确定

同单参考帧相比,多参考帧技术可有效提高图像编码质

到稿日期:2012-07-19 返修日期:2012-10-26 本文受江苏高校优势学科建设工程项目,江苏省政府留学奖学金项目,江苏省自然科学基金项目(BK2009623)资助。

仲伟波(1975-),男,博士,副教授,主要研究方向为信号与信息处理、模式识别与智能系统,E-mail:717605600@qq.com;孟艳茹(1988-),女,硕士生,主要研究方向为视频编解码、视频传输。

量,然而使用多少个参考帧最合适,这完全依赖于视频序列的内容。另一方面,由视频序列的特性可知,运动愈缓慢的序列中相邻帧间的时间相关程度愈强,也即当前帧与其紧邻的前一帧的相似度愈高,其紧邻的前一帧作为当前帧的最佳参考帧的可能性愈大,即便对于运动较为剧烈的视频序列来讲,可能需要更多的帧数作为最佳参考帧的候选,但从视频时间相关性上来讲,候选最佳参考帧多为紧邻当前帧的图像。

校验模型采用全搜索方法,编码器在每个参考帧中对当前宏块的所有模式进行运动估计与补偿,并计算其代价函数,最终确定代价函数最小的参考帧为最佳参考帧,其中代价函数定义为:

$$J_{MOTION}(mv, ref | \lambda_{MOTION}) = SAD[s, r(ref, mv)] + \lambda_{MOTION} [R(mv - pred) + R(ref)] \quad (1)$$

式中,  $s$  为当前块像素值,  $ref$  为要选择的参考帧,  $mv$  为参考帧的运动矢量,  $pred$  为预测矢量,  $r(ref, mv)$  为参考块的像素值,  $SAD$  为当前块与参考像素差值的绝对值,  $\lambda_{MOTION}$  是 Lagrange 参数,  $R(mv - pred)$ 、 $R(ref)$  为运动矢量与其预测值差值编码的比特和参考帧的编码比特。对于视频序列,若设最大参考帧数为 5,不同运动程度的视频序列中最佳参考帧的分布情况如表 1 所列。

表 1 不同序列中最佳参考帧的分布

序列		第 1 帧	第 2 帧	第 3 帧	第 4 帧	第 5 帧
		/%	/%	/%	/%	/%
运动缓慢	news	95.1	2.4	1.6	0.6	0.3
	silent	96.8	1.8	0.9	0.2	0.3
	平均	96.0	2.1	1.2	0.4	0.3
运动一般	foreman	70.6	14.4	8.9	3.3	2.8
	waterfall	70.2	11.8	11.5	3.8	2.7
	平均	70.4	13.1	10.2	3.5	2.8
运动剧烈	mobile	52.9	11.6	12.9	11.5	11.1
	tempete	47.7	15.3	14.7	11.6	10.7
	平均	50.3	13.4	13.8	11.6	10.9

表 2 本文算法确定的不同运动程度视频序列候选最佳参考帧分布

序列		$P > 0.95\%$	$0.5 < P < 0.95\%$	$P < 0.5\%$	$P > 0.9\%$	$0.5 < P < 0.9\%$	$P < 0.5\%$
		第 1 帧	第 1、2、3 帧	紧邻的前 5 帧	第 1 帧	第 1、2、3 帧	紧邻的前 5 帧
运动缓慢	news	88.6	11.4	0.0	92.7	7.3	0.0
	silent	91.7	8.3	0.0	93.8	6.2	0.0
运动一般	foreman	2.1	96.9	1.0	4.2	94.8	1.0
	waterfall	1.0	98.0	1.0	2.2	96.8	1.0
运动剧烈	mobile	0.0	9.3	90.7	0.0	9.3	90.7
	tempete	0.0	6.3	93.7	0.0	6.3	93.7

由表 2 中的数据可知,对于运动平缓的视频序列,本文直接将紧邻当前帧的前 1 帧作为最佳参考帧的概率平均超过 90%;将紧邻当前帧的 3 帧作为候选最佳参考帧的概率约在 10%;第 4、5 帧作为最佳参考帧的比例极小,近乎为 0。对于运动一般的视频序列,候选最佳参考帧为前 3 帧的概率约为 99%;第 4、5 帧作为候选最佳参考帧的概率较校验模型中全搜索方法得到的结果小。对于运动剧烈的视频序列,直接将紧邻当前帧的第 1 帧作为最佳参考帧的比例极小,近乎为 0;将前 3 帧作为最佳参考帧的比例也小于 10%。上述统计数据与校验模型中全搜索方法得到的结果基本吻合,本文所给出的候选最佳参考帧确定算法可行,结果有效且符合视频序列的运动特性。

## 2.2 宏块编码模式的自适应确定

H. 264 中定义了 8 种帧间预测宏块模式(Skip、 $16 \times 16$ 、 $16 \times 8$ 、 $8 \times 16$ 、 $8 \times 8$ 、 $4 \times 8$ 、 $8 \times 4$ 、 $4 \times 4$ )和 2 种帧内模式(Intra

在表 1 所列的校验模型中,全搜索方法得到的最佳参考帧分布数据与视频序列帧间具有时间强相关这一特性相吻合,运动缓慢的视频序列中,最佳参考帧为其紧邻前一帧的概率约为 96%,其他 4 帧仅占 4%;运动一般的视频序列中,最佳参考帧为其紧邻前一帧的概率约为 70.4%,前 3 帧的概率为 93.7%,第 4、5 帧仅占 6.3%;而在运动剧烈的视频序列中,最佳参考帧为其紧邻前一帧的概率仅为 50.3%,前 3 帧的概率也仅为 77.5%,且后 4 帧所占概率相对比较平均。

由此,本文参考帧数由紧邻当前帧的已编码前 5 帧中最佳参考帧为其紧邻第 1 帧的宏块的比例确定,该比例定义为:

$$P = \frac{num(ref_0)}{num(ref_0) + num(ref_1) + num(ref_2) + num(ref_3) + num(ref_4)} \quad (2)$$

式中,  $num(ref_i)$  为最佳参考帧为第  $i$  帧的加权宏块数,因为离当前帧愈近的帧,一般与当前帧的相关性愈强,所以其相应的权值愈大。

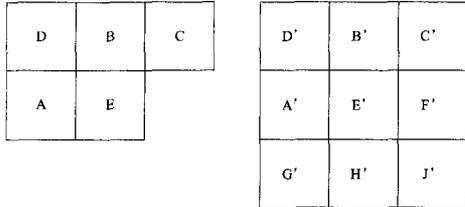
$$num(ref_i) = 0.5num_{P_0}(ref_i) + 0.25num_{P_1}(ref_i) + 0.15num_{P_2}(ref_i) + 0.06num_{P_3}(ref_i) + 0.04num_{P_4}(ref_i)$$

其中,  $num_{P_0}(ref_i)$ 、 $num_{P_1}(ref_i)$ 、 $num_{P_2}(ref_i)$ 、 $num_{P_3}(ref_i)$ 、 $num_{P_4}(ref_i)$  分别为紧邻当前帧的已编码前 5 帧中最佳参考帧为第  $i$  帧的宏块数目。若  $P > T_1$ ,则紧邻当前帧的前 1 帧为当前宏块的最佳参考帧;若  $T_2 < P < T_1$ ,则将紧邻当前帧的前 3 帧确定为候选最佳参考帧;若  $P < T_2$ ,则将  $T_1$  紧邻当前帧的前 5 帧都作为候选最佳参考帧。根据表 1 的统计数据,将设  $T_2$  定为 0.95 或 0.9,将设定为 0.5。表 2 为不同运动程度视频序列根据本文参考帧确定算法确定的候选最佳参考帧的分布情况。

$16 \times 16$ 、Intra  $4 \times 4$ )。若某宏块的最佳划分为  $16 \times 16$ ,最佳参考帧为当前帧的前 1 帧,通过运动搜索得到的最佳运动矢量与预测的运动矢量相同,残差经过变换量化后均为 0,则该宏块被定义为 Skip 编码模式。帧间大尺寸划分(Skip、 $16 \times 16$ 、 $16 \times 8$ 、 $8 \times 16$ )适用于编码平滑区域或者运动量小的区域;小尺寸划分( $8 \times 8$ 、 $8 \times 4$ 、 $4 \times 8$ 、 $4 \times 4$ )主要用于编码具有丰富细节或剧烈运动的区域。采用 Intra 编码模式主要是为了有效地阻止视频质量的下降,对于一般视频,宏块采用 Intra 编码模式的比例很小,甚至可以忽略不计;而对于运行剧烈、运动方式较为复杂的视频,采用 Intra 模式的宏块的比例会上升,但一般也仅占 10%左右<sup>[10]</sup>,故本文只采用帧间宏块模式。H. 264 采用了可变速运动估计和率失真优化模式判决,即编码时对每一种宏块模式都进行运动搜索,并利用率失真优化模式,这极大地增加了编码器的复杂度,很难实现实时编码。

根据视频序列的内容特性可知,视频序列相邻宏块间具有较强的时空相关性,其编码模式也有一定的相关性。即若当前宏块的相邻宏块采用大划分尺寸,则说明该区域较平滑或运动缓慢,当前宏块很可能也采用大的划分尺寸;同理,若当前宏块的相邻宏块采用小划分尺寸,则说明该区域图像纹理丰富或运动剧烈,当前宏块一般也应采用小的划分尺寸。

在当前帧中定义 5 个  $16 \times 16$  宏块,设 E 为当前宏块,A、B、C、D 分别表示当前帧当前宏块的左、上、右上和左上已编码的  $16 \times 16$  宏块,如图 1(a)所示;当前帧的前 5 帧定义 9 个  $16 \times 16$  宏块,其中 E' 的位置与当前宏块在帧中的坐标相同,如图 1(b)所示。



(a) 当前帧中宏块划分示意图 (b) 紧邻当前帧前 5 帧宏块划分示意图  
图 1

设定紧邻当前帧的前 5 帧中 45 个宏块和当前帧中的 A、B、C、D 4 个宏块都为当前宏块的关联宏块,采用校验模型全搜索方法计算、统计 49 个关联宏块中宏块的模式与当前宏块模式之间的关系。首先确定当前宏块模式及其最佳参考帧,然后统计当前宏块的 49 个关联宏块中宏块模式的加权个数,其中权值定义为:当前帧宏块权值为 6,紧邻的第 1 帧中与当前帧间坐标一致的宏块权值为 6,其他宏块权值为 5,第 2、3、4、5 帧中宏块的权值以此类推。对于不同运动状态的视频序列,统计所有宏块的宏块模式与其关联的 49 个宏块加权个数最多的宏块模式相同的比例,具体数据如表 3 所列。

表 3 不同运动程度视频中宏块模式与其关联宏块中加权个数最多的宏块模式关系

序列	模式相同的比例 %
运动缓慢	news 86.79
	silent 80.09
运动一般	foreman 65.80
	waterfall 69.63
运动剧烈	mobile 49.59
	tempete 52.13

由表 3 数据可知,当前宏块模式与其关联宏块的模式有较强的相关性,特别是对于运动缓慢的视频,这与宏块之间本身所具有的时、空间相关性相符。但是对于运动剧烈的视频来讲,相同的比例降低。若仅根据上述相关性把当前宏块的模式确定为其关联宏块中加权个数最多的宏块模式,则会有较大误差,从而影响视频的编码质量。虽然目前也有研究者利用宏块间的时空相关性来快速确定宏块的划分模式,但由表 3 统计的数据可知会存在较大误差,降低了视频编码质量<sup>[10,11]</sup>。本文在利用宏块时空相关性进行初步划分,然后通过自适应的方法快速确定宏块的划分模式及其对应的最佳参考帧。

本文宏块模式确定算法首先设定宏块模式与关联宏块加权个数最多的宏块模式相同,然后计算当前宏块在候选参考帧中的最小代价和所确定的最佳参考帧。设选定模式为  $16 \times 16$ ,若  $J_{mv} \leq \min(J_A, J_B, J_C, J_D, J_E)$ ,  $J_A, J_B, J_C, J_D, J_E$

分别为 A、B、C、D、E' 宏块的代价,则判定该宏块为 Skip 编码模式。否则,若  $J_{mv} \leq \max(J_A, J_B, J_C, J_D, J_E)$ ,则判定该宏块为设定的划分模式;若  $J_{mv} > \max(J_A, J_B, J_C, J_D, J_E)$ ,则计算当前宏块的其他 6 种划分下的宏块的代价及其相应的最佳参考帧,并将其中代价最小的划分及其对应的参考帧作为当前宏块的划分模式和最佳参考帧。

在对视频进行编码时,由于本文算法中应用到帧间、帧内宏块间的时空相关程度,故对视频的第 1 帧以及第 2、3、4、5 帧,每一帧的第一行、第一列以及最后一列宏块进行编码时不能使用本文算法,上述区域的宏块模式选择及其最佳参考帧通过校验模型全搜索方法确定,对于其他帧及帧内其他宏块本文算法适用。算法具体流程可概括为图 2 所示。

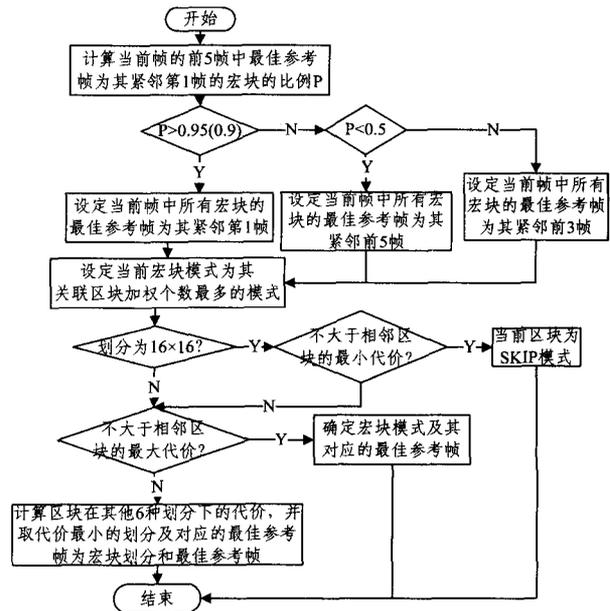


图 2 本文算法流程图

### 3 实验结果与分析

利用本文给出的快速算法对 6 种运动程度不同的  $352 \times 288$  视频序列进行了测试,每个序列为 100 帧。本文算法确定的宏块模式及其最佳参考帧与校验模型全搜索方法确定的结果如表 4 所列。

表 4 本文算法确定的宏块模式、最佳参考帧与校验模型全搜索方法的对比

序列	宏块模式不同 %	最佳参考帧不同 %	宏块模式不同 %	最佳参考帧不同 %	
					T1=0.9
运动缓慢	news	4.84	3.92	4.81	3.89
	silent	5.57	4.00	5.55	3.98
运动一般	foreman	7.71	4.47	7.69	4.46
	waterfall	7.90	4.51	7.89	4.49
运动剧烈	mobile	11.86	4.60	11.86	4.60
	tempete	10.00	4.54	10.00	4.54

由表 4 可知,本文算法与校验模型全搜索方法在宏块模式及其最佳参考帧上仍有一定的差距,特别是对于运动剧烈的视频序列,虽然超过 90% 的帧中都选用 5 帧作为候选最佳参考帧,但是由于宏块模式选择上的差异,造成最佳参考帧之间有约 4.5% 的差异。但总的来说,无论宏块模式还是最佳参考帧的差异都不太大,平均约在 10% 以内。

此外,还对上述6种不同视频进行实际编码,编码结构采用IPPP,熵编码采用CABAC,运动搜索范围为±16,最大参考帧数为5,量化参数为28,采用率失真优化方法,运动估计采用全搜索方法。编码结果与校验模型JM16.2的算法编码结果进行比较,比较指标有 $\Delta PSNR$ 、 $\Delta SSIM$ 、 $\Delta BR$ 及 $\Delta time$ 。

PSNR基于图像像素灰度值进行统计和平均计算,主要用来衡量信号的失真程度。结构相似法(SSIM)是一种基于结构信息衡量原始信号与处理后信号之间相似程度的方法。由于结构信息的改变与感知图像质量的变化非常接近,即若结构相似,则认为前后图像质量变化不大,也就是质量损失不大,结构相似法是目前常用的、与感知图像失真非常接近的一

种客观评价方法<sup>[12,13]</sup>。上述2个指标主要用来描述视频编码质量,而编码码率BR和编码时间time主要用来衡量算法的效率。表5给出了本文的改进算法与校验模型JM16.2的算法的比较数据。 $\Delta PSNR$ 、 $\Delta SSIM$ 、 $\Delta BR$ 及 $\Delta time$ 具体计算方法如下:

$$\Delta PSNR = PSNR[\text{校验模型}] - PSNR[\text{本文算法}]$$

$$\Delta SSIM = SSIM[\text{校验模型}] - SSIM[\text{本文算法}]$$

$$\Delta BR = \frac{BR[\text{校验模型}] - BR[\text{本文算法}]}{BR[\text{校验模型}]} \times 100\%$$

$$\Delta time = \frac{time[\text{校验模型}] - time[\text{本文算法}]}{time[\text{校验模型}]} \times 100\%$$

表5 本文算法与校验模型JM16.2的算法的比较

序列		$\Delta PSNR/dB$	$\Delta SSIM$	$\Delta BR/\%$	$\Delta time/\%$	$\Delta PSNR/dB$	$\Delta SSIM$	$\Delta BR/\%$	$\Delta time/\%$
		T1=0.9	T1=0.9	T1=0.9	T1=0.9	T1=0.95	T1=0.95	T1=0.95	T1=0.95
运动	news	0.018	0.0000	-0.35	79.8	0.010	0.0000	-0.32	75.4
缓慢	silent	0.040	0.0001	-0.60	81.7	0.039	0.0001	-0.57	79.4
运动	foreman	0.082	0.0003	-1.88	61.7	0.081	0.0003	-1.85	60.9
一般	waterfall	0.091	0.0003	-2.13	62.1	0.090	0.0002	-2.12	61.5
运动	mobile	0.170	0.0004	-2.56	40.3	0.170	0.0004	-2.56	40.3
剧烈	tempete	0.158	0.0004	-2.30	40.7	0.158	0.0004	-2.30	40.7
平均		0.093	0.0003	-1.63	61.1	0.091	0.0002	-1.62	59.7

由表5可以看出,本文提出的算法与校验模型相比,在编码质量基本不变的情况下,平均编码时间可以节省约60%。对于运动缓慢的视频序列,本文算法与校验模型全搜索方法相比, $\Delta SSIM$ 、 $\Delta PSNR$ 几乎不变, $\Delta BR$ 约有0.5%的增加,但是编码时间约有75%的节省。对于运动剧烈的视频序列, $\Delta SSIM$ 变化可以忽略不计, $\Delta PSNR$ 略有增加, $\Delta BR$ 约有2.5%的增加,编码时间减少40%以上。总的来讲,本文所给出的算法对视频序列的编码效率上有较大的提高,平均节省约60%的编码时间,同时视频质量几乎保持不变,只是编码码率稍有增加。上述结果与本文算法的假设基本吻合,即视频序列的时空相关性越强,本文算法的效果越好。

**结束语** 针对H.264中多参考帧技术为编码器带来的大量计算复杂度问题,本文基于视频序列帧间具有较强时间相关性、宏块间具有较强时空相关性,提出了一种多参考帧选择算法。该算法根据视频序列的帧间时间相关程度来确定参考帧个数;由相邻宏块的时空相关性自适应地确定当前宏块的编码模式及其最佳参考帧。实验结果表明,与校验模型JM16.2相比,该算法在编码质量几乎不变的情况下平均可节约60%的编码时间。本文算法略微增加了编码的码率,虽然提高了编码的实时性,但是增加了视频传输的带宽。为了降低视频采集、传输的要求,需要在保证视频质量的基础上,进一步研究提高编码实时性、降低编码码率的算法。

### 参考文献

[1] Wiegand T, Sullivan G, Bjontegaard G, et al. Overview of the H. 264/AVC video coding standard[J]. IEEE Transactions on Circuit and Systems for Video Technology, 2003, 13(7): 560-576

[2] Wiegand T. Draft ITU-T recommendation and final draft international standard of Joint Video Specification[S]. JVT-G050.

2003;1-253

[3] Chang J Y, Kim W J, Bae Y H, et al. Performance analysis for MPEG-4 video codec based on-chip network[J]. ETRI Journal, 2005, 27(5): 497-503

[4] Lee K J, Jeon G, Falcon R. An adaptive fast multiple reference frame selection algorithm for H. 264/AVC using reference region data[C]//IEEE Workshop on Signal Processing Systems. 2009:93-96

[5] Shen L Q, Liu Z, Zhang Z Y, et al. An adaptive and fast multi-frame selection algorithm for H. 264 video coding[J]. IEEE Signal Processing Letters, 2007, 11(14): 836-839

[6] 张淑芬, 李华. 基于H.264的多参考帧快速选择算法[J]. 电子学报, 2009, 37(1): 62-66

[7] Kuo T Y, Lu H J. Efficient reference frame selector for H. 264[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2008, 18(3): 400-405

[8] Sohn J S, Kim D G. Fast multiple reference frame selection method using correlation of sequence in JVT/H. 264[J]. IEICE Transactions on Fundamentals of Electronics, 2006, E89-A(3): 744-746

[9] 徐静, 周兵, 黄雪莉, 等. 基于空域特征的H.264快速多参考帧选择算法[J]. 通信学报, 2010, 31(7): 40-45

[10] 朱红, 吴成柯, 方勇. H.264宏块模式的一种快速判决方案[J]. 电子学报, 2005, 33(9): 1576-1580

[11] 崔晓磊. H.264/AVC中SVC宏块模式选择快速算法[D]. 西安:西安电子科技大学, 2010

[12] 佟雨兵, 胡薇薇, 杨东凯, 等. 视频质量评价方法综述[J]. 计算机辅助设计及图形学学报, 2006, 18(5): 735-741

[13] 佟雨兵, 张其善, 祁云平. 基于PSNR与SSIM联合的图像质量评价模型[J]. 中国图象图形学报, 2006, 11(12): 1758-1763